BOSTON UNIVERSITY

GRADUATE SCHOOL OF ARTS AND SCIENCES

Dissertation

NEURAL DYNAMICS OF SPEECH PERCEPTION AND PRODUCTION: FROM SPEAKER NORMALIZATION TO APRAXIA OF SPEECH

by

HEATHER AMES

B.A., University of California, Berkeley 2003

Submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

2009

ACKNOWLEDGEMENTS

The work described in this dissertation would not have been possible without the guidance and help from Steve Grossberg and Frank Guenther. Both Steve and Frank were invaluable resources as I progressed through the stages of graduate school. I would also like to thank Ennio Mingolla who was the first professor I met when I arrived at CNS and has always been there to offer advice.

In addition, support from CELEST helped me to pursue the work in this dissertation as well as many opportunities above and beyond my research. Serving CELEST as the founder of the student and post doc leadership group was a great experience that I will always treasure.

Many other individuals have been helpful over the years including Gail Carpenter and the whole tech lab, the speech lab, Jason Tourville, Anatoli Gorchetchnikov, Sai Gaddam, Arun Ravindran, Greg Amis, and Cindy Bradford. I would also like to thank the Jamaica Plain VA hospital and Carole Palumbo, Rupal Patel at Northeastern, and Edwin Maas for their assistance with the apraxia project.

I would also like to thank my family and friends for standing by my decision to be a perpetual student. Finally, I would like to thank Max for giving me the inspiration and confidence needed to complete this dissertation by offering me his never-ending love, support, and encouragement.

NEURAL DYNAMICS OF SPEECH PERCEPTION AND PRODUCTION: FROM SPEAKER NORMALIZATION TO APRAXIA OF SPEECH

(Order No.)

HEATHER AMES

Boston University Graduate School of Arts and Sciences, 2009

Major Professor: Stephen Grossberg, Wang Professor of Cognitive and Neural Systems

ABSTRACT

This dissertation seeks to enhance understanding of speech mechanisms by employing computational modeling in two key areas: understanding how the brain builds speaker-independent representations of heard speech sounds and why apraxic speakers are unable to effectively generate speech motor programs.

The first portion of the dissertation introduces the Neural Normalization Network model (NormNet) that has been developed to explain how the human brain is able to convert speaker-dependent acoustic information into speaker-independent language representations. NormNet is part of an emerging model of auditory streaming and speech categorization. Multiple strip representations and asymmetric competitive circuits are both used in the auditory streaming and speaker normalization parts of the model, thereby suggesting that these two circuits arose from similar neural designs.

NormNet is able to explain and accomplish speaker normalization by generating pitch-independent representations of speech sounds while preserving information about

speaker identity. The speaker-independent representations are categorized into unitized speech items, which input to sequential working memories whose distributed patterns can be rapidly categorized into syllable and word representations and stably remembered by Adaptive Resonance Theory circuits. Model simulations use synthesized steady-state vowels from the Peterson and Barney (1952) database. The model achieves accuracy rates similar to those achieved in human listeners.

The second portion of the dissertation investigates how brain lesions in patients with apraxia of speech (AOS) give rise to different behavioral characteristics. AOS is a disorder of the planning and/or programming of speech production without comprehension impairment and without weakness in the speech musculature. The DIVA model (Directions into Velocities of Articulators) and the GODIVA model (Gradient Order DIVA) provide a framework for theorizing about two possible subtypes of AOS. The first subtype is hypothesized to arise from damage to the inferior frontal sulcus region (IFS). This damage would result in fluent productions of erroneous or misplaced speech sounds. The second subtype is hypothesized to arise from damage to the frontal operculum region (FO). This damage would result in poorly articulated approximations of the desired syllables. These hypotheses are tested by investigating damage scenarios in DIVA and GODIVA. The results are compared to an apraxic patient case study.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ABA-2	Apraxia Battery for Adults, second edition
AI	Core area in monkey auditory cortex
AL	Lateral belt area in monkey auditory cortex
AMR	Alternating Motion Rate
ANOVA	ANalysis Of VAriance
AOS	Apraxia Of Speech (acquired)
ART	Adaptive Resonance Theory
ARTSPEECH	Adaptive Resonance Theory based SPEECH perception network
ARTSTREAM	Adaptive Resonance Theory based STREAMing network
ARTWORD	Adaptive Resonance Theory based WORD perception network
ASR	Automatic Speech Recognition
BA	Brodmann's Area
BA 6	Brodmann's Area 6 (premotor cortex)
BA 44	Brodmann's Area 44 (pars opercularis)
BA 45	Brodmann's Area 45 (pars triangularis)
BA 47	Brodmann's Area 47 (inferior prefrontal gyrus)
BNT	Boston Naming Test
BDAE	Boston Diagnostic Aphasic Examination
BVMT	Brief Visuospatial Memory Test
CAS	Childhood Apraxia of Speech
CBD	CorticoBasal Degeneration

CCCV	Consonant Consonant Consonant Vowel syllable
CELEX	Centre for LEXical information
CT	Computed Tomography
CV	Consonant Vowel syllable structure
CVC	Consonant Vowel Consonant syllable structure
CVCVCV	Consonant Vowel Consonant Vowel Consonant Vowel syllable
CVI	Cerebral Vascular Incident
dB	Decibel
DISC	Distinct Single Characters
DIVA	Directions Into Velocities of Articulators
ERB	Equivalent Rectangular Bandwidth
F0	Fundamental frequency
F1	First harmonic frequency
F2	Second harmonic frequency
F3	Third harmonic frequency
F5	Monkey equivalent of the inferior frontal gyrus
fMRI	functional Magnetic Resonance Imaging
FO	Frontal Operculum
Fuzzy ARTMAP	Fuzzy Adaptive Resonance Theory Map
GB	Gigabyte
GDS	Geriatric Depression Scale
GHz	Gigahertz

GODIVA	Gradient Order Directions Into Velocities of Articulators
HGARC	Harold Goodglass Aphasia Research Center
hVd	‘h’ Vowel ‘d’ syllable
HVLT	Hopkins Verbal Learning Test
Hz	Hertz
IFG	Inferior Frontal Gyrus
IFS	Inferior Frontal Sulcus
IPA	International Phonetic Alphabet
IRN	Iterated Rippled Noise
kHz	Kilohertz
LOFT	Lexical Orthographic Familiarity Test
MCA	Middle Cerebral Artery
ML	Lateral belt area in money auditory cortex
MRI	Magnetic Resonance Imaging
msec	Millisecond
NIST	Standard switchboard corpus
NormNet	Neural Normalization Network
PET	Positron Emission Tomography
PNFA	Progressive Non-Fluent Aphasia
R	Core area in monkey auditory cortex
RAM	Random-Access Memory
SMA	Supplementary Motor Area

SMR	Sequential Motion Rate
SPINET	Spatial PItch NETwork
SSM	Speech Sound Map
TAAWF	Test of Adolescent and Adult Word Finding
TONI-3	Test Of Nonverbal Intelligence, third edition
TMS	Transcranial Magnetic Stimulation
VA	Veteran's Affairs
VC	Vowel Consonant syllable
VOT	Voice Onset Time
VTLN	Vocal Tract Length Normalization
WMS-III	Wechsler Memory Scale, third edition
WMS-R	Wechsler Memory Scale, revised

CHAPTER 1

INTRODUCTION

How does the brain learn to generate speaker invariant representations of speech sounds from speaker-dependent acoustic information? How is this information used to stably learn speech sound categories and to chunk those sounds into larger language representations such as syllables and words? What brain mechanisms are responsible for sequencing and selection of the motor speech programs associated with the intended speech utterances? How does brain damage affect these processes? This dissertation begins to address these questions in an attempt to investigate common organizational properties and neural designs shared by both speech perception and production.

In order to better understand the mechanisms at work in two key areas of speech research – the ability of the human brain to build a speaker-invariant representation of heard speech sounds and the inability of apraxic patients to effectively generate motor speech programs – a computational modeling approach is employed. The two models presented in this paper are built upon a neuroanatomical foundation in which model mechanisms and components are associated with brain areas based on functional imaging studies in humans and neurophysiological experiments in non-human primates. In order to validate the models, comparisons are made to previously reported psychophysical studies, and a new apraxic case study is presented to investigate model predictions.

1.1 Building speaker-invariant representations

Auditory signals of speech are speaker-dependent, but representations of language meaning are speaker-independent. The transformation from speaker-dependent to speaker-independent language representations enables speech to be learned and understood from different speakers. In order to better understand how the human brain may accomplish this task, a speaker normalization model called the Neural Normalization Network or NormNet (Ames and Grossberg, 2008), has been developed.

NormNet transforms speaker-dependent information to speaker-independent language representations within the framework of an emerging model of auditory streaming and speech categorization. The speaker-invariant representations are categorized into unitized speech items, which input to sequential working memories whose distributed patterns can be categorized, or chunked, into syllable and word representations. The speaker normalization circuit does not require the storage of instances or exemplars of speech sounds and can learn a single categorical representation for each speech sound without the use of a teaching signal and with only local knowledge available to the system.

The auditory streaming and speaker normalization parts of the model both use multiple strip representations and asymmetric competitive circuits, thereby suggesting that these two circuits arose from similar neural designs. The normalized speech items are rapidly categorized and stably remembered by Adaptive Resonance Theory (ART; Grossberg, 1976a, 1976b, 1978, 1980) circuits. ART-based categorization can, where

task demands require, learn both specific, concrete categories, even individual exemplars, as well as general, abstract categories. This learning process enables ART models to selectively pay attention to the learned prototype of critical feature patterns that predict successful performance.

Simulations use synthesized steady-state vowels from the Peterson and Barney (1952) vowel database and achieve accuracy rates similar to those achieved by human listeners. Without normalization the model's performance in vowel categorization decreased. The addition of speaker invariance also creates more stable categories and better performance in speech categorization and understanding. These results are compared to behavioral data and other speaker normalization models.

1.2 Apraxia of speech impairs speech motor control

Apraxia of speech is a disorder of the planning and/or programming of speech production without comprehension impairment and without weakness in the speech musculature. Apraxia of speech (AOS) most often occurs following brain damage resulting from stroke, traumatic brain injury, or neurodegenerative disease (Kent, 2000; Maassen, 2002; Josephs et al., 2006). AOS can be distinguished from the dysarthias, which are motor speech disorders resulting from abnormalities of neuromuscular functioning, and from aphasias, which are language disorders. AOS affects primarily articulation and prosody and is diagnosed by the presence of a variety of speech errors using a battery of tests. Current diagnostic techniques are sparse and often involve co-diagnosis with aphasia. However, understanding exactly how the brain is damaged in apraxic patients can give rise to a better understanding of speech planning and

programming as well as the development of more effective treatment techniques for apraxia of speech.

The DIVA model (Directions into Velocities of Articulators; Guenther *et al.*, 2006; Guenther, 1994, 1995, 2006) makes many predictions about the organization and the structure of motor speech programs within its Speech Sound Map (SSM). The GODIVA model (Gradient Order DIVA; Bohland *et al.* in press) creates an additional module for DIVA to understand how speech sounds in the SSM are selected and sequenced for speech production. Modeling efforts making use of the DIVA and GODIVA framework have led to the hypothesis of two sub-types of AOS, one associated with SSM damage and the other associated with damage to the subsystem projecting to the SSM which is responsible for sequencing and selecting the speech sounds. Damage to the frontal operculum (FO) and ventral portion of the inferior frontal gyrus (IFG) corresponds to SSM damage and will most likely result in poorly articulated approximations of the desired syllables (referred to as AOS type 2 herein). The other AOS subtype involves damage to the posterior region of the inferior frontal sulcus (IFS) and dorsal regions of the IFG and will lead to fluent productions of the wrong speech sounds or misplacement of speech sounds within the motor program (referred to as AOS type 1 herein). This hypothesis is tested with computer simulations of apraxic damage and a comparison is made between the resulting model behaviors and human data.

Finally, a case study of an apraxic patient is presented. This patient suffered a stroke which resulted in damage to the IFG region. The lesion, however, spared the IFS

and the more superior portions of the IFG. Consistent with the predictions made by DIVA, this patient exhibits AOS type 2 errors. Comparisons of this patient's speech characteristics and the performance of the GODVIA and DIVA models are presented.

1.3 Organization of dissertation

The remainder of this dissertation is organized into three chapters. Chapter 2 will present the NormNet model, which includes a detailed description of how the model fits into a larger speech perception modeling framework, and the simulations performed to validate the model. Chapter 3 will discuss apraxia of speech. This chapter will include a description of behavioral symptoms, lesion data, frontal brain regions involved in speech articulation, and a summary of the functional interpretations of apraxia of speech. Chapter 4 will describe the modeling work done in the context of GODIVA and DIVA which was used to generate a hypothesis that there are two subtypes of apraxia of speech. This chapter will also describe an apraxic case study which was used to test the model hypothesis. Finally, Chapter 5 will summarize the research presented in this dissertation and suggest future work aimed at further developing a comprehensive understanding of speech perception and production.

CHAPTER 2

SPEAKER NORMALIZATION USING CORTICAL STRIP MAPS: A NEURAL MODEL FOR STEADY-STATE VOWEL CATEGORIZATION

2.1 Introduction: Speech learning, normalization, and imitation

Fundamental variations in speech exist both between speakers and within the speech of a single speaker. Intra-speaker variability is mainly concerned with the different pronunciations of the same phoneme by a single speaker. These variances can result from differences in phonemic context, including coarticulation effects, accent, and the emotions or stress level of the speaker. Inter-speaker variability concerns the variation of speech across speakers and these variations generally have a much larger effect on perception (Nearey, 1989). Despite this variability, a listener is able to identify and understand speech spoken by different speakers on the first encounter with a speaker and on nearly the first utterance. It seems that, not only does the listener's brain learn to store a speaker-invariant representation of speech, but that somehow the speech encountered is transformed, or *speaker normalized*, into a speaker-invariant representation for the purpose of understanding.

The process of speaker normalization enables a baby to begin to imitate sounds from adult speakers, notably parents whose spoken frequencies differ significantly from those that the baby can babble. A circular reaction from endogenously babbled to heard sounds enables a baby to learn a map between the auditory representations of its own heard babbled sounds to the motor commands that caused them (Piaget, 1963; Grossberg, 1978; Cohen *et al.,* 1988; Bullock *et al.,* 1993; Guenther, 1995; Guenther *et*

al., 2006). Speaker normalization enables sounds from adult caretakers to be filtered by this learned map and to thereby enable the baby to begin to imitate and refine heard sounds in its own language productions. Learning in such an imitative map needs to remain active for many years in order to enable an individual's changing voice through puberty and adulthood to continue to activate and update this map.

Speaker normalization also enables language meanings that were learned from one teacher's voice to be readily understood when uttered by another speaker. More generally, speaker normalization helps the brain to overcome a combinatorial explosion that would otherwise occur if the brain needed to store every instance of every speaker utterance in order to understand language meaning.

A similar problem of combinatorial explosion is overcome by the visual cortex as it learns to recognize visually perceived objects in the world. In vision, object category representations are learned that are relatively insensitive to object size, location, and orientation on the retina (Bradski and Grossberg, 1995; Ito *et al.*, 1995). Such invariance is built up across several processing stages, with invariance only appearing in the inferotemporal cortex and beyond. Likewise, speaker normalization is just one stage in the development of rate- and speaker-independent representations of language meaning. For example, Boardman *et al.* (1999) and Grossberg *et al.* (1997) have modeled how rate-invariance may develop across several processing stages.

Although speaker normalization and rate-invariance are important for learning to speak and understand language, humans and other animals are also exquisitely sensitive to the voice quality and prosody of individual speakers. A similar dichotomy

occurs during visual learning and recognition, where positionally-invariant object recognition categories coexist with cortical representations that enable manipulation of objects in space. In both audition and vision, this is accomplished through interactions across *what* and *where* cortical processing streams (Ungerleider and Mishkin, 1982; Goodale and Milner, 1992; Hickok and Poeppel, 2007; Fazl *et al.*, 2008). Such interactions have been predicted to compute computationally *complementary* properties (Grossberg, 2000): the properties needed to compute one such property prevent a complementary property from being computed in the same cortical processing stream, and conversely. Interactions between cortical processing streams that compute such complementary properties enable them to overcome their complementary deficiencies and thus to generate adaptive and creative behaviors. Thus, the present article's focus on speaker normalization is in no way contradicted by the fact that the brain can process many individual features of speaker identity and prosody. In fact, one property of the present speaker normalization model is that, while it generates a speaker-independent representation, it also marks the speaker-dependent frequencies of the selected speaker from which speaker-dependent properties can be computed in a complementary processing stream.

Another important issue concerns the specificity or generality of speaker-independent language categories. In certain environments, it is essential to distinguish fine differences between speaker utterances that can change language meaning across speakers. In other environments, considerable variability across utterances can lead to a similar meaning. Our model clarifies how such variability can naturally emerge during

incremental learning of individual language exemplars. Thus, the fact that humans and animals can sometimes distinguish individual speech exemplars does not imply that language learning is exemplar learning, with all the problems of combinatorial explosion and biologically implausible exhaustive search that such a model can create. Recognition categories can, instead, be learned in a way that tracks the task demands and statistics of each unique environment and that seems to conjointly maximize category generality and minimize predictive error. Neurophysiological data in the visual cortex support this viewpoint (Spitzer *et al*., 1988; Zoccolan *et al*., 2007) in a manner predicted by neural models (Carpenter and Grossberg, 1987, 1991, 2003; Carpenter, 1997; Grossberg, 1999; Fazl *et al*., 2008). Learned category prototypes in such models can represent either individual exemplars or abstract knowledge, as each learning situation uniquely demands. Vowel category simulations in the present article illustrate this property.

Speaker normalization is also an important technique used in engineering for building automatic speech recognition (ASR) systems. Vowel classification rates in ASR systems can be improved if features are speaker-normalized before classification (Nearey, 1989). The formant ratio theory is a foundation for many speaker normalization techniques. The formant ratio theory states that vowel quality depends on the log frequency intervals between formants (defined as ratios), and that shifting activations along a log frequency axis will generate the invariant representation (Lloyd, 1890a, 1890b, 1891, 1892; Peterson, 1961; Bladon *et al.*, 1984; Sussman, 1986; Syrdal

and Gopal, 1986; Miller, 1989; Sussman *et al.*, 1997). Formant ratios can be calculated by averaging across formant values for many utterances of a single speaker.

Despite its heuristic appeal, in its classical formulation, formant ratio theory faces two types of problems. First, no mechanism has been proposed to explain how the human auditory system could perform these calculations. How does the brain align cell activities corresponding to the formant frequencies for each utterance? Second, this method requires information contained in many speech samples of a single speaker and thus is not able to account for our ability to understand a speaker in the first utterance that we encounter. It is not biologically feasible for the brain to perform computations across all speech samples it encounters from a speaker, store this information, and then use it to normalize each new utterance encountered for each speaker.

The inability of ASR systems to understand speech in real situations and environments may be due to their lack of adherence to biological auditory principles. As Dusan and Rabiner (2005) pointed out, perhaps it is now time to take a closer look at how the brain performs speech recognition and apply these insights to design novel ASR systems. The modeling work presented in this chapter proposes a new method for speaker normalization that makes use of the functional architecture of the brain and builds upon previous modeling work that explains a large amount of data in acoustics, speech perception, and language.

The well-documented existence of tonotopic organization in the auditory cortex, which gives rise to *strips* of frequency selective cells, serves as the functional architecture within which the speaker normalization transformation is proposed to

occur; see Section 2.2. Frequency-selective strips provide more representational space within which finer computations can occur. They share some properties with the hypercolumn organization that is ubiquitous in the visual cortex (Hubel and Wiesel, 1962). In vision, hypercolumns occur in cortical maps that represent multiple features of visual objects in physical space. In audition, they occur in cortical maps that represent multiple features of acoustic objects in frequency space.

Such strip maps have earlier been shown capable of explaining key data about auditory streaming, or the separation of acoustic sources (Grossberg *et al*., 2004). Speaker normalization and streaming circuits may have arisen from similar underlying neural designs. In particular, the streaming model circuit and the speaker normalization circuit described herein use both strip maps and asymmetric competition across frequency-selective channels to realize a kind of *exclusive allocation*. During auditory streaming, exclusive allocation enables spectral information to be allocated to a specific source or stream (Bregman, 1990). Here we predict that the exclusive allocation properties that are familiar in streaming research also play a role in generating a speaker-independent representation of speech. Strip maps have also been used to explain other cortical processes, such as how place-value number systems may be learned (Grossberg and Repin, 2003). Strip maps may thus be a cortical design that has been specialized during brain evolution to accomplish multiple tasks.

As explained below, a simple transformation from speaker-dependent to speaker-independent speech information can be performed within strip maps in a way that is consistent with neurobiological data. The speaker-independent representations

are then categorized via a process of fast incremental learning. Results from synthesized steady-state vowel categorization simulations are presented to validate the performance of the speaker normalization model. The results have been reported in Ames and Grossberg (2006, 2007, 2008).

2.2 Tonotopic organization and multiple strip maps

The auditory system contains spatially organized maps of frequency selective cells called *tonotopic maps*. The frequency representations are arranged logarithmically. Tonotopy is preserved in the auditory system from the level of the cochlea to the auditory cortex of humans and other mammals (Tunturi, 1952; Merzenich and Brugge, 1973; Imig *et al.*, 1977; Reale and Imig, 1980; Romani *et al.*, 1982; Seldon, 1985; Luethke *et al.*, 1988; Pantev *et al.*, 1988; Morel and Kaas, 1992; Morel *et al.*, 1993; Heil *et al.*, 1994; Rauschecker *et al.*, 1995; Bilecen *et al.*, 1998; Wessinger *et al.*, 1998; Lockwood *et al.*, 1999; Talavage *et al.*, 2000, 2004; Rauschecker and Tian, 2004). In the auditory cortex, these tonotopic maps consist of iso-frequency contours which can be defined as *strips* of cortical cells that respond to a specific frequency, or best frequency.

In addition to spectral information that is explicitly in acoustic inputs, missing fundamental frequencies (F0) of harmonic sounds activate the tonotopic maps of the primary auditory cortex of mammals. Single-unit extracellular recordings in marmosets have shown that complex tones with missing fundamentals activate tonotopic areas corresponding to the missing fundamental (Bendor and Wang, 2005, 2006). These maps were found in the low frequency-selective areas on the border of core areas AI and R

and the lateral belt areas AL and ML, but did not extend into the entire tonotopic representation of any of these areas. Fishman *et al.* (1998) found an implicit representation of the missing fundamental in AI based on population neuronal responses in awake macaque monkeys. Missing fundamental activations have also been seen in auditory cortical areas of gerbils (Schulze *et al.*, 2002) and cats (Whitfield, 1980; Qin *et al.*, 2005).

In humans, fMRI has been used to show that the lateral Heschl's gyrus is sensitive to the F0 differences of iterated rippled noise (IRN) when subjects listened to noise with temporally varying patterns (Patterson *et al.*, 2002). Penagos *et al.* (2004) confirmed the existence of this F0-selective region by using fMRI to show that missing fundamental complex tones containing only low frequency harmonics causes a stronger activation in this region than if the tones contained only high frequency harmonics. This difference is attributed to the unresolvability of the high frequencies for listeners. Langner *et al.* (1997) found that a topographically ordered F0 map in human auditory cortex (where F0 was described as the periodicity of the complex sound) may be found orthogonally to the topographically ordered spectral map.

These data confirm that cells in auditory cortex respond selectively to frequencies and F0 in a spatially organized manner, but the exact placement of an F0-sensitive map with respect to a spectral map is unclear. For the purpose of our speaker normalization model, it is assumed that the F0-sensitive map may lie near or within the spectrally activated maps. Simulations were performed to manipulate the amount of energy at the F0 filters in order to test the role of F0 in the speaker normalization

transformation. The advantages and disadvantages of using F0 for speaker normalization are discussed below.

Our speaker normalization model builds upon the fact that multiple tonotopic maps of frequency-selective strips are found in the auditory cortex of both humans and other mammals (Merzenich and Brugge, 1973; Imig *et al.*, 1977; Morel and Kaas, 1992; Morel *et al.*, 1993; Hacket *et al.*, 1998; Kaas and Hackett, 1998; 2000; Formisano *et al.*, 2003; Rauschecker and Tian, 2004; Petkov *et al.*, 2006). Interactions between such maps are core design features of our speaker normalization model. Map boundaries are defined by frequency reversals such that the low frequency endpoint of one map is adjacent to the low frequency endpoint of the next map. The same occurs for the high frequency endpoints. Talavage *et al.* (2004) used fMRI and frequency-swept stimuli to identify six tonotopic mappings in the superior temporal plane, suggesting that there are at least five areas in the human auditory cortex that exhibit at least six tonotopic organizations. However, the number of maps that exist in the human brain is still uncertain and is difficult to determine with the resolution available in imaging technologies.

The speaker normalization model presented in this chapter assumes that at least two of these tonotopic strip maps have an orthogonal, or at least non-parallel, spatial arrangement. The overlapping interactions between these two spectral maps allow the spectral information from different speakers to be aligned along a diagonal map. This diagonal map arrangement underlies the computations needed to shift the speaker-dependent speech information into a speaker-independent representation.

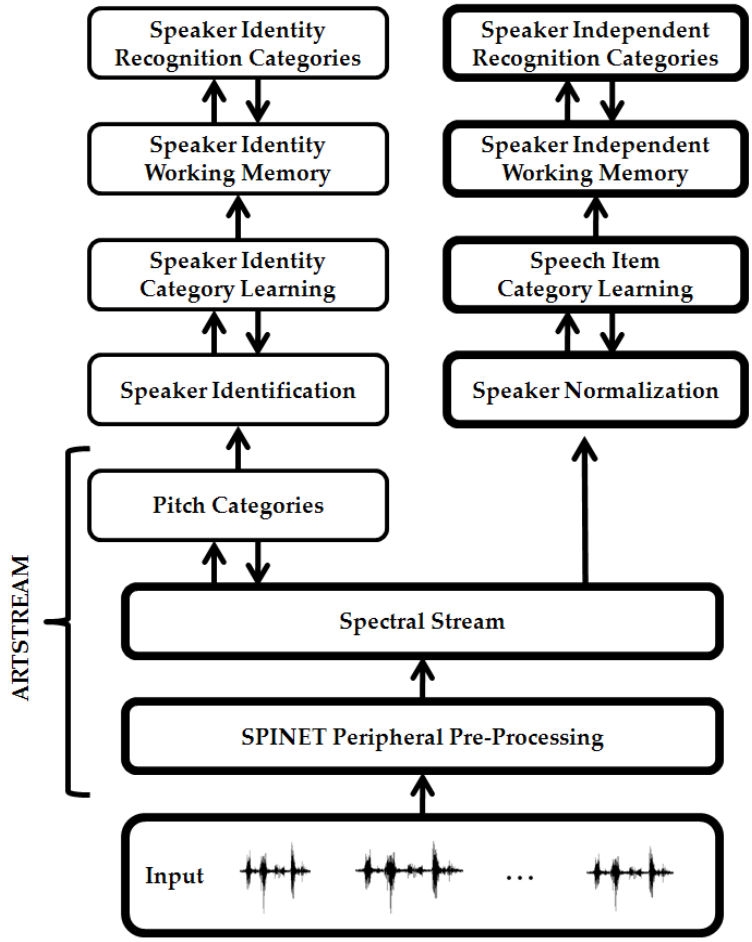


Figure 2.1: *Box diagram of the ARTSPEECH perception system.* The boldface boxes contain components discussed or simulated in this paper.

2.3 An emerging audition, speech, and language model

The model presented in this paper called the Neural Normalization Network, or NormNet for short. NormNet is part of an architecture for speech perception and recognition that is being developed by Grossberg and colleagues (Cohen *et al.*, 1995; Cohen and Grossberg, 1997; Grossberg *et al.*, 1997; Boardman *et al.*, 1999; Grossberg and Myers, 2000; Grossberg, 2003b; Grossberg *et al.*, 2004); see Figure 2.1. At the

periphery of this architecture, a Spatial PItch NETwork (SPINET) processes acoustic information and converts the temporally-occurring auditory signals into spatial representations of pitch (Cohen *et al.*, 1995; see Figure 2.2). Harmonically-related spectral components (see stages 6 and 7 in Figure 2.2) can activate a given pitch category through an adaptive filter. The selection of harmonics is due to learning that is driven by the natural grouping of frequencies in early auditory processing. SPINET hereby creates both spatial representations of pitch and harmonically related spatial activations. This mapping is a crucial feature for the proposed speaker normalization technique.

The SPINET model provides a natural front end for a more comprehensive model of pitch-based auditory streaming that is called the ARTSTREAM model (Grossberg *et al.*, 2004; see Figure 2.3). Both the spectral and pitch representations in SPINET are defined by strips of frequency and pitch. The frequency strips in the spectral maps are selective for a particular frequency and are ordered on a log frequency axis. These frequency selective strips are a key organizational structure in ARTSTREAM that allows the model to parse acoustical information into distinct auditory streams that intersect the strips at an orthogonal, or at least non-parallel, angle.

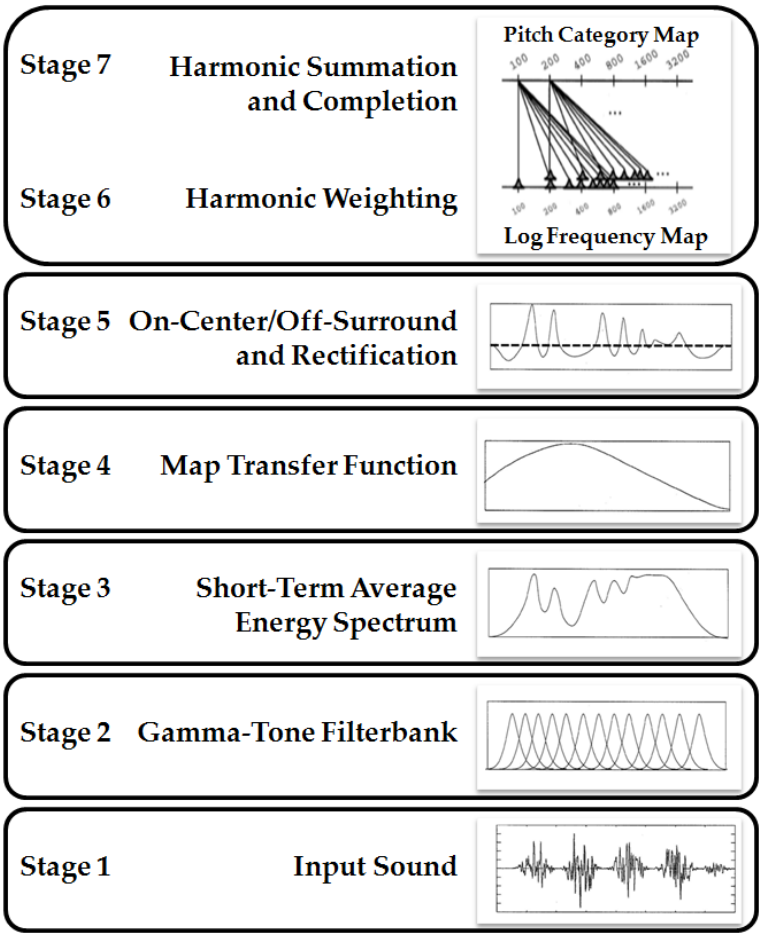


Figure 2.2: *SPINET Model.* The processing stages transform a sound stream into activations of spatially distributed pitch nodes. [Reprinted with permission from Cohen *et al.*, 1995]

ARTSTREAM derives its name from Adaptive Resonance Theory (ART; Grossberg, 1976a, 1976b, 1978, 1980). ART principles and mechanisms have been used to explain and predict data about visual and auditory perception, category learning and object recognition, cognitive information processing, cognitive and emotional interactions, and their underlying brain mechanisms (Carpenter and Grossberg, 1991; Grossberg, 1994, 1999, 2003a; Grossberg and Merrill, 1996; Chey *et al.*, 1997; Grossberg *et al.*, 1997;

Grunewald and Grossberg, 1998; Grossberg and Williamson, 1999; Vitevitch and Luce, 1999; Page, 2000; Grossberg and Myers, 2000; Bowers, 2002; Goldinger and Azuma, 2003; Hawkins, 2003; Fazl *et al.*, 2008; Grossberg and Versace, 2008). ART claims that resonant states between top-down expectations and bottom-up input drive stable learning of perceptual and cognitive representations, while preventing catastrophic forgetting of previously learned information.

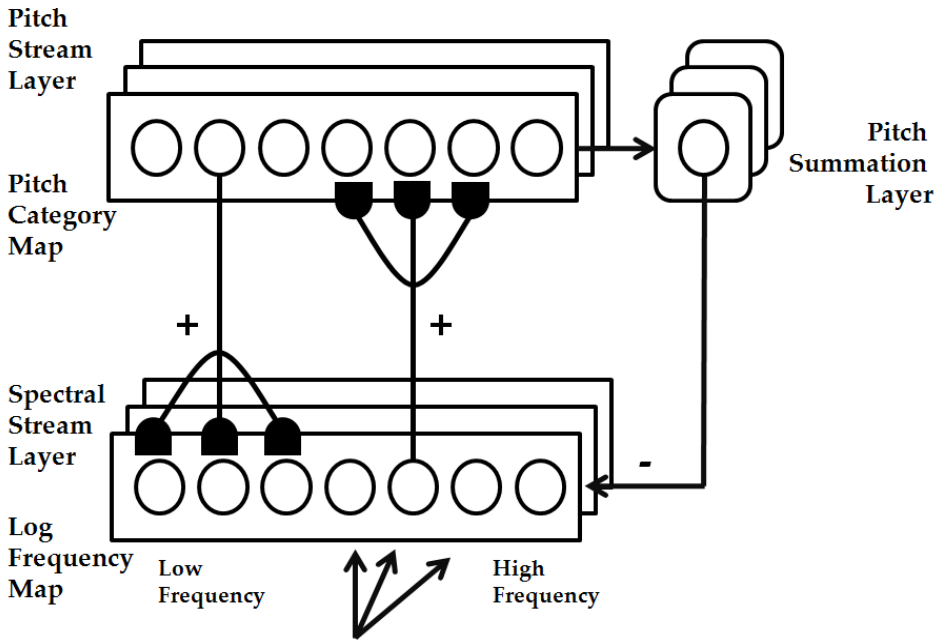


Figure 2.3: *ARTSTREAM Model*. The spectral and pitch layers of the SPINET model (layers 6 and 7) are elaborated in the ARTSTREAM model into multiple representations, or strips of cells, and top-down ART matching also occurs. Bottom-up signals group harmonically-related spectral components into activations of pitch categories. Inhibition within each pitch stream enables only one pitch category to be active at any time in a given stream. Asymmetric inhibition across streams in the pitch stream layer is biased so that the winning pitch cannot be represented in another stream. The winning pitch category feeds back excitation to its harmonics in the corresponding spectral stream. This stream also receives nonspecific top-down inhibition from the pitch layer. ART matching is hereby realized. It suppresses those spectral components that are not harmonically related to the active pitch. Inhibition across spectral streams then prevents the resonating frequency from being represented in other streams as well. [Reprinted with permission from Grossberg (2003b)]

In the domain of audition, speech perception, and language, models based on ART mechanisms have been used to explain, in addition to auditory streaming, word recognition and recall (Grossberg and Stone, 1986), manner distinctions in consonant perception (Boardman *et al.*, 1999), consonant integration and segregation in VC-CV syllables (Grossberg *et al.*, 1997), and interword integration and duration-dependent backward effects (Grossberg and Myers, 2000). These models mechanistically embody such design principles as storage in working memory of temporal order information derived from phonemic representations, automatic gain control to maintain rate invariance, and top-down matching of learned expectations with bottom-up patterns of information in order to focus attention on expected combinations of acoustic features and to stabilize fast auditory learning. See Grossberg (2003b) for a review.

ARTSTREAM includes a bottom-up adaptive filter, or "harmonic sieve," that groups together harmonics of an auditory source into learned pitch categories. In addition, a top-down filter encodes the expectations of the learned pitch categories. Each expectation consists of the harmonics of the learned pitch category, which competitively inhibit other frequencies. Both psychological and neurobiological data support the existence of such *biased competition* in the selection of attended data (Grossberg, 1980; Desimone, 1998; Grossberg, 1999; Kastner and Ungerleider, 2001; Grossberg, 2003a). An auditory stream forms when a bottom-up adaptive filter and its top-down expectation interact to generate a spectral-pitch resonant state. Through such resonant dynamics, ARTSTREAM is able to coherently select pitch-consistent frequencies, corresponding to both F0-related harmonics and formant frequencies, while

suppressing other frequencies. This ARTSTREAM process along with asymmetric competition across streams realizes the property of exclusive allocation (Bregman, 1990). Although ARTSTREAM has discussed only the special case of how pitch categories may be learned from spectral information and thereby used to separate distinct acoustic sources, the same ART mechanisms can learn to categorize other speaker-specific properties that can be used for speaker identification.

The spectral information in the selected stream can be the input to the speaker normalization model, since the spectral-pitch resonances, or other speaker-specific resonances, isolate different speaker's sounds from one another; see Figure 2.1. Moreover, the spatially organized frequency-selective strips of ARTSTREAM provide the computational substrate that is needed to initiate speaker normalization. The key design principle of frequency-selective strip maps allows these models to seamlessly connect and interact. Isolated vowels are the inputs to NormNet in the current simulations, so the stream-separating mechanisms are neither needed nor simulated.

After speaker normalization is accomplished and the invariant vowels are categorized by an ART network, the speaker-independent vowel categories are in a form that can naturally input to the ARTWORD model of variable-rate speech categorization and word recognition (Grossberg and Myers, 2000).

2.4 Model description

Peripheral processing in the NormNet model is based on the SPINET model (Figure 2.2) of Cohen *et al.*, (1995) with a few modifications. The gammatone filterbank (see stage 2 in Figure 2.2) consists of a cascade of fourth order gammatone filters (Holdsworth *et al.*, 1988; Patterson *et al.*, 1988; Cohen *et al.*, 1995):

$$GT(f) = [1 + j(f - f_i)/b(f_i)]^{-4} \quad . \qquad (2.1)$$

The center frequencies (f_i) of the filters range from 10 to 8000 Hz and are equally spaced in equivalent rectangular bandwidth (ERB) units (Patterson and Rice, 1987; Patterson *et al.*, 1987; Patterson *et al.*, 1988; Holdsworth *et al.*, 1988; Slaney, 1993, 1998). The dynamic range corresponds to data measuring the dynamic range in human listeners (Hudspeth, 2000; Plack *et al.*, 2005). The ERB of a filter at the center frequency (f_i) is a function of the filter center frequency (Glasberg and Moore, 1990):

$$ERB(f_i) = 24.7 + 0.108 * f_i \quad , \qquad (2.2)$$

and the bandwidth $b(f_i)$ of a filter is defined by:

$$b(f_i) = \frac{ERB(f_i)}{0.982} \quad . \qquad (2.3)$$

The output signal from the filterbank is then mapped onto a logarithmic scale, half-wave rectified, and low-pass filtered. This signal serves as the input to the speaker normalization model.

The speaker normalization transformation is proposed to occur in auditory cortex by using at least two intersecting tonotopic strip maps that are assumed, for simplicity, to align orthogonally. The names of these maps are the anchor log frequency

map (*anchor map*) and the stream log frequency map (*stream map*); see Figure 2.4. Because both maps are composed of strips of frequency-selective units, the activations in these maps spread along the strips into an *interstrip area* where strips from both maps are superimposed upon each other. Both the anchor map and stream map receive spectral information from the speech sound. In the full architecture, this spectral information is predicted to be the streamed output from a process like ARTSTREAM. In the current simplified model, SPINET preprocessing generates the model's spectral input pattern.

Asymmetric competition occurs in the anchor map to choose the cell with the lowest active frequency in the speech sound, which typically contains the largest amount of spectral energy (see Figure 2.4a). This cell is called the *anchor frequency coding cell*. As the anchor frequency coding cell wins the asymmetric competition, it inhibits any activations corresponding to higher frequencies in the anchor map. This form of exclusive allocation is predicted to be a key step in speaker normalization. The asymmetric competition is governed by the following on-center, off-surround shunting equation (Grossberg, 1973, 1980); see Figure 2.4a:

$$\frac{dx_{i0}}{dt} = -Ax_{i0} + (B - x_{i0})[I_{i0} + f(x_{i0})] - x_{i0}\sum_{i>k} f(x_{k0}) \quad , \qquad (2.4)$$

where x_{i0} is the activity of the i^{th} frequency-selective cell in the anchor map, and I_{i0} is the input to this cell in the anchor map. In equation (*2.4*), A = 0.1, B = 1, and $f(x) = x^2$. Since $f(x) = x^2$ is a faster-than-linearly increasing signal function, the activities of cells corresponding to the lowest frequency will increase as the activities of

the other cells decrease, resulting in contrast enhancement and winner-take-all choice of the cell whose activity corresponds to the lowest active frequency (Grossberg, 1973).

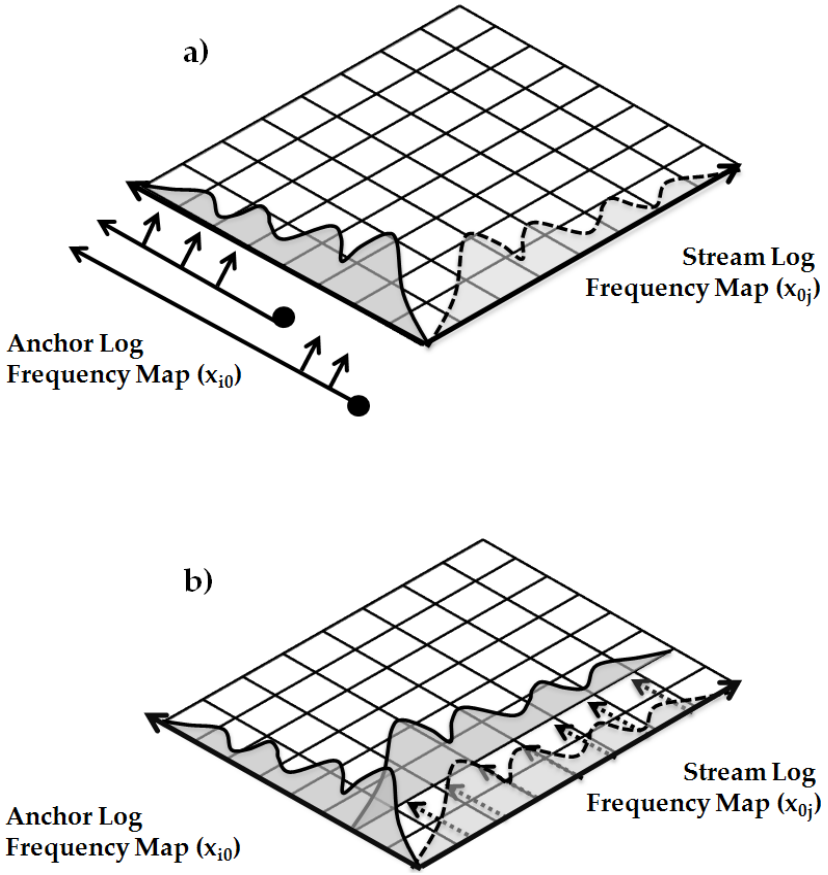


Figure 2.4: (a) *Anchor map and stream map*. These maps are organized orthogonally and superimpose on each other. Both maps receive spatially organized spectral information from the streamed sound. The activations spread along their corresponding strips into the inter-stream area. **(b)** *Coincidence detection*. The winning anchor frequency coding cell triggers coincidence detection along its anchor frequency strip. This coincidence detection moves the activations of the stream map into the anchor frequency strip.

The cell that codes the anchor frequency triggers coincidence detection along its strip in the inter-strip area where both the anchor map and the stream map activate their corresponding frequency-selective strips. The coincidence occurs in the strip

corresponding to the anchor frequency of the anchor map (i^{th} row) and all the active strips corresponding to spectral activations in the stream map (j^{th} columns); see Figure 2.4b. The activity, x_{ij}, of the cell in the i^{th} row and the j^{th} column obey:

$$\frac{dx_{ij}}{dt} = -Ax_{ij} + g(x_{i0})I_j, \qquad (2.5)$$

where x_{i0} is the activity of the i^{th} strip in the anchor map, I_j is the spectral representation of the speech sound at the j^{th} strip of in the stream map, the decay rate A= 0.1, and the anchor map sigmoid signal function

$$g(x_{i0}) = \frac{x_{i0}^b}{c^b + x_{i0}^b}, \qquad (2.6)$$

where the choices b = 100 and c = 0.5 enable $g(x_{i0})$ to approximate 1 at the anchor frequency and 0 elsewhere. Due to coincidence detection $g(x_{i0})I_j$ in equation *(2.5)*, the stream map shifts into the anchor frequency strip and becomes the *anchored stream*.

Because the anchor map and the stream map have connections which superimpose orthogonally, their coincidences can create diagonally connected strips; see Figure 2.5. These diagonal connections transform the anchored stream into a speaker-invariant representation, *S*. In particular, each cell in the *S* field sums inputs from all the cells along a diagonal created by the maps. The activity, s_m, of the m^{th} diagonal map cell is thus

$$s_m = \sum_{j=1}^{n} x_{j-m,j}\ , \qquad (2.7)$$

where n is the number of filters in the gammatone filterbank and m is the cell number in the *S* field. The speaker-independent spectrum is then categorized into unitized item

representations. These learned recognition categories are used for vowel identification.

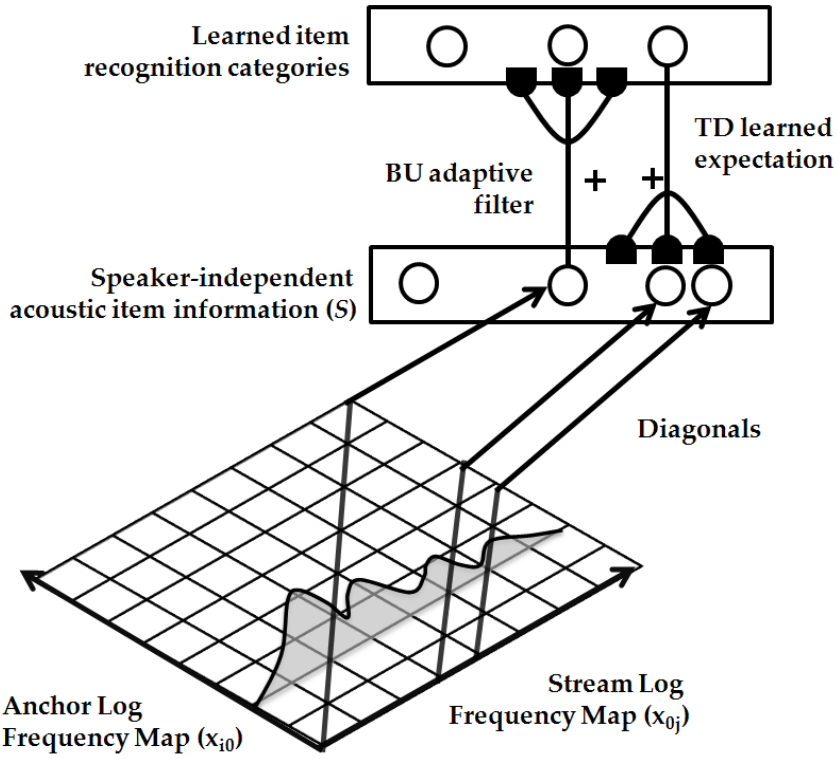


Figure 2.5: *Creation of speaker-independent working memory item information*. The diagonal strips are sampled to create the speaker-independent working memory item information which is then feed into an ART network which learns to categorize the item information.

In this paper, vowel categorization is carried out by a fuzzy ARTMAP network with default parameters (Carpenter *et al.*, 1992); see Figure 2.6. Fuzzy ARTMAP is a neural network that incorporates two fuzzy ART modules, ART_a and ART_b, where ART_a learns to map the speaker-independent vowel spectra to vowel categories. An intervening *map field*, F^{ab}, learns to associate the vowel categories to category names in ART_b. See Appendix A for the fuzzy ART equations.

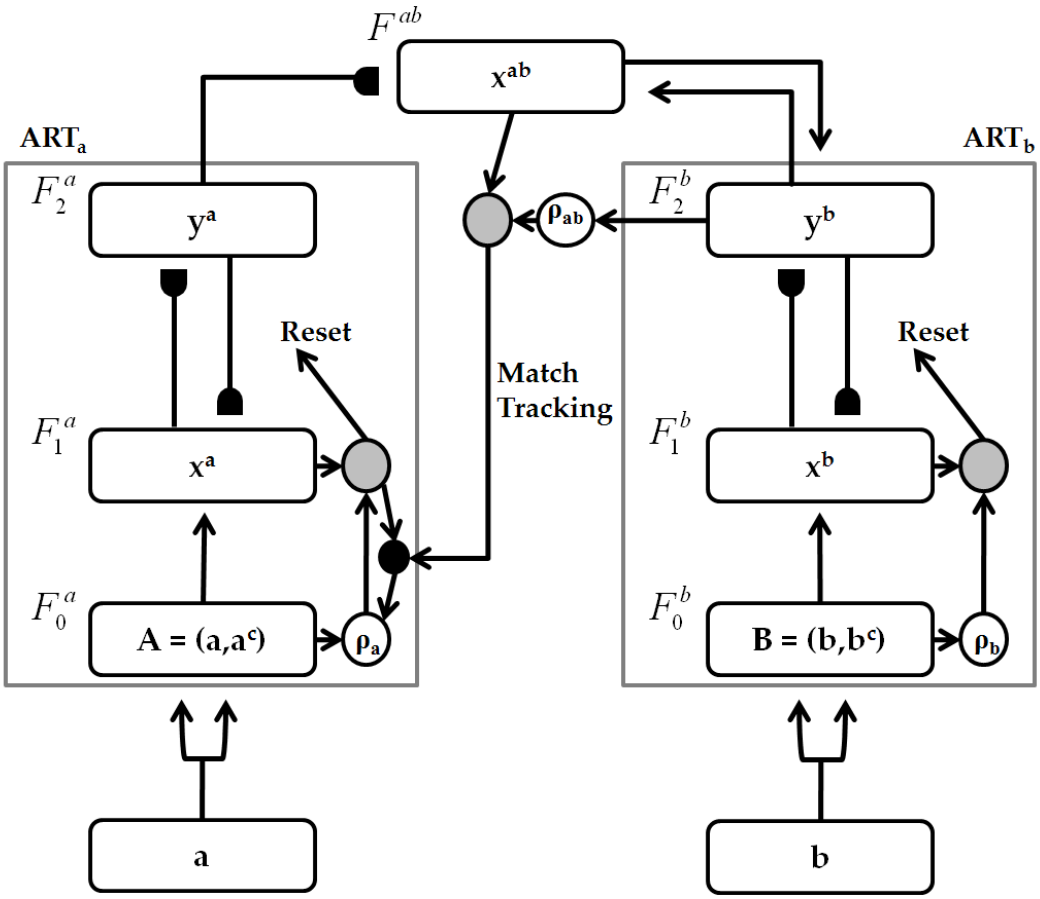


Figure 2.6: *Fuzzy ARTMAP*. The ART_a complement coding preprocessor transforms the M_a vector **a** into the $2M_a$ vector **A** = (**a**,**a**^{c}) at the ART_a field F_0^c. **A** is the input vector to the ART_a field F_1^c. Similarly, the input to F_1^b is the $2M_b$ vector **B** = (**b**,**b**^{c}). When a prediction by ART_a is disconfirmed at ART_b, inhibition of map field activation induces the match tracking process. Match tracking raises the ART_a vigilance (ρ_a) to just above the F_1^c to F_0^c match ratio $|\mathbf{x}^a|/|\mathbf{A}|$. This triggers an ART_a search which leads to activation of either an ART_a category that correctly predicts **b** or to previously uncommitted ART_a category node. [Reprinted with permission from Carpenter *et al.*, 1992]

When a vowel category is activated during category learning, it reads out a learned top-down expectation that is matched against the input vowel frequency spectrum. A *vigilance* parameter, ρ_a, determines whether the match is good enough. Learning occurs only if the match is good enough. Then a resonant state can develop that supports the

learning process. A predictive failure in category naming at ART_b increases ρ_a by the minimum amount needed to trigger a memory search. Such a memory search automatically leads to learning and/or selection of a new vowel category in ART_a that can better match the vowel frequency spectrum. This process is called *match tracking* (Carpenter *et al.*, 1992). It enables learning of the most general vowel categories that can minimize predictive errors in ART_b. Match tracking is realized in the F^{ab} map field.

The speaker-independent spectral input vectors **A** to the F_1^a field of ART_a are transformed into complement-coded vectors **A** = (**a**,**a**^{c}) before being further processed by ART_a. Complement coding means that both the activities **a** of the network's ON cells and the activities **a**^{c} = 1- **a** of its OFF cells form the input vector. The inputs **B** to F_1^b field of ART_b are complemented-coded representations of vowel names: **B** = (**b**,**b**^{c}). The values of all components in these input vectors lie between 0 and 1. The activity in the F_1^c field activates a vowel category node, *J*, in the F_2^c field which, in turn, sends top-down signals to the F_1^c field, where matching between the bottom-up input and the top-down weight vector, $\mathbf{w}_J^a$ of the expectation occurs. If the match is good enough, as determined by the vigilance criterion:

$$\rho_a |A| < |A \wedge w_J| \ . \qquad (2.8)$$

then learning occurs. Inequality (*1.8*) describes the balance between excitation and inhibition at a novelty-sensitive orienting system such as the nonspecific thalamus or hippocampus (Carpenter and Grossberg, 1993; Grossberg and Versace, 2008). The term $\rho_a |A|$ describes the total amount of excitation that reaches the orienting system. Term

$|A \wedge w_J|$ describes the total amount of inhibition that reaches the orienting system. In all, inequality (*2.8*) says that inhibition is greater than excitation, so that the orienting system does not fire. The matched patterns are thus allowed to resonate and learning is enabled. When excitation exceeds inhibition, the currently active category is reset and the network continues searching for a better category with which to encode the speaker-independent spectrum. The vigilance ρ_c value in equation (*2.8*) is a gain control term that determines the network's sensitivity to excitation from the bottom-up total input $|A|$. Increasing vigilance makes the network more sensitive to mismatches, and thus leads to finer categories. At very high vigilance, the network can learn individual exemplars. At low vigilance, it can learn abstract categories that enable many exemplars to be coded by the same recognition category.

During learning, the speaker-independent input pattern **A** is encoded by a vowel category in F_2^c, while the vowel name input pattern **B** is encoded by a name category *K* in F_2^t. In the present simulations, name category labels are directly input to F_2^t without loss of generality. The map field F^{ab} associates these categories unless *J* has previously learned to predict a different *K* category. If this occurs, then match tracking proceeds until an appropriate new ART_a category is chosen and learned. During testing, the speaker-independent input signal **A** activates a name category in ART_b through F^{ab}, which is the prediction of the system. Mathematical details about fuzzy ARTMAP are found in Appendix A.

2.5 Methods

2.5.1 Stimuli

In order to evaluate the performance of NormNet, the Peterson and Barney (1952) database (Peterson and Barney, 1952; Watrous, 1991) was chosen because it has been widely used as a benchmark database for studying vowel identification. Peterson and Barney originally tape recorded 76 speakers (33 males, 28 females, and 15 children) each speaking 10 vowels twice in a /h**V**d/ context, resulting in 1,520 tokens. The vowels used are found in Table 2.1. The recorded vowels were analyzed and the steady state measurements for F0, F1, F2, and F3 were preserved in the dataset. Listeners in this original study achieved 94% accuracy in recognition tasks when evaluating these vowels in /h**V**d/ context.

Number	**ARPAbet symbol**	**IPA Symbol**	**/hVd/**
1	IY	i	Heed
2	IH	ɪ	Hid
3	EH	ɛ	Head
4	AE	æ	Had
5	AH	ʌ	Hud
6	AA	ɑ	Hod
7	AO	ɔ	Hawed
8	UH	ʊ	Hood
9	UW	u	Who'd
10	ER	ɜ	Heard

Table 2.1. The ten vowels in the Peterson and Barney (1952) database.

Hillenbrand and Gayvert (1993) synthesized steady-state values corresponding to the values of the formants in the database in order to determine how well listeners can identify vowels based on static spectral cues. Seventeen listeners achieved 72.7%

accuracy for the synthesized vowels with flat F0 contours, which hold F0 constant for the duration of the sound stimulus. When F0 movement was added, performance only slightly improved to 74.8% correct. For the purposes of the simulations in this paper, the Hillenbrand and Gayvert (1993) performance will be used as a basis of comparison and the methods of these simulations will attempt to adhere to the methods presented in that paper.

2.5.2 Procedure

A vowel synthesizer (Slaney, 1998) was used to generate steady-state versions of all 1,520 tokens in the Peterson and Barney (1952) database. Formant frequencies and F0 were held constant for the full duration of the stimulus, similar to the synthesized vowels used by Hillenbrand and Gayvert (1993). The sampling frequency was set at 16 kHz and the formant bandwidth was set at 50 Hz. Waveforms for a sample synthesized vowel, IY for a man, woman, and child are shown in Figure 2.7.

In order to assess the performance of the system, several types of simulations were performed. Simulations were conducted by varying vowel lengths, the dynamic range and number of filters of the filterbank, the training set size, and spectral inputs with and without F0 information combined in the mappings.

In order to simulate the natural variances across human listeners, the dynamic range of the filterbank and the number of filters were varied for the simulations. The inputs were presented to the model in random order. The simulations were run on a workstation computer using a dual core AMD© Opteron Processor 246 with 1.99 GHz and 3.18 GB of RAM. Matlab v.7.1 and the auditory toolbox (Slaney, 1998) were used

to run the simulations. Statistical analysis was conducted using Statistics To Use (Kirkman, 1996) and Wessa.net (Wessa, 2007) software.

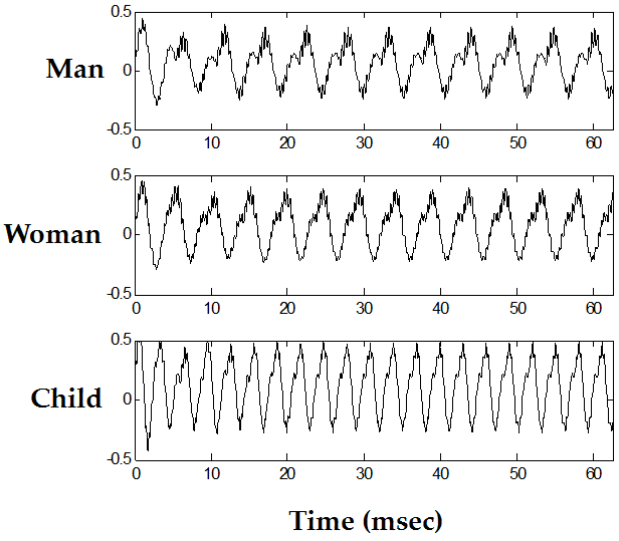


Figure 2.7: *Waveforms for synthesized steady-state vowel IY.* The top plot corresponds to a male, the middle to a female, and the bottom to a child.

2.6 Results

2.6.1 Number of filters for the filterbank

Simulations were performed by varying the number of filters (100, 150, 200, 250, 300, 350, 400) while keeping the filter range constant at 50-7500 Hz, 400 tokens in the training set, and three runs for each filterbank size. These simulations were performed both with and without adding F0 information to the input-activated spectral information in the anchor map. The results from these simulations are illustrated in Figure 2.8. Interestingly, adding F0 caused model performance to deteriorate by approximately 5%. An analysis of variance (ANOVA) did not show a significant effect for filterbank size ($F[6,14]=0.338$, $p < 0.91$).

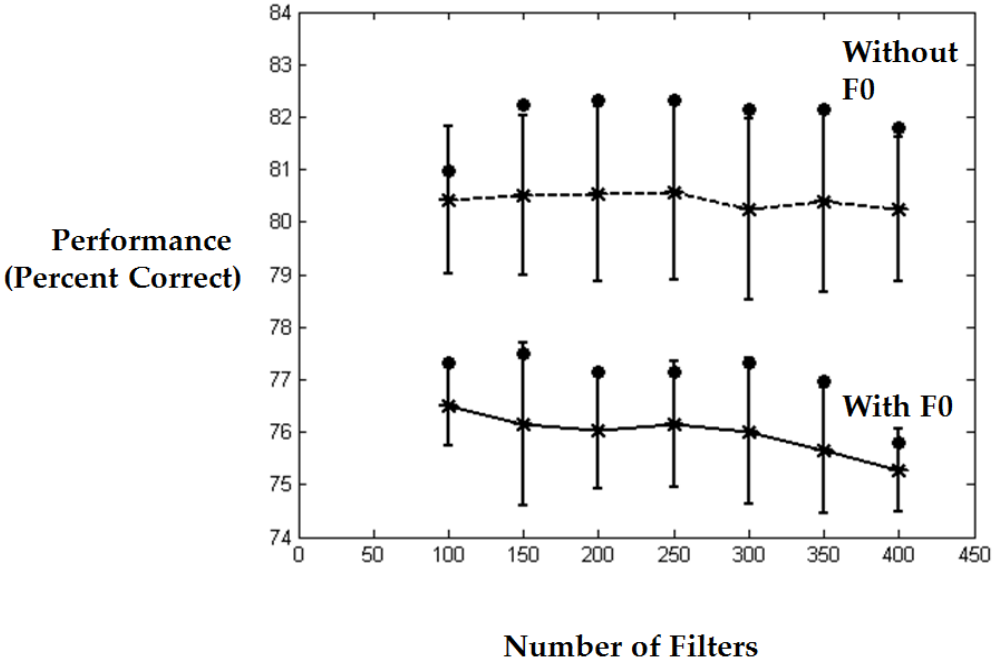


Figure 2.8: *Filterbank size.* The filterbank size was varied from 100-400 filters. The dataset was tested both with and without F0 information added to the spectral information of the Anchor Map. Three runs were performed with each filterbank size. The dynamic range was from 50-7500 Hz and the training set size was set to 400 tokens. The dashed line shows the results without F0 information and the solid line shows the results with F0 information in the spectral map. The * indicates the mean plus error bars and the • indicates the best performance for that filterbank size.

2.6.2 Dynamic range of filterbank

The dynamic range of the filterbank was tested; see Figure 2.9. The low frequency was varied from 20 to 180 Hz while the high end was held constant at 7500 Hz. The results of these simulations are illustrated in Figure 2.9a.

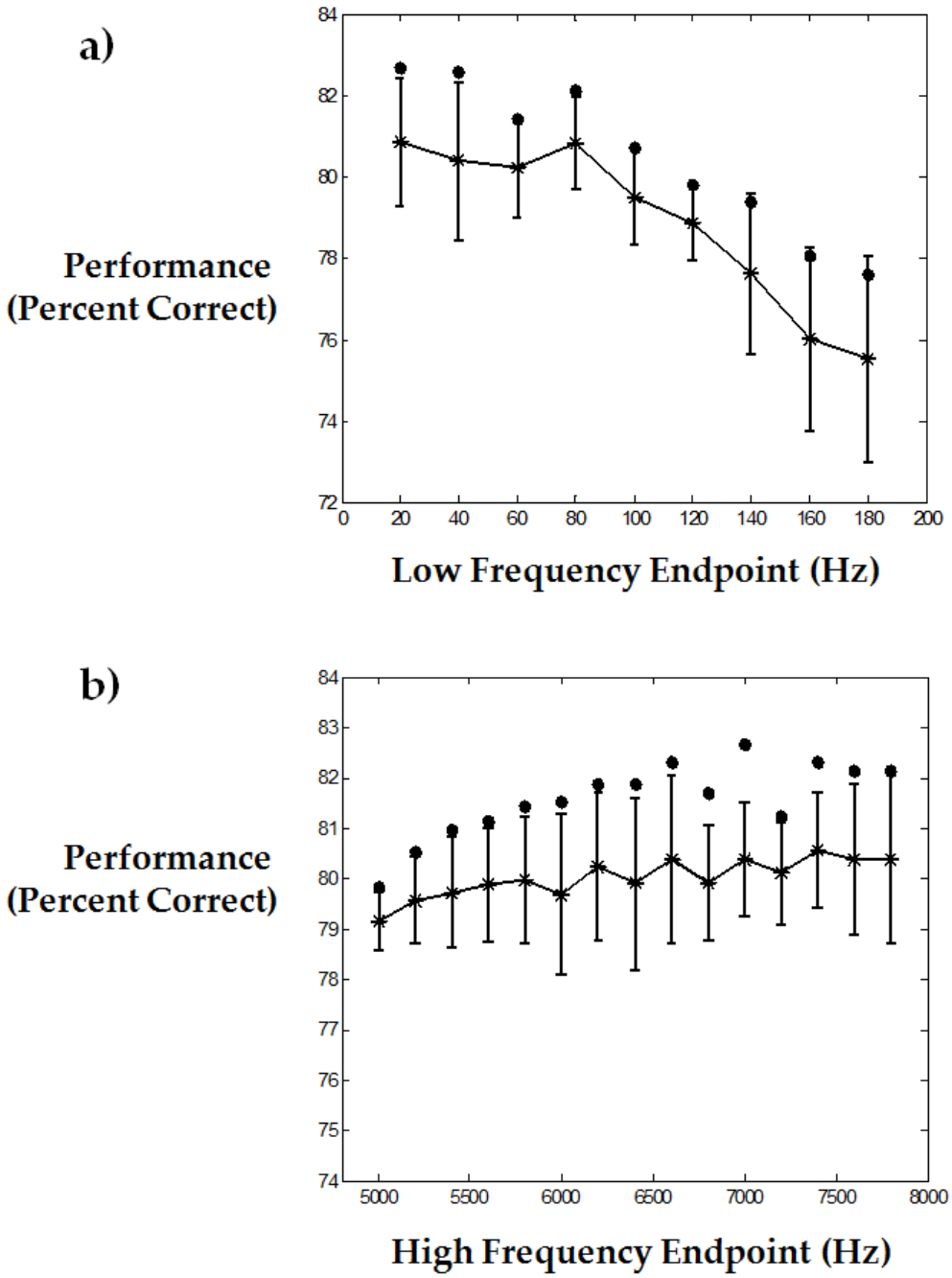


Figure 2.9: *Dynamic range of filterbank.* 250 filters were used in the filterbank, the training set contained 400 tokens, and three runs were performed for each variation. No F0 information was added to the spectrum for these simulations. The * indicates the mean plus error bars and the • indicates the best performance for that variation in the filterbank. **(a)** The low frequency endpoint was varied from 20 to 180 Hz while the high frequency endpoint was held constant at 7500 Hz. **(b)** The high frequency endpoint was varied from 5 to 8 kHz while the low frequency endpoint was held constant at 50 Hz.

These simulations were performed without adding F0 information, with 250 filters, and with a training set size of 400 vowel tokens. The performance of the system was best when the low frequency end was below 100 Hz (approximately 80% correct on average). These data were found to be well fit by a linear model with a negative slope (for the low frequency value: mean performance R^2 = 0.87 and best performance R^2 = 0.94). Performance deteriorated gradually as the lowest frequency was increased from 100 Hz because some of the lower frequency vowels in the databank contain frequency information below 100 Hz.

The high frequency was also varied from 5 to 8 kHz while the low end was held constant at 50 Hz. The results of these simulations are found in Figure 2.9b. The high frequency manipulation caused less of an effect. When these data were fit to a linear model (for the high frequency value: mean performance R^2 = 0.722 and best performance R^2 = 0.629), the slope was nearly zero (mean performance = 0.00036 and best performance = 0.00067) indicating that there is little change in performance across the different high frequency endpoint values. This is because the high frequency range takes into account information above the F3 values and the vowels used in these simulations were synthesized with only F0-F3 information. Generally, the effect of manipulation of the higher formants of the vowel (F3-F5) has a much a smaller effect on vowel identification (Johnson, 2005; Slawson, 1968; Nearey, 1989) and thus we expect that it would have less effect on the model even if we were to test items that contained such information. This prediction has yet to be tested, however.

2.6.3 Training set size

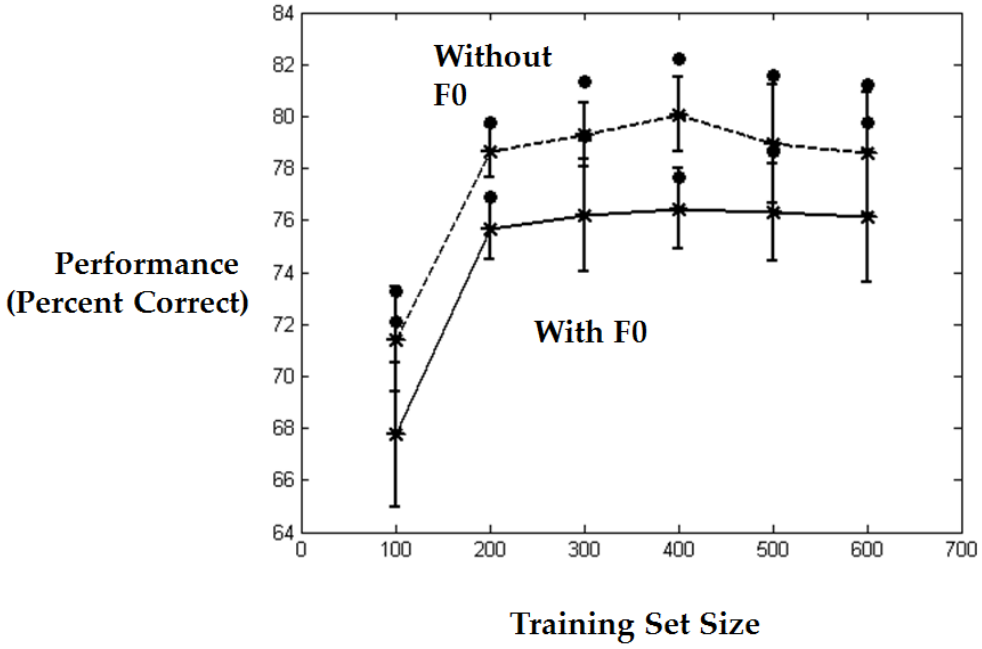


Figure 2.10: *Training set size.* The overall performance when the training set size is varied from 100 to 600 vowels. The dataset is also tested both with and without F0 information added to the spectral information of the signal at the anchor map. Five runs were performed for each training set size. The filterbank consisted of 250 filters ranging from 50-7500 Hz. The best performance, mean, and standard deviation results were recorded. The dashed line shows the results without F0 information added and the solid line shows the results with F0 information added to the spectral map. The * indicates the mean plus error bars and the • indicates the best performance for that training set size.

Training set size was varied (100, 200, 300, 400, 500, 600) with and without adding F0 information contained in the anchor map. The training set was chosen randomly without replacement. The remainder of the dataset was used for testing. The simulations used a filterbank that consisted of 250 filters ranging from 50-7500 Hz. Figure 2.10 shows the overall performance results from the different training set sizes. When the training set contained only 100 tokens, performance was the worst near 73% correct. Based on these results, a training set size of 400 achieves the best performance without adding F0

information (82.23% correct). Again, the model performed better without adding F0 information in order to anchor the spectral map.

2.6.4 Vowel duration

Vowel duration was varied to determine if steady state vowel duration affects model performance. Three vowel durations were tested (62.5, 300, 600 msec). No additional F0 information was used. The filterbank consisted of 240-260 filters, the dynamic range varied from 20 to 100 Hz on the low end and 7 to 8 kHz on the high end. The training set size was held constant at 400 vowel tokens. Four simulations were run for each vowel duration. Table 2.2 shows that varying vowel duration had little effect ($F[2,9] = 0.2428$, $p < 0.8$). However, the small effect may be because these vowels are steady-state. Varying the duration of naturally-produced vowels there may lead to a different outcome.

Vowel Duration (msec)	**Best Performance**	**Mean**	**Standard Deviation**
62.5	81.61	79.96	0.81
300	79.91	79.28	0.49
600	80.80	79.96	0.87

Table 2.2. The overall performance when the vowel duration varied from 62.5msec, 300 msec, and 600 msec. The filterbank varied from 240 to 260 filters, and the dynamic range varied from 10 to 100 Hz at the low end and 7 to 8 kHz at the high end. The training set size contained 400 tokens and four runs at each vowel duration were tested. No F0 information was added to the spectrum for these simulations. Percent correction classification in terms of best performance, mean, and standard deviation are recorded.

2.6.5 With and without adding F0 information

Three types of F0 simulations were performed. The first did not include any additional F0 information. In the second set, F0 information was only added to the anchor map and the spectral map received only spectral information. In the last type of simulations, F0 information was added to the spectral representation and this combination was redundantly mapped as input into both the anchor map and spectral map. F0 information was added to the spectrum to increase the energy at the filter corresponding to F0. These simulations used a training set of 400 vowel tokens of 62.5 msec in duration, a filterbank of 240-260 filters, a low frequency from 20-100 Hz, a high frequency of 7-8kHz, and 14 runs of each simulation type. Table 2.3 summarizes the results of these simulations. The best performance of 81.61% was found with no F0 information added. The simulations did find a highly significant effect across these conditions ($F[2,39] = 41.58$, $p < 0.001$) indicating that adding F0 information impaired performance.

F0 information	**Best Performance**	**Mean**	**Standard Deviation**
Without F0	81.61	79.96	0.81
Only in the Anchor Map	77.68	76.45	1.54
In both the Anchor and Stream Maps	78.04	77.33	0.64

Table 2.3. The overall performance without F0 information added to either map, F0 information only added to the anchor map, and F0 information added to the spectrum and redundantly mapped into both the anchor map and the stream map. The filterbank varied from 240 to 260 filters, and the dynamic range varied from 10 to 100 Hz at the low end and 7 to 8 kHz at the high end. The training set contained 400 tokens. The vowel duration was set at 62.5 msec. Fourteen runs were performed for each simulation type. Percent correction classification in terms of best performance, mean, and standard deviation are recorded.

2.7 Discussion

2.7.1 Comparison to human listeners

Comparisons of human identification rates and those of the model can only hope to show a qualitative correspondence, because human speakers come to such experiments with full knowledge of a language, and may thus contextually process even reduced speech cues in a way that a model that learns only those cues cannot. In particular, humans may experience filtering and competitive interference properties that a simple model may not. Despite this caveat, human/model comparisons illustrate many similar properties, as noted below.

Simulated identification rates for the synthesized vowels are shown in Table 2.4a along with the identification rates reported by Hillenbrand and Gayvert (1993) for flat F0 stimuli. Table 2.4b shows the results broken down across speaker groups (men, women, and children) from Hillenbrand and Gayvert (1993) and the model. The identification rates across speaker groups for the simulations was significant ($F[2,39] = 147.1105$, $p < 0.001$). The error rate was lowest for child speakers for both humans and the model. The model had the lowest error rate for the women speakers and the human listeners performed the best for male speakers. This difference may be due to learning in which the model was exposed to randomly chosen synthesized vowel samples from all three speakers groups whereas human listeners are exposed to a much wider variety of samples in varying context.

The confusion matrices for the Hillenbrand and Gayvert (1993) study are shown in Table 2.5a and for the simulations in Table 2.5b. The simulations performed in this

study found an overall accuracy measure of 79.96%, which is better than the 72.7% reported by Hillenbrand and Gayvert (1993).

Vowel	Simulation results	Hillenbrand and Gayvert (1993) Flat F0
IY	95.53 ± 1.10	96.2
IH	86.90 ± 4.91	67.0
EH	70.28 ± 3.13	65.8
AE	81.98 ± 1.53	63.2
AH	83.96 ± 3.81	74.7
AA	80.39 ± 1.83	55.0
AO	70.99 ± 7.03	67.2
UH	81.35 ± 1.97	62.0
UW	66.10 ± 3.84	89.1
ER	82.03 ± 4.11	86.6
TOTAL:	**79.96 ± 0.81**	**72.7**

(a)

Talker group	Simulation results	Hillenbrand and Gayvert (1993) Flat F0
Men	80.35 ± 1.78	74.4
Women	82.76 ± 0.54	72.2
Children	72.71 ± 2.10	70.0

(b)

Table 2.4. (a) Percent correct identification rates for the flat-formant synthesized vowels and **(b)** Percent correct identification rates for the three talker groups (men, women, and children) with flat F0s in both the simulations performed in this study and in the Hillenbrand and Gayvert (1993) study.

The confusion matrices for both the Hillenbrand and Gayvert (1993) study and the simulations reported here show that most errors occurred near the diagonal. The layout of the confusion matrix roughly corresponds to the layout of the vowels in F1/F2 space; see Figure 2.11. F1/F2 space is considered a rough perceptual mapping of vowels in that there is a relationship between the intended vowel and the formant frequency pattern (Peterson and Barney, 1952). Figure 2.11a shows the vowels that were classified

correctly and Figure 2.11b shows the vowels that were classified incorrectly. The ellipses are drawn based on the Peterson and Barney (1952) dataset. In Figure 2.11b, it is apparent that the majority of the vowel classification errors are near misses which occurred on vowel boundaries in the perceptual F1/F2 space. Fuzzy ARTMAP did a good job of correctly classifying vowels in overlapping F1/F2 space because the system does not cluster vowel boundaries in only F1/F2 space. Rather it takes into account the entire normalized spectra of the vowels.

Hillenbrand and Gayvert (1993) found that confusions between IY and IH, where IH was heard as IY; between UH and UW, where UH was heard as UW; and between EH and AE where AE was heard as EH are tense-lax asymmetries. They hypothesized that, when the subjects listened to vowel stimuli without durational cues, they had a tendency to misclassify the vowels as long rather than short vowels. The IH and IY confusions were consistently encountered with the model. However, the model was not as susceptible to the EH and AE confusions. This may be due to the shorter vowel duration (62.5 msec) used by the model as compared to the longer duration (300 msec) used by Hillenbrand and Gayvert (1993) such that the short AE was not misclassified as the long EH when presented with shorter stimuli.

	IY	IH	EH	AE	AH	AA	AO	UH	UW	ER
IY	96.2	3.1	0.6	0	0	0	0	0	0	0
IH	25.1	67.0	6.7	0.3	0.1	0	0	0.6	0	0.1
EH	1.3	23.7	65.8	7.2	0.3	0	0	0.4	0.1	1.1
AE	0.1	0.6	28.0	63.2	2.0	4.0	0	0.3	0	1.9
AH	0	0.1	0.9	0.7	74.7	12.8	6.8	2.7	0.1	1.2
AA	0	0	0.2	0.1	13.6	55.0	30.5	0.6	0.1	0
AO	0	0	0	0	8	5.9	67.2	13	5.9	0
UH	0	0.2	0.1	0	5.2	0.1	3.1	62.0	28.4	0.9
UW	0.2	0.2	0	0	0.7	0	0.7	9	89.1	0.2
ER	0.3	4.1	4.0	0.3	0.9	0	0	3	0.7	86.6

(a)

	IY	IH	EH	AE	AH	AA	AO	UH	UW	ER
IY	95.53	4.47	0	0	0	0	0	0	0	0
IH	7.24	86.90	4.28	0	0	0	0	0	0	1.58
EH	0	16.79	70.28	7.03	0	0	0	0	0	5.90
AE	0	0	9.25	81.98	7.76	0.31	0	0.12	0	0.56
AH	0	0	0	2.22	83.96	10.26	3.07	0.25	0	0.25
AA	0	0	0	2.40	13.50	80.39	2.90	0.81	0	0
AO	0	0	0	0.33	14.56	8.49	70.99	4.62	1.01	0
UH	0	0	0	0	0.90	0	4.48	81.35	9.02	4.25
UW	0	0.12	0.31	0	0	0	1.73	29.21	66.10	2.53
ER	0	2.72	3.06	3.54	2.72	0	0	5.53	0.41	82.03

(b)

Table 2.5. (a) Confusion matrix reported by Hillenbrand and Gayvert (1993) for synthesized steady state vowels with flat F0 contours. **(b)** Confusion matrix for synthesized steady state vowels with flat F0 contours generated with the model. The simulation results reported here are the mean results from fourteen runs. In each run, the filterbank was randomly chosen to be from 240 to 260 filters, and the dynamic range from 10 to 100 Hz at the low end and 7 to 8 kHz at the high end. The training set contained 400 tokens. The vowel duration was set at 62.5 msec.

In the simulations with differing vowel durations these confusions were reversed with UW heard as UH. These differences may be due to the initial stimulus set-up. When ANOVA is performed at the vowel durations of 62.5 msec, 300 msec, and 600 msec, it was found that the correct classification of UH differs significantly across the different

durations ($F[2,9] = 4.315$, $p < 0.05$) and almost significantly for the correct classification of UW ($F[2,9] = 3.320$, $p < 0.09$). The best classification was found at the 300 msec vowel duration, which is the same duration used by Hillenbrand and Gayvert (1993). Thus, the 300 msec stimuli seem to provide the best performance for classification of UH where the human subjects had a tendency to classify these vowels as long. When the model was presented with much shorter stimuli (62.5 msec), it classified these vowels as short, with performance improving at the longer vowel durations.

One other difference between Hillenbrand and Gayvert (1993) and the simulations concerns the classification of the vowel AA. Hillenbrand and Gayvert (1993) found that human listeners frequently misclassified AA as AO whereas the model frequently misclassified AA as AH, AH as AA and AO as AH. All three of these vowels are back vowels for which the tongue is placed near the back of the mouth and roughly corresponds to a smaller difference between F1 and F2 (Lindau, 1978). AA is differentiated from AH and AO in that it is slightly more open and unround, with AH unrounded and AO rounded. Thus, AA and AH differ only by a slight variation in openness, whereas AA and AO also differ in rounding. Finally, Figure 2.11 shows that all three of these vowels significantly overlap in F1/F2 space. The confusions made by both human listeners and the model are consistent with the close proximity of these vowels in perceptual space and both types of confusions are valid.

The last group of differences involves taking into consideration the better classification of AE and EH by the model versus the better classification of IY by

human listeners. It seems that the model is biased towards lower values for F1 than human listeners. This may be due to the fact that the low frequency endpoint of the filterbank was varied from 10 to 100 Hz, which may be lower than what is typically found in humans. These differences may also be attributed to the training effects where the model had more exposure to lower harmonics than do humans when learning to speak. In making this comparison, it is also worth noting that the model experienced only the vowel data, whereas humans respond with potential competition from the entire language.

The classification results show that the speaker normalization circuit helped to recognize vowels in the Peterson and Barney (1952) dataset. The training results of the fuzzy ARTMAP classifier also produce important information. During training, fuzzy ARTMAP learns categories corresponding to the vowel categories. In these simulations, only ten categories were learned, corresponding to the ten vowels in the dataset. Therefore, there is a one-to-one mapping between the learned categories and the vowel categories, which indicates the success of the system in creating invariant representations of the vowel categories. If the normalization scheme did not perform well, fuzzy ARTMAP would have learned to select many more categories corresponding to each vowel category. For example, without speaker normalization pre-processing, the classifier generated learned, on average, thirty categories and the vowel classification performance dropped to 71.95%. The classification rate of the system without normalization was slightly lower than human performance of 72.7% as reported by Hillenbrand Gayvert (1993).

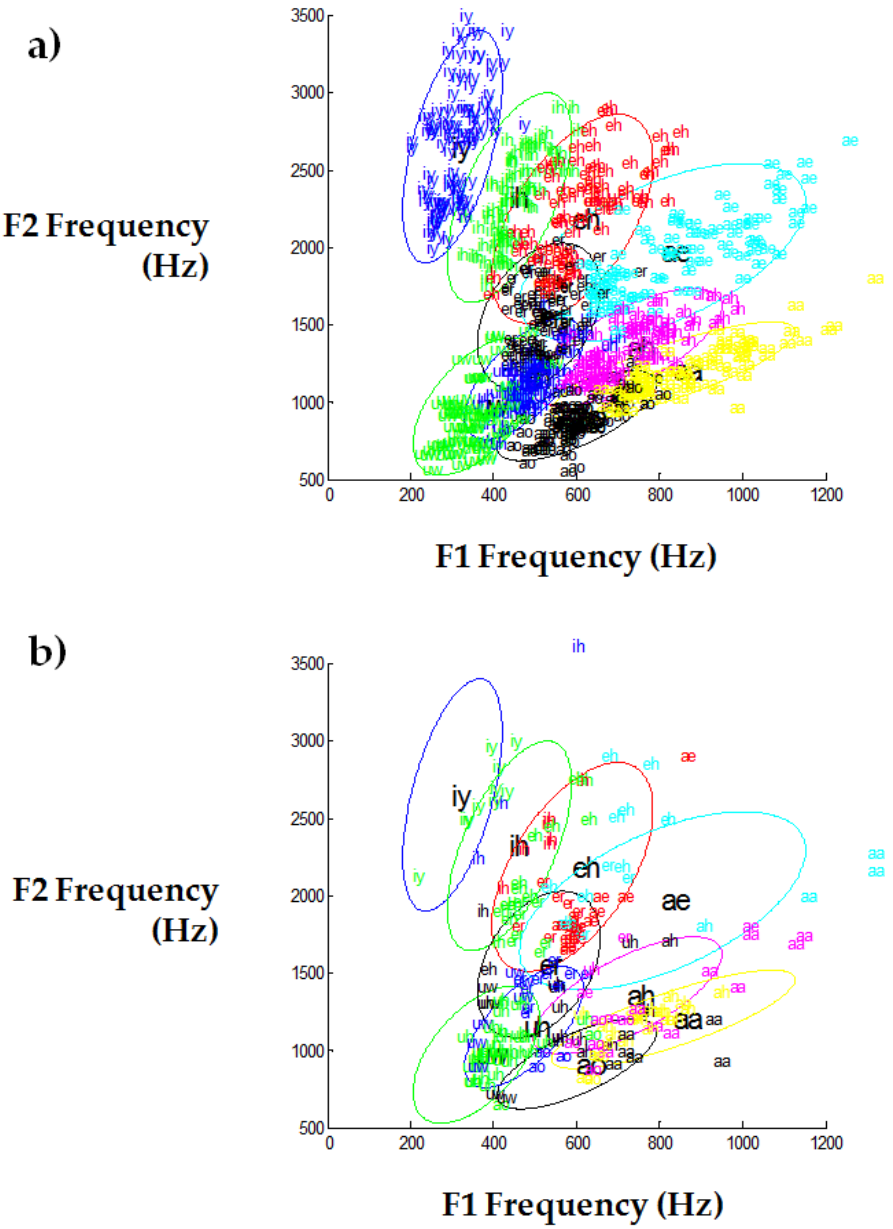


Figure 2.11: *F1/F2 vowel space.* The ellipses correspond to the confidence intervals reported over the entire Peterson and Barney (1952) database where IY is blue, IH is green, EH is red, AE is light blue, AH is pink, AA is yellow, AO is black, UH is dark blue, UW is green, and ER is black. **(a)** The individual data points correspond to the locations of the correctly classified vowels in the simulations. **(b)** The individual data points correspond to the misclassified vowels. The color of the data points corresponds to what the vowel should have been classified as and the vowel label at the data point corresponds to what it what misclassified as.

As summarized above, the vowel identification rates of the model are comparable to those reported by Hillenbrand and Gayvert (1993), with most of the misclassifications lying in adjacent F1/F2 space. The identification rate of 72.7% of Hillenbrand and Gayvert (1993) and the 79.96% reported in these simulations is still significantly less than the 94% accuracy reported by Peterson and Barney (1952) in a vowel recognition task because of the use of three-formant steady state synthesized vowels. These stimuli are more difficult for listeners and the model because of their lack of context, durational variation, and natural fluctuations of speech. Accuracy greatly improves from 57% to 95% for isolated vowel recognition to 83% to 96% in CVC context (Nearey, 1989). Listeners achieved 79% accuracy when listening to vowel recordings, but when the vowels were synthesized with a fixed duration and steady-state formants, only 61% accuracy was achieved (Lehiste and Meltzer, 1973). Listeners could achieve 89% accuracy when vowels were synthesized with their original formant trajectories, but would only achieve 74% accuracy when the vowels were synthesized with flat formants (Hillenbrand and Nearey, 1999).

The ART modeling framework, of which the NormNet forms a part, promises to generalize to more complex speech. Variability in human speech such as durational cues and context cues are best captured by a dynamical system that catches sources of variation in time, that is able to learn from local knowledge available in the speech signal, and that learns in real time. Previous modeling studies have quantitatively simulated speech categorization and word recognition data with these properties (e.g., Grossberg *et al*., 1997; Boardman *et al*., 1999; Grossberg and Myers, 2000). The

NormNet circuit, when integrated within this emerging theory, will enable it to begin to simulate variable speech perception properties of multiple speakers.

Many studies have shown that speaker normalization is an active process such that listeners are better and faster at word and vowel recognition in single talker lists than in multi-talker lists (Creelman, 1957; Summerfield and Haggard, 1973; Verbrugge *et al*., 1976) and that listeners retain memory of speaker-specific acoustic detail that influences their memory of previously spoken words (Palmeri *et al*., 1993; Church and Schacter, 1994; Goldinger, 1996, 1997). Although some authors have suggested that this is evidence against automatic speaker normalization, these results are consistent with the NormNet model. The increased reaction time and decreased performance in the multi-speaker lists can be attributed to the switching cost of moving to a new stream and to a new anchor and location within the maps. More generally, such effects may be due to a range of additional interactions between speaker-dependent and speaker-independent processes. This hypothesis is further supported by the study performed by Kato and Kakehi (1988), which showed that accuracy in syllable recognition monotonically increases from the first to the fifth presentation and that accuracy no longer increases on successive presentations. Furthermore, listeners are influenced by expectation of gender either through auditory cues (Eklund and Traunmüller, 1997; Johnson *et al*., 1999) or visual cues (Walker *et al*., 1995; Strand and Johnson, 1996; Schwippert and Benoit, 1997; Johnson *et al*., 1999). Taken together, these studies may probe how the anchored stream serves as a frame of reference for understanding a speaker and that mismatched expectations or switching between speakers affects

reaction time and performance as the model adapts to the variation in the location of the anchored stream.

Kraljic and Samuel (2007) recently reported that although listeners may adjust their internal representations of phonemic categories based on the speaker, this is not always the case. In fact, some of these categorical adjustments are not related to the speaker but may be primed by other cues (e.g., rate in voice onset time (VOT) of stop consonants). Therefore, it is necessary for the phonemic categories to learn in real time to adjust to fluctuations both within the speech of a single speaker and across speakers. Such adaptations can occur at multiple levels of the speech perception system.

Johnson (1997a, 1997b, 2005, 2006) suggests that the above-mentioned evidence is supportive of an episodic/exemplar coding model that bases speech perception on a set of stored exemplars that adapt to the perceived identity of the talker. The exemplar models do not make use of speaker normalization and instead rely on the large repertoire of exemplars to feed into category nodes describing the recognized speech sample. However, the limited data simulated with these models fails to show how they can scale up to natural speech and the large variety of speakers encountered in everyday life. To assume that the exemplars are created for each different mode of speaking the same speech sound would assume that the human brain has a massive capacity to store these exemplars, and a way to search among these exemplars in real time. The simulations reported by Johnson (1997b, 2006) do not report the number of exemplars or categories that are learned for each of the words tested. Furthermore, the learning of the word categories based on the set of exemplars requires the use of

teacher-based knowledge to create the mapping between the exemplars and the category nodes. The category nodes cannot be created based on local knowledge and in real time.

NormNet is not subject to these sources of criticism. The speaker normalization circuit does not require the storage of instances or exemplars of speech sounds and can learn a single categorical representation for each speech sound. The creation of the invariant speech representations described herein does not require the use of a teaching signal and can be performed in real time with only local knowledge available to the system. The simulation results without normalization show that vowel categorization was poorer (71.95%) and that more vowel categories were created (30.29 nodes generated). Hence, the addition of speaker invariance creates more stable categories and better performance.

More generally, ART-based categorization can, where task demands require, learn both specific, concrete categories, even individual exemplars, as well as general, abstract categories. This learning process enables ART models to selectively pay attention to the learned prototype of critical feature patterns that predict successful performance. High vigilance leads to concrete categories, whereas low vigilance leads to general categories (see Section 2.4 and Appendix A). Thus, one does not need to store all exemplars to learn fine distinctions when, in fact, a language requires them. In several senses, the NormNet circuit and the larger ART speech perception network to which it belongs (Figure 2.1), embody the phonological principle which states “that languages have basic building blocks, which are not meaningful in themselves, but which combine in different ways to make meaningful forms” (Pierrehumbert, 2006).

2.7.2 Comparison to other speaker normalization techniques

Other speaker normalization techniques have been applied to a variety of vowel recognition tasks using different classifiers. Two cues, vocal tract length and F0, are important in these speaker normalization techniques. Inter-speaker variability is often attributed to the difference in the shape and length of the vocal tracts, with males typically having longer vocal tracts then females (Lee and Rose, 1998; Stevens, 1998). The correlation between the vocal tract length and the position of the vowel formants contributes to differences perceived by the listener (Fant, 1973). Vocal tract length normalization (VTLN) is based on the assumption that the speech spectrum of one speaker differs from another due to stretching or compression along the frequency axis (Eide and Gish, 1996; Wegman *et al.*, 1996; Lee and Rose, 1996, 1998; McDonough and Byrne, 1999; Dognin and El-Jaroudi, 2003; Glavitsch, 2003). The speech sound is normalized by warping the frequency axis onto a standard vocal tract length. It is, however, unclear how speakers could estimate vocal tract length during naturally occurring language experiences.

Nonlinear (e.g. Eide and Gish, 1996), linear (e.g. Zahorian and Jagharghi, 1991), and bilinear (e.g. Glavitsch, 2003) transformations have all been used in VTLN techniques. The resulting transformation has the same Fourier transform as the original except that it is warped along the frequency axis. Both linear and bilinear transformations have led to increased performance in systems performing speech related tasks (Lee and Rose, 1996; Wegmann *et al.*, 1996; Zhan and Westphal, 1997; Zhan and Waibel, 1997).

Wegmann *et al.* (1996) used a VTLN method in which the frequency warping was done using a piecewise linear transformation of the frequency axis with fixed points at 0 kHz and the Nyquist frequency. Ten warp scales were constructed and each map scale was applied to the speech sound. The best warp scale was chosen through a comparison to a generic voiced speech model. Wegmann *et al.* (1996) reported a 12% reduction in word error rate as compared to unnormalized gender-independent models and a 6% reduction as compared to unnormalized gender-dependent models when tested on the standard Switchboard Corpus (NIST).

Zahorian and Jagharghi (1991) evaluated the effect of both a linear transformation of spectral features and a speaker-dependent frequency warping procedure to evaluate improvement on vowel classification. In both, the normalization parameters were chosen to minimize the mean squared error between the normalized features and the target features. They found an 8 to 15% increase in accuracy, where the accuracy level ranged from 69 to 91%.

It is difficult to compare the performance across these different speaker normalization techniques because of the different data sets and vowel classifiers that were used. A meaningful metric is to compare the performance of each technique to human listeners on comparable tasks. The Peterson and Barney (1952) database contains only steady state vowel information and human listeners are not as good at recognizing steady-state vowels as vowels containing durational and contextual cues. Taking this into account, if the results from the simulations of this paper, 79.96% correct, are compared to the human listeners of the Hillenbrand and Gayvert (1993)

study, 72.7% correct, the simulations reported by Nearey *et al*., (1979), with 81 to 92% correct, Syrdal and Gopal (1986), with 81.8 to 85.7% correct, and Turner and Patterson (2003), with 79 to 84%, it seems that these other systems may overfit human data, whereas the simulations from this paper adhere more closely to the reported human data.

2.7.3 Role of F0 in speaker normalization

F0 is determined by the rate of vibration of the vocal cords of the speaker and thus correlates with the size of the speaker's vocal folds (Titze, 1994). The average values of F0 are lowest in males, around 100 Hz, 200 Hz in females, and up to 400 Hz in infants (Kent and Read, 1992). Because the harmonics of the speech sound correspond to integer multiples of F0, F0 can, in principle, be inferred by the human brain from the spectrum of harmonics even if it is a *missing fundamental* in the signal (Pantev *et al.*, 1989; Ragot and Lepaul-Ercole, 1996).

The distance between F1 and F0 in critical bandwidth is an important cue for perception of vowel openness (Traunmüller, 1981). Speaker-dependent information contained in F0 has also been found to be important in both the recognition of vowels and Mandarin Chinese tones (Johnson, 1990; Moore and Jongman, 1997).

F0 varies greatly amongst speakers and the range of F0 in human speakers can vary from 50 to 800 Hz (Hess, 1983; Ferreira, 2007). The high end of this range, found in female and child speakers and in singing, can result in F0 being comparable to or higher than F1. In addition, it may be the case that the vowel's spectrum may contain a sufficiently small amount of energy for F0. In both of these situations, the asymmetric

competition in the model may be compromised. In order to test this concern, additional F0 information added into the speaker normalization model to anchor the spectral information, which caused performance to decrease. Thus, the F0 information contained in the original vowel spectrum is sufficient for speaker normalization.

F0 has been found to be slightly helpful in understanding speech in both speech recognition systems and human listeners (Glavitsch, 2003; Magimai-Doss *et al.*, 2003). For example, Nearey *et al.* (1979) classified the Peterson and Barney (1952) database with a linear discriminant classifier to identify vowels. They reported 81% correct when the system was trained on log-transformed F1 and F2, 86% correct when F0 and F3 were included, and 92% correct when speaker mean log formant values were subtracted from the individual log formant values. Syrdal and Gopal (1986) also used a linear discriminant classifier in which they achieved 81.8% correct when trained on F0 and F1-F3 and 85.7% correct when also trained on three bark-transformed spectral differences (F3-F2, F2-F1, and F1-F0). Turner and Patterson (2003) used a Mellin transform to look at the variation of vocal tract length and achieved 79% to 84% correct. These improvements using F0 are small or modest. In the case of Nearey *et al.* (1979), the classification used F0 information as an additional feature for the classifier rather than for use with the normalization scheme. Syrdal and Gopal (1986) used F0 information only to normalize the first formant. Thus, although F0 contains important cues that are useful for human listeners to evaluate the meaning of a speech utterance, F0 may not be needed to normalize typical speech sounds. However, F0 may be more useful in normalizing more extreme cases such as singing or infant-directed maternal

speech which in some cases has a very high F0. These cases have yet to be thoroughly tested in NormNet.

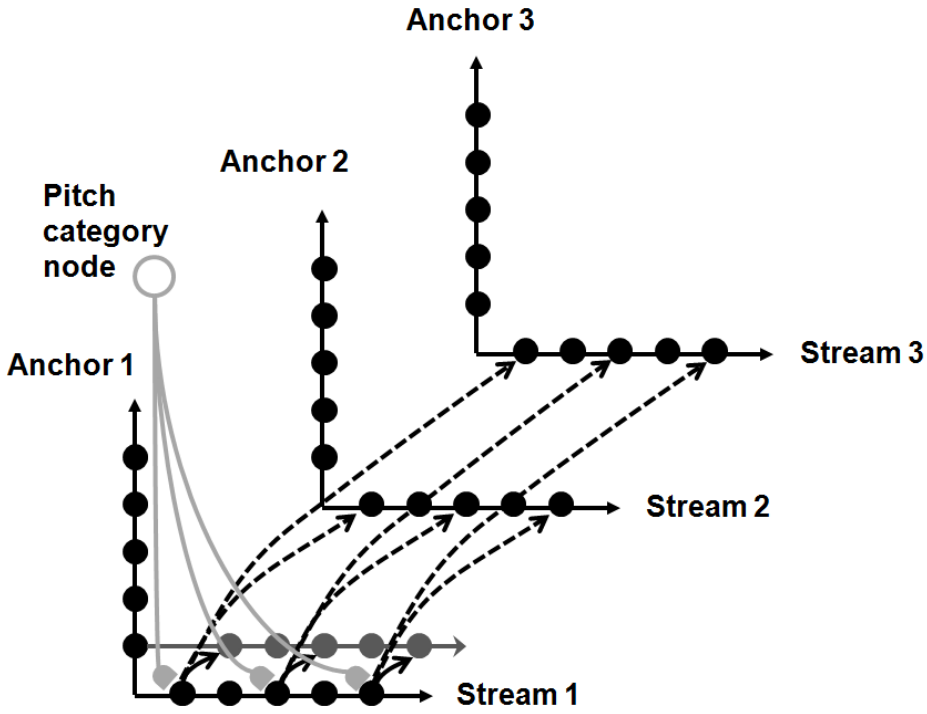


Figure 2.12: *Unification of multiple streams and their speaker normalization circuit.* Three potential streams and their anchor maps are illustrated. The first stream is chosen when its pitch category wins a competition with other streams and uses a harmonic top-down expectation to select the frequencies that are compatible with that pitch. Asymmetric competition with the other streams causes the selected frequencies to be exclusively allocated to that stream. Other streams select their frequency spectra from the remaining frequencies. This selection process determines the anchor frequency for each stream and thereby initiates speaker normalization within each stream by again using asymmetric competition to normalize each selected frequency spectrum in its stream.

2.8 Conclusion

The NormNet model provides a proof of principle for a new insight into how speaker normalization may be carried out by the brain. The speaker normalization model was able to achieve accuracy of 79.96% correct, on average, which is comparable to the

results obtained with human listeners identifying similar vowel stimuli. The model proposes that tonotopic strip maps of frequency-selective auditory cortical cells and asymmetric competitive interactions are used both to define the auditory streams that characterize acoustic sources, and to normalize the frequency spectra of these streams so that they can be understood across multiple speakers. Figure 2.12 depicts a hypothetical brain map that unifies multiple streams and the speaker normalization circuit. The way in which strip maps and asymmetric competition may be used in both streaming and speaker normalization circuits is a worthy topic of future research to clarify the predicted shared mechanisms that may be at work.

CHAPTER 3

APRAXIA OF SPEECH

Apraxia of speech (AOS) is one of many speech disorders that have been examined by speech pathologists, linguists, and neurologists. Some speech disorders impair speech perception and others impair speech production while some disorders involve damage to both systems. A large amount of literature is dedicated to differentiating symptoms of these disorders. Current research has begun to focus on how specific disorders relate to lesion sites in the brain, but brain imaging has yet to become common practice in the speech disorder diagnosis community.

Apraxia of speech is a disorder of planning and/or programming of speech production. There is generally neither comprehension impairment nor an impairment of the speech musculature. Apraxia of speech is diagnosed by the presence of a variety of speech errors using a battery of tests. Consequently, the speech pathology community is uncommitted to a core set of symptoms that define apraxia of speech, although more recently a set of inclusionary and exclusionary behavioral criteria has been proposed for its diagnosis (Wambaugh *et al*., 2006a). Typically, diagnosis of apraxia of speech has not involved neuroimaging. Thus, there is little agreement upon how the brain is damaged in apraxia of speech.

This chapter is intended to provide an overview of current research into apraxia of speech as a disorder as well as current research into brain imaging of this disorder. A review of relevant modeling efforts of speech production will follow. Finally, an attempt will be made to explain apraxia of speech based on these modeling efforts.

3.1 Behavioral characteristics of AOS

Apraxia of speech can be characterized as a motor speech disorder which involves a disruption to speech motor control. Speech motor control is defined as "the systems and strategies that regulate the production of speech, including the planning and preparation of movements and the execution of movement plans to result in muscle contractions and structural displacements" (Kent, 2000). Inputs to this system are phonological representations of language (e.g. phonemes or syllables) and outputs from this system are articulatory movements that result in speech. Thus, this system represents the transition from phonological encoding to articulation.

Ever since Fred Darley first described AOS, the disorder has been defined as a disorder of planning and/or programming of volitional speech movements, not due to abnormalities in strength, speed, and coordination of the speech musculature (Darley, 1968; Darley *et al*., 1975; Duffy, 2005). AOS is distinguished from the dysarthrias, which are motor speech disorders resulting from abnormalities of neuromuscular functioning, and from aphasias, which are language disorders. Although dysarthrias were originally seen as deficits in execution of movement only, more recent views of dysarthrias acknowledge motor programming deficits in most dysarthria types as well (Spencer and Rogers, 2005; Van der Merwe, 2007), with the exception of dysarthria due to flaccid paralysis (Duffy, 2005; McNeil *et al*., 2000; Van der Merwe, 2007). Unlike the dysarthrias, which may affect any or all systems of speech production (i.e. respiration, phonation, resonation, articulation, prosody), AOS affects primarily articulation and prosody.

Apraxia of speech can be divided into two types: childhood (CAS) and acquired (AOS). The difference between AOS and CAS is that AOS occurs as a result of brain damage, whereas CAS occurs developmentally and may also be genetic (Kent, 2000; Maassen, 2002). As a result, CAS is characterized not only by the symptoms seen in AOS patients, but also developmental speech problems (Maassen, 2002). The remainder of this chapter will focus on AOS instead of CAS.

Much debate, both theoretical and clinical, has revolved around the differentiation of AOS from dysarthrias and from aphasia (McNeil *et al*., 2000; Rosenbek, 2001). There are several reasons for this debate. First, many speech features associated with AOS may occur in all three disorders, though for different reasons. For instance, a perceived sound substitution (e.g., saying ‘ship’ instead of ‘sip’) could arise because the wrong target word was selected (a language problem), because the tongue movement necessary for ‘s’ was not planned or programmed correctly (an apraxic problem), or because weakness of the tongue muscles prevented the tongue protrusion needed for ‘s’ (a dysarthric problem). Second, AOS rarely occurs in pure form but typically co-occurs with aphasia and/or dysarthria, making it difficult to ascribe speech features with confidence to a single underlying cause. For instance, although AOS by itself is thought to involve intact auditory comprehension, the co-existence of aphasia may produce comprehension deficits in a patient with AOS. Third, the vocabulary to describe speech errors due to speech motor programming deficits has often been borrowed from the linguistic domain (e.g., phonemes), so that a description of errors in terms of phoneme substitutions or omissions is fundamentally incapable of

distinguishing between an aphasic (language) impairment and a speech motor programming (apraxic) impairment (Buckingham, 1979; Martin, 1974, 1975; Aten *et al*., 1971).

Further complicating the differentiation of AOS from other speech production deficits is the fact that a patient may exhibit only a subset of symptoms, yet still may be diagnosed with AOS. Finally, the set of symptoms used to diagnose AOS has not been consistent across different speech-language pathologists and different researchers. As McNeil *et al*. (2000) have noted, the literature on AOS is difficult to interpret in light of these various diagnostic criteria used to describe and diagnose the patients. Consequently, each study of AOS needs to be considered in terms of its AOS diagnostic techniques.

In an effort to remedy this state of affairs, a number of experts in the field of AOS have recently proposed a set of consensus criteria for the diagnosis of AOS (Wambaugh *et al*., 2006a). This set of criteria involves behavioral features, and includes both inclusionary and exclusionary criteria. In addition, a number of characteristics are proposed to be nondifferential or indicative of other disorders, as an explicit acknowledgement of the fact that AOS rarely occurs in isolation but typically occurs in the presence of other disorders such as dysarthria, limb apraxia, oral apraxia, or aphasia (McNeil and Doyle, 2004). There has been debate in the speech pathology community about whether to include lesion information as part of the diagnosis of AOS (e.g., Mlcoch *et al*., 1982), although this approach has generally been rejected because patients with seemingly similar symptoms have been found to have widely different

lesion sites. This latter observation strongly suggests a need for well-specified neural models of speech production as well as behavioral methods to test such models and differentiate between impairments. At present, the diagnosis of AOS remains purely behavioral, and the currently most widely accepted definition of AOS is also specified without regard to neural lesions (McNeil, *et al*., 2007):

> "Apraxia of speech is a phonetic-motoric disorder of speech production. It is caused by inefficiencies in the translation of well-formed and -filled phonological frames to previously learned kinematic information used for carrying out intended movements. These inefficiencies result in intra- and interarticulator temporal and spatial segmental and prosodic distortions." (p. 264)

An overview of the most recent consensus criteria for the diagnosis of AOS is presented in Table 3.1. Primary characteristics are those that, when present, suggest the presence of AOS, especially when they occur in combination. The primary characteristics of AOS are (1) overall slowed speech, due to both sound prolongations and increased intersound durations; (2) abnormal prosody with excess and equal stress (stress on normally unstressed syllables); (3) sound distortions and distorted substitutions as the predominant error type; and (4) consistent error types from trial-to-trial, although the presence of error is inconsistent (Wambaugh *et al*., 2006a).

Type of Characteristic	Characteristics
Primary (typical of AOS)	1. Slow speech rate, due to sound prolongation, increased intersegment durations, and insertion of schwa (e.g., plane → puhlane)
	2. Prosodic abnormalities, especially excess and equal stress on syllables
	3. Sound distortions and distorted substitutions
	4. Errors relatively consistent in type and location within the utterance, but errors not consistently present
Nondiscriminative (may also occur in other disorders)	1. Articulatory groping (audible and/or visible)
	2. Speech initiation difficulties
	3. Perseverative errors (e.g., pancakes → panpakes)
	4. Increasing errors with increasing word complexity
	5. Automatic speech better than volitional, propositional speech
	6. Islands of fluent, error-free speech
	7. Awareness of errors
	8. Anticipatory errors (e.g., pancakes → cancakes)
	9. Transposition errors (e.g., pancakes → canpakes)
Indicative of other disorders (may co-occur with AOS)	1. Limb apraxia
	2. Oral (nonspeech) apraxia
	3. Expressive-receptive speech/language discrepancy
Exclusionary	1. Normal speech rate
	2. Fast speech rate
	3. Normal prosody

Table 3.1. Characteristics for the diagnosis of AOS. [adapted from Wambaugh et al. 2006a]

3.1.1 Primary features

Slow speech rate is a perceptually prominent feature of AOS, and acoustic analysis has indicated that the reduction of speech rate is due to prolongation of consonants and vowels, increased duration of pauses between sounds, and insertion of transitional vowels (Kent and Rosenbek, 1983; Seddoh *et al*., 1996; Strand and McNeil, 1996). It appears that transitionalizing from one sound or syllable to the next is especially

challenging for speakers with AOS (Kent and Rosenbek, 1983). It has also been reported that individuals with AOS are incapable of increasing speech rate without increasing errors (McNeil *et al*., 2000; Robin *et al*., 1989). Reduced speech rate is not caused by slowness of movements: although some kinematic studies have observed reduced peak velocity of articulators (Bartle *et al*., 2007a; Itoh *et al*., 1980), most studies have found normal peak velocities (McNeil and Adams, 1991; McNeil *et al*., 1989; Robin *et al*., 1989; Van Lieshout *et al*., 2007).

Abnormal prosody, with more equalized stress on successive syllables, has been confirmed in perceptual and acoustic studies (Kent and Rosenbek, 1983; Masaki *et al*., 1991; Odell *et al*., 1991; Square *et al*., 1982). Odell *et al*. (1991) observed more perceived stress errors on multisyllabic words in speakers with AOS but not in speakers with conduction aphasia. Using acoustic analysis, Kent and Rosenbek (1983) noted that in addition to more equalized syllable durations, the differences in peak intensity values of syllables in a sentence were smaller than those of control speakers. Masaki *et al*. (1991) reported coordination difficulties between pitch contour and articulatory movements in a speaker with AOS. Not all aspects of prosody are impaired however. For instance, Kent and Rosenbek (1983) found that apraxic speakers did show the normal terminal fall in fundamental frequency at the end of the sentence. It should be noted that not all studies have found evidence for stress errors in AOS. For example, Marquardt *et al*. (1995) found few differences between apraxic and control speakers in terms of perceived stress errors nor in terms of intensity, fundamental frequency, or syllable duration, although they did find some neutralization of syllable durations.

Whether these prosodic abnormalities are primary (i.e., part of the underlying deficit) or secondary (i.e., compensatory behavior to cope with the primary deficit) remains unclear. For example, some have speculated that slowed speech rate is compensatory, in that it might allow for more time for processing feedback to achieve speech targets (Bartle *et al*., 2007a,b; Darley *et al*., 1975; Kent and Rosenbek, 1982, 1983). A study by Rogers *et al*. (1996) examined this possibility by comparing vowel duration in speakers with AOS and aphasia, with and without masking noise. The rationale was that if vowel lengthening was due to a reliance on auditory feedback to verify target attainment, then masking the auditory feedback should result in shorter (more normal) vowel durations because the auditory feedback strategy would not be possible under masking conditions. Rogers *et al*. (1996) found that the apraxic speakers did not differ from controls, and both increased vowel durations with masking noise. Thus, reliance on auditory feedback does not appear to be the reason for longer vowel durations in AOS.

An alternative possibility is that perhaps slowed speech rate and segmentation of speech reflect a primary deficit in generating speech plans of longer size (Kent and Rosenbek, 1983; Rogers *et al*., 1996; Rogers and Storkel, 1999; Square *et al*., 1982). Another possibility is that the unimpaired hemisphere (generally the right hemisphere) takes over some speech motor programming functions from the damaged (usually left) hemisphere, resulting in interference with its original function, which may involve control of prosodic aspects of speech production (see Baum and Pell, 1999; for a review). Despite this uncertainty, slow rate and dysprosody are now considered

essential for the diagnosis of AOS. In fact, normal or fast speech rate and normal prosody are currently considered to be the only exclusionary criteria for AOS (Wambaugh *et al*., 2006a).

Distortions of speech sounds are regarded as the predominant segmental error type in AOS (McNeil *et al*., 2007), based on a handful of studies involving pure AOS (Odell *et al*., 1990; Square *et al.*, 1982). Phoneme substitutions may also occur but these are generally also distorted. Perceptually, distortions sound like poor exemplars of the correct sound, although extreme distortions may be perceived as substitutions. Usually, the target word is recognizable, and perceived substitutions tend to differ from the target in only one or two articulatory features, such as voicing or place of articulation (Odell *et al*., 1990; Sugishita *et al*., 1987; Trost and Canter, 1974). Acoustic and kinematic studies have further corroborated the perceptual evidence, and have suggested both spatial and temporal deviations as being responsible for perceived distortions (Fromm *et al*., 1980; Haley, 2002; Itoh *et al*., 1979, 1982; Kent and Rosenbek, 1983; McNeil *et al*., 1989, 1994; Seddoh *et al*., 1996; Shuster and Wambaugh, 2000; Sugishita *et al*., 1987; Wambaugh *et al*., 1995; Ziegler and Von Cramon, 1986a,b).

Another primary feature is that the presence of a speech error may vary from attempt to attempt, but that the type of error (distortion) and location within the utterance (e.g., syllable-initial, medial, or final) tend to be relatively consistent (Mauszycki *et al*., 2007; McNeil *et al*., 1995; Sugishita *et al*., 1987; Wambaugh *et al*., 2004; Mlcoch *et al*., 1982). Although some investigators have reported that repeated productions of the same word or passage reduces the number of errors produced (Deal,

1974; Johns and Darley, 1970), others have reported no difference in number of errors across multiple repetitions (LaPointe and Horner, 1976). McNeil *et al*. (1995) further reported that their speakers with relatively pure AOS did not generally appear to benefit from repeated attempts, unlike speakers with conduction aphasia.

3.1.2 Nondiscriminative features

There are a number of characteristics that may occur in AOS but that are not discriminative (i.e., may also occur in other disorders such as phonological deficits). One often striking feature is the presence of articulatory groping, which refers to the occurrence of off-target articulatory positions that give the impression of searching for the correct articulatory configuration (e.g., labiodental contact or lip rounding for the target word ‘big’). Such attempts may be audible (McNeil *et al*., 1995), but they may also be silent and only discernible visually (Bartle *et al*., 2007b). Difficulty with initiating speech is also often present in speakers with AOS, although this may reflect a problem with motor programming rather than with reading out programmed movements, as suggested by the fact that when speakers with AOS are given time to prepare an utterance, initiation latencies do not differ from unimpaired speakers (Maas *et al*., 2008).

Errors involving serial misordering of phonemes may suggest either impairment at the level of phonological encoding, where the serial order of phonemes is thought to be specified (Levelt *et al*., 1999) or impairment at the level of motor sequencing and planning (Bohland *et al*., in press). Perseverative errors are not uncommon, and tend to be more common than anticipatory errors. Perseverative errors involve repeated

production of a speech sound or fragment after it has been used more or less recently, and may indicate getting stuck on that sound or fragment. Anticipatory errors are those in which a speech sound or fragment appears too soon, before its target location, and suggest availability of later parts of the utterance. In addition to perceptual clinical observations, acoustic evidence indicates reduced anticipatory coarticulation in speakers with AOS (McNeil *et al*., 1994; Ziegler and Von Cramon, 1985, 1986a,b).

A number of factors may affect speech performance, including utterance complexity, input modality, and automaticity. With respect to utterance complexity, multisyllabic words tend to result in greater slowing of speech rate than monosyllabic words (Haley and Overton, 2001; Kent and Rosenbek, 1983; Strand and McNeil, 1996) and tend to result in more errors (Deal and Darley, 1972; Johns and Darley, 1974; LaPointe and Horner, 1976; Square *et al*., 1982; Odell *et al*., 1990). Syllable structure complexity also influences performance, in that speakers with AOS generally produce more errors on consonant clusters than on single consonants (Aichert and Ziegler, 2004; LaPointe and Johns, 1975; Shankweiler and Harris, 1966; Square *et al*., 1982), and they produce fewer complex syllables in conversation (Edmonds and Marquardt, 2004).

In addition, the complexity of a syllable sequence also affects speech production, in that speakers with AOS tend to have greater difficulty with sequential motion rate (SMR), which involves rapid repetition of sequences comprised of different syllables (e.g., ‘patakapataka’), than with alternating motion rate (AMR), which involves rapid repetition of the same syllable (e.g., ‘papapapapapa’). Although speakers with phonological paraphasia may also produce more errors on SMR than on AMR,

speakers with ataxic dysarthria generally do not show such differential performance on SMR and AMR, making this difference one of the few features differentiating AOS and ataxic dysarthria (Duffy, 2005).

Regarding input modality, some investigators have reported that speakers with AOS produce fewer errors when asked to repeat after the examiner whose mouth is visible (auditory-visual mode) than when asked to repeat with only an auditory stimulus or when reading aloud without an auditory model (Johns and Darley, 1970). However, LaPointe and Horner (1976) reported that the audio-visual mode actually resulted in more errors than the auditory-only mode. These differences between studies likely reflect differences between patient characteristics; it seems that input modality affects speech production performance differently in different individuals. Some evidence also exists that repetition performance exceeds spontaneous production performance (Shankweiler and Harris, 1966; Trost and Canter, 1974; Wertz *et al*., 1984). Thus, it appears that providing an auditory model of the target improves production of the target. In addition, there is some suggestion that delaying auditory feedback of a speaker's own speech disrupts speech production more in patients with nonfluent aphasia (who likely also had AOS) than in patients with fluent aphasia (Chapin *et al*., 1981), although others have failed to find effects of delayed auditory feedback on speech production in AOS (Lozano and Dreyer, 1978).

With respect to automaticity, one sometimes striking feature of AOS is that more automatic, less volitional utterances are easier to produce than more volitional ones. For example, a patient may be unable to repeat the phrase 'Bless you' after the

examiner, but may produce the phrase spontaneously and without error in response to the examiner's sneeze. Such observations fit with the notion of apraxia in general as being an impairment of volitional movement control (Liepmann, 1900). However, the concept of volitionality is difficult to define operationally and quantify, thus leaving this aspect of apraxia of speech supported mainly by anecdotal clinical observations. One possible way to operationally define automaticity is in terms of frequency of occurrence of words or syllables, on the assumption that very frequent utterances should be more automatic than rare utterances.

Clinically, the automatic-volitional distinction is often examined by comparing production of familiar sequences (e.g., counting to 20, days of the week) in normal order versus reversed order (Dabul, 2000). A study by Aichert and Ziegler (2004) compared production of syllables that differed in terms of their frequency of occurrence. Their results indicated that speakers with AOS produce fewer errors on highly frequent syllables than on less frequent syllables, providing support for the notion that frequency of occurrence mediates performance in AOS. However, frequency of occurrence does not capture all aspects of this automatic-volitional distinction, as suggested for example by the difference between productions of the same utterance produced in repetition tasks versus in response to contextual cues (such as in the sneezing example above).

Individuals with AOS may also produce relatively fluent, error-free stretches of speech from time to time. Such *islands* of error-free speech do not alleviate the speech difficulties, however, and tend to be relatively infrequent. Again, this feature is based on clinical observations rather than controlled study, but such observations are important

because they indicate that the problem in AOS is not one of neuromuscular impairment (which would produce relatively consistent problems) but rather lies at a higher level of motor control and programming.

Speakers with AOS are generally quite aware of the errors they produce, as evidenced for example by repeated attempts at self-correction. Such error awareness suggests intact perceptual abilities as well as error-detection abilities. However, there has been relatively little systematic study of auditory or somatosensory perceptual abilities in AOS, nor of error-detection capabilities. Early studies of auditory perception in AOS showed that some speakers with AOS did not differ from controls on phoneme discrimination (Aten *et al*., 1971; Johns and Darley, 1970), although a majority of the patients with AOS in these studies did perform below normal limits (Martin, 1975). Nonetheless, Aten *et al*. (1971) reported that there was no rank order correlation between speech production and speech perception in their group of patients, suggesting that any concomitant perceptual deficits are not related to the speech production deficit, at least at the level of phonemic discrimination. Square, Darley, and Sommers (1981) reported that their four individuals with *pure* AOS did not differ from unimpaired speakers on a wide range of auditory perceptual tests, in contrast to individuals with aphasia (with or without AOS). Thus, auditory perceptual deficits do not seem to be the cause of AOS.

However, virtually all studies of auditory speech perception in AOS have utilized tasks in which phonemes had to be distinguished from other phonemes (Aten *et al*., 1971; Johns and Darley, 1970; Shankweiler and Harris, 1966; Square *et al*., 1981),

but there has been very little systematic study of auditory perceptual abilities at the subphonemic level. An early study by Blumstein *et al*. (1977) did report difficulties in identification and/or discrimination of synthesized plosives along a voice onset time continuum in some patients with *anterior aphasia*. However, although it is probable that some of these patients exhibited AOS, the description of these individuals was insufficient to be confident that this was the case. More recently, Jacks (2006) examined vowel perception in AOS using a synthetic vowel continuum. Jacks observed that four out of five individuals with AOS and mild concomitant aphasia showed difficulties in vowel identification, although vowel discrimination was intact. No correlation between vowel perception and vowel production was observed in this study, further supporting the notion that auditory perceptual impairments are not entirely responsible for the speech production impairments in AOS.

With respect to somatosensory perceptual abilities, only a few studies have been reported (e.g., Deutsch, 1981; Rosenbek *et al*., 1973), despite recurring speculations that inability or inefficiency in processing somatosensory feedback may in part underlie the deficit in AOS (Bartle *et al*., 2007a,b; Kent and Rosenbek, 1983). Rosenbek *et al.* (1973) conducted a study with 30 individuals with AOS (as defined by the presence of numerous phonemic errors, inconsistent errors, increasing errors with increasing word length, and/or effortful speech during volitional speech), 10 individuals with aphasia but without AOS, and 30 unimpaired adults. Three tests of oral sensory perception were used, including oral form identification (requiring identification of geometrical shapes of objects placed in the mouth), two-point discrimination (requiring a response

indicating perception of one or two points on upper lip, tongue tip, and tongue blade), and mandibular kinesthesia (requiring a response indicating whether passively placed mandibular position was larger or smaller than a reference opening width). Results showed that as a group, the AOS individuals differed from unimpaired and aphasic individuals on all three orosensory perception tasks (aphasic individuals did not differ from controls). Although most individuals with AOS showed this pattern, there were also a few who fell within the normal range. Deutsch (1981) examined oral form identification in 18 individuals with errors consistent with AOS and varying degrees of concomitant aphasia, and whose sole lesions were either frontal or temporoparietal. Comparison against normative data from another study revealed that both groups demonstrated a deficit in oral sensory perception; there were no differences involving lesion location (posterior vs. frontal) or number of speech errors produced.

Using acoustic analysis in a biteblock paradigm with patients with Broca's aphasia (and possibly AOS), Baum (1999) observed immediate compensation to the biteblock for the patients as well as the controls, for both vowels and sibilant consonants. Although auditory feedback was also available, compensation was already evident at the earliest measurement point in the vowel/consonant, suggesting that perhaps somatosensory information was used to compensate for the perturbation. Unfortunately, it is unclear whether these individuals had AOS or where their lesions were, so the issue of somatosensory perception remains unsettled. However, Jacks (2006) reported that adults with AOS have similar patterns of decreased vowel accuracy when compared to non-impaired adults with the introduction of the bite block. This

suggests that feedback control in AOS patients remains relatively intact. In sum, the evidence to suggest that individuals with AOS may have somatosensory perceptual deficits is unclear, but the bite block studies seem to suggest the contrary.

3.1.3 Features indicative of other disorders

A number of characteristics that sometimes co-occur with AOS are indicative of different disorders. There is evidence to suggest that at least some speakers with AOS retain the ability to apply phonological rules such as vowel shortening with increasing word length (e.g., 'thick-thicken-thickening'; Collins *et al*., 1983; Haley and Overton, 2001; Hough and Klich, 1998; Strand and McNeil, 1996) and vowel lengthening depending on the voicing status of the following consonant (e.g., 'bat-bad'; Haley, 2004; Rogers, 1997; Tuller, 1984). In addition, a difference between receptive and expressive speech/language abilities, while often present in AOS, is considered more indicative of dysarthria.

The presence of apraxic behaviors in limb movements or in nonspeech oral movements suggests limb apraxia or oral apraxia, respectively, and is not considered to be part of AOS. However, it is possible that with more refined measures of nonspeech motor performance, a common deficit underlying AOS and nonspeech oral or limb apraxia may be revealed for some individuals (Ballard *et al*., 2000; Ballard and Robin, 2007). In fact, there is growing evidence that individuals with AOS have consistent difficulties in certain nonspeech motor tasks such as visuomotor tracking (Ballard *et al*., 2000; Ballard and Robin, 2007; Hageman *et al*., 1994). For example, Ballard and Robin (2007) found that their individuals with AOS produced greater spatial and temporal

error than controls when tracking predictable target patterns (i.e. sinusoid waves of varying frequency) but not when tracking unpredictable patterns. Further, removal of the target pattern (effectively removing feedback of accuracy) resulted in reduced temporal differentiation between predictable target patterns for the AOS group relative to the controls. Feedback was beneficial for matching movement amplitude but not for movement timing. The authors interpret these findings in terms of impairment in the ability to develop stable, accurate motor programs, perhaps in part due to an inability to rapidly integrate on-line sensory information with the motor commands for the movement pattern (i.e. poor feedforward commands due to difficulty incorporating corrective commands). Although some have argued that such nonspeech deficits can be independent from speech deficits (Ziegler, 2003), others have suggested that AOS may be a manifestation in speech of a more general underlying deficit in motor control (Ballard *et al*., 2003), and that careful study of motor tasks that share properties with speech may shed light on the underlying deficit in this motor speech disorder.

The severity of AOS varies widely, from almost complete inability to produce speech to mild slowing of speech with minor sound distortions. Specific profiles of the deficit vary as well, ranging anywhere from specific articulatory symptoms to specific prosodic disturbances (Wertz *et al*., 1984). Such heterogeneity among the group of individuals considered to have AOS complicates both the understanding of the deficit as well as its potential contribution to understanding the nature and neurobiology of speech motor control. However, with increasingly refined neural computational models,

imaging techniques, and careful assessment and description of speech behaviors in AOS, such goals are becoming more attainable.

Finally, like diagnosis, treatment of AOS is behaviorally oriented. Treatment for AOS uses a symptom-based approach that generally focuses on articulatory placement and rate/rhythm control techniques (Wambaugh *et al*., 2006b). A recent review of the treatment literature for AOS concludes that individuals with AOS can learn to compensate for their impairments and thus benefit from treatment (Wambaugh *et al*., 2006b), although it has been noted that generalization to untrained sounds is often limited (Wambaugh and Doyle, 1994; Wambaugh and Nessler, 2004), and it often takes many sessions and trials to achieve reliable production of accurate speech targets. In addition, treatment is not effective for all people with AOS, but the factors that predict whether a given individual will benefit from treatment are still poorly understood.

3.2 AOS lesion localization

Defining brain lesions responsible for AOS has been a difficult endeavor confounded by several factors. As discussed in the previous section, AOS often co-occurs with other speech disorders. Thus, it is often difficult to dissociate not only the symptoms for each disorder, but it is also challenging to determine which areas of damage result in which symptoms. Additionally, lesions that result from stroke, as is often the case in AOS patients, are usually significant in that damage extends to different brain areas. Other confounding factors include group and individual functional and anatomical brain differences, etiology, time since lesion onset, lesion localization method used, and diagnostic criteria.

One additional, but important problem is that many different regions in the left frontal and parietal areas and their subcortical connections have been found to be damaged in patients with AOS. These regions are in close proximity to each other and in some cases are overlapping. Square, Roy, and Martin (1997) hypothesized that AOS could result from lesions to six regions within the frontal parietal and subcortical circuitry involved in speech production. These regions included: nonprimary motor cortex (Brodmann's area or BA 6), the inferior frontal gyrus pars opercularis (BA 44), white matter underlying Broca's area, the anterior insula, lenticular nucleus, and midparietal cortex. The parietal lesions in this study did not result in slowness of speech and thus the authors concluded that the AOS symptoms observed were more likely due to remote effects on frontal lobes (Square *et al*., 1997). Many other authors have also found that parietal lobe lesions are more than likely not associated with AOS (Dronkers, 1996; Hillis *et al*., 2004; Ogar *et al*., 2006). Therefore, this chapter will focus on several key regions within frontal areas that are most likely associated with the functions that are lost in AOS patients including the left inferior frontal gyrus (IFG), the left inferior frontal sulcus (IFS), the left frontal operculum (FO), the left anterior insula, and the left ventral premotor cortex.

3.2.1 Left inferior frontal gyrus

The left inferior frontal gyrus (IFG) contains Broca's area (BA 44/45), a cytoarchitecturally defined region in the dominant hemisphere[1]. Historically, Broca's area has been defined as the speech production element of the speech-language circuit in

[1] For the purposes of this chapter, the dominant hemisphere is assumed to be the left.

that damage to this region results in speech production deficits (Broca, 1861; Geschwind, 1967). Broca's area is considered an interface between speech perception and production as well as between basic functions of perceptual sequencing, action understanding, and imitation (Nishitani *et al*., 2005). Broca's area is proposed to be the storage site and access point for motor plans or programs for gestures or speech segments such as words, syllables, and phonemes (McNeil *et al*., 2000; Kimura, 1993; Kimura and Watson, 1989) where activation is more pronounced during speech rather than single-vowel production, suggesting its role in motor planning and preparation (Özdemir *et al*., 2006).

A few very early surgical excision studies indicated that Broca's area may not be as important for speech output as was initially thought. Mettler (1949) surgically excised areas BA 44 and BA 45 bilaterally in a patient and found that the patient, who was schizophrenic and mute before surgery, was able to speak a few months after the procedure with articulatory problems. However, few details regarding the speech of this patient were documented. A second patient with only BA 44 removed bilaterally had no detectable change in speech. Zangwill (1975) reported two patients in whom tumors were removed, resulting in removal of Broca's area, and/or damage to the fibers underneath. Both patients showed an initial transient speech disorder that was resolved within a month. Dejerine (1913) and von Monakow (1914) also evaluated the effect of lesions to Broca's area on speech, and came to the conclusion that lesions restricted to this region would result in a transitory effect whereas more extensive involvement

possibly into the anterior insula or operculum would result in a longer lasting deficit in speech.

Many neuroimaging studies have shown activation in Broca's area during a variety of speech perception and production tasks including naming (Salmelin *et al*., 1994), phonology judgments (Heim *et al*., 2003, Thierry *et al*., 1999), syntax judgments (Ben-Shachar *et al*., 2003; Friederici and Kotz, 2003; Friederici *et al*., 2003, Heim *et al*., 2003), phonetic monitoring tasks (Demonet *et al*., 1992; Zatorre *et al*., 1996), internal speech (Hinke *et al*., 1993), verb generation (Raichle, 1996), verbal fluency (Phelps *et al*., 1997; Schlösser *et al*., 1998), acquisition of grammatical rules, discrimination of speech sounds, word production, time interval estimation, and rhythm reproduction (Bookheimer, 2002). Activation is also found when subjects perform various sub-lexical tasks involving auditorily presented speech particularly when the task involves explicit segmentation (Burton *et al*., 2000; Zatorre *et al*., 1992, 1996). In addition, Broca's area can be active during reading tasks including reading visually presented words (Petersen and Fiez, 1993), when saying action words associated with visually presented written names of objects (Martin *et al*., 1995), sentence reading (Müller *et al*., 1997), and during silent reading (Bookheimer *et al*., 1995).

Broca's area may play a role in general motor planning and/or programming such that it may contain the representations for storage and access of motor plans and programs for limb gestures as well as speech segments (Luria, 1966). Broca's area (and the likely homologous area in monkey, F5) is hypothesized to contain mirror neurons which are activated both by watching a task being performed and while performing the

task (Rizzolatti and Arbib, 1998; Rizzolatti *et al*., 1996; Kent, 2004). Furthermore, both overt and covert motor imitation can engage Broca's area (Iacoboni *et al*., 1999; Bookheimer, 2002), and Broca's area is activated both during speech perception and production (Okada and Hickok, 2006). Some studies have shown preferential activation for both imitation and observation in IFG using TMS (Transcranial Magnetic Stimulation; Heiser *et al*., 2003) and fMRI (Koski *et al*., 2002). Hickok and Poeppel (2000, 2004) propose that this mirror property of the left inferior frontal gyrus plays a critical role in speech production by creating a link between perception and production in the auditory dorsal stream where sounds are mapped onto articulatory-based representations needed for auditory-motor integration.

Nishitani *et al*. (2005) suggested that Broca's area is a mosaic of functions such that there may be many subsystems within Broca's area involved in different functions including speech production and language, perceptual sequencing, action understanding, and imitation, just to name a few. The two major subregions of Broca's area are the pars opercularis (BA 44) and the pars triangularis (BA 45), which can be clearly identified based on anatomical landmarks, cytoarchitecture, and function. The pars opercularis (BA 44) is bordered by the inferior frontal sulcus superiorly and the anterior vertical (or ascending) ramus of the Sylvian fissure anteriorly, and immediately anterior to the ventralmost section of the precentral gyrus (Amunts *et al*., 1999; Mazziotta *et al*., 2001; Foundas *et al*., 1998; Keller *et al*., 2007). The pars triangularis (BA 45) is bounded by the inferior frontal sulcus dorsally, the anterior horizontal ramus inferiorally, and the anterior ascending ramus caudally (Mazziotta *et al*., 2001; Keller *et al*., 2007).

TMS studies have been used to differentiate the functions of BA 44 and BA 45. When TMS is applied to the anterior portion of Broca's area (BA 47/45) semantic processing is disturbed, thus suggesting that the anterior regions play a role in semantic processing (Gough *et al*., 2005). In contrast, when TMS is applied to the more posterior areas (BA 44), phonological processing is impaired (Nixon *et al*., 2004; Gough *et al*., 2005).

Anterior regions (includes pars orbitalis, BA 47, and pars triangularis, BA 45) are most often engaged during semantic tasks. Various semantic tasks have been shown to activate the anterior regions of BA 45 and BA 47 including tasks such as categorization, semantic generation, concrete/abstract judgments, and judgments of semantic similarity (Burton, 2001; Bookheimer, 2002; Buckner *et al*., 1995; Fiez, 1997; Petersen *et al*., 1989; Demonet *et al*., 1992; Kapur *et al*., 1994; Gitelman *et al*., 2005; Poldrack *et al*., 1999; Martin, 2003) and during word-level semantic processing such as making semantic decisions about words (Demb *et al*., 1995; Gabrieli *et al*., 1996; Kapaur *et al*., 1994; Wagner *et al*., 1998) or generating words based on semantic relationships (Klein *et al*., 1995; Petersen *et al*., 1988). Although rare, it is possible for language comprehension to be impaired by lesions in BA 47 (Dronkers *et al*., 2004). This sparing may be because the anterior region serves as an executive system for semantics in that it retrieves, stores, and manipulates semantic representations (Martin, 2003; Roskies *et al*., 2001; Fiez, 1997).

Posterior regions of Broca's area (including pars opercularis, BA44) are engaged during phonological tasks (Burton, 2001; Bookheimer, 2002; Buckner *et al*., 1995; Fiez,

1997; Demonet *et al*., 1992; Zatorre *et al*., 1992; Rumsey *et al*., 1997; Poldrack *et al*., 1999; Martin, 2003). These tasks include subvocal articulation (Demonet *et al*., 1994; Burton *et al*., 2000; Zatorre *et al*., 1992, 1996; Fiez *et al*., 1995; Thierry *et al*., 1999), reading of linguistically complex sentences (Friederici *et al*., 2006), phonologically based working memory (Nixon *et al*., 2004), and discrimination of temporal cues (Zaehle *et al*., 2008). BA 44 has been found to be activated also by orthographic decision making tasks but with more intense and widespread activation during phonological tasks (Rumsey *et al*., 1997) as well as during both propositional (self-contained speech) and non-propositional speech (consisting of overlearned sequences; Blank *et al*., 2002).

A dorsal-ventral functional asymmetry within IFG and Broca's area has recently begun to be studied. Dorsal regions are more activated by phonological processing of words and nonwords in a syllable counting task with greater activation for nonwords than words (Poldrack *et al*., 1999). Phonological working memory tasks activated the junction of the precentral sulcus and posterior IFG within the dorsal portion of the pars opercularis region (Gitelman *et al*., 2005). Dorsal and slightly more posterior regions (superior pars opercularis) exhibited activation that correlated with performance in delayed serial recall task of words and nonwords (Chein and Fiez, 2001; Chein *et al*., 2002). The left posterior superior IFG is activated during identification of sound units of language within a word (Burton *et al*., 2000; Demonet *et al*., 1992; 1994; Zatorre *et al*., 1992; 1996; Thierry *et al*., 1999; Paulesu *et al*., 1997). The superior portion of the IFG has been found to be more responsive to sequence rather than stimulus structure in both

matching and sequence manipulation tasks using both syllables and hummed notes (Gelfand and Bookheimer, 2003). On the other hand, the ventral region of BA 45 showed activations in a serial word/nonword recall task but was more sensitive to the nonword condition (Chein and Fiez, 2001). These studies suggest that the more dorsal regions may be involved more in phonological judgments and planning of motor programs whereas the more ventral region may be associated with activations of sound units.

The functional asymmetry along the dorsal-ventral axis of the inferior frontal gyrus has been further supported by Molnar-Szakacs *et al*. (2005) also dissociated functions along the dorsal-ventral axis when looking at brain activations that correlate with performing finger movements and observing finger movements were evaluated using fMRI. They found two peaks of activation in the pars opercularis for imitation: one in the dorsal and the other in the ventral region. The dorsal area was also activated during action observation. The pars triangularis was found to be activated during observation but not imitation. This suggests that the dorsal section of Broca's area contains a mirror neuron like system where activations are seen for both imitative actions and observations. The ventral section is only activated during imitation, suggesting that it may contain the motor representations of actions. Although this study did not look directly at speech, it can be posited that this type of functional specialization can be extended to speech because speech production and comprehension can be considered a highly developed form of action execution/observation matching (Nishitani *et al*., 2005). Furthermore, Chein *et al*. (2002) found a similar dorsal-ventral

asymmetry in that the dorsal region activations correlated more with task difficulty in verbal working memory whereas the ventral region was more sensitive to the lexical status of letter strings, further supporting the notion that this functional segregation within the dorsal-ventral axis of the pars opercularis is also found for speech. This dorsal region borders the inferior frontal sulcus whereas the ventral region lies in close proximity to the frontal operculum and the anterior insula. These regions will be evaluated in the following sections.

3.2.2 Left inferior frontal sulcus

The left inferior frontal sulcus (IFS) constitutes the superior border of Broca's area. This region has been found to be activated during phonological processing of words and pseudowords in a syllable counting task (Poldrack *et al*., 1999), in phoneme processing (Burton *et al*., 2000; Joanisse and Gati, 2003), and in processing of fast acoustic transients (Joanisse and Gati, 2003; Johnsrude *et al*., 1997; Poldrack *et al*., 2001). In addition, the left IFS has also been shown to be activated when subjects were asked to perform a semantic task aimed at probing knowledge associations between visually presented words (Vandenberghe *et al*., 1996), during tasks aimed at measuring activations corresponding to cognitive control in various tasks making use of visually presented digits and letters, and during stroop-like tasks (Derrfuss *et al*., 2004).

The posterior superior portion of IFG extending across the left IFS and into the left middle frontal gyrus showed increased activation when subjects made perceptual judgments regarding temporal order in the sequence of phonological and nonlinguistic vocal information (hummed notes) independent of whether the stimuli were linguistic in

nature (Gelfand and Bookheimer, 2003). The authors thus conclude that this region may be involved in sequencing that is not necessarily specific to phonemic content.

Bohland and Guenther (2006) found that by varying complexity of both syllables and syllable sequences, the IFS responds more for complex rather than simple sequences and that the response was larger when the sequences were composed of complex syllables. However, this response did not show a main effect of syllable complexity. The authors hypothesize, based on this evidence, that the IFS may hold a representation of the forthcoming utterance in a sequence and that the larger activations may correlate with larger representations for complexity.

These studies taken together provide evidence that the IFS may contribute to the functioning of the dorsal region of the IFG in phonological judgments and sequencing and planning of motor speech sound units. Thus, this functional region may extend from the dorsal IFG across the IFS and potentially into surrounding white matter regions.

3.2.3 Left frontal operculum

The left frontal operculum (FO) extends from the inferior portion of Broca's area anteriorly to ventral premotor cortex posteriorally in the superior bank of the Sylvian fissure. Both the left frontal operculum and adjoining anterior insula have been found to be important for speech production (Paulesu *et al*., 1993; Fiez and Petersen, 1998; Wise *et al*., 1999; Fox *et al*., 2000). FO has been considered to be an intersection between the cerebral system for language and the system for motor speech activity (Ojemann and Whitaker, 1978) because it contains terminal projections from the supplementary motor area (Jürgens, 1987) and bidirectional connections with posterior areas for language

including the superior temporal gyrus and neighboring areas of the inferior parietal lobe (Galaburda and Pandya, 1982). Stimulating FO can result in language impairments and speech arrest (Penfield and Roberts, 1959; Ojemann and Whitaker, 1978). In addition, Alexander *et al*., (1990) hypothesized based on an aphasic lesion study that FO integrates facilitatory, semantic, and motor programming aspects of language.

The left FO is engaged during both initiation and suppression of appropriate word choices in a sentence completion task (Nathaniel-James *et al*., 1997), during covert articulation (Rueckert *et al*., 1994), during inconsistent written and pronounced word tasks (e.g. 'lint' is not consistent when presented with 'pint', but is consistent when presented with 'hint'; Fiez and Petersen, 1998), during sentence reading when syntactic violations were encountered (Friederici *et al*., 2006), and in learning nonnative phonetic contrasts (Golestani and Zatorre, 2004). Indefrey *et al*. (2001) found that the left FO was activated based on the complexity of syntactic encoding during a restrictive scene description task, suggesting that this region may play a role in sentence-level and phrase-level syntactic encoding during speaking.

Bohland and Guenther (2006) have shown that the junction between the FO and the anterior insula bilaterally increases responsiveness for additional sequence or syllable activity. In their study, a strong interaction between sequence and syllable was found for activity in this region. The authors hypothesize that FO is involved in the representation of the speech plan at some level. Furthermore, it may be responsible for integrating lower level aspects of the speech motor plan with more abstract representations of speech sounds used in sequence planning.

Stimulation studies have shown that voicing and some aspects of speech control at a syllabic or phonemic level are mediated to some extent by this region (Kimura, 1993). In addition, stimulation of this region causes a higher frequency of arrest and speech hesitation than when its posterior counterpart is stimulated (Penfield and Roberts, 1959). Kimura (1993) theorizes that this region has more direct connections with descending pulses to control speech musculature. Thus, consistent with the role of the more inferior portions of IFG in holding representations of speech sound units or motor plans, the left FO may also play a role in this function.

3.2.4 Left anterior insula

The insula is the only cerebral lobe which cannot be seen on the surface of the brain because it is covered by the operculum (Guenot *et al*., 2004). There is mounting literature supporting the role of the left anterior insula in speech and language, but much of this work is relatively new and at times contradictory (Bamiou *et al*., 2003). However, the insula is most well known as being a paralimbic structure responsible for many visceral sensory functions including taste and olfaction; gastric motor functions including respirational and gastrointestinal; and cardiovascular control (Augustine, 1996). This evidence may suggest that the role of the insula in speech is more related to motivation, saliency, task demands, and/or emotions rather than to speech motor planning per se.

Patients with lesions to the insula have difficulties pronouncing phonemes in the proper order (Dronkers *et al*., 2000). Ojemann and Whitaker (1978) found that stimulation of the insula of epileptic patients during surgery leads to word finding

errors. In addition, several studies have observed insular damage in patients with conduction and Broca's aphasia, specifically when articulatory errors are prominent (Damasio and Damasio, 1980; Marie, 1906; Mazzocchi and Vignolo, 1979; and Vanier and Caplan, 1990). Paulesu *et al*. (1993) found increased blood flow in the insula during speech tasks. Phonological decision making when subjects were asked to compare pronunciations between real words and pseudowords, perform subvocal articulatory coding, and hold the speech item in working memory activated the border between BA 44 and the insula (Rumsey *et al*., 1997). The left anterior insula shows activation during both propositional and non-propositional speech (Blank *et al*., 2002). Wise *et al*. (1999) used functional neuroimaging to determine that activation was observed in the left anterior insula and lateral premotor cortex during articulation planning. However the data presented in both Blank *et al*. (2002) and Wise *et al*. (1999) are unclear as to whether activations are limited to the insula or also extend into the FO.

Riecker *et al*. (2000a) offered further evidence for the role of the insula in speech when they determined that the anterior insula was activated during overt and not covert speech, suggesting that this activation was more closely tied to speech execution and coordination rather than planning. The researchers hypothesized that the insula supports a checking mechanism between the phonetic code and the articulation module in that it is the last stage of checking before speech execution (Dogil *et al*., 2001; Riecker *et al*., 2000a; Ackermann and Riecker, 2004). However, in a study done by the same lab, it was found that repetition of stimuli of varying complexity (consonant vowel syllables CVs, CCCVs, CVCVCV non-word sequences, and CVCVCV words) did not

result in activation of the insula (Riecker *et al*., 2000b). On the other hand, Sörös *et al*. (2006) found no significant activation of the insula during a repetition task.

Bohland and Guenther (2006) showed activation in the left anterior insula in overt production of syllables that was correlated with increasing syllable complexity but not increasing syllable sequence complexity. Thus, the authors concluded that this region of the insula may be involved in overt speech rather than speech sequencing.

Nota and Honda (2003) provide an explanation for the ambiguity in the above-mentioned articulation results. They point out that the activation of the left anterior insula in articulation is largely dependent on the tasks used. This study evaluated brain activations of normal subjects performing two different session types each consisting of three parts: (1) producing speech based on a visual cue of 1, 2, or 3 corresponding to an utterance type, (2) a control condition in which the subjects were asked to think about the cue of 1, 2, or 3, and (3) a rest state in which the subjects stared at a fixation point. In the first session type was random in which there was a random presentation of cues and the second was a repetition type in which one cue was presented repeatedly. The left anterior insula was activated during the first session type (random) whereas it was not activated during the second session type (repetition).

The Nota and Honda (2003) repetition session followed a paradigm similar to the Riecker *et al*. (2000b). However, Nota and Honda (2003) showed that repetition does not activate the left anterior insula. The authors theorize that the left anterior insula plays a role in encoding and storing phonetic plans in the articulatory buffer rather than a more general role in articulation planning.

On the other hand, some authors have found that the insula may not be as important to speech production. For example, Bonilha *et al*. (2006) found that when subjects were asked to produce oral mouth movements and CV syllables, BA 44 was activated more during the speech movements than the nonspeech movements and the insula was not activated during speech movements at all. In addition, Duffau *et al*. (2001) found that removal of the entire left insula in a patient with a benign brain tumor did not elicit AOS. In a study of electrical stimulation of the insular cortex, it was found that speech disturbances (defined as speech arrest or blurred speech) were reported in less than 25% of the 14 patients when the insula was stimulated in 27 sites spreading across the anterior and posterior insular cortices in both the left and right hemisphere (Ostrowsky *et al*., 2000).

Nevertheless, much of the evidence seems to suggest that the left anterior insula may be activated during speech production depending on the task. Some authors have hypothesized that this role is tied directly to speech articulation planning by encoding and storing phonetic plans (Nota and Honda, 2003; Wise *et al*., 1999) while others have offered a less convincing theory that it is simply activated by processing speech sounds in general (Riecker *et al*., 2000a; Bohland and Guenther, 2006). Of course, fundamentally, it is known that this structure is paralimbic in nature and it thus cannot be excluded that the activation seen in the insula may suggest arousal in response to the task paradigms or simply motivation for speaking (Habib *et al*., 1995).

3.2.5 Left ventral premotor cortex

The premotor cortex (BA 6) is typically concerned with external cue dependent motor planning. Damage to the premotor cortex can lead to dissolution of learned, skilled purposeful movements (Liepmann, 1977; Square *et al*., 1997). Interoperative brain stimulation of the ventral premotor cortex has been shown to disrupt speech articulation (Duffau *et al*., 2003) and TMS applied to the ventral premotor cortex markedly diminishes the number of correct syllables produced during overt speech (Tandon *et al*., 2003). The ventral premotor cortex activity increases when an uttered speech sound is composed of different syllables as opposed to the same one (Bohland and Guenther, 2006) and with increasing the length and the number of syllables of words/psuedowords (Alario *et al*., 2006).

Ziegler (2002) theorizes that the premotor cortex may be the storage site for complex speech motor patterns. For example, after a person learns to say a specific syllable and uses that syllable frequently in everyday speech, it becomes stored in the premotor cortex. In this way, a lesion to the premotor cortex would lead to a disintegration of learned articulatory gestures (Ziegler, 2002). This lesion would damage some portion of the stored representations so that the syllables would then have to be articulated in a phoneme by phoneme manner. This hypothesized impairment following premotor cortex lesions is consistent with the symptoms expressed by AOS and thus suggests that it may play a role in speech articulation and/or planning.

3.2.6 AOS lesion studies

TMS studies have been used to try to dissociate the functions of the previously discussed regions by artificially simulating damage caused by brain lesions. Unfortunately, TMS studies have had limited success in reproducing production deficits associated with AOS and aphasia (Stewart *et al*., 2001; Flitman *et al*., 1998; Epstein *et al*., 1996, 1999). In order to look more directly at the effect of damage or interruptions to brain areas that produce behaviors consistent with AOS, we must look more closely at case studies of patients with behaviors consistent with AOS.

Some earlier work done by Mohr found correlations between Broca's area and AOS. Mohr (1976) found that damage limited to Broca's area and adjacent insular tissue did not lead to full Broca's aphasia but that these patients could be characterized by imprecise control of voicing or articulatory positioning and dysprosody. Furthermore, the patients with full-blown Broca's aphasia had infarctions spanning the upper division of the middle cerebral artery (MCA) and extending into the insula. Mohr *et al*. (1978) also found in postmortem analysis that infarctions to left anterior insula and pars opercularis rendered patients severely dysarthric or transiently mute and that isolated lesions of Broca's area were not sufficient for long-lasting impairments unless the lesion sufficiently involved the adjacent opercular areas. In other words, the lesion needed sufficient depth into the frontal operculum and/or insula for long-lasting aphasic symptoms.

Marquardt and Sussman (1984) found similar lesions in AOS patients. They found a correlation between lesion volume and AOS in 15 patients diagnosed with

Broca's aphasia and AOS such that lesions were primarily located in the left frontal lobe, extended deep into the cortex, and that several lesions included the insular cortex and frontal operculum. Alexander *et al*. (1990) reported a patient with a lesion in the FO and middle frontal gyrus exhibited impaired articulation and prosody. However, in comparison to other aphasic patients, the authors conclude that damage must include more than FO for deficits to be permanent. In addition, Tanji *et al*. (2001) reported a case of pure anarthia (defined as an improper articulation of words or sentences) in which that patient, who was right hemisphere dominant, had a lesion restricted to the right precentral gyrus and frontal portion of the right insula, possibly extending into the right frontal operculum.

Dronkers (1996) published a widely cited lesion overlap study which implicated the left precentral gyrus of the insula in AOS. This study looked at CT and MRI scans of twenty-five stroke patients with AOS and nineteen patients without AOS. Some of the more common articulation errors observed in the apraxic patients included: phoneme substitutions (usually on initial consonants), greater difficulty producing longer words or sentences, and greater difficulty with repetition than reading aloud (Dronkers, 1996). Dronkers (1996) found that in the apraxic patients, there was 100% lesion overlap in the left precentral gyrus of the insula and no overlap in this region in the non-AOS subjects. From this double dissociation, Dronkers (1996) concluded that not only does damage to the left precentral gyrus of the insula induce AOS, but this structure may play an important role in articulatory planning. However, looking closely

at the results also shows an approximately 98% overlap of lesions in the premotor cortex.

Ogar *et al*. (2006) also found that the left anterior insula is the common area of infarction in patients with AOS. It is important to note that many of the same patients evaluated by Dronkers (1996) were also used in this study. Here, performance in motor speech tasks including vowel prolongation, sequential and alternating diadochokinesis, single and multiple repetition of multisyllabic words, single repetition of monosyllabic words, production of words increasing in length, sentence repetition, and reading the Grandfather Passage was compared to lesion site. The authors concluded that although the common area of infarction in these patients was the anterior insula, the severity and extent of the AOS symptoms positively correlated with the size of the lesion where lesions that incorporated regions outside of the insula resulted in more severe forms of both AOS and aphasia.

The role of the anterior insula in AOS is not as clearly understood as was presented in Dronkers (1996) and Ogar *et al.* (2006). To the contrary, McNeil *et al.* (1990) found that only two of the four patients they diagnosed with pure AOS had damage to the insula. Robin *et al*. (2008) also point out that the lesion overlay method used by both Dronkers (1996) and Ogar *et al*. (2006) may not be sufficient in determining correlations between lesions and behaviors in these patients. Many of the behavioral criteria used in these studies were not necessarily consistent with the currently accepted behavioral symptoms (as discussed in the previous section). Many of these patients may have also shown signs of dysarthria and aphasia and these patients

were not differentially analyzed. Hillis *et al.* (2004) also called into question the role of the anterior insula in speech movement planning. This study evaluated lesion sites in stroke patients and found that damage to the insula is fairly common in stroke patients because of its close proximity to the MCA. Fifty percent of ischemic MCA territory strokes damage the insula (Fink *et al*., 2005). Furthermore, in forty patients with AOS symptoms, there was no correlation between AOS and left anterior insula damage, but a correlation was found between AOS and lesions in the posterior IFG (Hillis *et al*., 2004).

It is difficult to dissociate the involvement of the left anterior insula from left frontal cortical areas in AOS. This may be largely due to their close anatomical proximity: the insula is buried in the lateral sulcus covered by the operculum (Augustine, 1996). Some studies have even found lesions in both areas: Marien *et al.* (2001) conducted a case study that looked at an 83 year old patient with a diagnosis of AOS and phonological agraphia (phonological agraphia is a disorder of the spelling system in which the patient is unable to spell novel words and pronounceable nonwords). This patient had a lesion localized to the left intrasylvian frontal opercular cortex and possibly extending into the adjacent part of the left anterior insula. At onset, this patient exhibited mutism and 1 year post stroke, was able to greatly improve: the AOS and the phonological agraphia had disappeared.

There have been a few studies that have evaluated subcortical damage resulting in AOS and aphasia (Kertesz, 1984; Marquardt and Sussman, 1984; Alexander *et al*., 1987; Robin and Schinberg, 1990). Peach and Tonkovich (2004) reported a case study

of a patient with a basal ganglia hemorrhage and behavioral symptoms consistent with AOS including phoneme substitution errors most often occurring in the initial position of words as well as dysprosody and slow speech rate. However, this patient's lesion did extend into frontal regions. The authors reported that this damage did not extend into the lower primary motor cortex, the posterior portion of the IFG, the insula, or the postcentral opercular structures. Nevertheless, it is difficult to determine whether or not the speech deficit seen in this patient is due to the subcortical damage or the frontal damage that may have extended into more superior areas of the IFG or the IFS. Moreover, the description of the behavioral features of this patient makes it difficult to determine whether this patient had AOS (e.g., the patient made substitution errors but no mention is made of sound distortions and dysfluency appears to be defined in terms of grammatical structure rather than speech rate).

One interesting piece of information has begun to emerge in this lesion literature. Notably, the lesion responsible for AOS shows a correlation between its extent and depth and the severity or persistence of AOS symptoms (Mohr, 1976; Mohr *et al*., 1978; Ogar *et al*., 2006; Alexander *et al*., 1990). Furthermore, many of these lesions are reported to extend from the left IFG into the closely lying frontal opercular area, into the anterior insula, or into the IFS and surrounding regions.

3.2.7 Neurodegenerative and progressive cases

Most of the chapter has discussed brain lesions resulting from stroke. There is another class of brain lesions that have been found to cause AOS symptoms and these are the lesions associated with neurodegenerative diseases. In fact, AOS is often identified as

one of the first symptoms of neurodegenerative diseases such as corticobasal degeneration or non-fluent progressive aphasia (Josephs *et al*., 2006; Rosenfield, 1991; Blake *et al*., 2003; Gorno-Tempini *et al*., 2004a,b). Therefore, it is also important to include a discussion of the studies of progressive AOS symptoms possibly related to neurodegeneration and brain atrophy associated with those cases.

Nestor *et al*. (2003) looked at patients diagnosed with progressive non-fluent aphasia who present with similar articulation problems as seen in AOS patients. Progressive non-fluent aphasia (PNFA) occurs when "patients lose the ability to communicate fluently in the context of relative preservation of single word comprehension and non-linguistic cognitive abilities" (Nestor *et al*., 2003). Using PET and compared to normal subjects, patients with PNFA symptoms had a resting hypometabolism centered in the left anterior insula and the frontal operculum.

Broussolle *et al*. (1996) reported case studies of eight patients with progressive speech production deficits that consisted of an initial combination of AOS, dysarthria, dysprosody, and orofacial apraxia. Many of these patients progressed to mutism within 6-10 years. In all patients there was some degree of abnormality in the left inferior frontal region ranging from decreased blood flow to severe atrophy. Five of the patients had distinct atrophy to the left frontal operculum and two patients presented with bilateral opercular syndrome. This suggests that the frontal operculum atrophy contributed to the symptoms seen in these patients.

A four year longitudinal case study of a patient with corticobasal degeneration (CBD) evaluated the progression of the disease, the behavior symptoms, as well as the

atrophy in the brain (Gorno-Tempini *et al*., 2004b; Sanchez-Valle *et al*., 2006). A mild AOS was observed in the patient several years after the onset of symptoms and progressed to mutism over the next two years. Before AOS was present, CT and MRI scans showed minor atrophy in the inferior frontal regions. By the time the AOS symptoms were present, the atrophy had progressed to be significant with respect to controls in the pars opercularis of the left inferior frontal gyrus and the left insula. By the time the patient was mute, the atrophy had spread to the more anterior regions of the left insula and the medial frontal lobe. The appearance of AOS symptoms coincided with the atrophy to the pars opercularis and the left insula.

These progressive AOS cases associated with neurodegeneration are important to consider when evaluating the brain areas associated with AOS. The relatively slow onset of distinct symptoms coupled with cerebral degeneration allows researchers to gain a clearer understanding of how brain areas may be damaged, what the effect of that damage is on behaviors such as speech and how behavior symptoms change as atrophy increases.

This analysis of the neurophysiological data of brain areas that may be implicated in speech production and AOS is not exhaustive. There exists a large literature supporting the role of one brain area or another. The close proximity of the frontal areas discussed in this review suggests that they may have similar functionality in speech production and thus in AOS. Miller (2002) stated that just because it appears that lesions are linked to AOS, it does not establish a causal link. In order to provide that link, one must establish the function of the brain regions that may be impaired in

AOS. This link has not been convincingly created thus far in the lesion literature. In order to gain better understanding of how these brain areas may contribute to AOS, it is necessary to consider the functional modeling perspectives of AOS and how they relate to brain lesions and behavioral criteria. The next section will discuss these models.

3.3 Functional interpretations of AOS

As seen in the previous section, a lot of disagreement exists concerning the brain regions that are damaged in AOS patients and what the underlying function of those regions are in normal subjects. Speech production modeling may be able to provide an explanation of AOS that takes into account the underlying functions in the lesioned brain regions.

Models that provide details into phonological encoding and articulation will be able to provide an informative explanation of AOS. Two models, the Levelt model (Levelt, 2001; Levelt *et al*., 1999; Roelofs, 1997) and the DIVA model (Directions Into Velocities of Articulators; Guenther, 1995, 2001, 2006; Guenther and Ghosh, 2003; Guenther *et al*., 1998, 2003, 2006), provide the most convincing illustration of speech production mechanisms at a level that would offer insight into AOS. Both of these models attempt to computationally explain how speech sounds are produced. Although the Levelt model is able to explain how a speech sound is selected from the point of conceptualization, it is rather limited in its explanation of how the system controls the articulators. Conversely, the DIVA model provides a comprehensive treatment of articulator control, but it lacks the ability to explain speech sound selection at the level of linguistic representations.

The next two subsections will explore each of these models individually and attempt to explain AOS in terms of their architecture.

3.3.1 The Levelt model account of AOS

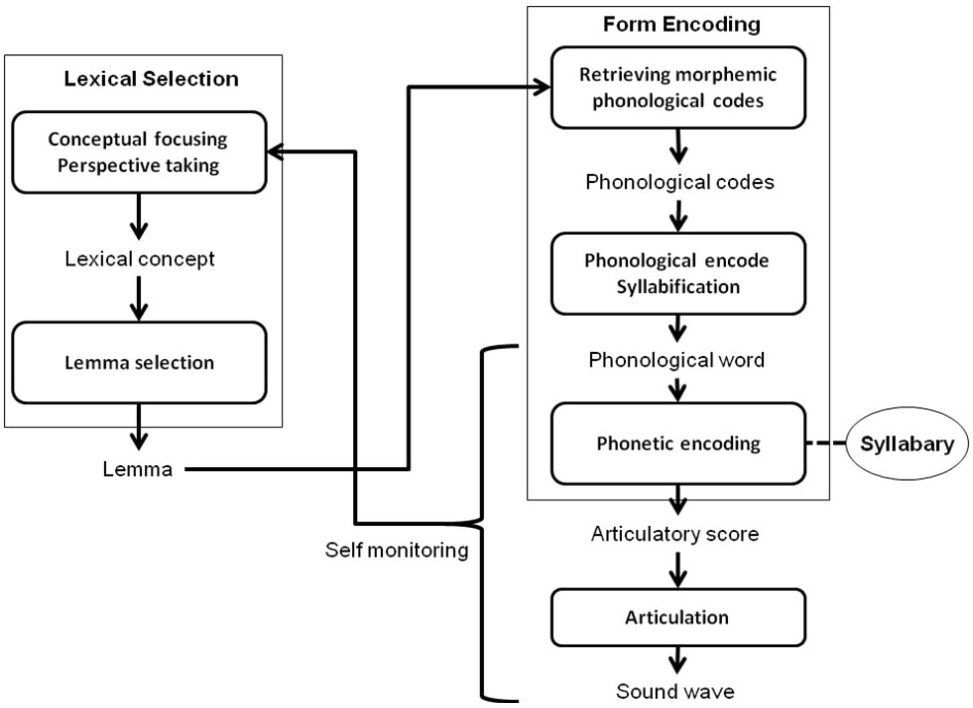


Figure 3.1. *Box diagram of the Levelt model.* [Reprinted with permission from Levelt *et al*., 1999]

The Levelt model represents a theory of lexical access in which speech production is described from the point of conceptualization of the target word to the initiation of articulation; see Figure 3.1. The computational framework for Levelt's model is embedded in the Weaver++ model, a spreading-activation-based network with a parallel object-oriented production system (Roelofs, 1997). The Levelt group has performed meta-analyses of neuroimaging studies to attempt to find neurological correlates of the different parts of the model. Most notably, the phonetic encoding, articulation, and

syllabification levels seem to overlap with the neurological data discussed in this chapter (Indefrey and Levelt, 2004; Levelt, 2001).

The model is marked by a serial two-system architecture: a *lexical selection system* and a *form encoding system*. The lexical selection system can be defined as the selection of the appropriate item from the lexicon. The form encoding system can be defined as the preparation of the item's articulatory shape (Levelt *et al*., 1999; Levelt, 2001). The lexical selection system has little involvement in AOS, so it will only be briefly introduced. The input to the model is called *perspective taking* in which the speaker focuses on concepts that will accomplish the communication goal (Levelt, 2001). From there, activation of all concepts that might accomplish the speaking goal occurs. For example, if a speaker wanted to say 'dalmatian', concepts such as 'dog' and 'animal' might also be able to accomplish the speaking goal. Thus all three concepts are activated and the target (ex. 'dalmatian') is selected through competition. The winning target is called a *lemma* and corresponds to not just the concept, but also its syntactic description (ex. number, gender, etc.; Levelt *et al*., 1999; Levelt, 1999, 2001). It is at this point that the form encoding system becomes activated.

Phonological encoding is the first stage of form encoding, which involves simultaneous but independent retrieval of segments (phonemes) and metrical frames that specify the number of syllables and the stressed syllable. Next, segments are inserted, left-to-right, into the slots of the frames to generate an output in terms of phonological words. A *phonological word* is a syllabified string of phonemes that contains one stressed syllable and a variable number of unstressed syllables. The second

stage is *phonetic encoding*, which is the model's speech motor programming stage. Phonetic encoding transforms the abstract phonological words into a specification of the speech motor programs (also called phonetic syllables), which are then temporarily stored in an articulatory buffer until all motor programs for the phonological word have been computed. Phonetic syllables are still relatively abstract, in that they specify articulatory goals (e.g., lip closure) rather than movement trajectories or positions. According to the model, phonetic syllables for infrequent syllables are constructed by concatenating phoneme-sized articulatory gestures, whereas phonetic syllables for frequent syllables are retrieved as prepackaged wholes from a memory store called the *mental syllabary* (Cholin *et al*., 2006; Laganaro and Alario, 2006; Levelt *et al*., 1999; Levelt and Wheeldon, 1994). The mental syllabary is one of the core assumptions in the model. It is the library of practiced articulatory gestures. The phonetic encoding level thus outputs an articulatory score which is a string of syllabic gestures that correspond to the target word.

Given the basic architecture of the Levelt model, what type of damage would predict AOS symptoms? First, this damage can be limited to the form encoding system because AOS does not involve a conceptual impairment. Within the form encoding system, this damage can be limited to the phonetic encoding stage such that damage occurs to 1) the representations in the syllabary or to the phoneme-sized motor programs used for producing low-frequency syllables, 2) the syllabary access mechanism, 3) the articulatory buffer which is involved in syllable sequencing, or 4) a combination of all three.

Within the context of the Levelt model, several specific accounts of AOS have been proposed, namely the *dual route hypothesis* (Varley and Whiteside, 2001a,b; Whiteside and Varley, 1998), the *reduced buffer capacity hypothesis* (Rogers and Storkel, 1999), and the *damaged motor program hypothesis* (Aichert and Ziegler, 2004). The dual route hypothesis places the locus of the deficit in the phonetic encoding stage of the model (Varley and Whiteside, 2001a,b; Whiteside and Varley 1998), specifically in the *direct* route of phonetic encoding (i.e., impaired syllabary access). Thus, speakers with AOS are forced to rely on the *indirect*, phoneme-by-phoneme route of phonetic encoding. This indirect route is more resource-intensive and error-prone, resulting in a greater number of speech errors, loss of automaticity, prolongations of segments, reduced coarticulation, increased variability, and reductions in speech rate.

In contrast, the reduced buffer capacity hypothesis places the locus of deficit in AOS at the level of the articulatory buffer (Rogers and Storkel, 1999). Specifically, AOS is thought to reflect a limitation of the buffer capacity to a single syllable, forcing speakers with AOS to program utterances one syllable at a time. This proposal accounts for core symptoms such as syllable segregation, dysprosody, and slow speech rate. Support for this hypothesis comes from a study in which speakers were asked to repeat as fast as possible two-word sequences such as ‘bug-pug’ and ‘bug-chug’. Feature similarity was varied in these pairs, such that some pairs involved a difference of only one feature (e.g., voicing: ‘bug-pug’) while others varied in several features (e.g., voicing, place, and manner: ‘bug-chug’).

Previous work had shown that when producing single words in list format, unimpaired speakers are slower to produce a target word whose initial phoneme shares many features with that of a preceding repeated prime word (‘bug’, ‘bug’, ‘bug’, ‘bug’, ‘bug’, ‘pug’) than when successive words share fewer features (‘bug’, ‘bug’, ‘bug’, ‘bug’, ‘bug’, ‘chug’; Rogers and Storkel, 1998). Rogers and Storkel (1998) argued that phonologically similar items cause interference in the reprogramming of the articulatory buffer. In their 1999 study, they used this finding to examine speech motor programming in AOS by asking speakers to repeatedly produce word pairs. Rogers and Storkel hypothesized that unimpaired speakers and aphasic speakers without AOS would be able to program both syllables and thus would not show the similarity effect on interword intervals (since no reprogramming would be required), but that speakers with AOS would show such similarity effects if they are unable to program both syllables and keep them in the articulatory buffer simultaneously. The results confirmed these predictions, suggesting impairment in maintaining more than one syllable or word in the articulatory buffer.

However, findings by Deger and Ziegler (2002) and by Maas *et al*. (2008) suggest that speakers with AOS are capable of programming more than one syllable at a time. For example, Maas *et al*. (2008) observed a sequence length effect on a reaction time measure that assesses preprogramming, suggesting that multiple syllables were being programmed by their patients with AOS. Deger and Ziegler (2002) observed longer simple reaction times in their AOS patients for sequences of different syllables

(e.g., ‘daba’) than for sequences of repeated syllables (e.g., ‘dada’), suggesting that at least two syllables resided in the buffer.

While both the dual route hypothesis and the reduced buffer capacity hypothesis can account for several characteristics of AOS, such as slow speech rate and segmentation, sound distortions are not easily captured. In addition, the dual route hypothesis predicts that syllable frequency should not affect performance in AOS, because all syllables would presumably have to be programmed using the indirect, phoneme-by-phoneme route of phonetic encoding. Evidence against this prediction comes from a study by Aichert and Ziegler (2004), who observed in a group of ten apraxic speakers that more frequent syllables resulted in fewer (perceptually defined) errors than did infrequent syllables, suggesting access to the syllabary. In addition, patients made more errors on consonant sequences when they were part of the same syllable, as in ‘star’, than when the consonants belonged to differed syllables, as in ‘mister’. This finding also argues against an interpretation of phoneme-by-phoneme programming, because such a programming strategy should be blind to higher level syllable structure. Aichert and Ziegler (2004) propose that rather than complete failure to access the syllabary, the representations themselves may be damaged, although the degree of damage may vary with syllable frequency and complexity. While this account can explain distortion errors in AOS and relatively consistent error type and location within the utterance, by itself it does not account for other features such as dysprosody and slow speech rate.

Thus, the Levelt model has been used to frame accounts of AOS, although at present none of these accounts appear capable of explaining all primary characteristics of the disorder. In part this state of affairs is due to underspecification and shortcomings of the model. First of all, the model is underspecified in terms of how the speech musculature is activated. This is particularly problematic when considering AOS because in AOS, there is a strict speech production impairment without conceptual or perceptual speech problems. Secondly, the syllabary holds distinct representations of learned syllables and not a continuum or distribution of learned syllables. Without a more distributed representation of the syllables, it is not clear how the model can account for the fact that the presence of an error is inconsistent in AOS. Finally, the Levelt model does not adequately address the issue of where in the brain the lesion responsible for AOS may reside. Taking into consideration the brain regions responsible for speech and AOS is an important part of thoroughly examining and understanding AOS.

3.3.2 The DIVA model account of AOS

Figure 3.2 schematizes the main components of the DIVA model, a neural network model of speech acquisition and production (Guenther, 1995, 2001, 2006; Guenther and Ghosh, 2003; Guenther *et al*., 1998, 2003, 2006). DIVA differentiates itself from the Levelt model by focusing on sensorimotor transformations underlying the control of articulator movements and thus focuses on the control of speech at the syllable and lower motor levels. In addition, DIVA differentiates itself from other models in that all

of its components have been associated with specific anatomical locations in the brain (see Guenther *et al*., 2006).

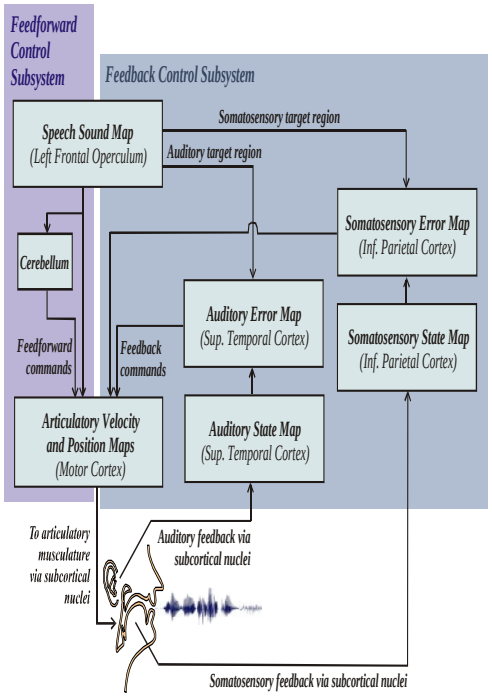


Figure 3.2. *Box diagram of the DIVA model of speech production.*

The production of a sound in DIVA begins with the activation of a cell in the *speech sound map* (SSM). This region is hypothesized to lie in the inferior posterior portion of Broca's area and/or the left frontal operculum and ventral premotor cortex. Each cell in the model can be described as a small population of neurons that fire together. These cells will be activated both when perceiving their corresponding speech sound and by producing the same sound. This property is also found in mirror neurons, which have been found in the frontal operculum or Broca's area homologue in monkeys, area F5 (Rizzolatti and Arbib, 1998; Rizzolatti *et al*., 1996). The speech sound itself can be

described as a phoneme, syllable, or word. For the sake of readability, the speech sound will be referred to as a syllable herein.

As currently implemented, each speech sound map (SSM) cell in the DIVA model represents an entire speech sound *chunk* that was presented to the model for learning. For example, if the model is presented with the syllable ‘ba’, a single SSM cell is chosen to represent that syllable. After learning, the model is capable of producing ‘ba’ in combination with any other learned speech sounds; however, it is incapable of using portions of ‘ba’, such as the vowel ‘a’, in other syllables. To learn ‘ta’, for example, the model has to learn a complete motor program for producing ‘ta’; it cannot reuse the portion of the motor program for ‘ba’ that specifies ‘a’ production. This inability stems from the fact that only a single SSM cell represents ‘ba’. In contrast, when infants start producing new words, they appear to prefer words that utilize previously learned production routines (McCune and Vihman, 2001; Schwartz, 1988), suggesting that they build new motor programs by incorporating previously learned programs.

The GODIVA model (Gradient Order DIVA; Bohland *et al*., in press) has introduced a mechanism for selecting these speech sounds. GODIVA adds higher-level representations for planned speech sounds and their serial order, and simulates various aspects of serial speech planning and production. Furthermore, the model follows recent instantiations of the DIVA model by proposing specific neuroanatomical substrates for its components. GODIVA contains a representation for sequencing and selection of SSM cells that is hypothesized to reside in the IFS. This IFS module contains two

layers: a *planning layer* for laying out the motor plan and sequence based on inputs from higher level cortical areas, and a *choice layer* used to select the appropriate SSM cells based on the motor plan and sequence.

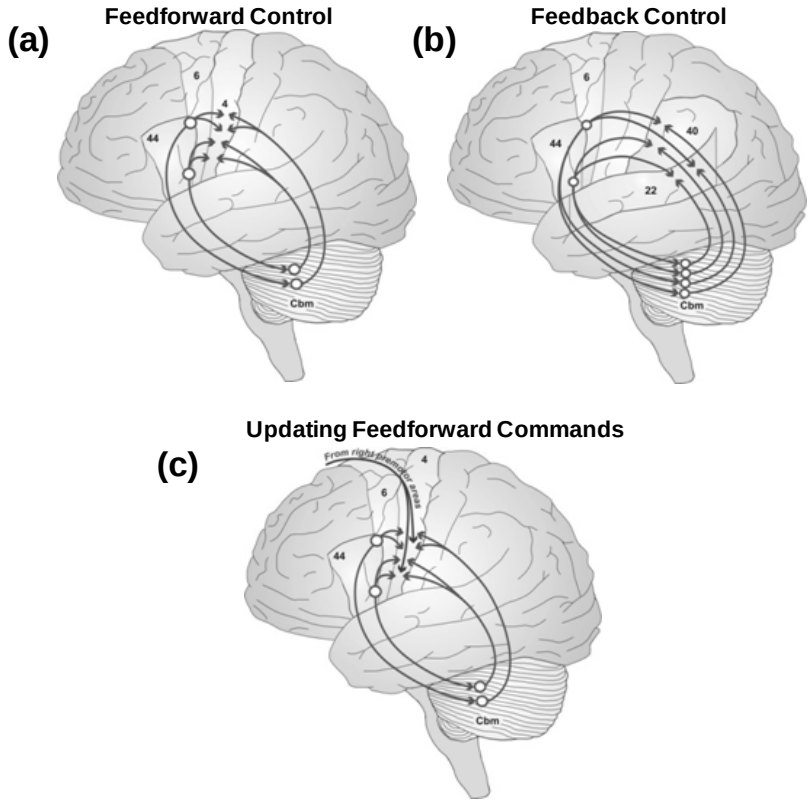


Figure 3.3. *Three different roles for SSM in the DIVA model*. The SSM is hypothesized to lie in left FO and/or the inferior posterior portion of the IFG and ventral premotor cortex. **(a)** Projections from speech sound map cells to primary motor cortex encode feedforward commands for speech sounds (syllables and phonemes). Damage to these projections results in an inability to access the feedforward commands (or motor programs) for speech sounds. **(b)** Projections to higher-order auditory and somatosensory cortices encode the sensory expectations for these sounds. Damage to these projections impairs the auditory and somatosensory feedback control systems for speech. **(c)** When the sensory feedback control systems send corrective commands to the primary motor cortex, the feedforward commands for the current sound, encoded by the synaptic projections to primary motor cortex, are updated to include these corrective commands. Damage to left FO and/or the inferior posterior portion of IFG and ventral premotor cortex thus impairs the ability to update the feedforward commands for speech sounds based on perceived speech errors.

After the SSM cell is activated, motor commands are sent to the motor cortex via two control subsystems: a *feedforward control system* and a *feedback control system*; see Figure 3.3. The feedback control system contains two components: an *auditory feedback control system* and a *somatosensory feedback control system*. These control systems are tuned during a learning phase analogous to infant babbling and early word imitation. Babbling causes semi-random articulator movements which work to tune the synaptic projections between the sensory error maps and the motor cortex.

Auditory feedback control is crucial for tuning the speech motor control system. DIVA contains projections from the SSM cells in the left frontal operculum to higher auditory cortical areas which contain the auditory target regions for the sounds being produced. When the model *hears* speech sounds, similar to the exposure of an infant to the sounds of his/her native language, these connections are tuned. The target is compared to incoming auditory information from the auditory periphery, and if the feedback is outside the target region, auditory error cells become active in the posterior superior temporal gyrus and planum temporale. These error cell activities are transformed into corrective motor commands through projections from higher auditory cortical areas to motor cortex. The motor cortex contains articulator velocity and position maps and sends execution signals that specify positions of the seven articulators that determine the vocal tract shape in the articulatory synthesizer (Guenther *et al.*, 2003; Guenther and Ghosh, 2003).

Once DIVA has learned the appropriate feedforward commands for producing speech sounds, it can correctly produce the sound using only the feedforward

commands. These feedforward motor commands consist of projections from the SSM in the left frontal operculum to the primary motor cortex. If no auditory error occurs during speech production, then the auditory feedback control subsystem will not be activated. However, if an external perturbation occurs (e.g. real time auditory warping of the speaker's feedback) that causes a perceived auditory error then the auditory error cells will become active and attempt to correct for the perturbation. For example, Tourville *et al.* (2005, 2008) showed that when auditory feedback is altered during speech production, subjects compensate for this perturbation and auditory areas such as the posterior superior temporal gyrus and planum temporale are activated during this compensation.

The somatosensory feedback control subsystem works alongside the auditory feedback control system. The model's somatosensory state map corresponds to the representation of tactile and proprioceptive information from the speech articulators in primary and higher-order somatosensory cortical areas in the postcentral gyrus and supramarginal gyrus. Cells in this map are active during speech if the speaker's tactile and proprioceptive feedback from the vocal tract deviates from the target region for the sound being produced. The output from the somatosensory error map propagates to motor cortex through synapses tuned during babbling, and the errors are transformed into motor commands to correct the errors. These functions were confirmed in an fMRI study in which the speaker's jaw was blocked during speech production (Tourville *et al.*, 2005).

Based on the DIVA model, some hypotheses can be made regarding damage to the different regions and control systems within the model. There may be at least two subtypes of AOS, one associated with damage to the SSM and the other with damage to the subsystem projecting to the SSM. Duffy (1995, p.264) has written about AOS subtypes, noting that one possible subtype "has been described as characterized by relatively well-formed articulatory errors (frank substitutions) with periods of normal prosody. It contrasts with another type of AOS characterized by distorted approximations of phonetic targets with relatively pervasive rate and prosodic abnormalities." DIVA predicts that damage to the subsystem projecting to the SSM and responsible for selecting the SSM cells will lead to the former subtype of AOS (referred to as AOS type 1), while damage to the SSM itself is expected to lead to the latter subtype (referred to as AOS type 2). In terms of lesion location, it can be hypothesized that damage to the frontal operculum (FO), posterior regions of IFG, and possibly extending into the ventral premotor cortex and anterior insula will most likely result in AOS type 2 errors. Damage to the superior areas of the IFG, extending into the inferior frontal sulcus (IFS) will lead to AOS type 1 errors. The main anatomical distinction between these two subgroups lies on the superior-inferior plane of IFG, the anatomical region proposed by Hillis *et al.* (2004) to be damaged in AOS patients.

According to these hypotheses, both subject groups should exhibit articulatory groping, initiation difficulties, increased response latencies, and self-correction attempts. AOS type 1 errors can be generally defined as fluent productions of the wrong speech sounds or misplacement of speech sounds within the motor program, including:

fluent substitutions, additions, transpositions, and omissions, few distortions, and few phonemic paraphasias. AOS type 1 patients should also exhibit relatively normal coarticulation and prosody. On the other hand, AOS type 2 errors can be classified as poorly articulated approximations of the desired syllables, including: slurring, distortions, substitutions, deletions, phonemic paraphasias, reduced co-articulation, nonfluent speech and dysprosody.

In terms of DIVA, AOS type 2 would result from damage to the SSM cells and thus impairing three potential functions of the SSM; see Figure 3.3. Damage to the feedforward control system would result in an inability to access the feedforward commands (or motor programs) for speech sounds such that the model may not be able to access the motor programs for the syllables corresponding to the damaged cells. In addition, damage could impair the ability of the model to update the feedforward commands for speech sounds based on perceived speech errors. This would result in difficulties in compensating for perceived speech errors and updating feedforward commands.

If a syllable's SSM representation is partially damaged, one would expect only the undamaged portion of the syllable's motor program to be read out[2] when the model tries to produce the syllable. The produced speech would be expected to be relatively slow, poorly articulated, poorly timed, and possibly missing phonetic units due to the loss of portions of the feedforward motor program. Assuming word-level prosody is represented in the SSM, prosodic disturbances would also be expected. If the syllable's

[2] By read out, the SSM cells are activated, causing the feedforward motor program represented in the projections from these cells to the motor cortex to be carried out.

representation is extensively or completely damaged, one would expect a *groping* error in which the model cannot activate any SSM cells and thus cannot produce any sounds; such groping episodes are a common aspect of AOS (e.g., Duffy, 1995).

Damage to projections from the left ventral premotor cortex areas to sensory errors could impair the auditory and somatosensory feedback control systems for speech by damaging the representation of auditory and somatosensory targets. However, both auditory feedback and somatosensory feedback are relatively preserved in AOS patients, at least based on the limited evidence available (see Section 3.1.2). For example, bite block studies have shown intact somatosensory feedback in AOS patients (Jacks, 2008). This seems to suggest that the right ventral premotor cortex, which is typically not damaged in AOS, may be more responsible for feedback control (Tourville *et al*., 2008).

On the other hand, damage to the subsystem *projecting to* the SSM (as opposed to damage to the SSM cells themselves) will most likely result in sequencing errors containing fluent misplacements of speech sounds in the motor plan as well as deletions or insertions. In this case, the erroneous utterances would be expected to be fluent (i.e., AOS Type 1 errors) since the motor programs projecting from the SSM cells would be intact if there were no damage to the SSM cells themselves.

The DIVA model offers a convincing neural network model of speech production that differs from previous models of speech production (e.g., Dell *et al*., 1997; Levelt *et al*., 1999) in that it includes a realistic motor control system that carries out the production of desired phonological units. The output of these previous models is

phonological (e.g., a string of phonemes) rather than a sequence of articulator movements and corresponding acoustic signal. Thus, these models do not account for AOS Type 2 errors, characterized by poor approximations of phonetic targets. Instead they produce errors in the ordering of phonemes (fluent paraphasias, including omissions and substitutions, but not poor productions). This situation prompted Ziegler (2002) to comment that "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".

DIVA is able to reconcile these issues and thus provides a hypothesis that there should be (at least) two different subtypes of AOS: (i) AOS due to damage to the left IFS and surrounding areas that degrades the sequencing and selection of the SSM cells, but not the SSM cells themselves, and yields AOS Type 1 errors (fluent productions of the wrong speech sounds or misplacement of speech sounds within the motor program), and (ii) AOS due to damage to the SSM cells in the frontal operculum that leads to AOS Type 2 errors (which include poorly articulated approximations of the desired syllables).

3.3.3 Outstanding issues

There are many outstanding issues that remain to be addressed when developing a functional hypothesis regarding AOS. By using the DIVA model as a guide, two hypotheses have been developed regarding damage seen in AOS patients. These two patient groups are differentiated both by behavioral characteristics and lesion location. However, very little research has looked into the different behavioral characteristics exhibited by specific lesions. In order to better assess the DIVA account of AOS,

patients with lesions to either the IFS or the FO need to be studied to determine if their behavioral features are consistent with the DIVA hypothesis. In addition, more analysis regarding how the effects of damage to the SSM can impair feedback and feedforward control systems as illustrated in Figure 3.3 is needed. More studies aimed at evaluating the brain damage with behavioral characteristics in speech disorders in general can help to further refine the AOS hypotheses presented here as well as to better develop motor speech models, such as DIVA. Ultimately, the speech pathology, neuroscience, and modeling communities need to come together to create studies aimed at finding the functional associations of these brain regions and how they are impacted in AOS.

3.4 Conclusion

Apraxia of speech (AOS) is a disorder of the speech planning and/or programming of speech production without comprehension impairment and without weakness in the speech musculature. Although much variability exists in how AOS has been defined and diagnosed, more recently a set of characteristics has been proposed on the basis of expert consensus that aims to reduce this variability and address overlap of symptoms with other disorders such as aphasia and dysarthria. Most speech pathologists and researchers now agree that AOS needs to be differentiated both from aphasias which involve language deficits (spanning modalities, such as comprehension, speaking, and writing) and from dysarthrias which involve speech deficits due to impaired neuromuscular functioning.

AOS onset is typically caused by brain trauma, yet speech pathologists rarely use brain imaging to help determine AOS diagnosis. Research has begun to surface that

attempts to define the neurological correlate of AOS impairment. Regions in the left inferior frontal lobe are most often cited as the location of damage in AOS patients. The exact regions implicated in AOS are difficult to determine given that these regions lie in close proximity with each other and thus, AOS may result from damage to all of these regions or a combination of them.

The analysis of AOS contained in this chapter is consistent with both the Levelt model and the DIVA model in that they both contain a representation of stored articulatory gestures. Further studies aimed at validating and refining the predictions of these models can provide us with an additional explanatory tool for apraxia of speech. Thus, functional modeling efforts may provide the speech pathology communities and the neurology communities with an explanation of AOS in such a way as to aid in diagnosis and treatment.

CHAPTER 4

SPEECH MOTOR CONTROL IN APRAXIA OF SPEECH: COMPARISON WITH A CASE STUDY

4.1 Introduction

In this chapter, a computational neural model is presented which describes how the brain represents, selects, and stores learned speech sounds as well as how these sound representations and mechanisms may be damaged in apraxic patients. The model is built upon the DIVA model (Directions Into Velocities of Articulators; Guenther, 1994, 1995; Guenther *et al*., 1998; Guenther and Ghosh, 2003; Tourville *et al*., 2005) and the GODIVA model (Gradient Order DIVA; Bohland *et al*., in press). For over 15 years, the DIVA model has been used to simulate a large number of speech production phenomena including motor equivalence, contextual variability, coarticulation, velocity and distance relationships, speaking rate effects, and speaking skill acquisition. The GODIVA model was built as an extension to DIVA to provide DIVA with a module for selection and sequencing of speech sounds. See Section 3.3.2 for a high level overview of DIVA.

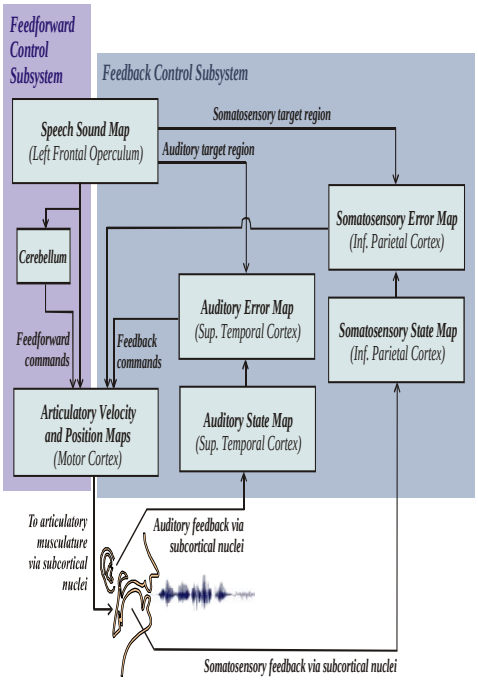
In order to further explore the hypothesis set forth in the previous chapter regarding damage to DIVA in apraxic patients, a simplified version of GODIVA is built onto DIVA. In addition, the Speech Sound Map (SSM) in DIVA is elaborated with a *distributed* representation of speech sounds in order to create a more biologically plausible representation of the SSM cells. A brief introduction of the model components is provided in Section 4.2. Computer simulations illustrating apraxic-like symptoms will

also be described. Section 4.3 will discuss a case study of an apraxic patient. Finally, the chapter will conclude with a discussion of the model hypotheses with respect to the computer simulations and a case study.

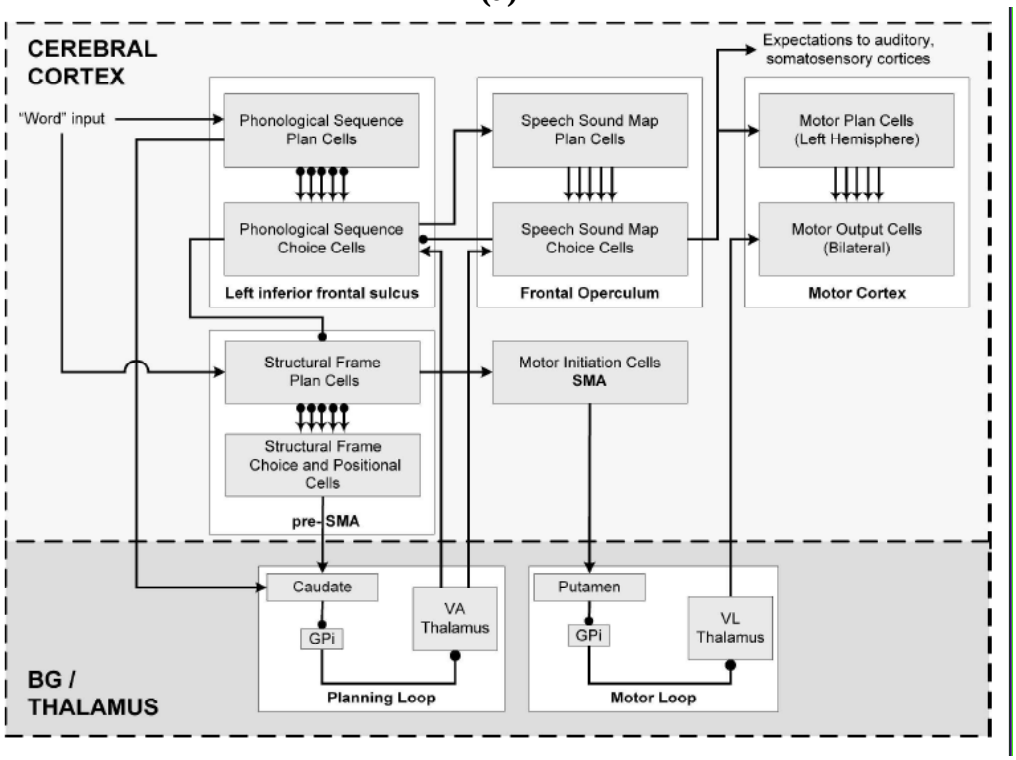
4.2 Computational modeling of apraxia of speech

4.2.1 Model description

The model presented in this chapter is built upon both the GODIVA and DIVA models. A schematic box diagram of each of these models is illustrated in Figure 4.1. Both of these models were introduced in the previous chapter (see Section 3.3.2). In this section, these models are simplified and combined to create the appropriate network needed to explain the mechanisms damaged in apraxia of speech (AOS) by focusing on the transition from phonological encoding to articulation. Notably, this chapter will focus on two key regions: the left inferior frontal sulcus (IFS) and the speech sound map (SSM); see Figure 4.2. Some components from DIVA and GODIVA have been excluded from discussion; see Guenther *et al.* (2006) and Bohland *et al.* (in press) for more extensive reviews of these components.



(a)



(b)

Figure 4.1. *Schematic box diagram of* ***(a)*** *DIVA and* ***(b)*** *GODIVA.* [Reprinted with permission from Guenther *et al.*, 2006 and Bohland *et al.*, in press, respectively]

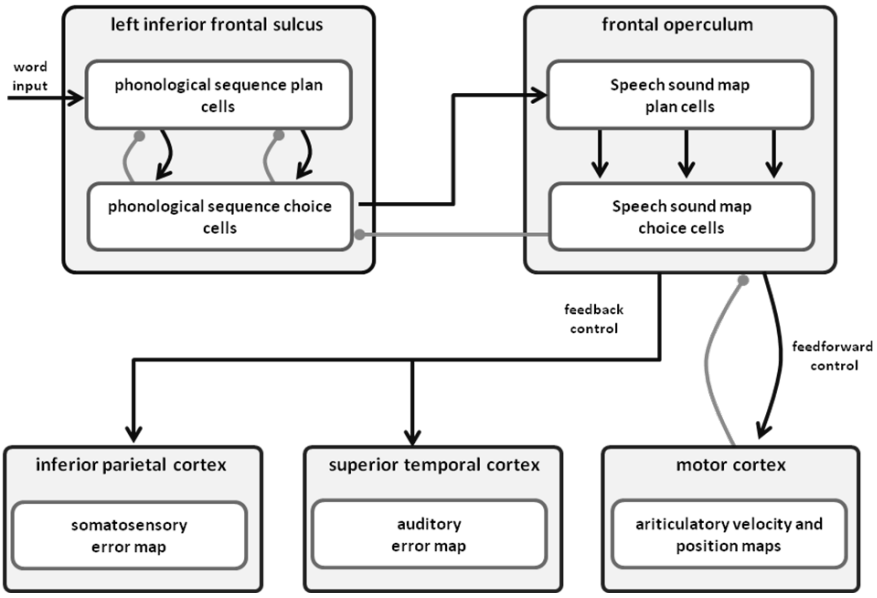


Figure 4.2. *Schematic box diagram of left inferior frontal sulcus and frontal operculum/speech sound map.* These two regions are the focus of the model and simulations in this chapter. This diagram also highlights the point of connection and crossover of DIVA and GODIVA.

4.2.1.1 Phonological content representation in the left inferior frontal sulcus

The *input* to the model during ordinary speech production codes lexical items (e.g. words) that arrive sequentially from high-level lexical/semantic processing areas, such as the inferior or ventrolateral prefrontal cortical regions and middle anterior temporal lobe. The inputs initiate the activation of the phonological content representation, which is hypothesized to exist in the left inferior frontal sulcus (IFS; see Bohland and Guenther, 2006; Bohland *et al*., in press). For simplicity, this model component will be referred to as the IFS herein.

The IFS makes use of a competitive queuing (CQ) model architecture (Houghton, 1990; Bullock and Rhodes, 2003). This class of models has been used to describe many 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), cursive handwriting production (Bullock *et al*., 1993), and language production (Dell *et al*., 1997). Further support for CQ models is found in neurophysiological data showing neural cell populations in the prefrontal cortex of monkeys behaving in a manner similar to what is proposed by CQ models during shape drawing tasks (Averbeck *et al.*, 2002, 2003).

The CQ architecture consists of two layers: a *planning layer* and a *choice layer*; see Figure 4.3. The planning layer cells maintain a primacy gradient of the input which serves to preserve both the items and their serial order, otherwise known as *item and order memories* (Grossberg, 1978a,b). An iterative choice process is employed in which the item with the highest activation is first chosen by the choice layer. This item's activation is then suppressed. This process repeats itself through the entire sequence.

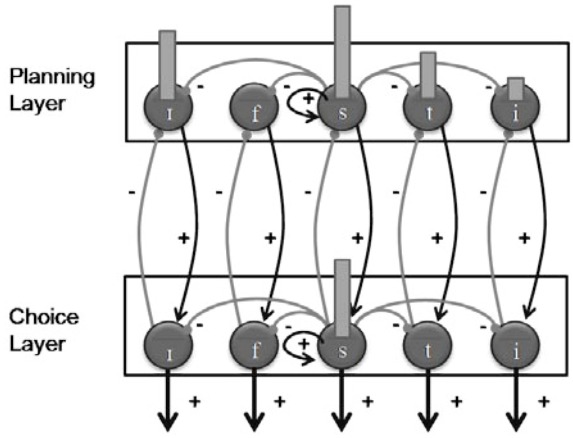


Figure 4.3. *Competitive Queuing (CQ) model architecture*. The example architecture shows the processing of the letter sequence /sɪ ti/ (city). The serial position of each letter is encoded by the strength (corresponding to bar height) of its representation in the planning layer (top). The choice layer chooses the strongest input through a winner take all competitive process (e.g. /s/). Once the winner is selected, its representation in the planning layer is suppressed and then the next most active representation (/ɪ/) is chosen. This process repeats itself through the entire sequence.

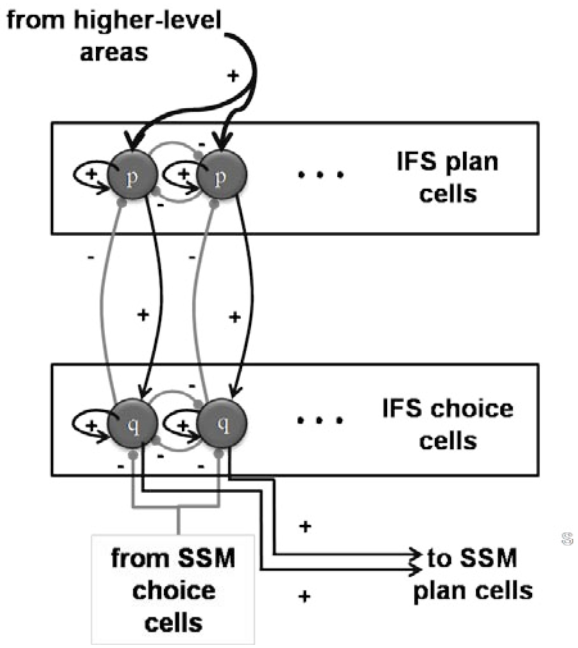


Figure 4.4. *Schematic diagram of the IFS component.* The top layer consists of plan cells, *p*, and the bottom layer consists of choice cells, *q*. Each of these cells codes a phoneme in a given syllable position. The choice cells undergo a winner-take-all competition and inhibit their corresponding plan cell once they have been activated. This process results in activation of a phonological syllable in the IFS choice field that can potentially activate matching syllable motor programs in the speech sound map. Once the correct program is chosen in the speech sound map, the choice cell activation is suppressed.

The IFS component is schematized in Figure 4.4. Both the planning and choice layers of the IFS contain topographic *zones* of cells which correspond to appropriate syllable positions in the forthcoming speech sound. For the purpose of this model, seven syllable positions are included which correspond to a generic syllable template (Fudge, 1969). This syllable template is able to accommodate a substantial number of English syllables. For simplicity, this model assumes that the fourth position is always used to represent the syllable nucleus, the vowel, and the remaining positions can be filled with

consonants depending on the syllable. The model contains representations for 53 phonemes, including 30 consonants and 23 vowels, derived from the CELEX (Centre for LEXical information) lexical database (Baayen *et al.*, 1995); see Appendix B for a table of these sounds.

The cells in the planning layer maintain a cortical code for the potential phonemes of the speech utterance and for laying out the motor plan and sequence. This layer also maintains a primacy gradient in order to co-temporally code for multiple forthcoming phonemes or syllables. The activity of cell $\boldsymbol{p}_{ij}$, representing the phoneme *i* at syllable position *j* in the planning layer matrix **p**, is governed by the following differential equation:

$$\frac{dp_{ij}}{dt} = -A_p p_{ij} + (B_p - p_{ij})(\alpha u_{ij}^{p} + [p_{ij} - \theta_p]^+) - \ldots$$
$$p_{ij}(\sum_{k \neq i} W_{ik} p_{kj} + 12 * f([q_{ij} - \theta_q]^+)) \quad , \qquad (4.1)$$

where the first term shows passive decay at a rate controlled by A_p = 0.1. The second term models excitatory input to the cell which drives activity in the positive direction. The multiplicative term ($B_p - p_{ij}$) enforces an upper bound B_p = 5 to the cell's activity. $\boldsymbol{u}_{ij}^{I}$ is the *word* input representation that excites the IFS plan layer. This input is gated by α = 0.4 to ensure that the activity of the cells receiving new inputs corresponding to words spoken later does not exceed the activity level of those spoken earlier. Thus, the correct order of speech output is maintained (c.f. Bradski *et al.*, 1994). The second term of the equation also contains a recurrent self-excitation which is thresholded by $\boldsymbol{\theta}_{I}$ = 0.01 and then half-wave rectified, $[\ ]^+$.

The third term models inhibitory inputs by other plan cells coding different phonemes at the same syllable position to the cell. This term drive activity of the cell in the negative direction to a lower bound of zero. These inhibitory inputs are weighted by the weight matrix **W**. p_{ij} also receives inhibition from cell q_{ij} which is the corresponding cell in the IFS choice layer and together create a cortical column of cells. This inhibition is thresholded by θ_c= 0.5 and is amplified by a faster than linear function (c.f. Grossberg, 1973):

$$f(x) = x^2. \qquad (4.2)$$

The general equation structure used in equation *(4.1)* where the first term represents passive decay, the second term represents excitatory inputs to the cell, and the third represents inhibitory inputs to the cell is often referred to as a shunting equation; see Figure 4.5. These shunting terms (Grossberg, 1973) are motivated by empirically-observed cell membrane properties (e.g. Hodgkin and Huxley, 1952). The shunting equation structure is repeated in other components of the model.

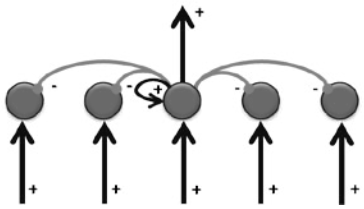


Figure 4.5. *Basic shunting equation architecture.* The arrows at the bottom indicate the input to the cells and the arrow at the top indicates the output from the cell. The light grey connections indicate the inhibitory term. The middle cell also shows self-excitation.

The choice layer is used to select the appropriate SSM cells based on the motor plan and sequence. Within the choice layer, strong competition exists such that there is a single

winning representation within each positional zone. After the choice cell becomes active, it suppresses its corresponding representation in the planning layer. The activity of a cell, q_{ij}, in the IFS choice layer **q** is given by:

$$\frac{dq_{ij}}{dt} = -A_q q_{ij} + (B_q - q_{ij})([p_{ij} - \theta_p]^+ + f(q_{ij})) - q_{ij}(\sum_{kj,k\neq i} W_{ik} f(q_{kj}) + \Gamma_{ij})) \quad , \quad (4.3)$$

where A_q = 1 is the passive decay parameter and B_q = 5 is the upper bound on cell activity. The excitatory term includes the function $f(q_{ij})$ which is self-exciting and governed by equation *(4.2)*, as well as selective input from IFS plan cells, p_{ij}, corresponding to appropriate positional zones. The inhibition is also received from within the same syllable position zone. The activation of the inhibitory cells is again governed by the faster than linear function in equation *(4.2)*.

After the choice cell becomes active, it will maintain its activation through recurrent interactions. The cell's activity can be quenched through a potentially strong inhibitory input, Γ_{ij}, which represents a *response suppression signal*. This signal originates in the SSM choice layer and arrives via interneurons. It suppresses the IFS choice cells by the chosen speech motor program cells in the SSM. The value of Γ_{ij} at time *t* is defined by:

$$\Gamma_{ij} = 10 Z_k^{ij} s_k(t) \ , \quad (4.4)$$

where Z_k^{ij} is active if the phoneme *i* is part of the syllable motor program *k* in syllable position *j*, and 0 otherwise; see the next subsection for a more descriptive definition.

The choice cells in the IFS create cortico-cortical synapses with neural populations in the SSM. These synapses allow for the set of winning choice cells in the IFS choice layer to activate a set of potential matching motor programs represented by the SSM plan cells. The better the match between the choice cells and the motor programs, the higher the received activation in the SSM plan cells. Since each IFS choice cell codes for each position in the syllable, these projections transform the phonological syllable into a speech motor program.

4.2.1.2 Speech sound representation in the speech sound map

The cells in the speech sound map (SSM) initiate the readout of motor programs along with the auditory and somatosensory expectations for learned speech sounds representing one thousand of some of the most frequently encountered syllables in the English language (Baayen *et al*., 1995). It is hypothesized that the SSM lies in the more ventral regions of the left inferior frontal gyrus (IFG), frontal operculum (FO) area, and/or the left ventral premotor cortex (Guenther *et al*., 2006). The SSM region contains two layers, a plan layer and a choice layer.

The SSM cells contain a distributed representation; see Figure 4.6. This distribution is based on articulation features of the phonemes that comprise the syllable. Each phoneme *i* in the syllable *k* contain a distributed feature vector, $\boldsymbol{x_i}$, which is comprised of a ranking for each of the three features of articulation; see Table 4.1. For

vowels, these features include tongue height, backness, and roundness[3]. For consonants, the features are place of articulation, manner of articulation, and voicing.

Phoneme	**Articulation feature type**	
Vowel	***Tongue height***	Close, near close, close mid, mid, open mid, near open, and open
	Tongue backness	Front, near front, central, near back, and back
	Roundedness	Round and unround
Consonant	***Place of articulation***	Bilabial, labiodentals, dental, alveolar, postalveolar, palatal, velar, and glottal
	Manner of articulation	Plosive, nasal, fricative, approximant, lateral approximant, and trill
	Voicing	Voiced and voiceless

Table 4.1 Phoneme features used in the distributed representations of the SSM cells.

The first two characteristics (for vowels: height and backness and for consonants: place and manner) are more useful in differentiating sounds than are the last characteristics (for vowels, roundness and for consonants, voicing; Lindau, 1978). Thus, the feature vectors were paired with a weight vector $\boldsymbol{w}_x$ = [0.4 0.4 0.2] in order to capture this distribution.

[3] For diphthongs, an average value for the two vowels that comprise the dipthong was calculated for each characteristic.

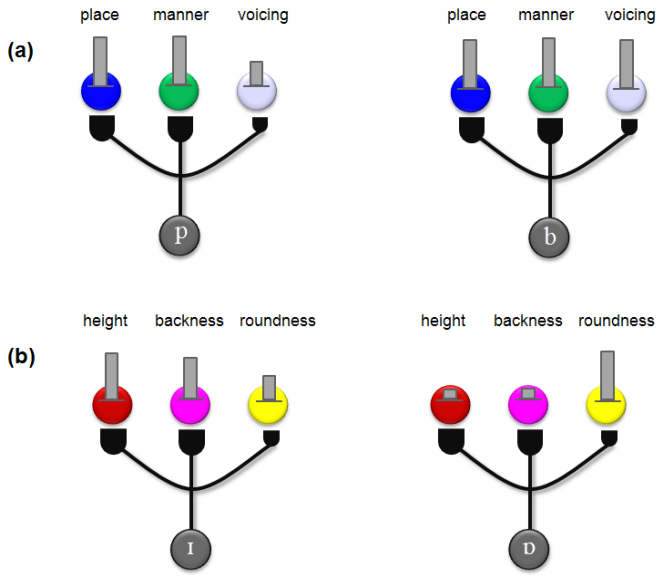


Figure 4.6. *SSM cell distribution.* **(a)** Consonants are differentiated based on place of articulation, manner of articulation, and voicing. Two very similar consonants, /p/ and /b/, are shown. They differ only in voicing. **(b)** Vowels are differentiated based on tongue height, tongue backness, and roundness. Two dissimilar vowels, /ɪ/ and /ɒ/, are shown which differ significantly on all articulation features.

The SSM plan layer differs from the IFS plan layer in that it does not code serial order. Instead it codes the *degree of match* between active IFS choice cells and stored motor programs in the SSM. The SSM plan layer receives projections from the IFS choice layer. The activity level of cell k in the SSM plan layer representation **r** is governed by:

$$\frac{dr_k}{dt} = -A_r r_k + (B_r - r_k)(\sum_i \sum_j Z_k^{ij} f([q_{ij} - \theta_q]^+) + [r_k - \theta_r]^+) - r_k(\sum_{n \neq k} r_n) \quad , \quad (4.5)$$

where A_r = 0.1 is the passive decay parameter and B_r = 5 is the upper bound on cell activity. The double sum in the excitatory term computes the net excitatory input from cells in the IFS choice layer **q** which is weighted by the synaptic strengths specified in the input weight matrix $\mathbf{Z}_k^{ij}$ and subject to the low noise threshold θ_q. Cell r_k also

receives self-excitatory feedback which is subject to a low noise threshold θ_r = 0.1. The third term models the lateral inhibitory inputs to the cell from all other cells in the SSM plan layer, which drive activity in the negative direction, with a lower bound of zero.

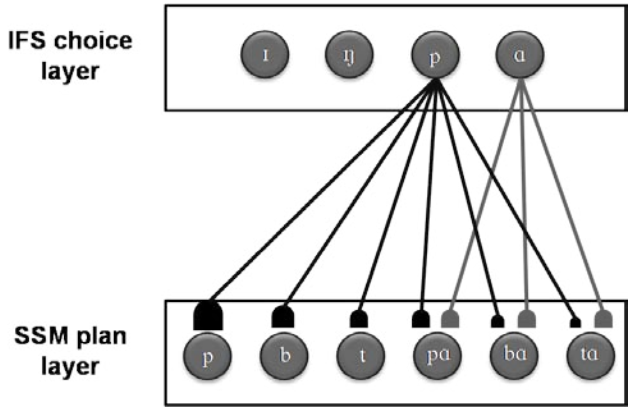


Figure 4.7. *Weights projecting between IFS choice layer and SSM plan layer*. For example, the IFS choice cell representing /p/ projects to /p/, /b/, /t/, /pɑ/, /bɑ/, and /tɑ/ in the SSM plan layer. The weight between /p/ in the IFS choice layer and /p/ in the SSM plan layer is the strongest, since it is a perfect match. The weights to similar phoneme syllables decrease in strength with decreasing similarity. The weight between /p/ in the IFS choice layer and /pɑ/ in the SSM plan layer is at half strength because /ɑ/ also projects to /pɑ/. The weight strength to /bɑ/ and /tɑ/ from /p/ is also much smaller and decreases with decreasing similarity.

The weights, $\mathbf{Z}_{ip}^{i}$, were hard-wired based on creating a weight representation between the phoneme representations in the IFS choice layer and all of the similar sound representations in the SSM plan layer. The similarity is determined by the articulation features in the distributed representations of the SSM cells. For example, /p/ in the IFS choice layer will project with full weight strength to /p/, /pɑ/ etc in the SSM plan layer because these syllables contain /p/; see Figure 4.7. /p/ will also project to sounds like /b/, /t/, /bɑ/, /tɑ/, etc at less strength because /p/ and /b/ only differ on voicing and /p/

and /t/ differ on place of articulation. Thus, /p/ and /b/ are considered highly similar phonemes and /p/ and /t/ are more dissimilar.

The dissimilarity $\mathbf{S}_k^i$ is the measure of dissimilarity between the phoneme *i* in position *j* in the IFS choice layer **q** and the syllable *k* in the SSM plan layer **r** and is determined by:

$$S_k^{ij} = \sum abs(x_{ij} - x_k).* w_x \quad , \qquad (4.6)$$

where w_x is the feature weight vector and x_{ij} and x_k are the feature vectors corresponding to the phoneme *i* in position *j* in the IFS choice layer **q** and the syllable *k* in the SSM plan layer **r** respectively. If the phoneme *i* is found at position *j* in the syllable *k* in the SSM, then $S_{qr} = 0$. If phoneme *i* at position *j* in the IFS choice layer is similar to the phoneme in position *j* in syllable *k* in the SSM plan layer, then $0 < S_{qr} < 1$. If it is not, then $S_{qr} > 1$. Thus, the synapse $\mathbf{Z}_k^i$ from IFS choice cell q_{ij} (which codes the phoneme *i* at syllable position *j*) to SSM plan cell r_k is given by:

$$\mathbf{Z}_k^{ij} = \begin{cases} 0 & \text{if } S_k^{ij} > 1 \\ (1 - S_k^{ij})(0.85 - 0.05\,j) & \text{if } r_k \text{ codes the single phoneme target } i \\ \dfrac{(1 - S_k^{ij})}{N_k} & \text{otherwise} \end{cases} . \qquad (4.7)$$

The second condition ensures that the single phoneme targets are weighted by the position of the IFS choice cells such that inputs from earlier positions in the syllable have greater efficacy. This allows 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). N_k is the number of phonemes in the syllable coded by r_k. This specification 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.

The SSM choice cells receive input from the SSM plan cells. Competitive interactions in the SSM choice layer lead to a winner and activation of the corresponding motor plans, which are hypothesized to reside in the left motor cortex (Guenther *et al*., 2006). These motor plans are co-activated by inputs from the supplementary motor area (SMA) which code for motor initiation. The activity level of cell *k* in the SSM choice layer representation **s** is governed by:

$$\frac{ds_k}{dt} = -A_s s_k + (B_k - s_k)(r_k + 10 f([s_k - \theta_s]^+)) - s_k(\sum_{j \neq k}[s_j - \theta_s]^+ + \Omega) \quad , \qquad (4.8)$$

where A_s = 10 and *f(x)* is a faster-than-linear signal activation function, resulting in winner-take-all dynamics within the layer **s**. B_k = 5 is linearly dependent upon the length of the speech segment *k*:

$$B_k = 4 + (|k| - 1) * 0.3 \quad . \qquad (4.9)$$

This equation ensures that the upper bound on the cell's activity is dependent on the length of the syllable coded by *k*. In other words, the longer the syllable *k*, the higher the excitatory activity of the cell s_k can be driven. This idea that speech sounds which represent longer lists of phonemes have a prewired competitive advantage over those that represent shorter lists, has been called a *masking field* (Cohen and Grossberg, 1986, 1987; Grossberg and Myers, 2000; Grossberg, 1978, 1986, 2003b). Thus, with all things

being equal, the longer lists are better predictors of subsequent events because they embody a more unique temporal context. As a result, there is an advantage in place for longer speech sounds which enable effective competition with shorter speech sounds.

Ω models a non-specific response suppression signal that arrives from the articulatory portion of the model indicating the impending completion of production of the current syllable motor program. When Ω is large, activity is quenched in **s**, and a new winner is then instantiated, corresponding to the most active SSM program in the plan layer **r**. This signal arises in the DIVA model prior to completion of articulation due to the delay between sending a motor command and the effect that the command has on the motor articulators (Guenther *et al*., 2006).

4.2.2 Model predictions

Section 3.2.2 described two different subgroups of AOS, one associated with damage to the SSM and the other with damage to the IFS region. Duffy (1995, p.264) has written about AOS subtypes, noting that one possible subtype "has been described as characterized by relatively well-formed articulatory errors (frank substitutions) with periods of normal prosody. It contrasts with another type of AOS characterized by distorted approximations of phonetic targets with relatively pervasive rate and prosodic abnormalities." The model predicts that damage to the IFS will lead to the former subtype of AOS (referred to as AOS type 1), while damage to the SSM would lead to the latter subtype (AOS type 2).

According to this hypothesis, both subject groups should exhibit articulatory groping, initiation difficulties, increased response latencies, and self-correction

attempts. AOS type 1 errors can be generally defined as fluent productions of the wrong speech sounds or misplacement of speech sounds within the motor program, including: fluent substitutions, additions, transpositions, and omissions, few distortions, and few phonemic paraphasias. AOS type 1 patients should also exhibit relatively normal coarticulation and prosody. On the other hand, AOS type 2 errors can be classified as poorly articulated approximations of the desired syllables, including: slurring, distortions, substitutions, deletions, some phonemic paraphasias, reduced co-articulation, nonfluent speech and dysprosody.

AOS type 2 would result from damage to part or all of a syllable's distributed SSM representation. Complete or extensive damage to a SSM cell's distribution will either prevent activation of any SSM cells (yielding a *groping* error) or yield activation of the incorrect cells in the SSM, producing erroneous utterances. Furthermore, partial damage to the distribution corresponding to a SSM cell, would result in only the undamaged portion of the syllable's motor program being read out when the model tries to produce the syllable. The produced speech would be expected to be relatively slow, poorly articulated, poorly timed, and possibly missing phonetic units due to the loss of portions of the feedforward motor program. Assuming word-level prosody is represented in the SSM, prosodic disturbances would also be expected.

On the other hand, damage to the IFS directly would result in AOS type 1 errors. These errors are defined as sequencing errors where there can be fluent misplacement of speech sounds in the motor plan as well as deletions and insertions. Basically, the model is unable to effectively generate the sequential motor plan for the intended utterance.

Longer speech sounds should generate more errors than shorter speech sounds. Furthermore, erroneous utterances would be expected to be relatively fluent since the motor programs projecting from the SSM cells would be intact if there were no damage to the SSM cells themselves.

4.2.3 Simulation results

In order to test the predictions put forth in the previous sections, simulations of the two apraxic subgroup damage types were performed on the model and the resulting behaviors were analyzed. The simulations were run on a workstation computer using an eight core Intel Xeon Processor with 2.33 GHz and 8.00 GB of RAM and using Matlab v.7.1.

The model was tested using a variety of syllabic sequences. For a more direct comparison to human performance, a subset of words from the Apraxia Battery for Adults – Second Edition (ABA-2; Dabul, 2000) were chosen for analysis. These words and their IPA spellings are listed in Table 4.2.

Word	**IPA spelling**	**Number of syllables**
Hope	/həʊp/	1
Please	/pleɪs/	1
City	/sɪ tiː/	2
Hopeful	/həʊp fʊl/	2
Pleasing	/pleɪ zɪŋ/	2
Hopefully	/həʊp fæ liː/	3

Table 4.2. Words tested by the model in the AOS conditions.

Figure 4.8 shows the IFS planning and choice layers for syllable positions 3, 4, and 5 and the SSM planning and output layer when the model is presented with the word ‘hopefully’ (/həʊp fæ liː/) and no damage is present. This figure illustrates how the model performs correctly in normal conditions.

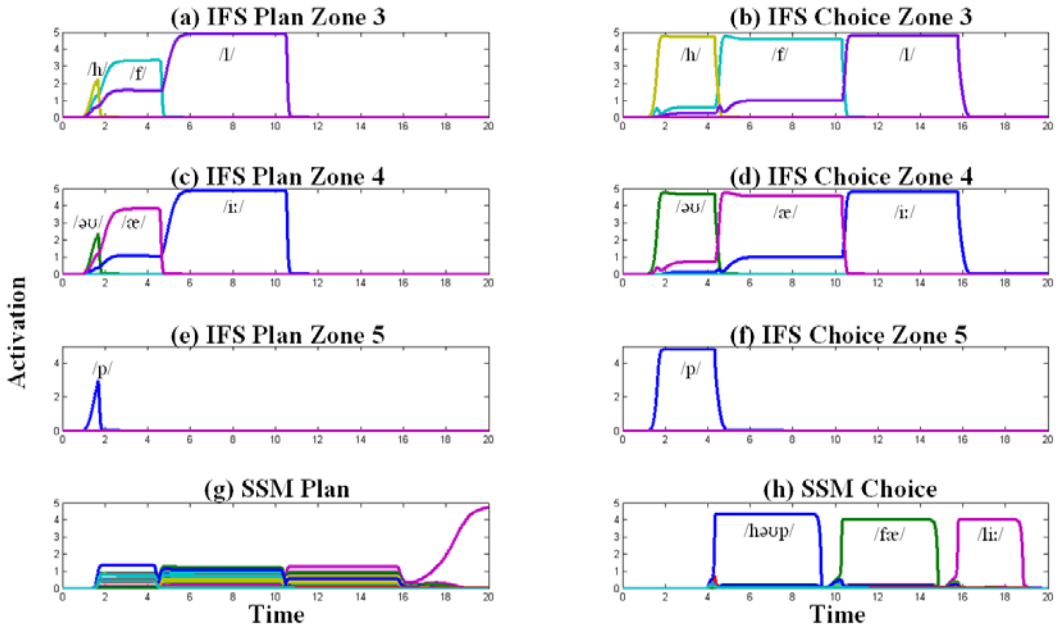


Figure 4.8. *Example without damage.* Simulation of ‘hopefully’ (/həʊp fæ lː/). **(a)**, **(c)**, and **(e)** are the IFS plan layer for syllable position zones 3, 4, and 5, respectively. **(b)**, **(d)**, and **(f)** show the IFS choice layer for syllable position zones 3, 4, and 5, respectively. **(g)** is the SSM plan layer. **(h)** is the SSM choice/output layer.

In Figure 4.8, the IFS plan layers show a temporal activation of the phonemes for the syllable position where zone 3 codes for the first consonant, zone 4 codes for the vowel, and zone 5 codes for the second consonant of the syllable. Only one activation peak is seen in zone 5, corresponding to the /p/ in the first syllable, /həʊp/. The IFS choice layers show the winning choice cells corresponding to the phonemes in the word. The SSM plan layer shows the activation of the various SSM cells that correspond to the

winning IFS choice cells. Finally, the SSM choice layer shows the winning speech sounds, corresponding to the syllables /həʊp/, /fæ/, and /lː/.

Score	Description
2	The response was correct, without hesitation, without struggle, and without articulatory error.
1	The response was delayed, corrected, displayed searching, contained one or more articulatory errors, but the correct number of syllables was maintained and there was general conformation of the word.
0	There was no response, an attempt with no word production, the wrong number of syllables were said, or it was so misarticulated that the word was no longer recognizable.

Table 4.3. Increasing word length score guide. [Reprinted with permission from Dabul, 2000]

Once damage is introduced to the model, the results of the simulations are analyzed by evaluating the effects of increasing word length and by taking an inventory of the articulation errors generated by the model. This evaluation is guided by the ABA-2 (Dabul, 2000). Table 4.3 shows the scoring for the increasing word length analysis. The scores for longer words are compared to those for shorter words.

Table 4.4 summarizes the errors made by the model. Word level, syllable level, and phoneme level errors are evaluated separately.

Error Type	Description	Example
Initiation Difficulties	Initial word groping or repetition of sound.	/hhəʊp/ instead of /həʊp/
Word Elongation	Increased number of syllables or phonemes or lengthening of vowel sounds.	/pleɪz zɪ ŋ/ instead of /pleɪ zɪŋ/
Word shortening	Decreased number of syllables or phonemes or shortening of vowel sounds.	/pleɪ/ instead of /pleɪs/
Move syllable boundary	Syllable boundary is moved.	/sɪt iː/ instead of /sɪ tiː/
Abnormal prosody	Difficulty in initiation, shortening or lengthening sounds, stress changes, and syllable boundary changes.	/həʊp/ instead of /həʊp fʊl/

(a)

Error Type	Description	Example
Anticipatory error	Syllable from later position is repeated earlier.	/liː fæ liː/ instead of /həʊp fæ liː/
Transposition error	Two syllables exchange position.	/zɪŋ pleɪ/ instead of /pleɪ zɪŋ/
Perseverative error	Two syllables are repeated.	/sɪ sɪ/ instead of /sɪ tiː/
Insertion error	Insert additional syllable.	/sɪ pʊt tiː/ instead of /sɪ tiː/
Deletion error	Deletion of a syllable.	/həʊp/ instead of /həʊp fʊl/

(b)

Error Type	Description	Example
Anticipatory error	Phoneme from later position is repeated earlier.	/pəʊp/ instead of /həʊp/
Transposition error	Two phonemes exchange position.	/pəʊh/ instead of /həʊp/
Perseverative error	Two phonemes are repeated.	/həʊh/ instead of /həʊp/
Voicing error	A consonant looses or gains voicing.	/həʊb/ instead of /həʊp/
Vowel error	Vowel error that is not anticipatory, perseverative, or transposition.	/hɪp/ instead of /həʊp/
Groping error	Grope or search for the correct phonemic sound.	/həʊ-əb/ instead of /həʊp/
Schwa insertion	Insert schwa (ə) between syllables or in consonant clusters.	/pəleɪs/ instead of /pleɪs/
Elongation of the vowel	Vowel sound is elongated or becomes a diphthong.	/seɪ tiː/ instead of /sɪ tiː/
Shortening of the vowel	Vowel sound is shorted or the diphthong is replaced with a single vowel.	/hʊp/ instead of /həʊp/
Consonant cluster reduction	Consonant cluster within a single syllable is reduced.	/peɪs/ instead of /pleɪs/
Insertion error	Insert a phoneme.	/həʊlp/ instead of /həʊp/
Deletion error	Delete a phoneme that is not a consonant cluster reduction.	/pleɪ/ instead of /pleɪs/
Consonant error	Consonant error not voicing, anticipatory, transposition, or perseverative.	/kleɪs/ instead of /pleɪs/

(c)

Table 4.4 Articulatory errors evaluated in the simulations. **(a)** is word level errors, **(b)** is syllable level errors, and **(c)** is phoneme level errors. Errors in italics indicate errors that are primary characteristics of AOS. Other errors can also be indicative of other disorders.

4.2.3.1 AOS type 1 simulations

In order to simulate AOS type 1 damage, cells in the IFS were directly damaged by randomly destroying the cells in both the planning and choice layers. The damage was administered to cells representing the same speech sound and the same syllable zone consistent with the idea that the damage destroyed the entire cortical column in the IFS. The amount of damage was parametrically varied from 5% to 100%. As expected, the model was unable to produce utterances when the damage is set to 100%.

When the model received lesions to the IFS component, the damage was exhibited in the IFS region and the SSM continued to function normally. Figure 4.9 illustrates an example utterance when the IFS of the model is 25% damaged and the input is 'hopeful' (/həʊp fʊl/). The resulting utterance was /həʊ fʊ/ which illustrates two deletion errors at the end of each syllable. The damage was apparent in both layers of IFS. In zone 4, /ʊ/ had difficulty in gaining enough energy for activation in the choice layer, but was able to become activated. In planning layer zone 5, both /p/ and /l/ became slightly activated but with not enough energy to activate their corresponding cells in the choice layer. In the SSM layers, it was apparent that functioning was still intact and speech sounds associated with the activated phoneme cells in the IFS choice layer were activated.

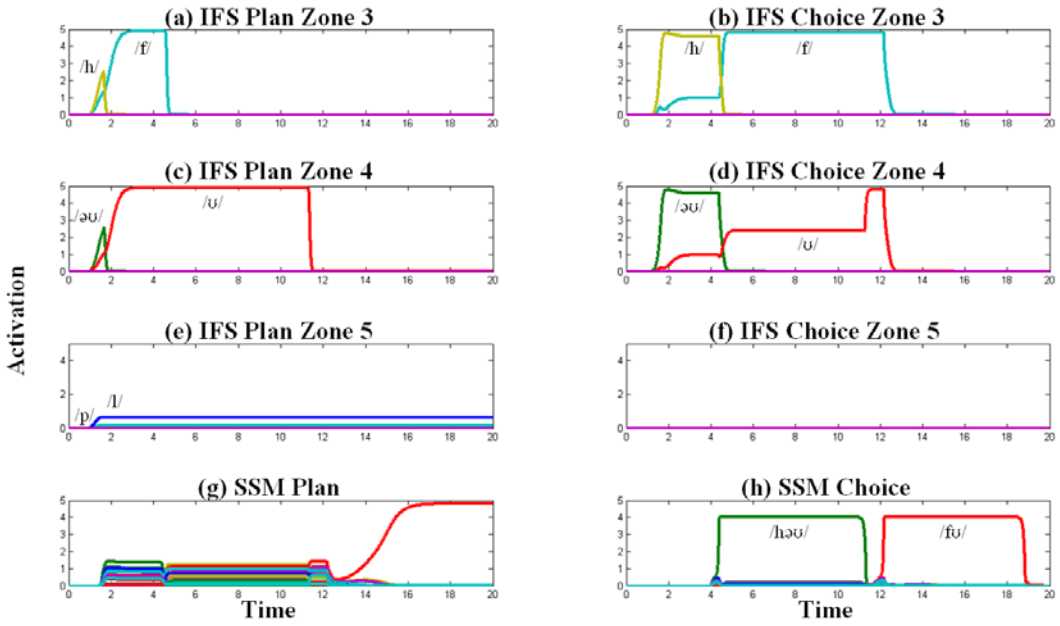


Figure 4.9. *IFS damage example*. The IFS plan layers for zones 3 **(a)**, 4 **(c)**, and 5 **(e)**, the IFS choice layers for zones 3 **(b)**, 4 **(d)**, and 5 **(f)** as well as the SSM plan **(g)** and choice **(h)** layers when the IFS was 25% damaged. The input word was 'hopeful' (/həʊp fʊl/). The resulting utterance was /həʊ fʊ/.

The average of the errors over the different IFS lesion scenarios of the model for each input word was analyzed; see Figure 4.10.

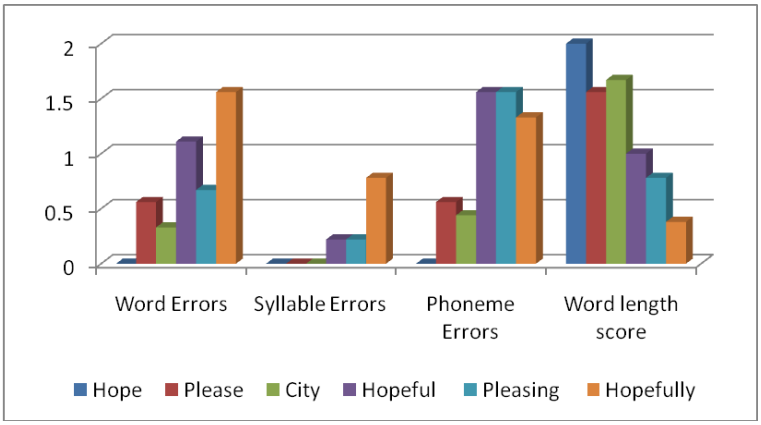


Figure 4.10. *Error analysis with IFS damage.* For each of the 6 input words, the average number of word, syllable, and phoneme level errors are illustrated. The bars on the far right represent the word length score.

As expected, Figure 4.10 shows that there were no syllabic errors for the single syllable words. The average number of errors for the input words increased with increasing word length. This was verified in the word length score (see the last group of bars in Figure 4.10) where the word length score decreased as the word length increased. Recall from Table 4.3 that a word length score of 2 indicates no errors and a word length score of 0 indicates the wrong number of syllables or an unintelligible utterance.

The number of errors increased with respect to the amount of damage to the IFS; see Figure 4.11. As a general trend, the number of word, syllable, and phoneme level errors increased as the amount of damage increased. In general there were the fewest number of syllable errors.

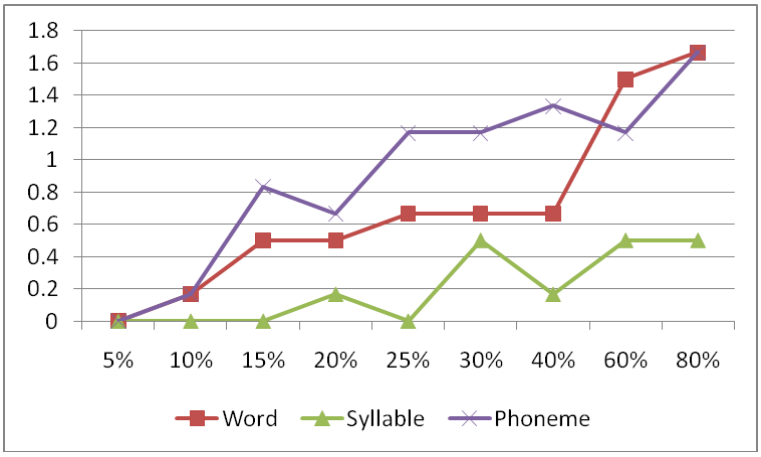


Figure 4.11. *Errors as IFS damage was varied.* The three plots show the average word, syllable, and phoneme level errors for the 6 input words as the damage was varied from 5% to 80%.

In order to take a closer look at the types of errors committed by the model, Table 4.5 summarizes the errors encountered for each error level type. The most commonly encountered word level error was word shortening followed by abnormal prosody. The syllabic level errors were only encountered for words with at least two syllables. The most common error seen at the syllabic level was deletion and transposition. The most common phonemic level errors were deletion and transposition as well. In addition, consonant cluster errors were only possible in words with consonant clusters within a single syllable. This only occurred in ‘pleasing’ and ‘please’. Whereas ‘please’ showed very few errors and no consonant cluster reductions, ‘pleasing’ showed more consonant cluster reduction errors.

Word	Word errors	Syllabic errors	Phonemic errors
Hope	NA	NA	NA
Please	*Abnormal prosody* *Word shortening* Delayed initiation	NA	*Deletion*
City	*Word shortening*	NA	*Deletion* Transposition
Hopeful	*Word shortening* *Abnormal prosody*	Transposition *Deletion*	*Deletion* Anticipatory Persevatory *Vowel error* *Vowel shortening*
Pleasing	*Word shortening* *Abnormal prosody*	Transposition *Deletion*	*Deletion* Consonant cluster reduction Transposition *Vowel elongation* *Insertion*
Hopefully	*Word shortening* *Abnormal prosody* Delayed initiation	*Deletion*	*Deletion* Transposition

Table 4.5 Encountered errors (in order from most to least encountered) when the IFS region was damaged. Errors in italics are highly indicative of AOS. The other errors may occur in AOS but are also found in other motor speech disorders.

4.2.3.2 AOS type 2 simulations

In order to simulate AOS type 2 lesions, the distributed representation corresponding to the SSM cells was randomly damaged. Such damage would imply that the SSM no longer contained fully distributed representations for each speech sound in the SSM. The amount of damage was parametrically varied from 5% to 100% damage. As expected, the model was unable to produce utterances when the damage was set to 100%.

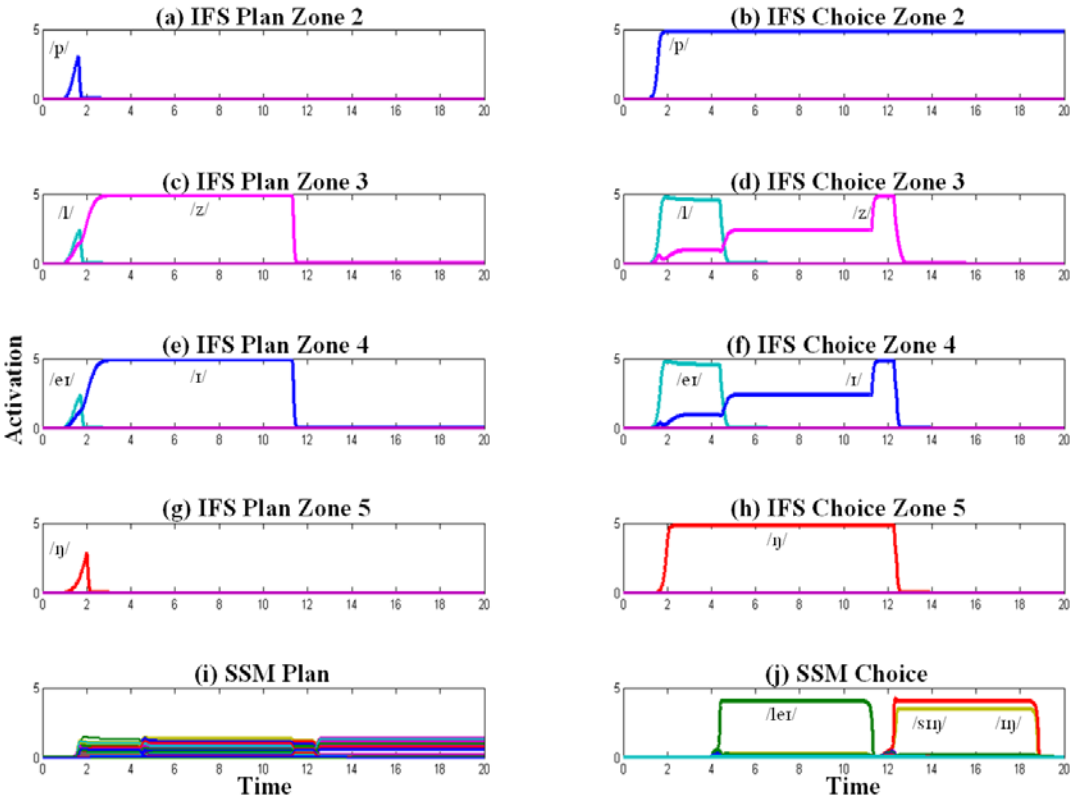


Figure 4.12. *Example SSM damage* shows the IFS plan layers for zones 2 **(a)**, 3 **(c)**, 4 **(e)**, and 5 **(g)**, the IFS choice layers for zones 2 **(b)**, 3 **(d)**, 4 **(f)**, and 5 **(h)** as well as the SSM plan **(i)** and choice **(j)** layers for SSM damage. The input word was 'pleasing' (/pleɪ zɪŋ/). The resulting utterance was (/leɪ sɪŋ/). In **(j)**, /zɪŋ/ corresponds to the top plot in red and /ɪŋ/ corresponds to the bottom plot in yellow.

When these lesions were applied, the model had difficulty in articulating the correct speech sounds. For example, when the model experienced 25% damage and was expected to produce the word 'pleasing' (/pleɪ zɪŋ/), the model produced /leɪ sɪŋ/; see Figure 4.12. Thus the model produced a groping error, a consonant cluster reduction, and a voicing error. As can be seen in Figure 4.12j, the groping error was found in the second syllable where both /sɪŋ/ and /ɪŋ/ are activated in the SSM. In addition, the IFS

layers showed normal operation. In the resulting utterance /p/ was deleted from the consonant cluster /pl/ resulting in a consonant cluster reduction and in the IFS choice layer, zone 2, the /p/ was activated, but its activity was never quenched. This problem occurred because the /pleɪ/ speech sound representation was damaged. Thus the model, although knowing it needed to produce the /p/ was unable to do so and chose the next most active speech sound in its place, /leɪ/. In addition, there is a voicing error on the first consonant of the second syllable where /z/ was replaced with /s/. This is a fairly common error in AOS and reflects damaged to the /z/ or /zɪŋ/ representation in the SSM.

The average of the errors over the different lesion scenarios of the model for each input word was analyzed in the same manner as the data was analyzed for damage to the IFS region; see Figure 4.13.

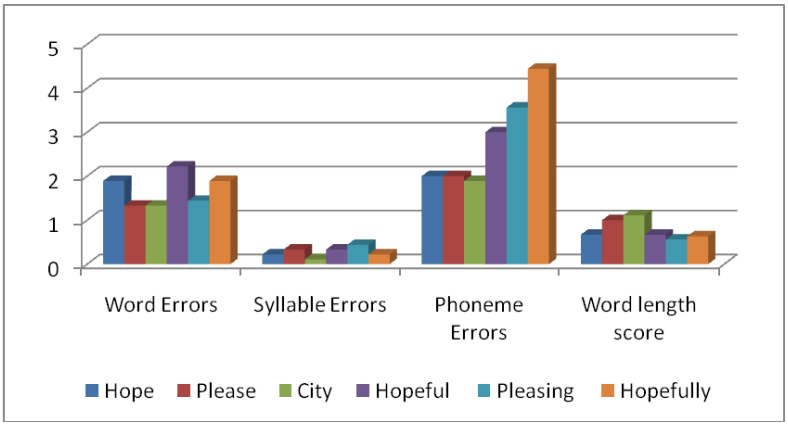


Figure 4.13. *Error analysis with SSM damage*. For each of the 6 input words, the average number of word, syllable, and phoneme level errors are illustrated. The bars on the far right represent the word length score.

Both the word level and syllabic level errors did not vary with word length. However, phonemic errors did vary with word length. In order to determine if this is a result of there being more phonemes in the longer words, the word length score was evaluated; see the last group of bars in Figure 4.13. The word length score confirms that the length effect in the phoneme errors was not a result of the word length, but rather of the increased number of phonemes. In fact both ‘please’ and ‘city’ scored the highest with 1 and 1.11 respectively. ‘pleasing’ scored the lowest with 0.56. Thus, damage to the SSM distribution of the model did not affect performance as a function of word length. Figure 4.14 shows further the amount of errors as they varied with the amount of damage in the SSM.

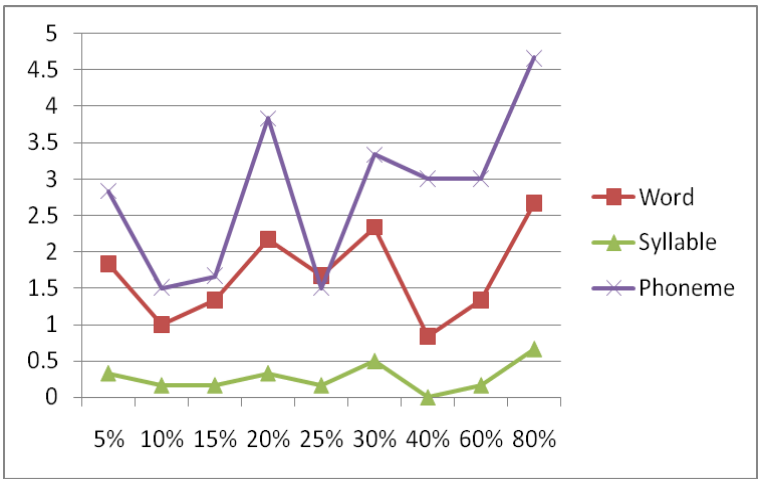


Figure 4.14. *Errors as SSM damage was varied*. The three plots show the average word, syllable, and phoneme level errors for the 6 input words as the damage was varied from 5% to 80%.

Word	Word errors	Syllabic errors	Phonemic errors
Please	*Abnormal prosody* *Word elongation* *Syllable boundary*	*Insertion*	Groping *Vowel elongation* Consonant cluster reduction *Deletion* Perseverative *Schwa insertion* *Vowel error* *Insertion* *Consonant error*
Hope	*Abnormal prosody* *Word elongation* Delayed initiation *Word shortening* *Syllable boundary*	*Insertion*	*Schwa insertion* *Voicing* *Insertion* *Deletion* *Consonant error* Groping *Vowel elongation*
Pleasing	*Word elongation* *Syllable boundary* *Abnormal prosody*	*Insertion*	*Voicing* Groping *Consonant error* Perseverative Consonant cluster reduction *Vowel error* *Vowel shortening* *Insertion* Transposition *Schwa insertion* *Deletion*
Hopeful	*Abnormal prosody* *Word elongation* *Syllable boundary* Delayed initiation	*Insertion*	*Voicing* Groping *Insertion* *Schwa insertion* *Consonant error* *Vowel error* Anticipatory Transposition Perseverative *Deletion*
City	*Abnormal prosody* *Word lengthening* *Syllable boundary* *Word shortening*	*Deletion*	*Voicing* *Insertion* Perseverative Groping *Consonant error*
Hopefully	*Abnormal prosody* *Syllable boundary* Delayed initiation *Word shortening* *Word elongation*	*Deletion*	Transposition *Voicing* *Insertion* Groping *Vowel error* *Deletion* *Consonant error* Anticipatory Perseverative *Vowel elongation*

Table 4.6. Encountered errors (in order from most to least encountered) when the SSM region was damaged. Errors in italics are highly indicative of AOS. The other errors may occur in AOS, but are also found in other motor speech disorders.

Similar to what was seen with IFS damage, there were fewer syllable level errors than word or phoneme errors. With SSM damage, however, there were clearly more phoneme level errors than any other type. Also, unlike the IFS damage, there was not a clear trend of increasing number of errors with increased damage.

In order to better understand the types of errors that are committed by the model when the SSM distribution was damaged, the error types are summarized in Table 4.6. Abnormal prosody and word elongation were the most commonly encountered word level errors. Very few syllabic level errors were encountered and these were only insertion and deletion errors. The majority of errors were phonemic level with the most common error types being voicing and groping. Consonant cluster reductions were seen for both ‘please’ and ‘pleasing’.

4.2.4 Discussion

The simulations presented in this chapter are supportive of the hypothesis put forth by the DIVA model to explain AOS. DIVA predicts that there are two subtypes of AOS: one with damage to the IFS resulting in AOS type 1 errors and one with damage to the SSM resulting in AOS type 2 errors. AOS type 1 errors can be defined as fluent productions of the wrong speech sounds or misplacement of speech sounds within the motor program. This includes fluent substitutions, additions, transpositions, and omissions, few distortions, and few phonemic paraphasias. AOS type 1 patients should also exhibit relatively normal coarticulation and prosody. On the other hand, AOS type 2 errors can be classified as poorly articulated approximations of the desired syllables,

including: slurring, distortions, substitutions, deletions, some phonemic paraphasias, reduced co-articulation, nonfluent speech and dysprosody.

The two hypothesized AOS subtypes may result from damage to two regions: the IFS and surrounding regions in the superior portion of the IFG and the FO and possibly extending into the inferior IFG as well as anterior portions of the insula for AOS type 1 and 2 respectively. This is consistent with AOS lesion studies suggesting that damage to these regions results in AOS behavioral characteristics (Hillis *et al*., 2004; Ogar *et al*., 2006; Dronkers, 1996; Mohr, 1976; Mohr *et al*., 1978; Marquardt and Sussman, 1984; Alexander *et al*., 1990; Marien *et al*., 2001). Furthermore, patients with neurodegenerative diseases that exhibit AOS symptoms often show atrophy in the IFG, IFS, and/or FO (Nestor *et al*., 2003; Broussolle *et al*., 1996; Gorno-Tempini *et al*., 2004; Sanchez-Valle *et al*., 2006). Unfortunately, many of these lesion studies have not explicitly differentiated lesions between the superior IFG/IFS region and the inferior IFG/FO/anterior insula region.

Nevertheless, studies looking at these brain regions in normal subjects suggest that there may be such a functional specialization within the IFG. Notably, the more superior regions are more activated by phonological tasks (Gitelman *et al*., 2005; Poldrack *et al*., 1999), sound sequence manipulations (Bohland and Guenther, 2006; Gelfand and Bookheimer, 2003), task difficulty (Chein *et al*., 2002), and during both motor imitation and observation (Molnar-Szakacs *et al*., 2005). Conversely, the more posterior regions are more sensitive to stimuli structure and complexity (Chein and Fiez, 2001; Bohland and Guenther, 2006; Fiez and Peterson, 1998) and only during

imitative motor tasks (Molnar-Szakacs *et al*., 2005). Thus, suggesting that the IFS and surrounding regions are concerned with phonological representations and sequential speech motor control whereas the FO region is more concerned with storing and representing the speech motor programs for articulation. This is consistent with the AOS behavioral characteristics associated with damage to each of these regions and exemplified in the simulations.

IFS and SSM lesions produced different patterns of behavioral characteristics. Many errors that resulted from damage to the IFS and not the SSM region were dependent on the length of the intended utterance such that longer words produced more errors. This is because the IFS damage led to more problems with sequencing of the speech sounds. There were also far fewer errors as a result of damage to the IFS such that damage to the IFS region, in general, did not have as great of an effect on performance: in nearly half of the simulations performed with IFS damage, no errors were exhibited by the model. However, the words evaluated in the simulations were short and relatively simple. It is likely that as the word length and difficulty increases, more errors as a result of IFS damage will occur.

Damage to IFS resulted in more syllabic level sequencing errors. Both IFS damage and SSM damage resulted in similar word level error patterns with most of those errors being prosodic abnormalities and word elongation. Prosodic abnormalities are a primary characteristic of AOS and are often marked by stress changes, word elongation, prolongation of consonants and vowels, increased duration of pauses between sounds, and insertion of transitional vowels, such as schwa insertions (Square

et al., 1982; Kent and Rosenbeck, 1983; Masaki *et al*., 1991; Odell *et al*., 1991; *Seddoh et al*., 1996; Strand and McNeil, 1996; Dabul, 2000). The damage to the IFS, for the most part, led to fluent productions, very few word level errors, and very few prosodic based phonemic disturbances. On the other hand, damage to the SSM led to largely nonfluent productions and frequent prosodic disturbances; in particular, the model exhibited schwa insertions, syllabic boundary changes, and vowel elongations. Prosodic disturbances are thus present in both AOS types, but are more often seen in AOS type 2.

Based on the ABA-2 (Dabul, 2000) a majority of the types of phonemic errors exhibited by the model are indicative of AOS. The damage to the SSM led to far more phonemic errors than did the IFS damage. Damage to the IFS resulted in phonemic errors more related to sequencing (e.g. deletions, insertions, etc) than to misapproximation errors (e.g. substitutions, voicing, and groping errors). This is most likely due to the role of the IFS in sequencing speech sounds. Due to the damage, the model may be attempting to parse the words on the phonemic level and thus we see such errors.

Damage to the SSM, on the other hand, directly affected the representation of the stored sounds in the SSM cells. Most of the errors are considered close approximations or distortions of the desired syllables. These types of errors are regarded as the predominant segmental errors type in AOS (McNeil *et al*., 2007). Voicing and groping (McNeil *et al*., 1995) are two common errors of this type. Usually, the target word is recognizable, and perceived substitutions tend to differ from the target in only one or two articulatory features such as voicing or place of articulation (Odell *et al*.,

1990; Sugishita *et al*., 1987; Trost and Canter, 1974). The vowel and consonantal errors produced by the model with SSM damage tended to follow this pattern. In addition, errors such as vowel elongations are indicative of groping for the correct speech sound. Vowel elongation errors were exhibited by apraxic patients (Rogers *et al*., 1996) and were more readily found in the AOS type 2 simulations.

Not only did the model exhibit errors on the word, syllable, and phoneme level but the increasing length and complexity of the word also allowed for variation in the errors. When the model contained lesions to the IFS, increasing the word length resulted in increasing the number of errors. On the other hand, damage to the SSM did not show this effect. The ABA-2 (Dabul, 2000) as well as patient studies (Deal and Darley, 1972; Johns and Darley, 1970; LaPointe and Horner, 1976; Square *et al*., 1982; Odell *et al*., 1990) indicate that increasing errors with increasing word length is a characteristic of AOS, yet this seems to only be the case in AOS type 1 errors.

Increasing the amount of damage also had different effects when the model had IFS or SSM lesions; as the amount of IFS damage increased, the amount of errors increased. IFS damage directly affected the sequencing of the speech sounds in the intended utterance. With SSM damage, there was not a clear effect on the number of errors based on the amount of damage. SSM damage affected the representation of the speech sounds in the SSM. Thus, errors were only exhibited when the damage had impacted the speech sounds that were being spoken. It may be that a stronger correlation between SSM damage and errors will emerge when the model simulates more words as when producing phrases or sentences.

The damage that the SSM encountered in the simulations will also lead to impairments in the feedforward commands in DIVA. The feedforward control system may be unable to access the motor programs for the damaged speech sounds in the SSM. Furthermore, these commands may not be able to be updated based on perceived speech errors which would result in difficulties in compensating for errors. Only partial damage to the SSM representation may lead to a partial readout of the motor program. The produced speech would be expected to be relatively slow, poorly articulated, poorly timed, prosodically abnormal, and possibly missing phonetic units due to the loss of portions of the feedforward motor program consistent with the simulation results.

The simulations also exhibited two key characteristics of AOS that are not readily studiable by functional models: islands of fluency and inconsistent error patterns. The simulations contained islands of fluency by occasionally producing the intended utterance perfectly even when the model was substantially damaged. This is consistent with the behaviors observed in apraxic patients. In addition, the presence of the speech errors varied from attempt to attempt but the type of error and its location within the utterance tended to be relatively consistent. This type of behavior is also often observed in AOS patients (Mauszycki *et al*., 2997; McNeil *et al*., 1995; Sugishita *et al*., 1987; Wambaugh *et al*., 2004; Mlcoch *et al*., 1982).

Three other accounts of AOS based on the Levelt model (Levelt, 2001; Levelt *et al*., 1999; Roelofs, 1997; see Section 3.3.1 for an overview) have been put forth as functional interpretations of this disorder: the *dual route hypothesis* (Varley and Whiteside, 2001a,b; Whiteside and Varley, 1998), the *reduced buffer capacity*

hypothesis (Rogers and Storkel, 1999), and the *damaged motor program hypothesis* (Aichert and Ziegler, 2004). The dual route hypothesis places the locus of the deficit in the phonetic encoding stage of the Levelt model (Varley and Whiteside, 2001a,b; Whiteside and Varley 1998), specifically in the *direct* route of phonetic encoding (i.e., impaired syllabary access). Thus, speakers with AOS are forced to rely on the *indirect*, phoneme-by-phoneme route of phonetic encoding. This indirect route is more resource-intensive and error-prone, resulting in a greater number of speech errors, loss of automaticity, prolongations of segments, reduced coarticulation, increased variability, and reduced in speech rate.

The simulations presented are not consistent with the dual route hypothesis (see Section 3.3.1) for several reasons. First of all, the simulations showed a significant amount of sound distortions. The dual route theory does not address these types of errors in AOS. Secondly, as was seen in Figures 4.9 and 4.12, many of the errors seen in AOS do not necessarily result from a phoneme by phoneme sequencing paradigm. In fact, this type of method was only used less than one percent of the time by the model in the simulations presented. Thus, it seems that AOS patients may do not construct words by an indirect phoneme by phoneme route.

The reduced buffer capacity hypothesis places the locus of deficit in AOS at the level of the articulatory buffer in the Levelt model (Rogers and Storkel, 1999). Specifically, AOS is thought to reflect a limitation of the buffer capacity to a single syllable, forcing speakers with AOS to program utterances one syllable at a time. This proposal accounts for core symptoms such as syllable segregation, dysprosody, and

slow speech rate. Similar to the dual route theory, this account of AOS does not take into consideration the large amount of sound distortion errors seen in AOS patients and shown in the simulations presented in this chapter. Furthermore, the simulations showed that the apraxic patient can hold the motor plan in the IFS region and that the SSM is able to effectively choose more than one syllable at a time for production. In addition, moving syllable boundary errors, as seen in the simulations, are inconsistent with this account because these errors require the two syllables to be held in the articulatory buffer at the same time. Last, anticipatory errors, both at the level of the syllable and the phoneme, as illustrated in these simulations and produced by apraxic patients, cannot be accounted for by the reduced buffer capacity hypothesis because access to more than one syllable is necessary to create such an error.

Finally, Aichert and Ziegler (2004) propose a damaged motor program hypothesis in which the syllabary representations themselves may be damaged, although the degree of damage may vary with syllable frequency and complexity. This account can explain distortion errors in AOS and relative consistent error type and location within the utterance. This hypothesis corresponds approximately to the AOS type 2 simulations which describe herein damage to the SSM cells. However, this hypothesis does not address the AOS type 1 errors.

The DIVA model predictions of AOS which are further clarified and supported by the simulations in this section provide the most comprehensive account of AOS. Not only does DIVA explain the errors as seen in AOS patients, but it also makes a further

prediction by defining two subgroups of patients based both on behavioral characteristics and brain lesion location.

4.3 Case study

In order to further test the predictions made by the DIVA model and the simulations presented in the previous section, we evaluated a patient with AOS in order to assess a correlation between lesion location and behavioral characteristics. The patient was recruited based on brain lesion characteristics, notably she has a large anterior left hemisphere lesion resulting from stroke. The lesion extends into the frontal opercular area but leaves the IFS region relatively intact. The extent of white matter damage is unknown. Based on this lesion evidence, it was hypothesized that this patient would exhibit AOS type 2 errors. A speech language evaluation was conducted to verify this hypothesis.

4.3.1 Methods

The subject evaluated in this study was recruited from the Harold Goodglass Aphasia Research Center (HGARC) at the Veteran's Affairs (VA) Boston Healthcare System. All subjects in the HGARC database have agreed to be contacted for participation in affiliated research projects. The subject was selected for participation based on a neurological assessment and a determination of a significant number of phonemic paraphasias and articulatory disturbances in her preliminary speech evaluations.

The behavioral study included the following elements: a brief case history, an oral-peripheral examination (this involved looking inside and outside the mouth to

determine if the structure and function of the oral articulators is intact), a cranial nerve examination, a series of motor speech tasks (e.g. word and sentence repetition tasks), and the ABA-2 (Dabul, 2000). The session was audio-taped and video-taped for offline analysis.

4.3.2 Case history

A 75 year old, right handed, native English speaking woman began to experience right-sided weakness and difficulties with walking and speaking while she was playing golf. An initial CT scan upon admittance to the hospital did not reveal hemorrhage. A follow-up CT scan after treatment for atrial fibrillation did indicate a large left middle cerebral artery (MCA) cerebral vascular incident (CVI), but no hemorrhage. An MRI performed the day after the stroke indicated that the patient had a large early subacute left MCA territory infarct involving the basal ganglia and the insula. The patient was transferred to a rehabilitation hospital one week post stroke and various other rehabilitative facilities until she returned home seven months post stroke.

The patient's medical history consisted of hypertension, high cholesterol, colon cancer (8 years before stroke onset), and tobacco use (quit 33 years before stroke onset). The patient presented at the initial physical examination as thin but in good health. She had no significant facial weakness. Cranial nerve examination showed no damage. The patient was noted to be aphasic with limited verbal output and had dysphagia (difficulty in swallowing). She had weakness of the right side and was flaccid in the right upper and lower extremities.

One month post stroke, the patient presented with severe aphasia, minimal oral-pharyngeal dysphasia, and oral apraxia of speech. Auditory comprehension was intact and she was able to follow simple two step commands and answer yes and no questions. Verbal expression was initially marked with stereotyped ‘no’ responses. Familiar single words and some familiar phrases were able to be produced spontaneously. Words were often marked with articulatory errors. Writing was also impaired, but reading comprehension was not. Oral reading was difficult and contained many articulatory errors.

The patient received occupational, physical, and speech therapy during her rehabilitative time. The therapy allowed the patient to return home and live independently. Seventeen months post stroke, the patient was able to cook, shop for groceries, do laundry, play golf, garden, socialize with friends, and drive.

4.3.3 Neurological evaluation

Figure 4.15 shows the MRI scans of the subject taken the day following stroke. The MRI scans contained some motion artifacts which lead to difficult evaluation. The scans indicate that there is massive damage to the peri-Sylvian area and into the adjacent white matter. Subcortical damage includes the entire putamen and globus pallidum, portions of the caudate (sparing some of the caudate head), the entire external capsule and large portions of the internal capsule. The lateral portion of the thalamus may have also been affected. Cortical damage included nearly all of the anterior and posterior insula, the posterior portion of the frontal operculum, and all of the central and parietal opercula.

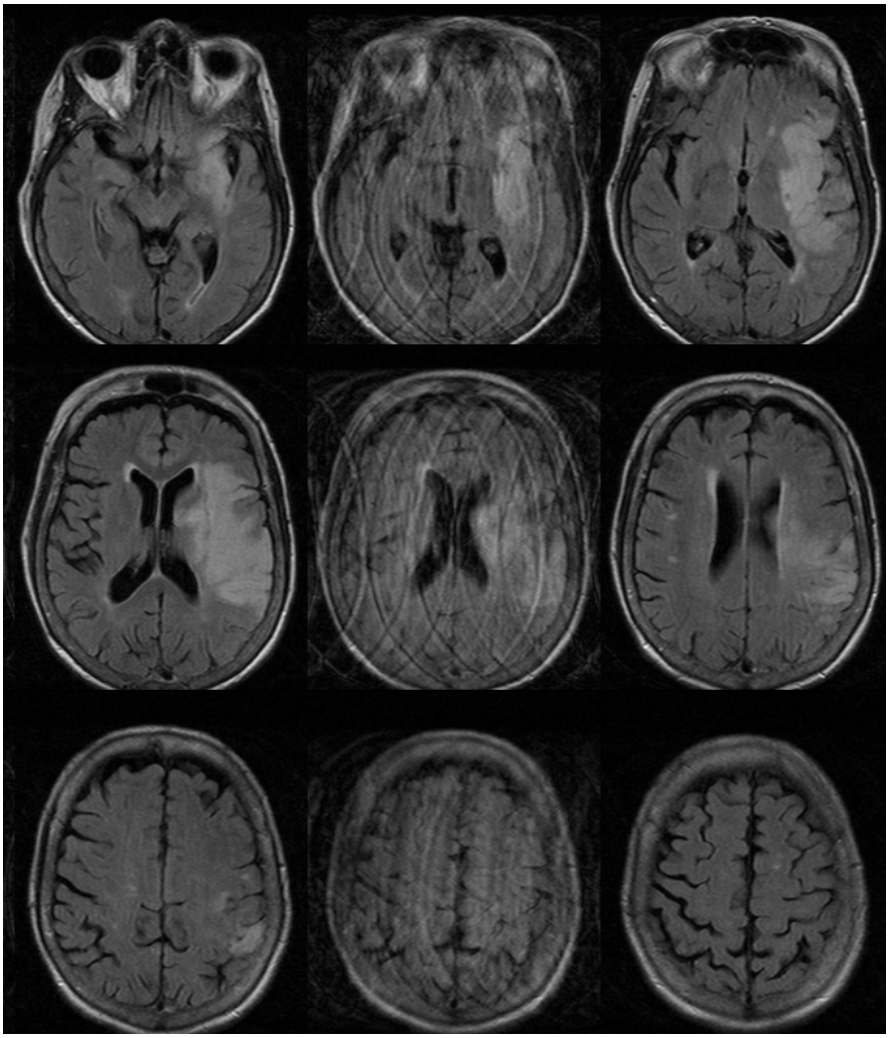


Figure 4.15. MRI scans demonstrating damage to the frontal operculum and adjacent anterior insula, with the inferior frontal sulcus region being intact.

Nearly the entire inferior parietal cortex was damaged, including the ventral somatosensory cortex. It is unclear how far the damaged region extends into the dorsal/lateral temporal lobe. Much of the ventral motor and premotor cortex appeared to be spared except for the ventral-most portions near the Sylvian fissure, although the MRI scans cannot confirm this finding with one hundred percent certainty. It also

appears that most of the lateral inferior frontal cortex was spared. The frontal operculum damage did not appear to extend to the lateral surface and much of the anterior inferior frontal sulcus was spared. Only the region right at the inferior frontal - precentral sulcal junction appeared to be potentially damaged.

Thus, in addition to a lot of damage outside the inferior frontal region, it appears that this patient has damage to the frontal operculum and adjacent anterior insula, with the inferior frontal sulcus region being intact. Based on the hypotheses set forth in the previous section, this patient would be considered an AOS type 2 patient.

4.3.4 Speech language evaluation

Approximately 16 months post stroke, the patient underwent a battery of neurological and speech exams to evaluate progress during her treatment and to reassess treatment plans. At this point in time, the patient had non-verbal output marked with occasional, complex, spontaneous fluent sentences. Her speech was also marked with many phonemic paraphasias and perseverations. She frequently interchanged ‘yes’ and ‘no’. She made some use of hand gestures but was unable to write. She had right-side hemiparesis and exhibited some impulsivity. Comprehension was relatively spared and she was able to follow complex, multistep commands.

Table 4.7 summarizes the results of the patient’s neuropsychological exam. Based on education level, age, and occupational attainment as well as performance on tests of intellect and premorbid functioning, it was determined that the patient functions at least in the high average range. The patient did not experience difficulty with inattention during testing but showed mild-moderate impairment for auditory-verbal

material. Verbal memory was difficult to assess due to the patient's limited speech output abilities. She performed within the normal range on a recognition test for structured material, but had difficulty with an unstructured task. The false positive errors suggest some executive dysfunction. Nonverbal memory was found to be in the low average to mildly impaired range for her age. Testing of executive functioning indicated an impulsive response style and a mild impairment in learning abstract rules and shifting sets when provided with minimal guidance and structure. The patient's visuospatial abilities seemed to be only impaired in the use of the non-dominant hand. The patient also exhibited mild depressive symptoms and attributed these thoughts to the stroke.

Subtest	**Raw Score**	**Standardized Score**	**Level of Impairment**
Intellectual			
TONI-3	16/45	SS=92	Average
LOFT	49/50	SS=125	Superior
Attention/Working Memory			
Spatial Span:	17/32	SS=15	Superior
Forward Span	6		
Backward Span	6		
Digit Span:	6/30	SS=4	Mild-moderate impairment
Forward Span	4		
Backward Span	2		
Auditory-Verbal Memory			
HVLT:			
Immediate Recognition	8/12, 2 false positives		
Delayed Recognition	8/12, 5 false positives		
WMS-R Logical Memory:			
Immediate Recognition	11/11		
Delayed Recognition	10/11		
Nonverbal Memory			
BVMT:			

Trial 1	1	T=32	Mild-moderate impairment
Trial 2	4	T=36	Mild impairment
Trial 3	8	T=49	Average
Total Recall	13	T=38	Mild impairment
Learning	7	T=67	Superior
Delayed Recall	5	T=39	Mild impairment
Percent Retained	62.5%	6-10th percentile	Mild impairment
Recognition Hits	5/6	>16th percentile	Average
False Positives	0/6	>16th percentile	Average
Copy	12		WNL
WMS-III Faces:			
Immediate Recognition	26/48	SS=6	Mild impairment
Delayed Recognition	27/48	SS=7	Low average
Executive Functioning			
Clock in the Box	6/8		Moderate impairment
Trails Part A	92", 1 error	T=18	Severe impairment
Trails Part B	d/c at 301", 2 errors	T=16	Severe impairment
Coin switch task	1/3		Mild impairment
Visuospatial Abilities			
Scan Jr:			WNL
Boston Visuospatial Quantitative Battery:			
Addition	4/6		
Subtraction	7/7		
Multiplication	2/2		
Division	3/3		
Depressive Symptoms			
GDS	11/30		Mild

Table 4.7. Results of neuropsychological evaluation taken 16 months post stroke. LOFT = Lexical Orthographic Familiarity Test, HVLT = Hopkins Verbal Learning Test (12 item list learning task), WMS-R is a test of memory of a logically organized story, BVMT = Brief Visuospatial Memory Test (nonverbal memory for shapes and their locations), and GDS = Geriatric Depression Scale.

At this time, the patient also participated in a speech and language evaluation that included the Boston Diagnostic Aphasic Examination (BDAE; see Table 4.8), Boston Naming Test (BNT; see Table 4.9), and the Test of Adolescent and Adult Word Finding (TAAWF; see Table 4.10).

Subtest	**Raw Score**	**Percentile**
Fluency		
Articulation Rating	5.5	45
Phrase Length	2-7 (variable)	25-70
Melodic Line	5 (variable)	40
Verbal Agility	5 (variable)	30
Nonverbal Agility	9	75
Auditory Comprehension		
Word Discrimination	70	90
Body-Part Identification	18	80
Commands	13	70
Complex Ideational Material	9	80
Naming		
Responsive Naming	10	50
Confrontation Naming	75	65
Animal Naming	6	80
Oral Reading		
Word Reading	18	65
Oral Sentence Reading	2	60
Repetition		
Repetition of Words	8	50
High-Probability	5	70
Low-Probability	0	40
Paraphasia		
Neologistic	3	35
Literal	16	15
Verbal	16	30
Other	3	30
Automatic Speech		
Automatized Sequences	5	55
Reciting	2	80

Reading Comprehension		
Symbol Discrimination	10	70
Word Recognition	7	60
Comprehension of Oral Spelling	1	40
Word-Picture Matching	9	60
Reading Sentences and Paragraphs	9	90
Writing		
Mechanics	3	40
Serial Writing	39	70
Primer-Level Dictation	14	80
Spelling to Dictation	5	80
Written Confrontation Naming	4	65
Sentences to Dictation	2	70
Narrative Writing	1	20
Music		
Singing	2	40
Rhythm	0	10

Table 4.8. Results of BDAE taken 16 months post stroke.

Subtest	**Raw Score**
Spontaneously correct	22/60
Correct following stimulus cue	0
Correct following phonemic cue	17

Table 4.9. Results of BNT taken 16 months post stroke.

Subtest	**Possible Points**	**Score**	**Percent Correct**
Picture Naming-Nouns	37	10	27%
Comprehension-Nouns	37	29	78%
Sentence Completion Naming	16	4	25%
Sentence Comprehension	16	13	81%
Descriptive Naming	12	1	8%
Word Comprehension	12	10	83%
Picture Naming-Verbs	21	3	14%
Comprehension-Verbs	21	17	81%
Category Naming	21	5	24%
Comprehension	21	20	95%

Table 4.10. Results of TAAWF taken 16 months post stroke.

The results of the BDAE found that the patient had an aphasia severity rating of 2 (60th percentile) which means that the patient is able to converse about familiar subjects with help from the listener. There may be frequent failures to convey the idea but the patient shares the burden of communication with the listener. Furthermore, administration of the BDAE found that:

1. The patient has a variable melodic line, but can carry intonation almost through an entire sentence.
2. The patient produces a variable phrase length that runs from two to seven words without interruption.
3. The patient's articulatory agility was mildly impaired.
4. The patient was able to produce grammatically correct constructions, but these were limited to stereotypical constructions.
5. The patient had few paraphasias present in running speech.
6. The information contained in the patient's utterances was proportional to fluency.
7. The patient had unimpaired auditory comprehension.

Overall, the BDAE confirmed that the speech errors exhibited by the patient were mild to moderately aphasic. The patient showed the most impairment on fluency, naming, repetition, and paraphasias. Nonfluency is a primary characteristic of AOS type 2. In addition, the impairments on naming, repetition, and paraphasias suggests that the patient may be exhibiting misarticulations on word productions. Both the BNT and the

TAAWF confirm that the patient has difficulty in naming. Both the BDAE and the TAAWF scores show that the patient exhibits no comprehension deficit.

In order to better understand the potential articulation difficulties exhibited by the patient, at twenty months post stroke, the ABA-2 (Dabul, 2000) was administered along with a oral peripheral and cranial nerve examination, a motor speech evaluation, and a hearing test. Her pure tone hearing thresholds were at or above 25 dB in both ears. Jaw, lip velopharyngeal and laryngeal function were within normal limits. Weakness and incoordination were noted for tongue movements. Limb movements were also within normal limits, although she did require a model for longer multi-step commands. She was ambulatory but still exhibited some right sided flaccidity. Her overall rate of speech was slightly slowed and often showed overt difficulty with word finding and formulation.

On the motor speech exam, the patient exhibited minor sound errors (e.g. /θ/ for /s/; /r/ for /l/; and /v/ for /b/) and occasional visible groping. Consonant clusters were reduced and simplified. At the sentence level, as her rate of speech increased, her articulatory precision decreased. Otherwise she was able to accurately produce even multisyllabic words and sentences with few/no errors. Especially noteworthy is the fact that increasing word length did not particularly affect the degree or type of errors.

On the ABA-2, the patient scored in the mild range of impairment for all but the increasing word length subtest B; see Table 4.11. Rather than indicating a word length effect, this subtest was more useful in testing very difficult words. Performance was relatively well preserved on polysyllabic words. Errors on longer words tended to be

substitutions or reduced consonant clusters; see Table 4.12 for a summary of some of the errors made by the patient. She also inserted the phoneme /s/ at the beginning of some multisyllabic words which could be viewed as a compensatory strategy for having difficulty with initiation of speech movements.

Subtest	Raw Score	Level of Impairment
Diadochokinetic rate	11	Mild
Increasing word length A	2	Mild
Increasing word length B	5	Moderate
Limb apraxia	45	Normal
Oral apraxia	40	Mild
Utterance time for polysyllabic words	40	Mild
Repeated trials	21	Mild
Inventory of articulation characteristics	12	

Table 4.11. Results of ABA-2 taken 20 months post stroke.

Increased practice with the same word did not always yield improvements in articulation. Her speech errors were considerably more pronounced when reading aloud. She had anticipatory errors and some voicing errors in spontaneous and read speech. Her automatic speech was relatively intact with only occasional phonemic errors. Her speech production was marked by audible and visual groping behaviors and interjections/fillers (e.g. ‘um’ and ‘uh’).

Word	Correct IPA spelling	Incorrect articulation	Error description
Thicken	/θɪk ən/	/θɪs ən/	Consonant error
Jabber	/dʒæb ər/	/zæb ər/	Consonant cluster reduction and error
Jig	/dʒɪg/	/zɪg/	Consonant cluster reduction and error
Pleasing	/pleɪ zɪŋ/	/peɪ zɪŋ/	Consonant cluster reduction
Catastrophe	/ka tæs trə fi/	/ta tæs trə fi/	Consonant error
Snowman	/snoʊ mæn/	/ʃoʊ mæn/	Consonant cluster reduction and error
Thigh	/θaɪ/	/saɪ/	Consonant error

Table 4.12. Some misarticulated word examples from the ABA-2.

Overall, the patient presented with mild apraxia of speech characterized by disturbances in speech rate (slow and labored productions), speech sound substitution errors, phonemic distortions, schwa and /s/ interjections, cluster reduction, and visible groping. The patient did not exhibit a word length effect and for the most part, her errors can be described as misarticulations of the wrong speech sound marked by substitution and reduction errors as well as groping. The patient did exhibit some prosodic abnormalities, but her speech was more characterized by approximations and fillers. She also did not exhibit any transpositions of speech sounds. Based on the behavioral analysis of the patient, her speech characteristics are consistent with mild AOS type 2 errors. This behavioral diagnosis is consistent with the lesion diagnosis.

4.4 General discussion

In this chapter, simulations of AOS making use of the GODIVA and DIVA models were presented along with a case study of an apraxic patient. The model simulations illustrated a clear distinction between two subtypes of apraxia of speech. The first subtype of AOS is defined by lesions to the IFS and superior portions of the IFG. This damage results in behavioral errors that can be characterized as fluent missequencing of the speech sounds in the intended utterance. This is because the damage results in impairment of the sequencing and planning mechanism of the human speech production system. The most common error types encountered in the simulations for this type of AOS were deletions and transposition errors on both the phoneme and syllable level. In addition, word level errors included some abnormal prosody and word shortening. The

number of errors increased with increasing word length, suggesting a word length effect for AOS type 1.

Conversely, AOS type 2 results from lesions to the FO and extending possibly into the anterior insula and the ventral regions of the IFG. This damage results in nonfluent misarticulations of speech sounds characterized by erroneous utterances including poor articulation, poor timing, missing phonetic units, frequent substitutions and insertions, and a significant number of prosodic disturbances. This is because the damage destroys the articulatory representations of the speech sounds. The most common error types encountered in the simulations for AOS type 2 was abnormal prosody, moving of syllable boundaries, word length changes, groping, and phoneme insertions and substitutions. In addition, many more errors were encountered with AOS type 2, but the errors did not vary with respect to word length. This indicates that AOS type 2 patients should not exhibit a word length effect.

The traditional, primary characteristics of AOS include: (1) overall slowed speech, due to sound prolongations, increased intersound durations, and schwa insertions; (2) abnormal prosody with excess and equal stress (stress on normally unstressed syllables); (3) sound distortions and distorted substitutions as the predominant error type; and (4) consistent error types from trial-to-trial, although the presence of the error is inconsistent (Wambaugh *et* al., 2006a). From this perspective, AOS type 2 will be more likely characterized as traditional AOS.

AOS type 1, on the other hand, only exhibited errors consistent with the second and the fourth general characteristic, although the other two may also be exhibited. The

word length effect, which is commonly associated with AOS diagnosis, is only seen in AOS type 1. Furthermore, because AOS type 1 errors can be considered serial misordering of speech sounds, these errors are often overlooked as speech planning deficits and instead classified as phonological encoding deficits (Levelt *et al*., 1999). As GODIVA has suggested, these types of errors can also be considered phonological sequencing and planning problems that are part of the speech motor control system rather than the language encoding system. Thus, although many errors seen in AOS type 1 may traditionally be viewed as aphasic errors, this may not necessarily be the case. In the clinical setting, however, the two primary characteristics used for determining the presence of AOS is sequencing and groping errors. Furthermore, the presence of groping is often used to rule out dysarthia as diagnosis. Thus, it seems that the clinical presentation of AOS contains characteristics of both AOS types.

The patient evaluated in the case study had experienced a large left MCA CVI which resulted in a large anterior left hemisphere lesion that extends into the frontal opercular region while leaving the IFS region relatively intact. Additional damage extends into the parietal lobe and subcortical structures including basal ganglia. Based on this lesion, the patient was initially classified as AOS type 2. A subsequent behavioral study of the patient concluded that the patient's speech was marked by disturbances in speech rate (slow and labored productions), speech sound substitution errors, phonemic distortions, schwa and /s/ interjections, cluster reduction, and visible groping. The patient did not exhibit a word length effect and for the most part, her errors can be described as misarticulations of the wrong speech sound marked by

substitution and reduction errors as well as groping. These findings were consistent with the errors exhibited by the model in the AOS type 2 simulations. In addition, the patient did not exhibit a word length effect, which is also consistent with the simulations. Thus, the results of the case study of the apraxic patient were consistent with the error predictions made by the model regarding AOS type 2.

This chapter introduced a functional hypothesis of AOS based on computational modeling and neuroanatomical data. Simulations were performed to confirm this hypothesis and the results were compared to an apraxic case study. This AOS hypothesis is the first analysis of AOS put forth that makes an attempt to dissociate different types of behavioral characteristics of AOS, relate those characteristics to particular lesions, and to correlate the damage with the functional attributes of the damaged brain regions. This level of detail is necessary to gain a better understanding of AOS and to aid in both diagnosis and treatment. Patients diagnosed with AOS type 1 should focus treatment efforts on restoring or compensating for sequencing and motor planning. AOS type 2 patients should focus treatment efforts on building new representations for stored speech sounds or compensatory mechanisms for accessing speech sounds in a more indirect route. Taken together AOS patients and speech pathologists can benefit greatly from this more detailed method of AOS diagnosis and treatment.

CHAPTER 5

CONCLUSION

This dissertation has combined computational modeling and behavioral analysis of both speech perception and production in order to provide new insights into the brain mechanisms involved in speech in both the normal and disordered brain. Furthermore, the research presented in this dissertation has attempted to investigate the common organizational properties and neural designs employed by both speech perception and production.

Two key areas of speech research were addressed in this dissertation: the ability of the human brain to build a speaker-invariant representation of heard speech sounds and the inability of apraxic patients to effectively generate motor speech programs. Both of these research projects were guided by computational modeling. NormNet, the computational model developed to explain speaker normalization, was confirmed through comparisons made to previously published psychophysical studies. Both GODIVA and DIVA models were used to explain apraxia of speech (AOS). These models were further elaborated and compared to an apraxic case study.

In Chapter 2, NormNet was presented and shown to provide a proof of principle for how the brain can carry out speaker normalization. It was able to achieve accuracy of nearly 80% correct, on average, which is comparable to the results obtained in human listeners identifying similar speech stimuli. NormNet supports the need of a computational modeling framework that makes use of both asymmetric competitive interactions and tonotopic strip maps of frequency-sensitive auditory cortical cells. In

auditory streaming, such an architecture is used to define auditory streams that characterize acoustic sources. In NormNet, this architecture is used to normalize the frequency spectra from streamed input. In turn, the model generates speaker-independent language categories in an adaptive resonance theory network. The adaptive resonance theory network is able to learn only a single category node for each vowel category in the simulations. When normalization was removed, the model generated many more category nodes and performance deteriorated. Thus, NormNet is able to understand speech across different speakers without creating a combinatorial explosion of speech representations in the brain.

A review of apraxia of speech (AOS) was presented in Chapter 3, and a functional study of AOS was presented in Chapter 4. Both the GODIVA and DIVA models were used to generate a key functional hypothesis regarding AOS and its relationship to brain lesions. The hypothesis defines two subtypes of AOS based on both brain lesion location and behavioral characteristics. AOS type 1 results from damage to the IFS including the superior portion of the IFG. The model predicts that damage would result in fluent missequencing errors including deletions and insertions. AOS type 1 does show a word length effect where errors increase as the word length increases. With respect to the model, AOS type 1 describes a breakdown in the sequencing and planning mechanism. AOS type 2 results from damage to the FO, possibly extending into the inferior portion of the IFG and the anterior insula. The models predict that damage would result in misarticulation errors including substitutions, insertions, groping, and abnormal prosody. No word length effect was

observed in the AOS type 2 simulations. With respect to the model, AOS type 2 results from damage to all or part of the distribution of the articulatory features describing the speech sound cells in the speech sound map.

An apraxic case study was presented and compared to the model predictions. This patient had a large anterior left hemisphere lesion that significantly damaged the frontal operculum region, but left the inferior frontal sulcus region was relatively spared. Thus, based on lesion information, this patient was classified as AOS type 2. This was validated with a behavioral study of the patient's speech errors. The patient's speech was characterized by disturbances in speech rate, speech sound substitution errors, occasional sequencing errors, groping, and a lack of a word length effect.

Although the results presented in Chapters 2-4 of this dissertation are significant, there are still many unanswered questions regarding speech perception and production. First of all, speaker normalization has been presented as a necessary and important component to speech perception; without it, listeners would not be able to understand speech from multiple speakers and on speech from the first encounter with a novel speaker. However, speaker normalization is also a key step in closing the loop between auditory representations of a baby's own heard babbled sounds to the motor commands that caused them (Piaget, 1963; Grossberg, 1978; Cohen *et al*., 1988). A further elaboration of NormNet that would allow for the closing of this loop would be an important step in bridging the gap between speech perception and production.

Additionally, the speaker normalization model makes use of an architecture based on tonotopic organization of frequency selective cells as well as asymmetric

competition. This architecture has also been proposed to be used in various other auditory cortical mechanisms, notably auditory streaming. The way in which tonotopic strip maps and asymmetric competition may be used in both streaming and speaker normalization circuits is a worthy topic of future research in order to clarify the importance of these mechanisms in auditory cortical processing.

The speaker normalization simulations evaluated NormNet's performance in a vowel categorization task. However, it would be important for future research to test the model with more complex stimuli including consonants, syllables, words, and even accented or whispered speech. Such simulations would be key in validating the performance of the model as well as the importance of speaker normalization in speech perception in general.

The predictions made by both GODIVA and DIVA regarding AOS need to be further validated and refined by looking at more patients with lesions to both the IFS and the FO regions. Currently, one apraxic case study has shown that the AOS type 2 predictions are consistent with lesion location and patient behavior. However, more patients need to be evaluated in order for the predictions to be validated.

One of the main characteristics of AOS, slowed speech rate, still needs to be captured and validated with simulations. The GODIVA model only explicitly represents order and not the precise timing or temporal dynamics of the speech utterances. By adding such temporal dynamics to the model, simulations would not only more realistically capture temporal characteristics of speech including prosody, stress

changes, and speech rate variations, but the model would be able to show how these temporal characteristics of speech are impaired in the apraxic patient.

Currently, the simulations presented to explain AOS only take into account two damage scenarios. They do not fully simulate how damage to the feedforward for feedback control systems may be impaired in AOS. These types of damage scenarios must be considered and simulated in order to evaluate how the entire speech production system may be affected in AOS patients.

The speech sound map in the DIVA model was elaborated to include distributed representations for the speech sounds based on three major articulatory features. This elaboration would be more realistic and more closely related to the errors seen in AOS patients if this distribution were spatially aligned such that speech sounds with common articulatory features are closer to each other.

Finally, the computational modeling work presented in this dissertation was used to evaluate both speech production and perception mechanisms. These models were used to present clearly testable predictions that were simulated and compared to human performance. The computational modeling work provides both a framework for future modeling studies and motivates future behavioral and neurophysiological studies aimed at discovering and explaining the common neural mechanisms and organizational properties in speech perception and production.

APPENDIX A

FUZZY ARTMAP EQUATIONS

ARTMAP is a neural network that is capable of both unsupervised and supervised incremental learning in response to sequences of binary input vectors presented in real time (Carpenter *et al*., 1991). Fuzzy ARTMAP can learn stable recognition categories in response to binary or analog input vectors (Carpenter *et al*., 1992). Learning always converges because all adaptive weights are monotonically increasing.

The fuzzy ARTMAP system consists of two adaptive resonance theory modules, ART_a and ART_b that are linked together by an inter-ART module, F^{ab}, called a map field (see Figure 2.6). During supervised learning, both modules receive a stream of input patterns: $\{\mathbf{a}^{(p)}\}$ and $\{\mathbf{b}^{(p)}\}$ where $\mathbf{b}^{(p)}$ is the correct prediction given $\mathbf{a}^{(p)}$. The inputs to the ART modules are $\mathbf{A} = (\mathbf{a},\mathbf{a}^c)$ for ART_a and $\mathbf{B} = (\mathbf{b},\mathbf{b}^c)$ for ART_b. These inputs are in a complement-coded form. Complement coding combines ON-cell and OFF-cell responses to prevent category proliferation by normalizing the amplitudes of the input feature vectors while preserving the amplitude of individual feature activations. To define complement coding, consider the ART_a module in which the input vector, **a**, is the ON-response. Then the complement of **a** is the OFF-response defined as:

$$\mathbf{a}_i^c = 1 - \mathbf{a}_i . \qquad (A.1)$$

Hence, the complement coded input **A** is a 2*M* dimensional vector:

$$\mathbf{A} = (\mathbf{a},\mathbf{a}^c) = (a_1,...,a_M,a_1^c,...,a_M^c) \qquad (A.2)$$

.

Each ART module contains a field, F_0^c and F_0^t, of cells that represent a current input vector **a** and **b**, respectively. The F_1^c and F_1^t feature fields receive complement-coded inputs **A** and **B** from F_0^c and F_0^t, respectively, and top-down learned expectations from the F_2^c and F_2^t active learned categories. The number of cell populations in each field is arbitrary. For ART$_a$, $\mathbf{x}^a = (x_1^a, ..., x_{2M_a}^a)$ is the F_1^c output vector, $\mathbf{y}^a = (y_1^a, ..., y_{N_a}^a)$ is the F_2^c output vector, and $\mathbf{w}_j^a = (w_{j1}^a, w_{j2}^a, ..., w_{j2M_a}^a)$ is the *j*th ART$_a$ adaptive weight vector. For ART$_b$, $\mathbf{x}^b = (x_1^b, ..., x_{2M_b}^b)$ is the F_1^t output vector, $\mathbf{y}^b = (y_1^b, ..., y_{N_b}^b)$ is the F_2^t output vector, and $\mathbf{w}_k^b = (w_{k1}^b, w_{k2}^b, ..., w_{k2M_b}^b)$ is the *k*th ART$_b$ adaptive weight vector. The adaptive weight vectors are associated with each F_2 category cell population j (j = 1,…, $2N_a$) for ART$_a$ and k (k = 1,…, $2N_b$) for ART$_b$. Each adaptive weight, or long-term memory (LTM) trace, of the weight vector is initially set to one indicating an uncommitted category. After the category is selected for coding, it becomes committed. In the present simulations, only the F_2^t field is implemented and its nodes are directly activated by category name labels.

The fuzzy ART module, ART$_a$, requires three parameters to be specified. These parameters are a choice parameter $\alpha > 0$, a learning rate parameter $\beta \in [0,1]$, and a vigilance parameter $\rho \in [0,1]$.

Category choice also occurs in both ART modules. The notation for the ART$_a$ module will be listed. The equations are the same for the ART$_b$ module except that the

superscript is *b* and the *j* subscript in the F_2 field is *k*. For each input **A** and F_2^c node *j*, the choice, T_j, is defined by:

$$T_j(\mathbf{A}) = \frac{|\mathbf{A} \wedge \mathbf{w}_j|}{\alpha + |\mathbf{w}_j|} \quad , \tag{A.3}$$

where the fuzzy and operator $\wedge$ is defined by:

$$(\mathbf{p} \wedge \mathbf{q})_i = \min(p_i, q_i) \quad , \tag{A.4}$$

and the norm | | is defined by:

$$|\mathbf{p}| = \sum_{i=1}^{M} |p_i| : \tag{A.5}$$

for any *M*-dimensional vectors **p** and **q**.

The ART$_a$ module makes a category choice when at most one F_2^c cell population is active at a given time. This category choice is indexed by *J*:

$$T_J = \max\{T_j : j = 1, \ldots, N\} \quad . \tag{A.6}$$

If more than one T_j is maximal, then the category *j* with the smallest index is chosen. These cells become committed in order of *j* = 1, 2, 3, … When the *J*th category is chosen, $y_J = 1$ and $y_j = 0$ for all $j \neq J$. The F_1^c activity vector, **x** = **A** when F_2^c is inactive and $\mathbf{x} = \mathbf{A} \wedge \mathbf{w}_J^a$ if the *J*th F_2^c node is chosen.

Resonance and reset are governed by the match value:

$$\frac{|\mathbf{A} \wedge \mathbf{w}_J^a|}{|\mathbf{A}|} \quad . \tag{A.7}$$

If *(A.7)* is greater than or equal to the vigilance, ρ, then resonance occurs and learning ensues, as defined below. Otherwise, mismatch reset occurs, which results in the choice

function T_J being set to zero for the duration of the input presentation to prevent persistent selection and learning of that category. A new index J is then chosen by *(A.6)* and the search continues until a chosen J achieves resonance.

Resonance triggers learning such that, once the search ends, the chosen weight vector, $\mathbf{w}_J^a$ is updated:

$$\mathbf{w}_J^{a(new)} = \beta(\mathbf{A} \wedge \mathbf{w}_J^{a(old)}) + (1-\beta)\mathbf{w}_J^{a(old)} \quad . \qquad (A.8)$$

Fast learning, as was used in the simulations in this paper, occurs when $\beta = 1$.

The map field, F^{ab}, links the two ART modules and is used to form predictive associations between ART_a and ART_b categories and to perform match tracking. The map field becomes active whenever one of the ART_a or ART_b categories is active, or when both are active only if ART_a predicts the same category as ART_b through the weights $\mathbf{w}_J^{ab}$. The output vector of the F^{ab} map field, $\mathbf{x}^{ab} = \mathbf{y}^b \wedge \mathbf{w}_J^{ab}$ if the Jth F_2^c category is active and F_2^b is active; $\mathbf{x}^{ab} = \mathbf{w}_J^{ab}$ if the Jth F_2^c category is active and F_2^b is inactive; $\mathbf{x}^{ab} = \mathbf{y}^b$ if F_2^c is inactive and F_2^b is active; and $\mathbf{x}^{ab} = \mathbf{0}$ if F_2^c is inactive and F_2^b is inactive. Thus, $\mathbf{x}^{ab} = \mathbf{0}$ when the prediction $\mathbf{w}_J^{ab}$ is disconfirmed by $\mathbf{y}^b$. This mismatch triggers an ART_a memory search, or hypothesis testing, for a better match via match tracking.

During match tracking, the vigilance parameter of ART_a, ρ_a, increases in response to a predictive mismatch with ART_b in order to ensure that predictive errors are not repeated on subsequent presentations of the input. The parameter ρ_a calibrates the minimum confidence that ART_a must have in a recognition category activated by

the input **A** in order for the ART_a module to accept that category. Smaller values of ρ_a lead to broader generalization and higher code compression. By match tracking, the minimum amount of generalization necessary to correct a predictive error is sacrificed. In other words, the ARTMAP system embodies a minimax learning rule in which the system strives to minimize predictive error while maximizing predictive generalization.

At the start of the input presentation, ρ_a equals the baseline vigilance and the map field vigilance parameter is ρ_{ab}. If

$$| \mathbf{x}^{ab} | < \rho_{ab} | \mathbf{y}^{b} | , \qquad (A.9)$$

then ρ_a is increased until it is slightly larger than the match value in *(A.7)*. Reset occurs and a memory search discovers the next ART_a category to learn. With fast learning, the map field weights $\mathbf{w}_{jk}^{ab} = 1$ for all time when *J* learns to predict the ART_b category name *K*.

APPENDIX B

CELEX LEXICAL DATABASE

DISC Vowel	IPA Symbol	Example	Height	Backness	Roundness
I	ɪ	pit	6	4	1
E	ɛ	pet	3	5	1
{	æ	pat	2	5	1
V	ʌ	putt	3	1	1
Q	ɒ	pot	1	1	2
U	ʊ	put	6	2	1
@	ə	another	4	3	1
i	iː	bean	7	5	1
#	ɑː	barn	1	1	1
$	ɔː	born	3	1	2
u	uː	boon	7	1	2
3	ɜː	burn	3	3	1
0[4]	æ̃ː	lingerie	2	5	1
1[5]	eɪ	bay	5.5	4.5	1
2[6]	aɪ	buy	3	4.5	1
4[7]	ɔɪ	boy	4.5	2.5	1.5
5[8]	əʊ	no	5	2.5	1
6[9]	aʊ	brow	4.5	4.5	1
7[10]	ɪə	peer	5	3.5	1
8[11]	ɛə	pair	3.5	4	1
9[12]	ʊə	poor	5	2.5	1
c[13]	ɐ̃	timbre	2	5	1
q[14]	ɑː	detente	1	1	1

Table B.1. Vowels and their features.

[4] 0 is the same as { in terms of features, but 0 is a longer and nasalized version of {
[5] 1 is a dipthong, the values for the features are averaged across those sounds
[6] 2 is a dipthong, the values for the features are averaged across those sounds
[7] 4 is a dipthong, the values for the features are averaged across those sounds
[8] 5 is a dipthong, the values for the features are averaged across those sounds
[9] 6 is a dipthong, the values for the features are averaged across those sounds
[10] 7 is a dipthong, the values for the features are averaged across those sounds
[11] 8 is a dipthong, the values for the features are averaged across those sounds
[12] 9 is a dipthong, the values for the features are averaged across those sounds
[13] c is the same as { in terms of features, but c is a nasalized version of {
[14] q is the same as # in terms of features, but q is a longer version of #

DISC Vowel	IPA Symbol	Example	Place	Manner	Voicing
p	p	pat	8	6	1
b	b	bad	8	6	2
t	t	tack	5	6	1
d	d	dad	5	6	2
k	k	cad	2	6	1
g	g	game	2	6	2
N	ŋ	bang	2	5	2
m	m	mad	8	5	2
n	n	nat	5	5	2
l	l	lad	5	2	2
r	r	rat	5	1	2
f	f	fat	7	4	1
v	v	vat	7	4	2
T	θ	thin	6	4	1
D	ð	then	6	4	2
s	s	sap	5	4	1
z	z	zap	5	4	2
S	ʃ	sheep	4	4	1
Z	ʒ	measure	4	4	2
j	j	yank	3	3	2
x	x	loch	2	4	1
h	h	had	1	4	1
C[15]	ŋ̩	bacon	2	5	1
F[16]	m̩	idealism	8	5	2
H[17]	n̩	burden	5	5	2
J[18]	tʃ	cheap	5	4.5	1
P[19]	l̩	dangle	5	2	2
R[20]	*	father	5	1	2
-[21]	dʒ	jeep	5	4.5	2
w	w	why	7	3	2

Table B.2. Consonants and their features.

[15] C is the same as N in terms of features, but C is stressed.
[16] F is the same as m in terms of features, but F is stressed.
[17] H is the same as n in terms of features, but H is stressed.
[18] J is an affricative so the Manner is an average of the two sounds.
[19] P is the same as l in terms of features, but P is stressed.
[20] R is the same as r in terms of features, but R is a linking r.
[21] - is an affricative so the Manner is an average of the two sounds.

REFERENCES

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.

Aichert, I. and Ziegler, W. (2004). Syllable frequency and syllable structure in apraxia of speech. *Brain and Language*, 88, 148-159.

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

Alexander, M.P., Naeser, M.A., and Palumbo, C.L. (1987). Correlations of subcortical CT lesion sites and aphasia profiles. *Brain*, 110, 961-991.

Ames, H.M. and Grossberg, S. (2006). Neural dynamics of auditory streaming, speaker normalization, and speech categorization. *Society for Neuroscience Abstracts*, Atlanta, GA.

Ames, H.M. and Grossberg, S. (2007). Speaker normalization using cortical strip maps: A neural model for steady state vowel identification. *Computational Cognitive Neuroscience Conference Abstracts*, San Diego, CA.

Ames, H. and Grossberg, S. (2008). Speaker normalization using cortical strip maps: A neural model for steady state vowel categorization. *Journal of the Acoustical Society of America*, 124(6), 3918-3936.

Amunts, K., Schleicher, A., Burgel, U., Mohlberg, H., Uylings, H.B., and Zilles, K. (1999). Broca's region revisited: cytoarchitecture and intersubject variability. *Journal of Computational Neuroscience*, 412, 319-341.

Aten, J.L., Johns, D.F., and Darley, F.L (1971). Auditory perception of sequenced words in apraxia of speech. *Journal of Speech and Hearing Research*, 14, 131-143.

Augustine, J.R. (1996). Circuitry and functional aspects of the insular lobe in primates including humans. *Brain Research Reviews*, 22, 229-244.

Averbeck, B.B., 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.

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 cell studies. *Experimental Brain Research*, 150, 127-141.

Baayen, R.H., Piepenbrock, R., and Gulikers, L. (1995). *The CELEX Lexical Database (Release 2)*. Philadelphia, PA: University of Pennsylvania, Linguistic Data Consortium.

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

Ballard, K.J. and Robin, D.A. (2007). Influence of continual biofeedback on jaw pursuit tracking in healthy adults and adults with apraxia plus aphasia. *Journal of Motor Behavior,* 39, 19-28.

Ballard, K.J., Robin, D.A., and Folkins, J.W. (2003). An integrative model of speech motor control: a response to Ziegler. *Aphasiology,* 17(1), 37-48.

Bamiou, D.E., Musiek, F.E., and Luxon, L.M. (2003). The insula (island of Reil) and its role in auditory processing literature review. *Brain Research Reviews,* 42(2), 145-154.

Bartle, C.J., Goozée, J.V., and Murdoch, B.E. (2007a). An EMA analysis of the effect of increasing word length on consonant production in apraxia of speech: A case study. *Clinical Linguistics and Phonetics,* 21(3), 189-210.

Bartle, C.J., Goozée, J.V., and Murdoch, B.E. (2007b). Preliminary evidence of silent articulatory attempts and starters in acquired apraxia of speech: A case study. *Journal of Medical Speech-Language Pathology,* 15(3), 207-223.

Baum, S.R. (1999). Compensation for jaw fixation by aphasic patients under conditions of increased articulatory demands: a follow-up study. *Aphasiology,* 13(7), 513-527.

Baum, S.R. and Pell, M.D. (1999). The neural bases of prosody: Insights from lesion studies and neuroimaging. *Aphasiology,* 13(8), 581-608.

Ben-Shachar, M., Hendler, T., Kahn, I., Ben-Bashat, D. and Grodzinsky, Y. (2003). The neural reality of syntactic transformations: evidence from functional magnetic resonance imaging. *Psychological Science,* 14, 433-440.

Bendor, D. and Wang, X. (2005). The neuronal representation of pitch in primate auditory cortex. *Nature,* 436, 1161-1165.

Bendor, D. and Wang, X. (2006). Neural representations of pitch in auditory cortex of humans and other primates. *Current Opinions in Neurobiology,* 16, 391-399.

Bilecen, D., Scheffler, K., Schmid, N., Tschopp, K., and Seelig, J. (1998). Tonotopic organization of the human auditory cortex as detected by BOLD-FMRI. *Hearing Research,* 126(1-2), 19-27.

Bladon, R.A., Henton, C.G., and Pickering, J.B. (1984). Towards an auditory theory of speech normalization. *Language and Communication*, 4(1), 59-69.

Blake, M., Duffy, J., Boeve, B.F., Ahlskog, E.J., and Maraganore, D.M. (2003). Speech and language disorders associated with corticobasal degeneration. *Journal of Medical Speech-Language Pathology*, 11(3), 131-146.

Blank, S.C., Scott, S.K., Murphy, K., Warburton, E., and Wise, R.J.S. (2002). Speech production: Wernicke, Broca, and beyond. *Brain*, 125(8), 1829-1838.

Blumstein, S.E., Cooper, W.E., Zurif, E.B., and Caramazza, A. (1977). The perception and production of voice onset time in aphasia. *Neuropsychologia*, 15, 371-383.

Boardman I. and Bullock, D. (1991). A neural network model of serial order recall from short-term memory. *Proceedings of the International Joint Conference on Neural Networks*, Seattle, WA, 879-884.

Boardman, I., Grossberg, S., Myers, C., and Cohen, M. (1999). Neural dynamics of perceptual order and context effects for variable-rate speech syllables. Perception and Psychophysics, 6, 1477-1500.

Bohland, J., Bullock, D., and Guenther, F. (in press). Neural representations and mechanisms for the performance of simple speech sequences. *Journal of Cognitive Neuroscience*.

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

Bonilha, L., Moser, D., Rorden, C., Baylis, G.C., and Fridriksson, J. (2006). Speech apraxia without oral apraxia: Can normal brain function explain the physiopathology? *Neuroreport*, 17(10), 1027-1031.

Bookheimer, S.Y. (2002). Functional MRI of language: New approaches to understanding the cortical organization of semantic processing. *Annual Review of Neuroscience*, 25, 151-188.

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

Bowers, J.S. (2002). Challenging the widespread assumption that connectionism and distributed representations go hand-in-hand. *Cognitive Psychology*, 45, 413-445.

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.

Bradski, G. and Grossberg, S. (1995). Fast learning VIEWNET architectures for recognizing 3-D objects from multiple 2-D views. *Neural Networks*, 8, 1053-1080.

Bregman, A.S. (1990). *Auditory Scene Analysis*. Cambridge, MA: MIT Press.

Broca, P.P. (1861). Perte de la parole, ramollissement chronique et destruction partielle du lobe antérieur gauch do cerveau. *Bulletin de la Société Anthropologique*, 2, 235-238.

Broussolle, E., Bakchine, S., Tommasi, M., Laurent, B., Bazin, B., Cinotti, L., Cohen, L., and Chazot, G. (1996). Slowly progressive anarthia with late anterior opercular syndrome: a variant form of frontal cortical atrophy syndromes. *Journal of the Neurological Sciences*, 144, 44-58.

Buckingham, H.W. (1979). Explanation in apraxia with consequences for the concept of apraxia of speech. *Brain and Language*, 8, 202–226.

Buckner, R.L., Petersen, S.E., Ojemann, J.G., Myezin, F.M., Squire, L.R., and Raichle, M.E. (1995). Functional anatomical studies of explicit and implicit memory retrieval tasks. *Journal of Neuroscience*, 15, 12-29.

Bullock, D., Grossberg, S., and Guenther, F.H. (1993). A self-organizing neural model of motor equivalent reaching and tool use by a multijoint arm. *Journal of Cognitive Neuroscience*, 5, 408-435.

Bullock, D., Grossberg, S., and Mannes, C. (1993). A neural network model for cursive script production. *Biological Cybernetics*, 70, 15-28.

Bullock, D. and Rhodes, B.J. (2003). Adaptive force generation for precision-grip lifting by a spectral timing model of the cerebellum. *Neural Networks*, 16(5-6), 521-528.

Burton, M.W. (2001). The role of inferior frontal cortex in phonological processing. *Cognitive Science*, 25, 695-709.

Burton, M.W., Small, S., and Blumstein, S.E. (2000). The role of segmentation in phonological processing : an fMRI investigation. *Journal of Cognitive Neuroscience*, 12, 679-690.

Carpenter, G.A. (1997). Distributed learning, recognition, and prediction by ART and ARTMAP neural networks. *Neural Networks*, 10, 1473-1494.

Carpenter, G.A. and Grossberg, S. (1987). A massively parallel architecture for a self-organizing neural pattern recognition machine. *Computer Vision, Graphics, and Image Processing*, 37, 54-115.

Carpenter, G.A. and Grossberg, S. (1991). *Pattern recognition by self-organizing neural networks.* Cambridge, MA: MIT Press.

Carpenter, G.A. and Grossberg, S. (1993). Normal and amnesic learning, recognition, and memory by a neural model of cortico-hippocampal interactions. *Trends in Neuroscience*, 16, 131-137.

Carpenter, G.A. and Grossberg, S. (2003). Adaptive resonance theory. In Arbib M.A. (ed) *The Handbook of Brain Theory and Neural Networks* (second edition). Cambridge, MA: MIT Press, pp. 87-90.

Carpenter, G.A., Grossberg, S., Markuzon, N., Reynolds, J.H., and Rosen, D.B. (1992). Fuzzy ARTMAP: A neural network architecture for incremental supervised learning of analog multi-dimensional maps. *IEEE Transactions on Neural Networks*, 3, 698-713.

Chapin, C., Blumstein, S.E., Meissner, B, and Boller, F. (1981). Speech production mechanisms in aphasia: A delayed auditory feedback study. *Brain and Language*, 14, 106-113.

Chein, J.M. and Fiez, J.A. (2001). Dissociation of verbal working memory system components using a delayed serial recall task. *Cerebral Cortex*, 11, 1003-1014.

Chein, J.M., Fissell, K., Jacobs, S., and Fiez, J.A. (2002). Functional heterogeneity within Broca's area during verbal working memory. *Physiology and Behavior*, 77, 635-639.

Chey, J., Grossberg, S., and Mingolla, M. (1997). Neural dynamics of motion grouping: From aperture ambiguity to object speed and direction. *Journal of the Optical Society of America*, 14, 2570-2594.

Cholin, J., Levelt, W.J.M., and Schiller, N.O. (2006). Effects of syllable frequency in speech production. *Cognition*, 99, 205–235.

Church, B.A. and Schacter, D.L. (1994). Perceptual specificity of auditory priming: implicit memory for voice intonation and fundamental frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(3), 521-533.

Cohen, M.A. and Grossberg, S. (1986). Neural dynamics of speech and language coding: Developmental programs, perceptual groping, and competition for short-term memory. *Human Neurobiology*, 5, 1-22.

Cohen, M.A. and Grossberg, S. (1987). Masking fields: A massively parallel neural architecture for learning, recognizing, and predicting multiple groupings of patterned data. *Applied Optics*, 26, 1866-1891.

Cohen, M.A., and Grossberg, S. (1997). Parallel auditory filtering by sustained and transient channels separates coarticulated vowels and consonants. *IEEE Transactions on Speech and Audio Processing*, 5, 301-318.

Cohen, M.A., Grossberg, S. and Stork, D.G. (1988). Speech perception and production by a self-organizing neural network. In Lee YC (ed) *Evolution, Learning, Cognition, and Advanced Architectures*. Hong Kong: World Scientific, 217-231.

Cohen MA, Grossberg S and Wyse L (1995). A spectral network model of pitch perception. *Journal of the Acoustical Society of America*, 98, 862-879.

Collins, M., Rosenbek, J.C., and Wertz, R.T. (1983). Spectrographic analysis of vowel and word duration in apraxia of speech. *Journal of Speech and Hearing Research*, 26, 224-230.

Creelman, C.D. (1957). Case of the unknown talker. *Journal of the Acoustical Society of America*, 29, 655.

Dabul, B.L. (2000). *Apraxia Battery for Adults - 2*. Austin, TX: Pro-Ed.

Damasio, H. and Damasio, A.R. (1980). Dichotic listening pattern in conduction aphasia. *Brain and Language*, 10(2), 281-286.

Darley, F.L. (1968). Apraxia of speech: 107 years of terminological confusion. Paper presented to the *American Speech and Hearing Association*, Denver, CO.

Darley, F.L., Aronson, A.E., and Brown, J.R. (1975). *Motor Speech Disorders*. Philadelphia: W.B. Saunders, Inc.

Deal, J.L. (1974). Consistency and adaptation in apraxia of speech. *Journal of Communication Disorders*, 7, 135-140.

Deal, J. and Darley, F.L. (1972). The influence of linguistic and situational variables on phonemic accuracy in apraxia of speech. *Journal of Speech and Hearing Research*, 15, 639-653.

Deger, K. and Ziegler, W. (2002). Speech motor programming in apraxia of speech. *Journal of Phonetics*, 30, 321-335.

Dejerine, J. (1913). Motor aphasia, anarthia, and apraxia. *Transactions of International Congress of Medicine,* 85-106.

Dell, G.S., Schwartz, M.F., Martin, N., Saffran, E.M., and Gagnan, D.A. (1997). Lexical access in aphasic and nonaphasic speakers. *Psychological Review,* 104(4), 801-838.

Demb, J.B., Desmond, J.E., Wagner, A.D., Vaidya, C.J., Glover, G.H., and Gabrieli, J.D.E. (1995). Semantic encoding and retrieval in the left inferior prefrontal cortex: A functional MRI study of task difficulty and process specificity. *Journal of Neuroscience*, 15, 5870-5878.

Demonet, J-F., Chollet, F., Ramsay, S., Cardebat, D., Nespoulous, J-L., Wise, R., Rascol, A., Frackowiak, R. (1992). The anatomy of phonological and semantic processing in normal subjects. *Brain*, 115, 1753-1768.

Demonet, J.F., Price, C., Wise, R., and Frackowiak, R.S.J. (1994). A PET study of cognitive strategies in normal subjects during language tasks: Influence of phonetic ambiguity and sequence processing on phoneme monitoring. *Brain*, 117, 671-682.

Derrfuss, J., Brass, M., and von Cramon, D.Y. (2004). Cognitive control in the posterior frontolateral cortex: evidence from common activations in task coordination, interference control, and working memory. *NeuroImage*, 23, 604-612.

Desimone, R. (1998). Visual attention mediated by biased competition in extrastriate visual cortex. *Philosophical Transactions of the Royal Society*, 353, 1245-1255.

Deutsch, S.E. (1981). Oral form identification as a measure of cortical sensory dysfunction in apraxia of speech and aphasia. *Journal of Communication Disorders*, 14, 65-73.

Dogil, G., Ackermann, H., Grodd, W., Haider, H., Kamp, H., Mayer, J., Riecker, A., and Wildgruber, D. (2001). The speaking brain: a tutorial introduction to fMRI experiments in the production of speech, prosody and syntax. *Journal of Neurolinguistics*, 15, 59-90.

Dognin, P.L. and El-Jaroudi, A. (2003). A new spectral transformation for speaker normalization. *Proceedings of EuroSpeech*, 1865-1868.

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

Dronkers, N.F., Pinker, S., and Damasio, A. (2000). Chapter 59: Language and the aphasias. In Kandel E.R., Schwartz J.H., and Jessell T.M. (eds) *Principles of Neural Science* (fourth edition).New York: McGraw Hill, pp. 1169-1186.

Dronkers, N.F., Wilkins, D.P., Van Valin Jr., R.D., Redfern, B.B., and Jaeger, J.J. (2004). Lesion analysis of the brain areas involved in language comprehension. *Cognition*, 92, 145-177.

Duffau, H., Bauchet, L., Lehericy, S., and Capelle, L. (2001). Functional compensation of the left dominant insula for language. *Neuroreport*, 12, 2159-2163.

Duffau, H., Capelle, L., Denvil, D., Gatignol, P., Sichez, N., Lopez, M., Sichez, J.P., and Effenterre, R.V. (2003). The role of dominant premotor cortex in language: a study using intraoperative functional mapping in awake patients. *NeuroImage*, 20, 1903-1914.

Duffy, J.R. (1995). *Motor Speech Disorders: Substrates, Differential Diagnosis, and Management*. St Louis, MO: Mosby-Year Book, Inc.

Duffy, J.R. (2005). *Motor Speech Disorders: Substrates, Differential Diagnosis, and Management* (second edition). St. Louis, MO: Mosby-Year Book, Inc.

Dusan, S. and Rabiner, L.R. (2005). Can automatic speech recognition learn more from human speech perception? *Proceedings of the Third Conference of Speech Technology and Human-Computer Dialogue*, 21-36.

Edmonds, L.A. and Marquardt, T.P. (2004). Syllable use in apraxia of speech: Preliminary findings. *Aphasiology*, 18(12), 1121-1134.

Eide E and Gish H (1996). A parametric approach to vocal tract length normalization. *Proceedings of the International Congress on Audition, Speech, and Signal Processing*, 1, 346-348.

Eklund, I. and Traunmüller, H. (1997). Comparative study of male and female whispered and phonated versions of the long vowels of Swedish. *Phonetica*, 54, 1-21.

Epstein, C.M., Lan, J.J., Meador, K., Weissman, J.D., Gaitan, L.E., and Dihenia, B. (1996). Optimum stimulus parameters for lateralized suppression of speech with magnetic brain stimulation. *Neurology*, 47, 1590-1593.

Epstein, C.M., Meador, K.J., Loring, D.W., Wright, R.J., Weissman, J.D., Sheppard, S., Lah, J.J., Puhalovich, F., Gaitan, L., and Davey, K.R. (1999). Localization and characterization of speech arrest during transcranial magnetic stimulation. *Clinical Neurophysiology*, 110(6), 1073-1079.

Fant, G. (1973). Stops in CV syllables. In Fant G. (ed) *Speech Sounds and Features*.. Cambridge, MA: MIT Press, pp. 110-139.

Fazl, A., Grossberg, S. and Mingolla, E. (2009). View-invariant object category learning, recognition, and search: How spatial and object attention are coordinated using surface-based attentional shrouds. *Cognitive Psychology*, 58, 1-48.

Ferreira, A.J.S. (2007). Static features in real-time recognition of isolated vowels at high pitch. *Journal of the Acoustical Society of America*, 122(4), 2389-2404.

Fiez, J.A. (1997). Phonology, semantics, and the role of the left inferior prefrontal cortex. *Human Brain Mapping*, 5, 79-83.

Fiez, J.A. and Petersen, S.E. (1998). Neuroimaging studies of word reading. *Proceeding of the National Academy of Science, USA*, 95, 914-921.

Fiez, J.A., Raichle, M.E., Miezin, F.M., Petersen, S.E., Tallal, P., and Katz, W.F. (1995). PET studies of auditory and phonological processing: Effects of stimulus characteristics and task design. *Journal of Cognitive Neuroscience*, 7, 357-375.

Fink, J.N., Selim, M.H., Kumar, S., Voetsch, B., Fang, W.C., and Caplan, L.R. (2005). Insular cortex infarction in acute middle cerebral artery territory stroke. *Archives of Neurology*, 62, 1081-1085.

Fishman, Y.I., Reser, D.H., Arezzo, J.C., and Steinschneider, M. (1998). Pitch vs. spectral encoding of harmonic complex tones in primary auditory cortex of the awake monkey. *Brain Research*, 786(1-2), 18-30.

Flitman, S.S., Grafman, J., Wassermann, E.M., Cooper, V., O'Grady, J., Pascual-Leone, A., and Hallett, M. (1998). Linguistic processing during repetitive transcranial magnetic stimulation. *Neurology*, 50, 175-181.

Formisano, E., Kim, D.-S., Di Salle, F., van de Moortele, P.F., Ugurbil, K., and Goebel, R. (2003). Mirror-symmetric tonotopic maps in human primary auditory cortex. *Neuron*, 40(4), 859-869.

Foundas, A.L., Eure, K.F., Luevano, L.F., and Weinberger, D.R. (1998). MRI asymmetries of Broca's area: the pars triangularis and pars opercularis. *Brain Language*, 64, 282-296.

Fox, P.T., Ingham, R.J., Ingham, J.C., Zamarripa, F., Xiong, J.H., Lancaster, J.L. (2000). Brain correlates of stuttering and syllable production. A PET performance-correlation analysis. *Brain*, 123, 1985-2004.

Friederici, A.D., Fiebach, C.J., Schlesewsky, M., Bornkessel, Ina D., and von Cramon, D.Y. (2006). Processing linguistic complexity and grammaticality in the left frontal cortex. *Cerebral Cortex*, 16, 1709-1717.

Friederici, A.D. and Kotz, S.A. (2003). The brain basis of syntactic processes: functional imaging and lesion studies. *NeuroImage*, 20, *Suppl* 1, S8-S17.

Friederici, A.D., Ruschemeyer, S.A., Hahne, A., and Fiebach, C.J. (2003). The role of left inferior frontal and superior temporal cortex in sentence comprehension: localizing syntactic and semantic processes. *Cerebral Cortex*, 13, 170-177.

Fromm, D., Abbs, J.H., McNeil, M.R., and Rosenbek, J.C. (1980). Simultaneous perceptual and physiological method for studying apraxia of speech. *Clinical Aphasiology*, 10, 251-262.

Fudge, E.C. (1969). Syllables. *Journal of Linguistics*, 5, 226-320.

Gabrieli, J.D.E., Desmond, J.E., Demb, J.B., and Wagner, A.D. (1996). Functional magnetic resonance imaging of semantic memory processes in the frontal lobes. *Psychological Science*, 7(5), 278-283.

Galaburda, A.M. and Pandya, D.N. (1982). Role of architectonics and connections in the study of primate brain evolution. In Amsterdam E. and Falks O. (eds) *Primate brain evolution: methods and concepts*. New York, NY: Plenum Publishing, pp. 203-216.

Gelfand, J.R. and Bookheimer, S.Y. (2003). Dissociating neural mechanisms of temporal sequencing and processing phonemes. *Neuron*, 38(5), 831-842.

Geschwind, N. (1967). Neurological foundations of language. In Myklebust I.H.R. (ed) *Progress in Learning Disabilities* (volume 1). New York, NY: Grune and Stratton.

Gitelman, D.R., Nobre, A.C., Sonty, S. Parrish, T.B., and Mesulam, M.M. (2005). Language network specializations: An analysis with parallel task designs and functional magnetic resonance imaging. *NeuroImage*, 26, 975-985.

Glasberg, B.R. and Moore, B.C. (1990). Derivation of auditory filter shapes from notched-noise data. *Hearing Research*, 47(1-2), 103-138.

Glavitsch, U. (2003). Speaker normalization with respect to F0: A perceptual approach. *TIK Report No 185*, Swiss Federal Institute of Technology Zurich.

Goldinger, S.D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 1166-1183.

Goldinger, S.D. (1997). Words and voices: Perception and production in an episodic lexicon. In Johnson K. and Mullennix J.W. (eds) *Talker Variability in Speech Processing*. San Diego: Academic Press, pp. 33-66.

Goldinger, S.D. and Azuma, T. (2003). Puzzle-solving science: The quixotic quest for units in speech perception. *Journal of Phonetics*, 31, 305-320.

Golestani, N. and Zatorre, R.J. (2004). Learning new sounds of speech: reallocation of neural substrates. *NeuroImage*. 21, 494-506.

Goodale, M.A. and Milner, D. (1992). Separate visual pathways for perception and action. *Trends in Neuroscience*, 15, 10-25.

Gorno-Tempini, M.L., Rankin, K.P., Woolley, J.D., Rosen, H.J., Phengrasamy, L., and Miller, B.L. (2004a). Cognitive and behavioral profile in a case of right anterior temporal lobe neurodegeneration. *Cortex*, 40(4-5), 631-644.

Gorno-Tempini, M.L., Murray, R.C., Rankin, K.P., Weiner, M.W., and Miller, B.L. (2004b). Clinical, cognitive and anatomical evolution from nonfluent progressive aphasia to corticobasal syndrome: A case report. *Neurocase*, 10(6), 426-436.

Gough, P.M., Nobre, A., and Devlin, J.T. (2005). Dissociating linguistic processes in the left inferior frontal cortex with transcranial magnetic stimulation. *The Journal of Neuroscience*, 25(35), 8010-8016.

Grossberg, S. (1973). Contour enhancement, short-term memory, and constancies in reverberating neural networks. *Studies in Applied Mathematics*, 52, 213-257.

Grossberg, S. (1976a). Adaptive pattern classification and universal recoding, I: Parallel development and coding of neural feature detectors. *Biological Cybernetics*, 23, 121-134.

Grossberg, S. (1976b). Adaptive pattern classification and universal recoding, II: Feedback, expectation, olfaction, illusions. *Biological Cybernetics*, 23, 187-202.

Grossberg, S. (1978). A theory of human memory: Self-organization and performance of sensory-motor codes, maps, and plans. In Rosen R. and Snell F. (eds) *Progress in theoretical biology* (volume 5)**.** New York, NY: Academic Press, pp. 233-374.

Grossberg, S. (1980). How does a brain build a cognitive code? *Psychological Review*, 87, 1-51.

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. (eds) *Pattern Recognition by humans and machines, V.1: Speech Perception*. New York, NY: Academic Press, pp. 187-294.

Grossberg, S. (1994). 3-D vision and figure ground separation by visual cortex. *Perceptual Psychophysics*, 55(1), 48-120.

Grossberg, S. (1999). The link between brain learning, attention, and consciousness. *Consciousness and Cognition*, 8, 1-44.

Grossberg, S. (2000). The complementary brain: Unifying brain dynamics and modularity. *Trends in Cognitive Science*, 4**,** 233-246.

Grossberg, S. (2003a). How does the cerebral cortex work? Development, learning, attention, and 3D vision by laminar circuits of visual cortex. *Behavioral and Cognitive Neuroscience Reviews*, 2, 47-76.

Grossberg, S. (2003b). Resonant neural dynamics of speech perception. *Journal of Phonetics*, 31, 423-445.

Grossberg, S., Boardman, I., and Cohen, M. (1997). Neural dynamics of variable-rate speech categorization. *Journal of Experimental Psychology: Human Perception and Performance*, 23, 418-503.

Grossberg, S., Govindarajan, K.K., Wyse, L., and Cohen, M.A. (2004). ARTSTREAM: A neural network model of auditory scene analysis and source segregation. *Neural Networks*, 17, 511-536.

Grossberg, S. and Merrill, J.W.L. (1996). The hippocampus and cerebellum in adaptively timed learning, recognition, and movement. *Journal of Cognitive Neuroscience*, 8, 257-277.

Grossberg, S. and Myers, C.W. (2000). The resonant dynamics of speech perception: Interword integration and duration-dependent backward effects. *Psychological Review*, 107, 735-767.

Grossberg, S. and Repin, D. (2003). A neural model of how the brain represents and compares multi-digit numbers: spatial and categorical processes. *Neural Networks*, 16, 1107-1140.

Grossberg, S. and Stone, G.O. (1986). Neural dynamics of attention switching and temporal order information in short-term memory. *Memory and Cognition*, 14, 451-468.

Grossberg, S. and Versace, M. (2008). Spikes, synchrony, and attentive learning by laminar thalamocortical circuits." *Brain Research*, 1218C, 278-312..

Grossberg, S. and Williamson, J.R. (1999). A self-organizing neural system for learning to recognize textured scenes. *Vision Research*, 39, 1385-1406.

Grunewald, A. and Grossberg, S. (1998). Self-organization of binocular disparity tuning by reciprocal corticogeniculate interactions. *Journal of Cognitive Neuroscience*, 10, 199-215.

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, 594-621.

Guenther, F.H. (2001). Neural modeling of speech production. *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, 350-365.

Guenther, F.H. and Ghosh, S.S. (2003). A model of cortical and cerebellar function in speech. *Proceedings of the XVth International Congress of Phonetic Sciences*, Barcelona.

Guenther, F.H., Ghosh, S.S., and Nieto-Castanon, Alfonso. (2003). A neural model of speech production. *Proceedings of the 6th International Seminar on Speech Production*, Sydney.

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, 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*, 102, 594-621.

Habib, M., Daquin, G., Milandre, L., Royere, M.L., Rey, M., Lanteri, A., Slamanon, G., Khalil, R. (1995). Mutism and auditory agnosia due to bilateral insular damage-Role of the insula in human communication. *Neuropsychologia*, 33(3), 327-339.

Hageman, C.F., Robin, D.A., Moon, J.B, and Folkins, J.W. (1994). Oral motor tracking in normal and apraxic speakers. *Clinical Aphasiology*, 22, 219-229.

Haley, K.L. (2002). Temporal and spectral properties of voiceless fricatives in aphasia and apraxia of speech. *Aphasiology*, 16(4/5/6), 595-607.

Haley, K.L. (2004). Vowel duration as a cue to postvocalic stop voicing in aphasia and apraxia of speech. *Aphasiology*, 18(5/6), 443-456.

Haley, K.L. and Overton, H.B. (2001). Word length and vowel duration in apraxia of speech: The use of relative measures. *Brain and Language*, 79, 397-406.

Hartley, T. and Houghton, G. (1996). A linguistically constrained model of short-term memory for nonwords. *Journal of Memory and Language*, 35, 1-31.

Hawkins, S. (2003). Roles and representations of systematic fine phonetic detail in speech understanding. *Journal of Phonetics*, 31, 373-405.

Heil, P., Rajan, J., and Irvine, D.R. (1994). Topographic representation of tone intensity along the isofrequency axis of cat primary auditory cortex. *Hearing Research*, 76(1-2), 188-202.

Heim, S., Opitz, B., and Friederici, A.D. (2003). Distributed cortical networks for syntax processing: Broca's area as the common denominator. *Brain Language*, 85, 402-408.

Heiser , M., Iacoboni, M., Maeda, F., Marcus, J., and Mazziotta, J.C. (2003). The essential role of Broca's area in imitation. *European Journal of Neuroscience*, 17, 1123-1128.

Hess, W. (1983). *Pitch Determination of Speech Signals-Algorithms and Devices*. Berlin: Springer.

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

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

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

Hillenbrand, J.M. and Gayvert, R.T. (1993). Identification of steady-state vowels synthesized from the Peterson and Barney measurements. *Journal of the Acoustical Society of America*, 94, 668-674.

Hillenbrand, J.M. and Nearey, T.M. (1999). Identification of resynthesized /hVd/ utterances: effects of formant contour. *Journal of the Acoustical Society of America*, 105(6), 3509-3523.

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

Hinke, R.M., Hu, X., Stillman, A.E., Kim, S.G., Merkle, H., Salmi, R., et al (1993). Functional magnetic resonance imaging of Broca's area during internal speech. *NeuroReport*, 4, 675-678.

Hodgkin, A.L. and Huxley, A.F. (1952). Currents carried by sodium and potassium ions through the membrane of the giant squid axon of Loligo. *Journal of Physiology*, 116(4), 449-472.

Holdsworth, J., Nimmo-Smith, I., Patterson, R., and Rice, P. (1988). Implementing a gammatone filterbank. *Annex C of the SVOS Final Report: Part A: The Auditory Filterbank.*

Hough, M.S. and Klich, R.J. (1998). Lip EMG activity during vowel production in apraxia of speech: Phrase context and word length. *Journal of Speech, Language, and Hearing Research*, 41, 786-801.

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. (eds) *Current Research in Natural Language Generation*. San Diego, CA: Academic Press, pp. 287-319.

Hubel, D. H., and Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiol*ogy, 160, 106-154.

Hudspeth, A.J. (2000). Chapter 30: Hearing. In Kandel E.R., Schwartz J.H., and Jessell T.M. (eds) *Principles of Neural Science* (fourth edition).New York: McGraw Hill, pp. 590-613.

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

Imig, T.J., Ruggero, M.A., Kitzes, L.M., Javel, E., and Brugge, J.F. (1977). Organization of auditory cortex in the owl monkey. *Journal of Comparative Neurology*, 171(1), 111-128.

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

Indefrey, P., Brown, C.M., Hellwig, F., Amunts, K., Herzog, H., Seitz, R.J., and Hagoort, P. (2001). A neural correlate of syntactic encoding during speech production. *Proceedings of the National Academy of Science*, 98(10), 5933-5936.

Itoh, M., Sasanuma, S., Hirose, H., Yoshioka, H., and Ushijima, T. (1980). Abnormal articulatory dynamics in a patient with apraxia of speech: X-ray microbeam observation. *Brain and Language*, 11, 66-75.

Itoh, M., Sasanuma, S., Tatsumi, I.F., Murakami, S., Fukusako, Y., and Suzuki, T. (1982). Voice onset time characteristics in apraxia of speech. *Brain and Language*, 17, 193-210.

Itoh, M., Sasanuma, S., and Ushijima, T. (1979). Velar movements during speech in a patient with apraxia of speech. *Brain and Language*, 7, 227-239.

Itoh, M., Tamura, H., Fujita, I. and Tanaka, K. (1995). Size and position invariance of neuronal responses in monkey inferotemporal cortex. *Journal of Neurophysiology*, 73, 218-226.

Jacks, A.P. (2006). *Vowel Targeting and Perception in Apraxia of Speech*. Dissertation, University of Texas, Austin.

Jacks, A. (2008). Bite block vowel production in apraxia of speech. *Journal of Speech, Language, and Hearing Research*, 51, 898-913.

Joanisse, M.F. and Gati, J.S. (2003). Overlapping neural regions for processing rapid temporal cues in speech and nonspeech signals. *NeuroImage*, 19, 64-79.

Johns, D.F. and Darley, F.L (1970). Phonemic variability in apraxia of speech. *Journal of Speech and Hearing Research*, 13, 556-583.

Johnson, K. (1990). The role of perceived speaker identity in F0 normalization of vowels. *Journal of the Acoustical Society of America*, 88, 642-654.

Johnson, K. (1997a). Speech perception without speaker normalization: an exemplar model. In Johnson K. and Mullennix J.W. (eds) *Talker Variability in Speech Processing*. San Diego: Academic Press, pp. 145-166.

Johnson, K. (1997b). The auditory/perceptual basis for speech segmentation. *Ohio State University Working Papers in Linguistics*, 50, 101-113.

Johnson, K. (2005). Speaker normalization in speech perception. In Pisoni D.B. and Remez R. (eds) *The Handbook of Speech Perception*. Oxford: Blackwell Publishers, pp. 363-389.

Johnson, K. (2006). Resonance in an exemplar-based lexicon: The emergence of social identity and phonology. *Journal of Phonetics,* 34, 485-499.

Johnson, K., Strand, E.A., and D'Imperio, M. (1999). Auditory-visual integration of talker gender in vowel perception. *Journal of Phonetics*, 27, 359-384.

Johnsrude, I.S., Zatorre, R.J., Milner, B.A., and Evans, A.C. (1997). Left-hemisphere specialization for the processing of acoustic transients. *NeuroReport*, 8, 1761-1765.

Josephs, K.A., Duffy, J.R., Strand, E.A., Whitwell, J.L., Layton, K.F., Parisi, J.E., Hauser, M.F., Witte, R.J., Boeve, B.F., Knopman, D.S., Dickson, D.W., Jack Jr, C.R., and Petersen, R.C. (2006). Clinicopathological and imaging correlates of progressive aphasia and apraxia of speech. *Brain*, 129(6), 1385-1398.

Jürgens, U. (1987). The efferent and afferent connections of the supplementary motor area. *Brain Research*, 300, 63-81.

Kaas, J.H. and Hackett, T.A. (1998). Subdivisions of auditory cortex and levels of processing in primates. *Audiology and Neurotology*, 3(2-3), 73-85.

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

Kapur, S., Rose, R., Liddle, P.F., Zipursky, R.B., Brown, G.M., Stuss, D., Houle, S., and Tulving, E. (1994). The role of the left prefrontal cortex in verbal processing: semantic processing or willed action? *NeuroReport*, 5, 2193-2196.

Kastner, S. and Ungerleider, L. G. (2001). The neural basis of biased competition in human visual cortex. *Neuropsychologia,* 39, 1263-1276.

Kato, K. and Kakehi, K. (1988). Listener adaptability to individual speaker differences in monosyllabic speech perception. *Journal of the Acoustical Society of Japan,* 44, 180-186.

Keller, S.S., Highly, J.R., Garcia-Finana, M., Sluming, V., Rezaie, R., and Roberts, N. (2007). Sulcul variability, stereological measurement and asymmetry of Broca's area of MR images. *Journal of Anatomy,* 211, 534-555.

Kent, R.D. (2000). Research on speech motor control and its disorders: a review and prospective. *Journal of Communication Disorders*, 33, 391-428.

Kent, R.D. (2004). Models of speech motor control: Implications from recent developments in neurophysiological and neurobehavioral science. In Maassen B., Kent R., and Peters H.F.M. (eds) *Speech Motor Control*, Oxford: University Press, pp. 3-28.

Kent, R.D. and Read, C. (1992). *Acoustic Analysis of Speech*. Madison, WI: Singular.

Kent, R.D. and Rosenbek, J.C. (1982). Prosodic disturbance and neurologic lesion. *Brain and Language*, 15, 259-291.

Kent, R.D. and Rosenbek, J.C. (1983). Acoustic patterns of apraxia of speech. *Journal of Speech and Hearing Research*, 26, 231-249.

Kertesz, A. (1984). Subcortical lesions and verbal apraxia. In Rosenbeck J.C., McNeil M.R., and Aronson A. (eds), *Apraxia of speech: Physiology, acoustics, linguistics, management*. London: College Hill Press, pp. 73-90.

Kimura, D. (1993). *Neuromotor Mechanisms in Human Communication*. New York, NY: Oxford University Press.

Kimura, D. and Watson, N. (1989). The relation between oral movement control and speech. *Brain and Language*, 37, 565-590.

Kirkman, T.W. (1996). *Statistics to Use*. http://www.physics.csbsju.edu/stats (1 Oct 2007).

Klein, D., Milner, B., Zatorre, R.J., Meyer, E., and Evans, A.C. (1995). The neural substrates underlying word generation: A bilingual functional-imaging study. *Proceedings of the National Academy of Science USA*, 92(7), 2899-2903.

Koski, L., Wohlschlager A., Bekkering, H., Woods, R.P. Dubeau, M.C., Mazziotta, J.C. and Iacoboni, M. (2002). Modulation of motor and premotor activity during imitation of target-directed actions. *Cerebral Cortex*. 12, 847-855.

Kraljic, T. and Samuel, A.G. (2007). Perceptual adjustments to multiple speakers. *Journal of Memory and Language*, 56, 1-15.

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.

Langner, G., Sams, M., Heil, P., and Schulze, H. (1997). Frequency and periodicity are represented in orthogonal maps in the human auditory cortex: Evidence from magnetoencephalography. *Journal of Comparative Physiology*, 181(6), 665-676.

LaPointe, L.L. and Johns, D.F. (1975). Some phonemic characteristics in apraxia of speech. *Journal of Communication Disorders*, 8, 259-269.

LaPointe, L.L. and Horner, J. (1976). Repeated trials of words by patients with neurogenic phonological selection-sequencing impairment (apraxia of speech): Stimulus mode and response condition revisited. *Clinical Aphasiology Conference Proceedings*, 261-277.

Lee, L. and Rose, R. (1996). Speaker normalization using efficient frequency warping procedures. *Proceedings of the International Conference on Audition, Speech, and Signal Processing*, 1, 353-356.

Lee, L. and Rose, R. (1998). A frequency warping approach to speaker normalization. *IEEE Transactions on Speech and Audio Processing*, 6(1), 49-60.

Lehiste, I. and Meltzer, D. (1973). Vowel and speaker identification in natural and synthetic Speech. *Language and Speech*, 16, 356-264.

Levelt, W.J.M. (1999). Models of word production. *Trends in Cognitive Sciences*, 3(6), 223-232.

Levelt, W.J.M. (2001). Spoken word production: A theory of lexical access. *Proceedings of the National Academy of Science*, 98(23), 13464-13471.

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

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

Liepmann, H. (1900). Das Krankheitsbild der Apraxie ('motorischen Asymbolie') auf Grund eines Falles von einseitiger Apraxie. *Monatschrift fur Psychiatrie und Neurologie*, 8, 15-44, 102-132, 182-197.

Lindau, M. (1978). Vowel features. *Language*, 54(3), 541-563.

Lloyd, R.J. (1890a). *Some Researches into the Nature of Vowel-Sound*. Liverpool, England: Turner and Dunnett.

Lloyd, R.J. (1890b). Speech sounds: Their nature and causation (I). *Phonetische Studien,* 3, 251-278.

Lloyd, R.J. (1891). Speech sounds: Their nature and causation (II-IV). *Phonetische Studien,* 4, 37-67, 183-214, 275-306.

Lloyd, R.J. (1892). Speech sounds: Their nature and causation (V-VII). *Phonetische Studien,* 5, 1-32, 129-141, 263-271.

Lockwood, A.H., Salvi, R.J., Coad, M.L., Arnold, S.A., Wack, D.S., Murphy, B.W., and Burkard, R.F. (1999). The functional anatomy of the normal human auditory system: Responses to 0.5 and 4.0 kHz tones and varied intensities. *Cerebral Cortex,* 9(1), 65-76.

Lozano, R.A. and Dreyer, D.E. (1978). Some effects of delayed auditory feedback on dyspraxia of speech. *Journal of Communication Disorders,* 11, 407-415.

Luethke, L.E., Krubitzer, L.A., and Kaas, J.H. (1988). Cortical connections of electrohpysiologically and architectonically defined subdivisions of auditory cortex in squirrels. *Journal of Comparative Neuroscience,* 268(2), 181-203.

Luria, A.R. (1966). *Human Brain and Psychological Processes*. New York, NY: Harper.

Maas, E., Robin, D.A., Wright, D.L., and Ballard, K.J. (2008). Motor programming in apraxia of speech. *Brain and Language,* 106, 107-118.

Maassen, B. (2002). Issues contrasting adult acquired versus developmental apraxia of speech. *Seminars in Speech and Language*, 23(4), 257-266.

Magimai-Doss, M., Stephenson, T.A., Bourlard, H. (2003). Using pitch frequency information in speech recognition. *Proceedings of EuroSpeech*.

Marie, P. (1906). Revision de la question de l'aphasie: La troisiéme circonvolution frontale gauce ne joue aucun role special dans la function du language. *Semaine Medicale,* 26, 241-247.

Marien, P., Pickut, B.A., Engelborghs, S., Martin, J.J., and De Dyn, P.P. (2001). Phonological agraphia following a focal anterior insula-opercular infarction. *Neuropsychologia,* 39, 845-855.

Marquardt, T.P, Duffy, G., and Cannito, M.P. (1995). Acoustic analysis of accurate word stress patterning in patients with apraxia of speech and Broca's aphasia. *American Journal of Speech-Language Pathology*, 4, 180-185.

Marquardt, T., and Sussman, H. (1984). The elusive lesion-apraxia of speech link in Broca's aphasia. In Rosenbeck J.C., McNeil M.R., and Aronson A. (eds), *Apraxia of speech: Physiology, acoustics, linguistics, management*. London: College Hill Press, pp. 91-112.

Martin, A.D. (1974). Some objections to the term "apraxia of speech." *Journal of Speech and Hearing Disorders*, 39, 53-64.

Martin, A.D. (1975). Letter: reply to Aten, Darley, Deal, and Johns. *Journal of Speech and Hearing Disorders*, 40, 421-423.

Martin, A., Haxby, J.V., Lalonde, F.M. Wiggs, C.L., and Ungerleider, L.G. (1995). Discrete cortical regions associated with knowledge of color and knowledge of action. *Science*, 270, 102-105.

Martin, R.C. (2003). Language processing: Functional organization and neuroanatomical basis. *Annual Review of Neuroscience*, 54, 55-89.

Masaki, S., Tatsumi, I.F., and Sasanuma, S. (1991). Analysis of the temporal relationship between pitch control and articulatory movements in the realization of Japanese word accent by a patient with apraxia of speech. *Clinical Aphasiology*, 19, 307-316.

Mauszycki, S.C., Dromey, C., and Wambaugh, J.L. (2007). Variability in apraxia of speech: A perceptual, acoustic, and kinematic analysis of stop consonants. *Journal of Medical Speech-Language Pathology*, 15(3), 223-242.

Mazziotta, J., Toga, A., Evans, A., Fox, P., Lancaster, J., Zilles, K., Woods, R., Paus, T., Simpson, G., Pike, B., Holmes, C., Collins, L., Thompson, P., MacDonald, D., Iacoboni, M., Schormann, T., Amunts, K., Palomero-Gallagher, N., Geyer, S., Parson, L., Narr, K., Kabani, N., Le Goualher, G., Boomsma, D., Cannon, T., Kawashima, R., and Mazoyer, B. (2001). A probabilistic atlas and reference system for the human brain: International Consortium for Brain Mapping (ICBM). *Philosophical Transactions of the Royal Society of London B Biological Science*, 356, 1293-1322.

Mazzocchi, F. and Vignolo L.A. (1979). Localisation of lesions in aphasia: clinical-CAT scan correlations in stroke patients. *Cortex*, 15(4), 627-653.

McCune, L. and Vihman, M.M. (2001). Early phonetic and lexical development: A productivity approach. *Journal of Speech, Language, and Hearing Research*, 44, 670-684.

McDonough, J. and Byrne, W. (1999). Speaker adaptation with all-pass transforms. *Proceedings of the International Conference on Audition, Speech, and Signal Processing*, 2, 757-760.

McNeil, M.R. and Adams, S. (1991). A comparison of speech kinematics among apraxic, conduction aphasic, ataxic dysarthric, and normal geriatric speakers. *Clinical Aphasiology*, 19, 279-294.

McNeil, M.R., Caligiuri, M., and Rosenbek, J.C. (1989). A comparison of labiomandibular kinematic durations, displacements, velocities, and dysmetrias in apraxic and normal adults. *Clinical Aphasiology*, 17, 173-193.

McNeil, M.R. and Doyle, P.J. (2004). Apraxia of speech: nature and phenomenology. In Kent R. (ed)*The MIT Encyclopedia of Communication Disorders,* Cambridge, MA: MIT Press, pp. 101-103.

McNeil, M.R., Doyle, P.J., and Wambaugh, J. (2000). Chapter 9: Apraxia of speech: a treatable disorder of motor planning and programming. In Nadeau S.E., Gonzalez-Rothi L.J., and Crosson B. (eds) *Aphasia and Language. Theory to Practice*, New York, NY: Guilford Press, pp. 221-266.

McNeil, M.R., Hashi, M., and Southwood, H. (1994). Acoustically derived perceptual evidence for coarticulatory errors in apraxic and conduction aphasic speech production. *Clinical Aphasiology*, 22, 203-218.

McNeil, M. R., Odell, K. H., Miller, S. B., and Hunter, L. (1995). Consistency, variability, and target approximation for successive speech repetitions among apraxic, conduction aphasic, and ataxic dysarthric speakers. *Clinical Aphasiology*, 23, 39-55.

McNeil, M.R., Robin, D.A., and Schmidt, R.A. (2007). Chapter 15: Apraxia of speech: definition and differential diagnosis. In McNeil M.R. (ed) *Clinical Management of Sensorimotor Speech Disorders* (second edition), New York, NY: Thieme Medical Publishers, pp. 249-268.

McNeil, M.R., Weismer, G., Adams, S., and Mulligan, M. (1990). Oral structure nonspeech motor control in normal dysarthic aphasic and apraxic speakers: Isometric force and static position control. *Journal of Speech and Hearing Research*, 33, 255-268.

Merzenich, M.M. and Brugge, J.F. (1973). Representation of the cochlear partition of the superior temporal plane of the macaque monkey. *Brain Research*, 50, 275-296.

Mettler, F.A. (1949). *Selective Partial Ablation of the Frontal Cortex*. New York, NY:Paul Hoeber.

Miller, J. (1989). Auditory-perceptual representation of the vowel. *Journal of the Acoustical Society of America,* 85, 2114-2134.

Miller, N. (2002). The neurological bases of apraxia of speech. *Seminars in Speech and Language*, 23(4), 223-230.

Mlcoch, A.G., Darley, F.L., and Noll, J.D. (1982). Articulatory consistency and variability in apraxia of speech. *Clinical Aphasiology*, 11, 235-238.

Mohr, J.P. (1976). Broca's area and broca's aphasia. In Whitaker H. (ed) *Studies in neurolinguistics* (volume 1). New York, NY:Thieme Medical Publishers.

Mohr, J.P., Pessin, M.S., Finkelstein, S., Funkenstein, H.H., Duncan, G.W., and Davis, K.R. (1978). Broca aphasia: pathologic and clinical. *Neurology*, 28, 311-324.

Molnar-Szakacs, I., Iacoboni, M., Koski, L., and Mazziotta, J.C. (2005). Functional segregation within pars opercularis of the inferior frontal gyrus: Evidence from fMRI studies of imitation and action observation. *Cerebral Cortex.* 15, 986-994.

Moore, C.B. and Jongman, A. (1997). Speaker normalization in the perception of Mandarin Chinese tones. *Journal of the Acoustical Society of America,* 102, 1864-1877.

Morel, A. and Kaas, J.H. (1992). Subdivisions and connections of auditory cortex in owl monkeys. *Journal of Comparative Neurology*, 318, 27-63.

Morel, A., Garraghty, P.E., and Kaas, J.H. (1993). Tonotopic organization, architectonic fields, and connections of auditory cortex in macaque monkeys. *Journal of Comparative Neurology*, 335, 437-459.

Müller, H.M., King, JW., and Kutas, M. (1997). Event-related potentials elicited by spoken relative clauses. *Cognitive Brain Research*, 5(3), 193-203.

Nathaniel-James, D.A., Fletcher, P., and Frith, C.D. (1997). The functional anatomy of verbal initiation and suppression using the Hayling Test. *Neuropsychologia*, 35(4), 559-566.

Nearey, T.M. (1989). Static, dynamic, and relational properties in vowel perception. *Journal of the Acoustical Society of America,* 85, 2088-2113.

Nearey, T.M., Hogan, J., and Rozsypal, A. (1979). Speech signals, cues and features. In Prideaux G.D. (ed) *Perspectives in Experimental Linguistics*, Amsterdam: John Benjamins.

Nestor, P.J., Graham, N.L., Fryer, T.D., Williams, G.B., Patterson, K., and Hodges, J.R. (2003). Progressive non-fluent aphasia is associated with hypometabolism centered on the left anterior insula. *Brain*, 126, 2406-2418.

Nishitani, N., Schurmann, M., Amunts, K., and Hari, R. (2005). Broca's region: From action to language. *Physiology*, 20, 60-69.

Nixon, P., Lazarova, J., Hodinott-Hill, I., Gough, P., and Passingham, R. (2004). The inferior frontal gyrus and phonological processing: An investigation using rTMS. *Journal of Cognitive Neuroscience*, 16(2), 289-300.

Nota, Y. and Honda, K. (2003). Possible role on the anterior insula in articulation. *Proceedings of the 6th International Seminar on Speech Production*, Sydney.

Odell, K., McNeil, M.R., Rosenbek, J.C., and Hunter, L. (1990). Perceptual characteristics of consonant production by apraxic speakers. *Journal of Speech and Hearing Disorders*, 55, 345-359.

Odell, K., McNeil, M.R., Rosenbek, J.C., and Hunter, L. (1991). A perceptual comparison of prosodic features in apraxia of speech and conduction aphasia. *Clinical Aphasiology*, 19, 295-305.

Ogar, J., Willock, S., Baldo, J., Wilkins, D., Ludy, C., and Dronkers, N. (2006). Clinical and anatomical correlates of apraxia of speech. *Brain and Language*, 97(3), 343-350.

Ojemann, G.A., and Whitaker, H.A. (1978). Language localization and variability. *Brain and Language*, 6, 239-260.

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

Ostrowsky, K., Isnard, J., Ryvlin, P., Guenot, M., Fischer, C., and Mauguiere, F. (2000). Functional mapping of the insular cortex: Clinical implication in temporal lobe epilepsy. *Epilepsia*, 41(6), 681-686.

Özdemir, E., Norton, A., and Schlaug, G. (2006). Shared and distinct neural correlates of singing and speaking. *NeuroImage*, 33 628-635.

Page, M. (2000). Connectionist modeling in psychology: A localist manifesto. *Behavior and Brain Science*, 23, 443-467.

Page, M.P.A. and Norris, D. (1998). The primacy model: A new model of immediate serial recall. *Psychological Review*, 105(4), 761-781.

Palmeri, T.J., Goldinger, S.D., and Pisoni, D.B. (1993). Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19(2), 309-328.

Pantev, C., Hoke, M., Lehnertz, K., Lutkenhoner, B., Anogianakis, G., and Wittkowski, W. (1988). Tonotopic organization of the human auditory cortex revealed by transient auditory evoked magnetic fields. *Electroencephalography and Clinical Neurophysiology*, 69(2), 160-170.

Pantev, C., Hoke, M., Lutkenhoner, B., and Lehnertz, K. (1989). Tonotopic organization of the auditory cortex: pitch versus frequency representation. *Science*, 246(4929), 486-488.

Patterson, R., Nimmo-Smith, I., Holdsworth, J., Rice, P. (1987). An efficient auditory filterbank based on the gammatone function. *Annex B of the SVOS Final Report: Part A: The Auditory Filterbank.*

Patterson, R., Nimmo-Smith, I., Holdsworth, J., and Rice, P. (1988). *Spiral VOS Final Report: Part A: The Auditory Filterbank.*

Patterson, R.D. and Rice, P. (1987). A preliminary study of the feasibility of a hardware version of the auditory filterbank. *Annex A of the SVOS Final Report: Part A: The Auditory Filterbank.*

Patterson, R.D., Uppenkamp, S., Johnsrude, I.S., and Griffiths, T.D. (2002). The processing of temporal pitch and melody information in auditory cortex. *Neuron*, 36(4), 767-776.

Paulesu, W., Frith, C.D., and Frackowiak, R.S. (1993). The neural correlates of the verbal component of working memory. *Nature*, 362, 342-345.

Paulesu, E., Goldacre, B., Scifo, P., Cappa, S.F., Gilardi, M.C. Castiglioni, I., Perani, D., and Fazio, F. (1997). Functional heterogeneity of left inferior frontal cortex as revealed by fMRI. *Neuroreport*, 8, 2011-2017.

Peach, R.K. and Tonkovich, J.D. (2004). Phonemic characteristics of apraxia of speech resulting from subcortical hemorrhage. *Journal of Communication Disorders*, 37(1), 77-90.

Penagos, H., Melcher, J.R., and Oxenham, A. (2004). A neural representation of pitch salience in nonprimary human auditory cortex revealed with functional magnetic resonance imaging. *Journal of Neuroscience*, 24(30), 6810-6815.

Penfield, W. and Roberts, L. (1959). *Speech and Brain Mechanisms*. Princeton, NJ: Princeton University Press.

Petersen, S.E. and Fiez, J.A. (1993). The processing of single words studied with positron emission tomography. *Annual Reviews of Neurosci*ence, 16, 509-530.

Petersen, S.E., Fox, P.T., Posner, M.I., Mintun, M., Raichle, M.E. (1989). Positron emission tomographic studies of the cortical anatomy of single-word processing. *Nature*, 331, 585-589.

Petersen, S.E., Fox, P.T., Posner, M.I., Mintun, M., Raichle, M.E. (1989). Positron emission tomographic studies of the processing of single words. *Journal of Cognitive Neuroscience*, 1, 153-170.

Peterson, G.E. (1961). Parameters of vowel quality. *Journal of Speech and Hearing Research*, 4, 10-29.

Peterson, G.E. and Barney, H.L. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America*, 24, 175-184.

Petkov, C.I., Kayser, C., Augath, M., and Logothetis, N.K. (2006). Functional imaging reveals numerous fields in the monkey auditory cortex. *Public Library of Science Biology*, e215.

Phelps, E.A., Fahmeed, H., Blamire, A.M., and Shulman, R.G. (1997). fMRI of the prefrontal cortex during overt verbal fluency. *Neuroreport*, 8(2), 561-565.

Piaget, J. (1963). *The Origins of Intelligence in Children.* New York, NY: Norton.

Pierrehumbert, J. (2006). The next toolkit. *Journal of Phonetics*, 34, 516-530.

Plack, C.J., Oxenham, A.J., Popper, A.N., and Fay, R.R. (2005). *Pitch: Neural Coding and Perception*. New York: Springer Verlag.

Poldrack, R.A., Temple, E., Protopapas, A., Nagarajan, S., Tallal, P., Merzenich, M., and Gabrieli, J.D. (2001). Relations between the neural bases of dynamic auditory processing and phonological processing: evidence from fMRI. *Journal of Cognitive Neuroscience*, 13, 687-697.

Poldrack, R.A., Wagner, A.D., Prull, M.W., Desmond, J.E., Glover, G.H., and Gabrieli, J.D.E. (1999). Functional specialization for semantic and phonological processing in the left inferior prefrontal cortex. *NeuroImage*, 10, 15-35.

Qin, L., Sakai, M., Chimoto, S. and Sato, Y. (2005). Interaction of excitatory and inhibitory frequency-receptive fields in determining fundamental frequency sensitivity of primary auditory cortex neurons in awake cats. *Cerebral Cortex*, 15, 1371-1383.

Ragot, R. and Lepaul-Ercole, E. (1996). Brain potentials as objective indexes of auditory pitch extraction from harmonics. *Neuroreport*, 7(4), 905-909.

Raichle, M.E. (1996). What words are telling us about the brain. *Cold Spring Harbor Symposia on Quantitative Biology*, 61, 9-14.

Rauschecker, J.P. and Tian, B. (2004). Processing of band-passed noise in the lateral auditory belt cortex of the rhesus monkey. *Journal of Neurophysiology*, 91, 2578-2589.

Rauschecker, J.P., Tian, B., and Hauser, M. (1995). Processing of complex sounds in the macaque nonprimary auditory cortex. *Science*, 268(5207), 111-114.

Reale, R.A. and Imig, T.J. (1980). Tonotopic organization in the auditory cortex of the cat. *Journal of Comparative Neurology*, 192(2), 265-291.

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, 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, 259-276.

Rizzolati, G. and Arbib, M.A. (1998). Language within our grasp. *Trends in Neuroscience*, 21, 188-194.

Rizzolatti, G., Fadiga, L., Matelli, M., Bettinardi, V., Paulesu, E., Perani, D., and Fazio, F. (1996). Localization of grasp representations in humans by PET: 1. Observation versus execution. *Experimental Brain Research*, 111, 246-252.

Robin, D.A., Bean, C., and Folkins, J.W. (1989). Lip movement in apraxia of speech. *Journal of Speech and Hearing Research*, 32, 512-523.

Robin, D.A., Jacks, A., and Ramage, A.E. (2008). The neural substrates of apraxia of speech as uncovered by brain imaging: A critical review. In Ingham R.J. (ed) *Neuroimaging in Communication Sciences and Disorders.* San Diego: Plural Publishing.

Robin, D.A. and Schienberg, S. (1990). Subcortical lesions and aphasia. *Journal of Speech, Language, and Hearing Research,* 55, 90-100.

Roelofs, A. (1997). The WEAVER model of word-form encoding in speech production. *Cognition*, 64, 249-284.

Rogers, M.A. (1997). The vowel lengthening exaggeration effect in speakers with apraxia of speech: compensation, artifact, or primary deficit? *Aphasiology*, 11(4/5), 433-445.

Rogers, M.A., Eyraud, R., Strand, E.A., and Storkel, H.L. (1996). The effects of noise masking on vowel duration in three patients with apraxia of speech and a concomitant aphasia. *Clinical Aphasiology*, 24, 83-96.

Rogers, M.A. and Storkel, H.L. (1998). Reprogramming phonologically similar utterances: the role of phonetic features in pre-motor encoding. *Journal of Speech, Language, and Hearing Research*, 41, 258-274.

Rogers, M.A. and Storkel, H.L. (1999). Planning speech one syllable at a time: the reduced buffer capacity hypothesis in apraxia of speech. *Aphasiology*, 13(9/10/11), 793-805.

Romani, G.L., Williamson, S.J., and Kaufman, L. (1982). Tonotopic organization of the human auditory cortex. *Science*, 216(4552), 1339-1340.

Rosenbek, J.C. (2001). Darley and apraxia of speech. *Aphasiology*, 15(3), 261-273.

Rosenbek, J.C., Wertz, R.T., and Darley, F.L. (1973). Oral sensation and perception in apraxia of speech and aphasia. *Journal of Speech and Hearing Research*, 16, 22-36.

Rosenfield, D.B. (1991). Speech apraxia in cortico-basal-ganglionic degeneration. *Annals of Neurology*, 30, 296.

Roskies, A.L., Fiez, J.A., Balota, D.A., Raichle, M.E., and Petersen, S.E. (2001). Task-dependent modulation of regions in the left inferior frontal cortex during semantic processing. *Journal of Cognitive Neuroscience*, 13, 829-843.

Rueckert, L., Appollonio, I., Grafman, J., Jezzard, P., Johnson, R. Jr., Le Bihan, D. and Turner, R. (1994). Magnetic resonance imaging functional activation of left frontal cortex during covert word production. *Neuroimaging*, 4(2), 67-70.

Rumsey, J.M., Horwitz, B., Donahue, B.C., Nace, K., Maisog, J.M., and Andreason, P. (1997). Phonological and orthographic components of word recognition: a PET-rCBF study. *Brain*, 120, 739-759.

Salmelin, R., Hari, R., Lounasmaa, O.V. and Sams, M. (1994). Dynamics of brain activation during picture naming. *Nature*, 368, 463-465.

Sanchez-Valle, R., Forman, M.S., Miller, B.L., and Gorno-Tempini, M.L. (2006). From progressive nonfluent aphasia to corticobasal syndrome: A case report of corticobasal degeneration. *Neurocase*, 12(6), 355-359.

Schlösser R., Hutchinson, M., Joseffer, S., Rusinek, H., Saarimaki, A., Stevenson, J., Dewey, S.L. and Brodie, J.D. (1998). Functional magnetic resonance imaging of human brain activity in a verbal fluency task. *Journal of Neurology, Neurosurgery, and Psychiatry*, 64, 492-498.

Schulze, H., Hess, A., Ohl, F., and Scheich, H. (2002). Superposition of horseshoe-like periodicity and linear tonotopic maps in auditory cortex of the Mongolian gerbil. *European Journal of Neuroscience*, 15(6), 1077-1084.

Schwartz, R.G. (1988). Phonological factors in early lexical acquisition. In Smith M.D. and Locke J.L. (eds) *The Emergent Lexicon: The Child's Development of a Linguistic Vocabulary*, New York, NY: Academic Press, pp. 185-222.

Schwippert, C. and Benoit, C. (1997). Audiovisual intelligibility of an androgynous speaker. *Proceedings of the Workshop Audio Visual, and Speech Processing: Cognitive Computational Approaches*, Rhodes, Greece, 81-84.

Seddoh, S.A.K., Robin, D.A., Sim, H.S., Hageman, C., Moon, J.B., and Folkins, J.W. (1996). Speech timing in apraxia of speech versus conduction aphasia. *Journal of Speech and Hearing Research*, 39, 590-603.

Seldon, H.L. (1985). The anatomy of speech perception: Human auditory cortex. In Peters A. and Jones E.G. (eds) *Cerebral Cortex 4*, New York: Plenum Press, pp. 273-327.

Shankweiler, D. and Harris, K.S. (1966). An experimental approach to the problem of articulation in aphasia. *Cortex*, 2, 277-292.

Shuster, L.I. and Wambaugh, J.L. (2000). Perceptual and acoustic analyses of speech sound errors in apraxia of speech accompanied by aphasia. *Aphasiology*, 14(5/6), 635-651.

Slaney, M. (1993). An efficient implementation of Patterson-Holdsworth auditory filter bank. *Apple Computer Technical Report*, #35.

Slaney, M. (1998). Auditory toolbox, version 2. *Interval Research Corporation Technical Report* #10.

Slawson, A.W. (1968). Vowel quality and musical timbre as functions of spectrum envelope and fundamental frequency. *Journal of Acoustical Society of America*, 43, 87-101.

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

Spencer, K.A. and Rogers, M.A. (2005). Speech motor programming in hypokinetic and ataxic dysarthria. *Brain and Language*, 94, 347-366.

Spitzer, H., Desimone, R., and Moran, J. (1988). Increased attention enhances both behavioral and neuronal performance. *Science*, 240, 338-340.

Square, P.A., Darley, F.L., and Sommers, R.K. (1981). Speech perception among patients demonstrating apraxia of speech, aphasia, and both disorders. *Clinical Aphasiology Conference Proceedings,* 83-88.

Square, P.A., Darley, F.L., and Sommers, R.K. (1982). An analysis of the productive errors made by apractic speakers with differing loci of lesions. *Clinical Aphasiology Conference Proceedings*, 245-250.

Square, P.A., Roy, E.A., and Martin, R.E. (1997). Chapter 11: Apraxia of speech: another form of praxis disruption. In Rothi L.J. and Heilman K.M. (eds) *Apraxia: The Neuropsychology of Action*. New York, NY: Psychology Press, pp.173-206.

Stevens, K.N. (1998). *Acoustic Phonetics*. Cambridge, MA: MIT Press.

Stewart, L., Walsh, V., Frith, U., and Rothwell, J.C. (2001). TMS produces two dissociable types of speech disruption. *NeuroImage*, 13(3), 472-478.

Strand, E.A. and Johnson, K. (1996). Gradient and visual speaker normalization in the perception of fricatives. In Gibbon D. (ed) *Natural Language Processing and Speech Technology: Results of the 3rd KONVENS Conference, Bielefeld*. Berlin: Mouton de Gruyter, pp. 14-26.

Strand, E.A. and McNeil, M.R. (1996). Effects of length and linguistic complexity on temporal acoustic measures in apraxia of speech. *Journal of Speech and Hearing Research*, 39, 1018-1033.

Sugishita, M., Konno, K., Kabe, S., Yunoki, K., Togashi, O., and Kawamura, M. (1987). Electropalatographic analysis of apraxia of speech in a left hander and in a right hander. *Brain*, 110, 1393-1417.

Summerfield, Q. and Haggard, M.P. (1973). Vocal tract normalization as demonstrated by reaction times. *Reviews of Speech Research Progress*, 2, 1-12.

Sussman, H.M. (1986). A neuronal model of vowel normalization and representation. *Brain and Language*, 28(1), 12-23.

Sussman, H.M., Bessell, N., Dalston, E., and Majors, T. (1997). An investigation of stop place of articulation as a function of syllable position. *Journal of the Acoustical Society of America*, 101(5), 2826-2838.

Syrdal, A.K. and Gopal, H.S. (1986). A perceptual model of vowel recognition based on the auditory representation of American English vowels. *Journal of the Acoustical Society of America*, 79, 1086-1100.

Talavage, T.M., Ledden, P.J., Benson, R.R., Rosen, B.R., and Melcher, J.R. (2000). Frequency-dependent responses exhibited by multiple regions in human auditory cortex. *Hearing Research*, 150, 225-244.

Talavage, T.M., Sereno, M.I., Melcher, J.R., Ledden, P.J., Rosen, B.R., and Dale, A.M. (2004). Tonotopic organization in human auditory cortex reveled by progressions of frequency sensitivity. *Journal of Neurophysiology*, 91, 1282-1296.

Tandon, N., Narayana, S., Lancaster, J.L., Brown, S., Dodd, S., Vollmer, D.G., Ingham, R., Ingham, J., Liotti, M., and Fox, P. (2003). CNS resident award: Role of the lateral premotor cortex in articulation. *Clinical Neurosurgery*, 50, 341-349.

Tanji, K., Suzuki, K., Yamadori, A., Tabuchi, M., Endo, K., Fujii, T., and Itoyama, Y. (2001). Pure anarthia with predominantly sequencing errors in phoneme articulation: A case report. *Cortex*, 37, 671-678.

Thierry, G., Boulanouar, K., Kherif, F., Ranjeva, J.P.., and Demonet, J.F. (1999). Temporal sorting of neural components underlying phonological processing. *Neuroreport* 10, 2599-2603.

Titze, I.R. (1994). Mechanical stress in phonation. *Journal of Voice*, 8(2), 99-105.

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. *Proceedings of the 11th Annual Meeting of the Organization for Human Brain Mapping, Toronto*, 26(S1), S49.

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

Traunmüller, H. (1981). Perceptual dimension of openness in vowels. *Journal of the Acoustical Society of America,* 69, 1465-1475.

Trost, J.E. and Canter, G.J. (1974). Apraxia of speech in patients with Broca's aphasia: A study of phoneme production accuracy and error patterns. *Brain and Language,* 1, 63-79.

Tuller, B. (1984). On categorizing aphasic speech errors. *Neuropsychologia,* 22(5), 547-557.

Tunturi, A.R. (1952). A difference in the representation of auditory signals from the left and the right ears in the isofrequency of the right middle ectosylvian auditory cortex of the dog. *American Journal of Physiology,* 168, 712-727.

Turner, R.E. and Patterson, R.D. (2003). An analysis of the size information in classical formant data: Peterson and Barney (1952) revisited. *Journal of the Acoustical Society of Japan*,33(9), 585-589.

Ungerleider, L.G. and Mishkin, M. (1982). Two cortical visual systems: Separation of appearance and location of objects. In Ingle D.L., Goodale M.A., and Mansfield R.J.W. (eds) *Analysis of Visual Behavior,* Cambridge MA: MIT Press, pp. 549-586.

Van der Merwe, A. (2007). A theoretical framework for the characterization of pathological speech sensorimotor control. In McNeil M.R. (ed) *Clinical Management of Sensorimotor Speech Disorders* (second edition), New York, NY: Thieme Medical Publishers, pp. 3-18.

Van Lieshout, P.H.H.M., Bose, A., Square, P.A., and Steele, C.M. (2007). Speech motor control in fluent and dysfluent speech production of an individual with apraxia of speech and Broca's aphasia. *Clinical Linguistics and Phonetics*, 21(3), 159-188.

Vandenberghe, R., Price, C.J., Wise, R., Josephs, O., Frackowiak, R.S.J. (1996). Functional anatomy of a common semantic system for words and pictures. *Nature*, 383, 254-256.

Vanier, M. and Caplan, D. (1990). CT scan correlates of agrammatism. In Menn L. and Obler L. (eds) *Agrammatic Aphasia,* Amsterdam: John Benjamin's, pp. 97-114.

Varley, R. and Whiteside, S.P. (2001a). What is the underlying impairment in acquired apraxia of speech? *Aphasiology*, 15(1), 39-49.

Varley, R. and Whiteside, S.P. (2001b). Reply: exploring the enigma. *Aphasiology,* 15(1), 78-84.

Verbrugge, R.R., Strange, W., Shankweiler, D.P., and Edman, T.R. (1976). What information enables a listener to map a talker's vowel space? *Journal of the Acoustical Society of America*, 60, 198-212.

Vitevitch, M.S. and Luce, P.A. (1999). Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language*, 40, 374-408.

von Monakow, C. (1914). *Die Localisation in Grosshirn*. Wiesbaden:J.F. Bergmann.

Wagner, A.D., Schacter, D.L., Rotte, M., Koutstaal, W., Maril, A., Dale, A.M., Rosen, B.L., and Buckner, R.L. (1998). Building memories: Remembering and forgetting of verbal experiences as predicted by brain activity. *Science*, 281, 1188-1191.

Walker, S., Bruce, V., and O'Malley, C. (1995). Facial identity and facial speech processing: Familiar faces and voices in the McGurk effect. *Perception and Psychophysics*, 57, 1124-1133.

Wambaugh, J.L. and Doyle, P.J. (1994). Treatment for acquired apraxia of speech: a review of efficacy reports. *Clinical Aphasiology*, 22, 231-243.

Wambaugh, J.L., Doyle, P.J., West, J.E., and Kalinyak, M.M. (1995). Spectral analysis of sound errors in persons with apraxia of speech and aphasia. *American Journal of Speech-Language Pathology*, 4, 186-192.

Wambaugh, J.L., Duffy, J.R., McNeil, M.R., Robin, D.A., and Rogers, M.A. (2006a). Treatment guidelines for acquired apraxia of speech: A synthesis and evaluation of the evidence. *Journal of Medical Speech Language Pathology*, 14(2), xv-xxxiii.

Wambaugh, J.L., Duffy, J.R., McNeil, M.R., Robin, D.A., and Rogers, M.A. (2006b). Treatment guidelines for acquired apraxia of speech: Treatment descriptions and recommendations. *Journal of Medical Speech Language Pathology*, 14(2), xxxv-lxii.

Wambaugh, J.L. and Nessler, C. (2004). Modification of sound production treatment for apraxia of speech: Acquisition and generalization effects. *Aphasiology*, 18, 407-427.

Wambaugh, J.L., Nessler, C., Bennett, J., and Mauszycki, S.C. (2004). Variability in apraxia of speech: A perceptual and VOT analysis of stop consonants. *Journal of Medical Speech-Language Pathology*, 12, 221-227.

Watrous, R.L. (1991). Current Status of Peterson-Barney vowel formant data. *Journal of the Acoustical Society of America*, 89, 2459-2460.

Wegman, S., McAllaster, D., Orloff, J., and Peskins, B. (1996). Speaker normalization on conversational telephone speech. *Proceedings of the International Conference on Audition, Speech, and Signal Processing*, 1, 339-341.

Wertz, R.T., LaPointe, L.L., and Rosenbek, J.C. (1984). *Apraxia of Speech in Adults: The Disorder and its Management*. Orlando, FL:Grune and Stratton Inc.

Wessa, P. (2007). *Free Statistics Software*, Office for Research Development and Education. Version 1.1.22-rl. http://www.wessa.net (1 Oct 2007).

Wessinger, C.M., Buonocore, M.H., Kussmaul, C.L., and Mangun, G.R. (1998). Tonotopy in human auditory cortex examined with functional magnetic resonance imaging. *Human Brain Mapping*, 5(1), 18-25.

Whiteside, S.P. and Varley, R.A. (1998). A reconceptualisation of apraxia of speech: a synthesis of evidence. *Cortex*, 34, 221-231.

Whitfield, I.C. (1980). Auditory cortex and the pitch of complex tones. *Journal of the Acoustical Society of America*, 67(2), 644-647.

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

Zaehle, T., Geiser, E., Alter, K., Jancke, L., and Meyer, M. (2008). Segmental processing in the human auditory dorsal stream. *Brain Research*, 1220, 179-190.

Zahorian, S.A. and Jagharghi, A.J. (1991). Speaker normalization of static and dynamic vowel spectral features. *Journal of the Acoustical Society of America*, 90(1), 67-75.

Zangwill, O. (1975). Excision of Broca's area without persistent aphasia. In Zulch K.J., Creutzfeldt O., and Galbraith G.C. (eds) *Cerebral Localization*, Berlin:Springer, pp. 258-263.

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

Zatorre, R.J., Meyer, E., Gjedde, A., and Evans, A.C. (1996). PET studies of phonetic processing of speech: review, replication, and reanalysis. *Cerebral Cortex*, 6, 21-30.

Zhan, P. and Waibel, A. (1997). Vocal tract length normalization for large vocabulary continuous speech recognition. *Technical Report CMU-CS-97-148*, School of Computer Science, Carnegie Mellon University.

Zhan, P. and Westphal, M. (1997). Speaker normalization based on frequency warping. *Proceedings of the International Conference on Audition, Speech, and Signal Processing*, 2, 1039-1041.

Ziegler, W. (2002). Psycholinguistic and motor theories of apraxia of speech. *Seminars in Speech and Language*, 23(4), 231-243.

Ziegler, W. (2003). Speech motor control is task-specific. Evidence from dysarthria and apraxia of speech. *Aphasiology*, 17(1), 3-36.

Ziegler, W. and Von Cramon, D. (1985). Anticipatory coarticulation in a patient with apraxia of speech. *Brain and Language*, 26, 117-130.

Ziegler, W. and Von Cramon, D. (1986a). Disturbed coarticulation in apraxia of speech: Acoustic evidence. *Brain and Language*, 29, 34-47.

Ziegler, W. and Von Cramon, D. (1986b). Timing deficits in apraxia of speech. *European Archives of Psychiatry and Neurological Sciences*, 236, 44-49.

Zoccolan, D., Kouh, M., Poggio, T., and DiCarlo, J.J. (2007). Trade-off between object selectivity and tolerance in monkey inferotemporal cortex. *Journal of* Neuroscience, 26, 13025-13026.

CURRICULUM VITAE

Heather Ames

Work Experience

2004-present **Boston University** Boston, MA

Research Assistant

- **Neural network modeling of speech perception**. This research aims to determine how the human brain creates a speaker and rate invariant representation of speech sounds and how the brain learns to recognize those signs and categorize them with their associated meaning. The model addresses filtering speech sounds by the ear, streaming acoustical sources, pitch perception, and learning and categorization.
- **Behavioral and modeling work on speech motor control of aphasics**. In this project, I am coordinating with Boston University, the VA hospital, and Northeastern University to perform a behavioral study investigating the speech motor control of patients with aphasia. My responsibilities include computational modeling, being the interface with IRBs at both the VA and BU, overseeing study design, patient recruitment, data analysis, obtaining consent from patients, and writing a book chapter and paper for submission to a peer reviewed journal.
- **Hi-tech software development project.** I am co-leading a team of seven CNS graduate students and post docs in a project aimed at developing an innovative software platform for advanced neural modeling. The software, the KDE Integrated NeuroSimulation Software (KInNeSS, www.kinness.org), has become the primary tool in many departmental research projects in the design and testing of complex neural systems.

2006-present **Neurala LLC** Boston, MA

Co-Founder and COO

- Oversee the day to day operations of Neurala LLC including accounting and business client communications.
- Lead engineer and project manager for defense related contract with Biomimetic Systems involving gunshot detection and classification.
- Co-author of business plan.
- Assist development of business goals and product design.
- Project Manager for general product development and contract work.

2000-2003 **Barons Jewelers** San Leandro, CA

Corporate Office Manager and Systems Administrator

- Managed office employees at four locations in both day to day operations and sales and accounting issues involving customers, business clients, other employees, and the computer system.
- Co-authored a business plan which successfully received multi-million dollar financing.
- Managed company LAN hardware and data transfer.
- Helped facilitate conversion to a computerized point of sales system.
- Conducted employee training sessions and meetings.
- Attained accounts receivable experience in billing and in-house finances.

1998-2000 **University of California, Berkeley** Berkeley, CA

Lab Assistant

- Work study position in a water ecology laboratory.
- Identified taxonomy of aquatic insects through sample analysis and processing.
- Organized data into spreadsheets.

Education

2003-2009 **Boston University** Boston, MA

- Candidate for PhD degree in Cognitive and Neural Systems, May 2009
- Cumulative GPA of 3.9
- *Concentration*: Biologically inspired speech perception and production neural network modeling.
- *PhD Advisor*: Dr Stephen Grossberg
- *Dissertation title:* Neural dynamics of speech perception and production: from speaker normalization to apraxia of speech

1998-2003 **University of California - Berkeley** Berkeley, CA

- Bachelor of Arts degree received in May 2003.
- Received departmental honors and cumulative GPA of 3.52.
- Majored in cognitive science with an emphasis in neuroscience.

Leadership

- President of the NSF sponsored CELEST (Center for Excellence in Education, Science, and Technology) Science of Learning Center Student Organization from 2006-2008
- Advisory board member for the NSF sponsored CELEST Science of Learning Center Student Organization from 2008-2009
- Co-leader of the NSF sponsored inter-Science of Learning Center (iSLC) and co-author of accepted grant to fund a student and postdoc led conference series
- Chairperson of the CELEST student led workshop planning committee for the International Conference on Cognitive and Neural Systems (ICCNS 2008 and 2009) in Boston, MA.
- Committee member for the CELEST student led Career Day organizational a planning committee (2007 and 2008).

Teaching Experience

Fall 2004 **Boston University** Boston, MA

- Teaching Fellow for CN 510: Introduction to Neural Network Modeling
- Recipient of 2004-2005 Outstanding Teaching Fellow Award

Awards Received

- UC Berkeley Undergraduate Grant Recipient
- Donald Schaefer Physical Sciences Scholar
- H. Wollenberg Grant Recipient
- Iowa Byrd Scholar
- RIA Federal Credit Union Scholarship Recipient
- Girl Scout Silver Award
- Ralston Purina Scholarship Recipient

Membership

- Student member of the Acoustical Society of America
- Student member of the International Neural Network Society
- Student member of the Society for Neuroscience

Invited Panelist

- Joint Fire Science Program Eastern Risk Roundtable 2007
- RUSLE2 & LiDAR Expert Panel 2008

Peer Reviewer

- Spatial Vision, Special Issue on Vision Science and Art
- NSF sponsored inter Science of Learning Centers Workshop (2008)
- International Joint Conference on Neural Networks (2008, 2009)
- Neuropsychologia

Skills

- Microsoft Word, Excel, and PowerPoint, Adobe Photoshop, PageMaker, FileMaker and strong internet skills.
- Operating systems: Unix, Linux, Windows 95/98, 2000, ME, XP.
- Programming languages: Matlab, LISP, Scheme, C++
- Accounts receivable

Related Coursework

- Computational Methods in Cognitive and Neural Systems
- Principles and Methods of Cognitive and Neural Modeling I
- Neural and Computational Models of Vision
- Neural and Computational Models of Adaptive Movement Planning and Control
- Neural and Computational Models of Recognition, Memory and Attention
- Neural and Computational Models of Speech Perception and Production
- Neural and Computational Models of Conditioning, Reinforcement, Motivation and Rhythm
- Advanced Topics in Neural Modeling
- Neural and Computational Models of Planning and Temporal Structure in Behavior
- Models of Visual Perception
- Topics in Sensory-Motor Control
- Qualitative Theory of Ordinary Differential Equations

Patents

- Gorchetchnikov A., Ames H.M., Versace M., and Santini F. (2006) Hardware, system and methods for acceleration of massively parallel computations (filed provisional patent application US60/826,892; September 2006).
- Gorchetchnikov A., Ames H.M., Versace M., and Santini F. (2007) Graphic Processor Based Accelerator System and Method(utility patent application US 11/860,254; September 2007).

Publications

Journal articles

- Versace M., Ames H.M., Leveille J., Fortenberry B., Mhatre H., and Gorchetchnikov A. (2008). Kinness: A modular framework for computational neuroscience. *Neuroinformatics*. 6(4), 291-309.
- Ames H. and Grossberg S. (2008). Speaker normalization using cortical strip maps: A neural model for steady-state vowel categorization. *Journal of the Acoustical Society of America*. 124(6), 3918-3936.

Abstracts

- Ames H.M. and Grossberg S. (2006). Neural dynamics of auditory streaming, speaker normalization, and speech categorization. Society for Neuroscience Abstracts, Atlanta, GA (SFN 2006).
- Ames H.M. and Grossberg S. (2006). Neural dynamics of auditory streaming, speaker normalization, and speech categorization. NSF SLC PI Meeting Abstracts, Washington, DC (NSF SLC 2006).
- Ames H.M. and Grossberg S. (2007). Speech categorization through auditory cortical interactions. CELEST External Advisory Board Review Abstracts, Boston, MA (EASRB CELEST 2007).
- Ames H.M. and Grossberg S. (2007). Cortical maps used in speaker normalization. NSF CELEST Annual Review Abstracts, Boston, MA (NSF CELEST 2007).
- Ames H.M. and Grossberg S. (2007). Neural dynamics of speaker normalization used in steady-state vowel identification. International Conference on Cognitive and Neural Systems Abstracts, Boston, MA (ICCNS 2007).
- Ames H.M. and Grossberg S. (2007). Speaker normalization during steady state vowel identification. NSF SLC PI Meeting Abstracts, Washington, DC (NSF SLC 2007).
- Ames H.M. and Grossberg S. (2007). Speaker normalization using cortical strip maps: A neural model for steady state vowel identification. Computational Cognitive Neuroscience Conference Abstracts, San Diego, CA (CCNC 2007).
- Ames H.M. and Grossberg S. (2007). Speaker normalization using cortical strip maps: A neural model for steady state vowel identification. Acoustical Society of America Abstracts, New Orleans, LA (ASA 2007).
- Ames H. and Grossberg S. (2008). Speaker normalization using cortical strip maps: A neural model for steady-state vowel identification. NSF sponsored inter-Science of Learning Centers Abstracts, Pittsburgh, PA (NSF iSLC 2008).
- Ames H. and Grossberg S. (2008). Speaker normalization using cortical strip maps: A neural model for steady-state vowel identification. CELEST External Advisory Board Review Abstracts, Boston, MA (EASRB CELEST 2008).
- Gorchetchnikov A., Ames H., and Versace M. (2008). Simulating biologically realistic neural models on graphic processing units. International Conference on Cognitive and Neural Systems Abstracts, Boston, MA (ICCNS 2008).
- Versace M., Ames H., Leveille J., Fortenberry B., Gorchetchnikov A. (2008). KInNeSS: A modular framework for computational neuroscience. International Conference on Cognitive and Neural Systems Abstracts, Boston, MA (ICCNS 2008).
- Versace M., Ames H., Leveille J., Fortenberry B., and Gorchetchnikov A. (2008). KInNeSS: A modular framework for computational neuroscience. CELEST Annual Review Abstracts, Boston, MA (NSF CELEST 2008).
- Ames H. and Grossberg (2008). Speaker normalization using cortical strip maps:

A neural model for steady-state vowel identification. CELEST Annual Review, Abstracts, Boston, MA (NSF CELEST 2008).

- Ames H. (2008). Learning technologies: embedding CELEST models in real world applications. NSF SLC PI Meeting Abstracts, Washington, DC (NSF SLC 2008).
- Ames H. and Grossberg S. (2008). Speaker normalization using cortical strip maps: A neural model for steady-state vowel identification. Auditory Perception, Cognition, and Action Abstracts, Chicago, IL (APCAM 2008).
- Ames H., Versace M., and Gorchetchnikov A. (2009). How can computational neuroscience benefit real world technological applications? NSF sponsored inter-Science of Learning Centers Abstracts, Seattle, WA (NSF iSLC 2009).

Presentations

- Ames H.M. (2007). Recognition through hearing. CELEST Education Summer Workshops.
- Ames H.M. and Grossberg S. (2007). Speaker normalization using cortical strip maps: A neural model for steady state vowel identification. Computational Cognitive Neuroscience Conference Abstracts, San Diego, CA (CCNC 2007).
- Ames H.M., Booth J.L., Hausmann R.G.M., Lee, T., Roll, I., and Zimmerman H. (2007). SLC student workshop presentation. NSF SLC PI Meeting, Washington, DC (NSF SLC 2007).
- Ames H.M., Booth J.L., Hausmann R.G.M., Lee, T., Roll, I., and Zimmerman H. (2008). First Annual iSLC workshop overview presentation. NSF sponsored inter-Science of Learning Centers, Pittsburgh, PA (NSF iSLC 2008).
- Ames, H. (2008). CELEST as a science of learning center. NSF sponsored inter-Science of Learning Centers, Pittsburgh, PA (NSF iSLC 2008).
- Ames H. and Versace M. (2008). Computing with neural interfaces introduction. International Conference on Cognitive and Neural Systems, Boston, MA (ICCNS 2008).
- Ames, H. (2008). How can neural networks help us? RUSLE2 & LiDAR Expert Panel Meeting, Nebraska City, NB (LiDAR 2008).
- Ames H., Kantak K., and Liederman J. (2008). Diversity activities within CELEST. CELEST Annual Review, Boston, MA (NSF CELEST 2008).
- Ames H. (2008). Learning technologies: embedding CELEST models in real world applications. NSF SLC PI Meeting, Washington, DC (NSF SLC 2008).
- Ames H. and Grossberg S. (2008). Speaker normalization using cortical strip maps: A neural model for steady-state vowel identification. Auditory Perception, Cognition, and Action, Chicago, IL (APCAM 2008).