

BOSTON UNIVERSITY
GRADUATE SCHOOL OF ARTS AND SCIENCES

Dissertation

**NEUROIMAGING AND COMPUTATIONAL MODELING OF
SYLLABLE SEQUENCE PRODUCTION**

by

JASON W. BOHLAND

B.S., University of Cincinnati, 1998
M.S., University of Cincinnati, 2000

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

2007

© Copyright by
JASON W. BOHLAND
2007

Approved by

First Reader

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

Second Reader

Daniel H. Bullock, PhD
Associate Professor of Cognitive and Neural Systems
and Psychology

Third Reader

Alec Marantz, PhD
Professor of Psychology and Linguistics, New York University

Acknowledgments

The mental trauma inherent to writing a dissertation was tempered by patience, wisdom, and friendship from my advisor, my dissertation committee, my labmates, my friends, and my family. Thank you for so much support.

Financial support was provided from the National Institute on Deafness and other Communication Disorders (R01 DC002852 and R01 DC007683, F. Guenther, PI) and the National Science Foundation (NSF SBE-0354378, S. Grossberg, PI). Functional magnetic resonance imaging at the Athinoula A. Martinos Center for Biomedical Imaging was made possible by grants from the National Center for Research Resources (P41RR14075) and the MIND institute.

SYLLABLE SEQUENCE PRODUCTION

(Order No.)

JASON W. BOHLAND

Boston University Graduate School of Arts and Sciences, 2007

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

Abstract

Fluent speech involves producing sound sequences that are composed from a finite alphabet of learned words, syllables, and phonemes. The brain thus requires machinery to organize and enact properly ordered and timed motor command sequences that correspond to the desired phonological plan. This dissertation seeks to provide an enhanced mechanistic understanding of this system through a combination of computational neural modeling and neuroimaging.

The first portion of the dissertation describes an experiment using sparse event-triggered functional magnetic resonance imaging (fMRI) to measure brain responses due to preparation and overt production of non-lexical three syllable sequences of varying complexity. The network of brain regions related to initiation, motor execution and hearing one's own voice was found to include the primary motor and somatosensory cortices, auditory cortices, supplementary motor area (SMA), insula, and portions of the thalamus, basal ganglia, and cerebellum. Additional stimulus complexity led to increased engagement of the basic speech network and recruitment of additional areas known to be involved in control of non-speech motor sequences, including the left hemisphere inferior frontal sulcus region and posterior parietal cortex, and bilateral regions at the junction of the anterior insula and frontal operculum.

the pre-SMA, basal ganglia, anterior thalamus, and cerebellum.

These experimental results as well as previous clinical, behavioral, and imaging data were used to guide the development of a neural model of speech syllable sequencing based on a "competitive queuing" architecture. The new GODIVA (Gradient Order DIVA) model extends the DIVA model of speech production, which describes how individual speech items are learned and produced, to include explicit parallel representations for forthcoming utterances. GODIVA posits detailed neuroanatomical substrates and neurobiologically plausible mechanisms for its components. The model can thus account for a database of clinical and neuroimaging results beyond the scope of previous non-biological models.

Finally, preliminary efforts using magnetoencephalography (MEG) and surface electromyography (EMG) to obtain neuroimaging data that complements fMRI results and offers further modeling constraints are described. A novel algorithm was applied to detect neural source components that could be used to reliably discriminate between stimuli that necessitated the preparation of one, two, or three syllable plans.

CONTENTS

1	Introduction	1
1.1	Functional neuroimaging	2
1.2	Neural modeling	4
1.3	Organization of dissertation	4
2	An fMRI investigation of syllable sequence production	7
2.1	Introduction	7
2.2	Materials and Methods	11
2.2.1	Subjects	11
2.2.2	Experimental Protocol	11
2.2.3	Data Acquisition	14
2.2.4	Data Analysis	15
2.3	Results	18
2.3.1	Acoustic analysis	18
2.3.2	Basic speech production network	18
2.3.3	Main effect of overt production	24
2.3.4	Main effect of sequence complexity	24
2.3.5	Main effect of syllable complexity	27
2.3.6	Interactions between factors	32
2.4	Discussion	35
2.4.1	Sensorimotor areas	42
2.4.2	Left hemisphere prefrontal areas	43

2.4.3	Anterior insula and frontal operculum	45
2.4.4	Temporal and parietal areas	47
2.4.5	Medial premotor areas	48
2.4.6	Cerebellum	51
2.4.7	Basal ganglia and thalamus	53
2.4.8	Sequencing and the FOXP2 gene	54
2.4.9	Conclusions	56
3	A computational neural model of speech sequence planning and production	59
3.1	Introduction	59
3.2	Models of serial behavior	62
3.2.1	Associative chaining	63
3.2.2	Positional coding	64
3.2.3	Parallel models of serial performance: Competitive queuing . .	65
3.3	Linguistic models	68
3.3.1	Syllables, phonemes, and features	68
3.3.2	Factorization of structure and content	70
3.3.3	Speech motor and phonological development	71
3.3.4	The WEAVER / WEAVER++ model	73
3.3.5	Other related models	74
3.4	Constraints on linguistic models	78
3.4.1	Speech error studies	78
3.4.2	Reaction time studies	82
3.4.3	Clinical studies	84
3.5	Neuroanatomical and neurophysiological modeling constraints	86
3.5.1	Neurophysiology of prefrontal cortical cells	86

3.5.2	SMA and pre-SMA	87
3.5.3	Cortico-striato-pallido-thalamo-cortical loops	90
3.6	The DIVA model of speech production	93
3.6.1	Speech sound map	94
3.6.2	Feedforward control system	96
3.6.3	Feedback control system	96
3.6.4	Limitations of the DIVA model	98
3.7	The Gradient Order DIVA (GODIVA) model	98
3.7.1	Computational methods	98
3.7.2	Functional overview of the model	99
3.8	Model specification	103
3.8.1	Phonological content representation in inferior frontal sulcus .	104
3.8.2	Structural “frame” representations in pre-SMA	112
3.8.3	Cortico-striato-pallido-thalamo-cortical “planning loop”	117
3.8.4	Speech sound map	120
3.8.5	Response release via the “motor loop”	123
3.9	Simulation results	125
3.9.1	Performance of a sequence of well-learned syllables	125
3.9.2	Performance from sub-syllabic targets	129
3.10	Discussion	131
3.10.1	The GODIVA model	131
3.10.2	Representations for serial order	133
3.10.3	Repeating elements	136
3.10.4	A general framework	137
3.10.5	Future extensions to GODIVA	138

4	Examining syllable sequence production using magnetoencephalography	144
4.1	Introduction to magnetoencephalography	145
4.1.1	Temporal components of MEG / EEG	147
4.1.2	Spectral components	148
4.2	Previous speech production studies using MEG / EEG	148
4.2.1	Artifacts due to speech-related movements	151
4.2.2	Source estimation methods	153
4.3	Materials and methods	154
4.3.1	Subject	154
4.3.2	Experimental protocol	155
4.3.3	Visual baseline stimuli	155
4.3.4	Syllable sequence stimuli	155
4.3.5	Data acquisition	157
4.3.6	EMG signal analysis	159
4.3.7	MEG signal preprocessing	159
4.3.8	Head model	160
4.4	A novel method for single-trial MEG analysis	162
4.4.1	Statistical tests	165
4.5	Results	167
4.5.1	Visual evoked fields	167
4.5.2	Syllable sequence production	169
4.6	Discussion	172
4.6.1	Visual evoked fields	173
4.6.2	Speech production	174

5 Conclusion	179
5.1 Summary of contributions	179
5.2 Future Directions	181
References	185
Curriculum Vitae	217

LIST OF TABLES

2.1	Acoustic production durations by condition	18
2.2	Activation peak summary for GO vs. baseline	21
2.4	Activation peak summary for GO vs. NOGO	25
2.5	Activation peak summary for main effect of <i>seq</i>	29
2.6	Activation peak summary for main effect of <i>syl</i>	30
2.7	Activation peak summary for <i>seq</i> \times <i>syl</i> interaction	34
3.1	Legend of symbols used in the GODIVA model	104
3.2	An algorithmic summary of GODIVA model operation	126
3.3	Summary of parameters used in the GODIVA model	131

LIST OF FIGURES

2·1	Design of syllable sequence stimuli	12
2·2	Time course of a single fMRI trial	14
2·3	Minimal speech production network	20
2·4	Main effect of overt production	26
2·5	Main effect of sequence complexity	28
2·6	Main effect of syllable complexity	31
2·7	Interactions between sequence and syllable complexity	33
2·8	Effects of sequence complexity by syllable type	36
2·9	Renderings of morphometric abnormalities in the 'KE' family	57
3·1	Outline of the "Nijmegen Model"	61
3·2	Associative chain model schematic	64
3·3	Positional coding model schematic	66
3·4	Basic competitive queuing architecture	67
3·5	Averbeck et al. (2002) recordings from prefrontal cortex	88
3·6	The DIVA model of speech production	95
3·7	Overview of the GODIVA model	100
3·8	Illustration of model IFS fields	106
3·9	Illustration of the IFS map planning layer	107
3·10	Illustration of pre-SMA structure and function in GODIVA	115
3·11	Illustration of basal ganglia planning loop	118
3·12	Simulation of the three syllable sequence " <i>gəv.di.və</i> "	127

3.13	Simulation of “ <i>gəv.di.və</i> ” from phonemic targets	132
4.1	Visual baseline checkerboard stimuli	156
4.2	Coregistration of coordinate frames for MEG analysis	161
4.3	Schematic depiction of the MEG analysis algorithm	166
4.4	Results of visual evoked fields analysis	168
4.5	Reaction times by condition estimated from EMG	169
4.6	Mean magnitude frequency response in time period of interest	170
4.7	Results of speech production analysis	171

LIST OF ABBREVIATIONS

AOS	Apraxia of Speech
BA	Brodmann's Area
BG	Basal Ganglia
BOLD	Blood Oxygenation Level Dependent
CALM	Continuously Adjusted Least-Squares Method
CDF	Cumulative Distribution Function
CQ	Competitive Queuing
CV	Consonant Vowel
DFT	Discrete Fourier Transform
DIVA	Directions Into Velocities of Articulators
DTI	Diffusion Tensor Imaging
EEG	Electroencephalography
EMG	Electromyography
EMMA	Electromagnetic Mid-Sagittal Articulometry
EOG	Electrooculogram
ERF	Event-related field
ERP	Event-related potential
FDR	False Discovery Rate
FIR	Finite Impulse Response
fMRI	Functional Magnetic Resonance Imaging
FO	Frontal Operculum

FWE	Family-wise Error
FWHM	Full-Width at Half Maximum
GABA	Gamma-aminobutyric acid
GODIVA	Gradient Order DIVA
GPI	Globus Pallidus Internal Segment
IFS	Inferior Frontal Sulcus
ISI	Inter-stimulus interval
K-S	Kolmogorov-Smirnov Test
LED	Light-emitting Diode
MEG	Magnetoencephalography
MNI	Montreal Neurological Institute
MR	Magnetic Resonance
MRI	Magnetic Resonance Imaging
MSN	Medium Spiny projection Neuron
MSR	Magnetically Shielded Room
PET	Positron Emission Tomography
ROI	Region of Interest
RT	Reaction Time
SMA	Supplementary Motor Area
SnPM	Statistical Non-parametric Mapping
SPM	Statistical Parametric Mapping
SSM	Speech Sound Map
SQUID	Superconducting Quantum Interference Device
T	Tesla
TE	Echo Time
TELOS	Telencephalic Laminar Objective Selector

TR	Repetition Time
TTL	Transistor-Transistor Logic
VCA	Vertical line transversing Anterior Commissure
VEF	Visual Evoked Field
VEP	Visual Evoked Potential

CHAPTER 1

INTRODUCTION

This dissertation describes three distinct but complementary investigations, each having the ultimate goal of achieving a more comprehensive understanding of the neural processes that underlie the preparation and production of syllable sequences. Fluent speech production requires phonemes and syllables to be arranged sequentially to form a coherent articulatory plan. It is in the consideration of this type of problem where “high-level” studies of speech planning and, more generally, language formulation, intersect with “low-level” theories of motor control for speech articulation. While each of these sub-fields has been studied in some detail, they have remained largely distinct. Furthermore, the vast majority of theoretical models developed to describe speech and language at either level have not addressed the underlying *neural* machinery that is ultimately responsible for the behaviors under examination.

The approach taken in this dissertation places particular emphasis on the neural substrates responsible for planning and producing speech sequences. The problem of serial order is of principle importance, and it is hypothesized (following Lashley, 1951) that speakers plan syllable sequences in parallel in a phonological space prior to the selection and initiation of corresponding sensorimotor programs. The combination of functional magnetic resonance imaging (fMRI), magnetoencephalography (MEG), and computational neural modeling is used to examine the questions of *how* such sequences can be represented and enacted, and *where* in the brain the relevant

representations and transformations can be found. This approach also seeks to unify multiple datasets and multiple theoretical and computational ideas. The primary result is a formal model that can simulate various aspects of serial speech production, that proposes neural representations for speech codes and their serial order, and that suggests what neural computations are performed during these behaviors. This model, itself informed by functional imaging results, can furthermore be used to generate experimental predictions to be tested by the application of these same experimental techniques.

1.1 Functional neuroimaging

The development of non-invasive technologies for measuring human brain function such as positron emission tomography (PET), functional magnetic resonance imaging (fMRI), electroencephalography (EEG), and magnetoencephalography (MEG) has led to a vast increase in the quantity and quality of neurological data available to the research community. Because speech is a behavior restricted only to humans (although modest parallels might be drawn to other animals such as the songbird; e.g. Doupe and Kuhl, 1999), single- or multi-unit neurophysiological recordings are only available in rare circumstances, for example in Parkinson’s Disease patients undergoing surgical implantation of stimulation units (e.g. Watson and Montgomery, 2006). The relative paucity of such direct measurements, which have been frequently used in the examination of other neural systems, means that neuroimaging methods are of critical importance in the study of speech and language.

The experimental portion of the research described in this dissertation makes use of fMRI and MEG methods. Blood oxygenation level dependent (BOLD) fMRI (Ogawa et al., 1992; Kwong et al., 1992) provides an indirect measure of neural activity during the performance of a task. The BOLD signal is based on relative

proportions of oxygenated and deoxygenated hemoglobin in blood vessels in the brain. These proportions are related to local neuronal activity because consumption of metabolic resources by active neurons leads to a local increase in blood flow to that region (Roy and Sherrington, 1890). This *hemodynamic response* delivers an oversupply of oxygenated blood (Fox and Raichle, 1986), resulting in a net decrease in the paramagnetic agent deoxyhemoglobin, leading to the net *increase* in BOLD contrast typically observed during task performance (relative to a rest condition) in fMRI experiments.

Functional MRI methods are able to deliver measurements with high spatial resolution relative to other brain imaging techniques. On the other hand, because the BOLD signal is based on the relatively slow hemodynamic response, the method can not offer particularly fine temporal resolution. Magnetoencephalographic methods exhibit the opposite resolution profile: high temporal but limited spatial resolution. MEG measures magnetic fields outside the skull produced by synchronized neuronal currents flowing within pyramidal cells in the cortex. Because magnetic fields are instantaneously related to current densities (by Maxwell's equations), the inherent temporal resolution of the technique is limited only by the measurement devices themselves. In practice, the sampling rate of MEG data can be above 1 kHz, and thus such data provide the opportunity to examine fine temporal and / or spectral characteristics of cortical responses during different tasks. The ability to localize the neural sources responsible for such responses, however, is limited by the ill-posed nature of the *MEG inverse problem*. Because of the reciprocal space-time resolution profiles of these two techniques, it appears to be advantageous, in terms of understanding neural mechanisms, to collect measurements using both modalities. The examination of brain responses during *overt* speaking tasks, however, raises methodological difficulties with either technology; such potential problems are specifically

addressed in the design of the experiments herein.

1.2 Neural modeling

The explosion of functional brain imaging studies in recent years has provided important data points to researchers in the speech and neuroscience communities, but these data in isolation are not sufficient to describe a complex neural system like the one responsible for planning and producing speech. A better understanding of the precise nature of neural representations and computations in a particular system can be achieved through the development of *computational neural models* whose components mimic the activity of neurons or groups of neurons in individual brain regions. To be successful, such models must assume the constraints from the known neurophysiology of particular brain regions, and from the known connectivity between these regions. Such models can provide a coherent framework within which to explore neural processes and to interpret experimental observations.

This approach has led to the previous development of such a neural model, the DIVA (Directions Into Velocities of Articulators) model of speech production (Guenther, 1994, 1995; Guenther et al., 1998, 2006), which describes speech motor control and acquisition. The modeling work presented here extends the DIVA model to allow for explicit parallel planning of multiple speech sounds prior to their production. In so doing, the extended model draws heavily on previous theoretical work in the general study of sequence memory and recall, which has led to the establishment of *competitive queuing* (Grossberg, 1978a,b; Houghton, 1990; Bullock and Rhodes, 2003) as a biologically plausible neural architecture for representing the order and identity of items to be recalled sequentially.

1.3 Organization of dissertation

The remainder of this dissertation is organized into three chapters containing the body of the research, followed by a chapter that summarizes the present contributions and identifies possible future directions for related research. Each of Chapters 2-4 includes a review of pertinent data and previous theories, models, or methods. There is inevitably some degree of repetition in these discussions across the three main chapters.

Chapter 2 describes an experimental study of syllable sequence production that was performed using fMRI. This study reveals how different cortical and subcortical brain regions respond to added complexity in simple non-lexical speech sequences during both preparation only and overt production conditions. The results are discussed in the broad context of the previous relevant experimental and clinical results for each region of interest in this study, and mechanistic interpretations of the various observed response profiles are explored.

Chapter 3 presents the development of a biologically-plausible computational neural model of syllable sequence planning and production. This model embeds various computational proposals, with an emphasis on competitive queuing, into a realistic and well-specified architecture with particular modules determined on the basis of the fMRI experiment described above, as well as previous findings. This model is formally specified by a set of equations, and simulations show that it is capable of representing and “reading out” arbitrary syllable sequences. The model interfaces with the current DIVA model of speech production by selecting and activating appropriate stored sensorimotor programs in the appropriate order.

Chapter 4 explores the use of magnetoencephalography as a tool in the study of speech sequence planning and production. Measuring MEG in overt speech production tasks is problematic due to potential contamination of the measured signal by

muscle related artifacts. By measuring surface EMG simultaneously with MEG, it is shown, using a novel analysis method, that the time period just prior to activation of the lip muscles (and to the onset of articulation) contains components that can be used to differentiate between speaking conditions in which one, two, or three syllable utterances were planned. While this analysis is preliminary and more data are needed, indications are that MEG can provide additional information regarding syllable sequence representations that could be used in developing the modeling work further.

Finally, in Chapter 5, the major contributions of this work are summarized, and further research is proposed to develop a more comprehensive understanding of the organization of sequences of speech sounds and of speech production processes in general.

CHAPTER 2

AN FMRI INVESTIGATION OF SYLLABLE SEQUENCE PRODUCTION

2.1 Introduction

Fluent speech requires a robust serial ordering mechanism to combine a finite set of discrete learned phonological units (such as phonemes or syllables) into larger meaningful expressions of words and sentences. Lashley (1951) posed the problem of serial order in behavior, asking how the brain organizes and executes smooth, temporally integrated behaviors such as speech and rhythmic motor control. His proposal for the “priming of expressive units,” or parallel, co-temporal activation of the items in a behavioral sequence prior to execution, has been supported in studies of speech production by bountiful data related to linguistic performance errors (e.g. MacKay, 1970; Fromkin, 1980; Gordon and Meyer, 1987), by reaction time experiments (e.g. Klapp, 2003), and by the demonstration of anticipatory and perseveratory coarticulation (e.g. Ohman, 1966; Hardcastle and Hewlett, 1999).

The problem of serial order in speech production can be considered at multiple levels. Phonemes, for example, might be manipulated to form syllables and words, where each phonemic token is learned and stored with corresponding auditory and/or orosensory consequences (see, for example, the DIVA model of speech production; Guenther, 1995; Guenther et al., 1998, 2006, which provides a computational account for how such tokens can be learned and produced). Various researchers have

suggested, on the basis of reaction time data, that syllable- or word-sized tokens can be learned such that they may be efficiently executed as single motor *chunks*, forming a *mental syllabary* (Levelt and Wheeldon, 1994; Levelt et al., 1999b; Cholin et al., 2006); these larger chunks might then serve as manipulable tokens for speech sequence planning.

In addition to organizing sequences of planned sounds within a memory buffer, speech production requires a mechanism to *initiate* or release items to the motor apparatus at precise times. Speakers can typically produce up to six to nine syllables (20 to 30 segments) per second, which is faster than any other form of discrete motor behavior (Kent, 2000). A system that coordinates the timed release of each discrete item in the planned sequence of speech is, therefore, of critical importance to fluent performance.

While the formulation of spoken language plans has been widely studied at a conceptual level (see e.g. Levelt, 1989; Levelt et al., 1999b), relatively little is known about the neural representations of those plans or about the cortical and subcortical machinery that guides the serial production of speech. Clinical studies have suggested that damage to the anterior insula or neighboring inferior frontal areas (Dronkers, 1996; Hillis et al., 2005; Tanji et al., 2001), supplementary motor area (Jonas, 1981, 1987; Ziegler et al., 1997; Pai, 1999), basal ganglia (Pickett et al., 1998; Ho et al., 1998), or cerebellum (Riva, 1998; Silveri et al., 1998) may lead to deficits in sequencing and/or initiation of speech plans. Such deficits appear in various aphasias, apraxia of speech (AOS), and stuttering. Literal or phonemic paraphasias, in which “well-formed sounds or syllables are substituted or transposed in an otherwise recognizable target word” (Goodglass, 1993), are observed in many cases of aphasia, including conduction aphasia and Broca’s aphasia. AOS, a speech-motor condition¹,

¹Apraxia of speech (AOS) as described by Darley et al. (1975) is a unique syndrome that affects motor speech production without diminished muscle strength. AOS has been associated

has been attributed to damage to the left precentral gyrus of the insula (Dronkers, 1996), as well as the inferior frontal gyrus, subcortical structures, or posterior temporal / parietal regions (Hillis et al., 2005; Peach and Tonkovich, 2004; Duffy, 1995). Ziegler (2002) presents an excellent review of theoretical models of AOS. Though different in many ways, stuttering, which affects approximately 1% of the adult population in the United States, shares with AOS the trait of improper initiation of speech motor programs without impairment of comprehension or damage to the peripheral speech neuromuscular system (Kent, 2000). Stuttering has also been linked to deficits in various phonological memory tasks (Bosshardt, 1993; Ludlow et al., 1997; Hakim and Ratner, 2004; Anderson et al., 2006), suggesting that individuals who stutter may not be able to *represent* speech utterances with the same level of quality as normal subjects.

Only a small portion of the large functional neuroimaging literature related to speech and language has dealt with overt speech production. Within that body, very few studies have explicitly addressed sequencing demands during overt speech. Riecker et al. (2000b) examined brain activations evoked by repetitive production of stimuli of varying complexity: consonant-vowel syllables (CV's), CCCV's, CVCVCV non-word sequences, and CVCVCV words. This study found that production of none of the stimulus types (compared to a resting baseline condition) resulted in significant activations in the SMA or insula; activation was instead largely restricted to the primary sensorimotor areas. Only production of the CCCV stimulus led to significant activation of the cerebellum. Also, production of the multi-syllabic items led to a more limited and lateralized expanse of activation in the banks of the central sulcus than did production of single syllables. These findings seem inconsistent with

with phoneme substitution errors similar to literal paraphasias (e.g. Wertz et al., 1984). The notion of the existence of AOS as a unique disorder, however, has been controversial (see Helm-Estabrooks, 2002) with some clinicians arguing that the condition actually reflects articulatory deficits associated with aphasia (e.g. Goodglass, 1993).

the existing literature (cf. Indefrey and Levelt, 2000), and one motivation for the present study was to clarify how additional complexity in the speech stimulus affects neural activity in the production network.

Shuster and Lemieux (2005) compared production (both overt and covert) of multi-syllabic and mono-syllabic words following the presentation of an auditory exemplar. For the overt speaking condition, additional activation was found in the left inferior parietal lobe, inferior frontal gyrus, and precentral gyrus for multi-syllabic versus mono-syllabic words. Mono-syllabic words resulted in greater activation of the left middle frontal gyrus (BA46). The results for covert speech were somewhat dissimilar; for example, in covert speech there was greater activation of the left middle frontal gyrus for multi-syllable words, and greater activation in the left precentral gyrus for mono-syllable words. The authors emphasize a consistent finding was that multi-syllable words caused additional activation in left inferior parietal areas (BA40), and suggest a role for this region in speech programming. In comparing the results of this study to that of Riecker et al. (2000b) it is difficult to develop a consistent account for the effects of sequential complexity on the speech production system.

The present experiment was designed to clarify how the speech system organizes and produces sequences of speech sounds. While the DIVA model of speech production makes predictions about brain activations in the executive speech motor system (Guenther et al., 2006; Guenther, in press) it does not address brain regions likely to be responsible for sequence planning. Based on clinical observations and studies of other non-speech sequential motor control tasks, it was expected that additional responses to additional stimulus complexity would be observed in a network of brain regions outside of the primary sensorimotor areas (and other regions treated by the DIVA model), including the prefrontal cortex, basal ganglia, anterior insula, supple-

mentary motor area and cerebellum. Blood oxygenation level dependent (BOLD) functional magnetic resonance imaging (fMRI; see Ogawa et al., 1990, 1992; Kwong et al., 1992) was used to measure responses to speech sequences of varying complexity at both the sub- and supra-syllabic levels, and in both preparatory and overt speech production tasks. An “event-triggered” design was employed with both GO and NOGO trials that offered many benefits over previous methods (see Discussion). The results are discussed in terms of the necessary mechanisms for sequencing and initiation in fluent speech production.

2.2 Materials and Methods

2.2.1 Subjects

Thirteen right-handed native English speakers (ages 22-50 years, mean 28.7 years, six females) participated in this study. Written informed consent was obtained according to the Boston University Institutional Review Board and the Massachusetts General Hospital Human Research Committee. No subjects reported a history of any neurological or speech, language, or hearing disorders.

2.2.2 Experimental Protocol

The experimental tasks consisted of preparing to produce (NOGO trials) and overtly producing (GO trials) non-lexical three syllable sequences. The linguistic content of the stimuli was determined by two factors: syllable complexity (*syl*) and sequence complexity (*seq*). Each of these two factors assumed one of two levels (*simple* or *complex*), thereby creating a 2×2 matrix of stimulus types (see Figure 2-1), where stimuli in the same row or column have the same level of *sequence complexity* or *syllable complexity*, respectively. Each stimulus type was used in both GO and NOGO trials, resulting in a full $2 \times 2 \times 2$ factorial design. This third factor is referred to

herein as *go*. Additionally, a baseline stimulus (three “xxx” syllables) was included which informed the subject that there was no speech to be planned or produced for this trial, but that (s)he should maintain fixation throughout the trial. Each of the stimulus conditions (nine in total) were encountered with equal probability in the experiment.

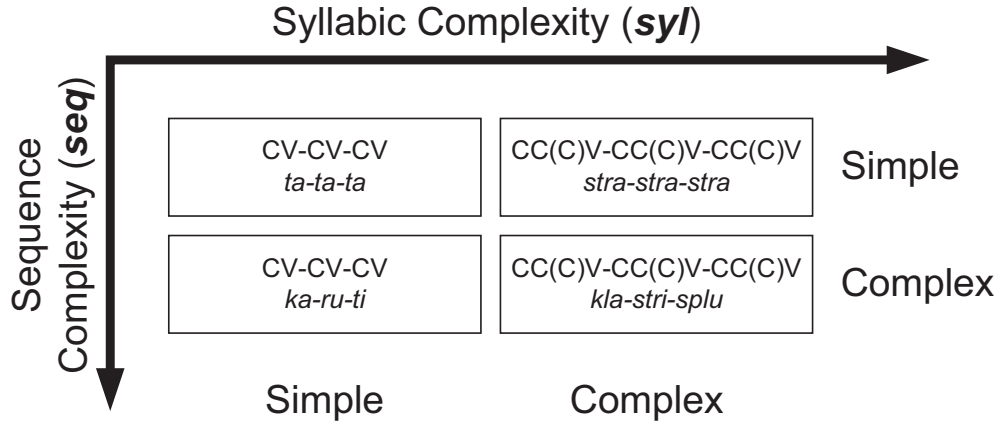


Figure 2.1: Design of syllable sequence stimuli. Sequences were each composed of three syllables presented in lower-case font and separated by hyphens. Four stimulus types were used; a schema for the construction of each type, as well as an example, is shown in the boxes above. Simple sequences (S_{seq}) were repetitions of the same syllable three times; Complex sequences (C_{seq}) contained three unique syllables. A similar complexity parameterization has been used to demonstrate sequence-related effects in previous studies using finger movements (e.g. Shibasaki et al., 1993; Gerloff et al., 1997). At the syllabic level, simple syllables (S_{syl}) were composed of a single consonant and a vowel (CV), whereas complex syllables (C_{syl}) began with a consonant cluster (CCCV or CCV) followed by a vowel. All syllables could be easily produced by speakers of American English; consonants used in S_{syl} were a subset of those used in C_{syl} {/s/, /p/, /t/, /k/, /r/, /l/}, and all vowels were chosen randomly from the English “point” vowels: {/a/, /i/, /u/}. Each stimulus type was used in both GO and NOGO trials.

Each approximately 20 minute-long functional run consisted of the presentation

of 80 stimuli², and subjects were asked to complete three runs. For two subjects only two runs were used due to technical difficulties. Stimuli were delivered using the PsyScope software package (Cohen et al., 1993). Each trial began with the visual (orthographic) presentation of a stimulus on a projection screen in the rear of the scanner³ (a single trial is schematized in Figure 2.2). After 2.5 s the syllables were removed and immediately replaced by a white fixation cross in the center of the visual field. Subjects were instructed to maintain fixation and to prepare to speak the syllable sequence that they had just read. In GO trials, after a short random duration (chosen uniformly from 0.5 - 2.0 s), the white cross turned green, signaling the subject to immediately produce the prepared sequence. Subjects were instructed to speak at a typical volume and rate and to speak monotonously (avoiding prosodic modulation). The scanner remained silent throughout the 2.5 s production period and was then *triggered* to acquire three functional volumes⁴ (see acquisition details below). In NOGO trials the fixation cross remained white throughout. Because of the random time jitter preceding the production period, subjects were unable to differentiate GO and NOGO trials until scanning had begun for a particular trial. Following the third volume acquisition, the fixation cross disappeared and was replaced by the next stimulus. The mean inter-trial interval was 13.75 s. Vocal responses were recorded using an MRI-compatible microphone; for this purpose custom modifications were made to the Shure[®] SM93 (Shure Inc., Niles, IL) lavalier condenser microphone.

²One subject performed 100 stimuli per run; all other aspects were equivalent to other subjects' sessions.

³The standard procedure for presenting visual stimuli in the Siemens Trio Scanner at the Martinos Center for Biomedical Imaging was used. This involves back-projecting the image onto a plexi-glass screen at the rear of the scanner, behind the subject. A mirror is fixed to the head coil and positioned to allow the subject to fully view the display on the screen.

⁴In GO trials, the first volume was acquired between 2.5 s and 5.0 s after the GO signal. Due to the hemodynamic delay (peaking $\sim 5 - 6$ s after task performance; Birn et al. 1999), the response in this volume is likely to be similar to the response to the NOGO task. The second and third volumes, however, are time aligned to capture the peak of the response to the GO task (5.0 to 10.0 s after the GO signal).

Utterance durations were estimated from the recorded signals, and means for each subject and condition were entered into paired t-tests to test the hypothesis that different stimulus conditions resulted in different utterance durations. Trials in which subjects produced incorrect utterances were removed from all analyses.

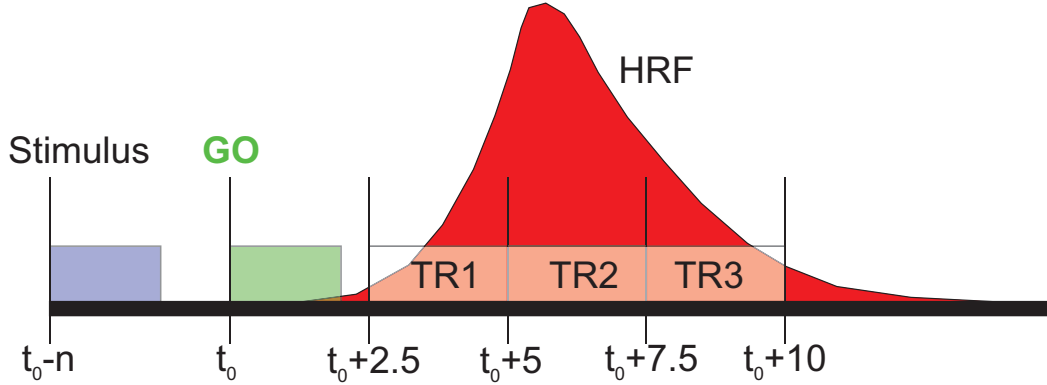


Figure 2-2: The time course of a single trial in the fMRI experiment. Each trial began with the presentation of the stimulus for 2.5 seconds (blue shaded area). After the stimulus was removed, a random delay period (between 0.5 and 2.0 s) was followed, on GO trials, by a GO signal; on NOGO trials, no GO signal was provided. In GO trials, subjects spoke overtly during the period shaded in green. Three functional volumes were acquired in the time interval from 2.5 s to 10.0 s after the GO signal (shaded in white). The red filled curve shows a schematized hemodynamic response curve corresponding to the response due to neural activity occurring just after the GO signal.

2.2.3 Data Acquisition

Subjects lay supine in a 3 Tesla Siemens Trio whole-body scanner (Siemens Medical Systems, Erlangen, Germany) with a Bruker head coil (Bruker BioSpin MRI Inc., Billerica, MA). Foam padding applied between the subject's head and the head coil helped to constrain head movement. A high-resolution anatomical volume (T1-weighted, 128 sagittal images, 256×256 matrix, 1 mm² in-plane resolution, 1.33 mm slice thickness, TR=2530 ms, TE=3.3 ms, flip angle 9°) was acquired for each

subject prior to the functional series. Functional images were acquired sparsely, and in three-volume clusters, triggered by TTL (Transistor-Transistor Logic) pulses delivered to the scanner at appropriate times from the stimulus computer (Macintosh iBook notebook computer). Pulses were sent via the serial port using a custom software and hardware extension to the PsyScope software developed for this project. 30 axial slices (5 mm thickness, 0 mm gap, 64×64 matrix, 3.125 mm^2 in-plane resolution) oriented parallel to the line between the anterior and posterior commissures were acquired in each functional volume using a T2* weighted gradient echo pulse sequence (TR=2500 ms, TE=30 ms, flip angle 90°). These slices were sufficient to cover the entire brain in all subjects. A T1-weighted anatomical volume was also acquired using the same slice parameters as the functional images and was used for between-modality co-registration.

2.2.4 Data Analysis

Functions from the SPM2 software package (Wellcome Department of Imaging Neuroscience, London, UK) were used for pre-processing and voxel-based analyses within MATLAB® (The MathWorks, Inc., Natick, MA). Functional images were realigned to the first image from each series by estimating and applying the parameters of a rigid-body transformation; these coefficients were also included as covariates of non-interest during model estimation. Images were then co-registered to the anatomical scans, non-linearly warped (spatially normalized) to a template in Montreal Neurological Institute (MNI) space (Evans et al., 1993), and smoothed using an isotropic Gaussian kernel with full width at half-maximum (FWHM) of 8 mm. Stimulus events were modeled as delta functions, and the hemodynamic response at each event was estimated using a finite impulse response (FIR) model with a single time bin. This method makes no assumptions about the shape of the hemodynamic response, and

is well suited for event-related studies (Henson et al., 2001). Differences in the global signal level between the three functional volumes in each acquisition cluster were accounted for through linear regression (covariates of non-interest).

A mixed-effects analysis was used. Statistical models were estimated individually for each subject at the first level. A non-parametric permutation test approach (Nichols and Holmes, 2001) was used to assess effects across subjects. This method makes weaker assumptions about the data than methods based on Gaussian Random Fields, and is particularly useful for second-level tests with low degrees of freedom (Nichols and Holmes, 2001). Using the assumption of exchangeability, condition labels were randomly permuted for each subject, resulting in $2^{\#of\ subjects} = 8192$ permutations for each contrast. Under the null hypothesis of no effect, “incorrect” (random) permutations of condition labels will yield roughly the same statistics as the “correct” (designed) labeling. Significance, therefore, was determined by comparing a test statistic for the “correct” labeling to the distribution of that statistic across all permutations. Variance estimates for each voxel were pooled across a $4 \times 4 \times 4 \text{ mm}^3$ volume, yielding additional degrees of freedom and a resulting *pseudo-T* statistical map.

In addition to these voxel-based inferences, a region-of-interest (ROI) analysis was performed (Nieto-Castanon et al., 2003) to provide supplementary information about the size and significance of effects in specific, anatomically-defined cortical areas. The FreeSurfer software package was used to reconstruct cortical surfaces from each subject’s anatomical scan (Dale et al., 1999; Fischl et al., 1999) and was trained to perform cortical parcellation (Fischl et al., 2004) according to a scheme based on anatomical landmarks and node points that was developed for speech-related studies (Tourville and Guenther, 2003). Previous tests revealed that the average overlap between regions assigned by FreeSurfer and regions assigned by a

trained neuroanatomist was approximately 74%, with most errors occurring near region boundaries (S.S. Ghosh, 2005, personal communication). fMRI data from each region in each subject were extracted, and dimensionality was reduced using a Fourier basis set. A mixed-effects analysis used the same design matrices as in the voxel-based analysis. Effects related to a particular contrast were considered significant for $P < 0.001$. The ROI tools were also used when possible to test for lateralization in particular ROIs. For this purpose, the effect sizes estimated for each subject in the left and right hemisphere for a particular ROI were entered into a one-tailed paired t-test. Lateralization was considered significant for $P < 0.05$.

Each of the individual speaking conditions was contrasted with the baseline condition. For these contrasts the False Discovery Rate (FDR) method (Benjamini and Hochberg, 1995; Genovese et al., 2002) was used to correct for multiple comparisons. A minimal speech production network was established by combining the statistical images for each overt speaking condition using a conjunction approach based on the “conjunction null” hypothesis (Nichols et al., 2005). A factorial analysis was used to estimate regions showing direct and/or interaction effects of each factor (*go*, *seq*, and *syl*). “Increasing” the level of each factor (from simple to complex or from NOGO to GO) was hypothesized to lead to additional activation in relevant areas. Effects in this “positive” direction are shown in the results. Inference used a combination of voxel height and cluster extent (Hayasaka and Nichols, 2004). The cluster-defining threshold was set at $\mu_c = 4$, approximately corresponding to $P < 0.001$ uncorrected. Height and extent tests were combined using the unweighted ($\theta = 0.5$) Tippet, Fisher, and cluster mass combining functions, and these were meta-combined in an additional permutation test (see Hayasaka and Nichols, 2004 for details). P-values from the individual and combined tests were corrected to control family-wise error rate (FWE). Areas which reached significance ($P_{FWE} < 0.05$) in the voxel test or

the combined voxel / cluster test are included in the results.



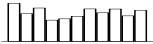

The “Automated Anatomical Labeling” atlas (Tzourio-Mazoyer et al., 2002) was used to identify region labels for activation peaks. Cerebellar labelings refer to the parcellation scheme of Schmahmann et al. (1999). For visualization results were rendered on partially inflated cortical surfaces, created by using FreeSurfer to segment and process the cortical surface of the canonical SPM brain. It should be noted that the analysis was performed volumetrically and resulting statistical maps were projected onto the cortical surface. This results, in some cases, in activations that are contiguous in the volume but non-contiguous on the surface, primarily due to voxel-based smoothing across the banks of a sulcus.

2.3 Results

2.3.1 Acoustic analysis

Table 2.1 shows the means and across-subject standard deviations of acoustic production durations by condition. The difference between S_seq , S_syl and C_seq , S_syl was not significant. All other pair-wise differences were significant ($p < 0.05$).

Table 2.1: Measured durations of acoustic signal resulting from production of utterances in each condition. From left to right, the table shows the condition, the mean duration across all subjects, the standard deviation across individual subject means, and a bar plot of individual subject means for that condition.

Condition	Mean Duration	Standard Deviation	Subjects
Simple seq / simple syl	993 ms	215 ms	
Complex seq / simple syl	1006 ms	186 ms	
Simple seq / complex syl	1195 ms	209 ms	
Complex seq / complex syl	1332 ms	155 ms	

2.3.2 Basic speech production network

Production of each of the stimulus types was individually contrasted with the baseline condition (passive viewing of “xxx-xxx-xxx” stimuli). Group results showed regions of significant activation that were largely overlapping across stimulus types. Table 2.2 summarizes strongly significant ($P_{FDR} < 0.01$) activations for each of the four GO conditions compared to baseline. The conjunction of activity across the four speaking conditions is shown in Figure 2.3.

The minimal network for overt production included, bilaterally, the central sulcus extending rostrally onto the precentral gyrus and caudally onto the postcentral gyrus (including ventral premotor cortex, ventral motor cortex, and ventral somatosensory cortex); the anterior insula; the superior temporal cortex extending posteriorly from the primary auditory cortex along the sylvian fissure to the parietal-temporal junction (including Heschl’s gyrus, planum temporale, and the posterior superior temporal gyrus); the medial premotor areas including the supplementary motor area (SMA) and extending antero-ventrally into the pre-SMA and cingulate sulcus; the basal ganglia (putamen / pallidum); the thalamus; and the superior cerebellar hemispheres (Lobule VI and Crus I). The frontal opercular region was activated and appeared to be somewhat left-lateralized. ROI analysis confirmed that the inferior frontal gyrus *pars opercularis* was significantly active ($P < 0.001$) in all speaking conditions but did not find significant left-lateralization. The anterior insula showed a strong left lateralization ($P < 0.02$). Additional lateralized responses emerged in the left inferior frontal sulcus (IFS) above the inferior frontal gyrus *pars triangularis*, and in the right inferior cerebellum (Lobule VIII). Finally, an activation focus was observed at the base of the pons on the right (not shown).

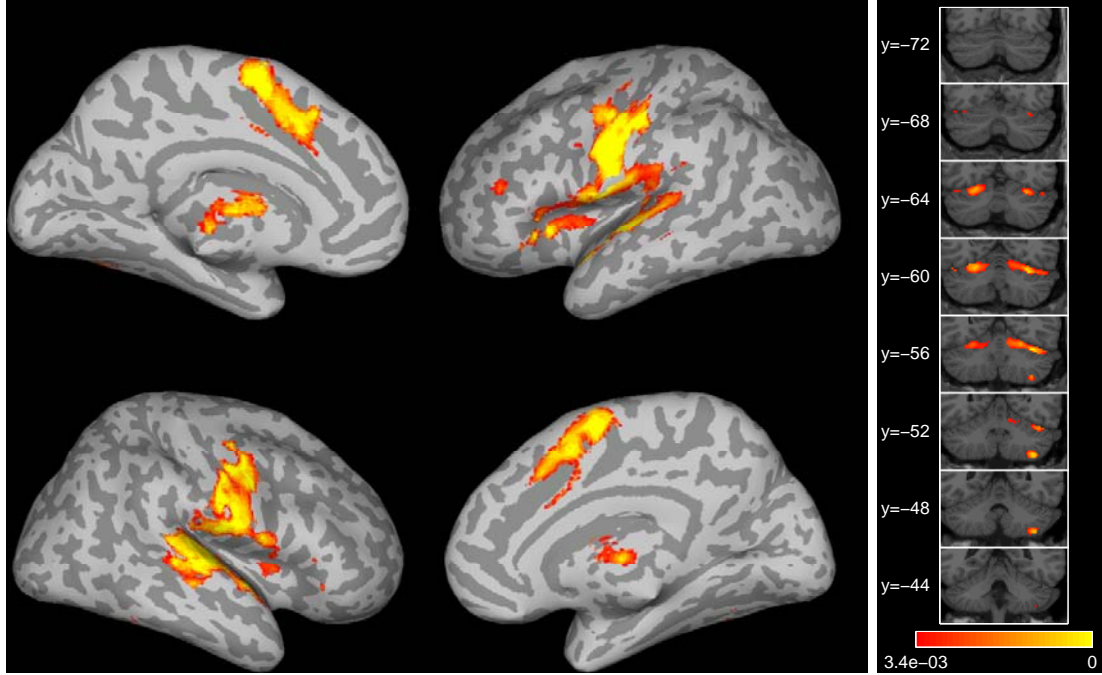


Figure 2-3: Minimal speech production network. These renderings show the conjunction of activations in the four overt speaking (GO) conditions compared to baseline. The map was thresholded to control false discovery rate at 5%. The color scale represents significance level (P-value) of activations, and results are rendered using a logarithmic scale ($-\log_{10}(P)$). *Left:* Significant activity rendered on semi-inflated cortical surface. Dark gray cortical areas represent sulci, lighter gray areas are gyri. *Right:* Significant activations rendered on coronal slices through the cerebellum at various depths. Anatomical sections are cropped versions of the canonical SPM T1 image, and follow neurological conventions (right hemisphere on the right side of image); y-values refer to planes in MNI-space. The color scale is common to both cortical and cerebellar renderings.

Table 2.2: Activation peak summary for each overt speaking condition versus baseline (False Discovery Rate (FDR) < 0.01), sorted by anatomical region. Left to right, the columns show the region label (Tzourio-Mazoyer et al., 2002), followed by pseudo-T value and MNI coordinates of activity peaks in that region for each of the four conditions.

Region	S_syl S_seq		S_syl C_seq		C_syl S_seq		C_syl C_seq	
	pseudo-T	MNI-coord	pseudo-T	MNI-coord	pseudo-T	MNI-coord	pseudo-T	MNI-coord
Precentral_L	3.54	(-46,-10,60)	4.78	(-44,-18,64)			7.04	(-60,0,30)
Precentral_R	4.1	(56,6,40)	4.2	(-34,-6,54)				
			5.05	(48,6,32)	3.83	(56,6,40)	6.55	(62,-4,42)
Postcentral_L	11.38	(-54,-12,40)	4.91	(56,8,32)	3.63	(50,-14,60)	5.14	(56,8,40)
			14.32	(-54,-10,40)	14.63	(-52,-12,40)	12.82	(-54,-10,44)
Postcentral_R	9.3	(-62,-4,24)	9.31	(-62,-4,22)	9.2	(-62,-4,24)		
			10.76	(64,-10,14)	8.11	(68,-6,26)	7.72	(60,-4,32)
Rolandic_Oper_L	6.21	(-42,-24,12)	10.16	(66,-4,24)	7.85	(56,-8,38)	6.19	(68,-4,22)
			10.11	(56,-6,34)				
Rolandic_Oper_R	8.54	(66,-10,12)	11.53	(-48,-26,14)	10.11	(-42,-24,12)	7.44	(-44,-24,14)
			8.86	(-64,-12,12)	7.49	(-52,0,4)		
Insula_L	7.17	(-32,20,0)	7.89	(-50,2,6)	8.4	(68,-8,12)		
			10.69	(54,-14,12)				
Insula_R	5	(40,10,6)	5.36	(38,-6,14)	6.94	(-44,4,0)	10.54	(-34,24,0)
			10.07	(-34,8,8)			8.23	(-40,14,4)
Heschl_L	4.09	(38,24,2)	9.48	(-34,-4,8)	5.44	(36,-22,6)	8.37	(40,24,0)
			4.27	(42,8,-14)			9.18	(-62,-12,8)
Heschl_R	4.52	(38,-22,8)						
			7.15	(-66,-24,8)	7.89	(-64,-10,6)	5.63	(-64,-30,12)
Temporal_Sup_L	7.87	(-62,-8,6)	4.72	(-40,-2,-14)	5.64	(-66,-22,8)		
			7.73	(-54,-4,4)	5.16	(-64,-30,12)		
Temporal_Sup_R	7.11	(56,-26,8)	9.92	(70,-26,8)	8.34	(58,-12,6)	7.33	(52,-24,10)
					8.26	(66,-22,8)	7.16	(60,-14,8)
Temporal_Pole_Sup_L	7.09	(-56,8,-6)			7.25	(50,-24,12)	6.58	(64,-28,2)
			6.37	(-54,10,-6)	6.53	(70,-30,16)		
					6.78	(-52,8,-4)		

Continued on Next Page...

Region	S_syl S_seq			S_syl C_seq			C_syl S_seq			C_syl C_seq		
	pseudo-T	MNI-coord		pseudo-T	MNI-coord		pseudo-T	MNI-coord		pseudo-T	MNI-coord	
Temporal_Pole_Sup_R Frontal_Inf_Oper_L	7.24	(64,6,2)		7.75	(64,6,2)		7.03	(64,6,2)		6.38	(64,8,0)	
	4.13	(50,4,-8)		7.29	(-46,6,28)					7.6	(-54,12,0)	
				5.81	(-54,14,32)					7.09	(-48,10,28)	
Frontal_Inf_Oper_R Frontal_Inf_Tri_L	4.09	(48,14,18)								8.71	(50,20,-6)	
	3.26	(-34,36,12)		8.12	(-38,24,2)					7.09	(-56,16,30)	
							3.75	(-38,32,14)		4.8	(-46,28,24)	
Frontal_Inf_Tri_R Frontal_Mid_L				6.68	(40,26,4)							
				5.11	(50,20,0)							
	1.77	(-32,46,12)		3	(-36,40,32)					4.37	(-30,-6,50)	
Frontal_Mid_R Frontal_Motor_Area_L				2.89	(-26,46,20)					4.04	(-42,46,20)	
				5.51	(56,-8,54)							
	10.13	(-2,-2,68)		9.39	(0,0,66)		10.16	(2,0,66)		15.94	(0,0,70)	
Supp_Motor_Area_R Cingulum_Mid_L	6.59	(-2,6,50)		6.3	(-6,10,52)							
	8.89	(4,4,70)										
	7.34	(-4,14,40)		8.66	(-6,14,42)		6.53	(-2,14,42)		10.61	(-2,18,38)	
Cingulum_Mid_R Parietal_Sup_L Parietal_Inf_L				4.11	(-8,-12,42)							
				3.3	(8,-12,40)							
				5.97	(-28,-52,52)					7.8	(-28,-52,50)	
Parietal_Inf_R SupraMarginal_L				4.23	(-52,-36,50)					5.14	(-52,-38,54)	
				3.08	(40,-48,48)					4.4	(-48,-36,46)	
				2.48	(42,-44,46)					3.52	(34,-56,52)	
Caudate_L Caudate_R Putamen_L				5.47	(-60,-40,30)		5.18	(-58,-38,28)				
	6.56	(-12,2,10)		3.18	(-46,-30,32)							
										9.46	(-12,0,10)	
Putamen_R Pallidum_R Thalamus_L	8.33	(-24,2,-10)		9.84	(-22,2,6)		8.09	(-24,0,-8)		9.64	(12,2,8)	
	8.01	(-20,12,4)		9.84	(-30,-8,-4)		7.03	(-22,4,6)		7.6	(-22,-2,6)	
				9	(-28,-16,10)							
	7.91	(18,10,6)		10.48	(32,-6,-2)		6.99	(20,8,6)		4.53	(32,-20,0)	
	6.84	(32,-6,-2)		8.69	(24,4,6)		6.94	(32,-16,-2)				
	6.58	(24,4,-4)					6.48	(32,-4,-4)				
							5.31	(30,-4,8)				
							6.27	(24,2,-6)				
	11.18	(-8,-20,-2)		10.89	(-10,-16,0)		9.35	(-10,-16,4)		10.14	(-10,-14,6)	

Continued on Next Page...

Region	S_syl S_seq			S_syl C_seq			C_syl S_seq			C_syl C_seq		
	pseudo-T	MNI-coord		pseudo-T	MNI-coord		pseudo-T	MNI-coord		pseudo-T	MNI-coord	
Thalamus_R	7.45	(12,-22,0)		10.48	(14,-16,4)		7.8	(14,-16,2)		8.96	(14,-12,6)	
	6.67	(14,-12,6)								6.62	(8,-20,0)	
	6.24	(20,-16,10)								4.54	(22,-22,-2)	
	5.92	(8,-6,4)										
Cerebellum_6_L	5.63	(-26,-60,-22)		9.21	(-20,-60,-22)		7.2	(-28,-60,-24)		7.73	(-28,-62,-28)	
Cerebellum_6_R	4.75	(40,-56,-26)		10.45	(20,-60,-20)		7.39	(20,-56,-20)		6.38	(-14,-62,-18)	
	4.5	(34,-62,-24)		4.21	(40,-68,-22)					9.15	(22,-62,-20)	
	4.7	(24,-56,-20)								7.96	(34,-62,-26)	
										5.47	(8,-68,-20)	
Cerebellum_Crus1_R				6.81	(44,-54,-28)		6.87	(46,-56,-28)		7.04	(42,-54,-28)	
Cerebellum_8_R	7.02	(36,-54,-54)		8.31	(36,-54,-54)		5.98	(36,-52,-56)		6.82	(38,-52,-56)	
				5.54	(16,-62,-44)					5.06	(36,-40,-52)	
				4.87	(38,-40,-50)					4.74	(20,-64,-46)	
										3.47	(4,-46,-16)	
Vermis_3										4.6	(4,-56,-24)	
Vermis_6										5.83	(-46,-60,-20)	
Fusiform_L	4.28	(-44,-60,-20)					4.04	(-46,-60,-20)				

2.3.3 Main effect of overt production

Figure 2.4 shows the main effect of overt production ($GO > NOGO$; $P_{FWE} < 0.05$)⁵. GO trials resulted in significantly increased responses bilaterally in the primary motor and somatosensory cortices, the superior temporal plane, the anterior insula, and the medial premotor areas, particularly focused in the supplementary motor area near the superior convexity, but also including portions of the pre-SMA and anterior cingulate sulcus. ROI analysis confirmed that both the SMA and pre-SMA bilaterally were more active for GO than for NOGO trials. The anterior cingulate showed the same trend but was not significant. No active cortical ROI's showed significant lateralization for the effect of *go*. Subcortically, the putamen / globus pallidus and two regions of the thalamus (one anterior, one posterior) showed an additional bilateral response. The superior cerebellar cortices (Lobule VI) bilaterally were more active for GO trials, as was a small region in the right inferior cerebellum (anterior Lobule VIII). This latter region was significant in the voxel-based test but not in combined voxel-cluster inference. Table 2.4 summarizes activations for the main effect of *go*.

2.3.4 Main effect of sequence complexity

Figure 2.5 shows the main effect of sequence complexity ($C_seq > S_seq$; $P_{FWE} < 0.05$). The medial premotor areas were more active bilaterally for complex sequences. Region-level testing showed an effect in both hemispheres in the pre-SMA but no effect in the SMA or anterior cingulate. The lateral frontal cortex, including premotor and prefrontal areas and extending along the inferior frontal sulcus was also more

⁵The results shown for main effects and interactions are unidirectional according to the hypothesis that increasing the level of a factor will result in an increase in BOLD response. Regions that showed significant activations in the other direction were typically not active in the baseline contrasts and not areas for which there were no *a priori* hypotheses. Discussion of these areas, which included the angular gyrus, precuneus, and anterior prefrontal regions, is therefore omitted for the sake of brevity.

Table 2.4: Significant ($P < 0.05$, corrected for multiple comparisons) activation peak summary for the main effect of overt production (GO > NOGO). Left to right, columns show the size of contiguous clusters, the P-value for that cluster using combined cluster extent-voxel height inference, the P-value based only on cluster extent, and the voxel-wise P-value, pseudo-T value, MNI coordinates, and anatomical region label for activation peaks within the cluster. All P-values are corrected to control family-wise error.

Cluster-size	P(combo)	P(cluster)	P(voxel)	pseudo-T	MNI (x,y,z)	Region Label
3682	0.00037	0.00171	0.00012	13.14092	(-54,-12,40)	Postcentral_L
			0.00012	11.95341	(-44,-24,12)	Rolandic_Oper_L
			0.00037	10.31301	(-64,-8,20)	Postcentral_L
			0.00037	9.89571	(-62,-6,4)	Temporal_Sup_L
			0.01318	6.81526	(-48,-16,2)	Heschl_L
			0.01648	6.62541	(-50,10,-6)	Temporal_Pole_Sup_L
			0.02441	6.29443	(-60,-30,12)	Temporal_Sup_L
			0.02454	6.28383	(-44,6,-2)	Insula_L
			0.02966	6.1484	(-48,-14,60)	Precentral_L
			0.00024	11.59105	(60,-12,10)	Rolandic_Oper_R
6079	0.00037	0.00073	0.00037	9.79065	(64,8,0)	Temporal_Pole_Sup_R
			0.00122	8.48157	(62,-4,28)	Postcentral_R
			0.00122	8.35654	(50,-22,12)	Rolandic_Oper_R
			0.00281	7.87694	(12,-16,4)	Thalamus_R
			0.00378	7.78941	(46,-14,0)	Temporal_Sup_R
			0.00378	7.74591	(0,-6,12)	Thalamus_Mid
			0.0127	6.83599	(10,0,10)	Caudate_R
			0.01379	6.72545	(68,-26,4)	Temporal_Sup_R
			0.01917	6.49536	(-10,-16,4)	Thalamus_L
			0.03809	5.98882	(-24,0,-8)	Putamen_L
			0.04089	5.93748	(-20,4,2)	Pallidum_L
			0.06079	5.68653	(30,0,-6)	Putamen_R
			0.08899	5.38535	(40,8,4)	Insula_R
			0.09436	5.34178	(-10,-14,16)	Thalamus_L
			0.11584	5.19571	(20,8,4)	Putamen_R
			0.13843	5.06274	(14,-16,16)	Thalamus_R
			0.19312	4.82253	(34,-12,-2)	Putamen_R
			0.39014	4.24658	(48,2,-10)	Temporal_Sup_R
490	0.01111	0.0127	0.01416	6.7078	(32,-66,-22)	Cerebellum_6_R
482	0.01135	0.01294	0.03003	6.14265	(20,-58,-18)	Cerebellum_6_R
			0.01953	6.45726	(-26,-60,-22)	Cerebellum_6_L
1162	0.00635	0.00598	0.02075	6.39888	(-14,-60,-16)	Cerebellum_4_5_L
			0.39856	4.22681	(-8,-58,-2)	Lingual_L
			0.02136	6.37997	(0,0,68)	Supp_Motor_Area_R
			0.02222	6.34191	(2,-6,72)	Supp_Motor_Area_R
			0.08215	5.44359	(0,2,50)	Supp_Motor_Area_R
			0.08728	5.40461	(2,-4,52)	Supp_Motor_Area_R
			0.11011	5.2412	(2,18,40)	Frontal_Sup_Medial_R
			0.14331	5.04149	(-4,-14,78)	Paracentral_Lobule_L
53	0.06458	0.10913	0.04102	5.93599	(38,-48,-56)	Cerebellum_8_R

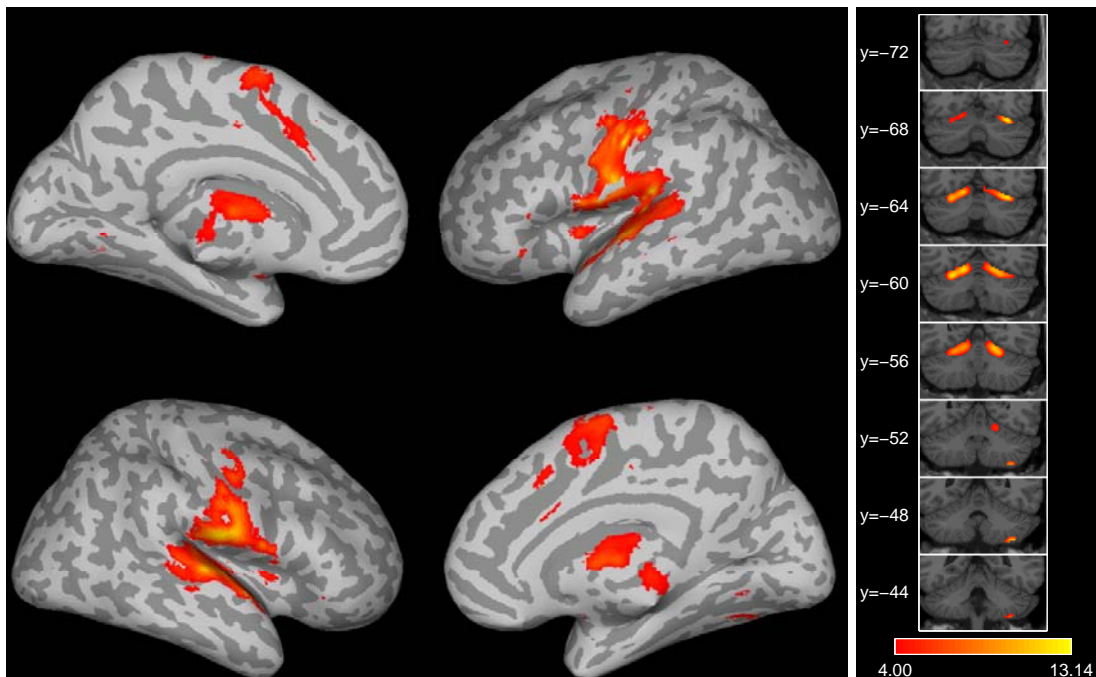


Figure 2.4: Main effect of overt production: areas that showed a significantly greater response for GO trials than for NOGO trials, averaged across other factors. The statistical image was thresholded at $P_{FWE} < 0.05$. Color scale represents voxel-wise pseudo-T value for significant voxels. See *methods* for further details. *Left:* Significant activity rendered on semi-inflated cortical surface. Dark gray cortical areas represent sulci, lighter gray areas are gyri. *Right:* Significant activations rendered on coronal slices through the cerebellum at various depths. Anatomical sections are cropped versions of the canonical SPM T1 image, and follow neurological conventions (right hemisphere on the right side of image); y-values refer to planes in MNI-space. The color scale is common to both cortical and cerebellar renderings.

active. These activations were strikingly left-lateralized in the voxel-based results. The lateralization test for the ventral premotor cortex and the inferior frontal gyrus *pars opercularis* showed very strong left lateralization ($P < 0.001$); however, none of the ROI's in the parcellation scheme (Tourville and Guenther, 2003) corresponded well to the inferior frontal sulcus region, and thus it was not possible to explicitly test this using the current set of available ROI tools. Regions at the junction of the anterior insula and the frontal operculum were engaged bilaterally by sequence complexity. The ROI analysis confirmed that the activation included both the anatomically defined anterior insula and frontal operculum ($P < 0.001$). The effect was significantly greater in the left anterior insula than in the right; no such lateralization effect was found in the frontal operculum. The posterior parietal lobe, left lateralized ($P < 0.05$), and the inferior posterior temporal lobes also showed the sequence complexity effect. The cerebellum demonstrated strong effects bilaterally (although somewhat stronger in the right hemisphere) in the superior areas (Lobule VI, Crus I, Crus II) and unilaterally in the right inferior cerebellar cortex (Lobule VIII). The superior cerebellar activations extended more laterally than those related to the main effect of *go* (see above), and also included portions of the vermis. The anterior thalamus and caudate nucleus also showed a main effect for sequence complexity bilaterally. Table 2.5 summarizes activations for the main effect of *seq*.

2.3.5 Main effect of syllable complexity

Figure 2.6 shows the main effect of syllable complexity ($C_{syl} > S_{syl}$; $P_{FWE} < 0.05$). The medial premotor areas showed additional activation in the voxel-based analysis; region-level testing showed a significant effect isolated to the pre-SMA bilaterally, with no significant difference in the effect size between hemispheres. The junction of the frontal operculum and anterior insula was engaged bilaterally; in the

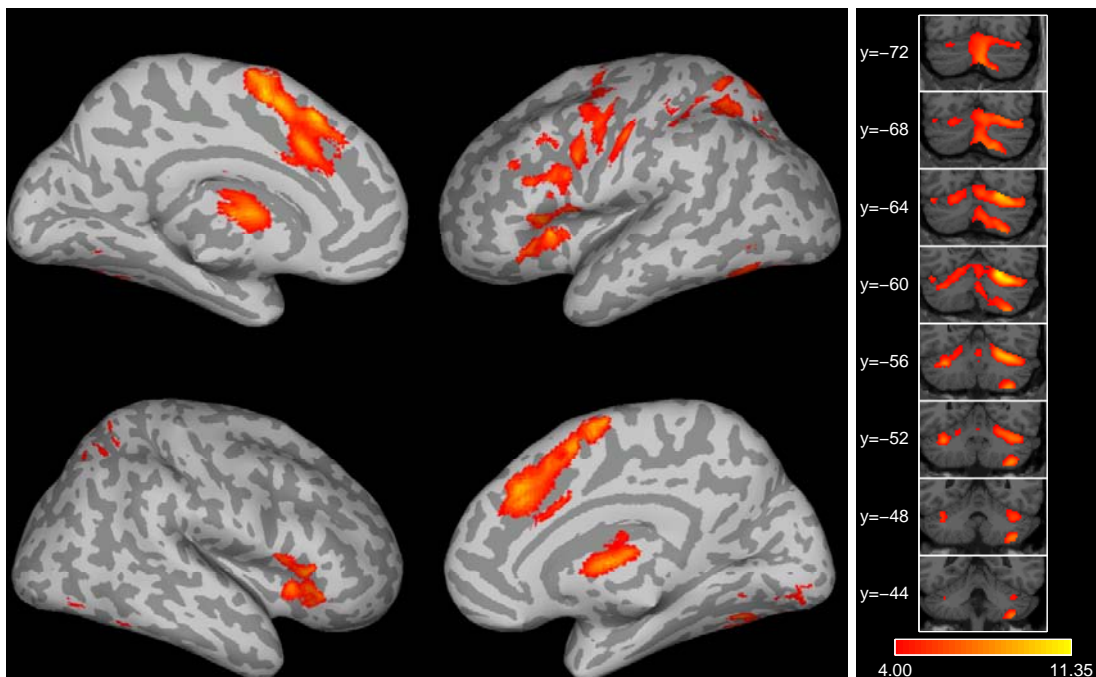


Figure 2.5: Main effect of sequence complexity: areas that showed a significantly greater response to complex sequences than to simple sequences, averaged across other factors. The statistical image was thresholded at $P_{FWE} < 0.05$. Color scale represents voxel-wise pseudo-T value. See *methods* for details. *Left:* Significant activity rendered on semi-inflated cortical surface. Dark gray cortical areas represent sulci, lighter gray areas are gyri. *Right:* Significant activations rendered on coronal slices through the cerebellum at various depths. Anatomical sections are cropped versions of the canonical SPM T1 image, and follow neurological conventions (right hemisphere on the right side of image); y-values refer to planes in MNI-space. The color scale is common to both cortical and cerebellar renderings.

Table 2.5: Significant ($P < 0.05$, corrected for multiple comparisons) activation peak summary for the main effect of *seq*. Left to right, columns show the size of contiguous clusters, P-value for that cluster using combined cluster extent-voxel height inference, P-value based only on cluster extent, and the voxel-wise P-value, pseudo-T value, MNI coordinates, and region label for activation peaks within the cluster. All P-values are corrected to control family-wise error.

Cluster-size	P(combo)	P(cluster)	P(voxel)	pseudo-T	MNI (x,y,z)	Region Label
4920	0.00024	0.00012	0.00049	9.3025	(22,-60,-20)	Cerebelum_6_R
			0.00061	8.6905	(32,-60,-26)	Cerebelum_6_R
			0.0061	7.13077	(-34,-56,-32)	Cerebelum_6_L
			0.00708	7.00493	(36,-54,-56)	Cerebelum_8_R
			0.0083	6.89034	(26,-32,-46)	Cerebelum_8_R
			0.00964	6.8132	(6,-74,-38)	Cerebelum_Crus2_R
			0.00977	6.80131	(16,-70,-48)	Cerebelum_8_R
			0.01575	6.54791	(30,-62,-56)	Cerebelum_8_R
			0.01843	6.43769	(36,-44,-54)	Cerebelum_8_R
			0.03589	6.07315	(-44,-58,-10)	Temporal_Inf_L
			0.04578	5.92515	(6,-68,-18)	Cerebelum_6_R
			0.06006	5.76695	(30,-38,-50)	Cerebelum_8_R
			0.06995	5.6757	(4,-80,-18)	Vermis_6
			0.13599	5.24158	(-24,-64,-22)	Cerebelum_6_L
			0.16626	5.10358	(-48,-64,-22)	Fusiform_L
			0.17029	5.08838	(22,-82,-18)	Fusiform_R
			0.17712	5.0637	(-16,-62,-16)	Cerebelum_6_L
			0.21021	4.94853	(-30,-78,-22)	Cerebelum_6_L
			0.30566	4.65947	(-22,-84,-22)	Cerebelum_Crus1_L
			0.46948	4.31036	(6,-88,-10)	Lingual_R
			0.47888	4.2941	(36,-38,-40)	Cerebelum_Crus2_R
2294	0.00037	0.00061	0.00024	11.3493	(0,6,56)	Supp_Motor_Area_R
			0.00049	9.32545	(8,30,34)	Cingulum_Mid_R
			0.00049	9.25186	(-2,18,46)	Supp_Motor_Area_L
			0.00061	8.66842	(0,2,68)	Supp_Motor_Area_R
			0.00073	8.53792	(0,-6,70)	Supp_Motor_Area_R
			0.00122	8.14325	(-2,22,36)	Frontal_Sup_Medial_L
1736	0.00061	0.00098	0.00281	7.64762	(-48,4,30)	Precentral_L
			0.0061	7.12261	(-56,-8,46)	Postcentral_L
			0.01782	6.46693	(-50,28,24)	Frontal_Inf_Tri_L
			0.02063	6.34655	(-54,16,32)	Frontal_Inf_Oper_L
			0.05212	5.84752	(-50,-6,54)	Precentral_L
			0.05823	5.77984	(-54,6,42)	Precentral_L
			0.08655	5.54831	(-32,-4,64)	Frontal_Sup_L
			0.1167	5.35413	(-42,-2,44)	Precentral_L
			0.18481	5.03744	(-32,-4,52)	Precentral_L
			0.21655	4.92606	(-58,10,20)	Frontal_Inf_Oper_L
			0.23328	4.86595	(-26,-6,50)	Frontal_Sup_L
1153	0.00061	0.00305	0.00061	8.71686	(0,-6,12)	Thalamus
			0.00098	8.27126	(-8,-2,10)	Caudate_L
			0.23267	4.86751	(18,-8,20)	Caudate_R
1031	0.00061	0.00354	0.00061	8.71972	(-32,22,4)	Insula_L
			0.00452	7.2841	(-42,16,6)	Insula_L
			0.00854	6.88164	(-48,14,2)	Frontal_Inf_Oper_L
			0.02576	6.24065	(-48,20,-6)	Frontal_Inf_Orb_L
830	0.00171	0.00476	0.00195	7.84031	(40,22,2)	Insula_R
1063	0.0022	0.0033	0.00391	7.36148	(50,20,-2)	Frontal_Inf_Oper_R
			0.00684	7.02179	(-30,-54,58)	Parietal_Sup_L
			0.00757	6.95509	(-26,-60,56)	Parietal_Sup_L
			0.01013	6.76815	(-30,-48,46)	Parietal_Inf_L
			0.11938	5.33424	(-48,-32,46)	Postcentral_L
			0.41821	4.40675	(-26,-68,38)	Parietal_Sup_L
130	0.07507	0.04443	0.45251	4.34567	(-52,-34,52)	Postcentral_L
			0.14087	5.21796	(26,-64,64)	Parietal_Sup_R
			0.23376	4.86263	(32,-56,52)	Parietal_Inf_R

ROI test, the effect was significant in the anatomically defined frontal operculum (FO) in both hemispheres, but the effect was below significance in the anterior insula in both hemispheres. Additionally, the left posterior parietal cortex, near the intraparietal and postcentral sulci demonstrated an effect due to *syl*. Cerebellar effects were much more focal when compared with the effect of *seq*, with significant increased activity limited to the right superior cerebellar cortex (Lobule VI) near the vermis, and generally posterior to the areas showing an effect of *seq* (see Figure 2.5). Table 2.6 summarizes activations for the main effect of *syl*.

Table 2.6: Significant ($P < 0.05$, corrected for multiple comparisons) activation peak summary for the main effect of syllable complexity (*syl*). Left to right, columns show the size of contiguous clusters, the P-value for that cluster using combined cluster extent-voxel height inference, the P-value based only on cluster extent, and the voxel-wise P-value, pseudo-T value, MNI coordinates, and anatomical region label for activation peaks within the cluster. All P-values are corrected to control family-wise error.

Cluster-size	P(combo)	P(cluster)	P(voxel)	pseudo-T	MNI (x,y,z)	Region Label
1106	0.00159	0.00488	0.00061	8.38733	(0,18,46)	Supp_Motor_Area_L
			0.0094	7.00759	(0,4,62)	Supp_Motor_Area_R
			0.01013	6.95133	(0,0,70)	Supp_Motor_Area_R
			0.04236	5.97899	(4,24,38)	Cingulum_Mid_R
510	0.00879	0.01306	0.00623	7.20664	(50,22,-6)	Frontal_Inf_Orb_R
			0.09216	5.4626	(42,20,-12)	Frontal_Inf_Orb_R
			0.0979	5.42468	(38,26,0)	Insula_R
			0.125	5.24541	(38,24,-6)	Insula_R
346	0.02197	0.02063	0.021	6.40769	(-26,-62,52)	Parietal_Sup_L
			0.05579	5.7753	(-30,-54,52)	Parietal_Inf_L
			0.12891	5.22414	(-48,-40,52)	Parietal_Inf_L
			0.3396	4.44609	(-20,-66,66)	Parietal_Sup_L
380	0.02026	0.01855	0.42749	4.23381	(-38,-44,44)	Parietal_Inf_L
			0.05469	5.78835	(-34,26,0)	Frontal_Inf_Tri_L
			0.06726	5.6656	(-34,22,4)	Insula_L
			0.11047	5.33845	(-50,12,0)	Frontal_Inf_Oper_L
178	0.07104	0.04468	0.16602	5.02891	(22,-76,-20)	Cerebellum_6_R
			0.19812	4.89095	(26,-62,-18)	Cerebellum_6_R

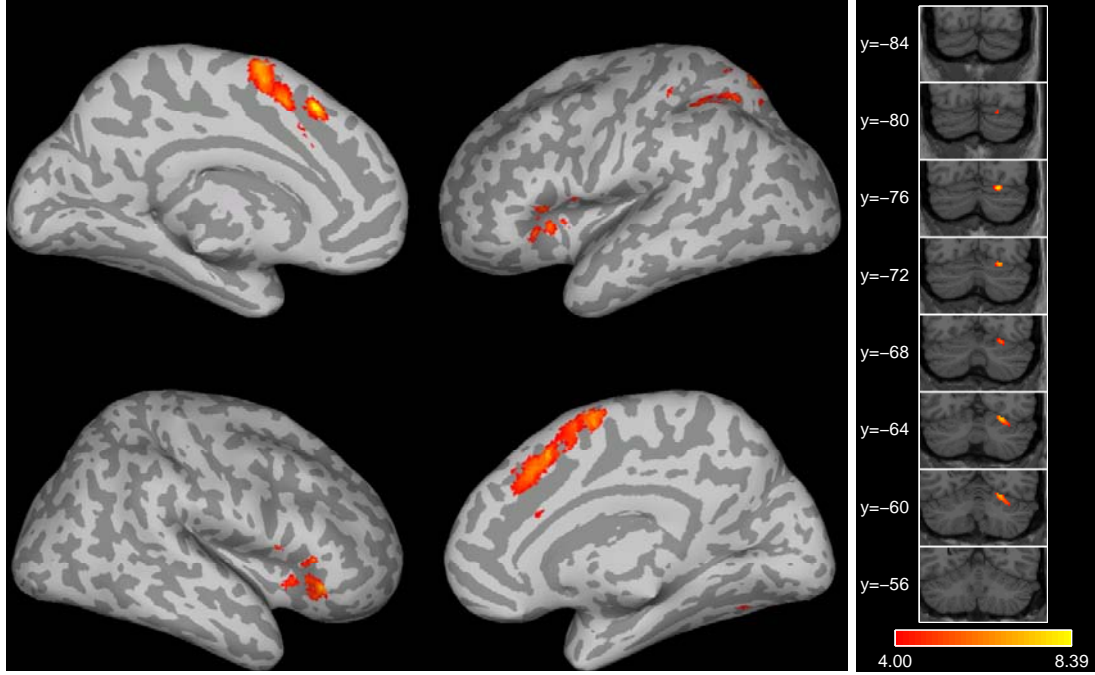


Figure 2-6: Main effect of syllable complexity: areas that showed a significantly greater response for sequences comprised of complex syllables than for sequences comprised of simple syllables, averaged across other factors. The statistical image was thresholded at $P_{FWE} < 0.05$. Color scale represents voxel-wise pseudo-T value. See *methods* for details. *Left:* Significant activity rendered on semi-inflated cortical surface. Dark gray cortical areas represent sulci, lighter gray areas are gyri. *Right:* Significant activations rendered on coronal slices through the cerebellum at various depths. Anatomical images are cropped versions of the canonical SPM T1 image, and follow neurological conventions (right hemisphere on the right side of image); y-values refer to planes in MNI-space. The color scale is common to both cortical and cerebellar renderings.

2.3.6 Interactions between factors

No significant ($P_{FWE} < 0.05$) interaction effects were found for $go \times seq$, $go \times syl$, or for the three-way interaction $go \times seq \times syl$. There was, however, a strong interaction between the factors seq and syl . Figure 2-7 shows brain areas that demonstrated a significant positive-direction interaction between sequence complexity and syllable complexity (i.e. $\{C_syl, C_seq - C_syl, S_seq\} > \{S_syl, C_seq - S_syl, S_seq\}$). These areas included the medial premotor cortices (SMA / pre-SMA / cingulate sulcus), the junction of the frontal operculum and anterior insula bilaterally, the left posterior parietal cortex, the anterior thalamus, the superior cerebellum, and regions of the precentral gyrus and prefrontal cortex in and surrounding the inferior frontal sulcus, primarily in the left hemisphere. Results from region-level testing showed that the medial activations only produced a significant effect in the pre-SMA (and not SMA), bilaterally. The effects in the ventral premotor cortex, inferior frontal gyrus *pars opercularis*, and superior parietal lobe were significantly ($P < 0.05$) left-lateralized. Table 2.7 summarizes activations for the $seq \times syl$ interaction. A further investigation of interactions between syl and seq is also available in the online supplementary materials.

The finding that sequence complexity (seq) and syllable complexity (syl) had a significant interaction in certain areas warranted further investigation. Portions of the left prefrontal cortex, for example, showed a main effect for seq but not for syl , but also showed a strong interaction effect. It was useful, then, to determine how the effect of sequence complexity (Figure 2-5) differed for the two levels of syllable complexity; that is, how the additional activity required for sequencing multiple unique syllables was modulated by the phonetic / articulatory complexity of each syllable. The effect of seq was tested individually within each of the two levels of syllable complexity (simple syllables, complex syllables). Figure 2-8 shows the

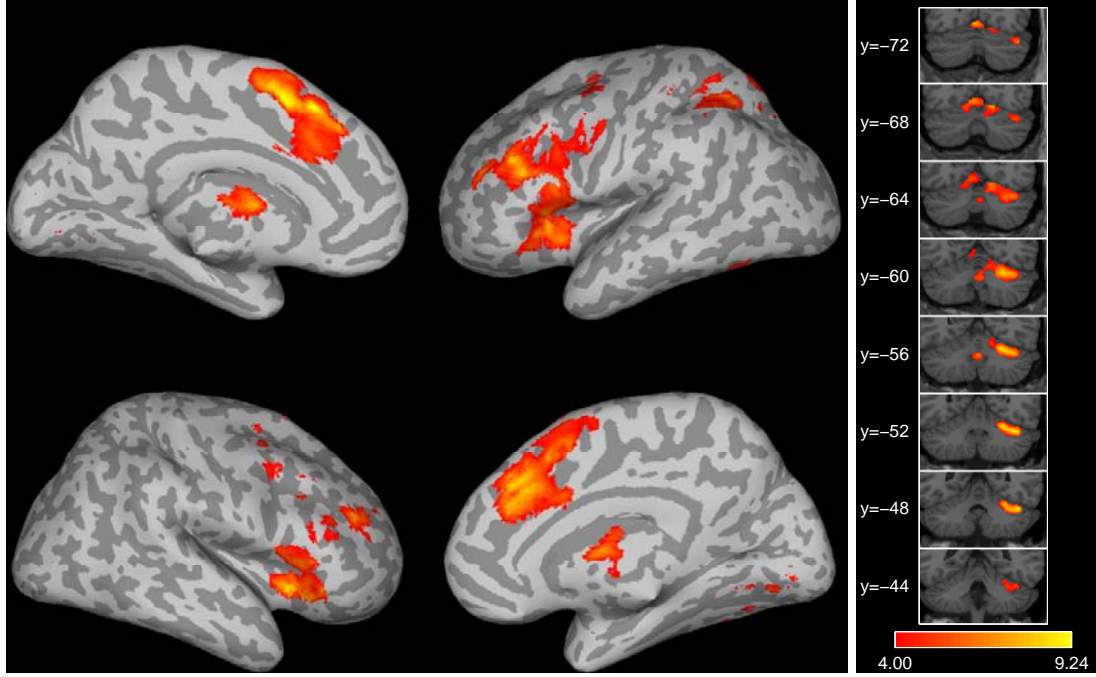


Figure 2-7: Interactions between sequence complexity and syllable complexity. The statistical image was thresholded at $P_{FWE} < 0.05$. Color scale represents voxel-wise pseudo-T value. See *methods* for details. *Left:* Significant activity rendered on semi-inflated cortical surface. Dark gray cortical areas represent sulci, lighter gray areas are gyri. *Right:* Significant activations rendered on coronal slices through the cerebellum at various depths. Anatomical images are cropped versions of the canonical SPM T1 image, and follow neurological conventions (right hemisphere on the right side of image); y-values refer to planes in MNI-space. The color scale is common to both cortical and cerebellar renderings.

Table 2.7: Significant ($P < 0.05$, corrected for multiple comparisons) activation peak summary for the positive interaction effect of syllable complexity \times sequence complexity ($seq \times syl$). Left to right, columns show the size of contiguous clusters, the P-value for that cluster using combined cluster extent-voxel height inference, the P-value based only on cluster extent, and the voxel-wise P-value, pseudo-T value, MNI coordinates, and anatomical region label for activation peaks within the cluster. All P-values are corrected to control family-wise error.

Cluster-size	P(combo)	P(cluster)	P(voxel)	pseudo-T	MNI (x,y,z)	Region Label
2768	0.00012	0.00037	0.00012	9.24008	(0,16,48)	Supp_Motor_Area_L
			0.00037	8.85036	(-8,8,62)	Supp_Motor_Area_L
			0.00037	8.32387	(2,34,36)	Frontal_Sup_Medial_R
			0.00073	8.07759	(8,26,34)	Cingulum_Mid_R
			0.03589	6.11734	(0,16,66)	N/A
			0.06482	5.77648	(2,14,32)	Cingulum_Mid_R
			0.08374	5.61253	(-6,24,28)	Cingulum_Ant_L
2101	0.00012	0.00049	0.00012	9.15435	(34,22,-8)	Frontal_Inf_Orb_R
			0.00195	7.77868	(38,44,24)	Frontal_Mid_R
			0.00891	6.97827	(52,20,-4)	Frontal_Inf_Orb_R
			0.01501	6.60839	(40,20,10)	Frontal_Inf_Tri_R
			0.13525	5.30193	(52,34,26)	Frontal_Inf_Tri_R
			0.31763	4.67356	(58,24,14)	Frontal_Inf_Tri_R
			0.00037	8.41327	(-42,30,24)	Frontal_Inf_Tri_L
3187	0.00024	0.00037	0.00305	7.42877	(-30,24,6)	Insula_L
			0.00439	7.31329	(-42,46,22)	Frontal_Mid_L
			0.01282	6.68387	(-36,16,-8)	Insula_L
			0.01404	6.64428	(-58,14,18)	Frontal_Inf_Oper_L
			0.04053	6.06169	(-52,16,14)	Frontal_Inf_Tri_L
			0.04272	6.04246	(-44,14,4)	Insula_L
			0.12463	5.35462	(-62,6,28)	Precentral_L
			0.15784	5.19544	(-40,12,26)	Frontal_Inf_Tri_L
			0.27173	4.80442	(-52,10,44)	Frontal_Mid_L
			0.31409	4.68594	(-50,4,36)	Precentral_L
			0.01111	6.81014	(42,-50,-30)	Cerebellum_Crus1_R
			0.0166	6.5356	(28,-52,-24)	Cerebellum_6_R
			0.02649	6.27152	(32,-52,-28)	Cerebellum_6_R
			0.0271	6.24839	(36,-56,-28)	Cerebellum_6_R
1686	0.00134	0.00085	0.03821	6.08466	(-2,-72,-8)	Vermis_6
			0.05945	5.82793	(14,-66,-12)	Cerebellum_6_R
			0.11084	5.43767	(42,-72,-28)	Cerebellum_Crus1_R
			0.13684	5.29047	(2,-56,-32)	Vermis_9
			0.31763	4.67387	(14,-58,-20)	Cerebellum_6_R
			0.52759	4.17515	(14,-54,-14)	Cerebellum_4_5_R
			0.00317	7.39145	(16,-6,14)	Caudate_R
			0.00415	7.32411	(-10,0,10)	Caudate_L
			0.01111	6.81383	(10,-2,12)	Caudate_R
			0.03857	6.08045	(-4,-10,14)	Thalamus_L
856	0.00244	0.00366	0.07166	5.71289	(8,-8,2)	Thalamus_R
			0.01379	6.65005	(-30,-52,50)	Parietal_Inf_L
			0.10303	5.48829	(-40,-44,54)	Parietal_Sup_L
			0.1759	5.12255	(-52,-40,56)	Postcentral_L
			0.18689	5.08376	(-36,-48,42)	Parietal_Inf_L
			0.48474	4.26502	(-24,-72,46)	Parietal_Sup_L
			0.51648	4.19948	(-18,-68,64)	Parietal_Sup_L
292	0.0282	0.01501	0.07263	5.69839	(34,2,58)	Frontal_Mid_R
			0.19836	5.0424	(34,2,38)	Frontal_Mid_R
			0.21497	4.98342	(34,4,44)	Frontal_Mid_R
			0.23511	4.92222	(44,12,38)	Frontal_Mid_R
			0.43384	4.36841	(34,0,48)	Precentral_R
			0.0166	6.53542	(-44,-58,-16)	Fusiform_L
114	0.03137	0.06018	0.10193	5.49464	(-32,0,52)	Frontal_Mid_L
221	0.0354	0.02271	0.15063	5.22492	(-38,0,62)	Precentral_L

effects of additional sequence complexity within GO trials for both syllable types (C_seq / S_syl minus S_seq / S_syl and C_seq / C_syl minus S_seq / C_syl) as well as the intersection (conjunction) of these comparisons rendered on a single brain. The pseudo-T map in Figure 2-8 was subjected to a less stringent threshold than the other figures. Because the comparison involved many fewer trials for each subject, the statistical power was insufficient to allow for corrections for multiple comparisons. Nevertheless, the uncorrected statistical map provides some insight into the interactions between the two complexity factors.

2.4 Discussion

This study was designed to provide additional insight into the neural substrates for planning and producing sequences of simple speech sounds, a faculty that is ubiquitous in normal discourse. This topic has received relatively little attention in the neuroimaging literature to date, with most studies of language production focusing on aspects of word generation and production (reviewed in Indefrey and Levelt, 2000; Turkeltaub et al., 2002), or on other aspects of verbal output such as speaking rate (Wildgruber et al., 2001; Riecker et al., 2005) or prosody (Riecker et al., 2002). Previous computational studies in the Guenther laboratory have led to the implementation of a neural model that is capable of learning and producing (by means of a computer-simulated vocal tract) simple speech sounds (Guenther, 1994, 1995; Guenther et al., 1998, 2006). More recently hypotheses regarding the neuroanatomical locations of various processing components and representations in the model have been developed and published (Guenther et al., 2006; Guenther, 2006). Currently, however, the model does not treat sequencing or explicit planning beyond a single “chunk.” This experiment investigated the neural substrates for representing speech items (and their serial order) within planned sequences, and for

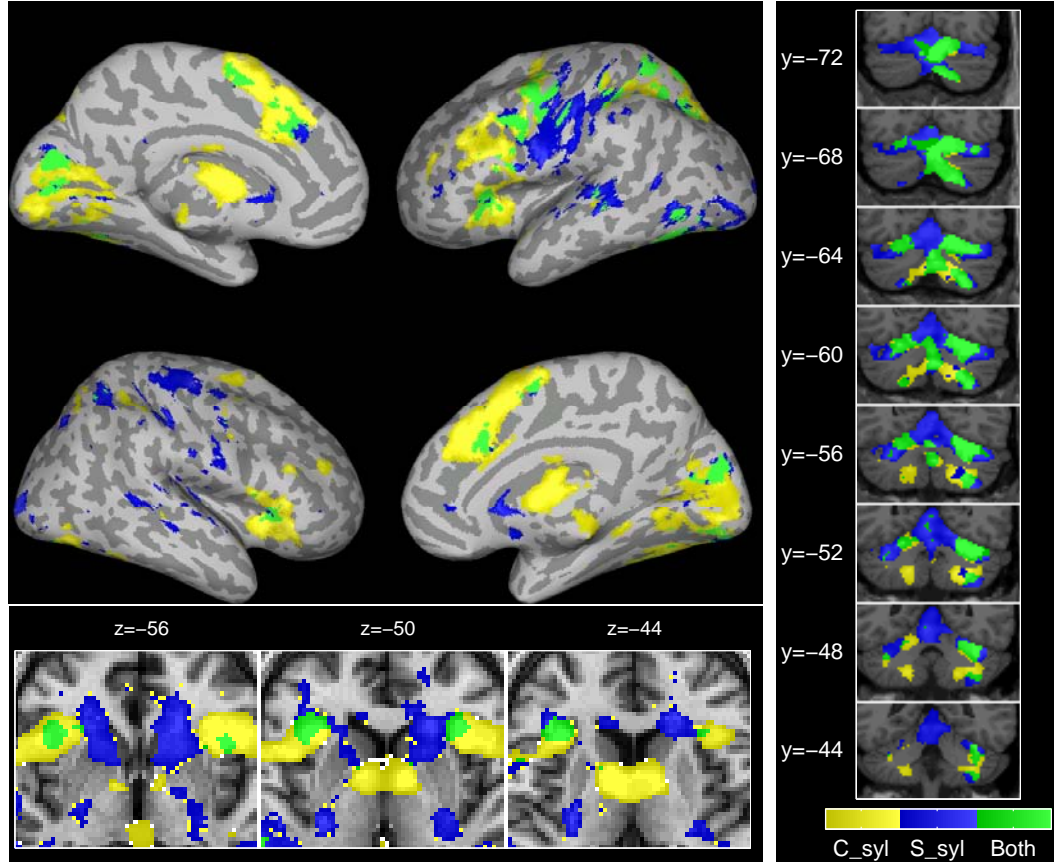


Figure 2-8: Effect of sequence complexity by syllable type during GO trials ($P < 0.01$ uncorrected). Blue patches show the effects of sequence complexity for simple syllables, yellow patches for complex syllables, and green the intersection (conjunction) between syllable types. *Left Top:* Significant activity rendered on semi-inflated cortical surface. Dark gray cortical areas represent sulci, lighter gray areas are gyri. *Left Bottom:* Significant activity rendered on axial slices through the basal ganglia and thalamus at various depths. *Right:* Significant activations rendered on coronal slices through the cerebellum at various depths. y- and z-values refer to planes in MNI-space. Anatomical images are cropped versions of the canonical SPM T1 image, and follow neurological conventions (Left hemisphere on the left side of image); The color scale is common to each sub-figure.

initiating and coordinating the serial production of these items (e.g. Lashley’s *action syntax* problem; Lashley, 1951).

Subjects spoke or prepared to speak non-word sequences of three syllables. The use of non-lexical items served to eliminate semantic effects, which were not a focus of interest in this study⁶. Because related modeling work is not tied to a particular level of phonological representation (the current DIVA implementation is capable of learning phonemes, syllables, or multi-syllabic words), and because the research community has not arrived at a consensus on planning “units” in speech, the stimuli were parameterized by two complexity factors: within each syllable (syllable complexity or *syl*) and across the syllables in the sequence (sequence complexity or *seq*). Many previous authors have considered the importance of the syllable as a unit in speech production (Sevald et al., 1995; Ferrand and Segui, 1998; Ziegler and Maassen, 2004; Cholin et al., 2006), and in the present study the presentation of stimuli as three one-syllable items separated by hyphens likely encouraged participants to treat syllables as chunks (see for example Klapp, 2003, who demonstrated a similar chunking effect dependent on how the stimuli were structured). Although syllable-sized units are probably involved at some level(s) of the speech planning process, the relevance of phonemic units is also supported by slips of the tongue, phonemic paraphasias, and deficits in disorders such as apraxia of speech. “Slots and fillers” (Shattuck-Hufnagel, 1979, 1983, 1987) or “Frame and Content” (MacNeilage, 1998) theories of speech production postulate that syllables and the phonemes which comprise them may have separate representations. In such proposals, the abstract syllable frame often serves to indicate the eligibility of phonemes in particular “slots” or serial positions. Such models have been particularly useful in addressing speech error data.

⁶It has been suggested (Gupta et al., 2005), however, that non-words repetition and word list recall may share common sequencing mechanisms. The use of non-words was intended to simplify possible interpretations of the experimental results and still sheds light on mechanisms involved in more typical language production.

In the $2 \times 2 \times 2$ factorial analysis performed here, the complexity-related effects have important interpretations in understanding the representations of forthcoming speech plans. A main effect of *seq* was observed when a region showed a greater response due to the demands of representing three unique syllables compared to just one. Increasing sequence complexity also *necessarily* led to an increase in the number of unique sub-syllabic targets. A main effect of *syl* occurred when a region’s response increased due to the demands for representing sub-syllabic complexity *at the level of a single syllable*. Because the syllable complexity comparison was made without regard for sequence complexity, it does not always reflect the necessity to plan more articulatory targets over the entire forthcoming utterance; instead it is always true that increasing *syl* increases the structural complexity of the individual syllable-sized items being planned. A *seq* \times *syl* interaction occurred when increasing sequence or syllable complexity increased the size of the effect of the other factor (e.g. if the effect of sequence complexity was greater when the syllabic items were complex).

The experimental protocol used was different in several ways from most other neuroimaging studies of speech production. First, the utilization of a sparse scanning procedure (see also Eden et al., 1999; Birn et al., 2004; Schmithorst and Holland, 2004; Nebel et al., 2005) that took advantage of the hemodynamic delay enabled the use of overt speech production while avoiding movement-related artifacts (Birn et al., 1998; Barch et al., 1999), and allowed subjects to produce utterances in relative silence. Other authors have dealt with movement artifacts by excluding images obtained during articulation from their analyses (e.g. Riecker et al., 2002), but this approach still requires subjects to speak with loud background noise due to the scanner gradients. While such important issues associated with imaging overt speech have been discussed in the literature (Munhall, 2001; Gracco et al., 2005), they are often disregarded due to technical limitations or other priorities (although see

de Zubizaray et al., 2000 and Abrahams et al., 2003) Also, in the present design, stimuli were drawn randomly from different conditions in each trial, eliminating adaptation and habituation effects that can occur with blocked presentation. Finally, the inclusion of a random-duration wait period between stimulus presentation and the GO signal enabled the imaging of pre-articulatory preparation for speech as well as the articulation period without cueing the subject about the trial type beforehand. This design is similar to simple reaction time tasks (e.g. Sternberg et al., 1978; Klapp, 2003) as well as electrophysiological studies of motor sequence performance in non-human primates (e.g. Shima and Tanji, 2000; Lu and Ashe, 2005). In the latter studies, cells in many regions of the frontal cortex show anticipatory activity related to the forthcoming sequence during the wait period. Here fMRI was used in an attempt to measure analogous responses in the wait period prior to articulation of syllable sequences.

While the NOGO task used in this experiment shares common elements with *covert* speech, it is not equivalent to that task, which has been used in many speech imaging studies. In our task, there is no explicit instruction other than to “be prepared to immediately speak” the most recently presented sequence upon viewing a GO signal. It was assumed that subjects use the stimulus display as a “precue,” loading the sequence into a working memory buffer prior to the arrival of the GO signal. This notion is supported by the classical finding in reaction time studies that choice reaction time (in which the GO signal itself informs the subject of the stimulus) is longer than simple reaction time (in which the precue provides the stimulus, as in the present study; Donders, 1969).

The minimal network used for producing syllable sequences was assessed by performing a conjunction analysis (Nichols et al., 2005) between the four individual speaking conditions compared to the baseline. This method based on the maximum

P -statistic provides a conservative estimate (Friston et al., 2005) of the speech production system (see Figure 2·3). Overt production of syllable sequences of all types resulted in significant activation that extended beyond the central sulcus, involving also the medial premotor areas, the frontal operculum and anterior insula, the anterior thalamus, and the cerebellum. The only differences between speaking conditions were in the phonological composition of the sequences. Very generally, we observed that increasing the complexity of the stimulus led to additional activity in this minimal speech production network and beyond. Average utterance durations varied moderately but significantly across conditions. Although these differences could, themselves, lead to variable brain responses, one would expect duration-specific responses to be focused in the primary sensorimotor and auditory regions. Differences observed across conditions in “higher-order” regions are unlikely to have been a simple effect of speaking duration.

The results observed here conflict with the findings of Riecker et al. (2000b), who examined the effects of articulatory/phonetic complexity on the speech production system. In that study, none of the stimuli elicited significant activation of the anterior insula, frontal operculum, or SMA, and only production of complex syllables (using the terminology adopted herein) activated the cerebellum. There were several differences between experimental designs. In Riecker et al. (2000b), stimuli were spoken repeatedly (at syllable production rates between approximately 1 and 2 Hz) for one minute periods. For single syllables, this amounted to simple repetitions over the full minute; for the multi-syllabic utterances, subjects attempted to equally space the individual syllables at the same rate as the single syllable stimuli, and repeated the set of three until the minute was complete. In our protocol, a sequence was presented then removed during a delay period, forcing subjects to load the sequence into a working memory buffer in anticipation of the GO signal. A three

syllable utterance was prepared and/or produced just once in a trial, and the next trial involved a new stimulus. In a previous study in our laboratory (Ghosh et al., 2003), production of even simple vowel sounds activated areas beyond those observed in Riecker et al. (2000b); furthermore, in that experiment, syllables were produced immediately upon visual presentation, so the activation of those areas cannot be merely attributed to the verbal working memory requirements in the present study. The limited activation patterns for complex speech stimuli in Riecker et al. (2000b) may result from the blocked paradigm used. The authors' suggestion that polysyllable tokens might be organized as higher-order units posing fewer demands on the motor system seems unlikely. In English, for example, there are approximately 500 very commonly used syllables. If arbitrary non-lexical combinations of these syllables were stored as higher-level motor memories, this would result in an unlikely combinatorial explosion. Rather, as Lashley (1951) noted, the human brain must be able to arrange smooth sequences of behavior from a finite alphabet of learned acts. The additional activations observed in the present study due to increasing stimulus complexity supports the notion that these utterances were “assembled” and not simply executed from a single motor memory.

The basic speech production network observed is in general agreement (although activated regions differ depending on the precise conditions and baselines used) with many other studies of overt production of various speech stimuli (Murphy et al., 1997; Wise et al., 1999; Riecker et al., 2000a; Fiez, 2001; Blank et al., 2002; Riecker et al., 2002; Shuster and Lemieux, 2005; Riecker et al., 2005, see also Indefrey and Levelt, 2000 and Turkeltaub et al., 2002 for meta-analyses of word production experiments). Many of the regions within and beyond the minimal speech production network (Figure 2.3) showed complexity-related response variations. Our results show that sequence and syllable complexity interacted strongly in many of the regions in which

a main effect of *seq* was observed. This is likely due to the hierarchical relationship between syllables and the phonemes or phonetic targets that comprise them. In this study, a complex sequence of simple syllables (e.g. ka-ru-ti) could contain up to four more distinct phonetic targets than a simple sequence of simple syllables (e.g. ka-ka-ka), whereas a complex sequence of complex syllables (e.g. kla-tri-splu) could contain up to eight more targets than a simple sequence of complex syllables (e.g. kla-kla-kla). Thus the two factors were inherently intertwined, and an interaction would be anticipated if a region represented the full forthcoming speech plan at a sub-syllabic level, or if the representation of complex syllables was simply larger (e.g. greater BOLD response). In assessing the main effect of *seq*, complex sequences were compared to simple ones regardless of the complexity of the individual syllables within. While more *syllables* had to be represented for complex sequences, subjects also had to plan more sub-syllabic targets because these stimuli *always* contained more unique phonemes than did simple sequence stimuli. If an area showed a main effect for *seq*, but did *not* show an interaction between *seq* and *syl*, this would indicate that the area likely was used to represent “chunks” without regard for the complexity of the chunk. In the present study, the only region that showed the main effect of *seq* but did not also show the $seq \times syl$ interaction was the right inferior cerebellum (Lobule VIII). The fact that the remaining regions showing a main effect for *seq* also showed a $seq \times syl$ interaction is informative because it indicates that in most portions of the speech planning system, sub-syllabic detail plays an important representative role.

A major motivation for this study was to provide additional constraints for models of the speech production system. The following sections discuss the patterns of responses obtained for various anatomical structures, review previous pertinent data, and develop hypotheses concerning how these structures may each contribute to the

planning and production of sequences of syllables and, moreover, fluent speech.

2.4.1 Sensorimotor areas

Overt production of all stimulus types resulted in significant bilateral activation (compared to baseline) of the primary sensorimotor areas in and surrounding the central sulcus. These areas showed a main effect for *go*, indicating that they were, on average, more active for performance than for preparation. In both comparisons, the activity maps roughly follow the motor / sensory homunculus representations of the lips, jaw, tongue, and larynx (see Guenther et al., 2006, for a review of the estimated anatomical locations of the components of the speech motor system). These results suggest that the primary motor and somatosensory cortices, bilaterally, are engaged in the online control of the articulators and registration of orosensory feedback. This result was, of course, expected since sensorimotor cortical activity is seen in all neuroimaging studies involving articulated speech.

The degree to which activation in these areas is lateralized for speech is of interest. Significant left lateralization at the level of the precentral gyrus has previously been demonstrated for *covert* speech (Wildgruber et al., 1996; Riecker et al., 2000a). Riecker et al. (2000a) found bilateral activation (with moderate left-hemisphere bias) when the speaking task was made overt. In the present study, a similar lateralization of motor cortex activity was observed for the preparation-only trials. ROI analyses revealed significant ($P < 0.05$) left lateralization in the ventral motor cortex during NOGO trials. For GO trials, this region's activation was on average stronger in the left, but this trend was not significant. The effect of (*seq*) was also significantly stronger in the left hemisphere ventral motor and ventral premotor cortices. These results, coupled with the previous observations for covert speech, suggest a special role for the left hemisphere motor cortex. It is hypothesized that preparation for

speaking “primes” motor cortical cells primarily in the *left* hemisphere that drive execution of learned motor programs, but that the right hemisphere motor cortex becomes active when overt speech is initiated in order to more generally aid in the online control of the articulators.

2.4.2 Left hemisphere prefrontal areas

In this study a strong left-lateralized response to additional sequence complexity (see Figure 2.5) was observed in the left precentral gyrus and prefrontal cortex along the inferior frontal sulcus. The left IFS region also showed a strong positive interaction effect between *seq* and *syl* (see Figure 2.7). In other words, the BOLD response near the IFS was greater for complex vs. simple sequences, but the amount of the additional signal was larger when the sequences were composed of complex syllables. This conclusion is also supported by the effect of sequence complexity, evaluated individually for each syllable type (see Figure 2.8), which showed greater magnitude and extent of activation in left prefrontal areas when the sequences were composed of complex syllables. This region did not show a main effect of *syl*.

The lateral prefrontal cortex has been implicated in a great number of studies of language and working memory (Gabrieli et al., 1998; Kerns et al., 2004; Fiez et al., 1996; D’Esposito et al., 1998) and in serial order processing (Petrides, 1991; Averbeck et al., 2002, 2003). The complexity-related activity observed here is near the human homologue of a region that Averbeck et al. (2002, 2003) recorded from (BA46) while monkeys planned serial drawing movements. Averbeck et al. (2002) demonstrated that prior to initiating a planned sequence of movements, there existed a parallel co-active representation of each of the components of the forthcoming sequence. The relative activity level in groups of cells that coded for the component movements corresponded to the order in which the movements would be produced. Based on

the results of the present study, I hypothesize that planning memory-guided syllable sequences also necessitates such a parallel representation; coding for three distinct syllable “chunks” requires more neural and metabolic resources than coding for a sequence that contains only one syllable “chunk” repeated three times. Specifically, a standing parallel representation of the forthcoming utterance is suggested to be located in or near the inferior frontal sulcus. The presence of a strong interaction between *seq* and *syl* suggests that complex syllables may require the activation of multiple phonological units in the inferior frontal sulcus, or that complex or less frequently utilized syllables have a larger representation in this area than simple syllables.

An alternative hypothesis regarding IFS activity was proposed by Crosson et al. (2001) who found that, in an inner speech task, IFS activity was modulated by the amount of semantic processing required. The authors speculated that the IFS is involved in word generation from semantic cues. In a follow-up study of covert word generation, Crosson et al. (2003) found left IFS activity only when word generation required the use of semantic knowledge. In the present study IFS activity was modulated by the phonological composition of non-lexical syllable sequences. The stimuli were designed to remove semantic effects completely but IFS activation and stimulus-dependent modulation was still observed. This suggests that this region, at least in part, plays a non-semantic role in representing speech plans.

Activity was also observed in the present study within the left posterior inferior frontal gyrus *pars opercularis* (BA44) and neighboring premotor areas related to *seq*. In previous work this area (in the left hemisphere) has been associated with the *Speech Sound Map* component of the DIVA model (Guenther et al., 2006). The effect of *seq* in both the ventral premotor cortex and the inferior frontal gyrus *pars opercularis* was significantly greater in the left hemisphere. A prediction of the model,

which suggests that Speech Sound Map cells read out motor plans for well-learned speech “chunks,” is that there should be additional activity when multiple chunks are activated. Because production of complex sequences requires the activation of multiple speech sound map cells, one would expect to observe additional activity with BOLD fMRI, thus accounting for the complexity-related activation of posterior BA44 observed here.

2.4.3 Anterior insula and frontal operculum

Recently the role of the anterior insula in speech production has received great attention (Dronkers, 1996; Wise et al., 1999; Nagao et al., 1999; Ackermann and Riecker, 2004; Hillis et al., 2005). Dronkers (1996) identified the precentral gyrus of the left-hemisphere insula as the common site of lesion overlap in a group of patients diagnosed with apraxia of speech; this region was preserved in an aphasic control group without AOS. Wise et al. (1999) found a similar region involved in articulated but not covert speech. In this study activation was observed in or near the precentral gyrus of the insula in *both hemispheres* during all GO conditions (Figure 2-3); these areas were not significantly active for NOGO trials, and did not show significant effects for the factors *seq* or *syl*. This portion of the anterior insula, believed to be analogous to that found by Wise et al. (1999), is, therefore, suggested to be engaged only for the overt production of speech and is not explicitly involved in sequence representation. The involvement of the right anterior insula in overt speech is somewhat surprising (cf. Wise et al., 1999; Riecker et al., 2000a). Ackermann and Riecker (2004) suggested that the left and right insula might act on different time scales in vocal control; this study involved supra-segmental sequences, but subjects were specifically instructed to avoid prosodic modulation, which has been attributed to right hemisphere structures. It is possible that in previous experiments the right in-

sula was involved but failed to reach significance and/or the present use of non-lexical stimuli may have further engaged the right hemisphere.

Nota and Honda (2003) hypothesized that the anterior insula may be involved in encoding and buffering phonetic plans for articulation. This suggestion was based on results showing insular involvement when the spoken utterance was changed randomly throughout the session but not when the same utterance was repeatedly spoken. The present result, that the precentral gyrus region of the insula was active in all GO trials, is consistent with this suggestion because stimuli were chosen randomly per trial, and thus subjects always needed to “reload” the speech plan. The lack of a complexity effect, however, suggests that it is unlikely to play a role in the representation of the phonological / phonetic plan. Furthermore, this area became active due to *overt* speech, not merely by reloading a speech plan as in the NOGO trials. Insular damage has previously been found to lead to deficits in speech initiation (Shuren, 1993) and motivation to speak (Habib et al., 1995). Based on our results, this portion of the insula is more likely involved in these functions than in speech encoding or sequence buffering.

A separate focus of activity, at the junction of the anterior insula and frontal operculum bilaterally, showed a consistent activation pattern that was quite different from that discussed above. Increased responses were observed for additional sequence or syllable complexity. This area also showed a strong interaction between *seq* and *syl* and showed no significant difference for GO vs. NOGO trials. It is likely, therefore, that this region is involved in representation of the speech plan at some level. It may perhaps serve as a substrate for the integration of lower-level aspects of the speech motor plan with more abstract representations of speech sounds used in sequence planning. In addition to providing the proper speech units to the motor apparatus at appropriate times, a system for organizing fluent speech must also integrate affective

and linguistic prosody, for example. The anterior insula is well connected with the medial premotor areas and the temporal and parietal lobes, and gives projections to the frontal operculum as well as the prefrontal cortex (Augustine, 1996; Flynn et al., 1999). It is therefore in a position to provide contextual information to the *speech sound map* allowing flexible production of learned motor programs. This notion is similar to one discussed by Van der Merwe (1997) who likened motor programs to computer subroutines, which can be supplied with parameters by other parts of the speech / language system. Alternatively, this region may be a portion of the speech sound map itself.

2.4.4 Temporal and parietal areas

The temporal lobe activity observed in this study can be primarily attributed to subjects hearing their own voices while speaking. Compared with the baseline, the overt speaking (GO) conditions conjointly activated bilateral areas along the supratemporal plane, including Heschl’s gyrus and planum temporale, as well as the posterior superior temporal gyrus. Each of these areas also was significantly more active for GO trials than for NOGO trials, and none showed effects for the other factors.

A region in the parietal lobe along the intraparietal sulcus near the junction with the postcentral sulcus responded to additional complexity, demonstrating effects for *seq* and *syl*, and a *seq*×*syl* interaction. These effects were significantly stronger in the left hemisphere. This area was not a part of the minimal network required for performance of any of the sequence types (see Figure 2.3), but did become active (compared to the baseline condition) for complex sequences (Table 2.2). No significant differences were found between GO and NOGO trials. The intraparietal sulcus divides the superior parietal lobule (BA 7) and the supramarginal gyrus (BA 40). The latter area has been associated with the “phonological store” portion of Badde-

ley’s (1986) phonological loop model (Paulesu et al., 1993; Awh et al., 1996; Jonides et al., 1998); in Baddeley’s model this module contains phonological representations which can be temporarily activated by incoming verbal information. Henson et al. (2000) found activity in BA 7 and BA 40 (near the focus of activation in this study) when comparing a delayed matching task involving letters to one involving non-verbal symbols. They suggest that these areas participate in phonological recoding of visually presented verbal materials. Crottaz-Herbette et al. (2004) found nearby areas along the left intraparietal sulcus to be more active in a verbal working memory task when stimuli were presented visually than when they were presented auditorily.

These results suggest that the activation of primarily left hemisphere parietal areas in this study is likely related to the translation of the orthographic display of the stimuli into manipulable phonological codes used in speech planning. Because stimuli of increasing complexity at both the syllable and sequence level would presumably require further encoding, the complexity effects in these areas are naturally accounted for. The absence of a main effect for *go* indicates that this activity is not significantly augmented during production. This makes sense if the activation is due to orthographic to phonological translation, which can be performed immediately upon stimulus presentation in both GO and NOGO trials.

2.4.5 Medial premotor areas

The role of the SMA in speech production has been studied since stimulation experiments in patients by Penfield and colleagues elicited speech arrest or prolongation of vowel sounds (Penfield and Welch, 1951; Penfield and Roberts, 1959). Many studies have shown that the medial aspect of Brodmann’s Area 6 comprises at least two sub-regions that can be distinguished on the basis of cytoarchitecture, connectivity, and function: the pre-SMA located rostral to the vertical line passing through the

anterior commissure (VCA line), and the SMA-proper located caudally (Picard and Strick, 1996). Additional motor-related zones also lie in the anterior portions of the cingulate sulcus (BA32) and have been associated with complex movements (Picard and Strick, 1996). Although most lesion and brain imaging studies have failed to delineate these regions, Tanji and colleagues have collected a wealth of data in monkeys that suggest that the SMA and pre-SMA are both crucially involved in the representation of movement sequences, with the pre-SMA likely serving a higher-order role than the SMA (Matsuzaka et al., 1992; Tanji and Shima, 1994; Shima et al., 1996; Shima and Tanji, 1998, 2000; Tanji, 2001). The two regions have different patterns of connectivity with cortical and subcortical areas in monkeys (Jürgens, 1984; Lupino et al., 1993), and diffusion tensor imaging results verify disparate connections in humans (Johansen-Berg et al., 2004; Lehericy et al., 2004). While the pre-SMA is well-connected with the prefrontal cortices and the anterior striatum, the SMA is more connected with the motor cortex and the posterior striatum. This suggests a role more generally associated with planning for the pre-SMA and with motor performance for the SMA.

Various case studies of speech emission in patients with SMA lesions have been described in the literature (Jonas, 1981, 1987; Ziegler et al., 1997; Pai, 1999). Following a transient period of total mutism, patients generally suffer from a reduced propositional (self-initiated) speech with non-propositional speech (automatic speech; e.g. counting, repeating words) nearly intact. Such a deficit is often termed *transcortical motor aphasia*. Other problems include involuntary vocalizations, repetitions, paraphasias, echolalia, lack of prosodic variation, stuttering-like behavior, and variable speech rate, with only rare occurrences of distorted articulations. Micro-stimulation in humans (Penfield and Welch, 1951; Fried et al., 1991) has yielded vocalization, repetitions of words or syllables, speech arrest, slowing of speech, or hesitancy. Jonas

(1987) and Ziegler et al. (1997) arrived at similar conclusions regarding the role of the SMA in speech production, suggesting that it aids in sequencing and initiating speech sounds, but probably not in determining their content. This conclusion is consistent with the *Frame-Content Theory* of speech production (MacNeilage, 1998), which assigns motor control of the “frame” to the medial areas and determination of “content” to the lateral areas. These proposals do not, however, delineate separate roles for the pre-SMA and SMA, despite evidence for distinct roles in sequential motor control.

In this study portions of the SMA, pre-SMA, and cingulate motor areas were activated in all speaking conditions (Figure 2-3, Table 2.2). The “SMA-proper” activity was primarily located very near the VCA line (consistent with somatotopic representation of the face; Fried et al. 1991; Picard and Strick 1996). The main effect of *go* primarily involved the SMA-proper (Figure 2-4). Consistent with electrophysiological studies, it is hypothesized that this portion of the medial wall is responsible, in part, for properly-timed *initiation* of an overt production. This may occur through known projections to the motor cortex, basal ganglia, or anterior insula/frontal opercular regions (Jürgens, 1984; Luppino et al., 1993). In region-level analyses, the SMA only showed a main effect for *go* and not for *seq* or *syl*. This further supports the proposal that the SMA-proper is related more to initiation of speech production than to planning.

The pre-SMA showed an effect for *go*, but also showed strong effects for *seq* and *syl* as well as an interaction between the two factors. Shima and Tanji (2000) showed that the pre-SMA contains cells that code for an entire sequence to be produced. If the separation of syllabic frames and phonemic content (e.g. MacNeilage, 1998; Shattuck-Hufnagel, 1983) is realized in the brain, then a possible role for the anterior pre-SMA is to represent syllable or word-sized frames, and to coordinate serial position / timing

signals with the motor apparatus via the SMA. The pre-SMA was one of a small set of regions (relative to those showing effects of *seq*) that demonstrated a main effect of *syl*; this indicates that it was more active when the structure of individual syllables in the speech plan was complex regardless of the complexity of the overall sequence. This would be expected if complex syllable frames necessitate larger representations than simple frames. These results are also consistent with the suggestion of Krainik et al. (2003), that there is a “rostrocaudal shift,” whereby the SMA is associated with vocal sound production and the pre-SMA with “complex verbal demands.”

2.4.6 Cerebellum

Across all stimulus types, overt production of speech sequences activated the superior cerebellar hemispheres (Lobule VI, Crus I) bilaterally, and the right inferior cerebellar cortex (Lobule VIII). Speech deficits due to 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 characterized by inaccurate articulation, prosodic excess, and phonatory-prosodic insufficiency (Darley et al., 1975). Cerebellar damage also results in increased duration of sentences, words, syllables, and phonemes (Kent et al., 1997; Ackermann and Hertrich, 1994). It is also implicated in the control of motor sequences (Inhoff et al., 1989), possibly in translating a discrete programmed sequence into fluent motor action (Braitenberg et al., 1997; Ackermann et al., 2004). Damage to the cerebellum may additionally lead to deficits in short-term verbal rehearsal and planning for speech production (Silveri et al., 1998).

Portions of superior Lobule VI were more active bilaterally during production than during preparation (Figure 2.4). Grodd et al. (2001) localized activation during lip pursing and vertical tongue movements to nearby parts of lobule VI. Activation

in right inferior Lobule VIII was also significantly greater at the voxel-level but not at the combined voxel-cluster level. It can be hypothesized that the superior regions are particularly involved in ongoing control of the articulators through crossed thalamo-cortical projections to the motor cortex and/or direct connections with the periphery. This is consistent with the notion that superior cerebellar artery stroke causes dysarthria. Additional syllable complexity caused greater activity in the right superior cerebellar cortex (Lobule VI; see Figure 2.6), posterior to the differences observed for the main effect of *go*. Riecker et al. (2000b) also found activation of right hemisphere Lobule VI for repetitions of the syllable “stra” but not for “ta,” suggesting that articulation of consonant clusters engages this region. Wildgruber et al. (2001) also suggested a special role for this cerebellar region for speaking in “time-critical conditions.” The cerebellum is implicated in adaptively timed motor responses (e.g. Perrett et al., 1993); these adaptive timing mechanisms centered in the superior cerebellum may be used for feedforward control and anticipatory coarticulation in speech production (e.g. Guenther et al., 2006). The alternative possibility, that superior cerebellar activations were related to auditory perception of one’s own voice, however, can not be ruled out; similar areas have been reported to be related to speech and auditory perception (Callan et al., 2004; Petacchi et al., 2005).

Both the superior and inferior cerebellum showed responses related to *seq* (Figure 2.5). The inferior focus was right-lateralized, did not show a main effect for *syl*, and did not show a *seq* × *syl* interaction effect. The superior portions, also moderately right-lateralized, extended more laterally than the focus related to syllable complexity, which corresponds to the general notion that more lateral portions of the cerebellum are involved in higher-order processes compared to more medial regions (e.g. Leiner et al., 1993). In the right hemisphere, lateral superior regions also

showed a *seq* × *syl* interaction. The right hemisphere cerebellar bias paralleled the left hemisphere fronto-cortical bias observed for sequence complexity (Figure 2·5). Both the superior lateral and inferior cerebellar regions demonstrating complexity effects are in close proximity to regions studied by Desmond and colleagues (Desmond et al., 1997; Chen and Desmond, 2005; Kirschen et al., 2005). Desmond et al. (1997) showed that both a superior lateral portion (corresponding to Lobule VI/Crus I as in the present study), and an inferior portion of the cerebellum (right lateralized Lobule VIIB, just lateral to our observations) showed load-dependent activations in a working memory task, but only the superior portions showed load-dependent effects in a motoric rehearsal task that lacked working memory storage requirements. Chen and Desmond (2005) extended these results to suggest that Lobule VI/Crus I works in concert with frontal regions for mental rehearsal, and that Lobule VIIB works in concert with the parietal lobe (BA40) as a phonological memory store. This division of labor is reasonable in the context of our current experiment which involved a phonological storage component that might engage the same network that Chen and Desmond (2005) suggest. No syllable complexity effects or interactions in the inferior region were observed, which may indicate that this system works with abstract chunks without regard for their complexity.

In summary, it is suggested that the right inferior cerebellum, perhaps in concert with the left parietal lobe, was used to maintain a chunk-based working memory of the to-be-spoken utterance in this experiment. The lateral superior aspects contribute to sequence organization in both sub-vocal rehearsal and overt production. The superior regions near the vermis (Figure 2·4) are more closely related to motor execution.

2.4.7 Basal ganglia and thalamus

Frontal cortical areas form the input to multiple cortico-striato-thalamo-cortical loops (Alexander et al., 1986; Alexander and Crutcher, 1990; Middleton and Strick, 2000). It has been proposed that the architecture of the basal ganglia make these loops suitable for selectively enabling one output from a set of competing alternatives (Mink and Thach, 1993; Mink, 1996; Kropotov and Etlinger, 1999; Brown et al., 2004). During action sequence performance the selection of a single component movement (or syllable) from a parallel sequence plan requires this type of mechanism. Pickett et al. (1998) reported the case of a woman with bilateral damage to the putamen and head of the caudate nucleus. She suffered from an articulatory sequencing deficit, with a particular inability to rapidly switch from one articulatory target to the next, consistent with a basal ganglia role for selecting movements in a sequence.

In the present study overt production increased activation of the putamen bilaterally. This coincided with additional motor cortical activation and likely represents a portion of the motor executive loop. Additional sequence complexity led to an increased activation in the anterior thalamus and/or the caudate nucleus. These areas also showed a $seq \times syl$ interaction, indicating that the phonological makeup of the items in the sequence modulated this additional activation. The anterior thalamus, however, showed no main effect of syl , suggesting that it was not the complexity of individual items that engaged this region, but rather the complexity of the overall speech plan. Crosson (1992) previously made note of the similarities between electrical stimulation effects in the caudate nucleus and anterior thalamic nuclei. Schaltenbrand (1975) reported that stimulation of the anterior nuclei of the thalamus sometimes caused compulsory speech that could not be inhibited. Stimulation of the dominant head of the caudate has also evoked word production (Van

Buren, 1963), and Crosson (1992) describes the similarities in the language evoked from stimulation of the two areas as “striking.” This suggests that the areas serve similar functions, and that they are involved in the release of a speech / language plan. A comparison of the effects of *seq* for each syllable type (available in online supplementary materials) indicated a possible different focus of activation based on syllable type that warrants further study.

2.4.8 Sequencing and the FOXP2 gene

Approximately half of the members of the three-generation ‘KE’ family suffer from a severe speech and language disorder (Hurst et al., 1990; Gropnik and Cargo, 1990). This family has recently been the subject of copious study in the research community, which has led to the identification of FOXP2, which is mutated in affected family members, as the first gene known to be involved in the development of speech and language capabilities (Lai et al., 2001). The syndrome that affects members of the KE family is characterized by developmental verbal dyspraxia and other deficits in language processing and grammatical skills. Watkins et al. (2002a) administered a battery of linguistic and non-linguistic examinations in an attempt to establish a behavioral phenotype for this disorder. They found that affected members of the KE family could be successfully discerned from non-affected members according to a test of repetition of non-words with complex articulations (containing consonant clusters), ranging in length from one to four syllables. The effects of the disorder worsened with the number of syllables. Watkins et al. (2002b) described the disorder as “best characterized as a deficit in the sequencing of articulation patterns rendering speech sometimes agrammatical and often unintelligible” (p. 466).

A voxel-based morphometric analysis using MRI has been performed in order to compare regional grey matter volumes in affected family members to those in

unaffected members and age-matched controls (Watkins et al., 2002b). Considering the difficulty that affected family members have with a multi-syllabic non-word repetition task and that task’s resemblance to the task investigated herein, it was hypothesized that the regions showing morphological abnormalities would at least partially overlap with regions related to sequence complexity in the present study. The similarities are indeed compelling. The published coordinates (Tables 1 and 2 in Watkins et al., 2002b) of regions showing a significant difference in grey matter volume between affected family members and (i) age- and gender-matched controls and (ii) unaffected family members were used to create cortical and cerebellar renderings using the same methods described above (Figure 2·9). These renderings plot regions of activity that were created by centering a 3-D isotropic Gaussian with FWHM of 12 mm at each of the published coordinates. A comparison of Figure 2·9 with the maps of areas showing a significant main effect for sequence complexity (Figure 2·5) and/or an interaction between sequence and syllable complexity (Figure 2·7) is of interest. Areas that showed abnormal grey matter volume for affected family members versus either of the control groups and also demonstrated *seq* and/or *seq* \times *syl* complexity effects in the present study were in the prefrontal cortex along the inferior frontal sulcus (two regions), in the left SMA, at the junction of the anterior insula and frontal operculum, in the caudate, and the right inferior cerebellum. Although further investigation is necessary, it is possible that members of the KE family who suffer speech disturbances due to this inherited speech and language disorder may do so, in part, due to structural abnormalities in a sequencing circuit for speech production as revealed in the present study.

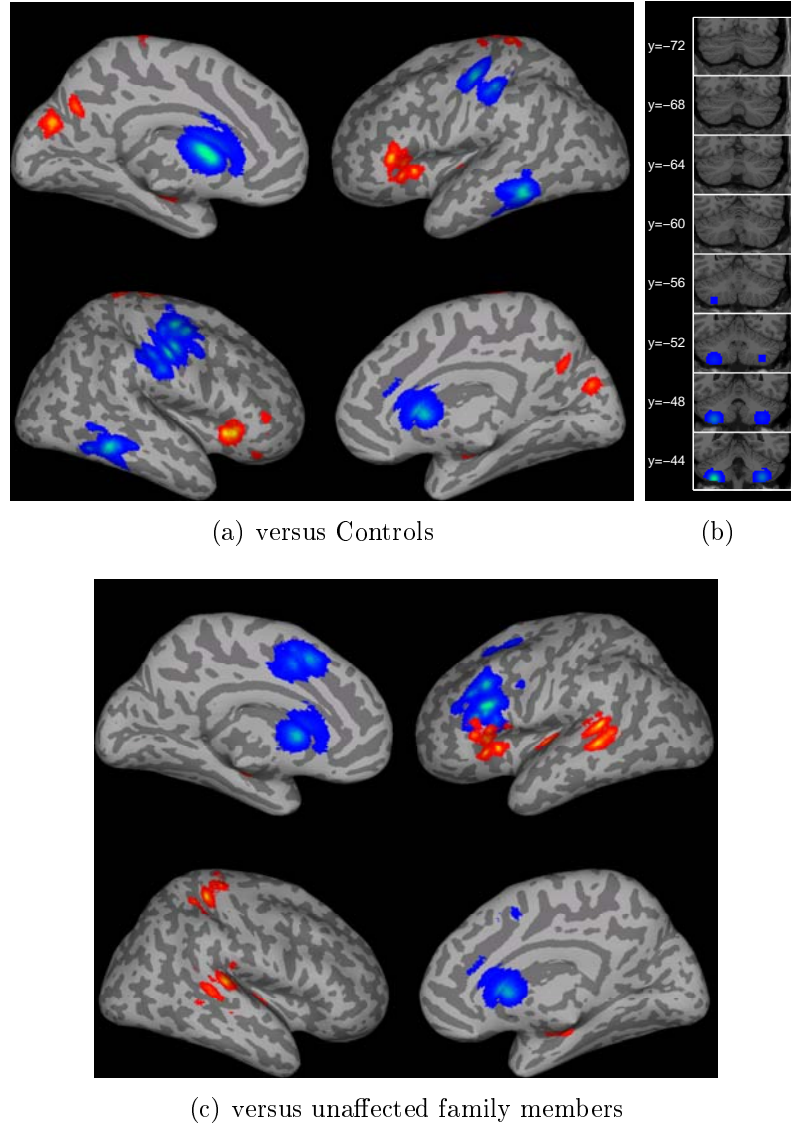


Figure 2.9: Renderings of brain areas found to have significantly different grey matter density in a voxel-based morphometric analysis (Watkins et al., 2002b) between affected KE family members and (Panels a and b) age- and sex-matched controls, and (Panel c) unaffected family members. Renderings were created by applying a Gaussian spread with FWHM of 12 mm centered at the coordinates specified in Tables 1 and 2 of Watkins et al. (2002b), then rendering the figures with the same methods used for the results of the present study. 'Hot' colors indicate regions where affected family members had more grey matter volume than the other group; 'cool' colors indicate less grey matter volume in affected family members.

2.4.9 Conclusions

The basic experimental hypothesis prior to this investigation was that both added *sequence complexity* and *syllable complexity* would further engage the speech production system and recruit areas beyond the primary sensorimotor cortices known to be involved in non-speech motor sequencing. The results confirmed this hypothesis, showing areas of the left hemisphere including the inferior frontal sulcus and the posterior parietal cortex, as well as bilateral regions in the anterior insula and frontal operculum, the basal ganglia, thalamus, and cerebellum to be further engaged by additional stimulus complexity. A strong interaction was found between the two types of complexity studied, and the areas showing this interaction largely overlapped with areas showing a main effect of *seq*. This suggested that sub-syllabic information was important in many areas involved with representing a forthcoming speech sequence. A much more limited set of areas showed the main effect of *syl*; these areas are hypothesized to be especially concerned with the structural complexity of individual syllables in the sequence. This study provides a wealth of data regarding sequential organization in speech production, though further experiments are necessary to test functional hypotheses and guide construction of a more comprehensive model of speech production.

CHAPTER 3

A COMPUTATIONAL NEURAL MODEL OF SPEECH SEQUENCE PLANNING AND PRODUCTION

3.1 Introduction

In this chapter a computational neural model is presented which describes how the brain represents and enables the production of sequences of simple, learned speech sounds. In particular, this model addresses the question of how, using a finite inventory of learned speech motor units, a speaker can produce seemingly arbitrary utterances that fall within the phonotactic and linguistic rules of his or her language. This modeling study develops the *phonological* level of representation, implementing a pair of complementary subsystems corresponding to the *structure* and *content* of planned speech utterances in a neurobiologically realistic architecture that models cortical and subcortical structures and their interactions. This phonological level of representation is hypothesized to serve as an interface between the higher-level conceptual and morpho-syntactic language processing areas and the lower-level speech motor control system which implements a limited set of learned motor programs. The results of the imaging experiment described in Chapter 2 and in Bohland and Guenther (2006) were used to a great extent to guide the development of the model, and in particular, its functional architecture.

Much theoretical research has focused on the processes involved in language pro-

duction. A popular general approach has been to delineate abstract stages through which a communicative concept is subjected to the rules of a language and ultimately transformed into a series of muscle activations used for speech production. This approach was championed by Garrett (1975) in his analysis of *sentence production*, which laid a foundation for much future work. Perhaps the most widely referenced theoretical framework of this type has been developed by Willem Levelt and colleagues (Levelt, 1989; Levelt et al., 1999b). This framework, hereafter referred to as *The Nijmegen Model* (after the city in which it was developed), is schematized in Figure 3-1. The Nijmegen model proposes processing stages (the boxes in Figure 3-1) which receive an input representation of a certain form (or at a certain linguistic *level*) and output a representation of the speech plan in a different form (at a *lower level*). The modeling work presented in this chapter deals with the proposed *phonological encoding* and *phonetic encoding* stages, and interfaces with an existing model that describes the stage of *articulation*. Its primary focus is on the ongoing parallel representation of a speech plan as it cascades through these stages of production. While the model does not explicitly address higher level linguistic processing or representation stages, the proposed architecture appears to be capable of being extended to address these stages quite naturally.

The development of the present model continues a “bottom-up” approach toward the formal description and implementation of biologically-realistic neural systems for planning and controlling the production of speech. The model is an extension to a previously developed model of speech production, the DIVA (Directions into Velocities of Articulators) model (Guenther, 1995; Guenther et al., 1998, 2006), which describes how motor programs for *speech sounds*¹ can be learned and exe-

¹In the DIVA model, a speech sound can be a phoneme, a syllable, an entire word, etc. For the purposes of the current model, it is assumed that speech sounds are syllables and individual phonemes. The notion that syllables can be used as a performance unit in speech is supported by reaction time studies, discussed in Section 3.4.2.

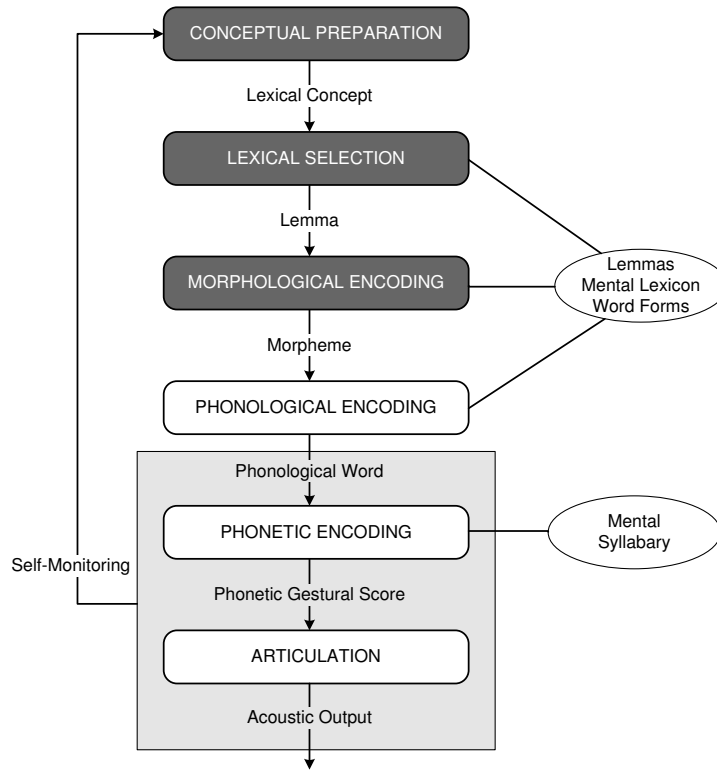


Figure 3.1: Outline of the “Nijmegen Model.” Illustration adapted from Levelt et al. (1999b). Boxed labels indicate processing stages in the model, and unboxed labels indicate proposed input and output representations. Shaded boxes (top three) are not treated in the present model (GODIVA). The large gray box enclosing the final two stages of processing and their inputs and outputs represents stages and representations that Levelt et al. (1999b) suggest are subject to self-monitoring. The phonetic encoding stage is suggested to interface with a “mental syllabary” which contains learned motor programs for frequently used syllables. Earlier stages access the mental lexicon to retrieve lemmas and word forms.

cuted. A limitation of the DIVA model is that it contains no explicit representations for speech *planning* beyond the simple activation of a single speech sound’s stored representation, nor does it specifically address the related issue of appropriately releasing planned speech sounds to the motor apparatus (referred to herein as *initiation*). The extended model (called the GODIVA or **G**radient **O**rders **D**IVA model) adds higher-level sequential representations for planned speech sounds and simulates various aspects of serial speech planning and production. Furthermore, the model continues in the spirit of recent instantiations of the DIVA model (Guenther et al., 2006; Guenther, 2006) by proposing specific neuroanatomical substrates for its components, thereby improving its testability through state-of-the-art neuroimaging methods.

This chapter first includes a historical review of theoretical models of general serial behavior, then describes additional constraints that are placed on models that attempt to explain aspects of speech and language production. This is followed by a description of known neuroanatomical and neurophysiological data that must guide the development of any realistic model of how *the brain* organizes speech. After a brief introduction to the DIVA model, the new GODIVA model is presented, including a formal description of the dynamical equations that control its operation and the hypothesized neural substrates for those components. This specification is followed by example simulations and a discussion of results, limitations, and possible extensions to the model.

3.2 Models of serial behavior

The production of prescribed movement sequences underlies much of human behavior, and has been studied by psychologists, cognitive scientists, and neuroscientists for hundreds of years. Although the problem of how *order* is represented in the

brain is often not addressed in modeling endeavors, a number of proposals have emerged, and are briefly introduced here. Several relevant review articles are also recommended to the interested reader (Houghton and Hartley, 1996; Rhodes et al., 2004; Bullock, 2004).

3.2.1 Associative chaining

Associative chaining theories, which have been considered for many decades (e.g. Ebbinghaus, 1913), postulate that serial order is stored through learned connections between nodes (or neurons) representing successive elements in a sequence. The activation of each node thereby causes activation of the associated subsequent node, facilitating the serial read-out of the sequence (see Figure 3.2). Associative chaining is a derivative of stimulus-response theory, where early proposals suggested that the feedback generated from one response could provide the stimulus required to generate the next. Lashley (1951) recognized a problem for such models when one stimulus (in this context, the activation of a node) could lead to multiple different responses (the activation of different nodes); that is, these simplest models could not learn to unambiguously read-out different sequences defined over the same alphabet of component items. In the (in)famous speech production model of Wickelgren (1969) this problem was overcome by introducing many context-sensitive allophones (e.g. /_kæ_t/ for the phoneme /æ/ when preceded by /k/ and followed by /t/) as items in the set of nodes through which a sequence chain might proceed. This type of model however, encapsulates no relationship between same phonemes in different contexts and suffers from a combinatorial explosion in the number of necessary nodes when allowing for different speech sequences that can overlap over a string of several phonemes.

More recent neural network models (e.g. Jordan, 1986; Lewandowsky and Mur-

dock, 1989; Elman, 1990; Beiser and Houk, 1998) proposed revisions to the associative chaining theory while retaining the principle of node to node or state to state links as the basis for their dynamics. Models of this type rely on a series of sequence-specific internal states that must be learned in order to allow for the recall of any sequence. Although these networks allow more than one sequence to be learned over the same set of elements, there is no basis for performance of novel sequences, learning is often unrealistically slow with poor temporal generalization, and internal recall of a sequence remains an iterative sequential operation (Henson et al., 1996; Page and Norris, 2000; Wang et al., 1996). Finally, for any model based on associative chaining, recreation of cognitive error data is problematic. This is due to the fact that if a “wrong link” is followed in error, the model has no means to recover from the error and, for example, produce the remaining items in the original sequence (e.g. Henson et al., 1996).

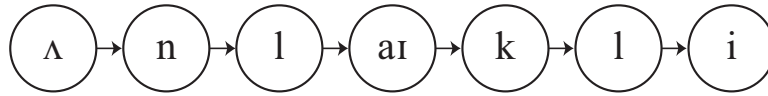


Figure 3.2: Schematic of a simple associative chain model for the production of the word “unlikely”. Read out of the phoneme sequence is initiated by activating the left-most phoneme node. Activation from this node is transferred across the link to the second node, and so forth. This example demonstrates one problem for associative chaining theories in that it requires two distinct nodes for the phoneme /l/. If only one /l/ node existed, it would be unclear which link to follow (to the /aɪ/ or to the /i/ node) without further information.

3.2.2 Positional coding

The development of the serial computer led to the use of many computer metaphors to describe brain function. Computers typically represent order by using successive slots in memory that can contain arbitrary bytes of data. A program then proceeds

in a pre-determined linear succession in order to “perform” the stored program. The influential memory model by Atkinson and Shiffrin (1971), for example, similarly represented items in memory as binary activations in memory “slots.” Conrad (1965) developed a model of human short-term memory in which it was suggested that there exists an ordered set of “boxes” into which individual items in a sequence could be placed. Sequence performance then simply involved stepping through the series of boxes (which are themselves ordered) and performing each associated component item.

One problem with such “slot” models is that there is no obvious *neural* mechanism to allow the insertion of an arbitrary memory (or memory pointer) into a particular “slot.” Such models either require the ability to “label” a positional node with a certain representation or require a set of all possible representations to be available at all serial positions, which is infeasible according to a combinatorial argument in most cases. More recent positional models hypothesize serial order to be associated with some contextual signal such as the state of an oscillatory circuit or some other time-varying function (Henson, 1998; Brown et al., 2000; Burgess and Hitch, 1999). Recall then involves “replaying” this contextual signal which, in turn, preferentially activates the items associated with the current state of the signal. This type of model assumes the ability to form these associations between context signal and component item through “one-shot” learning in order to allow for the performance of novel sequences. A subset of the recent positional models also incorporate aspects of competitive queuing systems (see below) in their architecture.

3.2.3 Parallel models of serial performance: Competitive queuing

Lashley (1951) can be credited with the insight that revealed that associative chaining models could not sufficiently describe sequence performance, and that serial behavior

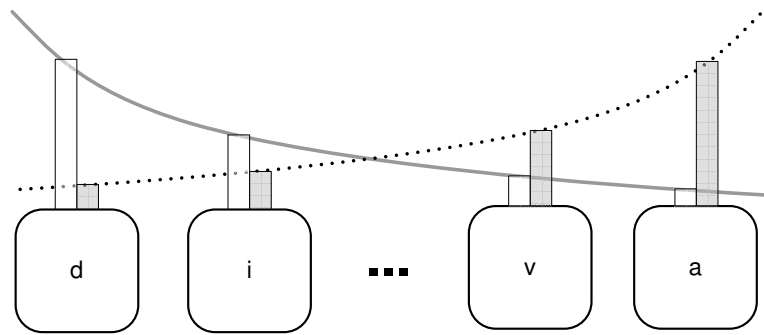


Figure 3-3: Schematic illustration of an example of a positional model representation for the letter sequence “diva.” In positional models, items are associated with explicit positions (e.g. the ordered set of boxes), and/or with context signals that vary with serial position. A set of two such signals are shown here (solid gray curve and dotted curve), with each position being coded by a tuple corresponding to the value of the two curves shown above. Such a positional model would code for position relative to the beginning and end of the entire sequence (cf. Henson, 1998). Frame-based speech production models can also be considered examples of positional coding models.

might instead be performed based on an underlying *parallel* planning representation. Townsend (1974) showed formally how a parallel system that begins processing elements simultaneously could yield equivalent reaction time predictions to a serial system in which elements are processed one after another. Grossberg (1978a,b) was the first to fully develop a computational theory of short-term memory of sequences in which items and their serial order are stored via a *primacy gradient* utilizing the simultaneous parallel activations of a set of nodes. Grossberg’s proposal was motivated by the question of how sequences in short-term memory could be stably coded in long-term memory without destabilizing previously learned codes. In this model the relative activation levels of content-addressable nodes code for their relative order in the sequence. This parallel working memory plan, which is isomorphic to a *spatial pattern* of activation in a neuronal map, can be converted to serial performance

through an iterative competitive choice process in which:

1. The item with the highest activation is chosen for performance.
2. The chosen item's activation is suppressed.
3. The above process is repeated until the sequence reaches completion.

Many similar models that employ a parallel planning layer coupled with an iterative choice process have been developed to account for various aspects of serial behavior including the recall of novel lists (Boardman and Bullock, 1991; Page and Norris, 1998), word recognition and recall (Grossberg, 1986; Hartley and Houghton, 1996; Gupta and MacWhinney, 1997), spelling (Glasspool and Houghton, 2005), cursive handwriting production (Bullock et al., 1993), imitation of unfamiliar movements (Agam et al., 2005), and language production (Dell et al., 1997; Ward, 1994). These types of constructions have collectively come to be labeled *competitive queuing* (CQ) models (Houghton, 1990; Bullock and Rhodes, 2003). Figure 3·4 illustrates the basic CQ model architecture.

Recent evidence from neurophysiology (Averbeck et al., 2002, 2003) as well as from a comparative modeling investigation (Farrell and Lewandowsky, 2004) has lent substantial support to CQ-like models of serial order. A CQ-compatible architecture forms the basis of various representations used in the GODIVA model.

3.3 Linguistic models

While the majority of serial order theories and models have taken aim at data from short-term memory experiments without explicit treatment of linguistic units, several computational models have been introduced to account for the processing of such representations for word production. These models generally follow from the theoretical work of Garrett (1975), Levelt (1989), and others. For the present purposes,

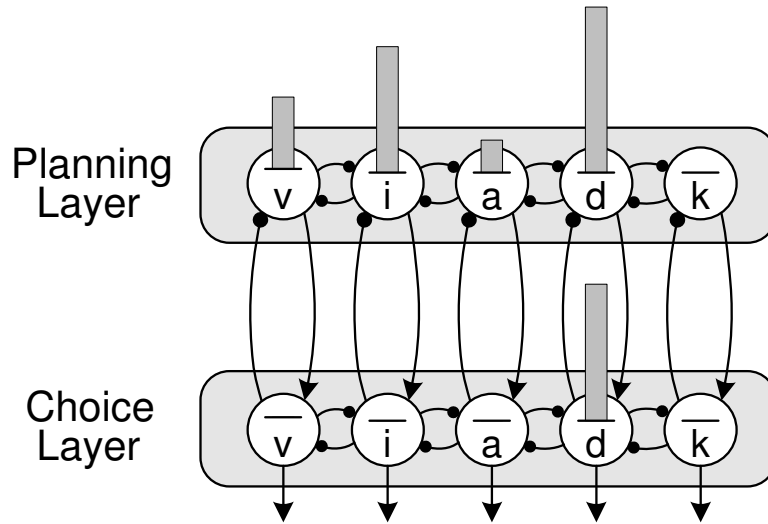


Figure 3.4: Competitive Queuing (CQ) model architecture for the representation and performance of the letter sequence *diva*. The serial position of each letter is encoded by its strength of representation (height of bar) in the planning layer (top). The choice layer (bottom) realizes a competitive (winner take all) process which allows only the strongest input to remain active, in this case “*d*.” Upon selection of “*d*”, its representation in the planning layer would be suppressed, leaving “*i*” as the most active node. This entire process iterates through time, enabling performance of the entire letter sequence “*diva*.”

however, *computational* models are distinguished from these *conceptual* models which typically take the form of “boxes and arrows” but lack a formal level of description or the ability to explicitly simulate the outcome of various task conditions. These linguistic models typically seek to address one or both of two major types of results: i) patterns observed in speech error data or ii) chronometric data concerning reaction times. The major findings from both of these lines of study are outlined in Section 3.4. Before describing these models and data, however, it is necessary to briefly describe a few basic constructs from speech science and phonology.

3.3.1 Syllables, phonemes, and features

Within the domain of phonology, several “units” of production are frequently referred to, and these units are arranged hierarchically, with a *syllable* comprising one or more *phonemes* and a phoneme comprising one or more *features*. The syllable is generally accepted as a phonological unit that structurally binds a set of phonemes. Nevertheless, there is no precise agreed upon definition for the syllable. One definition suggests that syllables are parsed such that there is one *sonority peak* per syllable, where sonority refers to the relative intensity of a sound, or, alternatively, the relative openness of the vocal tract. Some theories (e.g. MacNeilage, 1998) have emphasized the relationship between the motor act of an open-closed jaw alternation and the syllable unit. This is a useful relationship presently in that it implies that there is a behaviorally-relevant, naturally occurring motor *frame* that roughly demarcates syllable boundaries. Such a motor frame might be useful in delineating what the motor system can learn as a single “chunk” and what it needs to organize into sequences of chunks.

Importantly, the phonological syllable does not simply circumscribe a set of phonemes, but also appears useful to describe, for example, the “rules” governing

what phonemes can occur in what serial position within the syllable (e.g. Fudge, 1969). To this end, the syllable can be broken into, at least, an *onset* and a *rime*, the latter of which contains sub-elements *nucleus* and *coda*. The onset and coda can consist of a consonant or cluster of consonants, whereas the nucleus consists of the most sonorant phoneme, typically a vowel. Syllable structure in various languages allow all or some pieces of this syllable structure tree. Interestingly, the CV syllable, consisting of a single consonantal phoneme followed by an open-mouth vowel, exists in nearly all languages, and is the first syllable type acquired by infants (Levelt et al., 1999a).

The phoneme is generally considered to be the most basic contrastive sound element in a language. Phonemes are defined without reference to their syllable or word positions or their phonetic context. Phonemes are categorical, exhibiting a many-to-one relationship between acoustic signals and phoneme labels, and with all realizations of a particular phoneme being cognitively equivalent. Although the mental reality of the phoneme has been a controversial topic for many years (e.g. Sapir, 1949), it is difficult to argue against the notion that language users organize incoming and outgoing speech materials into contrastive categories that have semantic implications. While those categories, as implemented neurally, may not perfectly match a list of phonemes in a phonetic transcription chart, such a relatively small, discrete alphabet of communicatively important categories would be extremely beneficial in terms of efficiency of the production system. The modeling work presented herein begins with an assumption of the reality of phoneme categories; further evidence for the development of such categories in children is discussed in Section 3.3.3.

A *feature* is a distinctive property of a speech sound, either acoustic (Jakobson et al., 1952; Stevens, 1998) or articulatory (Chomsky and Halle, 1968), at the sub-phonemic level. These may describe, for example, whether or not a phoneme is

voiced, or the place of articulation. Segmental theories of speech production postulate that phonological segments (phonemes) have their own abstract representation in planning. According to this view, the representation of a phoneme may *reference* the lower-level features of the segment but can be manipulated at an independent level of processing. Featural theories, on the other hand, suggest that segmental representations are merely the aggregation of appropriate feature-level representations, which are themselves planning units in speech (e.g. Mowrey and MacKay, 1990). Features tend to be particularly useful in describing the similarity between categorical speech sounds; similarity tends to have a large effect, for example, in biasing the types of exchanges that are made in slips of the tongue.

3.3.2 Factorization of structure and content

The majority of theories of phonological encoding and/or serial organization of speech sounds propose some form of factorization of the structure and the phonological content of the utterance. This often takes the form of syllabic frames and phonemic content (although the frame-content division is often extended to higher linguistic levels as well, which are not addressed herein). Such a division is motivated, in large part, by the pattern of errors observed in spontaneously occurring slips of the tongue (see Section 3.4.1). MacNeilage (1998) further suggests that speech *evolved* the capability to program syllabic frame productions with phonological (segmental) content elements, and that every speaker learns to make use of this capacity during his or her own period of speech acquisition.

3.3.3 Speech motor and phonological development

It is well beyond the scope of this dissertation to characterize the development of speech and language capabilities in infants, but it is informative to point out data

that specifically relate to linguistic representations used in the modeling work presented here. Early speech acquisition is characterized by several stages (e.g. Oller, 1980; Stark, 1980). At approximately 2 to 3 months, children exhibit a *cooing* or *gooing* stage, dominated by vowel sounds and some velar constrictions. This is followed by a vocal play stage characterized by yells, whispers, squeals, growls, and occasional rudimentary syllable productions. At approximately 7 months, children enter a *canonical babbling* stage in which they rhythmically alternate an open and closed vocal tract configuration while phonating, resulting in repeated utterances such as “babababa.” MacNeilage and Davis (1990) have suggested that these productions represent “pure frame” productions, and form the basis of a suggestion that frames are acquired *prior* to content. These reduplicated babbles dominate the early canonical babbling stage, but are largely replaced by *variegated babbling* at around 10-13 months of age. This stage involves modifications of the consonant-like and vowel-like sounds in babbles, resulting in syllable strings such as “bagidabu.” MacNeilage and Davis (1990) suggest that this stage may represent the earliest period of “content” development.

Locke (1997) presents a theory of neurolinguistic development involving four stages: i) Vocal Learning, ii) Utterance Acquisition, iii) Structure Analysis and Computation, and iv) Integration and Elaboration. The second and third stages are particularly interesting to the present study. Locke suggests that in Stage 2, “every utterance [children] know is an idiom, an irreducible and unalterable ‘figure of speech.’” This notion of indivisibility is supported by the interesting finding that very young children make far fewer slips of the tongue than adult speakers (Warren, 1986). It is only with the onset of Stage 3, perhaps at 18 to 20 months, that children have the ability to analyze the structure of their utterances, locating, for example, recurring elements. This is suggested to be the stage that provides the child with

the units needed for phonology, and enables generativity and the efficient storage of linguistic material. Importantly, at around 18 months of age, the rate of word acquisition in children may quadruple (Goldfield and Reznick, 1990). The timing of this explosion in a child’s available vocabulary also coincides with a development in the perceptual system at approximately 19 months, at which time children can effectively discriminate the phonetic categories in their language (Werker and Pegg, 1992).

The position taken in the work presented herein is that the stages of speech acquisition up to and including babbling are particularly important for tuning speech-motor mappings such as those used in the DIVA model of speech production (Guenther, 1995; Guenther et al., 1998). These stages also provide a *protosyllabary* of motor programs that are “purely motoric,” having little to no linguistic significance (Levelt et al., 1999b). A later stage, such as Stage 3 described by Locke (1997), leads to the development of *phonological representations* that can become associated with the *phonetic* programs that realize those phonemes. It is suggested that this development also allows the learning speaker to insert content items into common learned syllable frames, thus offering an explanation for the rapid increase in the child’s vocabulary at this time. Furthermore, this representation of the common sound elements in a speaker’s language should remain largely unchanged following learning, and can be used by the adult speaker to interface both words and non-words with a more plastic speech motor system. In a sense, this representation provides a *basis* for representing any utterance in the language. The GODIVA model describes the speech system *after* the development of this stage. It should, therefore, be considered an *adult* rather than child speech neural model.

3.3.4 The WEAVER / WEAVER++ model

The WEAVER / WEAVER++ model (Roelofs, 1997; Levelt et al., 1999b) is, broadly, a computer implementation of the Nijmegen Model (Figure 3.1). WEAVER (Word-form Encoding by Activation and VERification) addresses the stages subsequent to lexical selection (see Figure 3.1), whereas WEAVER++ includes modeling of the lexical selection process as well. Of interest to the present work are the stages following morphological encoding. Specifically, in WEAVER (Roelofs, 1997), a selected morpheme activates nodes representing its constituent phonemes as well as a *metrical structure* which specifies the number of syllables and stress pattern. The serial order of the activated segments is assumed to be indicated by the links between the morpheme node and the phoneme nodes; likewise, links between phoneme nodes and syllable nodes that represent phonetic syllables (e.g. motor programs) are also “labeled” with positional information (in this case indicating onset, nucleus, or coda). The activated set of phoneme nodes constitutes a *phonological word* and is the domain of syllabification. In the WEAVER model, and the Nijmegen model more generally, syllabification is a late-occurring process. Morphemes do not specify syllable boundaries, but only number of syllables. A rule-based system instead computes syllabification in order to account for *resyllabification*, a phenomenon in which syllable boundaries can transcend morpheme or word boundaries.

While the WEAVER / WEAVER++ model is an important formalization of an influential language production model and shares certain similarities with the GODIVA model described below, it has several limitations. The model is designed primarily to account for reaction time data (see Section 3.4.2), and has difficulty explaining, for example, typical speech error patterns. Additionally, its use of rule-based labeling of nodes and links is difficult to conceive of in terms of actual brain mechanisms. The flow of information in the model is, furthermore, not linked to

processing regions and pathways in the cortex, and, therefore, the ability to make inferences about *neural* function on the basis of this model is severely limited. The GODIVA model is intended to bridge this gap between theoretical information processing and the neural substrates that implement such processes.

3.3.5 Other related models

(Dell, 1986) presented one of the first models of language production based on the connectionist principle of *spreading activation*. Dell’s model offers a formal explanation for a variety of speech error data, and represents the archetypal “frame-based” model. The proposal makes use of representations at various hierarchically-organized linguistic levels (cf. Garrett, 1975) such that a node at one level receives top-down input from the nodes one level higher in the model. Representations of the forthcoming utterance are built through a process of *tagging* most active nodes at each level, and this process is done largely in parallel, ultimately forming a *loosely yoked* production system (Reich, 1977). Nodes in the model are *labeled* with linguistic categories. In the process of phonological encoding, for instance, phonemes are labeled as comprising the onset, nucleus, or coda position in a syllable. An abstract syllable frame, or ordered set of categories, is constructed and used to *tag* the most active nodes *within the appropriate categories* at a particular representational level. In this sense, the frame dictates not which exact elements are tagged, but simply which items are eligible to be tagged. Connections in the model are bi-directional, which facilitates the explanation of multiple phenomena observed in naturally occurring speech errors, including various similarity effects.

Dell’s (1986) model was important in that it formalized many theoretical proposals made as possible explanations for speech error data in a computer model. It also further emphasized the frame-content complementarity and used a connection-

ist (cf. neural network) architecture, which brought the theoretical proposal a step closer to biology. Dell et al. (1997) proposed a more general frame-based model that addressed serial order in language production, including the relative proportions of anticipatory and perseveratory errors made by normal and aphasic speakers.

Ward (1994) developed a comprehensive language generator called FIG, which is fully embedded in a structured connectionist network. Concepts, words, and syntactical constructs form nodes that are interconnected (through weighted links) to reflect relational information, and the network operates iteratively and in parallel through simple activation spreading rules. In FIG, there is no central process which plans the serial order of words, but rather order emerges as a gradient of activity across word nodes, much like the parallel representation of order in the CQ architecture. The model follows the simple rule to “select and emit the most highly activated word,” and then that node’s activation is suppressed. FIG, therefore, can be viewed as a demonstration that a large-scale generative language production model operating at multiple levels can be achieved using a connectionist or neural network architecture with a CQ-compatible framework.

Hartley and Houghton (1996) proposed a competitive-queuing based model that also exploits a division of frame and content to explain learning and recall of unfamiliar non-words in verbal short-term memory. In the model, individual syllables are represented in terms of their constituent phonemes and the “slots” that they use in a generic *syllable template* adapted from Fudge (1969). A pair of nodes is allocated for each syllable presented for recall, representing (structurally) the syllable onset and rime. A one-shot learning rule is used to form temporary associative links between these syllable node pairs and both the appropriate syllable template slots and the appropriate phoneme content nodes for each syllable presented for recall. When recalling a sequence of syllables, an endogenous control signal causes a gradient of

activation across the syllable nodes, with the immediately forthcoming syllable receiving highest activation as in the CQ mechanism described by Burgess and Hitch (1992). The most active syllable pair is chosen for output, and gives its learned input to the syllable template and phoneme nodes. As each syllable slot becomes activated (iteratively), phoneme nodes also become activated, with the most active nodes generally corresponding to phonemes from forthcoming syllables that occupy the same slot. The most active phoneme node is then chosen for “output.” After any phoneme or syllable node in the network is chosen for output, its representation is suppressed to allow the system to iterate through the sequence, and to prevent perseveration (see Section 3.2.3). The above mechanisms allow the model to “repeat” non-word sequences and, with the addition of noise, to account for the syllable position constraint in speech error data (see Section 3.4.1). Hartley and Houghton (1996) is an advancement on earlier models such as that of Dell (1986) because of its capacity for single-trial learning of a novel sequence. It additionally explicitly employs the CQ architecture, adding further specificity of mechanistic details as well as biological plausibility.

Vousden et al. (2000) presented a model that is similar in spirit to that of Hartley and Houghton (1996), and that is derivative of a previous model of serial recall (Brown et al., 2000). Vousden et al. (2000) propose the existence of a dynamic, semi-periodic control signal (the *phonological context signal*) that largely drives the model’s operation. A major goal of Vousden et al. (2000) was to eliminate the necessity for syllable position-specific codes for phonemes; in Dell (1986), for instance, phoneme nodes are assigned to a positional category (onset, nucleus, rime), and phonemes which can appear in multiple positions² have nodes for *each* position. This “simplification” occurs, however, at the expense of creating a complex

²Many consonants can occur in either onset or coda position.

9-dimensional time-varying context signal, with each element formed by multiplying the states of a subset of 32 independent oscillators with different frequency and phase characteristics. The successive states of the signal are designed to have similarity peaks (i.e. autocorrelation) at a specific temporal separation, reflected by the period of a low-frequency set of the oscillators. In the simulations performed, this periodicity is always of length 3, which allows each of the three states in a single period to become associated with an onset, a nucleus, or a coda phoneme. The recall of a sequence in the model then depends on *learning* a set of large weight matrices that code associations between the context signal and a matrix³ constituting the *phoneme representation*. At recall, the phonological context vector is *reset* to its original state and “played back,” resulting in a gradient of activations at the phoneme level for each successive state in the context signal. The most active phoneme representation is then chosen for output, and its representation temporarily suppressed. Again, with the addition of noise, the model is then able to recreate various speech error data including positional and similarity effects.

Several concerns arise from the model’s timing, association and recall processes (see also the critique of this class of models in Lewandowsky et al., 2006; Agam et al., 2005). First, in the model, the production of *any* syllable sequence requires building an association between the phonological context signal and the phonemes in the sequence; this seems implausible given the ease with which speech sequences are produced and the relative complexity of the associations that must be made within the model. After such associations are made for a given word, for example, they could perhaps be stored, but this would require many additions to the model. Second, it is unclear how it can be assured that, during the association process, exactly one phoneme is made available for exactly one time step in the dynamic

³In the model, N phonemes are each coded by a 17-dimensional feature vector; placing each of these vectors into columns yields a $17 \times N$ matrix.

control signal. As this signal is built from physiologically motivated oscillators that occur with specific and consistent periodicity, it would appear that some additional mechanism must enforce a strict temporal scheduling to ensure that this encoding process remains precisely synchronized; otherwise, if phonemes were made available at a different frequency than that at which the phonological context signal updates, the entire benefit of such associations could be lost. Third, all simulations performed by Vousden et al. (2000) were of six syllable CVC sequences, which clearly correspond to the number of elements and periodicity of the phonological control signal used. It is unclear that the model is capable of simulating different syllable types without modifications.

The above proposals have provided a baseline upon which the present modeling efforts build. Importantly, each of the aforementioned models makes use of at least some aspects of a competitive queuing compatible architecture, including co-temporal activation of potential production units, winner-take-all choice processes, and post-output response suppression. A major shortcoming of the previous proposals is that such theories have failed to account for how linguistic behavior might emerge from neural structure (Nadeau, 2001). The present efforts make use of many of the same information-processing notions of these and other models, but embed these constructs in a biologically-realistic architecture with specific proposals about cortical and subcortical substrates. In so doing, the model offers the ability to explain additional data sets that are not, at least directly, available to the previous models. In particular, anatomical-region-level effects observed in functional imaging and lesion studies can be related to specific components of the GODIVA model developed herein.

3.4 Constraints on linguistic models

3.4.1 Speech error studies

The examination of naturally occurring speech errors has been a topic of abundant research for over 100 years, beginning with the publication and examination of a German language error corpus (Meringer and Mayer, 1895), and including, for example, the celebrated works of Sigmund Freud (Freud, 1914). Healthy adult speakers make errors in the serial order of speech sounds at a rate of approximately 0.1-0.2% (Garham et al., 1981), errors which include anticipations, perseverations, and exchanges. Early researchers realized that these “slips of the tongue” did not occur randomly, but rather showed regularities that could be useful in understanding the pre-articulatory stages of speech production. MacKay (1970) examined such regularities in speech errors called “spoonerisms,” named after Reverend William Archibald Spooner (1844-1930) who frequently made (often intentional) serial order errors. Spoonerisms are defined as involuntary reversals (or *exchanges*) in the serial order of speech items. For example, a speaker intending to say “left hemisphere” might produce the slip “hemisphere left.” MacKay (1970) noted several regularities in the patterns of these types of errors, including:

1. The within-syllable position of exchanged phonemes was almost always the same. (*Syllable Position Constraint*).
2. Consonants in the onset position of syllables and words were particularly prone to exchanges. (*Syllable / Word Onset Effect*).
3. Features of exchanged phonemes were often similar with the exception of place of articulation. (*Phonemic Similarity Effect*).
4. Consonants were more frequently exchanged than vowels. (*Consonant Effect*).

5. Exchanged phonemes were usually close together within a sentence. (*Transposition Distance Constraint*).
6. The exchanged phonemes often occurred before or (as often) after identical phonemes in the target utterance. (*Repeated Phoneme Effect*).

These basic patterns have been repeated across many studies. The syllable position constraint appears to be perhaps the strongest of the observed patterns. Shattuck-Hufnagel (1979), for example, found that 207 of 211 exchange errors involved transpositions to and from similar syllabic positions; MacKay (1970) similarly found 98 of 100 consonantal movement errors moved to the same syllable position. More recently, Vousden et al. (2000) provided an excellent and detailed analysis of a large speech error corpus collected by Trevor Harley over several years. This analysis found that approximately 90% of consonant movement errors followed this constraint. Treiman and Danis (1988) found that, also during non-word repetition, most errors are phonemic substitutions that preserve syllable structure.

Misorderings in speech errors can be classified as word-level, morpheme-level, or sound-level errors (Dell, 1986). Garrett (1975) hypothesized that the occurrence of slips involving linguistic items at different levels could be used to demarcate processing stages in his model of sentence production. The work presented here takes into consideration only *sound-level errors*, which usually take the form of misorderings of phonemes or sets of phonemes. Nooteboom (1969) found that $\sim 89\%$ of sound-level errors were phoneme errors, with an additional 7% involving entire consonant clusters, and only $\sim 4\%$ of another form. In order to determine the most likely *unit* involved in sound-level errors, Shattuck-Hufnagel and Klatt (1979) examined exchange errors occurring between phonemes that differed by more than one feature. They reasoned that, if the feature constituted a true unit of planning, then there should exist many errors in which only a *single* feature was involved in the exchange (e.g.

a *partial* phoneme substitution). They found that such single feature exchanges, however, occurred in only 3 of the 72 such examples in their database, suggesting a limited role for articulatory features in speech planning. To the contrary, some recent articulometric data demonstrate that in some errors, two phonemes may be produced simultaneously and/or intrusion errors may occur due to the activation of additional inappropriate speech gestures (Poupier and Hardcastle, 2005; Goldstein et al., in press). In the report by Goldstein et al. (in press) such errors were elicited in a syllable repetition task in which repeating syllables differed by only the initial phoneme (e.g. *cop top*). Error rates (and particularly for co-production or “intrusion” type errors) were much larger when speech rate was increased. A possible explanation for such non-phoneme intrusions, in the light of previous evidence in favor of mostly phoneme-sized errors, is that it is task-specific; repeating such similarly formed words, particularly at a high rate, could lead, for example, to a failure of convergence to a single active item in the choice layer of a CQ model prior to initiation. Even if the *units* of the CQ representation were phonemes, coactivation of two phonemes in the choice layer (e.g. /k/ and /t/) could reasonably lead to co-production of these consonants. Such an explanation has been advanced to explain co-production of two otherwise competing actions in the phenomenon of “saccadic averaging” (Brown et al., 2004). Until such non-phonemic slips are demonstrated to occur in a broader context, it would appear that the most parsimonious explanation remains that phoneme-like units are the important unit of content during phonological encoding.

Although several potential problems have been identified with the collection of and utilization of speech errors as evidence for language production processes (e.g. Cutler, 1982), such data can be taken as evidence for the types of representations used in planning, and for how the normal production system breaks down. Such evidence

is particularly important in attempts to model a neural system. By analogy, the study of optical illusions, where the visual system can be considered to break down in that the viewer perceives what is not actually true in the physical world, has been of great importance in understanding the mechanisms of the visual system (e.g. Cornsweet, 1970). It should also be noted that normal slips of the tongue and other serial order errors of linguistic output (e.g. errors in writing or typing) share at least some similarities with paraphasic errors made by aphasic patients (e.g. Berg, 2006). To the extent that such errors are similar, this suggests that the pathological case may involve a severe disruption to the same circuitry that occasionally misfires in normal speakers.

3.4.2 Reaction time studies

Two major paradigms have been used to gather data concerning the time required to initiate a behavior, or *reaction time* (RT). In both paradigms, RT is measured by beginning a timer at the delivery of an imperative stimulus that signals the subject to “go” and ends with the onset of the behavior. For the present purposes, the behavior in question is speech production and its onset is measured either acoustically or with some measure of the start of articulatory movements. The two RT paradigms differ in the point in time during a trial at which the subject is informed of the specific response to be made. In the *choice reaction time* paradigm the imperative signal itself informs the subject of the response. In the *simple reaction time* paradigm the subject is informed of the response by an earlier precue and given time to prepare or “load” the response prior to delivery of the imperative (GO signal). This preparatory period in the simple RT protocol causes a reduction in simple RT relative to choice RT (Donders, 1969) and is believed to indicate the utilization of an output “buffer” which allows the prepared response to be active for a short time period.

These RT paradigms have been used to measure reaction times for utterances of varying complexity (e.g. Eriksen et al., 1970; Sternberg et al., 1978; Klapp et al., 1973; Klapp, 1974; Klapp et al., 1981; Klapp, 2003). When subjects are given time to prepare the utterance prior to the GO signal, as in simple reaction time, an approximate linear relationship has been observed between the number of words planned and reaction time (Sternberg et al., 1978). This has been called the *sequence length effect on latency*. When a response is required immediately upon the presentation of the stimulus as in choice RT, the results are less clear. Some studies have found that choice reaction time varies with the number of syllables in a planned utterance (e.g. Eriksen et al., 1970; Klapp et al., 1973; Santiago et al., 2000), while others have failed to find such an effect (e.g. Bachoud-Lévi et al., 1998). In Klapp (2003) various manipulations of syllable sequences were performed which resulted in the establishment of two patterns of reaction time. For responses in which subjects treated an entire syllable sequence as a pseudoword (a single “chunk”), choice RT increased with increasing number of syllables (N) whereas simple RT was independent of N . This replicates Klapp et al. (1973), and is intuitively explained as follows. During simple RT, the utterance can be prepared in advance of the imperative stimulus, thus removing the need for one component process that must still occur *after* the imperative in the choice RT paradigm. In choice RT, the time required to encode the utterance is revealed, and that time depends on the complexity of the pseudoword. When subjects treated utterances as a sequence of individual chunks, however, the RT pattern was very different. In this case reaction times in the simple RT paradigm increased with the number of chunks whereas choice RT was independent of this number.

Schönle et al. (1986) performed a reaction time study in which subjects repeated *simple* and *complex* syllable sequences of similar composition to those used in the fMRI experiment described herein (Chapter 2). They found that, after controlling for

sensory processing of the syllable sequences, simple sequences (of the type *ba-ba-ba*) were produced with a significantly shorter latency than complex sequences (of the type *ba-da-ga*). A mean difference of 102 ms was suggested to reflect the additional time required to program the complex sequence compared to the simple sequence.

A *syllable frequency effect* has been the topic of much discussion in recent speech production literature. This effect, in which syllables that are frequently encountered in a speaker's language are produced with a shorter latency than syllables that are uncommon (but legal), has been reported by several researchers (Levelt and Wheeldon, 1994; Carreiras and Perea, 2004; Alario et al., 2004; Cholin et al., 2006; Laganaro and Alario, 2006). While Levelt and Wheeldon (1994) first argued that the syllable frequency effect implied the use of stored syllable motor programs, it has been difficult to rule out the possibility that the effect was due to higher-level phonological processing. Laganaro and Alario (2006) used a delayed naming task with and without an interfering task in an attempt to determine the stage at which the syllable frequency effect arises. In a delayed naming task (using both words and non-words), it was found that, with a sufficient delay period, the syllable frequency effect was not observed. However, when the delay period was filled by an articulatory suppression task thought to interfere with *phonetic encoding*, the effect reappeared. This study thus provides the strongest evidence that the syllable frequency effect is localized to phonetic encoding, and that individual syllables might be encoded as units at this late processing stage.

3.4.3 Clinical studies

A number of communication disorders, including aphasia, apraxia of speech (AOS), and stuttering include deficits in the proper sequencing of speech sounds. Phonemic paraphasias are observed in most aphasic patients, and most commonly in conduction

aphasics. Conduction aphasia is caused by damage to the inferior parietal cortex, the underlying white matter tracts including the arcuate fasciculus, or the insula (Palumbo et al., 1992; Damasio and Damasio, 1980). It is classically considered a disconnection syndrome in that language-receptive regions in the inferior parietal lobe are thought to be disconnected from frontal motor speech regions. Conduction aphasia is most frequently characterized as a repetition disorder, although literal paraphasias occurring in all types of speech, not only repetitions, are viewed by many as the defining symptom (Kohn, 1992). Literal paraphasias seen in conduction aphasics and also sometimes in Broca’s aphasics often present themselves as errors of speech output much like those associated with the speech-motor condition apraxia of speech. Several authors (McNeil et al., 2004; Van der Merwe, 1997; Ziegler, 2002) have noted the importance of establishing well-specified models of normal and disordered speech to help provide differential diagnoses and treatment options for these conditions.

Outside of the classical anterior and posterior language zones, lesions to specific brain sites can give rise to speech sequencing and initiation difficulties. A number of cases studies of speech output in patients with lesions of the supplementary motor area (SMA) have been reported (e.g. Jonas, 1981, 1987; Ziegler et al., 1997; Pai, 1999). These cases usually result in reduced propositional (self-initiated) speech with non-propositional speech (e.g. counting; repeating words) largely intact. Other problems include involuntary vocalizations, echolalia, lack of prosody, stuttering, and variable speech rate. Both Jonas (1987) and Ziegler et al. (1997) suggest that the SMA plays a role in sequencing and self-initiation of speech sounds. The basal ganglia are also involved in sequencing motor acts (e.g. Harrington and Haaland, 1998). Basal ganglia speech pathologies generally take the form of hypokinetic or hyperkinetic dysarthrias, often symptomatic of either Parkinson’s Disease or Huntington’s Disease respectively

(Kent, 2000; Murdoch, 2001). A study of speech in patients suffering with Parkinson’s Disease also revealed sequencing deficits, particularly when subjects were asked to read multi-syllabic sequences involving movements that were heterogeneous in place of articulation (Ho et al., 1998). Pickett et al. (1998) report the case study of a woman with bilateral striatal damage in the putamen and head of the caudate nucleus. They noted a general articulatory sequencing deficit, with a particular inability to rapidly switch from one target to the next. These and other data from case studies that specify the locations of focal lesions that affect speech sequencing abilities greatly help to inform models of speech production in normal and patient populations.

While pathological speech data are abundant, parsimonious explanations for differential syndromes remain elusive. Many authors have described the need for mechanistic models to generate testable hypotheses about normal and disordered speech. In just this context, while describing the DIVA model of speech production (Guenther, 1995; Perkell et al., 1997, 2000; Guenther et al., 1998, 2006), McNeil et al. (2004) writes “While this model addresses phenomena that may be relevant in the differential diagnosis of motor speech disorders, in its current stage of development it has not been extended to make claims about the relationship between disrupted processing and speech errors in motor speech disorders.” The GODIVA model specified here begins to extend the DIVA model in precisely this way.

3.5 Neuroanatomical and neurophysiological modeling constraints

Chapter 2 included a review of data concerning the potential roles of several cortical and subcortical regions in speech planning and production. This section is intended to further elaborate on the neuroanatomical and neurophysiological considerations

involved in the development of the GODIVA model. Such data are presented with an emphasis on modeling.

3.5.1 Neurophysiology of prefrontal cortical cells

The left prefrontal cortex, specifically in and surrounding the ventral inferior frontal sulcus, was shown in Chapter 2 to increase its activity in a memory-guided speech production task when the underlying complexity of the to-be-spoken utterance was increased. This increase in activity is consistent with the hypothesis that this region contains a parallel representation of the content (phonemes) in the forthcoming speech plan. That is, when additional items were required to be held in phonological working memory prior to a GO signal, activation of this area showed a corresponding increase.

Averbeck et al. (2002, 2003) recorded single cell activity from the right hemisphere prefrontal cortex near the posterior extension of the principal sulcus in macaque monkeys. The recording sites were within approximately 5 mm of the ventral portion of the arcuate sulcus (see Averbeck et al., 2003 for precise electrode placements), which has been proposed to be the monkey homologue to the inferior frontal sulcus in humans (Rizzolatti and Arbib, 1998). The monkeys were trained to copy geometrical shapes using a joystick held in the left hand while utilizing a specific ordered set of strokes to copy each shape. Cell ensembles that were found to code for specific segments in the shape (Averbeck et al., 2003) were recorded from in a delay period prior to the first stroke as well as throughout the performance of the stroke sequence. In the delay period, a parallel cotemporal representation of *all* of the forthcoming segments was found, and the relative strength of activity in each neuron ensemble was found to predict the order in which the segments were performed. During the movement period, after a segment was performed, the activation of its corresponding

ensemble representation was strongly reduced, and the other ensembles' activations increased. Figure 3.5 shows the activity patterns observed for four shape drawing sequences. Additionally, studies have shown a partial normalization of total activation distributed among the representation for planned items (Averbeck et al., 2002; Cisek and Kalaska, 2002). Total activity grows slower than the number of planned items in a sequence, eventually saturating. This property, which is replicated in the CQ planning layer dynamics (Grossberg, 1978a,b), explains why there is a limit to the number of items in a sequence that can be planned and remembered in the correct order prior to performance (see Cowan, 2000).

Taken together, these electrophysiological findings provide compelling evidence for CQ-like dynamics in the prefrontal cortex, in a location near the possible homologue for human inferior frontal sulcus. The GODIVA model posits that such parallel processing takes place for planned phonemes in the inferior frontal sulcus region in the left hemisphere.

3.5.2 SMA and pre-SMA

The medial premotor cortices have been implicated in sequencing as well as the production of speech for many years. Numerous studies have provided evidence for a separation of the medial wall premotor cortical areas previously collectively described as the "supplementary motor area" into a posterior area termed the SMA proper (referred to here as SMA) and an anterior area termed the pre-SMA as suggested by Matsuzaka et al. (1992). This parcellation is suggested on the basis of neuroanatomical and neurophysiological differences observed between the regions (see Picard and Strick, 1996; Tanji, 1996, for reviews).

It had been suggested many years ago (Vogt and Vogt, 1919) that the traditionally defined medial area 6 was composed of two cytoarchitectonically distinct zones (*6aa*

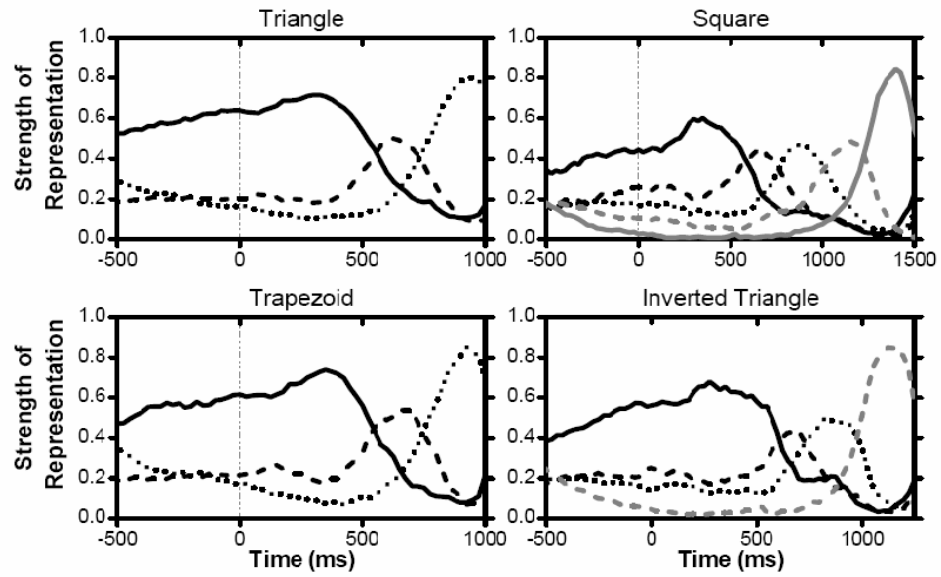


Figure 3.5: Plots showing the strength of representation for cells encoding one of four shapes performed in sequence. Each black or gray data trace (solid, dashed, dotted) represents the relative activation level in monkey area 46 of a small neural ensemble that represents one element of a 3-, 4-, or 5-element sequence used to draw a geometric form. [Adapted from Averbeck et al. (2002)].

and $6aa\beta$). This was confirmed by Matelli et al. (1991) in monkeys who divided the regions into fields F3 (corresponding to pre-SMA) and F6 (corresponding to SMA) and by Zilles et al. (1996) in humans. Vorobiev et al. (1998) reported the existence of *three* cytoarchitectonically distinct regions in humans. Many connectivity studies in primates have also shown strong differences in afferent and efferent projections of the pre-SMA and SMA (Jürgens, 1984; Wiesendanger and Wiesendanger, 1985; Luppino et al., 1993; Matelli and Luppino, 1996; Inase et al., 1999). Such differences have also been found in humans using diffusion tensor imaging (DTI; Johansen-Berg et al., 2004; Lehericy et al., 2004).

Matsuzaka et al. (1992) delineated the two regions on the basis of evoked potentials due to electrical stimulation of motor cortex and due to different observed cell responses during a trained movement task. M1 stimulation led to responses in the SMA-proper only. Furthermore, pre-SMA cells were more likely to i) respond to sensory cues, 2) show preparatory buildup, and 3) show time-locked activity to a GO signal than were SMA cells.

Shima and Tanji (2000) identified *sequence selective* cells in both the SMA and pre-SMA that fired selectively for a particular *entire* sequence of three movements being planned; this activity occurred during a delay period prior to the onset of the first movement in the sequence when the movements were arranged in a particular serial order but not when the same movements were to be performed in a different order. This activity ceased when the first movement was made. This study also identified *interval selective* cells, mostly in the SMA, that fired in the time between two particular component movements. Finally, *rank order selective* cells were found, primarily in the pre-SMA whose activity increased prior to the n^{th} movement in the sequence, regardless of what that particular movement was.

The finding of rank-order selective cells in pre-SMA was a replication of a result

found in a study by Clower and Alexander (1998), which found cells in both the pre-SMA and SMA that reflected the numerical order of a specific component in a well-learned movement sequence independent of the prior or subsequent movement (i.e. relational order). Cells with these properties were nearly twice as prevalent in the pre-SMA. Importantly, in all but one case, the cells coded for spatial variables rather than for the movement itself, indicating the operation of these cells at a higher-level of abstraction than, for example, motor cortex. This notion was, again, supported by the finding of Shima and Tanji (2000) that only 6% of pre-SMA cells recorded from were selective to particular movements; this was in contrast to 61% of SMA cells that were specific to a particular movement.

The overall findings suggest that i) the pre-SMA operates at a higher level in the motor hierarchy than does the SMA, ii) that both the pre-SMA and SMA contain cells that code for more abstract dimensions of the motor plan than the movements themselves, and iii) that the pre-SMA contains cells that code for serial positions or *slots* in learned sequences. These data motivate the cells proposed to exist in the medial premotor regions in the GODIVA model.

3.5.3 Cortico-striato-pallido-thalamo-cortical loops

Interactions between the cerebral cortex and subcortical regions are organized into multiple loop circuits (Alexander et al., 1986; Alexander and Crutcher, 1990; Middleton and Strick, 2000). The complete circuitry within these basal ganglia loops is quite complex (see e.g. Parent and Hazrati, 1995; Bolam et al., 2000); here a simplified view in line with the present modeling efforts is presented.

The striatum, classically considered the input region of the basal ganglia, consists of the caudate nucleus and the putamen. Both convergence and divergence have been observed in cortico-striatal projections, with one cortical area projecting to multiple

striatal patches, and multiple cortical areas projecting to the same striatal patch (Flaherty and Graybiel, 1994). The majority of neurons in the striatum are GABA-ergic medium spiny neurons (MSNs), also the principle projection neurons that send axons outside the striatum (Kemp and Powell, 1971). These cells are hyperpolarized and normally silent at rest, requiring coordinated convergent input from a number of cortical cells to become active⁴ (Wilson, 1993, 1995). The striatum also has a less prevalent set (only 2-3% of striatal cells in rats, but perhaps as high as 23% in humans; Graveland, 1985) of various interneurons, many of which exhibit high firing rates and receive predominantly cortical input (Kawaguchi, 1993; Tepper et al., 2004). These cells, rather than recurrent connections between the MSNs themselves, have been suggested to provide feed-forward surround inhibition in the striatum (Jaeger et al., 1994; Plenz and Kitai, 1998).

MSNs in the striatum give inhibitory projections to two segments of the pallidum, the GPi (Globus Pallidus Internal Segment) and GPe (Globus Pallidus External Segment). These projections form the basis of the classically defined *direct pathway* and *indirect pathway*, respectively (Albin et al., 1989). The GPe then gives inhibitory projections to the GPi, thus opposing the direct pathway⁵. Finally, cells in the GPi (and another output nucleus, the substantia nigra pars reticulata) are tonically active, and give inhibitory input to cells in the thalamus that project back to cortex (e.g. Deniau and Chevalier, 1985). The net effect of the dual pathway view of the basal ganglia is that the *direct pathway is excitatory* and the *indirect pathway is inhibitory*. Hikosaka and Wurtz (1989) found that basal ganglia output neurons

⁴MSNs have been characterized as having a “down-state” corresponding to the hyperpolarized resting state and an “up-state” corresponding to a more depolarized membrane potential that entails more sensitivity to cortical inputs (Wilson, 1993, 1995). This level of detail is beyond the scope of the present modeling investigation.

⁵The original conceptualization of the indirect pathway also included the subthalamic nucleus, which receives inhibitory projections from GPe, and sends excitatory projections to GPi. More recently researchers have acknowledged the existence of, and possibly greater role for, a shorter “indirect pathway” consisting of striatum to GPe to GPi projections (see e.g. Levy et al., 1997).

enable voluntary saccades by means of a pause in the normally tonic inhibition delivered to targets in the superior colliculus and motor thalamus. Such a (spatially specific) pause response can be generated by focused excitation of the direct pathway.

Mink and Thach (1993) and Mink (1996) outlined a conceptual model of basal ganglia function, with the basic principle suggesting that BG loops are used to selectively enable a motor program for output among competing alternatives. This could be achieved via two pathways through the basal ganglia - one a *focused, convergent pathway* and the other a *divergent pathway*. They proposed that i) the convergence of cortical inputs onto striatal cells and of striatal cells onto GPi cells, ii) local inhibitory interneuron circuits in the striatum, and iii) learned dopaminergic modulation of cortico-striatal synapses provide the basis for the convergent pathway. The divergent pathway was posited to comprise fast excitatory projections from cortex to STN followed by a highly divergent (Parent and Hazrati, 1993) connection from STN to GPi.

The net effect in the Mink et al. basal ganglia is an off-center, on-surround network with output at the GPi/SNr which can be used to selectively enable a movement while inhibiting competing movement plans. Mink (1996) suggested that the basal ganglia do not generate movements, but rather select and enable them. Such models of the basal ganglia can be classified as *Action Selection* models. Mink (1996) also alluded to a role for such a circuit in movement sequencing, suggesting that for a sequence of movements, each component movement must be selected while other potential movements are inhibited. Many other researchers have proposed basal ganglia-based models of action selection in this same spirit (e.g. Kropotov and Etlinger, 1999; Redgrave et al., 1999; Gurney et al., 2001a,b).

Brown et al. (2004) described a detailed computational neural model (TELOS; TElencephalic Laminar Objective Selector) for the control of saccades. The model

includes a number of components, but cortico-BG loops subserve a major coordinative computational role, acting as a “large set of programmable gates.” These gates help to choose among competing cortical plans, enable output of a selected plan, or defer the execution of a selected plan. In the model the striatum receives many cortical inputs from plan cells in the superficial layers of various gateable cortical zones (GCZs). These zones also receive the action of output projections from thalamus. Brown et al. (2004) stress that it is implausible for action selection models to suppose that loops through the BG have sufficient selectivity to choose specific actions represented in cortex; instead, the notion of GCZs is implemented such that an entire topographic region of cortex can be selected via the BG, but competition among specific plans should be implemented within cortical circuits. In TELOS, the direct pathway through the BG provides a means for a cortical plan to bid for release by the phasic inhibition (pause) of the tonically active GPi/SNr, thus “opening a gate.” The indirect pathway (striatum→GPe→GPi) is hypothesized to serve as a STOP signal that can be trained based on thalamo-striatal feedback projections. This STOP signal can defer the release of a cortical plan even if that plan’s activity would ordinarily excite the direct pathway sufficiently to open the thalamic gate. A final pathway (*the hyperdirect pathway*; cortex→STN→GPi) is used during movement to provide an excitatory *lockout* for GPi/SNr resources, prohibiting other movements from interfering with ongoing performance.

The BG loop model proposed within GODIVA is vastly simplified, but makes use of several of these previous theoretical proposals. For example, GODIVA uses feed-forward striatal inhibition in the striatum, and requires a phasic dip in GPi activity in the direct pathway to enable a cortical zone. As in TELOS, these zones correspond to patches of cortex rather than to specific cortical representations.

3.6 The DIVA model of speech production

The DIVA (Directions Into Velocities of Articulators) model is a neural network model of speech motor control and acquisition first described by Guenther (1994) and advanced by Guenther and colleagues over the past 12+ years. Computer simulations using a simulated vocal tract (based on Maeda, 1990) have proven capable of offering unified explanations for a large number of speech phenomena including motor equivalence, contextual variability, speaking rate effects, and coarticulation (Guenther, 1995; Guenther et al., 1998). The model's simulations have also been compared directly to speakers' articulator movements (Guenther et al., 1999; Perkell et al., 2004) measured using electromagnetic midsagittal articulometry (EMMA; Perkell et al., 1992). Additionally, the DIVA control scheme has been shown to provide stable control in spite of dramatic developmental changes in the vocal tracts of young children (Callan et al., 2000) and has been used to describe possible abnormalities that lead to dysfluency in stuttering (Max et al., 2004). Recent versions of the DIVA model have additionally hypothesized neural correlates for the representations and mappings that form the model (Guenther, 2001; Guenther and Ghosh, 2003; Guenther et al., 2006; Guenther, 2006).

Figure 3-6 illustrates the components of the most recent instantiation of the DIVA model (Guenther et al., 2006). Each block in Figure 3-6 represents a module hypothesized to correspond to a set of neurons in a particular anatomical region of the brain. Such modules are hypothesized to form a representation in a particular coordinate system. Directed connections between blocks indicate synapses or neural pathways, through which one representation is transformed into the next. Below the major components of the model are briefly reviewed.

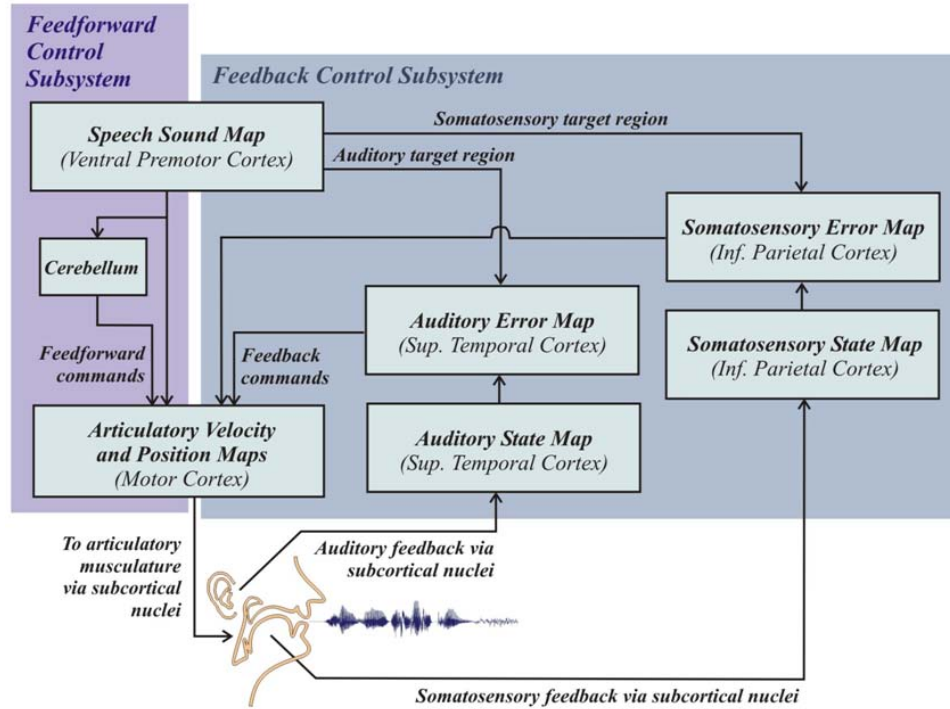


Figure 3.6: Schematic overview of the DIVA model of speech production. Boxes correspond to representations hypothesized to exist in specific cortical and subcortical regions. Arrows correspond to neural pathways that map one representation into another. This version of the model (Guenther et al., 2006) introduces a factorization of the circuit into feedforward and feedback control subsystems (purple and gray boxes in the left and right of the figure, respectively).

3.6.1 Speech sound map

The cells in the DIVA model’s *Speech Sound Map* (SSM) module correspond to specific well-learned speech sounds. This region is hypothesized to lie in the left ventral premotor cortex and/or posterior inferior frontal gyrus *pars opercularis*. SSM representations are functionally similar to a *mental syllabary* (Crompton, 1982; Levelt and Wheeldon, 1994; Levelt et al., 1999b), suggested by Levelt et al. (1999b) to consist of a “repository of gestural scores for the frequently used syllables of the language (p. 5).” In different terminology, SSM representations can be thought of as *motor chunks* or *motor programs*, learned higher-order representations of frequently specified spatiotemporal motor patterns. Section 3.4.2 described the *syllable frequency effect*, which suggested that such motor chunks could correspond to syllable-sized units. Laganaro and Alario (2006) provided additional evidence that this effect has a phonetic (rather than phonological) locus, which is consistent with the placement of the Speech Sound Map within the DIVA (and GODIVA) model.

3.6.2 Feedforward control system

Activation of a Speech Sound Map cell corresponding to a single specific speech sound “reads out”, through projections to articulatory velocity and position cells in motor cortex (see Figure 3-6), a time sequence of articulatory gestures. This series of motor commands results, under normal speaking conditions and for a well-learned sound, in the desired acoustic output for that speech sound. This *feedforward control system* is hypothesized to contain an additional trans-cerebellar pathway to motor cortex that contributes primarily to the temporal details of the motor program, and possibly to account for temporal processing delays inherent in the speech system (Ghosh, 2005). The feedforward system is responsible for learning the motor chunks corresponding to specific learned sounds, where learning is performed on the basis of error signals

generated in a complementary *feedback control system*.

3.6.3 Feedback control system

Model Speech Sound Map cells additionally project through modifiable synapses to secondary auditory and somatosensory cortical regions. These projections encode auditory and somatosensory expectations for the activated sound. Rather than specifying a precise point in auditory or somatosensory space (at each point in time), the model postulates that convex *target regions* are encoded; these correspond to acceptable ranges (in acoustic, motor, or somatotopic space) for the target sound, allowing the model to exhibit flexibility in its productions, including, for example, contextual variability effects.

The purpose of the *feedback control system* that receives these projections is to compare the ongoing sensory state with the expected sensory state, and to issue compensatory motor commands whenever an error is detected in the production of the selected sound. This is accomplished in the auditory and somatosensory *error maps* in the model (see Figure 3-6). In these maps the current sensory feedback is compared to the expectation; under normal conditions, no error is detected, and the two incoming signals (expectation and sensory state) “cancel each other out,” leaving no residual activity in the error map. When an error is detected, however, the residual activity results in a projection through tuned synapses to motor cortex that sums with the ongoing motor command from the feedforward control system. These projections from auditory and somatosensory error maps constitute *inverse models* in that they must encode an inverse kinematic transformation from sensory to motor coordinate frames. Ultimately, these projections lead to a change in motor velocities in order to impart a change in sensory directions. This is, in fact, the basis of the model’s name. If compensatory commands are required due to errors

generated by an inaccurate feedforward system, then such corrective commands are also incorporated into a modified feedforward command; thus the model continues to improve its ability to function properly without relying on sensory feedback.

The synapses that encode auditory to motor transformations can be learned through a simulated babbling phase that precedes the development of individual Speech Sound Map motor programs. This allows the model to learn the mappings between the Error Maps (in sensory space) and the Motor Cortex (in articulator space). Mathematically, these mappings are an approximation to the Moore-Penrose pseudoinverse of the Jacobian of the non-linear function that relates articulator positions to the corresponding sensory state. Ultimately such mappings are required to allow the model (and the developing speaker) to navigate acoustic space by intelligently manipulating the articulators.

3.6.4 Limitations of the DIVA model

In previously published versions of DIVA, the activation of Speech Sound Map cells is algorithmic and specified by the researcher running a simulation. Performing sequences of well-learned sounds thus requires the ad hoc specification of SSM cell activations. Furthermore, Guenther et al. (2006) acknowledge that the model's components only correspond to those used for the production of a simple speech sound, e.g. a single syllable. The model does not address planning for sequences of connected speech or the regions of the brain that are likely to be involved in those processes. Chapter 2 described an fMRI experiment, which was motivated by these limitations, that provided insight into these additional brain regions and their responses to the preparation and overt production of syllable sequences of varying complexity. Below, a computational neural model is specified that begins to extend DIVA to address these brain regions and associated planning and initiation processes.

3.7 The Gradient Order DIVA (GODIVA) model

This section describes the methods used to specify and implement the GODIVA model as well as a high-level functional overview of the model’s operation. Section 3.8 gives the more detailed formal specification of the new model, including the equations that govern its operation. Table 3.2 then provides an algorithmic summary of the steps the model takes to produce a syllable sequence. Example simulation results are presented in Section 3.9.

3.7.1 Computational methods

The GODIVA model is formally described as a system of differential equations that characterize the activity through time of simulated neurons or assemblies of neurons. The model was implemented using MATLAB® (The MathWorks, Inc., Natick, MA), and the differential equations were numerically integrated using 4th and 5th order Runge-Kutta methods with an adaptive time step.

Notation

The formal description of the model (Section 3.8) makes use of certain typical conventions for mathematical description. Neurons in a particular representation are specified by a lower-case letter to indicate the representation (layer), and subscripted to indicate the particular neuron in that representation. For example, x_i indicates the i^{th} neuron in layer x . Representations that have two descriptive dimensions have two subscripts (e.g. x_{ij}). An entire layer is, at times, referred to as a vector, which appears in lower-case, with bold-face font (e.g. \mathbf{x}). Derivatives with respect to time appear in “dot” notation; that is, $\dot{x} = \frac{dx}{dt}$. Connectivity between cells is represented with multiple weighted adjacency matrices; matrices appear as bold-faced upper-case letters (e.g. \mathbf{W}), whereas an individual synaptic weight from cell i to cell j appears

as W_{ij} . Upper-case letters which appear in the equations in normal (non-bold) face as well as Greek symbols are scalar parameters of the model. In figures that schematize the neural network for a particular modeled region, lines with arrows represent excitatory projections, lines with filled circles indicate inhibitory projections, and lines with filled semi-circles indicate multiplicative or “gating” projections.

3.7.2 Functional overview of the model

This section provides a high-level overview of how the GODIVA model functions. It is followed by a much more detailed description and precise specification of its implementation. An overall schematic view of the model is shown in Figure 3-7.

The “input” to the GODIVA model during ordinary speech production arrives from high-level lexical/semantic and or syntactic processing areas, possibly including the inferior or ventrolateral prefrontal regions of the cortex. In most cases, these inputs are thought to code lexical items (words) or short phrases, and likely arrive at the present model’s inputs sequentially as incremental processing is completed by higher-level linguistic modules. These inputs serve to initiate the activation of two parallel and complementary representations of a forthcoming utterance, a *phonological content representation* hypothesized to exist in the left hemisphere inferior frontal sulcus (IFS), and a *structural frame representation* hypothesized to exist in the pre-supplementary motor area (pre-SMA). Both representations constitute planning spaces or forms of working memory, where representative neurons or populations of neurons maintain a cortical code for the potential phonemes (in the IFS) or abstract syllable frames (in the pre-SMA) that define the utterance. Furthermore, both representations in the model can simultaneously, co-temporally code for multiple forthcoming phonemes and syllables by use of a *primacy gradient*, in which relative activation level codes for the serial order in which the items are to

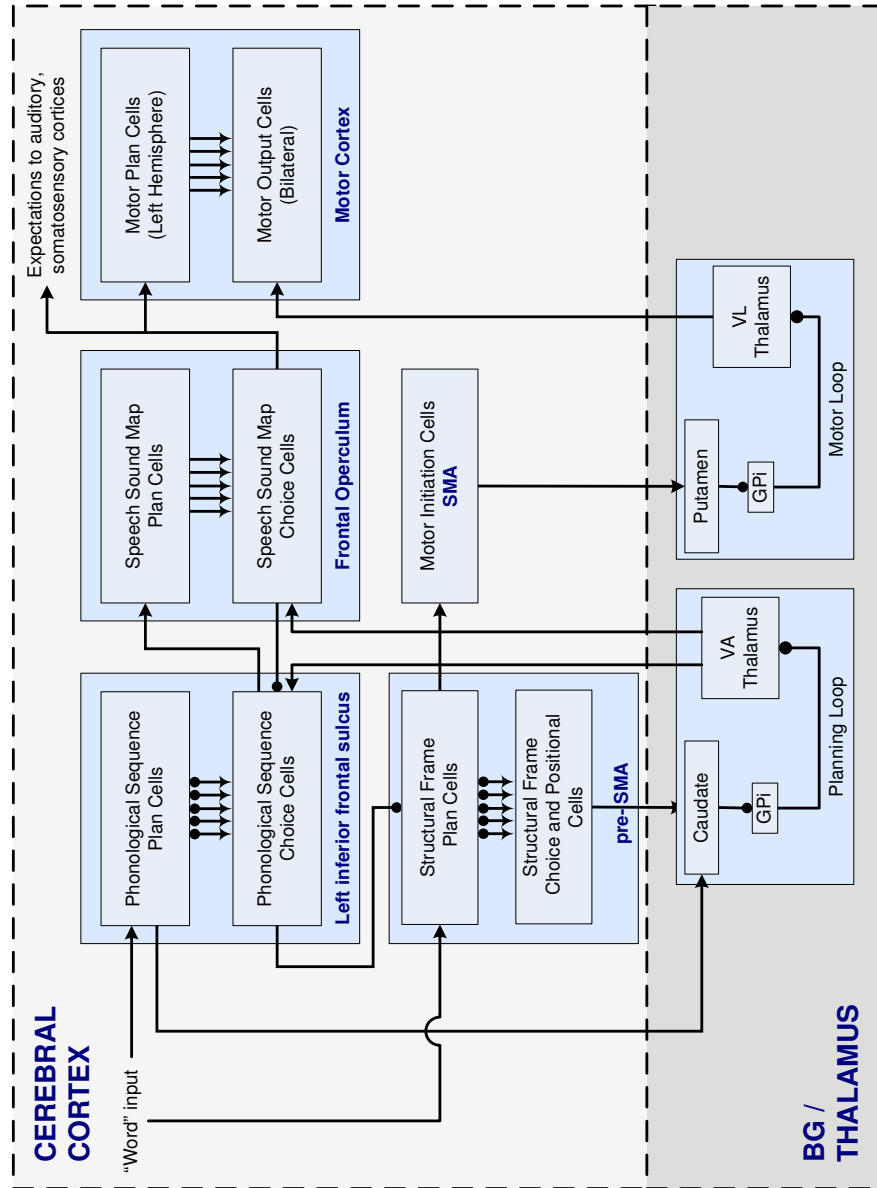


Figure 3.7: “Box-and-arrow” schematic of the primary GODIVA model components and their hypothesized cortical and subcortical correlates. Lines with arrows represent excitatory pathways, and lines with filled circles represent inhibitory pathways. Lines with both arrowheads indicate that connectivity between these modules features top-down excitatory connections and bottom-up inhibitory feedback connections. The inhibitory pathways shown in the cortical portion of the model are feedback pathways that suppress planning representations after their corresponding action has been taken.

be produced. These gradients over *plan cells* can be maintained through recurrent neural dynamics for a short duration throughout processing, and can robustly and appropriately handle new inputs as they arrive without disruption of ongoing performance, up to a certain item capacity limit determined by the signal to noise ratio of the representation. Both the IFS and pre-SMA plan layers thus take the form of “item and order memories” (Grossberg, 1978b,a) or, equivalently, planning layers in a competitive queuing circuit (Bullock and Rhodes, 2003).

In GODIVA the production process begins with the selection of the most active frame in the pre-SMA within a second pre-SMA layer (the choice layer). The breakdown of cortical representations into plan and choice layers with a columnar architecture is repeated throughout the model (see Figure 3.7). The activation of a pre-SMA choice cell initiates the firing of a chain of cells also in the pre-SMA, each corresponding to an abstract *position* (but not a specific phoneme) in the forthcoming syllable. These pre-SMA cells give input to a basal ganglia-mediated *planning loop*, which serves as an input *gate* to i) a distinct population of *choice cells* in the IFS region, and ii) choice cells in the *Speech Sound Map*, a component of the current DIVA model that is further specified in GODIVA. This planning loop selectively enables topographic *zones* of cells in the IFS choice layer that correspond to appropriate syllable positions for the forthcoming syllable only, as well as to selectively enable learned Speech Sound Map programs that match the abstract frame structure. Strong competition amongst IFS choice cells in each positional zone then results in a single “winning” representation within each active positional zone. As is standard in all CQ-based models, choice cells in both the IFS and pre-SMA selectively suppress their planning representations after becoming active. This allows for ongoing sequence performance.

Choice cells in the IFS form cortico-cortical synapses with cell populations in

the SSM that allow for the “read out” of motor programs as well as auditory and somatosensory expectations for simple learned speech sounds. The SSM is hypothesized to occupy the left inferior frontal gyrus / frontal operculum (BA44) and/or left ventral premotor cortex (Guenther et al., 2006). The IFS→SSM synapses are suggested to be learned at a somewhat late stage of development, after a child has developed well-defined phonetic / phonological perceptual categories for his or her language (see Section 3.3.3). These learned synapses (which are defined algorithmically in the model) allow the set of winning choice cells in the IFS choice layer to activate a set of potential “matching” motor programs represented by Speech Sound Map *plan cells*, with better matching programs receiving higher activations. Because one IFS choice cell is active *for each position* in the forthcoming syllable, this projection transforms a *phonological syllable* into a speech motor program.

Speech Sound Map plan cells give input, gated by the *planning loop*, to SSM choice cells. Competitive interactions amongst activated choice cells then lead to a “winner” being chosen for output to the motor apparatus. The model accounts for an additional basal ganglia loop (*Motor Loop* in Figure 3·7) that deals with the appropriate release of planned speech sounds to the motor execution system. The chosen SSM output cell is hypothesized to activate motoric plan cells primarily in the left-hemisphere motor cortex that, together with inputs from the SMA, bid for motor *initiation*. A new motor program will be initiated only upon completion of the previous program, for example. The uncoupling of the selection of motor programs from the timing of initiation allows the system to proceed with pre-motor selection prior to the completion of or, in some cases, even the initiation of the previous chosen program. This also allows for a simple mechanistic explanation of the differences between preparation and production and between *covert* and *overt* speech.

3.8 Model specification

This section discusses the various components of the GODIVA model in further detail, including the specification of a set of differential equations that controls the operation of model simulations. For reference, Table 3.8 provides a list of the symbols used in the various equations and the model representations to which they correspond.

In an attempt to reduce the complexity of the model, cortico-cortical inhibitory projections, which likely involve a set of intervening interneurons between two sets of excitatory neurons, are not explicitly modeled. Instead, the excitatory cortical neuron \rightarrow inhibitory interneuron \rightarrow excitatory cortical neuron disynapse is modeled as a single inhibitory synapse from a cortical neuron that, in the model, can also give excitatory projections.

It is important to note that the present model is “hand-wired.” That is, weights that are assumed to be modifiable through learning are algorithmically set within the range of values that learning must achieve for proper operation. Possible methods by which these modifiable synaptic weights can be learned are suggested in the Discussion section below (Section 3.10).

3.8.1 Phonological content representation in inferior frontal sulcus

Neurons in the region in and/or surrounding the inferior frontal sulcus (IFS) in the left hemisphere (BA 44/45/9) are hypothesized to be used in the short-term maintenance of the phonemes contained in a planned speech utterance. The IFS representation consists of two layers, one containing *plan cells* and one containing an identical corresponding set of *choice cells*. A plan cell and the corresponding

Table 3.1: Legend of symbols used to refer to cell populations in the GODIVA model specification.

Cell Type	Symbol
External Input to IFS	u^p
External Input to preSMA	u^f
IFS Phonological Content Plan Cells	p
IFS Phonological Content Choice Cells	q
pre-SMA Frame Plan Cells	f
pre-SMA Frame Choice Cells	g
pre-SMA Positional Chain Cells	h
Planning Loop Striatal Projection Cells	b
Planning Loop Striatal Interneurons	\underline{b}
Planning Loop GPi Cells	c
Planning Loop Anterior Thalamic Cells	d

choice cell are thought to represent a (simplified) cortical column⁶. This breakdown of representations into two layers constituting plan and choice cells is a repeated element throughout the model (see Figure 3-7). Figure 3-8 illustrates two such IFS columns from a single positional zone as well as their inputs and outputs.

The idealized cortical columns in this IFS representation are hypothesized to be tuned to a particular phoneme and to a particular abstract position in a syllabic frame. The IFS *map*, therefore, can be thought of as a two-dimensional grid, where each row corresponds to a particular phoneme and each column to a particular syllable position (see Figure 3-9). For the purposes of the model, 7 syllable positions are included. These positions correspond to a generic syllable template, such as that introduced by Fudge (1969) and also used in the model of short-term memory for

⁶This simplified breakdown of the layers in a cortical column is similar to the breakdown utilized in the detailed model of BG function of Brown et al. (2004). The two-layer simplification allows the model to incorporate two major empirical generalizations regarding cortico-BG and cortico-cortical projections. First, the dominant cortico-striatal projection is from layers 5a or above ("superficial") whereas the cortico-thalamic and cortico-sub-thalamic projections are from deeper layers (5b, 6). Second, the cortico-cortical projections are either from deep layers to superficial layers or from superficial layers to deep layers; cortico-cortical projections between layers of equivalent depth appear to be excluded (e.g. Barbas and Rempel-Clower, 1997).

words and non-words introduced by Hartley and Houghton (1996). Almost any English syllable can be represented in this template by assigning particular phonemes to particular template slots. In the GODIVA model, the middle (4th) position is always used to represent the syllable nucleus (vowel), and preceding consonants are loaded into preceding template positions, and succeeding consonants into succeeding template positions⁷. Within a particular syllable position (corresponding to the long axis in Figure 3-9), a gradient of activity across plan cells defines the serial order of the phonemic elements. For example, Figure 3-9 schematizes the representation of the planned utterance “gəv.di.və” in the IFS phonological planning layer. Competitive interactions in the IFS map model are restricted to *within position* interactions; in essence, therefore, this representation can be thought of as having multiple queues, one for each syllable position.

The model includes representations for 53 phonemes (30 consonants and 23 vowels) derived directly from the CELEX lexical database (Baayen et al., 1995). The set of cells in the IFS phonological content representation form an efficient *categorical* basis set for representing arbitrary speech sequences from a speaker’s language. This is an important principle of the GODIVA model in that it allows the model to represent and ultimately produce both often-encountered (and hence well learned) utterances and novel phonological “words” for which the speaker has no stored motor associations. Additionally, this representation allows the speaker to simultaneously plan *multiple* forthcoming syllables using this learned categorical space, a faculty that is crucial to fast fluent performance. It is important to note, however, that as depicted thus far, the representation fails to handle repeating elements in a speech plan. For example, in the syllable sequence “ta.ka”, the /a/ is repeated; if there existed just

⁷Due to the phonotactic rules of English, not all phonemes are eligible at all positions. For simplicity, this notion was not explicitly incorporated in the model, but its implications are worthy of further consideration.

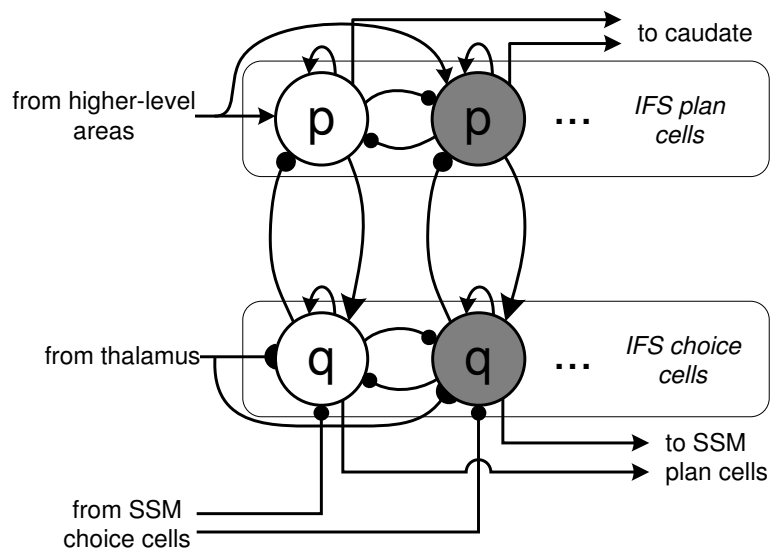


Figure 3-8: Schematic illustration of the structure of the GODIVA model's inferior frontal sulcus phonological content representation. The region is hypothesized to consist of a layer of plan cells (**p**; top) and a layer of choice cells (**q**; bottom), arranged into columns, each of which code for a planned phoneme in a given syllable position. The plan cells are loaded in parallel from other cortical or cerebellar regions. Choice cells, whose input from plan cells is gated by a syllable position-specific signal from the anterior thalamus, undergo a winner-take-all process within each gated zone. The activation of a choice cell suppresses its corresponding plan cell. This process results in the activation of a *phonological syllable* in the IFS choice field that can activate potentially matching syllable motor programs in the Speech Sound Map. Choice cell activations can be suppressed upon the selection of a specific Speech Sound Map motor program.

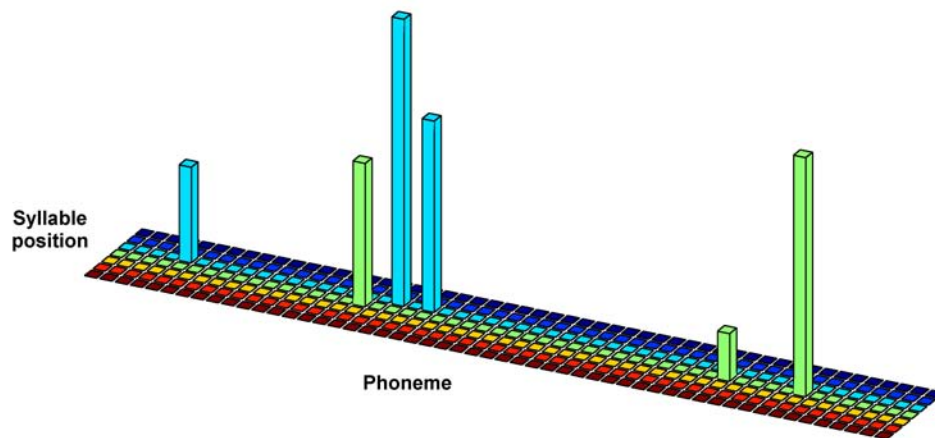


Figure 3-9: This figure schematizes the layout of cells in the IFS phonological content representation. Both plan and choice layers in the region use the same representation; shown here is the plan layer, which has dynamics that allow multiple parallel items to be coterminally active. The long axis in the IFS map corresponds to specific phonemes, and the short axis corresponds to abstract serial positions in a generic syllable template. Cells compete with one another through lateral inhibition along the long axis. This map illustrates an idealized plan that corresponds to the syllable sequence “*gəv.di.və.*” The height of the vertical bar at a particular entry in the map corresponds to a cell’s activation level. Note that entries in the schematic of the same color indicate these cells code for the same syllable position; in this representation, there are 3 active cells at each of syllable positions 3 and 4 in the template, corresponding to three [CV] syllables.

one cell to code this phoneme in the nucleus position, it would be impossible, using the gradient approach, to represent the *order* of two occurrences of that phoneme. It is therefore assumed that for each cell in the 53×7 representation depicted in Figure 3.9, there exist multiple “copies” of that unit. This augmentation requires some additional ad hoc mechanisms, particularly during response suppression, which are discussed at appropriate places in the specification below. For simplicity of presentation, the equations below make reference to just one copy of each representational cell.

The activity of a cell p_{ij} , representing phoneme i at syllable position j in the two-dimensional IFS phonological planning layer matrix \mathbf{p} , is governed by the shunting differential equation (with respect to time):

$$\dot{p}_{ij} = -A_p p_{ij} + (B_p - p_{ij}) (\alpha u_{ij}^p + [p_{ij} - \theta_p]^+) - p_{ij} \left(\sum_{k, k \neq i} W_{ik} p_{kj} + 10y ([q_{ij} - \theta_q]^+) \right) \quad (3.1)$$

In this equation, the first term is a passive decay. In the absence of any inputs, a cell’s activity will decay to resting state (identically zero for all cells) at a rate determined by the scalar constant A_p . The second term models excitatory inputs to the cell, which drive activity in the positive direction. The initial multiplicative term $(B_p - p_{ij})$ enforces an upper bound B_p to cell activity. Such multiplicative or shunting terms (Grossberg, 1973) are motivated by cell membrane properties (e.g. Hodgkin and Huxley, 1952). The final term, which also enforces a lower activity bound (of zero) models the inhibitory inputs to the cell, which drive activity in the negative direction.

In (3.1) there are two sources of excitatory input. First, u_{ij}^p is an external input that corresponds to a “word⁸” that gives input, in parallel, to the IFS phonological

⁸Here the term *word* is used loosely to indicate a portion of a planned utterance that is *at least* as large as a syllable. This could represent a real word, a morpheme, or a pseudoword, for example.

plan representation. This input is assumed to arrive from one or more of three brain areas not explicitly treated in the model:

1. A higher-level linguistic processing area involved in the morpho-syntactic processing of an internally generated communicative concept.
2. A phonological processing region likely in the parietal cortex that can load the phonological output system when the task is, for instance, reading, or repetition.
3. The inferior right-hemisphere cerebellum, which is hypothesized to assist in “fast-loading” of phonological content into this buffer. (See Discussion in Chapter 2).

The role of this external input is to provide a pulsed input to the IFS planning layer that instantiates a gradient across its units, which represents the ordered set of phonemes in the input “word.” This input is *gated* by a multiplicative term α that can be used to ensure that the activity of cells receiving new inputs is not allowed to exceed the activity level of currently active plans in the IFS, thus maintaining the order of planned speech elements (see e.g. Bradski et al., 1994). The second excitatory input to cell p_{ij} is from itself. Here θ_p is a noise threshold set to a low value and $[]^+$ indicates half-wave rectification. Such recurrent self-excitation allows this layer to maintain a loaded plan over a short duration even in the absence of external inputs.

Cell p_{ij} receives inhibitory inputs from cells representing other phonemes in the planning layer *within the same syllable position*. These inhibitory inputs are weighted by entries in the adjacency matrix \mathbf{W} . In the simplest case, entry W_{ik} is simply 1 for $i \neq k$ and 0 for $i = k$; these weightings can be modified, however, to incorporate a notion of physical distance in the cortex, allowing for at least a partial explanation

of phonemic similarity effects (see Discussion). In the simulations presented here, this simple weighting using 0's and 1's was utilized. Finally, cell p_{ij} also receives a strong inhibitory input from cell q_{ij} , the corresponding cell in the *IFS Choice Layer*. This input is thresholded by term θ_q , and subjected to a faster-than-linear *activation function* (cf. Grossberg, 1973), chosen to be $y(x) = x^2$. This activation function can be thought of as a non-linear response (e.g. spike rate varies non-linearly with membrane potential) inherent to choice cell neurons. The same activation function also guides self-excitatory activity amongst the choice cells in (3.2).

The activity of a cell q_{ij} in the IFS choice layer \mathbf{q} is given by:

$$\dot{q}_{ij} = -A_q q_{ij} + (B_q - q_{ij}) (d_j [p_{ij} - \theta_p]^+ + y(q_{ij})) - q_{ij} \left(\sum_{kj, k \neq i} W_{ik} y(q_{kj}) + \Gamma_{ij} \right) \quad (3.2)$$

where A_q is again a passive decay parameter and B_q is again an upper bound on cell activity. The excitatory inputs include a recurrent self-excitatory term ($y(q_{ij})$) and selective input from the IFS plan cells in the same cortical column. This input is gated by the multiplicative term d_j , which represents a signal hypothesized to arise from the ventral anterior thalamus as the output of the basal ganglia mediated *Planning Loop* (see Figure 3.7). The dynamics of this loop are specified in Section 3.8.3. Ultimately, the signal d_j serves as a gate to a particular cell population in the IFS choice layer that, when opened, allows a winner-take-all competition to occur within cells in that zone. In the model, these *gateable zones* (cf. Brown et al., 2004) correspond to the positional representations in the IFS map (i.e. the short axis in Figure 3.9).

The IFS choice cell q_{ij} receives inhibitory inputs from all other cells within the same gateable zone (syllable position). The action of the inhibitory cells is again subjected to the faster-than-linear signal function y . The resulting dynamics of this layer are such that it is typically quiescent, but when a thalamic input gates open

a positional zone, inputs from the IFS plan cells activate their corresponding choice cells, which in turn compete via non-linear lateral inhibition, resulting in a competitive choice (winner-take-all) process (cf. the competitive layer in a competitive queuing model; Figure 3.4) within a positional zone. Once a choice cell becomes active, it will maintain that activation through the use of recurrent interactions. The cell's activity can be quenched via the potentially strong inhibitory input Γ_{ij} . This term represents a *response suppression signal* which arrives via interneurons from the *Speech Sound Map choice layer*. The dynamics of Speech Sound Map cells are specified in Section 3.8.4. The value of Γ_{ij} at time t is given by:

$$\Gamma_{ij}(t) = 10Z_k^{ij}s_k(t) \quad (3.3)$$

where Z_k^{ij} is 1 if phoneme i is a part of syllable motor program k in syllable position j , and 0 otherwise. Γ_{ij} , therefore, models the suppression of IFS phonological choice cells by chosen speech motor program cells in the Speech Sound Map. It is important to note that only the phonemes that are part of the chosen motor program in the Speech Sound Map are suppressed. This allows the model to produce a novel or not well learned syllable from targets representing its constituent segments. This issue is discussed further below.

3.8.2 Structural “frame” representations in pre-SMA

Sets of neurons in the pre-SMA are hypothesized to serve as representations for structural *frames* for common syllable types and for their abstract “slots” or *positions*. For example, the model pre-SMA contains cells that code for the entire abstract syllable type CVC (consonant-vowel-consonant) as well as for C in onset position, V in nucleus position, and C in coda position. These representations are assumed

to be acquired through the extraction of structural regularities that occurs due to perceptual and motor experience with a language. Acquiring this discrete set of representations is feasible because few syllable frames are necessary to account for all of the syllable types used in a language. In English, based on an analysis of frequency of usage tables in the CELEX lexical database (Baayen et al., 1995), just 8 different syllable frames account for over 96% of all syllable productions.

During phonological encoding, the model's frame representations in the pre-SMA are activated in parallel with the activation of the IFS phonological content representation (see Figure 3.7). As is the case with phoneme representations in the IFS planning layer, multiple pre-SMA frame cells can be active co-temporally in its planning layer. The use of two layers representing an idealized cortical column is repeated in the model's pre-SMA. As in the CQ framework, the relative activation level of the pre-SMA plan cells codes for the serial order of the forthcoming syllable frames, with more activity indicating that a frame will be used earlier. In essence, the model loads forthcoming speech plans into two parallel and complementary queues, one in the IFS and one in the pre-SMA. This helps to solve a combinatorial problem that would result from requiring a representation that could code *all possible* combinations of frame and content. Such a representation would require tremendous neural resources in comparison to a representation such as the one proposed, which separates the representational bases into two relatively small discrete sets. Another advantage to learning abstracted representations of syllable frames separately from representations of phonemic content is that it appears to facilitate the rapid acquisition of speech motor programs through a substitution process (MacNeilage, 1998). Specifically, a child can potentially use the same *frame* with different eligible phonemes at each position to quickly visit the space of potential syllable-sized speech motor productions.

While the separation of frame and content offers computational advantages, it also necessitates a processing stage in which the two representations are brought back together in order to select and enable appropriate motor programs. In the model this occurs by way of a basal ganglia mediated “planning loop”, described specifically in Section 3.8.3, which enables the selection of the constituent phonemes from the IFS planning layer corresponding to each position in the forthcoming syllable *only*. This process results in the parallel activation of a *phonological syllable* (cf. Cholin et al., 2004) representation in the IFS choice cells.

For the sake of computational efficiency, the number of available frame types implemented in the model was limited to 7. These included [CV], [CVC], [VC], [CVCC], [CCV], [CCVC], and [VCC], which are the most common types according to the CELEX database. Syllables of other types were not eligible for selection by the model. As was required to allow for encoding the serial order of repeating phonemes, the model pre-SMA actually contains multiple “copies” of each syllable frame cell.

The model pre-SMA contains not only cells that code for the entire abstract frame of a forthcoming syllable, but also *chains* of cells that fire rapidly in sequence, which code for the individual abstract serial positions within the syllable frame. These two types of cells, one that codes for a whole sequence (in this case a sequence refers to the sequence of constituent syllable positions within a syllable frame), and another type that codes for a specific serial position within that sequence, are similar to cells that have been identified in the pre-SMA in monkey studies (see Section 3.5.2). In the GODIVA model, the selection of a syllable frame cell (e.g. activation of a pre-SMA choice cell) also initiates the readout of the chain of cells coding its constituent structural positions (but not specific phonemes). The structure and operation of the pre-SMA in the model are schematized in Figure 3.10.

For a single syllable, the temporal pattern of activity in the pre-SMA proceeds

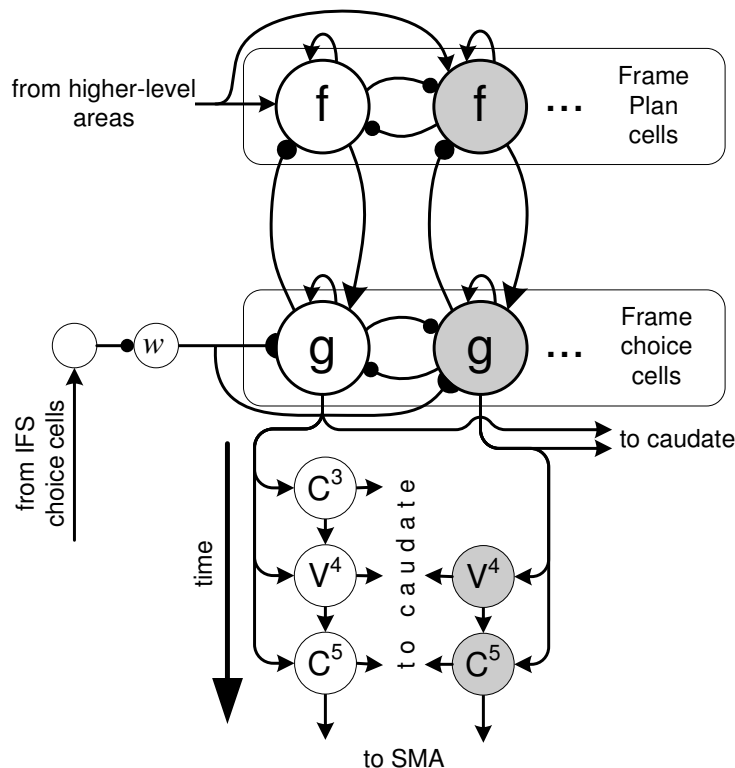


Figure 3-10: Schematic illustration of the structure and function of model cells hypothesized to exist in the pre-SMA. This region consists of a layer of plan cells (top) and a layer of choice cells arranged into columns, each of which correspond to the same abstract syllable frame (CV structure). When a pre-SMA choice cell is activated (the forthcoming frame is chosen), the cell gives inputs to a chain of cells, each of which corresponds to a position within the abstract syllable frame. These cells fire rapidly and in order, according to the vertical arrow labeled “time” (bottom left). In this schematic, the first pre-SMA cortical column codes for the syllable frame type [CVC], and the second column codes for the frame type [VC]. Note that the inputs to caudate are aligned such that the [V] position in both cases gives input to the same caudate channel (corresponding to positional zone 4). Cell w gates the pre-SMA frame choice process.

as follows. First, a single choice cell is activated, corresponding to the most active syllable frame among pre-SMA plan cells; upon the instantiation of this choice, the corresponding pre-SMA plan cell is suppressed. Next, the choice cell activates the first position cell in the positional chain corresponding to this syllable type. This cell and subsequent cells in the positional chain give their inputs to *zones* in the caudate which have a one-to-one correspondence with i) positions in the syllable template, and ii) gateable positional zones in the IFS. Such cortico-striatal projections form the inputs to the planning loop of the basal ganglia, which eventually enables the selection of the forthcoming syllable's constituent phonemes in the IFS choice field. When the positional chain has reached its completion, the last cell activates a corresponding cell in the SMA-proper which, effectively, informs the *motor* portion of the circuit that the planning loop has prepared a new phonological syllable.

The pre-SMA cells which code for entire syllable frames are modeled by equations very much resembling those that govern IFS cell activity (see Equations (3.1) and (3.2)). These layers, again, mimic the competitive queuing architecture, and take the form of shunting equations with three terms: the first a passive decay, the second the excitatory inputs, and the third the inhibitory inputs to this cell. This form is maintained throughout the specification of the model's equations. The activity of the i^{th} frame cell in the pre-SMA plan layer \mathbf{f} (see Figure 3-10) is given by:

$$\dot{f}_i = -A_f f_i + (B_f - f_i) \left(\alpha u_i^f + [f_i - \theta_f]^+ \right) - f_i \left(\sum_{k \neq i} f_k + 10y ([g_i - \theta_g]^+) \right) \quad (3.4)$$

where $\mathbf{U}^{\mathbf{k}}$ is the external input to the pre-SMA, assumed to arrive from the same source that provides input $\mathbf{U}^{\mathbf{f}}$ to the IFS. Equation (3.4) has a nearly identical form to that which governs IFS plan cells (3.1).

The activity of pre-SMA choice cell g_i is governed by:

$$\dot{g}_i = -A_g g_i + (B_g - g_i) (\omega [f_i - \theta_f]^+ + y(g_i)) - g_i \left(\sum_{k \neq i} y(g_k) \right) \quad (3.5)$$

Here the scalar ω gates the pre-SMA frame choice process. ω is modeled as a binary input, where its value is 1 when the IFS choice field is *empty*, and 0 when there are significantly active cells within that field. Without such a gate, the pre-SMA choice process could proceed without pause through selection of each of the syllable frames represented in the graded pattern \mathbf{f} . Instead, this gate ω requires the pre-SMA module to *wait* until the currently active syllable has been chosen for production on the *motor* side of the circuit. At this time, the choice of the *next* frame may proceed. This gating is implemented algorithmically but can be achieved through a cortico-cortical projection between IFS and pre-SMA by way of an inhibitory interneuron. This is schematized in Figure 3-10, where it is assumed that tonically active cell ω is suppressed fully when *any* IFS choice cells are active above some low threshold.

As noted above, the activation of a pre-SMA choice cell also initiates a serial chain of cells that code for individual abstract positions in the syllable. The activity of the j^{th} cell in the positional chain corresponding to syllable frame k is specified algorithmically by:

$$h_j^k(t) = \begin{cases} 1 & \text{if } (t_0 + (j-1)\tau) \leq t \leq (t_0 + j\tau) \\ 0 & \text{otherwise} \end{cases} \quad (3.6)$$

where t_0 is the time at which the pre-SMA choice cell g_k exceeds a threshold θ_g (the time at which it is “chosen”) and τ is a short duration for which each cell in the chain is uniquely active. Each of these cells gives input to a cell in the striatum corresponding to the same positional zone (see below). The deactivation of the final

cell in the chain activates an SMA-proper cell that codes for the appropriate syllable type k .

3.8.3 Cortico-striato-pallido-thalamo-cortical “planning loop”

Following the evidence presented in Section 3.5.3, the model proposes that two distinct basal ganglia loop circuits form competitive gating mechanisms for cortical modules during the production of syllable sequences. The first loop, the *planning loop*, serves to enable cortical zones in the choice layer of the model’s left inferior frontal sulcus. The planning loop receives inputs from the IFS *plan cells* (\mathbf{p}) as well as from the pre-SMA positional cells (\mathbf{h}). Following Brown et al. (2004), the GODIVA model uses the one-to-many projection from thalamic output cells to cells in the cortex to perform this gating function. The model’s basal ganglia circuitry is much simplified in comparison to other detailed BG models, but remains compatible with, for example, Brown et al. (2004). Although there is significant convergence within the cortico-striatal-pallido pathway, the model treats these sets of synapses as a set of competitive *channels*, each represented by 1 striatal projection neuron (b), 1 striatal interneuron (\underline{b}), and 1 pallidal (GPi) cell (c). This highly idealized circuitry is depicted in Figure 3-11. There is a one to one correspondence between these channels in the GODIVA model’s planning loop and the gateable cortical zones in the IFS choice layer. Furthermore, these zones, again, correspond directly to the 7 abstract syllable positions in a syllable template.

The activity of the striatal projection neuron in BG channel j is given by:

$$\dot{b}_j = -A_b b_j + (B_b - b_j) \left(h_j \wedge \left[\sum_k p_{kj} - \delta \right]^+ \right) - b_j \left(\sum_{k \neq j} y(\underline{b}_k) \right) \quad (3.7)$$

where the symbol \wedge is the logical AND operator, assumed here to output 1 when both of its operands have value greater than zero, and 0 otherwise. It is used here

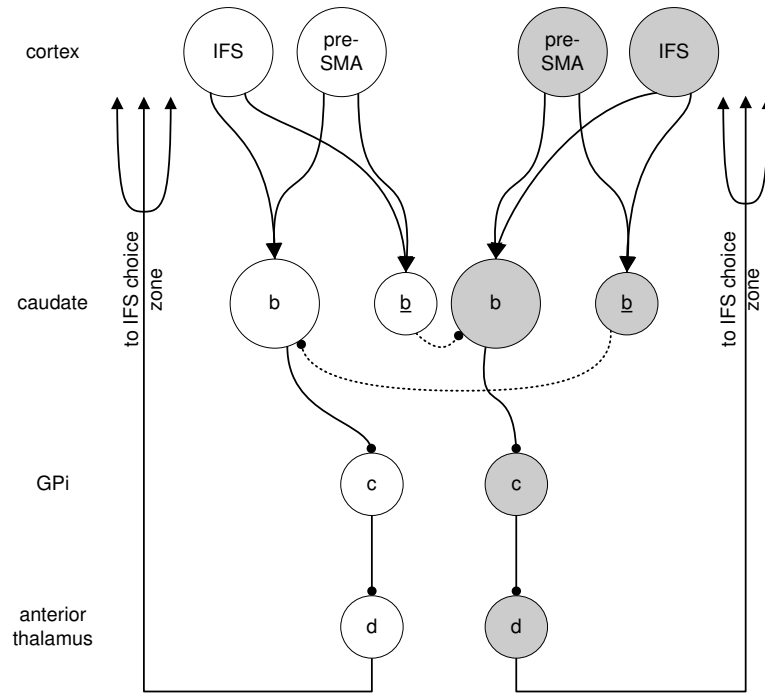


Figure 3-11: Schematic illustration of “channel” architecture through the basal ganglia planning loop. Each channel corresponds to an abstract serial position in the generic syllable template. The modeled caudate consists of one projection neuron (b) and one inhibitory interneuron (\bar{b}) in each channel. The channels compete via feedforward inhibition in the caudate. Caudate projection neurons give inhibitory projections along the *direct pathway* to a modeled GPi cell (c). The GPi cell, in turn, inhibits the anterior thalamic cell d . The successful activation of a channel disinhibits its specific thalamic cell, which in turn “opens the gate” to a zone in the inferior frontal sulcus phonological choice layer through a multiplicative interaction.

to indicate that *both* supra-threshold activity in one or more IFS plan cells tuned to position j and significant input from pre-SMA cells coding for position j are required to drive positive activation of this striatal projection cell. The cell b_j also receives strong (modeled as faster-than-linear) feed-forward inhibition from striatal interneurons \underline{b}_k in the other BG channels ($k \neq j$).

The activity of a striatal inhibitory interneuron in channel j is governed by the very similar equation:

$$\dot{\underline{b}}_j = -A_{\underline{b}}\underline{b}_j + (B_{\underline{b}} - \underline{b}_j) \left(h_j \wedge \left[\sum_k p_{kj} - \delta \right]^+ \right) - \underline{b}_j \left(\sum_{k \neq j} y(\underline{b}_k) \right) \quad (3.8)$$

Thus, the model's corticostriatal cells in both the IFS and pre-SMA give identical inputs to the projection neurons and inhibitory interneurons in the model's caudate.

The striatal projection neurons connect to model GPi cells within the same BG channel by means of an inhibitory synapse. The activity of the GPi cell c_j , which is itself inhibitory to a corresponding thalamic cell d_j , is given by:

$$\dot{c}_j = -A_c c_j + \beta_c (B_c - c_j) - c_j (b_j) \quad (3.9)$$

where β_c and B_c control the level of spontaneous tonic activation of the GPi cell. Such tonic activation is required for the BG model to be able to achieve a correct *net* effect within a channel. Specifically, the corresponding thalamic cell (d_j) should be typically silent, but should become phasically activated upon the selective competitive activation of BG channel j . To achieve this result, GPi cells are tonically active, but show a *pause* response when they receive inhibitory input from the striatal projection neuron in the same channel. Thus, because the projection from the GPi cell c_j to the anterior thalamus cell d_j is inhibitory, a pause response in c_j will *disinhibit* d_j , and thereby enable the cortical selection process in zone j of the IFS

choice layer. The activity of the thalamic cell d_j , which diffusely projects to zone j in the IFS choice layer is given by:

$$\dot{d}_j = -A_d d_j + \beta_d (B_d - d_j) - d_j (c_j) \quad (3.10)$$

Here β_d and B_d control the amplitude of the rebound excitation of the thalamic cell. A transient decrease in the inhibitory input c_j thus leads to transient activation of d_j , enabling the cortical selection process for syllable position j in the IFS choice field. It is interesting to note that such thalamic rebound excitation has recently been shown in the homologous BG – thalamic circuit controlling birdsong production (Person and Perkel, 2005).

3.8.4 Speech sound map

The Speech Sound Map (SSM) is a component of the DIVA model (Guenther, 1995; Guenther et al., 1998, 2006) that is hypothesized to contain cells that “read out” motor programs and sensory expectations for well-learned speech sounds. In the DIVA model, the tonic activation of an SSM cell (or ensemble of cells) is required to read out the stored sensory and motor programs throughout the production of the sound. To properly couple the system described herein with the DIVA model, GODIVA must provide this selective, sustained excitation to the appropriate SSM cells.

As is the case with its other cortical representations, the GODIVA model posits a breakdown of SSM cells into two layers, again labeled *Plan* and *Choice* cells (see Figure 3.7). In this representation, each idealized cortical column corresponds to a well-learned syllable or phoneme. Unlike the plan layers in the IFS and pre-SMA, the activation pattern across SSM plan cells does not code for *serial order*, but rather

codes for the *degree of match* between the set of active phonological cells in the IFS choice layer (e.g. the forthcoming phonological syllable) and the stored sensorimotor programs associated with the Speech Sound Map cells. This match is computed via an inner product of the IFS choice layer inputs with synaptic weights that are assumed to be learned between these cells and the SSM plan cells. In the current implementation of the model, these weights are simply “hand-wired” such that the synapse Z_k^{ij} from IFS choice cell q_{ij} (which codes phoneme i at syllable position j) to Speech Sound Map plan cell r_k is given by:

$$Z_k^{ij} = \begin{cases} \frac{1}{N_k} & \text{if } r_k \text{ includes phoneme } i \text{ at syllable position } j \\ 0 & \text{otherwise} \end{cases} \quad (3.11)$$

where N_k is the total number of phonemes in the syllable coded by SSM plan cell r_k . Such a specification of synaptic weights indicates that an SSM plan cell receives equally weighted input from each IFS choice cell that codes its constituent phonemes in their proper syllabic positions, and receives no input from other IFS choice cells. Furthermore, the sum of synaptic weights projecting to any syllable program in the SSM plan layer is equal to 1. Mathematically, the L1-norm of vector \mathbf{Z}_k is 1. Learning rules that conserve synaptic strength in this way have been proposed elsewhere (e.g. von der Malsburg, 1973; Grossberg, 1976), and similar conservational principles have been observed empirically (Markram and Tsodyks, 1996). An exception to the synaptic weight rule (3.11) is made for SSM cells that code single phoneme targets (as opposed to entire syllables). In the model implementation, these cells

have synaptic inputs set equal to:

$$Z_k^{ij} = \begin{cases} 0.85 - 0.05j & \text{if } r_k \text{ codes phoneme } i \\ 0 & \text{otherwise} \end{cases} \quad (3.12)$$

This algorithmic specification dictates that the input to SSM plan cells that code for single phoneme targets is *weighted by the position* in which the pre-synaptic IFS choice cell is active, such that inputs from earlier positions in the syllable have greater weight. This specification allows the SSM plan cell inputs to maintain the serial order of the constituent phonemes in the IFS choice field in the case that the syllable must be produced from sub-syllabic motor programs (e.g. when there is no matching syllable sized SSM representation for the forthcoming phonological syllable).

The activity level of cell k in the SSM plan layer representation \mathbf{r} is governed by the shunting equation:

$$\dot{r}_k = -A_r r_k + (B_r - r_k) \left(\sum_i \sum_j Z_k^{ij} y([q_{ij} - \theta_q]^+) + [r_k - \theta_r]^+ \right) - r_k \left(\sum_{n \neq k} r_n \right) \quad (3.13)$$

The double sum in the excitatory term above computes the net excitatory input from cells in the IFS choice field (\mathbf{q}) to the cell r_k , which is weighted by the synaptic strengths specified in the input weight matrix \mathbf{Z}_i . Cell r_k also receives self-excitatory feedback (subject to a low noise threshold θ_r) and lateral inhibitory input from all other cells in the SSM plan layer. The dynamics determined by (3.13) are such that, as in the other *plan* layers in the model, multiple cells can sustain their activation cotemporally.

The SSM plan cell r_k gives specific excitatory input to the SSM choice cell s_k

within the same idealized cortical column. The activation of s_k is given by:

$$\dot{s}_k = -A_s s_k + (B_s - s_k) (r_k + 10y([s_k - \theta_s]^+)) - s_k \left(\sum_{j \neq k} [s_j - \theta_s]^+ + \Omega \right) \quad (3.14)$$

where y is again a faster-than-linear signal activation function, ultimately resulting in winner-take-all dynamics within the layer \mathbf{s} . Ω models a non-specific response suppression signal that arrives from the DIVA model (the articulatory portion of the circuit) indicating the *impending* completion of the production of the current syllable motor program. When Ω is transiently large, the result is to quench activation of the current winning cell in \mathbf{s} , followed by the re-instantiation of a new winner, corresponding to the most active SSM program in the plan layer \mathbf{r} . The DIVA model can provide such a suppression signal prior to actual completion of articulation because of the inherent delays between sending a motor command and the effect that that motor command has on the articulators. Such delays in the production model have been considered by Guenther et al. (2006). Alternatively, in *covert* or internal speech, this completion signal can arrive from elsewhere, allowing the model to sequence through SSM programs without actually overtly articulating them.

3.8.5 Response release via the “motor loop”

The initiation or release of chosen speech motor programs for overt articulation is hypothesized to be controlled by a second loop through the basal ganglia, the *motor loop*. The proposal that *two* loops through the basal ganglia, one mediated by the head of the caudate nucleus, and one mediated by the putamen, are important in cognitive and motor control of speech production respectively, is supported by intraoperative stimulation results (Robles et al., 2005). In the model, the motor loop receives convergent input from the SMA and motor cortex and gates choice (or

execution) cells in the motor cortex (see Figure 3.7). The motor loop through the basal ganglia receives inputs at the putamen, as opposed to the planning loop, which receives its inputs, which arrive from “higher-level” prefrontal regions, at the caudate nucleus (cf. Alexander et al., 1986; Alexander and Crutcher, 1990). The motor loop also gives output to the ventrolateral thalamus, as opposed to the ventral anterior thalamic targets of the model’s planning loop.

Currently, the motor loop in the GODIVA model is not specified with the same level of detail as the previously discussed planning and selection mechanisms in the model. To achieve the same level of detail, it will be necessary to *fully* integrate the circuits described above with the existing DIVA model (e.g. Guenther et al., 2006). Nevertheless, a conceptual description of these mechanisms is possible, and follows from the general architecture of the higher-level portions of the model. Specifically, the activation of an SSM choice cell representing the forthcoming speech motor program is hypothesized to activate plan cells in the left motor cortex. These plan cells do not directly drive movement of the articulators, just as plan cell activity in other modules in GODIVA does not drive activity beyond that cortical region. Instead, overt articulation in the model requires the *enabling* of motor cortex choice cells via the BG-mediated motor loop. To “open the gate” and initiate articulation, the motor loop requires convergent excitatory inputs from the motor cortex plan cells and from the SMA-proper. This notion is based on three major findings from the fMRI study described in Chapter 2, which have also been described elsewhere in the literature: i) that *overt* articulation involves specific additional engagement of the SMA-proper, ii) that the *putamen* is particularly involved when speech production is overt, and iii) that the left hemisphere motor cortex may become active for covert speech or for motor preparation, but when speech is made overt, the motor cortex in both hemispheres is additionally engaged. These findings are discussed in more

detail in Chapter 2.

Table 3.2 provides a summary of the sequence of steps that the model goes through in order to produce a sequence of syllables.

3.9 Simulation results

This section describes simulations performed to verify that the model described above performs as designed. The model has been successfully tested for a variety of syllable sequences. Figure 3-12 and Figure 3-13 demonstrate the time courses of activity in several key components of the model during the planning and production of the syllable sequence “*gəv.di.və*” under two different assumptions about the model’s initial state.

3.9.1 Performance of a sequence of well-learned syllables

In the first simulation, the model is tasked with producing this sequence with the assumption that each individual syllable (“*gəv*”, “*di*”, and “*və*”) has a learned representation that is stored in the model’s Speech Sound Map. Sensorimotor programs for these syllables must be acquired by the DIVA portion of the circuit; this learning process is not explicitly simulated here. In this simulation, the 1000 most common syllables from the CELEX database (Baayen et al., 1995) (which include the three syllables to be performed here) are included in the model’s Speech Sound Map representation. The “input” to this simulation is a pulse that activates the two complementary gradients in the pre-SMA and IFS plan layers. This pulse is applied at the time indicated by the first arrow in each sub-figure in Figure 3-12. This input activates a gradient across the /g/, /d/, and /v/ phoneme cells in syllable position 3 (onset consonant) and a gradient across the /əv/, /i/, and /ə/ phoneme cells in syllable position 4 (vowel nucleus) in the IFS plan layer, as well as a gradient across

Table 3.2: A concise algorithmic summary of the steps that the GO-DIVA model takes to perform a syllable sequence.

1. Complementary activity gradients are loaded into the IFS plan and pre-SMA plan layers.
2. The most active syllable frame, corresponding to the 1st syllable in the sequence becomes active in the pre-SMA choice layer.
3. The corresponding cell in the pre-SMA plan layer is suppressed.
4. The active pre-SMA choice cell initiates the serial readout of a chain of cells corresponding to its abstract positions.
5. The active positional cell activates a BG planning loop channel, disinhibiting a thalamic cell, and enabling the appropriate positional zone in the IFS choice layer.
6. The most active phoneme in the IFS plan layer for this positional zone becomes active in the IFS choice field.
7. The corresponding phoneme cell in the IFS plan layer is suppressed.
8. Steps 5-8 are repeated for each serial position in the chosen syllable frame.
9. The now simultaneous activation of one phoneme for each syllable position in the IFS choice layer activates potential sensorimotor program “matches” in the SSM plan layer.
10. The best-matching SSM program is activated in the SSM choice layer.
11. Motor cortical plan cells are activated in the left motor cortex.
12. This program’s constituent phonemes are suppressed in the IFS choice layer.
13. If this action *empties* the IFS choice field, then Steps 2-9 can be performed for the *next* syllable.
14. Convergent SMA and M1 plan cell activity allows overt production to be initiated for the currently active SSM choice cell.
15. A completion signal transiently suppresses SSM choice cell activity, quenching the currently active program and allowing a new winner to be chosen.
16. Steps 2-15 are repeated until no cells are active in the pre-SMA (and IFS) planning layers.

three “copies” of the [CV] frame cell in the pre-SMA. In Figure 3-12 (A and B) it can be seen that the activation levels of the phonemes in these positional zones rise from the initial state of 0 and begin to equilibrate with each cell taking on a different activation level. The different time courses of activation in these plots are labeled by the phoneme that each cell represents. These resulting activation gradients, of course, are essential to the model performing the sequence in the correct order.

After the first CV frame representation is chosen via the pre-SMA choice layer, positional zones 3 and 4 are *enabled* in the IFS choice layer in rapid succession. This allows for the choice of the most active phoneme in each IFS plan layer positional zone. Figure 3-12 (C and D) shows this choice being made, resulting in the strong, sustained activation of the phonemes /g/ and /əʊ/ in IFS choice zones 3 and 4, respectively. The choice is made in zone 4 at a slightly later time than in zone 3. By comparing the sub-plots in Figure 3-12, it can be seen that, immediately following the choice of /g/ and /əʊ/ (panels C and D), the representations for each phoneme in the IFS plan representation (panels A and B) are rapidly suppressed. Activity in the IFS plan layer also then re-equilibrates, leaving only two phonemes in each zone active with a now larger difference in relative activation levels.

The cotemporal activation of /g/ and /əʊ/ in the IFS choice layer (panels C and D) causes activity to arise in the model’s SSM plan cells (panel E). It can be seen that multiple representations become active as there are multiple partially matching sensorimotor programs stored in the model’s Speech Sound Map. The most active SSM plan cell, however, codes for the best matching syllable (in this case “gəʊ”). This allows this syllable representation to become active in the SSM choice layer (panel F). As soon as “gəʊ” becomes active in (F), its constituent phonemes in the IFS choice layer (panels C and D) are suppressed. The resulting *lack of activity* in the IFS choice layer gates the choice of the *next* CV syllable frame in the pre-SMA

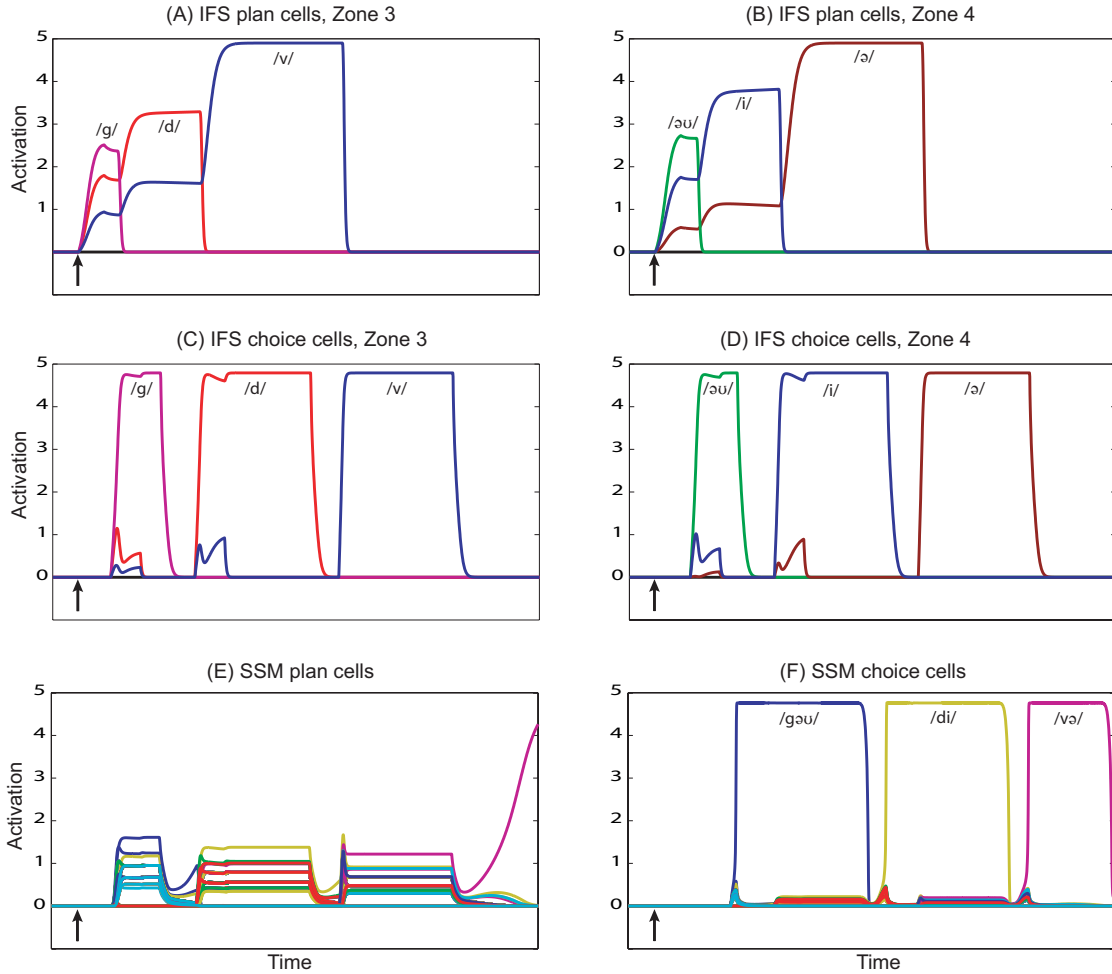


Figure 3.12: Simulation result showing the production of the three syllable sequence “*gəv.di.və*”. In this simulation, each of the three syllables has a corresponding stored Speech Sound Map representation. Each plot shows time courses of cell activity in different model components. The x-axis in each plot is time, and the y-axis is activation level (both in arbitrary model units). The arrows in each plot indicate the onset of the external input pulse, effectively the start of the simulation. These simulation results are described in detail in the text.

(not shown), allowing the model to begin *preparing* the syllable “*di*” (all the way to the stage of activating potential SSM matches in the SSM plan cells) while it is still *producing* the syllable “*gəv*” (compare panels C, D, and E to panel F). The syllable “*di*,” however, can only be chosen in the SSM choice layer (panel F) upon the receipt of a non-specific suppression signal arriving from the articulatory control circuit. The effect of this suppression signal is to transiently quench all activity in the SSM choice layer, which can be seen by the fast decrease in activation of the cell coding for “*gəv*” in panel F. Upon removal of this suppression signal, “*di*,” the most active SSM plan representation is chosen in the SSM choice layer. This entire process iterates until there are no remaining active cells in the pre-SMA or IFS plan layers.

It can be seen from Figure 3-12 (F) that the syllable motor programs corresponding to the desired syllables are activated in the proper order. This is precisely what is required in order to interface GODIVA with the DIVA model, which can then be used to control a computer-simulated vocal tract to realize the desired acoustic output for each syllable.

3.9.2 Performance from sub-syllabic targets

In the development of the GODIVA model, an emphasis was placed on the desired faculty to represent *arbitrary* syllable sequences that fall within the rules of the speaker’s language, and to allow these sequence representations to interface with and select for production the most appropriate available sensorimotor programs. By planning in the phonological space encompassed by the IFS and pre-SMA representations, the GODIVA model does not rely on having acquired phonetic or motor knowledge for every syllable it is capable of planning and/or producing. This point is addressed in a simulation that parallels the one described above, but makes different

assumptions about the initial state of the model’s Speech Sound Map.

Figure 3·13 demonstrates the GODIVA model again producing the syllable sequence *gəv.di.və*, but in this case, the syllables *gə* and *və* have each been removed from the model’s Speech Sound Map. Since this version of the model no longer has sensorimotor representations for these syllables, it must produce the syllables from smaller stored programs / targets, corresponding to the individual phonemes in the “missing” syllables. It can be seen in Figure 3·13 (F) that the model activates SSM choice cells corresponding to the constituent phonemes, in the correct order, for the first and third syllables of the planned utterance. The SSM program associated with the second syllable, *di*, remains as a possible match in the Speech Sound Map, and, hence, is chosen for production at the appropriate time.

Panels C, D, and E in Figure 3·13 demonstrate how the model operates differently when it must produce syllables from smaller stored programs as compared to the case where all planned phonological syllables correspond exactly to stored SSM programs (Figure 3·12). By comparing Panel C to Panel D it is apparent that the IFS choice cell representing the first phoneme (/g/ of the syllable “*gəv*” is suppressed prior to the suppression of the phoneme /ə/. This is because the suppression of IFS choice cells is dictated by what sensorimotor program is chosen in the SSM choice layer. Because no SSM cell matches “*gəv*” exactly, the *best* matching cell (as determined by the dot product of IFS choice layer activity with each SSM plan cell’s stored synaptic weights; see Section 3.8.4) codes for the phonetic representation of the phoneme /g/. Thus, this cell is chosen for activation in the SSM choice field (see panel F), and inhibits only the representation for the phoneme /g/ in position zone 3 of the IFS choice layer (panel C). Because the phoneme /əv/ remains active in IFS choice field zone 4 after this point in time (panel D), the preparation of the *next* syllable can not yet begin. Instead, the activity in SSM plan cells (panel E) automatically adjusts

to activate better potentially matching sensorimotor programs corresponding to the *remaining* phonological representation in the IFS choice field (in this case the single phoneme /əʊ/). Once the non-specific quenching signal arrives at the SSM choice field to indicate the impending completion of the motor program for /g/, the motor program for /əʊ/ is chosen. At this point, the entire IFS choice field (in both zones 3 and 4; panels C and D) is empty, which allows the pre-SMA to choose the next syllable frame and continue the sequencing process.

Table 3.3: Summary of the values of parameters used to perform both simulations described in Section 3.9.

IFS		pre-SMA		SSM		BG Loop	
Parameter	Value	Parameter	Value	Parameter	Value	Parameter	Value
A_p	0.1	A_f	0.1	A_r	10.0	A_b	1.0
B_p	5.0	B_f	5.0	B_r	5.0	B_b	5.0
θ_p	0.01	θ_f	0.01	θ_r	0.1	$A_{\bar{b}}$	0.5
A_q	1.0	A_g	1.0	A_s	10.0	$B_{\bar{b}}$	10.0
B_q	5.0	B_g	5.0	B_s	5.0	A_c	1.0
						B_c	5.0
						β_c	1.0
						A_d	1.0
						B_d	5.0
						β_d	1.0

3.10 Discussion

3.10.1 The GODIVA model

This chapter has presented the development of a neurobiologically plausible computational model that describes how arbitrary syllable sequences can be planned and produced by adult speakers. This model builds on much previous theoretical work, beginning first and foremost with the seminal contributions of Lashley (1951). Lashley’s ideas can be viewed as a precursor to competitive queuing proposals (Grossberg, 1978b,a; Houghton, 1990; Houghton and Hartley, 1996; Bullock and Rhodes, 2003),

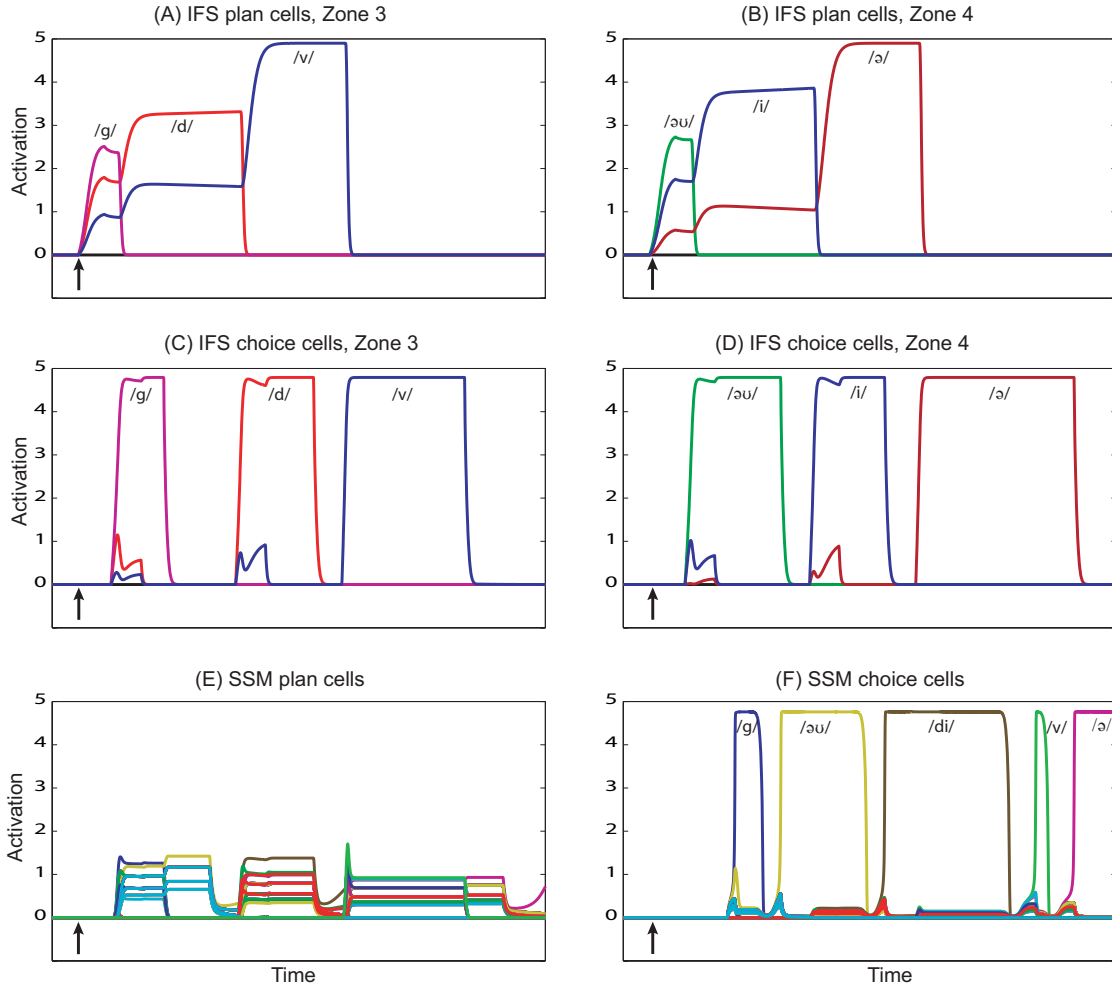


Figure 3-13: Simulation result showing the production of the syllable sequence “*gəv.di.və*” using piece-wise sensorimotor programs. In this simulation, only the second syllable (*di*) has a corresponding representation in the Speech Sound Map. The model must perform the first and third syllables, therefore, by sequentially activating targets for the constituent phonemes in those syllables. Each plot shows time courses of cell activity in different model components. The x-axis in each plot is time, and the y-axis is activation level (both in arbitrary model units). The arrows in each plot indicate the onset of the external input pulse, effectively the start of the simulation. These simulation results are described in detail in the text.

which the GODIVA model incorporates in multiple places. The use of a primacy gradient to represent serial order is a fundamental prediction of CQ-style models that has recently been confirmed in experimental studies (Averbeck et al., 2002, 2003). Such order-encoding activity gradients over representative units are also one of two modeling concepts that underlie the choice of the model’s name GODIVA (i.e. Gradient Order DIVA). The other equally important interpretation of this moniker is that the model provides ‘GO’ signals to the DIVA model. That is, the modules described in this chapter largely operate at a higher level in the speech production hierarchy than the existing DIVA model; these modules serve to select and activate the proper sensorimotor programs, and to initiate the production of speech sounds. Actual ongoing motor control of these speech sensorimotor programs, as well as their acquisition, is the function of the DIVA model itself which, although reviewed briefly in this chapter, has been described in detail elsewhere (Guenther, 1994, 1995; Guenther et al., 1998, 2006; Guenther, 2006).

That the GODIVA model was not developed in isolation, but rather as a continuation of a bottom-up approach (beginning with DIVA) to understanding the brain’s speech and language production circuits is an important characteristic. Although future work is necessary to fully integrate GODIVA with DIVA, the groundwork for perhaps the most comprehensive *computational* treatment of speech sound planning and production has been laid. Importantly, the model described here is not simply a computational or information processing treatment, but rather addresses the planning and production processes from a neurobiological perspective. To this end, each component of the GODIVA model, following previous efforts with the DIVA model (Guenther et al., 2006; Guenther, 2006), has a hypothesized cortical and/or subcortical correlate. The GODIVA model appears to represent the first thorough treatment of the sequential organization and production of speech sounds that is

described both formally and with substantive reference to known neuroanatomy and neurophysiology.

3.10.2 Representations for serial order

While the CQ architecture plays a fundamental role in GODIVA, it is not the only representation of order used within the model. The IFS representation combines elements of CQ with elements of positional models. Specifically, the minor axis of this two-dimensional map (see Figure 3-9) is proposed to code for abstract serial position within a syllable. The notion of using cells like those modeled in the IFS, which code for both a particular phoneme and a particular syllable position may at first seem unappealing; the use of multiple “copies” of nodes that code for the same phoneme but at different positions (e.g. Dell, 1986) has been often criticized for failing to encapsulate any relationship between phonemes with the same categorical identity that appear in different positions of a syllable. In the proposed IFS representation, to an extent, that relationship *is* encapsulated by the fact that these “copies” of the same phoneme will always appear topographically near one another so long as the 2-D grid is mapped continuously onto to the cortical sheet. Additionally, and perhaps more importantly, this position specific representation, which was motivated in this and other models on the basis of the very strong *syllable position constraint* in speech errors (see Section 3.4.1), is useful computationally. Because, in the model, IFS cells only interact with one another within a positional zone, the IFS representation can be thought of as one containing multiple queues. The capacity of a queue (i.e. a planning layer) in the CQ model is limited due to noise; as additional elements are “added” to the queue, the difference between activation levels of any two elements to be performed successively becomes smaller. If zero-mean Gaussian noise is added to these activation levels, the probability of recovering the wrong order at “read out”

then also becomes larger with additional elements. By separating the set of phonemes into multiple queues, of course, fewer phonemes are represented per queue, and noise is less of a problem. Effectively then, this representation increases the *overall capacity* to represent speech sounds during planning in comparison to a system with only one queue, assuming the same levels of noise.

The idea of representative units with serial position-specific tuning properties, while useful and supported by behavioral data in syllable production, is less appealing for modeling list memory, general movement planning, and other sequential behaviors because the number of “slots” is less well-determined, and the number of possible component movements or list items that must be able to be represented at any position could be quite large. Phonotactic constraints, while not specifically implemented in GODIVA, reduce the number of items that must be represented at any given position. GODIVA also includes “serial chain” representations, which are localized to the model’s pre-SMA module. The inclusion of these specific chains as a modeling element does not, however, invite all of the same criticisms that exist regarding associative chaining as a general theory of serial order. This is because the total number of sequences that must be encoded is small, and well established; these correspond one-to-one with the number of abstract structural syllable frames available to the speaker. As discussed in Section 3.8.2 just 8 syllable frames account for about 96% of all syllable productions. Moreover, the speaker has no need to *arbitrarily* order abstract syllable positions in a sequence. This leads to a general guiding principle that appears to be useful in modeling hierarchical sequential behavior. When sequence production must be *generative*⁹, the use of serial chains becomes extremely problematic, whereas the use of CQ-type activation gradients to

⁹Here, the term *generative* is used to mean that, in the behavior in question, the generation of novel, and perhaps arbitrary, sequences is crucial. In speech, for instance, combining words or syllables into a sequence that has never been performed before is simple and commonplace.

encode order is much preferred. When a sequence or a small set of sequences becomes highly stereotyped, however, readout by serial or “synfire chains” (e.g. Abeles, 1991; Pulvermüller, 1999, 2002), or by an “outstar avalanche” (Grossberg, 1969) may offer greater efficiency. The GODIVA model thus makes use of different representations of order as appropriate.

In a similar vein, Dell (1986) speculated that a principle explanation for the presence of speech errors in normal speakers is the need for productivity / generativity. In order to produce novel sequences within the language, it is necessary to “fill” slots in a sequence, and this inevitably results in the possibility of error due to potential difficulty with the “filling-in” mechanism(s). As put succinctly by Dell (1986), the set of possible phonemes is closed (after language acquisition), whereas the set of possible phoneme *combinations* is open. CQ provides an excellent and physiologically plausible mechanism for representing this open set of combinations. Furthermore, it makes both intuitive and computational sense that the units that “slip” during production of sequences should be the units that form the bases in CQ-type networks. This supports the GODIVA proposal that position-specific phonemes are represented by a CQ mechanism in the left inferior frontal sulcus region.

3.10.3 Repeating elements

One of the weaknesses for competitive queuing theories in general is in representing elements that repeat within a sequence. Because a cell codes for an item and that cell’s relative activation codes for its relative serial order, it is difficult to represent the relative order of the *same* item occurring twice in the planned sequence. The present model employs perhaps the simplest (but not necessarily best) solution to handle repeating elements. This is by including multiple “copies” of each representative cell in the IFS and pre-SMA representations. With the addition of such copies,

order can be maintained simply by using a different copy of the specific phoneme or frame cell for each occurrence of that phoneme or frame in the sequence. For example, the sequence “pa-ta-ka” would require the use of three different copies of the “/a/” phoneme cell in positional zone 4 of the IFS planning representation. In order to implement such a scheme, a bit of additional ad hoc machinery is required, which is implemented algorithmically in the model. Specifically, it is required that the model’s external input, when targeting a particular phoneme in the IFS plan layer or frame type in the pre-SMA plan layer, activate a cell of that type that is not already active. Response suppression within each representation is handled without additional circuitry, assuming the “copies” are arranged in the columnar CQ architecture. Response suppression of the IFS choice cells, however, arrives from the Speech Sound Map choice layer. To ensure that the correct copy of the phoneme cell is suppressed, the SSM choice layer→IFS choice layer quenching signal projects to *all* of the copies of a particular phoneme in a particular serial position.

When entire syllables (performance units), on the other hand, are to be repeated by the model (e.g. “ta-ta-ta”), a different assumption is made. On the basis of reaction time data from Schönle et al. (1986), as well as fMRI observations described in Chapter 2, it appears that producing the same syllable N times is fundamentally different from producing N different syllables. It is, therefore, assumed that planning a sequence such as “ta-ta-ta” only requires the phonological syllable “ta” to be represented in the complementary IFS and pre-SMA planning layers *once*. An additional simple mechanism is postulated to iterate the *production* portion of the circuit N times without the need to specify the phonological representation again each time.

3.10.4 A general framework

While the current modeling project does not deal with higher-level aspects of language production, the general architecture proposed here has the potential for reuse throughout the language system. The organization of basal ganglia into largely parallel loops (Alexander et al., 1986; Alexander and Crutcher, 1990) offers the possibility for cascaded processing stages that enable linguistic selections from competing alternatives; these selections (cf. choice layer activations) can then activate lower-level representations through cortico-cortical pathways (as IFS choice cells, for example, activate SSM plan cells). Such loops might be able to be nested to account for various levels of language production (e.g. Garrett, 1975; Ward, 1994). The model architecture presented here also offers a neurobiologically-plausible computational account for how learned structural patterns can be combined with an alphabet of “content” items (see theoretical development of this factorization of structure and content in Section 3.3.2). In the same way that abstract CV structure combines with representative phoneme units, syntactical structure might, for instance, combine with word units from different grammatical categories (cf. different positional zones). There is evidence that basal ganglia loops might indeed aid in selection mechanisms for higher level aspects of language. For instance, damage to portions of the caudate gives rise to *semantic paraphasias* (Kreisler et al., 2000), a condition marked by the wrongful selection of words, but such that the selected word has related meaning to the one desired. Crinion et al. (2006) also suggested that the caudate might subserve selection of words from a bilingual lexicon.

3.10.5 Future extensions to GODIVA

There are many limitations to the current version of the GODIVA model and many trajectories which future work can take. First, it will be important to establish re-

alistic mechanisms for *learning* the various representations and connections posited by the model. Currently, phonological representations in IFS and pre-SMA are assumed to have been learned, and connections between the IFS and SSM modules are hand-wired. Additionally, several brain regions whose BOLD activations were demonstrated to covary with the complexity of a planned syllable sequence are not included in the present model.

Brain regions not modeled

In particular these include the cerebellum¹⁰ and the anterior insula. It was hypothesized in Chapter 2 that the right inferior cerebellum could be used for “fast-loading” of well-learned phonological chunks into the IFS plan layer (cf. Rhodes et al., 2004). Many studies have now implicated the cerebellum in phonological coding (Paulesu et al., 1993; Desmond et al., 1997; Chen and Desmond, 2005; Silveri et al., 1998; Vallar et al., 1997; Justus et al., 2005; Ravizza et al., 2006).

The anterior insula is also of great interest in future work. Chapter 2 discussed a dissociation between two regions of the anterior insula, one which responded when speech was overt but that did not covary with the complexity of the prescribed syllable sequence, and another more anterior region (near the junction of the frontal operculum) where activation was similar during the preparation only and production conditions but covaried with stimulus complexity. It has been proposed that the anterior insula acts as a *phonetic buffer* during speech production (Nota and Honda, 2003). The GODIVA model currently does not contain a module capable of representing multiple *phonetic* plans simultaneously. Instead, the model’s Speech Sound Map choice layer can only code one winning cell, which codes for the best

¹⁰Portions of the superior cerebellum are addressed in modeling work related to single speech sound production in the DIVA model (Guenther and Ghosh, 2003; Ghosh, 2005), but further work is necessary to address this structure’s computational role in sequencing.

matching sensorimotor program for the currently selected *phonological* syllable. The SSM plan layer is able to maintain activity in multiple cells, but the gradient of activity in this region does not code for order, but rather for degree of phonological match. Thus, if such a phonetic buffer element is required as the model is further developed, a potential neural correlate seems to be the anterior insula. Data from the chronometric studies discussed in Section 3.4.2 appear to suggest that loading multiple sensorimotor programs into a phonetic buffer may indeed be possible. Much further consideration, however, is required before establishing specific roles for these additional brain regions in the model.

Speech error patterns

As currently formulated, the model has a limited capacity to recreate the rich patterns observed in naturally occurring slips of the tongue (see Section 3.4.1). The CQ architecture, however, is extremely well-suited for explaining data related to transposition errors (see, e.g. Farrell and Lewandowsky, 2004). The basic mechanical explanation for the three major error types observed in slips of the tongue are as follows: i) perseverations (e.g. “left hemisphere”) can occur when the IFS plan cell representation for a particular phoneme is not suppressed following its selection in the IFS choice layer; ii) anticipations (e.g. “left hemisphere”) can occur when the cell coding for the intruding phoneme (/h/) becomes more active than the proper phoneme, and its plan cell representation is not suppressed¹¹; and iii) exchanges (e.g. “left hemisphere”) can occur when the relative activation levels of the syllable onset phonemes /h/ and /l/ become reversed in the IFS plan layer.

Modeling such errors requires the addition of noise to the (currently determin-

¹¹Often errors are classified as anticipations when a speaker stops his or her utterance upon realizing their error (e.g. “left – (pause).” In these cases it is unclear whether, had the subject continued, the next word would have been produced as “hemisphere” or as “lemisphere,” which would be classified as an *exchange* error

istic) dynamics of the IFS phonological representation. With this simple addition, the model necessarily produces constrained errors. GODIVA accounts for the syllable position constraint because phoneme-coding cells only compete with each other within a positional zone, and the IFS choice selection process is zone specific. This is a hard constraint in the model; to replicate the actual data, which suggest that, on occasion, speech errors *do* occur across syllable positions the selection process could also be relaxed, allowing, for example, the wrong syllable position zone to be enabled with some low probability. As currently formulated the GODIVA model predicts a *syllable onset effect* but not of the magnitude reported in analyses of error databases (e.g MacKay, 1970; Vousden et al., 2000).

Another common observation in speech error data is that exchanged phonemes often share features (the *phonemic similarity effect*). On the surface, explaining such an effect when the current GODIVA model does not contain any explicit representation of features seems difficult. This might be addressed, however, by making use of physical space in the modeled cortex. In particular, each positional zone in the IFS planning layer could be organized as a *phonotopic* map, where cells that represent similar (in terms of shared articulatory features) phonemes are close together, and cells representing dissimilar phonemes are distant. Then, with the inclusion of a simple biologically reasonable assumption, that the magnitude of inhibition conveyed from cell p_{ij} to cell p_{kj} falls off with the distance between the cells, the effect will be that a particular phoneme has stronger competitive interactions with similar phonemes than with dissimilar phonemes. Since these competitive interactions ultimately determine the relative activation levels in the IFS plan cells, the net result of such organization would be that the most often exchanged phonemes would be the most similar phonemes. The notion of topographic organization is prevalent in neural computation, with many examples of models that learn 2-D mappings of stim-

ulus features (e.g. von der Malsburg, 1973; Grossberg, 1976; Kohonen, 1982), and Kohonen (1988) has previously applied the self-organizing feature map architecture to develop a 2-D map of phoneme space.

Such a topographically organized model, therefore, while not *explicitly* representing featural information, does represent featural similarity between planned segments *implicitly*. Other theories of speech production have proposed that phonological planning representations either i) specify no featural information (Levelt et al., 1999b; Roelofs, 1997; Dell, 1986; Shattuck-Hufnagel, 1987; MacKay, 1987), ii) fully specify all featural information (Wheeler and Touretzky, 1997), or iii) specify only non-default features (Levelt, 1989; Stemberger, 1991). The current proposal hypothesizes that featural information is not retrieved until the articulation stage, but that featural similarity can play a role, as described, at the planning level.

Communication disorders

Many researchers and clinicians have stressed the usefulness of comprehensive models in the study of communication disorders (e.g. Van der Merwe, 1997; Ziegler, 2002; McNeil et al., 2004). At present, however, models of speech production have largely been unable to shed light on disorders such as apraxia of speech (AOS) because "theories of AOS encounter a dilemma in that they begin where the most powerful models of movement control end and end where most cognitive neurolinguistic models begin" (Ziegler, 2002). The GODIVA model is the first step in an attempt to bring the DIVA model (the "model of movement control") into a broader neurolinguistic setting. In doing so, the hope is that communication disorders such as AOS and stuttering can be better understood in terms of pathological *mechanisms* within the model. For example, in GODIVA, the symptoms of apraxia of speech, particularly groping and difficulty reaching appropriate articulations, might be explained by at

least two mechanistic accounts. The first possibility is that the motor programs for desired sounds are themselves damaged. In the model, this amounts to damage to the Speech Sound Map (lateral premotor cortex / BA44) or its projections to the motor cortex. An alternative explanation could be that these sensorimotor plans are intact, but the mechanism for *selecting* the appropriate plan is defective. This would occur in the model with damage to connections between the IFS choice layer and the Speech Sound Map. A major focus of future research within this modeling framework should be the consideration of speech disorders.

Generating experimental predictions

As a closing note in the discussion of this model, it is important to realize that, ultimately, almost every model of a system as complex as that considered here, will eventually be found to have flaws. One of the most useful aspects of any model that can be simulated under various conditions is to generate experimental predictions. Through the generation of testable predictions the model may be proven invalid, but new proposals will arise from this knowledge that further our understanding of the system. The GODIVA model makes many such predictions. For example, GODIVA hypothesizes that the set of IFS choice layer to Speech Sound Map plan layer connections implements a selection process whereby the strength of input to an SSM plan cell depends on how strongly the speech sound that cell codes for matches the currently planned syllable in IFS. This proposal makes the corresponding prediction that when *many* cells in the SSM code for sounds that partially match the syllable being planned in IFS, the overall activation of the SSM will be larger than when there are few partial matches. More broadly speaking, planning and producing syllables with dense *phonological neighborhoods* is predicted to result in greater activation of the Speech Sound Map than planning and producing syllables with sparse

neighborhoods. This type of prediction seems to be readily testable using a cleverly designed fMRI or PET experiment. A continued program of model development combined with experimental neuroimaging is crucial to better understanding speech production.

CHAPTER 4

EXAMINING SYLLABLE SEQUENCE PRODUCTION USING MAGNETOENCEPHALOGRAPHY

This chapter describes preliminary efforts to examine aspects of syllable sequence preparation and production using magnetoencephalography (MEG). This sub-project was motivated by the fact that MEG can be used to measure neural signals that provide complementary information to those measured with fMRI. Experiments using magnetoencephalography that involve overt speech production, however, are technically challenging because activation of facial muscles can contaminate measurements. This chapter begins with a brief review of MEG and previous MEG studies using overt speech production. The methodological difficulties due to myogenic artifacts are explored, and these issues are addressed in the present work by recording surface electromyography (EMG) from facial muscles concurrently with MEG. This allows the time series to be parcellated into periods of interest demarcated by, for example, the stimulus onset, the GO signal, and the onset of muscle activity related to production. The temporal window between the GO signal and the onset of the EMG response is of particular interest herein. A novel algorithm is developed to recover neural sources whose estimated activity in a particular frequency range provide a means to discriminate between three syllable sequence production conditions. This algorithm is applied to data recorded during passive viewing of visual stimuli as a test-case and to data from a speech production task.

4.1 Introduction to magnetoencephalography

Magnetoencephalography (Cohen, 1972) is a non-invasive neurophysiological technique used to measure magnetic fields outside the skull caused by current flows in groups of neurons inside the brain¹. Such magnetic fields are extremely small in magnitude, typically on the order of 10^{-14} or 10^{-13} Tesla, many orders of magnitude smaller than the Earth's static magnetic field or the fields caused by typical urban electromagnetic noise. The measurement of such low-amplitude fields only became possible with the development of the highly sensitive Superconducting QUantum Interference Device (SQUID; Zimmerman et al., 1970) following the discovery of the Josephson Effect in superconducting materials (Josephson, 1962).

In contrast to BOLD fMRI (Ogawa et al., 1992; Kwong et al., 1992), MEG is a *direct* measure of neural activity in that the magnetic fields detected are instantaneously related to neural current flows through Maxwell's equations. Furthermore, temporal precision is not limited by the slow, delayed blood-flow response that gives rise to the BOLD signal but instead only by the sampling capabilities of the MEG instrumentation. Magnetoencephalography thus can provide neurophysiological measurements at a high temporal resolution, typically sampling the field patterns at ~ 1 ms intervals. The disadvantages that MEG has compared with fMRI in imaging brain activity are due to spatial resolution and certainty. These problems are exacerbated by the fact that magnetic fields can only be simultaneously measured from a limited number of sensor locations (typically hundreds in modern systems) positioned outside the head; this number is orders of magnitude smaller than the number of potential neural source locations in the cortex. The estimation of spatially localized current sources *within the brain* that give rise to an observed pattern of field

¹A full review of the theory behind magnetoencephalographic methods is beyond the scope of this chapter. Several excellent review articles, however, are highly recommended to the interested reader (Hämäläinen et al., 1993; Baillet et al., 2001; Vrba and Robinson, 2001)

measurements at the set of MEG sensors (or magnetometers) *outside the brain* is referred to as the *MEG Inverse Problem*. Helmholtz (1853) showed over 150 years ago that this kind of problem in the study of electromagnetism has no unique solution and is, hence, ill-posed (Hadamard, 1923).

Magnetoencephalography is a complementary method to its predecessor, electroencephalography (EEG). While MEG was first measured only about thirty-five years ago (Cohen, 1972) and systems for measuring whole-brain MEG have only been available since 1992, EEG has been measured for almost eighty years (Berger, 1929). The EEG method requires the attachment of surface electrodes to the subject's head in order to measure electric potentials at different locations on the scalp. These potentials are, like the magnetic fields measured by MEG, caused by currents flowing through neural cell assemblies. The measurements obtained by EEG and MEG are orthogonally related, and the two methodologies have unique *sensitivity distributions* or lead field properties (Malmivuo and Plonsey, 1995). MEG has been suggested to offer a higher practical spatial resolution compared to EEG; this is because the electrical potentials measured with EEG are strongly influenced by inhomogeneities in the tissues comprising the head, whereas the magnetic fields measured with MEG are largely unaffected (Hämäläinen et al., 1993).

Detectable MEG and EEG signals are believed to be generated by tens of thousands of cortical pyramidal neurons firing in synchrony (Okada, 1993; Murakami and Okada, 2006). The signal arises because these pyramidal cells within a patch of cortex have apical dendrites that are roughly oriented in parallel, and in the direction normal to the local cortical surface tangent. The co-activation of many spatially clustered pyramidal cells leads to a spatio-temporal superposition of activity that gives rise to a small, but detectable, magnetic field. Magnetoencephalography is relatively insensitive to sources oriented radially to the sensors such as, in some cases, those

on the crests of gyri. The method is, therefore, particularly useful for measuring activity within and surrounding cortical sulci.

4.1.1 Temporal components of MEG / EEG

Traditionally, MEG and EEG analyses have relied upon the time-locking of neural responses to particular stimuli or internal events across experimental trials. *Evoked responses* can be observed at characteristic delays relative to sensory stimulation. For example, visual evoked potentials (VEPs) or visually evoked fields (VEFs) can be recorded with EEG or MEG, respectively, by stimulating a subject's visual system with a high contrast image such as a black and white checkerboard pattern. Averaged over many presentations, the resulting measurements show a stereotyped response beginning approximately 80-100 ms after the onset of the visual stimulus, corresponding to the time of activation of the primary visual cortex (e.g. Ahlfors et al., 1992). In addition to responses that are time-locked to sensory stimulation, characteristic temporal responses have been shown to reliably arise due to higher-level processes such as expectation of a particular stimulus or decision-making. These findings have led to a large field of study of the responses related to particular events (e.g. event-related electric potentials or ERPs, or event-related magnetic fields, ERFs; Rugg and Coles, 1997; Hillyard and Kutas, 2002). Studies of such event-related responses tend to rely on averaging the responses over tens or hundreds of trials to improve the signal to noise ratio and thus reveal the response.

In studies of speech and language, one such typical response, the *M170*, appears bilaterally over the temporal-occipital region sensors approximately 150-200 milliseconds after the onset of a visually presented word and is associated with letter-string processing (Tarkiainen et al., 1999). Later components are also consistently found to be related to the process of word recognition. Such responses have been shown to

vary, for example, with sub-lexical frequency or lexical neighborhood (e.g. Pytkänen and Marantz, 2003). Studies of this type are of great interest in terms of language formulation and comprehension, but have to date provided little insight into the processes behind organizing and producing sequences of speech sounds.

4.1.2 Spectral components

The frequency characteristics of EEG / MEG recordings have also been the subject of abundant research. It is believed that the brain contains functional networks that operate within intrinsic frequency bands defined by neural circuitry and cell properties. For example, activity in the α -band ($\sim 8 - 12$ Hz) appears to have a source in the calcarine fissure that is suppressed when the eyes are open, but which increases oscillations when the eyes are closed (Hämäläinen et al., 1993). Spontaneous μ -band activity (~ 21 Hz) over the sensorimotor regions is damped by motor activity such as clenching the fist (Tiihonen et al., 1989). The dampening or strengthening of electromagnetic oscillations due to a stimulus or action is termed event-related synchronization or desynchronization (Pfurtscheller and Lopes da Silva, 1999). Recent research has additionally shown that increased activity in particular frequency bands might be used for encoding stimuli in working memory tasks (e.g. Jensen and Tesche, 2002; Jensen et al., 2002; Leiberg et al., 2006).

4.2 Previous speech production studies using MEG / EEG

Several previous studies have investigated evoked magnetoencephalographic signals during picture naming tasks (e.g. Salmelin et al., 1994; Levelt et al., 1998; Sörös et al., 2003). Levelt et al. (1998) conducted an overt picture naming study in order to explore the time courses of MEG signals hypothesized to correspond to the conceptual processes specified in the *Nijmegen Model* (Levelt, 1989; Levelt et al., 1999b). Using

a multiple dipole source analysis (see Section 4.2.2), they found that, generally, the activation of sources progressed from occipital areas for early visual processing to parietal and middle temporal areas for phonological code retrieval to left inferior frontal gyrus and left mid-superior temporal gyrus for phonological encoding. Model processes were delineated using time windows relative to stimulus onset determined on the basis of various previously measured chronometric data sets. The study did not include an examination of MEG signals following the activation of a *voice key* (indicating the start of the subject’s acoustic response), but the authors suggested that the immediately preceding interval (of approximately 150 ms) could be used to probe “phonetic and articulatory processing.” This, however, is problematic, because the onset of muscle activity can precede the onset of acoustic responses by up to hundreds of milliseconds, and these muscle activations can strongly contaminate the MEG recordings (see Section 4.2.1).

Despite these concerns, the temporal pattern of activations observed in the picture naming task described by Levelt et al. (1998) is generally consistent with hypotheses about functional roles for cortical regions activated in the fMRI study discussed in Chapter 2, and with the hypotheses of the GODIVA model (Chapter 3). These results, however, do not address how stimuli that are differentiated by some measure of planning complexity might elicit different time courses or different time or frequency response signatures in particular regions of the brain. This particular question, to date, appears not to have been addressed using magnetoencephalography.

Kuriki et al. (1999) measured simultaneous MEG and EMG from subjects producing a list of numbers. In this paradigm, subjects counted *covertly*, paced by a blinking LED (light-emitting diode), beginning at “one.” At a random time between the fourth and eighth number, another cue was given that informed subjects to produce the next number *overtly*. This procedure was suggested to aid in the time

alignment of brain processes relevant to production. A broad MEG response was observed, beginning approximately 100 ms prior to EMG perioral muscle activation, that was localized roughly to the left superior insular cortex.

In a series of studies, Riita Salmelin and colleagues have examined overt speech movements (although, in some cases, without phonation) using MEG (Salmelin et al., 2000; Salmelin and Sams, 2002; Saarinen et al., 2006). Salmelin et al. (2000) conducted a group study of single word production involving both fluent speakers and individuals who stutter. In normal speakers, a sequence of source activity was observed originating in left inferior frontal regions and advancing to the left lateral central sulcus and dorsal premotor cortex within approximately 400 ms of stimulus onset. In individuals who stutter, activity was localized in motor cortex prior to activation of the inferior frontal gyrus, suggesting possible abnormal motor preparation for speech. Speech-related suppression of a 20-Hz rhythm associated with motor cortical activity was observed bilaterally in both groups, but was right-hemisphere dominant in the stuttering group and left-lateralized in fluent speakers. This 20-Hz suppression phenomenon was again studied by Salmelin and Sams (2002) who compared silent speech production to non-verbal lip and tongue movements. This study revealed a left-lateralized post-suppression rebound (event-related synchronization) of 20-Hz activity in motor cortex following single word utterances, compared with a less focal and less lateralized rebound for non-verbal movements. Finally, several related results were reported by Saarinen et al. (2006). In particular, it was noted that 16-24-Hz suppression localized to the face area of motor cortex was tied to the onset of a visual stimulus and not to the onset of movement. Left hemisphere suppression preceded right, even for non-speech movements. Furthermore, the magnitude of 20-Hz suppression and rebound was related to the complexity of the movements performed, with greater modulation for sequences of non-speech gestures or for pseu-

dowords than for words, and with greater modulation for movement sequences than for single movements in isolation. Finally, as observed by Salmelin and Sams (2002), 20-Hz modulation was localized to a more focal area of the motor cortex for speech movements than for non-speech movements. Salmelin and Sams (2002) suggested that the 20-Hz event-related synchronizations and desynchronizations are only observed along the central sulcus, so this method cannot address higher-order speech or language processing areas. It is important to note that in each of these studies, the issue of possible face-muscle artifacts (see Section 4.2.1) was disregarded, with the suggestion that orofacial muscle signals operate outside the frequency range of interest (corresponding to the desynchronization frequency band; 16-24 Hz).

While still more experiments involving overt speech production with MEG have been reported in the literature, these tend to address research questions outside of the focus of the present project, including auditory cortical activations during self-produced speech (Gunji et al., 2000, 2001; Houde et al., 2002; Heinks-Maldonado et al., 2006) or the control of fundamental frequency in vowel-like utterances (Gunji et al., 2003). The present study sought to determine whether or not magnetoencephalography would be useful to reveal specific neural signatures related to planning simple non-lexical syllable sequences of differing complexity. While Saarinen et al. (2006) observed some changes in the modulation of μ -rhythms in motor cortex due to speech sequence complexity, the above question appears not to have been addressed systematically and across the entire speech production network using MEG or EEG.

4.2.1 Artifacts due to speech-related movements

Researchers have attributed electrical potentials and/or magnetic fields recorded extra-cranially during speech production to cortical activity since at least 1967 (Ertl and Schafer, 1967; Schafer, 1967). Only shortly after these initial publications, it

was realized that potentials generated outside the brain, particularly from the facial musculature, could easily have been mistaken as having cortical origin; this led to a published retraction of Ertl and Schafer's findings (Ertl and Schafer, 1969). McAdam and Whitaker (1971), several years after Ertl and Schafer, published results suggesting a slow, left-lateralized potential preceding speech at electrode sites near Broca's area and left premotor cortex. Shortly afterward, that study too, was criticized on the basis that the data could also be explained by artifacts caused by activation of the speech musculature (Morrell and Huntington, 1971; Grabow and Elliott, 1974). Szirtes and Vaughan (1977) published a summary of simultaneously acquired electrical recordings taken from cranial (using EEG) and facial (using surface EMG) locations prior to and during overt speech. Their analysis of recordings from their own laboratory led these authors to suggest that:

“the results reported here lend strong support to suggestions that scalp recorded speech-related potentials either represent activity of solely extracranial origin or are heavily contaminated by such activity (Szirtes and Vaughan, 1977, p. 391).”

This conclusion was based on three key findings: i) electrical potentials recorded over frontal locations showed substantial morphological changes with changes in the utterance; ii) maxima in potential distributions overlaid the lower face region with observed polarity inversions near the mouth region; and iii) observed “speech-related” potentials showed similar form and distribution to those observed during non-speech mouth movements.

Despite evidence that EEG recordings during speech articulation (including a short period prior to vocalization that involves muscle activation) can be highly contaminated due to muscle activations, some researchers have continued to publish findings obtained from recordings during this problematic time period. Salmelin and

Sams (2002) and Saarinen et al. (2006) provide results of MEG studies in which data were measured concurrent with movements of the lips and tongue. The issue of contamination of the MEG from muscle activation is neglected, relying on an assumption that if the data are high-pass filtered (above ~ 16 Hz in these studies), then “mouth movement artefacts are negligible (Salmelin and Sams, 2002, p. 83).” Despite this claim, much data suggests that EMG signals due to facial muscle activity have broad frequency distributions, with substantial power at frequencies well above such cutoff frequencies (e.g. van Boxtel, 2001; Goncharova et al., 2003). A recent study, for example, demonstrated that activity due to tongue muscle activation contributed to MEG recordings in the 25-70 Hz band (Furlong et al., 2004).

In the preliminary study presented here, the problem of facial muscle activity contaminating the MEG signal was considered carefully. It was determined that the best approach was to measure simultaneous EMG from relevant face muscles, and to discard the portion of the time series following initial muscle activation. This decision followed unsuccessful efforts to apply blind source separation with independent component analysis to separate muscle activity from cortical activity (Vigário et al., 1998; Hyvarinen and Oja, 2000). The difficulty with such techniques was in identifying which of N separated components were related to muscle and which were related to cortical activity; without a *model* of muscle artifacts, such decomposition techniques were extremely subjective. By analyzing only the period of time up to EMG onset in each trial, questions regarding ongoing articulatory control could not be addressed, but questions related to speech planning were accessible within the context of a typical speech production task (as opposed to, for example, covert speech tasks).

4.2.2 Source estimation methods

Many methods for estimating localized components (in time or frequency) observed in MEG sensor measurements have been proposed. It is beyond the scope of this chapter to describe these in detail, but the basics are discussed briefly here.

The *MEG forward problem*, which solves for the magnetic field pattern produced at the magnetometers for a known source distribution, is straight-forward, but can be implemented with varying degrees of complexity and accuracy (Mosher and Leahy, 1999). In practice the solution to the forward model yields a *lead field matrix* that represents a linear transformation that maps a vector in source space into a vector in sensor space.

All *inverse methods* make use of such a forward model. The simplest methods, *equivalent current dipole* methods, assume either one or a small number of focal dipolar sources are assumed to be responsible for the *entire* observed field pattern. These methods attempt to find the location and orientation of the source(s) that minimize (typically in a least-squares sense) the discrepancy between the observed sensor data and the sensor data predicted from the forward model. *Imaging* methods provide magnitude estimates for *many* fixed dipoles distributed densely in the source space. Such methods typically require a more precise head model than do equivalent current dipole procedures in order to specify the dipole locations, which are usually placed along a reconstructed cortical surface grid created from an MRI scan of the subject's brain (e.g. Dale and Sereno, 1993). Due to the ill-posedness of the MEG inverse problem, these methods require the addition of certain *a priori* assumptions to arrive at a unique solution. Most commonly used algorithms seek to minimize the L2- or L1-norm of the resulting current estimate (minimum-norm or minimum-current estimates).

Spatial filtering or *beamforming* methods effectively circumvent the biomagnetic

inverse problem by applying data-driven spatial filters to the sensor measurements. These filters, derived from the lead field matrix, are designed to pass signals from a location of interest in source space, and to block signals from other locations. Various formulations for designing such spatial filters have been proposed (van Veen et al., 1997; Robinson and Vrba, 1999; Sekihara et al., 2001, 2002). Estimation of a cortical source, then, is simply a matter of applying the appropriate linear filter to the observed sensor data. In the method developed in this chapter, a very simple spatial filtering approach is utilized.

4.3 Materials and methods

4.3.1 Subject

One right-handed adult American English speaker (male, 25 years old) with no history of neurological, speech, language, or hearing problems participated in this study. Several related pilot sessions involving additional subjects were also conducted but are not reported here.

4.3.2 Experimental protocol

The subject participated in two experiments: one “baseline” experiment involving simple visual stimulation and one speech production experiment involving overt production of sequences of nonsense syllables. In both experiments, task-relevant stimuli as well as digital “triggers” sent to the MEG recording channels (see below) were delivered using the DMDX Version 3 software package (Forster and Forster, 2003). Stimuli were presented visually on an approximately 18 cm \times 18 cm projection screen located approximately 24 cm from the subject’s head.

4.3.3 Visual baseline stimuli

Two experimental runs were performed using a visual evoked field paradigm. High-contrast checkerboard patterns were rapidly presented on the projection screen, each for a duration of 1000 ms followed by a blank screen for 500 ms, resulting in an overall inter-stimulus interval (ISI) of approximately 1.5 seconds. Stimuli consisted of full disc-shaped checkerboard patterns and hemi-disc patterns presented on only the left or right portion of the visual display. The background pattern was a solid medium-intensity gray. Each run consisted of the presentation of 100 total stimuli. Due to a technical problem involving trigger channels, however, only 40 right-field and 80 full-field checkerboard trials (see Figure 4-1) could be used in the analysis.

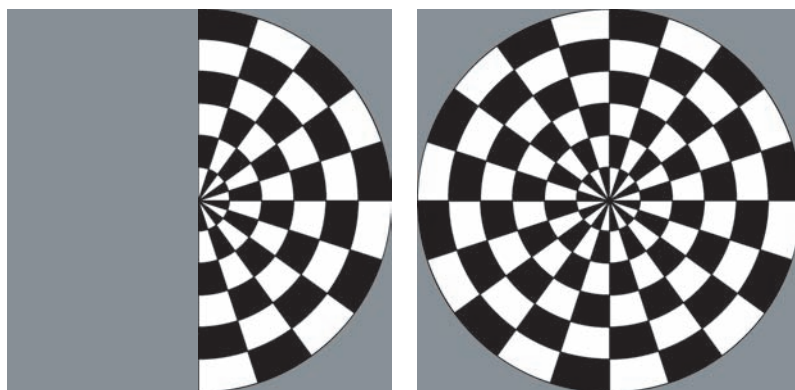


Figure 4-1: Right-field (left figure) and full-field (right figure) checkerboard stimuli used for visual baseline trials. The stimuli were presented randomly, for a duration of approximately 1.0 s, with a 1.5 s inter-stimulus interval.

4.3.4 Syllable sequence stimuli

Stimuli used during the speech production task consisted of one-, two-, or three-syllable sequences presented orthographically on the projection screen. All syllables were of the simple CV (consonant-vowel) syllable structure; vowels were selected

pseudo-randomly from $\{/a/,/i/,/u/,/ə/\}$ with the condition that no vowel was repeated in a single stimulus (that is, in a three-syllable stimulus, the three vowels were unique). Consonants for the *initial* syllable in a sequence were labials chosen from $\{/b/,/m/,/v/\}$; these labial consonants were chosen following pilot sessions, which showed that EMG could be used to provide reliable estimates of movement onset for these sounds (see EMG details below). The consonant phonemes in non-initial syllables were chosen from $\{/d/,/g/,/k/,/t/\}$. Again, no consonant appeared twice within a single stimulus. Trials were similarly formatted to those utilized in the corresponding fMRI experiment (Chapter 2; Bohland and Guenther, 2006), although all trials herein were GO (overt production) trials and the inter-stimulus interval was much reduced². Specifically, a single trial began with the visual presentation of the stimulus, chosen randomly from the three conditions. The stimulus was projected for 2.5 seconds and then replaced by a white cross. The subject was instructed to fixate on the white cross without blinking. After a random interval (uniformly chosen from 0.5 to 2.0 seconds), the white cross changed color slightly, to a light gray³. This instructed the subject to immediately vocalize the most recently presented stimulus. The subject was instructed to refrain from facial or head movements throughout the experiment, and to avoid eye blinks in the interval between stimulus onset and the completion of production of each stimulus. The subject was given an opportunity to blink and/or swallow following production and was instructed to return to a neutral mouth position with the jaw closed but not clenched before the start of the next trial.

²In fMRI trials, it was necessary to have a long inter-stimulus interval in order to capture the delayed hemodynamic response to that event and to allow the response to decay before the presentation of the next stimulus. In MEG there is no delay associated with measurements; rather, the magnetic fields measured reflect simultaneously occurring neural activity.

³The use of a more subtle visual change to indicate the GO signal in the MEG trials as compared to fMRI trials was designed to reduce the visual onset response in MEG which was not of interest in the study and threatened to overwhelm the MEG signals recorded following GO.

During each trial, two “trigger” signals were sent to two different MEG recording channels, one synchronous with the onset of the visual stimulus (the syllable sequence), and the other synchronous with the appearance of the GO signal. These were used off-line for segmenting the trial data.

4.3.5 Data acquisition

MEG measurements were acquired using a 160-channel whole-head axial gradiometer (with 50 mm baseline) system (Kanazawa Institute of Technology, Japan) located at the Massachusetts Institute of Technology. The system is a recumbent setup with a fixed dewar (the helmet that contains the SQUIDS) and sits within a magnetically-shielded room (MSR; Vacuumschmelze, Hanau, Germany) with active magnetic noise cancellation. Three measurement channels were used as reference sensors for additional offline noise reduction (see Section 4.3.7).

All measurements (including MEG, reference, trigger, and EMG channels) were sampled at 1 kHz and filtered online with a low-pass filter with cutoff at 200 Hz, and a band-stop filter with notch at 60 Hz (to eliminate electrical noise). The analog signals were digitized using a 12-bit analog to digital converter (Eagle Technology, Cape Town, South Africa). Five “marker coils” were affixed to the subject’s head. Between experimental runs, a small prescribed current is passed through the marker coils to be used to localize their positions (and hence the subject’s head position) relative to the locations of the sensors.

The subject’s head shape was digitized using the Polhemus Fastrak Digital Tracker (Polhemus, Colchester, VT) with 3 receivers in conjunction with Locator software (Source Signal Imaging Inc., San Diego, CA). Approximately 1000 locations on the head surface were sampled using a hand-held stylus. The positions of fiducial points, marker coil locations, and electrodes used for surface EMG were also

recorded.

The subject's vocal responses during the experiment were recorded using an Audio-Technica (Tokyo, Japan) ATM10a omni-directional condenser microphone placed in a shielded aperture in the MSR wall. The recorded audio and MEG measurements were synchronized by the simultaneous delivery of trigger pulses to specified MEG recording channels and to one input of a multi-channel stereo mixing device (Behringer Eurorack MX602A; Behringer International, Willich, Germany) used in the audio recording setup. The merged audio and trigger signals were digitized and recorded on a notebook computer (Dell, Inc., Round Rock, TX).

In the speech production experiment, 9 mm tin cup surface electrodes (Electro-Cap International, Inc., Eaton, OH) were used to measure electrooculogram (EOG) and electromyographic (EMG) signals from the face musculature. Speech-related EMG Signals were recorded from the orbicularis oris and temporalis muscles on the left. Signals were amplified using an electrically isolated 24-channel bioelectric amplifier (SA Instrumentation Co., Encinitas, CA) and sampled and recorded simultaneously with the MEG measurements.

4.3.6 EMG signal analysis

EMG signals were filtered using a 2nd order Butterworth bandstop filter with notch at 60 Hz then a 5th order bandpass Butterworth filter with low frequency cutoff at 20 Hz and high frequency cutoff at 400 Hz. They were then full-wave rectified and smoothed using a median filter with sliding window of length 7 ms (7 samples). Finally, signals were integrated over a moving 40 ms window.

The mean EMG signal from a 150 ms baseline period was extracted in each trial. An onset was detected when the mean value of the processed EMG signal across a sliding 30ms window exceeded 3 times this baseline mean. Similar procedures have

been applied elsewhere (see, e.g. Hodges and Bui, 1996). Trials with reaction times of less than 100 ms were considered outliers and were removed from the analysis.

The EMG recordings from the *orbicularis oris* (lower lip) muscle were found to be the most reliable indicators of movement for the particular labial consonants being produced, and thus were used as the EMG channel of interest for determining muscle activation onsets.

4.3.7 MEG signal preprocessing

A noise reduction algorithm, the Continuously Adjusted Least-Squares Method (CALM; Adachi et al., 2001) was applied to all MEG data. This method removes low frequency noise by eliminating correlations between three orthogonal reference channels located away from the subject’s brain and the data channels.

Following noise reduction, time series for each MEG sensor were extracted from each trial; these series were defined as the raw signals between the GO signal and the estimated EMG. Because the subject’s reaction time varied from trial to trial, the length (in time) of these extracted series also varied. Data analysis was thus performed in the frequency domain.

The series from each sensor and each trial were multiplied by an L -point Hanning window, where L is the length of the extracted time series (in millisecond samples) for that trial. The windowed time series were then transformed into the frequency domain using the Discrete Fourier Transform (DFT). Components above 200 Hz (the low-pass cutoff frequency for the data acquisition filter) were discarded. Only the magnitudes of the resulting frequency components were used in the analysis.

4.3.8 Head model

A high-resolution structural MR scan (T1-weighted, 128 sagittal images, 256×256 matrix, 1 mm^2 in-plane resolution, 1.33 mm slice thickness, TR=2530 ms, TE=3.3 ms, flip angle 9°) was acquired using a Siemens 3 Tesla Trio scanner. Freesurfer (Dale et al., 1999; Fischl et al., 1999) was used to extract the outer skull surface and both pial and white matter cortical surfaces. A new cortical surface was constructed corresponding to the midpoint between the two cortical surfaces⁴. This surface was used to generate the biomagnetic forward model.

Three distinct coordinate frames must be coregistered prior to the construction of the forward model. These coordinate frames are given by:

1. The subject's MRI scan / cortical reconstruction,
2. The subject's digitized head shape,
3. The locations of the magnetometers in the dewar.

The latter two frames were realigned by solving for the optimal (in a least-squares sense) parameters of a rigid-body transformation that brings the positions of the marker coil locations in the digitizer coordinate frame into alignment with the estimated locations of the markers in the sensor coordinate frame. The MRI coordinate system is then brought into alignment with the sensor coordinate frame by the use of an interactive surface-matching tool that attempts to minimize the disparity between the surface defined by the subject's digitized head shape and the subject's skull surface extracted from MRI (e.g. Kozinska et al., 2001). The cortical surface is then "brought along" by applying the affine transformations. Figure 4.2 shows a sample result of the realignment process.

⁴Previous experience in our laboratory has shown that this gray-white "midpoint" surface tends to provide a better forward model than either the gray or white matter surface.

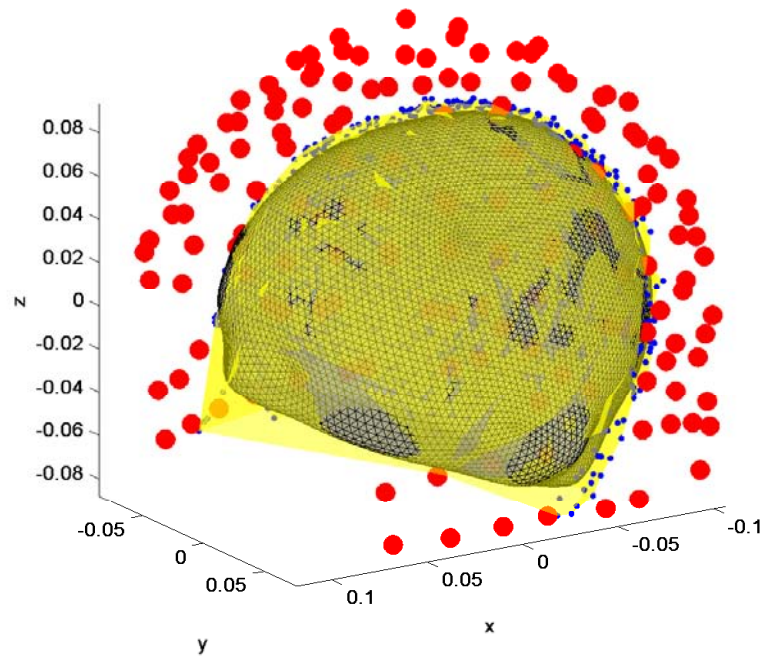


Figure 4.2: Example of surface-based alignment of coordinate frames from structural MRI space, head digitizer space, and MEG system space. Blue dots indicate positions at which the head shape was digitized; the translucent yellow surface fits these points. The gray mesh surface is a smoothed head shape extracted from structural MR for this subject. The large red dots indicate the positions of the MEG sensors relative to the subject's head.

The MEG forward model, which calculates the expected magnetic field at each sensor caused by a current source at a particular location and orientation, was calculated using functions from the BrainStorm software package (<http://neuroimage.usc.edu/brainstorm/>). Specifically, a sensor-weighted overlapping spheres approach (see Huang et al., 1999) was used to calculate the lead field matrix \mathbf{A} . \mathbf{A} has dimensions 157 (# of sensors) \times 1504 (# of sources). The 1504 sources were located at vertices sampled across a decimated version of the reconstructed cortical surface described above. Each column of \mathbf{A} , therefore, maps the activity of a putative cortical source into an expected field pattern across the 157 SQUID magnetometers.

4.4 A novel method for single-trial MEG analysis

A new algorithm was developed that utilizes single-trial MEG data (as opposed to averaged data) with the objective of finding cortical responses that reliably differ across experimental conditions. Specifically, the algorithm finds cortical sources whose estimated strengths at a particular time or frequency provide a means to discriminate between the designed experimental conditions, at a significantly greater than chance level. The algorithm can be applied in either the time or frequency domain, and simply requires different pre-processing of the sensor data. The details of frequency-domain preprocessing are given in Section 4.3.7. The procedures described were implemented in MATLAB®.

The following specification of the algorithm assumes that there exist M magnetometers (sensors), N total experimental trials drawn from K different experimental conditions, and P potential source locations⁵.

⁵Source locations are equivalent to vertices sampled in a reconstructed cortical surface for a subject. These vertices are the same locations used to calculate the biomagnetic forward model (described above).

The algorithm computes a *fitness value* associated with each potential source location for each particular time or frequency interval. The fitness value indicates how strongly the estimated contribution from a cortical source is associated with the experimental condition labels assigned to each trial. The strength of a cortical source is determined by the projection of the observed sensor data for a particular time or frequency range onto the lead field vector for that source. In essence, a high fitness value will be assigned to sources whose estimated strengths cluster according to condition labels; that is, if the estimated strengths are similar for trials of the same condition, and different from trials of a different condition.

Specifically, the contribution of the i^{th} source location to the observed sensor measurements within a particular time or frequency range is estimated for every trial. The $1 \times N$ vector of these values (for time or frequency range τ) is given by:

$$\mathbf{x}_i^\tau = \frac{\mathbf{a}_i'}{\|\mathbf{a}_i\|} \mathbf{Y} \quad (4.1)$$

where \mathbf{a}_i is the i^{th} column of the lead field matrix A , and \mathbf{Y} is an $M \times N$ matrix of time or frequency windowed measurements at each sensor for each trial, where:

$$Y_{mn}^\tau = \sum_h w_h s_{\tau+h}^{mn}$$

Here, h indexes the elements of the vector \mathbf{s}^{mn} corresponding to the (time or frequency) measurements at sensor m during trial n .

The *fitness* of source i over range τ is determined by how well the values in \mathbf{x}_i^τ are associated with the labeling of experimental conditions across trials, also a $1 \times N$ vector \mathbf{c} . To test this correspondence, \mathbf{x}_i^τ is sorted by value to obtain the vector $\hat{\mathbf{x}}_i^\tau$. The condition labels \mathbf{c} are sorted with the same indices to obtain $\hat{\mathbf{c}}$.

Then, an empirical cumulative distribution function (cdf; \mathbf{z}) is calculated for this

source and time or frequency range for each experimental condition. This function is a sum, accumulated across the sorted condition vector $\hat{\mathbf{c}}$, such that for condition k :

$$z_{n+1}^k = \begin{cases} z_n^k, & \text{if } \hat{c}_n \neq k \\ z_n^k + 1, & \text{if } \hat{c}_n = k \end{cases}$$

and $z_0^k = 0$ for all k . Each of these cumulative distribution vectors is then normalized to account for differences in the number of trials per condition.

$$\mathbf{z}^k = \frac{\mathbf{z}^k}{\max(\mathbf{z}^k)}$$

Finally, a scalar fitness value for this i^{th} source over range τ , f_i^τ , is calculated from the set of \mathbf{z}^k vectors. This value indicates the degree to which the cumulative distribution functions are separated from one another. Specifically, f_i^τ is the maximum Euclidean distance of a point in \mathbb{R}^K defined by $(z_n^1, z_n^2, \dots, z_n^K)$ from a point defined by the cdf mean value in all dimensions:

$$f_i^\tau = \max_n \left(\sqrt{\sum_{k=1}^K (z_n^k - \bar{z}_n)^2} \right)$$

This method is similar to procedures used in the Kolmogorov-Smirnov (K-S) Test, which can be used to determine if two samples were drawn from the same distribution. Whereas the empirical cumulative distribution functions used in the K-S test are defined over the actual values in the sample, the analogous cdf's used here are defined over indices.

The entire process described above is iterated over all P potential source locations and time or frequency ranges of interest, resulting in P fitness values for every time or frequency range. The larger the fitness value for a particular source component,

the better that source component is able to discriminate the experimental conditions. Figure 4.3 shows a schematic representation of the algorithm for determining a single fitness value.

4.4.1 Statistical tests

While the fitness values f_i^τ give a relative idea of how well a source component is able to provide a measure of discrimination between experimental conditions, these values have little meaning without a statistical framework. The algorithm thus uses permutation tests to obtain P -values for the source components with high fitness.

In order to obtain the required permutation data, the entire algorithm described above is repeated many times, but with random permutations of the vector \mathbf{c} indicating the condition labels for each trial. For N trials, consisting of the same number of trials, $J = N/K$, for each of the K conditions, there are $N!/(J!)^K$ possible unique relabelings. This number will grow extremely rapidly with the number of trials if K is small; assessing all relabelings is thus computationally infeasible, so a random subset is chosen for evaluation. This results in the calculation of a fitness value matrix \mathbf{F} for each chosen permutation. These matrices ultimately provide a sampling distribution of the fitness values that arise under the null hypothesis (i.e. by chance). The fitness values obtained from the *correct* condition labeling are then compared with the distribution of *maximum* fitness values obtained for the random labelings. A corresponding P -value for source i at time or frequency range τ is then calculated simply as:

$$P_i^\tau = \frac{q}{Q}$$

where Q is the total number of random permutations performed, and q is the number of those permutations for which the maximum fitness value found for time or frequency range τ is greater than the f_i^τ found for the correct labeling.

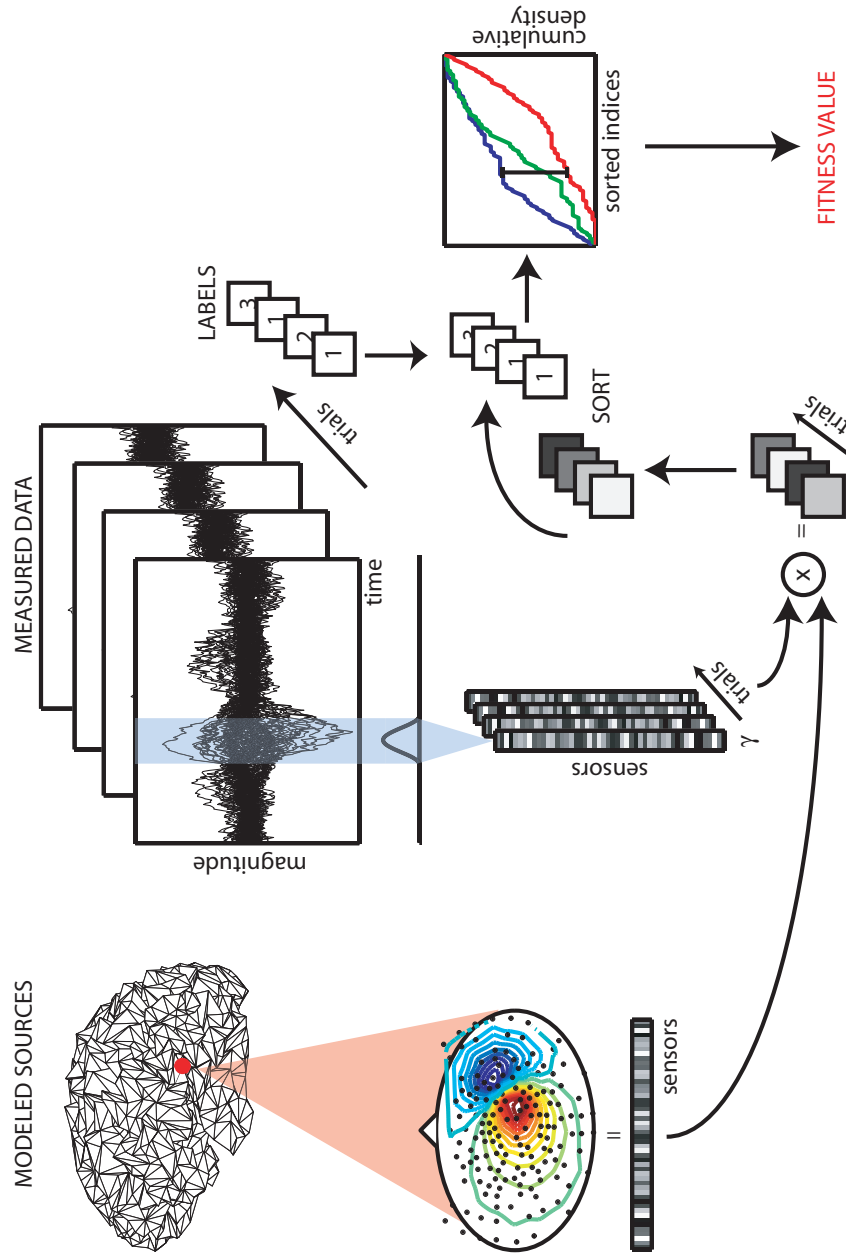


Figure 4-3: Schematic depiction of the process of determining the *fitness value* for a particular source location over a particular time or frequency window. A spatial filter determined for each source (bottom left) is applied to the windowed data from each trial, resulting in N estimated source strengths for N trials. These values are sorted, and the sort indices applied to the set of condition labels (top right). From the sorted labels, a cdf is computed for each label, and the fitness value is computed as a measure of maximum distance between the cdf's.

4.5 Results

4.5.1 Visual evoked fields

MEG sensor recordings from 120 trials (40 trials using right-field checkerboard, condition label 1; 80 trials using full-field checkerboard stimuli, condition label 2) were analyzed using the approach developed in Section 4.4. The goal of this application was to find cortical sources that reliably responded differently over a small time window to the right-field versus full-field conditions. Single trial time series were extracted from the onset of the visual stimulus until 250 ms post-onset. Data were binned by multiplying the time series over a 10 ms window by a 10-point Hanning window. The window was then slid across the time series using a 5 ms step size. In order to calculate statistics related to the computed fitness values, 250 additional iterations of the method were performed using random permutations of the per-trial condition labels. Figure 4.4(a) shows the results of this algorithm. Neural sources whose estimated strengths discriminated between the two conditions were found to be largely clustered (in time) around a central peak in discriminability in the temporal window centered around 115 ms following stimulus onset. The earliest component was found during the time window centered at 85 ms post-onset, and this source was located in the primary visual cortex. The largest (significant) fitness values were found in the 115 ms post-onset window. Figure 4.4(b) shows a rendering of the locations of the significant components found during this time window. Notably, nearly all of the sources found by the algorithm were localized in the primary and secondary visual cortices, and those that were outside of this region tended to show strongest discriminability between conditions later in each trial. Figure 4.4(c) shows all significant discriminatory sources, and the time at which each source achieved highest fitness.

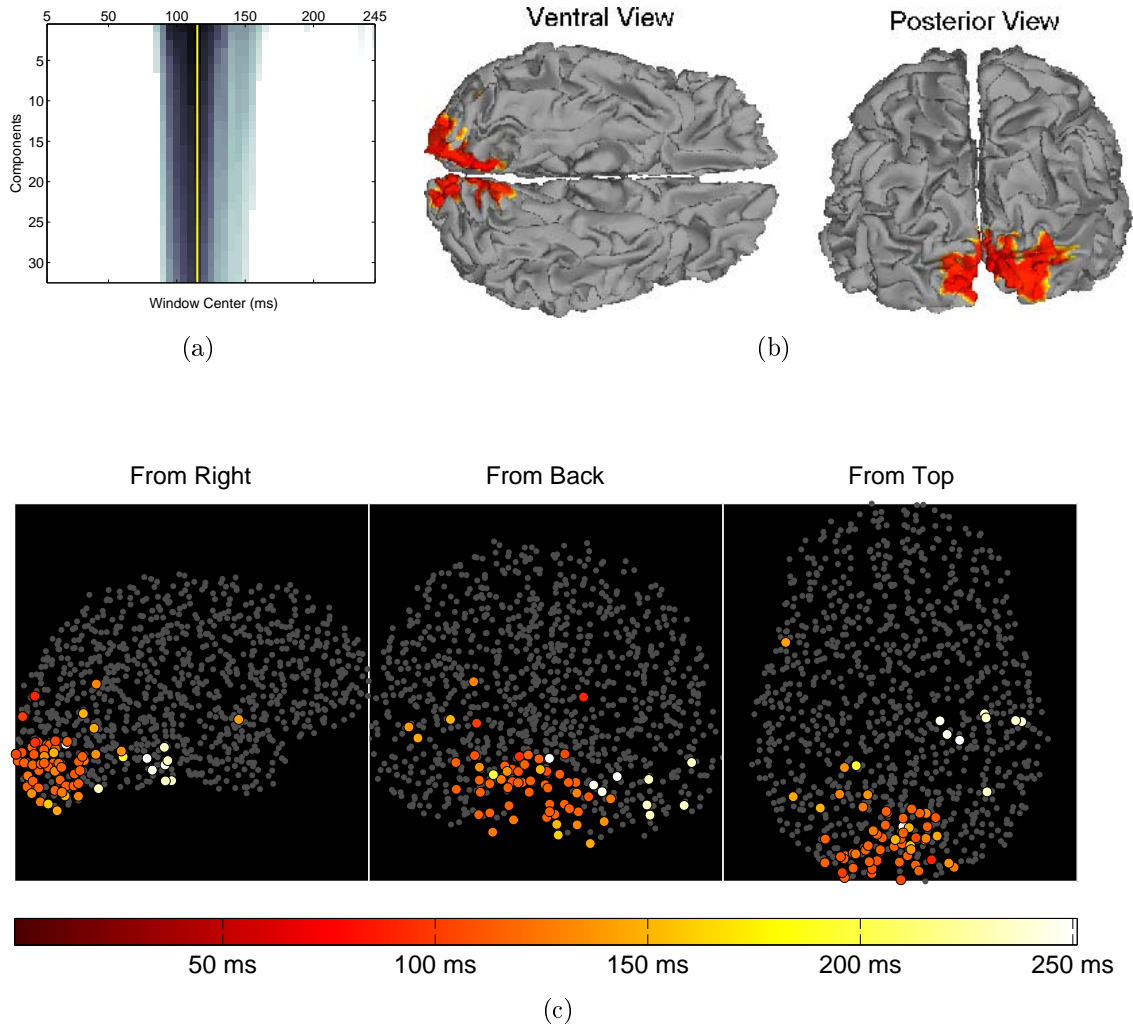


Figure 4.4: (a) Neural source components (top 32 components \times time windows) recovered by the algorithm. Only significant components ($P < 0.01$) are shown; that is, the color white at position (x,y) indicates the lack of a significant y^{th} component at time window x. Color of significant entries indicates the *fitness value*. A yellow line is drawn through the time window centered at 115 ms after stimulus onset, the time of the peak fitness value. (b) Renderings of all 32 significant source locations found at the peak time window. The results are rendered on a high-resolution cortical surface for the subject. (c) “Glass brain” plot showing the locations of all potential sources (gray) projected into each Cartesian plane. Sources that discriminated conditions significantly above chance are shown in color and with larger circles. The colors of the circles represent the time at which that source’s fitness value peaked (see color bar).

4.5.2 Syllable sequence production

EMG onsets were estimated using the procedures described above. Subtracting the time of occurrence of the GO signal from the estimated time of muscle activation onset yielded a *reaction time* for each trial. The means and standard deviations of estimated reaction times for each of the three speaking conditions are plotted together in Figure 4-5. A one-tailed T-test revealed that reaction times for both 2 and 3 syllable sequences were significantly longer than for 1 syllable utterances ($P < 0.001$); reaction times for 3 syllable and 2 syllable sequences were not significantly different.

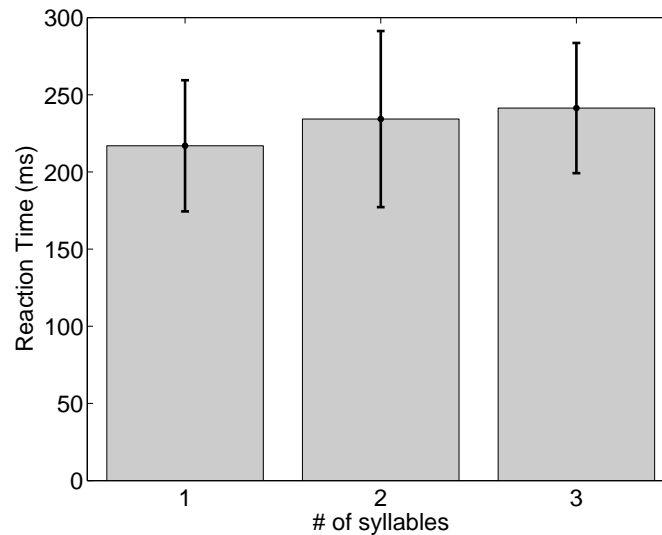


Figure 4-5: Reaction time per condition estimated from EMG. Across trial means (bar height) and standard deviations (error bars) are displayed.

Single trial data from the time period of interest (which varied in number of samples from trial to trial) were transformed into the frequency domain. Figure 4-6 shows the mean (across trials) magnitude spectra for each measurement channel. The *single trial* spectra served as inputs to the MEG analysis algorithm described

in Section 4.4. Spectra from 440 trials, with frequency bins of approximately 2 Hz in width were used in the analysis. Figure 4-7(a) shows a frequency \times rank plot of the components that were determined to significantly differentiate the three speaking conditions.

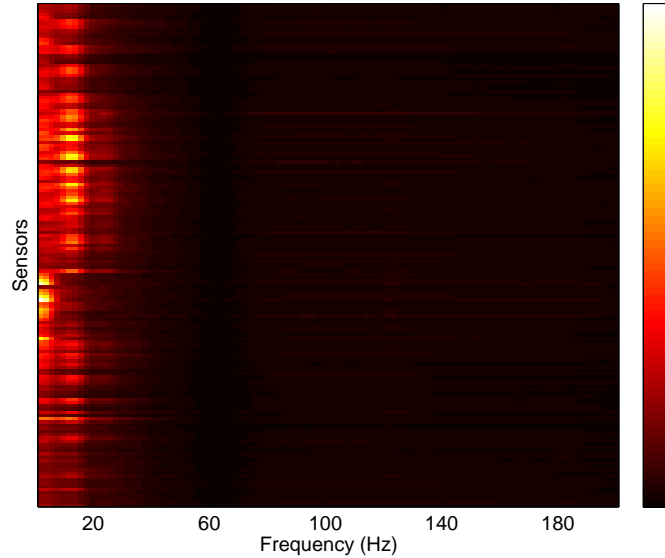


Figure 4-6: Mean frequency-domain response across all trials for the time period between the GO signal and the onset of muscle activation as estimated from lip EMG. Each row in the image corresponds to a measurement channel. Brighter colors indicate a higher magnitude response. It can be seen that the majority of the energy in the signal is at low frequencies, below ~ 40 Hz.

A band of significant components between ~ 10 and ~ 14 Hz were of particular interest because this frequency range strongly overlapped with high spectral density in the single trial spectra (see Figure 4-6). The locations of each of the components in this frequency range (indicated by a green outline in Figure 4-7(a)) are rendered on the subject's reconstructed cortical white matter surface in Figure 4-7(b). The majority of these significant components were found to be in the left lateral prefrontal cortex, including the area around the left inferior frontal sulcus (IFS). Because of

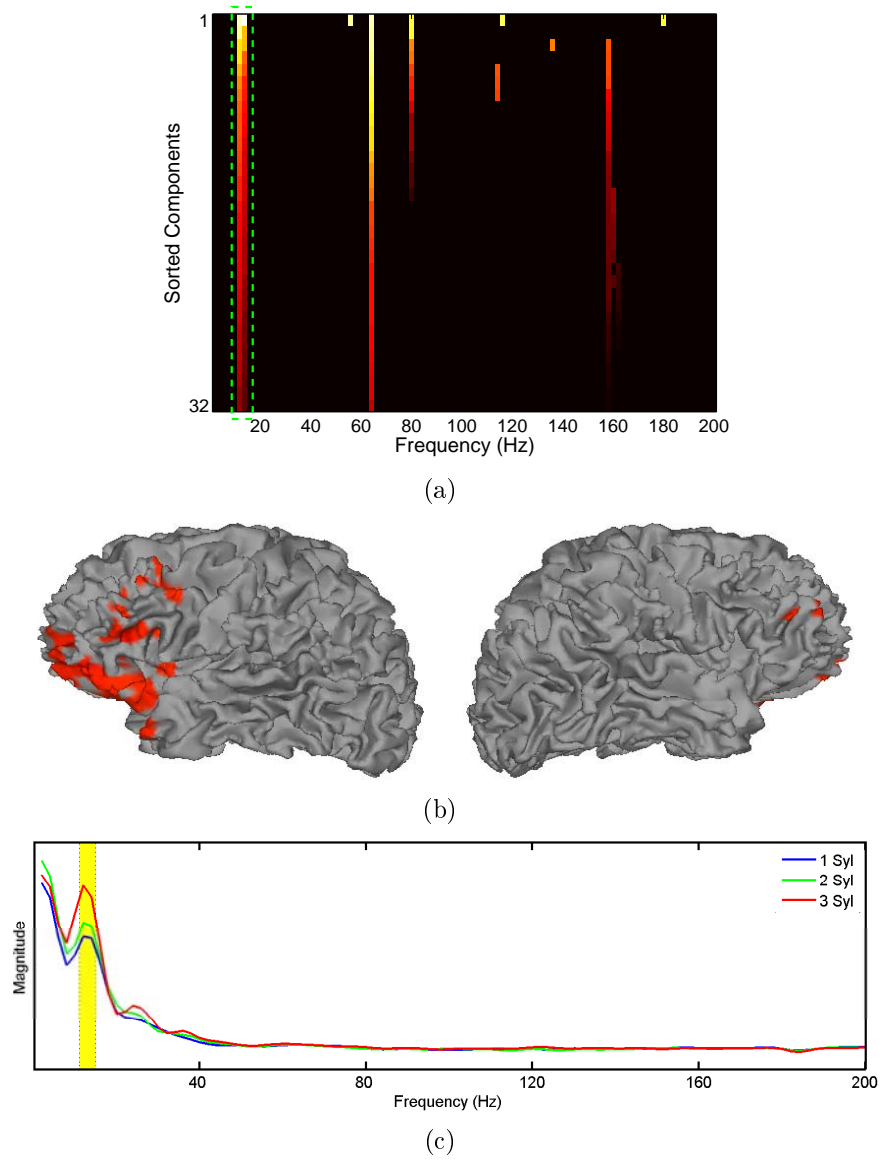


Figure 4-7: Results of MEG analysis for syllable sequence production. (a) Components found to discriminate between the 1, 2, and 3 syllable sequence conditions. The plot indicates the top 32 components \times frequency, ordered by the estimated *fitness value*. Components in black were not significant in the permutation test; components in color were significant for $P < 0.025$. (b) Significant neural source components in the ~ 10 -14 Hz range (those enclosed in the green box in (a)) rendered on the subject's reconstructed cortical surface. (c) The mean estimated frequency response during the time period of interest for each condition for a significant cortical source selected from the left inferior frontal sulcus region.

results obtained using fMRI reported in Chapter 2, the left IFS area was a region of interest. While the analysis performed indicates that the estimated strengths of these components are distributed differently across the three speaking conditions, it does not necessarily imply a rank-ordering of the *mean component strengths* that corresponds to the relative complexity of the conditions. Calculating these means for each condition, therefore, can provide additional information. Figure 4-7(c) shows the mean estimated frequency response of a characteristic component selected from the left IFS region of interest. It can be seen that in the ~ 10 -14 Hz frequency band, the strength of this source's mean estimated response to each condition follows the condition complexity, indexed by the number of syllables in each condition. Several other components were found to have significant discriminatory ability; for example, the band of components at approximately 65 Hz (see Figure 4-6) localized primarily to the right hemisphere temporal lobe and postcentral gyrus. Because these regions were not regions for which a priori hypotheses existed, and because the single-trial MEG recordings had very little overall energy in these higher frequency bands, only the low frequency (10-14 Hz) components are discussed presently.

4.6 Discussion

This chapter presented preliminary efforts using magnetoencephalography with the end goal of facilitating the inclusion of additional neurological datasets in the study of sequencing in speech production. While previous studies have used MEG to examine speech production, these have tended to focus on different aspects of the speech system. Additionally many previous studies have ignored or not fully addressed potential problems due to muscle-related artifacts in the MEG recordings. The efforts presented here led to the development of an algorithm for finding cortical sources with components in time or frequency whose estimated strengths varied along

with the experimental conditions. The algorithm was first applied to “standard” measurements of visual evoked fields, and then applied to data from a single subject in an overt syllable sequence production task.

4.6.1 Visual evoked fields

A simple “baseline” experiment was performed to measure visually evoked magnetic fields. Such fields were first reported by Brenner et al. (1975). Typical analyses of VEFs involve averaging the sensor signals over tens or even hundreds of trials. Here all analyses were performed using single trial data. In this experiment, the subject viewed either 1) full-field or 2) right-field only⁶ black and white checkerboard patterns (see Figure 4.1. The method used here, when presented with raw⁷ time series containing 250 samples (250 ms) following stimulus onset, located neural sources relevant for discriminating the two conditions primarily between 85 and 150 ms post-stimulus onset, with most sources localized to the primary and secondary visual cortices (see Figure 4.4). This result is consistent with the known data concerning visual evoked responses, which suggest that the primary evoked response begins at around 80-100 ms post-stimulus onset (Ahlfors et al., 1992). Furthermore, the algorithm found more discriminatory sources in the *right* hemisphere than in the left. This is an expected result since the right hemisphere visual cortex, which preferentially processes left visual field inputs should show a large difference between right-field checkerboard patterns (where there is less left-field stimulation) and full-field checkerboard patterns (where there is full stimulation of the left visual field). It is worth noting that the algorithm did not find significant components prior to 80

⁶Note that the subject also viewed left-field checkerboard patterns, but due to a technical problem those trials could not be used.

⁷The data presented to the algorithm were noise-reduced using the CALM algorithm (Adachi et al., 2001) and low-pass filtered by the acquisition system at 200 Hz, but were not additionally post-processed.

ms post-onset which might have been considered false positives. It is also of interest that, generally, the later (in time) that significant sources reached peak *fitness*, the more likely they were to be outside the early visual cortices (see Figure 4.4(c)). This is consistent with the notion that visual information is processed first in the primary and secondary visual cortices, then projected to many additional cortical regions which might also show a different response that depends on earlier processing. The results from this visual evoked field experiment provide evidence that the methods developed and utilized herein produce results that are consistent with other analysis procedures and theoretical expectations, at least for simple tasks.

4.6.2 Speech production

Magnetoencephalography and surface electromyography were used in conjunction to study the brain responses for planning the overt production of one, two, or three syllable sequences. Because overt speech production involves electrical activation of the facial muscles, non-cortical sources can strongly influence the signals recorded at MEG sensors (e.g. Szirtes and Vaughan, 1977; Loose et al., 2001; Zimmerman and Scharein, 2004; Furlong et al., 2004). In the present investigation steps were taken to exclude signals recorded from the period of facial muscle activation, while still allowing the use of a natural *overt* production paradigm. This is an important consideration because differences have been found in the neural processing of covert speech compared to overt speech (e.g. Riecker et al., 2000a; Munhall, 2001; Shuster and Lemieux, 2005).

The time period of interest chosen in this analysis was between the onset of the GO signal and the onset of EMG activity recorded from the *orbicularis oris* muscle. The duration of this interval was shown to vary systematically with the number of syllables being planned, with mean durations between 200 and 250 ms. Increased

reaction time with increasing number of planned elements (or the *sequence length effect on latency*) is a fundamental prediction of the competitive queuing architecture (e.g. Boardman and Bullock, 1991), which forms the basis of the GODIVA model (see Chapter 3). The pattern of latencies to initiate syllable sequence production as determined by EMG onset, therefore, were consistent with the modeling work presented here. The time period following the GO signal but prior to initiation of the utterance was thought to be an interval in which the speech plan would be maintained *in parallel* in the cortex at the *lowest level* of representation prior to articulation. Examining earlier intervals, such as the time period immediately following stimulus onset, might probe different processes or working memory representations, such as the process of sensory encoding of the stimuli. In the GODIVA model, the lowest-level parallel representation of *content* (phonemes) is hypothesized to occur in the left inferior frontal sulcus region, and the corresponding representation of *structure* (abstract syllable frames) is hypothesized to occur in the pre-SMA.

The primary question addressed by this preliminary study was whether a small set of neural sources could be found, such that these provided a measure of *discrimination* between the three speaking conditions in the time period just prior to the onset of articulation of the first speech sound. The GODIVA model predicts that the overall activation level of sources in the left IFS and pre-SMA will vary as a function of the number of elements represented; thus, there should be a larger response for the 3-syllable condition than for the 2-syllable condition, and a larger response for the 2-syllable condition than for the 1-syllable condition. The analysis method developed herein was applied to single trial measurements in order to find cortical sources whose estimated strengths provided a means to discriminate between the three speaking conditions, at a significantly above chance level. The individual time series were transformed into magnitude spectra in the frequency domain for

two reasons: i) time series varied in duration from trial to trial leading to ambiguity as to the optimal procedures for temporal alignment, and ii) previous studies have suggested that power within particular frequency ranges covaries with load during the maintenance of a neural representation, for instance in working memory tasks (e.g. Jensen and Tesche, 2002; Jensen et al., 2002; Leiberg et al., 2006).

Analysis of the extracted data from a single subject revealed discriminatory source components with differential energy in the approximately 10-14 Hz frequency band that localized to the left prefrontal cortex, including regions in and surrounding the inferior frontal sulcus (IFS). This band of frequencies is close to the border of the classically defined alpha (α) and low beta (β) bands. Several other studies have found alpha band enhancement as a function of memory load in variations on the classical Sternberg task using EEG (Klimesch et al., 1999; Jensen and Tesche, 2002; Busch and Hermann, 2003; Schack and Klimesch, 2002). Leiberg et al. (2006) found a monotonic increase in spectral amplitude at approximately 13-Hz over *right* prefrontal sensors as a function of memory load in a study of auditory working memory of speech sounds using MEG. This study did not involve a speech production component; rather, subjects simply had to report whether or not a probe stimulus (a syllable spoken by a female voice) was a member of a set of serially-presented syllables heard just previously. This suggests that the left hemisphere components revealed in the present study may represent *output codes* or representations that are preferentially used when articulation is required, whereas a homologous representation in the right hemisphere might be useful for working memory in non-production tasks. Interestingly, no components of interest in the present analysis localized to the pre-SMA. This could be because all syllables in the stimuli used were of the same abstract frame structure, perhaps requiring similar resources even as the number of items sharing that structural frame increased. It is also possible that the particu-

lar choices made in the methodological development throughout this investigation resulted in reduced sensitivity to the medial premotor regions.

The distribution of the identified sources in the left hemisphere prefrontal cortex (shown in Figure 4.7(b)) is quite diffuse. This is likely a consequence of the rather simplistic spatial filtering method used to estimate individual cortical source strengths. Improved spatial resolution is achievable by adding a more sophisticated beamforming approach, for example, to this estimation step (Barnes et al., 2004). The analysis method described herein has no dependence on a particular choice of spatial filtering technique. Additionally, the calculation of the forward model included the assumption that potential sources were oriented in the direction normal to the local cortical surface tangent. Such anatomical information is useful in constraining the ill-posed MEG inverse problem (Dale and Sereno, 1993). Because, however, only ~ 1500 possible sources were used, the sampling of the cortical surface is relatively sparse, introducing possible errors in the approximation of surface normals. Hillebrand and Barnes (2003) showed that the introduction of relatively small errors in these estimates can eliminate the benefit of using anatomical constraints in the forward model calculation and can introduce potentially large errors.

In summary, this chapter described preliminary efforts to apply the technique of magnetoencephalography to the study of syllable sequence planning and production. By measuring simultaneous EMG during MEG acquisition, it was possible to parcellate the times series into intervals of interest that would not be contaminated by possible myogenic artifacts. A method was developed in order to examine the time series trial by trial, with the goal of identifying cortical source components that showed differential estimated responses across the experimental conditions. The algorithm revealed that, between the time of the GO signal and the onset of articulation, the strength of left prefrontal activity in the ~ 10 -14 Hz frequency range (high alpha /

low beta) was related to the number of syllables planned. This is thought to be an MEG correlate of results obtained in a similar task using fMRI (Chapter 2), and to support one key prediction of the GODIVA model of speech production presented in Chapter 3. Finally, it must be emphasized that these results are from a single subject and therefore cannot be considered to be reflective of the population at large. Still, this chapter has outlined a promising approach to using MEG to study speech production. Further investigation is necessary to improve these procedures, and more subjects will need to be tested before these results can be considered conclusive.

CHAPTER 5

CONCLUSION

The combination of well-designed experimental studies using functional neuroimaging and the development of neurobiologically realistic computational models offers a framework for extending our understanding of the normal and disordered function of the speech production system. In this dissertation, these methods have been applied to the study of syllable sequencing; that is, how does the speaker represent, organize, and enable the appropriate production of arbitrary syllable sequences from his or her language? This theoretical question, although fundamental in the study of speech production, has been either neglected or addressed with treatments that lack neurobiological specificity or plausibility. Previous pertinent experimental findings have been sparse and inconsistent, owing perhaps to methodological issues and to a general lack of focus on the sequencing problem itself.

5.1 Summary of contributions

The fMRI study discussed in Chapter 2 (see also Bohland and Guenther, 2006) provides perhaps the most thorough existing examination of the effects of variations in the serial complexity of simple speech utterances on brain activity during the speech production process. This investigation utilized a combination of modern imaging procedures designed to optimize detection of effects of interest. Specifically, this included the use of sparse *event triggered* image acquisition and random (non-

blocked) stimulus presentation, non-parametric statistical analyses at the group level, permutation testing that combined both voxel height and cluster extent into a single statistical map for each effect of interest, and region-of-interest (ROI) level tests to improve anatomical specificity and to test for hemispheric lateralization.

The results of this experiment showed conclusively that, as an utterance that must be planned and produced by the speaker becomes more complex in terms of its serial composition, a network of cortical and subcortical brain regions, largely outside of the network responsible for simple articulation of a speech sound, becomes additionally engaged. This network included the left hemisphere inferior frontal sulcus and posterior parietal cortex, and bilateral anterior insula and frontal operculum, medial premotor cortices, basal ganglia, anterior thalamus, and cerebellum. When contrasting production trials with preparation only trials (averaged across all stimulus conditions), the *basic* speech production network, largely including the regions treated in the DIVA model (Guenther et al., 2006) as well as the supplementary motor area, was instead highlighted.

Based on these experimental findings as well as previously published reports, a new model, GODIVA, was proposed and specified that extended the DIVA model to include explicit parallel planning of forthcoming utterances. The modeling of such planning representations was based largely on previous work using the biologically plausible *competitive queuing* architecture (Grossberg, 1978a,b; Houghton, 1990; Bullock and Rhodes, 2003). It was hypothesized that the left inferior frontal sulcus region coded for the *content* (e.g. phonemes) of a forthcoming syllable sequence, whereas the pre-SMA coded for the abstract syllable frames in the utterance. This complementarity was proposed on the basis of differing response profiles in these regions in the present fMRI study as well as on the basis of behavioral data from speech errors and previous theoretical proposals (see especially MacNeilage, 1998). A *planning*

loop through the basal ganglia was proposed to coordinate activity between these two representations. Finally, it was hypothesized that these phonological representations interface with sensorimotor programs represented in the Speech Sound Map component of the DIVA model; simulations showed that this interface is capable of selecting appropriate sensorimotor targets for both syllables that are present and not present in the model's Speech Sound Map.

Chapter 4 presented a preliminary investigation using MEG. Magnetoencephalography is an attractive tool to investigate speech production because of its high temporal resolution relative to fMRI or PET, but presents difficulties due to potential contamination of measurements due to activation of the facial musculature during articulation. By measuring surface electromyography of the face muscles (in particular the *orbicularis oris*) simultaneously with MEG acquisition, it was possible to isolate the onset of such myogenic activity (the onset of articulation) in each trial. It was hypothesized that the time period between the GO signal informing the subject to overtly produce a planned utterance and the onset of articulation would contain a small set of components that would respond differentially based on the serial complexity (in this case the length) of the syllable sequence being performed. Using a novel procedure operating on single trial data, it was found that components in this time window within the $\sim 10\text{-}14$ Hz frequency range did exactly this, and these components were localized to the left lateral prefrontal cortex, consistent with the proposal that the left IFS contains a parallel representation of phonemic content in the forthcoming utterance. These results, while preliminary, are informative and unique, and warrant further investigation.

5.2 Future Directions

The results presented in this dissertation, while significant, leave many unanswered questions regarding how articulatory sequences are planned and represented. Even within the context of the GODIVA model as described, several topics should be addressed more fully. In particular, synaptic connections that are proposed to be adaptive were “hand-wired” in the model. It must be shown, for example, that the system can *learn* to form associations between the categorical phonemic representations in the model’s left IFS and the particular sensorimotor program representations in the Speech Sound Map that include those phonemes. The representation in IFS was suggested to arise as a result of *perceptual* learning, and thus it may be plausible that such associations are made as a child activates a developing motoric program, then categorizes the sounds that (s)he has just produced. To this end, it appears beneficial to model the learning of these synaptic weights in the broader context of neurolinguistic development.

While the model described has been based largely on observations from fMRI, it has not been *precisely fit* to the BOLD responses obtained in the experiment described in Chapter 2. Such a fit could be obtained by convolving the responses of the model’s components with an idealized hemodynamic response function as described in Guenther et al. (2006) and comparing the resulting “synthetic BOLD responses” across simulations that reflect the same speaking conditions as those used in the experiment. Such a comparison would help to validate the model as presently formulated. In addition this dissertation has presented data from speech error and reaction time studies, which proved beneficial in making design choices for the model. However, these rich data sets must now be explained by the model. A thorough treatment of the patterns observed in normally occurring slips of the tongue and of observations from reaction time studies will inevitably lead to modifications and

extensions to the present formulation of GODIVA.

The problem of sequencing speech sounds for production should also consider issues related to timing and prosody. At present, GODIVA only explicitly represents *order* but not precise timing or temporal relationships between, for instance, individual syllables. Speakers are, of course, readily capable of modulating the rate and rhythm of their overt productions, as well as, for example, stress patterns across a phrase. The basal ganglia and their connections with the medial premotor cortices have been frequently implicated in the regulation of motor timing (e.g. Harrington and Haaland, 1998; Ivry, 1996; Macar et al., 2002). To the extent that timing is a property of production that can be regulated independently of the speech sounds being produced, it is not surprising that such control processes would be regulated by the same or similar circuits as those hypothesized to control structural frames. One or more additional neuroimaging experiments is likely necessary to help elucidate the precise localization of such mechanisms in the production of syllable sequences.

A more comprehensive model of the representations of sequences of speech sounds and the interface between these representations and the articulatory system must also include the treatment of additional brain regions. In the experiment described in Chapter 2, for example, the superior parietal cortices, anterior insula, and inferior right cerebellum exhibited task-relevant modulations in activity that cannot be accounted for by the present GODIVA model. Possible mechanistic roles for these areas were discussed in Chapter 2, but it is likely that further experimentation will be required to validate these hypotheses.

An especially important future direction for research based upon the GODIVA model is in developing models of the *disordered* speech system. Several researchers have stressed the importance of uniting neurolinguistic models with models of speech motor control in helping to understand disorders such as apraxia of speech and stut-

tering. The GODIVA model appears to be among the best suited existing models for making contributions to this area of study. It was noted in Chapter 3 that specific damage to GODIVA (through the addition of noise or destruction of cells or connections) will lead to specific problems in productions. It is of great interest to explore how such simulated disorders can be related to observations in real patients. One such computational study is currently in progress (Oren Civier, personal communication), specifically investigating how manipulations to the model’s basal ganglia circuits might lead to stuttering behavior (cf. Alm, 2004). Projects of this nature will likely lead to experimental hypotheses about clinical populations, which can in turn also be tested using neuroimaging.

Finally, in continuing the combined experimental and computational approach championed here, it would be beneficial to explore additional methods for analyzing the resultant functional MRI and/or MEG data sets. In particular, covariance-based estimation procedures can be used to estimate *effective connectivity* between brain regions (e.g. McIntosh and Gonzalez-Lima, 1994; Horwitz et al., 2001; Friston et al., 2003; Harrison et al., 2003). Rather than modeling the responses of each brain region (or voxel) independently, such methods assess the functional integration of a network of interacting regions. This allows the experimenter to test aspects of a model that go beyond simple questions of whether a region should, for example, be more active in Task A than in Task B. Instead, given a network of interacting regions defined by a well-specified model, one can determine which pathways drive activation in a particular region, and how the relevant pathways change as experimental conditions change. The framework provided by a computational model will help greatly to focus such research questions.

REFERENCES

- Abeles, M. (1991). *Corticonics - Neural circuits of the cerebral cortex*. Cambridge University Press, Cambridge, UK.
- Abrahams, S., Goldstein, L. H., Simmons, A., Brammer, M. J., Williams, S. C. R., Giampietro, V. P., Andrew, C. M., and Leigh, P. N. (2003). Functional magnetic resonance imaging of verbal fluency and confrontation naming using compressed image acquisition to permit overt responses. *Human Brain Mapping*, 20(1):29–40.
- 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., Mathiak, K., and Ivry, R. B. (2004). Temporal organization of "internal speech" as a basis for cerebellar modulation of cognitive functions. *Behavioral and Cognitive Neuroscience Reviews*, 3(1):14–22.
- Ackermann, H. and Riecker, A. (2004). The contribution of the insula to motor aspects of speech production: a review and a hypothesis. *Brain and Language*, 89:320–328.
- Ackermann, H., Vogel, M., Petersen, D., and Poremba, M. (1992). Speech deficits in ischaemic cerebellar lesions. *Journal of Neurology*, 239(4):223–227.
- Adachi, Y., Shimogawara, M., Higuchi, M., Haruta, Y., and Ochiai, M. (2001). Reduction of non-periodic environmental magnetic noise in MEG measurement by continuously adjusted least squares method. *IEEE Transactions on Applied Superconductivity*, 11:669–672.
- Agam, Y., Bullock, D., and Sekuler, R. (2005). Imitating unfamiliar sequences of connected linear motions. *Journal of Neurophysiology*, 94(4):2832–2843.
- Ahlfors, S. P., Ilmoniemi, R. J., and Hämäläinen, M. S. (1992). Estimates of visually evoked cortical currents. *Electroencephalography and Clinical Neurophysiology*, 82(3):225–236.
- Alario, F.-X., Ferrand, L., Laganaro, M., New, B., Frauenfelder, U. H., and Segui, J. (2004). Predictors of picture naming speed. *Behavior Research Methods, Instruments, and Computers*, 36:140–155.

- Albin, R. L., Young, A. B., and Penney, J. B. (1989). The functional anatomy of basal ganglia disorders. *Trends in Neurosciences*, 12(10):366–375.
- Alexander, G. E. and Crutcher, M. D. (1990). Functional architecture of basal ganglia circuits: neural substrates of parallel processing. *Trends in Neurosciences*, 13:266–271.
- Alexander, G. E., DeLong, M. R., and Strick, K. L. (1986). Parallel organization of functionally segregated circuits linking basal ganglia and cortex. *Annual Reviews of Neuroscience*, 9:357–381.
- Alm, P. A. (2004). Stuttering and the basal ganglia circuits: a critical review of possible relations. *Journal of Communication Disorders*, 37:325–369.
- Anderson, J. D., Wagovich, S. A., and Hall, N. E. (2006). Nonword repetition skills in young children who do and do not stutter. *Journal of Fluency Disorders*, 31:177–199.
- Atkinson, R. C. and Shiffrin, R. M. (1971). The control of short term memory. *Scientific American*, 225:82–90.
- Augustine, J. R. (1996). Circuitry and functional aspects of the insular lobe in primates including humans. *Brain Research Reviews*, 22(3):229–244.
- Averbeck, B. B., Chafee, M. V., Crowe, D. A., and Georgopoulos, A. P. (2003). Neural activity in prefrontal cortex during copying geometrical shapes. I. Single cells encode shape, sequence, and metric parameters. *Experimental Brain Research*, 150(2):127–141.
- Averbeck, B. E., Chafee, M. V., Crowe, D. A., and Georgopoulos, A. P. (2002). Parallel processing of serial movements in prefrontal cortex. *Proceedings of the National Academy of Sciences*, 99(20):13172–13177.
- Awh, E., Jonides, J., Smith, E. E., Schumacher, E. H., Koeppel, R. A., and Katz, S. (1996). Dissociation of storage and rehearsal in verbal working memory. *Psychological Science*, 7:25–31.
- Baayen, R. H., Piepenbrock, R., and Gulikers, L. (1995). *The CELEX Lexical Database (Release 2) [CD-ROM]*. Linguistic Data Consortium, University of Pennsylvania, Philadelphia, PA.
- Bachoud-Lévi, A. C., Dupoux, E., Cohen, L., and Mehler, J. (1998). Where is the length effect? a cross-linguistic study. *Journal of Memory and Language*, 39:331–346.
- Baddeley, A. D. (1986). *Working Memory*. Oxford University Press, Oxford.

- Baillet, S., Mosher, J. C., and Leahy, R. M. (2001). Electromagnetic brain mapping. *IEEE Signal Processing Magazine*, pages 14–30.
- Barbas, H. and Rempel-Clower, N. (1997). Cortical structure predicts the pattern of corticocortical connections. *Cerebral Cortex*, 7:635–646.
- Barch, D. M., Sabb, F. W., Carter, C. S., Braver, T. S., Noll, D. C., and Cohen, J. D. (1999). Overt verbal responding during fMRI scanning: empirical investigations of problems and potential solutions. *NeuroImage*, 10(6):642–657.
- Barnes, G. R., Hillebrand, A., Fawcett, I. P., and Singh, K. D. (2004). Realistic spatial sampling for MEG beamformer images. *Human Brain Mapping*, 23:120–127.
- Beiser, D. and Houk, J. (1998). Model of cortical-basal ganglionic processing: encoding the serial order of sensory events. *Journal of Neurophysiology*, 79:3168–3188.
- Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B, Methodological*, 57:289–300.
- Berg, T. (2006). A structural account of phonological paraphasias. *Brain and Language*, 96(3):331–356.
- Berger, H. (1929). Über das electrenkephalogramm des menschen. *Archiv. für Psychiatrie und Nervenkrankheiten*, 7:527–570.
- Birn, R. M., Bandettini, P. A., Cox, R. W., Jesmanowicz, A., and Shaker, R. (1998). Magnetic field changes in the human brain due to swallowing or speaking. *Magnetic Resonance in Medicine*, 40:55–60.
- Birn, R. M., Bandettini, P. A., Cox, R. W., and Shaker, R. (1999). Event-related fMRI of tasks involving brief motion. *Human Brain Mapping*, 7:106–114.
- Birn, R. M., Cox, R. W., and Bandettini, P. A. (2004). Experimental designs and processing strategies for fMRI studies involving overt verbal responses. *NeuroImage*, 23(3):1046–1058.
- Blank, S. C., Scott, S. K., Murphy, K., Warburton, E., and Wise, R. J. (2002). Speech production: Wernicke, Broca and beyond. *Brain*, 125(Pt 8):1829–1838.
- Boardman, I. and Bullock, D. (1991). A neural network model of serial order recall from short-term memory. In *Proceedings of the International Joint Conference on Neural Networks (Seattle)*, volume II, pages 879–884, Piscataway, NJ.

- Bohland, J. W. and Guenther, F. H. (2006). An fMRI investigation of syllable sequence production. *NeuroImage*, 32(2):821–841.
- Bolam, J. P., Hanley, J. J., Booth, P. A. C., and Bevan, M. D. (2000). Synaptic organisation of the basal ganglia. *Journal of Anatomy*, 196:527–542.
- Bosshardt, H. G. (1993). Differences between stutterers’ and nonstutterers’ short-term recall and recognition performance. *Journal of Speech and Hearing Research*, 36(2):286–293.
- Bradski, G., Carpenter, G. A., and Grossberg, S. (1994). STORE working memory networks for storage and recall of arbitrary temporal sequences. *Biological Cybernetics*, 71:469–480.
- Braitenberg, V., Heck, D., and Sultan, F. (1997). The detection and generation of sequences as a key to cerebellar function: experiments and theory. *Behavioral and Brain Sciences*, 20(2):229–245.
- Brenner, D., Williamson, S. J., and Kaufman, L. (1975). Visually evoked magnetic fields of the human brain. *Science*, 190(4213):480–482.
- Brown, G. D. A., Preece, T., and Hulme, C. (2000). Oscillator-based memory for serial order. *Psychological Review*, 107(1):127–181.
- 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.
- Bullock, D. (2004). Adaptive neural models of queuing and timing in fluent action. *Trends in Cognitive Science*, 8(9):426–433.
- Bullock, D., Grossberg, S., and Mannes, C. (1993). A neural network model for cursive script production. *Biological Cybernetics*, 70(1):15–28.
- Bullock, D. and Rhodes, B. (2003). Competitive queuing for serial planning and performance. In Arbib, M., editor, *Handbook of brain theory and neural networks*, 2ed, pages 241–244. MIT Press, Cambridge, MA.
- Burgess, N. and Hitch, G. J. (1992). Toward a network model of the articulatory loop. *Journal of Memory and Language*, 31(4):429–460.
- Burgess, N. and Hitch, G. J. (1999). Memory for serial order: a network model of the phonological loop and its timing. *Psychological Review*, 106(3):551–581.
- Busch, N. A. and Hermann, C. S. (2003). Object-load and feature-load modulate EEG in a short-term memory task. *NeuroReport*, 14(13):1721–1724.

- Callan, D. E., Jones, J. A., Callan, A. M., and Akahane-Yamada, R. (2004). Phonetic perceptual identification by native- and second-language speakers differentially activates brain regions involved with acoustic phonetic processing and those involved with articulatory-auditory/orosensory internal models. *NeuroImage*, 22:1182–1194.
- 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*, 43:721–736.
- Carreiras, M. and Perea, M. (2004). Naming pseudowords in Spanish: effects of syllable frequency. *Brain and Language*, 90:393–400.
- Chen, S. H. and Desmond, J. E. (2005). Cerebrocerebellar networks during articulatory rehearsal and verbal working memory tasks. *NeuroImage*, 24(2):332–338.
- Cholin, J., Levelt, W. J. M., and Schiller, N. O. (2006). Effects of syllable frequency in speech production. *Cognition*, 99(2):205–235.
- Cholin, J., Schiller, N. O., and Levelt, W. J. M. (2004). The preparation of syllables in speech production. *Journal of Memory and Language*, 50:47–61.
- Chomsky, N. and Halle, M. (1968). *The sound pattern of english*. MIT Press, Cambridge, MA.
- Cisek, P. and Kalaska, J. F. (2002). Simultaneous encoding of multiple potential reach direction in dorsal premotor cortex. *Journal of Neurophysiology*, 87:1149–1154.
- Clower, W. T. and Alexander, G. E. (1998). Movement sequence-related activity reflecting numerical order of components in supplementary and presupplementary motor areas. *Journal of Neurophysiology*, 80:1562–1566.
- Cohen, D. (1972). Magnetoencephalography: Detection of the brain's electrical activity with a superconducting magnetometer. *Science*, 175:664–666.
- Cohen, J. D., MacWhinney, B., Flatt, M., and Provost, J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments, and Computers*, 25(2):257–271.
- Conrad, R. (1965). Order error in immediate recall of sequences. *Journal of Verbal Learning and Verbal Behavior*, 4:161–169.
- Cornsweet, T. N. (1970). *Visual perception*. Academic Press, New York.

- Cowan, N. (2000). The magical number 4 in short-term memory: a reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, 24:87–185.
- Crinion, J., Turner, R., Grogan, A., Hanakawa, T., Noppeney, U., Devlin, J. T., Aso, T., Urayama, S., Fukuyama, H., Stockton, K., Usui, K., Green, D. W., and Price, C. J. (2006). Language control in the bilingual brain. *Science*, 312(5779):1537–1540.
- Crompton, A. (1982). Syllables and segments in speech production. In Cutler, A., editor, *Slips of the tongue and language production*, pages 109–162. Mouton, Berlin.
- Crosson, B., Benefield, H., Cato, M. A., Sadek, J. R., Moore, A. B., Wieranga, C. E., Gopinath, K., Soltysik, D., Bauer, R. M., Auerbach, E. J., Gökçay, D., Leonard, C. M., and Briggs, R. W. (2003). Left and right basal ganglia and frontal activity during language generation: contributions to lexical, semantic, and phonological processes. *Journal of the International Neuropsychological Society*, 9(7):1061–1077.
- Crosson, B., Sadek, J. R., Maron, L., Gökçay, D., Mohr, C. M., Auerbach, E. J., Freeman, A. J., Leonard, C. M., and Briggs, R. W. (2001). Relative shift in activity from medial to lateral frontal cortex during internally versus externally guided word generation. *Journal of Cognitive Neuroscience*, 13:272–283.
- Crosson, B. A. (1992). *Subcortical functions in language and memory*. Guilford Press, New York.
- Crottaz-Herbette, S., Anagnoson, R. T., and Menon, V. (2004). Modality effects in verbal working memory: differential prefrontal and parietal responses to auditory and visual stimuli. *NeuroImage*, 21(1):340–351.
- Cutler, A. (1982). The reliability of speech error data. In Cutler, A., editor, *Slips of the tongue and language production*, pages 7–28. Mouton, Amsterdam.
- Dale, A., Fischl, B., and Sereno, M. I. (1999). Cortical surface-based analysis. I. Segmentation and surface reconstruction. *NeuroImage*, 9(2):179–194.
- Dale, A. M. and Sereno, M. I. (1993). Improved localization of cortical activity by combining EEG and MEG with MRI cortical surface reconstruction: A linear approach. *Journal of Cognitive Neuroscience*, 5(2):162–176.
- Damasio, H. and Damasio, A. R. (1980). The anatomical basis of conduction aphasia. *Brain*, 103:337–350.
- Darley, F. L., Aronson, A. E., and Brown, J. R. (1975). *Motor speech disorders*. Saunders, Philadelphia.

- de Zubizaray, G. I., Zelaya, F. O., Andrew, C., Williams, S. C., and Bullmore, E. T. (2000). Cerebral regions associated with verbal response initiation, suppression and strategy use. *Neuropsychologia*, 38(9):1292–1304.
- Dell, G. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, 93(3):283–321.
- Dell, G. S., Burger, L. K., and Svec, W. R. (1997). Language production and serial order: a functional analysis and a model. *Psychological Review*, 104(1):123–147.
- Deniau, J. M. and Chevalier, G. (1985). Disinhibition as a basic process in the expression of striatal functions. II. The striato-nigral influence on thalamocortical cells of the ventromedial thalamic nucleus. *Brain Research*, 334:227–233.
- 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.
- D’Esposito, M., Aguirre, G. K., Zarahn, E., Ballard, D., Shin, R. K., and Lease, J. (1998). Functional MRI studies of spatial and nonspatial working memory. *Cognitive Brain Research*, 7(1):1–13.
- Donders, F. C. (1969). Over de snelheid van psychische proessen (On the speed of mental processes). *Acta Psychologica*, 30:412–431. Originally published in 1868.
- Doupe, A. J. and Kuhl, P. K. (1999). Birdsong and human speech: common themes and mechanisms. *Annual Review of Neuroscience*, 22:567–631.
- Dronkers, N. F. (1996). A new brain region for coordinating speech articulation. *Nature*, 384(6605):159–161.
- Duffy, J. R. (1995). *Motor Speech Disorders*. Mosby, St. Louis, MO.
- Ebbinghaus, H. (1885/1913). *Memory: A contribution to experimental psychology*. Teachers College, Columbia University, New York.
- Eden, G. F., Joseph, J. E., Brown, H. E., Brown, C. P., and Zeffiro, T. A. (1999). Utilizing hemodynamic delay and dispersion to detect fMRI signal change without auditory interference: the behavior interleaved gradients technique. *Magnetic Resonance in Medicine*, 41(1):13–20.
- Elman, J. L. (1990). Finding structure in time. *Cognition*, 14:179–211.
- Eriksen, C. W., Pollack, M. D., and Montague, W. E. (1970). Implicit speech: mechanism in perceptual encoding? *Journal of Experimental Psychology*, 84:502–507.

- Ertl, J. and Schafer, E. W. P. (1967). Cortical activity preceding speech. *Life Sciences*, 6:473–479.
- Ertl, J. and Schafer, E. W. P. (1969). Erratum. *Life Sciences*, 8:559.
- 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. In *Proc. IEEE-Nuclear Science Symposium and Medical Imaging Conference*, pages 1813–1817.
- Farrell, S. and Lewandowsky, S. (2004). Modelling transposition latencies: Constraints for theories of serial order memory. *Journal of Memory and Language*, 51:115–135.
- Ferrand, L. and Segui, J. (1998). The syllable’s role in speech production: Are syllables chunks, schemas, or both? *Psychonomic Bulletin and Review*, 5(2):253–258.
- Fiez, J. A. (2001). Neuroimaging studies of speech: an overview of techniques and methodological approaches. *Journal of Communication Disorders*, 34(6):445–454.
- Fiez, J. A., Raife, E. A., Balota, D. A., Schwarz, J. P., Raichle, M. E., and Petersen, S. E. (1996). A positron emission tomography study of the short-term maintenance of verbal information. *Journal of Neuroscience*, 16(2):808–822.
- Fischl, B., Sereno, M. I., and Dale, A. M. (1999). Cortical surface-based analysis II. Inflation, flattening, and a surface-based coordinate system. *NeuroImage*, 9(2):195–207.
- 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.
- Flaherty, A. W. and Graybiel, A. M. (1994). Input-output organization of the sensorimotor striatum in the squirrel monkey. *Journal of Neuroscience*, 13:1120–1137.
- Flynn, F. G., Benson, D. F., and Ardila, A. (1999). Anatomy of the insula - functional and clinical correlates. *Aphasiology*, 13(1):55–78.
- Forster, K. I. and Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, and Computers*, 35(1):116–124.

- Fox, P. T. and Raichle, M. E. (1986). Focal physiological uncoupling of cerebral blood flow and oxidative metabolism during somatosensory stimulation in human subjects. *Proceedings of the National Academy of Sciences*, 83:1140–1144.
- Freud, S. (1914). (A. A. Brill, Trans.) *The psychopathology of everyday life*. MacMillan, New York. Originally published in German in 1901.
- Fried, I., Katz, A., McCarthy, G., Sass, K. J., Williamson, P., Spencer, S. S., and Spencer, D. D. (1991). Functional organization of the human supplementary motor cortex studied by electrical stimulation. *Journal of Neuroscience*, 11(11):3656–3666.
- Friston, K. J., Harrison, L., and Penny, W. (2003). Dynamic causal modelling. *NeuroImage*, 19(4):1273–1302.
- Friston, K. J., Penny, W. D., and Glaser, D. E. (2005). Conjunction revisited. *NeuroImage*, 25(3):661–667.
- Fromkin, V. (1980). *Errors in Linguistic Performance: Slips of the Tongue, Ear, Pen, and Hand*. Academic Press, London.
- Fudge, E. C. (1969). Syllables. *Journal of Linguistics*, 5:226–320.
- Furlong, P. L., Hobson, A. R., Aziz, Q., Barnes, G. R., Singh, K. D., Hillebrand, A., Thompson, D. G., and Hamdy, S. (2004). Dissociating the spatio-temporal characteristics of cortical neuronal activity associated with human volitional swallowing in the healthy adult brain. *NeuroImage*, 22(4):1447–1455.
- Gabrieli, J. D. E., Poldrack, R. A., and Desmond, J. E. (1998). The role of left prefrontal cortex in language and memory. *Proceedings of the National Academy of Sciences*, 95:906–913.
- Garnham, A., Shillcock, R. C., Brown, G. D. A., Mill, A. I. D., and Cutler, A. (1981). Slips of the tongue in the London-Lund corpus of spontaneous conversation. *Linguistics*, pages 805–817.
- Garrett, M. F. (1975). The analysis of sentence production. In Bower, G. H., editor, *The Psychology of Learning and Motivation*, volume 9. Academic Press, New York.
- Genovese, C. R., Lazar, N. A., and Nichols, T. (2002). Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *Neuroimage*, 15:870–878.

- Gerloff, C., Corwell, B., Chen, R., Hallett, M., and Cohen, L. G. (1997). Stimulation over human supplementary motor area interferes with the organization of future elements in complex motor sequences. *Brain*, 120(9):1587–1602.
- Ghosh, S. S. (2005). *Understanding cortical and cerebellar contributions to speech production through modeling and functional imaging*. PhD thesis, Boston University, Boston, MA.
- Ghosh, S. S., Bohland, J. W., and Guenther, F. H. (2003). Comparisons of brain regions involved in overt production of elementary phonetic units [abstract]. *Presented at 9th Annual Conference on Functional Mapping of the Human Brain, New York*. Available on CD-ROM in NeuroImage, Vol. 19, No. 2.
- Glasspool, D. W. and Houghton, G. (2005). Serial order and consonant-vowel structure in a graphemic output buffer model. *Brain and Language*, 94(3):304–330.
- Goldfield, B. A. and Reznick, J. S. (1990). Early lexical acquisition: Rate, content, and the vocabulary spurt. *Journal of Child Language*, 17:171–183.
- Goldstein, L., Pouplier, M., Chen, L., Saltzman, E., and Byrd, D. (in press). Dynamic action units slip in speech production errors. *Cognition*.
- Goncharova, I. I., McFarland, D. J., Vaughan, T. M., and Wolpaw, J. R. (2003). EMG contamination of EEG: spectral and topographical characteristics. *Clinical Neurophysiology*, 114:1580–1593.
- Goodglass, H. (1993). *Understanding Aphasia*. Academic Press, San Diego.
- Gordon, P. C. and Meyer, D. E. (1987). Hierarchical representation of spoken syllable order. In Allport, A., MacKay, D. G., and Prinz, W., editors, *Language Perception and Production*, pages 445–462. Academic Press, London.
- Grabow, J. D. and Elliott, F. W. (1974). The electrophysiologic assessment of hemispheric asymmetries during speech. *Journal of Speech and Hearing Research*, 17:64–72.
- Gracco, V. L., Tremblay, P., and Pike, B. (2005). Imaging speech production using fMRI. *NeuroImage*, 26(1):294–301.
- Graveland, G. A. (1985). A golgi study of the human neostriatum: neurons and afferent fibers. *Journal of Comparative Neurology*, 234:317–333.
- Grodd, W., Hülsmann, E., Lotze, M., Wildgruber, D., and Erb, M. (2001). Sensorimotor mapping of the human cerebellum: fMRI evidence of somatotopic organization. *Human Brain Mapping*, 13:55–73.

- Gropnik, M. and Cargo, M. B. (1990). Familial aggregation of a developmental language disorder. *Cognition*, 39:1–50.
- Grossberg, S. (1969). Some networks that can learn, remember, and reproduce any number of complicated space-time patterns, I. *Journal of Mathematics and Mechanics*, 19:53–91.
- Grossberg, S. (1973). Contour enhancement, short-term memory, and constancies in reverberating neural networks. *Studies in Applied Mathematics*, 52:213–257.
- Grossberg, S. (1976). Adaptive pattern classification and universal recoding I: Parallel development and coding of neural feature detectors. *Biological Cybernetics*, 23:121–134.
- Grossberg, S. (1978a). Behavioral contrast in short term memory: Serial binary memory models or parallel continuous memory models? *Journal of Mathematical Psychology*, 17:199–219.
- Grossberg, S. (1978b). A theory of human memory: self-organization and performance of sensory-motor codes, maps, and plans. In Rosen, R. and Snell, F., editors, *Progress in Theoretical Biology*, volume 5, pages 233–374. Academic Press, New York.
- Grossberg, S. (1986). The adaptive self-organization of serial order in behavior: Speech, language, and motor control. In Schwab, E. C. and Nusbaum, H. C., editors, *Pattern Recognition by Humans and Machines*, volume 1: Speech Perception. Academic Press, New York.
- Guenther, F. H. (1994). A neural network model of speech acquisition and motor equivalent speech production. *Biological Cybernetics*, 72:43–53.
- 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. (2001). Neural modeling of speech production. In *Proceedings of the 4th International Nijmegen Speech Motor Conference*, Nijmegen, The Netherlands.
- Guenther, F. H. (2006). Cortical interactions underlying the production of speech sounds. *Journal of Communication Disorders*, 39(5):350–365.
- Guenther, F. H., Espy-Wilson, C. Y., Boyce, S. E., Matthies, M. L., Zandipour, M., and Perkell, J. S. (1999). Articulatory tradeoffs reduce acoustic variability during American English /r/ production. *Journal of the Acoustical Society of America*, 105:2854–2865.

- Guenther, F. H. and Ghosh, S. S. (2003). A model of cortical and cerebellar function in speech. *15th International Congress of Phonetic Sciences*.
- 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.
- Gunji, A., Hoshiyama, M., and Kakigi, R. (2001). Auditory response following vocalization: a magnetoencephalographic study. *Clinical Neurophysiology*, 112(3):514–520.
- Gunji, A., Kakigi, R., and Hoshiyama, M. (2000). Spatiotemporal source analysis of vocalization-associated magnetic fields. *Cognitive Brain Research*, 9(2):157–163.
- Gunji, A., Kakigi, R., and Hoshiyama, M. (2003). Cortical activities relating to modulation of sound frequency: how to vocalize? *Cognitive Brain Research*, 17(2):495–506.
- Gupta, P., Abbs, B., and Lin, P. (2005). Serial position effects in nonword repetition. *Journal of Memory and Language*, 53:141–162.
- Gupta, P. and MacWhinney, B. (1997). Vocabulary acquisition and verbal short-term memory: computational and neural bases. *Brain and Language*, 59:267–333.
- Gurney, K., Prescott, T. J., and Redgrave, P. (2001a). A computational model of action selection in the basal ganglia. I. A new functional anatomy. *Biological Cybernetics*, 84:401–410.
- Gurney, K., Prescott, T. J., and Redgrave, P. (2001b). A computational model of action selection in the basal ganglia. II. Analysis and simulation of behavior. *Biological Cybernetics*, 84:411–423.
- Habib, M., Daquin, G., Milandre, L., Royere, M. L., Rey, M., Lanteri, A., Slamanon, G., and Khalil, R. (1995). Mutism and auditory agnosia due to bilateral insular damage - role of the insula in human communication. *Neuropsychologia*, pages 327–339.
- Hadamard, J. (1923). *Lectures on Cauchy’s problem in linear partial differential equations*. Yale University Press, New Haven, CT.
- Hakim, H. B. and Ratner, N. B. (2004). Non-word repetition abilities of children who stutter: an exploratory study. *Journal of Fluency Disorders*, 29(3):179–199.

- Hämäläinen, M., Hari, R., Ilmoniemi, R., Knuutila, J., and Lounasmaa, O. (1993). Magnetoencephalography - theory, instrumentation, and applications to noninvasive studies of the working human brain. *Reviews of Modern Physics*, 65:1–93.
- Hardcastle, W. J. and Hewlett, N., editors (1999). *Coarticulation*. Cambridge University Press, Cambridge, United Kingdom.
- Harrington, D. L. and Haaland, K. Y. (1998). Sequencing and timing operations of the basal ganglia. In Rosenbaum, D. A. and Collyer, C. E., editors, *Timing of behavior*, pages 35–61. MIT Press, Cambridge, MA.
- Harrison, L., Penny, W. D., and Friston, K. (2003). Multivariate autoregressive modeling of fMRI time series. *NeuroImage*, 19:1477–1491.
- Hartley, T. and Houghton, G. (1996). A linguistically constrained model of short-term memory for nonwords. *Journal of Memory and Language*, 35(1):1–31.
- Hayasaka, S. and Nichols, T. E. (2004). Combining voxel intensity and cluster extent with permutation test framework. *NeuroImage*, 23(1):54–63.
- 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.
- Helm-Estabrooks, N. (2002). Diagnostic and treatment issues of apraxia. *Seminars in Speech and Language*, 23:219–220.
- Helmholtz, H. (1853). Ueber einige Gesetze der Vertheilung elektrischer Ströme in körperlichen Leitern, mit Anwendung auf die thierisch-elektrischen Versuche. *Annals of Physics and Chemistry*, 89:211–233, 353–377.
- Henson, R. N., Burgess, N., and Frith, C. D. (2000). Recoding, storage, rehearsal and grouping in verbal short-term memory: an fMRI study. *Neuropsychologia*, 38:426–440.
- Henson, R. N. A. (1998). Short-term memory for serial order: The start-end model. *Cognitive Psychology*, 36:73–137.
- Henson, R. N. A., Norris, D. G., Page, M. P. A., and Baddeley, A. D. (1996). Unchained memory: Error patterns rule out chaining models of immediate serial recall. *Quarterly Journal of Experimental Psychology*, 49A(1):80–115.
- Henson, R. N. A., Rugg, M. D., and Friston, K. J. (2001). The choice of basis functions in event-related fMRI. *NeuroImage*, 13(6):127.

- Hikosaka, O. and Wurtz, R. H. (1989). The basal ganglia. In Wurtz, R. and Goldberg, M., editors, *The neurobiology of saccadic eye movements*, pages 257–281. Elsevier, Amsterdam.
- Hillebrand, A. and Barnes, G. R. (2003). The use of anatomical constraints with MEG beamformers. *NeuroImage*, 20(4):2302–2313.
- Hillis, A. E., Work, M., Barker, P. B., Jacobs, M. A., Breese, E. L., and Maurer, K. (2005). Re-examining the brain regions crucial for orchestrating speech articulation. *Brain*, 127:1479–1487.
- Hillyard, S. A. and Kutas, M. (2002). Event-related potentials and magnetic fields in the human brain. In Charney, D., Coyle, J., Davis, K., and Nemeroff, C., editors, *Neuropsychopharmacology: The Fifth Generation of Progress*. Lippincott, Williams and Wilkins, Baltimore.
- Ho, A. K., Bradshaw, J. L., Cunnington, R., Phillips, J. G., and Iansek, R. (1998). Sequence heterogeneity in parkinsonian speech. *Brain and Language*, 64:122–145.
- Hodges, P. W. and Bui, B. H. (1996). A comparison of computer-based methods for the determination of onset of muscle contraction using electromyography. *Electroencephalography and Clinical Neurophysiology*, 101:511–519.
- Hodgkin, A. L. and Huxley, A. F. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. *Journal of Physiology*, 117:500–544.
- Horwitz, B., Friston, K. J., and Taylor, J. G. (2001). Neural modeling and functional brain imaging: an overview. *Neural Networks*, 13:829–846.
- 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.
- Houghton, G. (1990). The problem of serial order: a neural network model of sequence learning and recall. In Dale, R., Mellish, C., and Zock, M., editors, *Current Research in Natural Language Generation*, pages 287–319. Academic Press, London.
- Houghton, G. and Hartley, T. (1996). Parallel models of serial behaviour: Lashley revisited. *Psyche*, 2(25).
- Huang, M. X., Mosher, J. C., and Leahy, R. M. (1999). A sensor-weighted overlapping-sphere head model and exhaustive head model comparison for MEG. *Physics in Medicine and Biology*, 44:423–440.

- Hurst, J. A., Baraitser, M., Auger, E., Graham, F., and Norel, S. V. (1990). An extended family with a dominantly inherited speech disorder. *Dev. Med. Child Neurol.*, 32:352–355.
- Hyvarinen, A. and Oja, E. (2000). Independent component analysis: algorithms and applications. *Neural Networks*, 13:411–430.
- Inase, M., Tokuno, H., Nambu, A., Akazawa, T., and Takada, M. (1999). Corticostriatal and corticosubthalamic input zones from the presupplementary motor area in the macaque monkey: comparison with the input zones from the supplementary motor area. *Brain Research*, 833:191–201.
- Indefrey, P. and Levelt, W. J. (2000). The neural correlates of language production. In Gazzaniga, M., editor, *The New Cognitive Neurosciences, 2nd edition*, pages 845–865. MIT Press, Cambridge, MA.
- Inhoff, A., Diener, H., Rafal, R., and Ivry, R. (1989). The role of cerebellar structures in the execution of serial movements. *Brain*, 112(Pt 3):565–581.
- Ivry, R. B. (1996). The representation of temporal information in perception and motor control. *Current Opinion in Neurobiology*, 6(6):851–857.
- Jaeger, D., Kita, H., and Wilson, C. J. (1994). Surround inhibition among projection neurons is weak or non-existent in the rat neostriatum. *Journal of Neurophysiology*, 72:2555–2558.
- Jakobson, R., Fant, G., and Halle, M. (1952). Preliminaries to speech analysis: The distinctive features and their correlates. Technical Report 13, MIT Acoustics Laboratory, Cambridge, MA.
- Jensen, O., Gelfand, J., Kounious, K., and Lisman, J. E. (2002). Oscillations in the alpha band (9-12 hz) increase with memory load during retention in a short-term memory task. *Cerebral Cortex*, 12:877–882.
- Jensen, O. and Tesche, C. D. (2002). Frontal theta activity in humans increases with memory load in a working memory task. *European Journal of Neuroscience*, 15:1395–1399.
- Johansen-Berg, H., Behrens, T. E. J., Robson, M. D., Drobniak, I., Rushworth, M. F. S., Brady, J. M., Smith, S. M., and Matthews, P. M. (2004). Changes in connectivity profiles define functionally distinct regions in human medial frontal cortex. *Proceedings of the National Academy of Sciences*, 101:13335–13340.
- 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 Perecman, E., editor, *The Frontal Lobes Revisited*, pages 241–250. IRBN Press, New York.
- 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.
- Jordan, M. I. (1986). Serial order: a parallel distributed processing approach. Technical Report 8604, University of California, San Diego, La Jolla, CA.
- Josephson, B. D. (1962). Possible new effects in superconductive tunneling. *Physics Letters*, 1:251–253.
- Jürgens, U. (1984). The efferent and efferent connections of the supplementary motor area. *Brain Research*, 300:63–81.
- Justus, T., Ravizza, S. M., Fiez, J. A., and Ivry, R. B. (2005). Reduced phonological similarity effects in patients with damage to the cerebellum. *Brain and Language*, 95(2):304–318.
- Kawaguchi, Y. (1993). Physiological, morphological, and histochemical characterization of three classes of interneurons in rat neostriatum. *Journal of Neuroscience*, 10:3421–3438.
- Kemp, J. M. and Powell, T. P. S. (1971). The structure of the caudate nucleus of the cat: Light and electron microscopy. *Philosophical Transactions of the Royal Society of London. B.*, 262:383–401.
- Kent, R. D. (2000). Research on speech motor control and its disorders: a review and prospective. *J. Commun. Disord.*, 33(5):391–427.
- Kent, R. D., Kent, J. F., Rosenbek, J. C., Vorperian, H. K., and Weismer, G. (1997). A speaking task analysis of the dysarthria in cerebellar disease. *Folia Phoiatrica et Logopaedica*, 49(2):63–82.
- Kerns, J. G., Cohen, J. D., Stenger, V. A., and Carter, C. S. (2004). Prefrontal cortex guides context-appropriate responding during language production. *Neuron*, 43:283–291.
- Kirschen, M. P., Chen, S. H., Schraedley-Desmond, P., and Desmond, J. E. (2005). Load- and practice-dependent increases in cerebro-cerebellar activation in verbal working memory: an fMRI study. *NeuroImage*, 24(2):462–472.
- Klapp, S. T. (1974). Syllable-dependent pronunciation latencies in number-naming, a replication. *Journal of Experimental Psychology*, 102:1138–1140.

- Klapp, S. T. (2003). Reaction time analysis of two types of motor preparation for speech articulation: action as a sequence of chunks. *Journal of Motor Behavior*, 35(2):135–150.
- Klapp, S. T., Anderson, W. G., and Berrian, R. W. (1973). Implicit speech in reading reconsidered. *Journal of Experimental Psychology*, 100:368–374.
- Klapp, S. T., Greim, D. M., and Marshburn, E. A. (1981). Buffer storage of programmed articulation and articulatory loop: Two names for the same mechanism or two distinct components of short-term memory? In Long, J. B. and Baddeley, A. D., editors, *Attention and Performance IX*, pages 459–472. Lawrence Erlbaum Associates, Hillsdale, NJ.
- Klimesch, W., Doppelmayr, M., Schwaiger, J., Auringer, P., and Winkler, T. (1999). 'paradoxical' alpha synchronization in a memory task. *Cognitive Brain Research*, 7:493–501.
- Kohn, S. E. (1992). *Conduction Aphasia*. Lawrence Erlbaum Associates, Hillsdale, NJ.
- Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, 43(1):59–69.
- Kohonen, T. (1988). The neural phonetic typewriter. *IEEE Computer*, 27(3):11–22.
- Kozinska, D., Carducci, F., and Nowinski, K. (2001). Automatic alignment of EEG/MEG and MRI data sets. *Clinical Neurophysiology*, 112:1553–1561.
- Krainik, A., Lehericy, S., Duffau, H., Capelle, L., Chainay, H., Cornu, P., Cohen, L., Boch, A. L., Mangin, J. F., LeBihan, D., and Marsault, C. (2003). Postoperative speech disorder after medial frontal surgery: role of the supplementary motor area. *Neurology*, 60(4):587–594.
- Kreisler, A., Godefroy, O., Delmaire, C., Debachy, B., Leclercq, M., Pruvo, J. P., and Leys, D. (2000). The anatomy of aphasia revisited. *Neurology*, 54:1117–1123.
- Kropotov, J. D. and Etlinger, S. C. (1999). Selection of actions in the basal ganglia-thalamocortical circuits: review and model. *International Journal of Psychophysiology*, 31(3):197–217.
- Kuriki, S., Mori, T., and Hirata, Y. (1999). Motor planning center for speech articulation in the normal human brain. *Neuroreport*, 10:765–769.
- 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., and Turner, R. (1992). Dynamic magnetic resonance imaging of human brain activity during

- primary sensory stimulation. *Proceedings of the National Academy of Sciences*, 89(12):5675–5679.
- Laganaro, M. and Alario, F.-X. (2006). On the locus of the syllable frequency effect in speech production. *Journal of Memory and Language*, 55:178–196.
- Lai, C. S. L., Fisher, S. E., Hurst, J. A., Vargha-Khadem, F., and Monaco, A. P. (2001). A forkhead-domain gene is mutated in a severe speech and language disorder. *Nature*, 413:519–523.
- Lashley, K. S. (1951). The problem of serial order in behavior. In Jeffress, L., editor, *Cerebral Mechanisms in Behavior*, pages 112–136. John Wiley and Sons, Inc.
- Lehéricy, 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.
- Leiberg, S., Lutzenberger, W., and Kaiser, J. (2006). Effects of memory load on cortical oscillatory activity during auditory pattern working memory. *Brain Research*, 1120(1):131–140.
- Leiner, H. C., Leiner, A. L., and Dow, R. S. (1993). Cognitive and language functions of the human cerebellum. *Trends in Neurosciences*, 16(11):444–447.
- Levelt, C. C., Schiller, N. O., and Levelt, W. M. (1999a). A developmental grammar for syllable structure in the production of child language. *Brain and Language*, 68:291–299.
- Levelt, W. J., Praamstra, P., Meyer, A. S., Helenius, P., and Salmelin, R. (1998). An MEG study of picture naming. *Journal of Cognitive Neuroscience*, 10(5):553–567.
- Levelt, W. J., Roelofs, A., and Meyer, A. S. (1999b). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22(1):1–38.
- Levelt, W. J. and Wheeldon, L. (1994). Do speakers have access to a mental syllabary? *Cognition*, 50:239–269.
- Levelt, W. J. M. (1989). *Speaking: From Intention to Articulation*. MIT Press, Cambridge, MA.
- Levy, R., Hazrati, L. N., Herrero, M. T., Vila, M., Hassani, O. K., Mouroux, M., Ruberg, M., Asensi, H., Agid, Y., Feger, J., Obeso, J., Parent, A., and Hirsch, E. C. (1997). Re-evaluation of the functional anatomy of the basal ganglia in normal and parkinsonian states. *Neuroscience*, 76:335–343.

- Lewandowsky, S., Brown, G. D. A., Wright, T., and Nimmo, L. M. (2006). Timeless memory: evidence against temporal distinctiveness models of short term memory for serial order. *Journal of Memory and Language*, 54:20–38.
- Lewandowsky, S. and Murdock, B. B. (1989). Memory for serial order. *Psychological Review*, 96(1):25–57.
- Locke, J. L. (1997). A theory of neurolinguistic development. *Brain and Language*, 58:265–326.
- Loose, R., Hamdy, S., and Enck, P. (2001). Magnetoencephalographic response characteristics associated with tongue movement. *Dysphagia*, 16:183–185.
- Lu, X. and Ashe, J. (2005). Anticipatory activity in primary motor cortex codes memorized movement sequences. *Neuron*, 45(6):967–973.
- Ludlow, C., Siren, K., and Zikira, M. (1997). Speech production learning in adults with chronic developmental stuttering. In Hulstijn, W., Peters, H., and Van Lieshout, P., editors, *Speech production: Motor control, brain research and fluency disorders*, pages 221–230. Elsevier, New York.
- 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*, 338:114–140.
- Macar, F., Lejeune, H. F., Ferrara, M., Pothaus, V., Vidal, F., and Maquet, P. (2002). Activation of the supplementary motor area and of attentional networks in temporal processing. *Experimental Brain Research*, 142:475–485.
- MacKay, D. G. (1970). Spoonerisms: the structure of errors in the serial order of speech. *Neuropsychologia*, 8(3):323–50.
- MacKay, D. G. (1987). *The organization of perception and action: A theory for language and other cognitive skills*. Springer Verlag, New York.
- MacNeilage, P. F. (1998). The frame/content theory of evolution of speech production. *Behavioral and Brain Sciences*, 21:499–511.
- MacNeilage, P. F. and Davis, B. (1990). Acquisition of speech production: Frames, then content. In Jeannerod, M., editor, *Attention and Performance XIII: Motor representation and control*, pages 453–476. Erlbaum, Hillsdale, NJ.
- Maeda, S. (1990). Compensatory articulation during speech: Evidence from the analysis and sythesis of vocal tract shapes using an articulatory model. In Hardcastle, W. J. and Marchal, A., editors, *Speech Production and Speech Modeling*, pages 131–149. Kluwer Academic Publishers, Boston.

- Malmivuo, J. and Plonsey, R. (1995). *Bioelectromagnetism: principles and applications of bioelectric and biomagnetic fields*. Oxford University Press, New York.
- Markram, H. and Tsodyks, M. (1996). Redistribution of synaptic efficacy between neocortical pyramidal neurons. *Nature*, 382(6594):807–810.
- Matelli, M. and Luppino, G. (1996). Thalamic input to mesial and superior area 6 in the macaque monkey. *Journal of Comparative Neurology*, 372:59–87.
- Matelli, M., Luppino, G., and Rizzolatti, G. (1991). Architecture of superior and mesial area 6 and the adjacent cingulate cortex in the macaque monkey. *Journal of Comparative Neurology*, 311(445-462).
- Matsuzaka, Y., Aizawa, H., and Tanji, J. (1992). A motor area rostral to the supplementary motor area (presupplementary motor area) in the monkey: neuronal activity during a learned motor task. *Journal of Neurophysiology*, 68(3):653–62.
- 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.
- McAdam, D. W. and Whitaker, H. A. (1971). Language production: electroencephalographic localization in the normal human brain. *Science*, 172(3982):499–502.
- McIntosh, A. R. and Gonzalez-Lima, F. (1994). Structural equation modelling and its application to network analysis in functional brain imaging. *Human Brain Mapping*, 2:2–22.
- McNeil, M. R., Pratt, S. R., and Fossett, T. R. D. (2004). The differential diagnosis of apraxia. In Maasen, B., Kent, R., Peters, H. F. M., van Leishout, P. H. H. M., and Hulstijn, W., editors, *Speech Motor Control in normal and disordered speech*, pages 389–413. Oxford University Press, New York.
- Meringer, R. and Mayer, C. (1895). *Versprechen und Verlesen: Eine Psychologisch-linguistische Studie*. Göschense verlagsbuchhandlung, Stuttgart.
- Middleton, F. A. and Strick, P. L. (2000). Basal ganglia and cerebellar loops: motor and cognitive circuits. *Brain Research Reviews*, 31:236–250.
- Mink, J. W. (1996). The basal ganglia: focused selection and inhibition of competing motor programs. *Progress in Neurobiology*, 50:381–425.
- Mink, J. W. and Thach, W. T. (1993). Basal ganglia intrinsic circuits and their role in behavior. *Current Opinion in Neurobiology*, 3:950–957.

- Morrell, L. K. and Huntington, D. A. (1971). Electrocortical localization of language production. *Science*, 174(4016):1359–1361.
- Mosher, J. C. and Leahy, R. M. (1999). Source localization using recursively applied and projected (RAP) MUSIC. *IEEE Transactions on Signal Processing*, 47:332–340.
- Mowrey, R. A. and MacKay, I. R. A. (1990). Phonological primitives: electromyographic speech error evidence. *Journal of the Acoustical Society of America*, 88:1299–1312.
- Munhall, K. G. (2001). Functional imaging during speech production. *Acta Psychologica*, 107:95–117.
- Murakami, S. and Okada, Y. (2006). Contributions of principal neocortical neurons to magnetoencephalography and electroencephalography signals. *Journal of Physiology*, 575(3):925–936.
- Murdoch, B. E. (2001). Subcortical brain mechanisms in speech and language. *Folia Phoniatrica et Logopaedica*, 53:233–251.
- Murphy, K., Corfield, D. R., Guz, A., Fink, G. R., Wise, R. J. S., Harrison, J., and Adams, L. (1997). Cerebral areas associated with motor control of speech in humans. *Journal of Applied Physiology*, 83(5):1438–1447.
- Nadeau, S. E. (2001). Phonology: a review and proposals from a connectionist perspective. *Brain and Language*, 79:511–579.
- Nagao, M., Takeda, K., Komori, T., Isozaki, E., and Hirai, S. (1999). Apraxia of speech associated with an infarct in the precentral gyrus. *Neuroradiology*, 41:356–357.
- Nebel, K., Stude, P., Wiese, H., Muller, B., deGreiff, A., Forsting, M., Diener, H. C., and Keidel, M. (2005). Sparse imaging and continuous event-related fMRI in the visual domain: a systematic comparison. *Human Brain Mapping*, 24(2):130–143.
- Nichols, T. E., Brett, M., Andersson, J., Wager, T., and Poline, J. B. (2005). Valid conjunction inference with the minimum statistic. *Neuroimage*, 25:653–660.
- Nichols, T. E. and Holmes, A. P. (2001). Nonparametric permutation tests for functional neuroimaging: A primer with examples. *Human Brain Mapping*, 15:1–25.
- 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.

- Nooteboom, S. G. (1969). The tongue slips into patterns. In Sciarone, A. G., van Essen, A. J., and van Read, A. A., editors, *Leyden studies in linguistics and phonetics*, pages 114–132. Mouton, The Hague.
- Nota, Y. and Honda, K. (2003). Possible role of the anterior insula in articulation. In Palethorpe, S. and Tabain, M., editors, *Proceedings of the 6th International Seminar on Speech Production*, pages 191–194, Sydney.
- Ogawa, S., Lee, T. M., Kay, A. R., and Tank, D. W. (1990). Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proceedings of the National Academy of Sciences*, 87:9869–9872.
- Ogawa, S., Tank, D. W., Menon, R., Ellermann, J. M., Kim, S. G., Merkle, H., and Ugurbil, K. (1992). Intrinsic signal changes accompanying sensory stimulation: functional brain mapping with magnetic resonance imaging. *Proceedings of the National Academy of Sciences*, 89(13):5951–5955.
- Ohman, S. (1966). Coarticulation in VCV utterances: spectrographic measurements. *Journal of the Acoustical Society of America*, 39:151–168.
- Okada, Y. (1993). Neurogenesis of evoked magnetic fields. In Williamson, S. H., Romani, G. L., Kaufman, L., and Modena, I., editors, *Biomagnetism: an interdisciplinary approach*, pages 399–408. Plenum Press, New York.
- Oller, D. K. (1980). The emergence of the sounds of speech in infancy. In Yeni-Komshian, G. H., Kavanagh, J. F., and Ferguson, C. A., editors, *Child Phonology, Volume 1: Production*, pages 93–112. Academic Press, New York.
- Page, M. and Norris, D. (1998). The primacy model: A new model of immediate serial recall. *Psychological Review*, 105(4):761–781.
- Page, M. P. and Norris, D. (2000). Connectionist modeling in psychology: a localist manifesto. *Behavioral and Brain Sciences*, 23:443–467.
- Pai, M. C. (1999). Supplementary motor area aphasia: a case report. *Clinical Neurology and Neurosurgery*, 101:29–32.
- Palumbo, C. L., Alexander, M. P., and Naeser, M. A. (1992). CT scan lesion sites associated with conduction aphasia. In Kohn, S. E., editor, *Conduction Aphasia*, pages 51–76. Lawrence Erlbaum Associates, Hillsdale, NJ.
- Parent, A. and Hazrati, L. N. (1993). Anatomical aspects of information processing in primate basal ganglia. *Trends in Neurosciences*, 16:111–116.

- Parent, A. and Hazrati, L. N. (1995). Functional anatomy of the basal ganglia. I. The cortico-basal ganglia-thalamo-cortical loop. *Brain Research Reviews*, 20:91–127.
- Paulesu, E., Frith, C. D., and Frackowiak, R. S. (1993). The neural correlates of the verbal component of working memory. *Nature*, 362:342–245.
- Peach, R. K. and Tonkovich, J. D. (2004). Phonemic characteristics of apraxia of speech resulting from subcortical hemorrhage. *Journal of Communication Disorders*, 37:77–90.
- Penfield, W. and Roberts, L. (1959). *Speech and brain mechanisms*. Princeton University Press, Princeton, NJ.
- Penfield, W. and Welch, K. (1951). The supplementary motor area of the cerebral cortex; a clinical and experimental study. *Archives of Neurology and Psychiatry*, 66(3):289–317.
- Perkell, J., Cohen, M., Svirsky, M., Matthies, M., Garabieta, I., and Jackson, M. (1992). Electro-magnetic midsagittal articulometer (EMMA) systems for transducing speech articulatory movements. *Journal of the Acoustical Society of America*, 92:3078–3096.
- Perkell, J. S., Guenther, F. H., Lane, H., Matthies, M. L., Perrier, P., J., V., Wilhelms-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*, 28:233–272.
- Perkell, J. S., Matthies, M. L., Lane, H., Guenther, F. H., Wilhelms-Tricarico, R., Wozniak, J., and Guiod, P. (1997). Speech motor control: Acoustic goals, saturation effects, auditory feedback and internal models. *Speech Communication*, pages 227–250.
- Perkell, J. S., Matthies, M. L., Tiede, M., Lane, H., Zandipour, M., and Marrone, N. (2004). 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.
- Perrett, S. P., Ruiz, B. P., and Mauk, M. D. (1993). Cerebellar cortex lesions disrupt learning-dependent timing of conditioned eyelid responses. *Journal of Neuroscience*, 13(4):1708–1718.
- Person, A. L. and Perkel, D. J. (2005). Unitary IPSPs drive precise thalamic spiking in a circuit required for learning. *Neuron*, 46:129–140.

- Petacchi, A., Laird, A. R., Fox, P. T., and Bower, J. M. (2005). Cerebellum and auditory function: An ALE meta-analysis of functional neuroimaging studies. *Human Brain Mapping*, 25:118–128.
- Petrides, M. (1991). Functional specialization within the dorsolateral frontal cortex for serial order memory. *Proc. R. Soc. Lond. B Biol. Sci.*, 246(1317):299–306.
- Pfurtscheller, G. and Lopes da Silva, F. H. (1999). Event-related EEG/MEG synchronization and desynchronization: basic principles. *Clinical Neurophysiology*, 110:1842–1857.
- Picard, N. and Strick, P. L. (1996). Motor areas of the medial wall: a review of their location and functional activation. *Cerebral Cortex*, 6(3):342–53.
- Pickett, E. R., 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 nucleus: a case study. *Neuropsychologia*, 36:173–188.
- Plenz, D. and Kitai, S. T. (1998). Up and down states in striatal medium spiny neurons simultaneously recorded with spontaneous activity in fast-spiking interneurons studied in cortex-striatum-substantia nigra organotypic cultures. *Journal of Neuroscience*, 18(1):266–283.
- Poupier, M. and Hardcastle, W. (2005). A re-evaluation of the nature of speech errors in normal and disordered speakers. *Phonetica*, 62(2-4):227–243.
- Pulvermüller, F. (1999). Words in the brain’s language. *Behavioral and Brain Sciences*, 22:253–336.
- Pulvermüller, F. (2002). A brain perspective on language mechanisms: from discrete neuronal ensembles to serial order. *Progress in Neurobiology*, 67:85–111.
- Pylkkänen, L. and Marantz, A. (2003). Tracking the time course of word recognition with MEG. *Trends in Cognitive Sciences*, 7(5):187–189.
- Ravizza, S. M., McCormick, C. A., Schlerf, J. E., Justus, T., Ivry, R. B., and Fiez, J. A. (2006). Cerebellar damage produces selective deficits in verbal working memory. *Brain*, 129:306–320.
- Redgrave, P., Prescott, T., and Gurney, K. (1999). The basal ganglia: a vertebrate solution to the selection problem? *Neuroscience*, 89(4):1009–1023.
- Reich, P. A. (1977). Evidence for a stratal boundary from slips of the tongue. *Forum Linguisticum*, 2:211–217.

- Rhodes, B. J., Bullock, D., Verwey, W. B., Averbeck, B. B., and Page, M. P. (2004). Learning and production of movement sequences: behavioral, neurophysiological, and modeling perspectives. *Human Movement Science*, 23(5):699–746.
- 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., Mathiak, K., Wildgruber, D., Erb, M., Hertrich, I., Grodd, W., and Ackermann, H. (2005). fMRI reveals two distinct cerebral networks subserving speech motor control. *Neurology*, 64:700–706.
- 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.
- Riva, D. (1998). The cerebellar contribution to language and sequential functions: evidence from a child with cerebellitis. *Cortex*, 34(2):279–287.
- Rizzolatti, G. and Arbib, M. (1998). Language within our grasp. *Trends in Neurosciences*, 21(5):188–194.
- Robinson, S. E. and Vrba, J. (1999). Functional neuroimaging by synthetic aperture magnetometry (SAM). In Yoshimoto, T., Kotani, M., Kuriki, S., and Karibe, H., editors, *Recent Advances in Biomagnetism*, pages 302–305. Tohoku University Press, Sendai.
- Robles, S. G., Gatignol, P., Capelle, L., Mitchell, M. C., and Duffau, H. (2005). The role of dominant striatum in language: a study using intraoperative electrical stimulations. *Journal of Neurology, Neurosurgery, and Psychiatry*, 76:940–946.
- Roelofs, A. (1997). The WEAVER model of word-form encoding in speech production. *Cognition*, 64:249–284.
- Roy, C. and Sherrington, C. (1890). On the regulation of the blood supply of the brain. *Journal of Physiology*, 11:85–108.
- Rugg, M. D. and Coles, M. G. H., editors (1997). *Electrophysiology of Mind: Event-related brain potentials and cognition*, volume 25. Oxford University Press, Oxford.

- Saarinen, T., Laaksonen, H., Parviainen, T., and Salmelin, R. (2006). Motor cortex dynamics in visuomotor production of speech and non-speech mouth movements. *Cerebral Cortex*, 16(2):212–222.
- Salmelin, R., Hari, R., Lounasmaa, O. V., and Sams, M. (1994). Dynamics of brain activation during picture naming. *Nature*, 368:463–465.
- Salmelin, R. and Sams, M. (2002). Motor cortex involvement during verbal versus non-verbal lip and tongue movements. *Human Brain Mapping*, 16(2):81–91.
- Salmelin, R., Schnitzler, A., Schmitz, F., and Freund, H. J. (2000). Single word reading in developmental stutterers. *Brain*, 123:1184–1202.
- Santiago, J., MacKay, D. G., Palma, A., and Rho, C. (2000). Sequential activation processes in producing words and syllables: evidence from picture naming. *Language and Cognitive Processes*, 15(1):1–44.
- Sapir, E. (1949). The psychological reality of the phoneme. In Mandelbaum, D. G., editor, *Selected writings of Edward Sapir in Language, Culture and Personality*. University of California Press, Berkeley, CA.
- Schack, B. and Klimesch, W. (2002). Frequency characteristics of evoked and oscillatory electroencephalographic activity in a human memory scanning task. *Neuroscience Letters*, 331:107–110.
- Schafer, E. W. P. (1967). Cortical activity preceding speech: semantic specificity. *Nature (London)*, 216:1338–1339.
- Schaltenbrand, G. (1975). The effects on speech and language of stereotactical stimulation in thalamus and corpus callosum. *Brain and Language*, 2:70–77.
- Schmahmann, J. D., Doyon, J., McDonald, D., Holmes, C., Lavoie, K., Hurwitz, A. S., Kabani, N., Toga, A., Evans, A., and Petrides, M. (1999). Three-dimensional MRI atlas of the human cerebellum in proportional stereotaxic space. *NeuroImage*, 10(3):233–260.
- Schmithorst, V. J. and Holland, S. K. (2004). Event-related fMRI technique for auditory processing with hemodynamics unrelated to acoustic gradient noise. *Magnetic Resonance in Medicine*, 51(2):399–402.
- Schönle, P. W., Hong, G., Benecke, R., and Conrad, B. (1986). Aspects of speech motor control: programming of repetitive versus non-repetitive speech. *Neuroscience Letters*, 63:170–174.

- Sekihara, K., Nagarajan, S. S., Poeppel, D., Marantz, A., and Miyashita, Y. (2001). Reconstructing spatio-temporal activities of neural sources using an MEG vector beamformer technique. *IEEE Transactions on Biomedical Engineering*, 48:760–771.
- Sekihara, K., Nagarajan, S. S., Poeppel, D., Marantz, A., and Miyashita, Y. (2002). Application of an MEG eigenspace beamformer to reconstructing spatio-temporal activities of neural sources. *Human Brain Mapping*, 15(4):199–215.
- Sevald, C. A., Dell, G. S., and Cole, J. S. (1995). Syllable structure in speech production: Are syllables chunks or schemas? *Journal of Memory and Language*, 34:807–820.
- Shattuck-Hufnagel, S. (1979). Speech errors as evidence for a serial order mechanism in sentence production. In Walker, E., editor, *Sentence Processing: Psycholinguistic Studies Presented to Merrill Garrett*, pages 295–342. Erlbaum, Hillsdale, NJ.
- Shattuck-Hufnagel, S. (1983). Sublexical units and suprasegmental structure in speech production planning. In MacNeilage, P., editor, *The Production of Speech*, pages 109–136. Springer-Verlag, New York.
- Shattuck-Hufnagel, S. (1987). The role of word-onset consonants in speech production planning: New evidence from speech error patterns. In Keller, E. and Gopnik, M., editors, *Motor and Sensory Processes of Language*, pages 17–51. Lawrence Erlbaum, Hillsdale, NJ.
- Shattuck-Hufnagel, S. and Klatt, D. (1979). The limited use of distinctive features and markedness in speech production: evidence from speech error data. *Journal of Verbal Learning and Verbal Behavior*, 18:41–55.
- Shibasaki, H., Sadato, N., Lyshkow, H., Yonekura, Y., Honda, M., and Nagamine, T. (1993). Both primary motor cortex and supplementary motor area play an important role in complex finger movement. *Brain*, 116:1387–1398.
- Shima, K., Mushiake, H., Saito, N., and Tanji, J. (1996). Role for cells in the pre-supplementary motor area in updating motor plans. *Proceedings of the National Academy of Sciences*, 93(16):8694–8.
- Shima, K. and Tanji, J. (1998). Both supplementary and presupplementary motor areas are crucial for the temporal organization of multiple movements. *Journal of Neurophysiology*, 80(6):3247–60.
- Shima, K. and Tanji, J. (2000). Neuronal activity in the supplementary and presupplementary motor areas for temporal organization of multiple movements. *Journal of Neurophysiology*, 84(4):2148–60.

- Shuren, J. (1993). Insula and aphasia. *Journal of Neurology*, 240:216–218.
- Shuster, L. I. and Lemieux, S. K. (2005). An fMRI investigation of covertly and overtly produced mono- and multisyllabic words. *Brain and Language*, 93(1):20–31.
- 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.
- Sörös, P., Cornelissen, K., Laine, M., and Salmelin, R. (2003). Naming actions and objects: cortical dynamics in healthy adults and in anomic patient with a dissociation in action/object naming. *NeuroImage*, 19:1787–1801.
- Stark, R. E. (1980). Stages of speech development in the first year of life. In Yeni-Komshian, G. H., Kavanagh, J. F., and Ferguson, C. A., editors, *Child Phonology, Volume 1: Production*, pages 73–92. Academic Press, New York.
- Stemberger, J. (1991). Radical underspecification in language production. *Phonology*, 8:73–112.
- Sternberg, S., Monsell, S., Knoll, R. L., and Wright, C. E. (1978). The latency and duration of rapid movement sequences: Comparisons of speech and typewriting. In Stelmach, G., editor, *Information processing in motor control and learning*, pages 117–152. Academic Press, New York.
- Stevens, K. N. (1998). *Acoustic phonetics*. MIT Press, Cambridge, MA.
- Szirtes, J. and Vaughan, H. G. (1977). Characteristics of cranial and facial potentials associated with speech production. *Electroencephalography and Clinical Neurophysiology*, 43:386–396.
- Tanji, J. (1996). New concepts of the supplementary motor area. *Current Opinion in Neurobiology*, 6(6):782–787.
- Tanji, J. (2001). Sequential organization of multiple movements: involvement of cortical motor areas. *Annual Review of Neuroscience*, 24:631–51.
- Tanji, J. and Shima, K. (1994). Role for supplementary motor area cells in planning several movements ahead. *Nature*, 371(6496):413–6.
- Tanji, K., Suzuki, K., Yamadoir, A., Tabuchi, M., Endo, K., Fuji, T., and Itoyama, Y. (2001). Pure anarthria with predominantly sequencing errors in phoneme articulation. *Cortex*, 37(5):671–678.

- Tarkiainen, A., Helenius, P., Hansen, P. C., Cornelissen, P. L., and Salmelin, R. (1999). Dynamics of letter string perception in the human occipitotemporal cortex. *Brain*, 122:2119–2131.
- Tepper, J. M., Koos, T., and Wilson, C. J. (2004). GABAergic microcircuits in the neostriatum. *Trends in Neurosciences*, 27(11):662–669.
- Tiihonen, J., Kajola, M., and Hari, R. (1989). Magnetic mu rhythm in man. *Neuroscience*, 32:793–800.
- Tourville, J. A. and Guenther, F. H. (2003). A cortical and cerebellar parcellation system for speech studies. Technical Report CAS/CNS-03-022, Boston University, Boston, MA.
- Townsend, J. (1974). Issues and models concerning the processing of a finite number of inputs. In Kantowitz, B., editor, *Tutorials in performance and cognition*, pages 133–185. Erlbaum, Hillsdale, N.J.
- Treiman, R. and Danis, C. (1988). Short-term memory errors for spoken syllables are affected by the linguistic structure of the syllables. *Journal of Experimental Psychology. Learning, memory, and cognition*, 14(1):145–152.
- 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):765–780.
- Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., and Joliot, M. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage*, 15(1):273–289.
- Vallar, G., DiBetta, A. M., and Silveri, M. C. (1997). The phonological short-term store-rehearsal system: patterns of impairment and neural correlates. *Neuropsychologia*, 35(6):795–812.
- van Boxtel, A. (2001). Optimal signal bandwidth for the recording of surface EMG activity of facial, jaw, and neck muscles. *Psychophysiology*, 38:22–34.
- Van Buren, J. M. (1963). Confusion and disturbance of speech from stimulation in vicinity of the head of the caudate nucleus. *Journal of Neurosurgery*, 20:148–157.
- Van der Merwe, A. (1997). A theoretical framework for the characterization of pathological speech sensorimotor control. In McNeil, M., editor, *Clinical Management of Sensorimotor Speech Disorders*, pages 1–25. Thieme, New York.

- van Veen, B. D., van Dronglen, W., Yuchtman, M., and Suzuki, A. (1997). Localization of brain electrical activity via linearly constrained minimum variance spatial filtering. *IEEE Transactions on Biomedical Engineering*, 44:867–880.
- Vigário, R., Jousmäki, V., Hämäläinen, M., Hari, R., and Oja, E. (1998). Independent component analysis for identification of artifacts in magnetoencephalographic recordings. In *Advances in Neural Information Processing Systems*, volume 10, pages 229–235. MIT Press.
- Vogt, O. and Vogt, C. (1919). Ergebnisse unserer Hirnforschung. *Journal of Psychology and Neurology*, 25:277–462.
- von der Malsburg, C. (1973). Self-organization of orientation sensitive cells in the striate cortex. *Kybernetik*, 14(2):85–100.
- Vorobiev, V., Govoni, P., Rizzolatti, G., Matelli, M., and Luppino, G. (1998). Parcellation of human mesial area 6: cytoarchitectonic evidence for three separate areas. *European Journal of Neuroscience*, 10(6):2199–2203.
- Vousden, J. I., Brown, D. A., and Harley, T. A. (2000). Serial control of phonology in speech production: a hierarchical model. *Cognitive psychology*, 41:101–175.
- Vrba, J. and Robinson, S. E. (2001). Signal processing in magnetoencephalography. *Methods*, 25:249–271.
- Wang, D. L., Liu, X. M., and Ahalt, S. C. (1996). On temporal generalization of simple recurrent networks. *Neural Networks*, 9(7):1099–1118.
- Ward, N. (1994). *A connectionist language generator*. Ablex Publishing, Norwood, NJ.
- Warren, H. (1986). Slips of the tongue in very young children. *Journal of Psycholinguistic Research*, 15(4):309–344.
- Watkins, K. E., Dronkers, N. F., and Vargha-Khadem, F. (2002a). Behavioural analysis of an inherited speech and language disorder: comparison with acquired aphasia. *Brain*, 125(Pt 3):452–464.
- Watkins, K. E., Vargha-Khadem, F., Ashburner, J., Passingham, R. E., Connelly, A., Friston, K. J., Frakowiak, R. S. J., Mishkin, M., and Gadian, D. G. (2002b). MRI analysis of an inherited speech and language disorder: structural brain abnormalities. *Brain*, 125(Pt 3):465–478.
- Watson, P. and Montgomery, E. B. J. (2006). The relationship of neuronal activity within the sensori-motor region of the subthalamic nucleus to speech. *Brain and Language*, 97(2):233–240.

- Werker, J. F. and Pegg, J. E. (1992). Infant speech perception and phonological acquisition. In Ferguson, C. A., Menn, L., and Stoel-Gammon, C., editors, *Phonological acquisition*, pages 285–311. York Press, Timonium, MD.
- Wertz, R. T., LaPointe, L. L., and Rosenbek, J. C. (1984). *Apraxia of Speech in Adults: The Disorder and its Management*. Grune and Stratton Inc., Orlando.
- Wheeler, D. W. and Touretzky, D. S. (1997). A parallel licensing model of normal slips and phonemic paraphasias. *Brain and Language*, 59(1):147–201.
- Wickelgren, W. A. (1969). Context-sensitive coding, associative memory, and serial order in (speech) behavior. *Psychological Review*, 76:1–15.
- Wiesendanger, R. and Wiesendanger, M. (1985). The thalamic connections with medial area 6 (supplementary motor cortex) in the monkey (*Macaca fascicularis*). *Experimental Brain Research*, 59:91–104.
- 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., Ackermann, H., Klose, U., Kardatzki, B., and Grodd, W. (1996). Functional lateralization of speech production at primary motor cortex: a fMRI study. *Neuroreport*, 7(15-17):2791–2795.
- Wilson, C. J. (1993). The generation of natural firing patterns in neostriatal neurons. In Arbuthnott, G. W. and Emson, P. C., editors, *Progress in brain research: chemical signaling in the basal ganglia*, pages 277–298. Elsevier, Amsterdam.
- Wilson, C. J. (1995). The contribution of cortical neurons to the firing patterns of striatal spiny neurons. In Houk, J., Davis, J., and Beiser, D., editors, *Models of information processing in the basal ganglia*, pages 29–50. MIT Press, Cambridge, MA.
- Wise, R. J., Greene, J., Buchel, C., and Scott, S. K. (1999). Brain regions involved in articulation. *Lancet*, 353(9158):1057–1061.
- Ziegler, W. (2002). Psycholinguistic and motor theories of apraxia of speech. *Seminars in Speech and Language*, 23(4):231–243.
- 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.

- Ziegler, W. and Maassen, B. (2004). The role of the syllable in disorders of spoken language production. In Maassen, B., Kent, R., Peters, H., Lieshout, P. V., and Hulstijn, W., editors, *Speech motor control in normal and disordered speech*, pages 415–445. Oxford University Press, Oxford.
- Zilles, K., Schlaug, G., Geyer, S., Luppino, G., Matelli, M., Qu, M., Schleicher, A., and Schormann, T. (1996). Anatomy and transmitter receptors of the supplementary motor areas in the human and nonhuman primate brain. In Luders, H. O., editor, *Supplementary Sensorimotor Area*, pages 29–43. Lippincott-Raven, Philadelphia.
- Zimmerman, J. E., Thiene, P., and Harding, J. T. (1970). Design and operation of stable RF-biased superconducting point-contact quantum devices and a note on the properties of perfectly clean metal contacts. *Journal of Applied Physics*, 41:1572–1580.
- Zimmerman, R. and Scharein, E. (2004). MEG and EEG show different sensitivity to myogenic artifacts. *Neurology and Clinical Neurophysiology*, 78.

CURRICULUM VITAE

Jason W. Bohland

21 Bay St.
Cambridge, MA 02139
(617) 821-8319
e-mail: jbohland@cns.bu.edu
web: <http://cns.bu.edu/~jbohland>

EDUCATION

2000 – Present	Ph.D., Cognitive and Neural Systems Boston University, Boston, MA
1998 – 2000	M.S., Electrical Engineering (Honors) University of Cincinnati, Cincinnati, OH
1994 – 1998	B.S., Computer Engineering (cum laude) University of Cincinnati, Cincinnati, OH

EXPERIENCE

2001 – Present	Research Assistant, Boston University Speech Lab
2002 – 2003	Teaching Assistant, Boston University
1998 – 2000	Teaching Assistant / Instructor, University of Cincinnati
1999 – 2000	Partner, InsideCorner.com, LLC, Cincinnati, OH
1996 – 1997	Product Development Engineer Co-op, Structural Dynamics Research Corporation, Milford, OH

AWARDS

2004	Student Travel Award, Organization for Human Brain Mapping
2000	Presidential University Graduate Fellowship, Boston University
2000	Special Student Service Award, University of Cincinnati
1994 – 1998	University Honors Scholarship, University of Cincinnati

PUBLICATIONS AND PRESENTATIONS

1. Bohland, J. W. and Guenther, F. H. (2006). An fMRI investigation of syllable sequence production. *NeuroImage*, 32 (2): 821-841.
2. Bohland, J.W., Guenther, F.H., and Bullock, D. (2006). Modeling and imaging of sequencing in speech production. *10th International Conference on Cognitive and Neural Systems*, Boston, MA. [oral presentation]
3. Bohland, J.W. (2006). Neural substrates for syllable sequence planning and production. *Massachusetts Institute of Technology Speech Communication Group Seminar Series*. [invited oral presentation].
4. Bohland, J.W. (2006). An fMRI investigation of syllable sequence production. *Massachusetts Institute of Technology Neurolinguistics Group Seminar*. [invited oral presentation].
5. 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*, Austin, TX.
6. Ghosh, S.S. and Bohland, J.W. (2005). A speech recording setup for fMRI with online reduction of scanner noise. *NeuroImage (11th Annual Meeting of the Organization for Human Brain Mapping, Toronto)*, 26 (S1): S42. [poster].
7. 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. [poster].
8. 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 (75th Meeting of the Acoustical Society of America, San Diego)*, 116 (4): 2631. [poster].
9. Bohland, J.W. and Guenther, F.H. (2004). An fMRI investigation of the neural bases of sequential organization for speech production. *NeuroImage (10th Annual Meeting of the Organization for Human Brain Mapping, Budapest, Hungary)*, 22 (S1): S41. [poster].
10. 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 Annual Meeting of the Organization for Human Brain Mapping, New York)*, 19 (2): S57. [poster].

11. Guenther, F.H., Tourville, J.A., and Bohland, J.W. (2003). Modeling the representation of speech sounds in auditory cortical areas. *Program of the 1435th Meeting of the Acoustical Society of America, Journal of the Acoustical Society of America*, 113 (4) Pt. 2: 2210. [invited oral presentation].
12. Guenther, F.H., and Bohland, J.W. (2002). Learning sound categories: A neural model and supporting experiments. *Acoustical Science and Technology*, 23(4): 213-220. Japanese-language version appeared in *Journal of the Acoustical Society of Japan*, 58 (7): 441-449, 2002.
13. Bohland, J.W. and Minai, A.A. (2001). Efficient associative memory using small-world architecture. *Neurocomputing*, 38-40: 489-496.
14. Bohland, J.W. and Minai, A.A. (2000). Small-World model of associative memory. *Proceedings of the International Joint Conference on Neural Networks*, 5: 597-604

MEMBERSHIPS AND SERVICE

Student Member, Organizaion for Human Brain Mapping

Student Member, Institute of Electrical and Electronics Engineers (IEEE)

Ad hoc reviewer for *NeuroImage* and *IEEE Transactions on Neural Networks*