BOSTON UNIVERSITY

SCHOOL OF MEDICINE

Dissertation

NEURAL REPRESENTATIONS USED BY BRAIN REGIONS UNDERLYING SPEECH PRODUCTION

by

JENNIFER ANNE SEGAWA

B.A., Johns Hopkins University, 2004

Submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

2013

Approved by

First Reader _____

Frank H. Guenther, Ph.D.

Professor of Speech, Language, & Hearing Sciences and

Biomedical Engineering

Second Reader _____

Daniel Bullock, Ph.D.

Professor of Psychology

Third Reader _____

David W. Gow, Ph.D.

Assistant in Neuropsychology, Massachusetts General Hospital

Professor of Psychology, Salem State College

NEURAL REPRESENTATIONS USED BY BRAIN REGIONS UNDERLYING SPEECH PRODUCTION

(Order No.)

JENNIFER SEGAWA

Boston University School of Medicine, 2013

Major Professor: Frank H. Guenther, Ph.D., Professor of Speech, Language, & Hearing Sciences and Biomedical Engineering

Abstract

Speech utterances are phoneme sequences but may not always be represented as such in the brain. For instance, electropalatography evidence indicates that as speaking rate increases, gestures within syllables are manipulated separately but those within consonant clusters act as one motor unit. Moreover, speech error data suggest that a syllable's phonological content is, at some stage, represented separately from its syllabic frame structure. These observations indicate that speech is neurally represented in

multiple forms. This dissertation describes three studies exploring representations of speech used in different brain regions to produce speech.

The first study investigated the motor units used to learn novel speech sequences. Subjects learned to produce a set of sequences with illegal consonant clusters (e.g. GVAZF) faster and more accurately than a similar novel set. Subjects then produced novel sequences that retained varying phonemic subsequences of previously learned sequences. Novel sequences were performed as quickly and accurately as learned sequences if they contained no novel consonant clusters, regardless of other phonemic content, implicating consonant clusters as important speech motor representations.

The second study investigated the neural correlates of speech motor sequence learning. Functional magnetic resonance imaging (fMRI) revealed increased activity during novel sequence productions in brain regions traditionally associated with non-speech motor sequence learning – including the basal ganglia and premotor cortex – as well as regions associated with learning and updating speech motor representations based on sensory input – including the bilateral frontal operculum and left posterior superior temporal sulcus (pSTs). Behavioral learning measures correlated with increased response for novel sequences in the frontal operculum and with white matter integrity under the pSTs, implicating functional and structural connectivity of these regions in learning success.

The third study used fMRI to understand the neural representations of syllabic frame structure and phonological content. The right lateral cerebellum – implicated in

movement timing – was sensitive to syllabic frame structure dissociated from phonological content. The right anterior cerebellum, right posterior superior temporal cortex, and left supplementary motor area – all associated with sensory-motor functions – were sensitive to phonological content.

Taken together, these results shed light on different representations used across the brain network underlying speech production.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

adSTs	Anterior dorsal superior temporal sulcus
aFO	Anterior frontal operculum
alCb	Anterior lateral cerebellum
aIFs	Anterior inferior frontal sulcus
aINS	Anterior insula
aMFG	Anterior middle frontal gyrus
aSMG	Anterior supramarginal gyrus
aSTg	Anterior superior temporal gyrus
Bilat	Bilateral
BOLD	Blood oxygen level-dependent
Cau/Caud	Caudate
dIFo	Dorsal inferior frontal gyrus, pars opercularis
dIFt	Dorsal inferior frontal gyrus, pars triangularis
DIVA	Directions into velocities of articulators
DTI	Diffusion tensor imaging
FA	Fractional anisotropy
FO	Frontal operculum
fMRI	Functional magnetic resonance imaging
fMRI-RS	Functional magnetic resonance imaging repetition suppression
FWHM	Full-width half-maximum

GLM	General linear model
GODIVA	Gradient order directions into velocities of articulators
GP	Globus pallidus
GPi	Internal segment of the globus pallidus
Hg	Heschl's gyrus
HRF	Hemodynamic response function
IFo	Inferior frontal gyrus, pars opercularis
IFS	Inferior frontal sulcus
ITO	Inferior temporal-occipital cortex
midPMC	Middle premotor cortex
OSCAR	Oscillator-based associative recall
Pal/Pall	Pallidum
pCO	Posterior central operculum
pdSTs	Posterior dorsal superior temporal sulcus
PFC	Prefrontal cortex
pFO	Posterior frontal operculum
pIFs	Posterior inferior frontal gyrus
PMC	Premotor cortex
PO	Parietal operculum
PP	Planum polare
preSMA	Presupplementary motor area

pSTg	Posterior superior temporal gyrus
pSTs	Posterior superior temporal sulcus
Put	Putamen
PT	Planum temporale
ROI	Region of interest
SFC	State feedback control
SMA	Supplementary motor area
smCb	Superior medial cerebellum
SPL	Superior parietal lobule
SSC	Subsyllabic constituent
SSM	Speech sound map
Thal	Thalamus
VA	Ventral anterior nucleus of the thalamus
vIFo	Ventral inferior frontal gyrus, pars opercularis
vIFt	Ventral inferior frontal gyrus, pars triangularis
VL	Ventral lateral nucleus of the thalamus
vMC	Ventral motor cortex
vSCC	Ventral somatosensory cortex
vPMC	Ventral premotor cortex
WEAVER	Word-form encoding by activation and verification

1. INTRODUCTION

The primary goal of this dissertation is to test the validity of various representations of speech possibly used by the brain for speech production. Previous behavioral and neuroimaging evidence suggests that speech is represented in different forms across speech production planning stages and across regions of the brain. A secondary aim of this project is to incorporate theories from a variety of fields studying speech or movement – e.g. motor sequence learning, evolutionary linguistics, phonology – that have yet to be applied to speech or to be applied to the brain. This research will attempt to provide a neural basis for these hypotheses as they apply to speech motor sequencing. To these ends, this work describes three studies that use behavioral and neuroimaging methodologies to understand representations of speech used in the brain. A fourth study using neuroimaging is also presented that had underlying methodological flaws.

1.1. Neuroimaging

A researcher's available methodologies are limited when studying speech. Humans are the only animals capable of language. For this reason, invasive methodologies – such as pharmacological lesions or single cell recordings – are unavailable. An exception is electrocorticography and cortical stimulation mapping; these data can be collected from electrodes placed directly on the cortical surface of epileptic patients undergoing brain surgery. However, the neural processes of these patients may not reflect those of healthy subjects. Moreover, electrode placement is primarily for clinical observation and

may be indicative of areas of abnormal activity. To study neural function in healthy subjects, only non-invasive imaging methods are currently available to researchers.

1.1.1. Functional magnetic resonance imaging

Functional magnetic resonance imaging (fMRI) is one such noninvasive method used to study brain activity in healthy subjects. Subjects are exposed only to magnetic fields in this methodology. FMRI measures the difference in magnetic resonance between oxygenated and deoxygenated blood. This difference in oxygenation is associated with the energy expenditure needed for increased neural activity and is used to infer the blood oxygen level-dependent (BOLD) hemodynamic response associated with performing an experimental task. Therefore, local differences in BOLD response during the performance of one cognitive task compared to another are believed to reflect the differences in the neural activity required for those tasks.

Beyond its non-invasive nature, fMRI is a useful methodology because it has relatively fine spatial resolution. One study estimated the point spread function of the BOLD response using a 3 Tesla scanner and gradient-echo echo-planar imaging with parameters similar to those used in this dissertation; the authors found a full width half maximum (FWHM) of 3.9 ± 0.7mm (Parkes et al., 2005). In comparison, electroencephalography (EEG) has a spatial resolution of tens or hundreds of millimeters, at least an order of magnitude less precise than fMRI (Ferree, et al., 2001); MEG has a similar spatial resolution (da Silva, et al., 1991; Liu, et al., 2002; Molins, et al., 2008).

MEG and EEG data can be combined with anatomical MRI volumes to constrain the dipole estimates based on individual subjects' sulci and gyri locations and orientations (Dale & Sereno, 1993). However, even with this technique, MEG/EEG is still limited in spatial accuracy compared to fMRI. Depending on the specific algorithm used, estimates are error-prone in particular anatomical regions. For instance, the minimum norm estimate approach (Hamalainen & Ilmoniemi, 1984, Dale & Sereno, 1993) is vulnerable to displacement and depth errors in the Rolandic operculum and medial brain surfaces, while dynamic statistic parametric maps (Dale, et al., 2000) – a measure based on the statistical significance of estimates – is more vulnerable to error in medial frontal regions, anterior temporal cortex, and inferior frontal cortex (Lin, et al., 2006). In both techniques, some of the affected brain regions include those important to understanding the speech network.

Another strength of fMRI is that it can be used to examine neural activity across the entire brain, including subcortical activity. Transcranial magnetic stimulation (TMS) – which transiently disrupts neural activity in awake subjects by inducing a current with a rapidly changing magnetic field – can only be applied to one area of the brain at a time. Moreover, because the current is applied across the scalp, TMS is limited to surface structures of the brain like the outside radius of the cortex and some of the cerebellum; direct stimulation of subcortical nuclei is not possible, and indirect stimulation must pass through overlying cortical tissue (see Bolognini & Ro, 2010). Similarly, MEG and EEG cannot capture signals from subcortical structures because they are too distant from the sensors.

FMRI does have limitations. Its temporal resolution is coarse – 3 or more orders of magnitude less precise – compared to methods measuring electromagnetic activity such as MEG and EEG. The hemodynamic response takes seconds to peak and varies from subject to subject (Aguirre, et al., 1998; Kim et al., 1997), so while fMRI allows researchers to easily infer locations of activity in the brain, there is less that can be gleaned about the timing of this activity.

Another limitation of fMRI is that data interpretation rests on the assumption that the BOLD response is tightly coupled with neuronal activity. However, the exact mechanism of the BOLD response is still unclear (see Ekstrom, 2010). There is some evidence that local field potentials (LFPs) – reflecting the perisynaptic activity of neural populations – contribute the most strongly to the hemodynamic response (see Logothesis, 2008 for review). However, glial cells – in particular, astrocytes – have been shown to affect the hemodynamic response (Attwell, et al., 2010; Schulz, et al., 2012), as has neuronal spike rate (Kida et al., 2006; Mukamel et al., 2005; Nir et al., 2007). To complicate matters, the BOLD signal in the hippocampus appears to be dissociated from both LFPs and spiking rate (Angenstein, et al., 2009; Ekstrom, et al., 2009). Moreover, neural oscillations at certain frequencies (see Singh, 2012 for review) and baseline γ amino-butyric acid concentrations also appear to be correlated with the BOLD signal (Donahue, et al., 2010; Muthukumaraswamy, et al., 2012). In short, there are many candidates for the physiological basis of the BOLD response, but so far, all appear to be correlational, and none have been shown to be definitively and uniquely causal. Researchers using

fMRI to study the brain measure fMRI “activity” without precisely knowing where that “activity” comes from.

Moreover, because the origin of the BOLD response is unclear, researchers cannot definitively characterize the neural circuitry underlying an fMRI finding. Thus far, it is impossible to decisively link increased BOLD to neuronal excitation or decreased BOLD to neuronal inhibition. For instance, proportional increases to both inhibitory and excitatory cells’ activity can lead to an increase in the BOLD response, and purely inhibitory activity can lead to increased O_2 metabolism (Logothesis, 2008). Despite these limitations, fMRI remains one of the few available tools for studying the neural bases of speech of healthy subjects and allows researchers to precisely localize differences in metabolic activity correlated with different cognitive tasks.

1.1.2. Repetition suppression

1.1.2.1. Origin of repetition suppression

Two studies presented in this work use fMRI repetition suppression (fMRI-RS) paradigms. This technique uses fMRI to measure the local reduction in the hemodynamic response after repeated presentation of a stimulus encoded by that area. There are currently 3 major theories on how and why repetition suppression occurs: sharpening, synchrony, and facilitation.

The sharpening model is based on population coding in which some cells in a neuron population are weakly tuned to a stimulus and some are tightly tuned. The model posits

that with repeated presentations, the activity of weakly tuned neurons drops off, leaving only neurons that best represent the stimulus to respond (Desimone, 1996; Wiggs & Martin, 1998). This reduces the overall activity across a neuron population, but maintains the important information that is encoded in the response. Single cell recordings only show population-wide activity reduction after weeks or months of stimulus exposure. Short term exposure on the order of seconds – as is seen in fMRI-RS –shows selective reduction of neural activity only in the neurons best representing the stimulus (Baker, et al., 2002; Freedman, et al., 2006; Li, et al., 1993; MacMahon & Olson, 2007). It should be noted that the studies presented in this dissertation rely on a similar short term stimulus exposure of approximately 10 s to evoke fMRI-RS.

The synchrony theory hypothesizes that repetition suppression occurs because neurons fire with more efficient timing after repeated exposures to a stimulus (Gotts, et al., 2012). If pre-synaptic neurons fire synchronously, their combined efforts are more likely to depolarize a post-synaptic neuron enough to fire, than if they fired at different times. In this way, synchronously firing pre-synaptic neurons can fire less than those firing isochronously, while still causing the post-synaptic neuron to fire. So far, however, evidence for the synchrony theory has largely been limited to evidence from biologically-plausible computational models (Gotts, 2003; Bazhenov, et al., 2005). A study using magnetoephelography on human subjects also supports this theory, showing increased phase-locking in alpha between frontal and temporal cortices for repeated over novel stimuli. This suggests that repetition leads to increased coupling between brain regions (Ghuman et al., 2008). However, there is currently no evidence from single-cell recordings that directly support (or refute) this model.

The facilitation theory and its more complex counterpart, the top-down bottom-up theory hypothesize that repetition suppression occurs because repeated stimuli presentations lead to faster, overall time course of activity, which require shorter periods of sustained neural activity, and in turn, reduced BOLD responses. The facilitation model, also known as the accumulation model, posits that with repeated presentations of a stimulus, neural activity time courses occur faster. (James, et al., 1999; James & Gauthier, 2006). However, this does not appear to occur in single cell recordings (Anderson et al., 2008; Pedreira et al., 2010). A more nuanced version of the facilitation/accumulation model is based on top-down/bottom-up theories. This theory posits that "higher-level" brain areas exert top-down expectations on "lower-level" brain areas (Henson, 2003; James & Gauthier, 2006). When a stimulus is first presented, the top-down expectations and bottom-up sensory information are mismatched, but, with repeated presentations, the expectations and sensory information match. This leads to a more efficient network that either reduces or hastens neural activity time courses. Perhaps the most convincing evidence for this theory comes from Ewbank et al. (2011) who used an fMRI-RS paradigm to study the body-sensitive portion of the visual cortex. During blocks when subjects repeatedly saw the same image of a human body, dynamic causal modeling[1] demonstrated both top-down and bottom-up connectivity. However, when the image changed within a block, either by varying the perspective on or the size of the body, only bottom-up connectivity was found.

[1] Dynamic causal modeling (DCM) of neuroimaging data is a controversial method (e.g. Lohmann et al., 2012) for many reasons including limited model-space, However, some researchers (e.g. Roebroeck et al., 2011) believe that with careful methodology – particularly, a rigorous selection of brain regions included in the model – DCM can provide useful results.

It should be noted that in response to Gotts, Chow, and Martin's (2012) characterization of these three theories of the neural basis of repetition suppression, both Henson (2012) and Friston (2012) suggested that the theories are not mutually exclusive; they could be descriptions of different levels of the same effect. Synchrony could be a low-level description of how repetition suppression occurs between neurons. Sharpening could be a result of that effect, and top-down-bottom up could be the higher-level goal of these changes. While it appears that the origin of fMRI-RS is still unclear, this type of theory that addresses many levels of neural representation and research appears to be honing how researchers think about the effect.

1.1.2.2. Studies using fMRI-RS

Despite a lack of consensus on the neural basis of the repetition suppression effect, many researchers have successfully used this technique to study neural processes (e.g., Henson, et al., 2002; Ishai, 2004). While much of the founding literature focused on object and word recognition processes (Dehaene et al., 2001; Desimone, 1996; James et al., 1999), more recent studies have successfully applied this paradigm to motor tasks (e.g., Dehaene-Lambertz et al., 2006; Hamilton & Grafton, 2009; Hasson, et al., 2006; Heim, et al., 2002; Majdandžić, et al., 2009; Orfanidou et al., 2006). Of these studies, only a few have focused on understanding speech representations in the brain. In an fMRI-RS study of speech production, Graves and colleagues (2008) used a parametric analysis that manipulated the number of pseudoword repetitions produced during a given fMRI trial. They found decreases in activation concomitant with increasing

repetitions in the left posterior superior temporal gyrus (pSTg), as well as other cortical and subcortical areas. Combining this with a previous finding (not using fMRI-RS, Graves et al., 2007) that found increasing activity in the pSTg for increasing word frequency, the authors hypothesized that the pSTg processes phonological representations for words, an idea upheld by other literature (Gow, 2012)

Vaden et al. (2009) also used a parametric fMRI analysis with a speech perception task in which they manipulated the amount of phonological repetition within a word list. They compared blocks in which subjects listened to lists with low (e.g., JUG, KNIT), medium (e.g., CAB, CALF), and high phonological repetition (e.g., HIP HIP). They found that decreasing activity levels in the bilateral superior temporal sulcus correlated with increasing phonological repetition. This suggests, like the previous study, that the superior temporal cortex processes phonological representations.

An fMRI-RS study by Peeva and colleagues (2010) sought to identify the representations of speech processed in various neural regions during production. Most fMRI-RS analyses use traditional fMRI contrasts between experimental conditions or parametric estimations to quantify repetition suppression. This, however, limits the number of representations that can be studied in a given experiment. In contrast, the subjects in Peeva et al. read pairs of pseudowords that varied according to how often each type of speech representation – phoneme, syllable, or whole pseudoword – was repeated between the pairs. For instance, in the reordered condition, subjects alternated between 2 pseudowords that contained the same 2 syllables in a varied order – e.g., ZEKLO and KLOZE. In this condition, the authors expected to see fMRI-RS in

brain regions representing phonemes and syllables as these remained consistent between the two pseudowords. However, they did not expect to see fMRI-RS in brain regions with suprasyllabic representations as these varied between the pseudowords.

For each speech representation – phonemic, syllabic, suprasyllabic, and phonologically insensitive – the authors hypothesized a relative pattern of fMRI-RS across the speaking conditions. They found activity matching the hypothesized phonemic representation pattern in the left pSTg, supplementary motor area (SMA), pallidum, and superior lateral cerebellum. They found activity matching syllabic representations in the left ventral premotor cortex (vPMC), and activity matching supra-syllabic representations in the right superior lateral cerebellum. Taken together, these results demonstrated an fMRI-RS methodology – which studies in this dissertation will emulate – to identify multiple speech representations within a single paradigm.

While it is possible to identify many speech representations with this methodology, only a finite number of conditions can be presented in a single study, and therefore a finite number of speech representations can be studied. For instance, syllabic frame structures (see Chapter 1.3) did not have a definitive pattern of across-condition activity pattern in Peeva et al. A similar issue arises with speech representations of intermediate sizes. For instance, the Peeva et al. study looked for phonemic and syllabic representations. If a region represented consonant clusters – a component larger than a phoneme and smaller than a syllable – as a single unit, the region's activity would inaccurately match the syllabic representation pattern. The work in this

dissertation aims to elaborate on the methodology introduced in Peeva et al. to investigate further representations of speech.

1.2. Subsyllabic constituents

Subsyllabic constituents (SSCs) are phonemes or groups of phonemes forming syllabic structure units that are intermediate between syllables and phonemes. An SSC may be composed of a single phoneme or a string of phonemes, depending on the composition of the syllable. For example, the *nucleus* of the syllable usually consists of the vowel or vowels. (In English, sonorous consonants can be the syllable nucleus, although infrequently). The *onset* contains any consonants before the nucleus, and the *coda* contains any consonants after the nucleus. While not all syllables in English contain an onset or a coda, a nucleus is obligatory. These three elements can also be hierarchically combined for additional SSCs. A nucleus and a coda form a *rime*; a syllable can consist of an onset and rime. Similarly, an onset and a nucleus form the *body*; a syllable can consist of a body and a coda. In this text, the term SSC refers to onset, nucleus, and coda SSCs unless otherwise noted.

1.2.1. Phonological evidence for subsyllabic constituents

Phonotactic constraints describe a language's restrictions on phoneme sequences and may provide evidence for SSCs as a unit of speech representation. Sequences within SSCs – e.g., orderings of phonemes within onsets or nuclei – are strictly limited, but between SSC sequences are more flexible. For instance, /br/ is a legal (allowable) onset cluster in English, as in BRAINS, but /rb/ is not. This is a stringent rule, and it is

difficult for a native speaker of English to produce an /rb/ onset. In contrast, a speaker can easily produce novel between-SSC sequences – like the diphthong /ai/ with the coda /ŋ/ – even if this phoneme pair also does not co-occur within any English syllable. This suggests that SSCs are, at some stage or stages of speech processing, single units of representations that can be combined with each other to form new sequences.

There are two caveats to this. First, while the allowable phoneme orderings for SSCs vary between languages, a principle governing consonant cluster formation across languages has been proposed. The Sonority Sequencing Principle (SSP, Selkirk, 1984, Clements, 1990) hypothesizes that sonority – the loudness or resonance of a sound, with voiceless consonants being the least sonorant sounds in English and low vowels being the most sonorant – rises to the syllable nucleus and then falls. In other words, SSCs are arranged such that syllable-edge phonemes are the least sonorant sounds and the nucleus is the most sonorant. The SSP hypothesizes that for a given language, phonotactically legal sequences of sounds in a syllable are restricted by the allowable differences in sonority between adjacent sounds in specific syllabic positions. For example, in complex English onsets and codas, consonants at the syllable edge must have a lower sonority ranking than those that are more internal to the syllable. In contrast, Russian consonants in clusters may have the same sonority ranking (i.e. a plateau, as in the onset /zv/).

A second caveat is that in English, some between-SSC restrictions may exist. For example, each of the consonant clusters, /sk/, /st/, and /sp/, do not appear in a single word as both the onset and coda clusters (Fudge, 1969). However, this type of void may

be due to accidental gaps in the lexicon from historical artifacts. Within-rime restrictions are more common. In the example above, /ŋ/ does not appear before diphthongs or long vowels even though other nasals – with the same sonority ranking – do appear in this context (e.g. ‘sound’ or ‘point’, Fudge, 1969).These restrictions may be used to argue for the onset-rime SSC organization discussed below.

Another piece of evidence for SSCs as a unit of speech representation is that a phonotactically legal onset in a given language may not be a legal coda, even though it consists of the same phonemes in the same order. For instance, in English, /str/ is a legal onset, but an illegal coda. This suggests that complex onsets and codas are not only sequences of phonemes, but also units of representation in their own right.

1.2.2. Psycholinguistic evidence for subsyllabic constituents

Treiman and colleagues hypothesized that onset and rime SSCs are the primary units of speech in American English. Treiman and Danis (1988) found that elicited speech errors are most likely to cause breaks at the onset-rime boundary and maintain the content of the onset or rime. Similar findings with word games (Fowler, et al., 1993; Treiman et al., 1995) show that subjects respond significantly faster and more accurately when asked to replace the entire onset or rime than when asked to replace only part of those SSCs. For example, in the utterance /balf/, subjects are faster to replace the /b/ with another consonant (C) or the /alf/ with another vowel-consonant (VC), than to exchange the /bal/ with another CVC or /f/ with another C.

Recordings from electrodes positioned on the palate – electropalatography – can measure the timing and position the tongue's contact with the palate. This method can discriminate between individual articulatory gestures of a phoneme sequence and can measure the variation in the timing of these gestures across multiple productions. During the production of an illegal consonant cluster of consonants across a word boundary (e.g., /kp/ in "jackpot"), the percentage of overlap between the consonants increases with speaking rate. However, during the production of a legal consonant cluster, the total utterance duration may change, but the temporal relationship between the articulatory movements remains the same (Byrd, 1996; Byrd & Tan, 1996; Byrd & Choi, 2006; Loevenbruck et al., 1999).

A similar contrast exists between the consonants of a CVC pseudoword and an analogous CC consonant cluster (Loevenbruck et al., 1999). For example, as speaking rate increases, the gestures of the syllable /kɛl/ act as separate entities – with increasing articulatory overlap (otherwise known as coarticulation) between the gestures of each phoneme with increasing speaking rate – while the gestures of the consonant cluster /kl/ acts as one motor unit. This stability appears to be stronger for onset clusters than for coda clusters (Browman & Goldstein, 1988; Byrd, 1996; Byrd & Choi, 2006), but this may be due to study confounds such as the use of word-edge sonorant phonemes in the coda clusters (e.g., /kl/) or the use of codas that could be interpreted as multi-morphemic (e.g., /dz/). It is interesting that this evidence (particularly Byrd, 1996) has been used to argue that the entire syllable is a cohesive structure (Levelt et al., 1999). While these data do not discredit the idea of the syllable as a key unit of speech representation, they appear to support the theory of SSC representations more directly.

Evidence from language development also supports the theory that SSCs are cohesive units of speech representation. Children acquire new syllable structures in a highly predictable order (for a review, see McLeod et al., 2001). In English, CV syllables are acquired first, then CVC, V, and VC syllables (Levelt, et al., 1999). If children were limited only by the ability to produce longer words or if they simply acquired new syllable frame structures regardless of SSCs, one would predict that the acquisition of syllabic frames with onset and coda clusters could be interleaved (e.g., CCV, then VCC, then CCVC). But when children begin producing words with consonant clusters, there is a bifurcation in their development patterns (Levelt, et al., 1999). Some children acquire onset clusters first, and some acquire coda clusters first. If a child produces onset clusters first, they learn CCV words, then CCVC words, and then they learn words with coda clusters. If a child produces coda clusters first, they learn CVCC words, then VCC words, and then they learn words with onset clusters. In other words, children do not simply acquire new syllabic frame structures; they master the articulation of increasingly complex onset or coda clusters.

Shattuck-Hufnagel (1986) observed that in speech errors, some multi-phonemic nuclei – e.g., vowel-liquid pairs and diphthongs – act as a single unit. Fromkin (1971) made a similar observation about vowel-approximate pairs and diphthongs, and she hypothesized that these constitute complex vowels. Similarly, Shattuck-Hufnagel (1983) and Fromkin (1971) both note that consonant cluster errors often constitute a single error unit (and that seems to hold more strongly for onset clusters).

However, while there is evidence for the role SSCs as units of speech representation in English, the specific SSCs implicated may not apply cross-linguistically. In particular, in English, the onset-rime organization hypothesized by Treiman and colleagues may be a product of the bigram frequency of the phoneme elements; it is easier to remember and produce pairs or sequences of phonemes that occur together frequently in the lexicon of the language (Lee & Goldrick, 2008). It is also possible that the cohesion of the nucleus and coda within the rime may be the result of the notoriously variable phonology-to-orthography relationship in English. Nucleus pronunciation from orthography is more strongly predicted by coda context than by the onset context (Treiman et al., 2003). These ideas, however, do not exclude the theory that SSCs are still important units of representation in speech production.

1.3. Theories of syllabic frame and syllabic representations of speech

1.3.1. Slot/filler theory

Shattuck-Hufnagel and Klatt (1979) examined spontaneous substitution and exchange speech errors. Substitution errors occur when an intended target sound is mistakenly replaced by an intrusion sound: “first part” becomes “pirst part”. Exchange errors, also known as Spoonerisms, occur when two target sounds switch positions: “dear old queen” becomes “queer old dean.” The authors assembled a speech error corpus and catalogued occurrences of all single consonantal phonemes in each error role (the target phoneme – /f/ in the intended “first” – or the intrusion phoneme - /p/ in the resulting “pirst”). They determined that phonemes, not distinctive features, constitute the error

units of substitution and exchange errors. They also found, with few exceptions, that phonemes are equally likely to be the target or intrusion phoneme.

Based on these findings, Shattuck-Hufnagel and Klatt (1979) proposed the "slot/filler" theory of speech planning. They hypothesized that the phonological content of an intended utterance is selected separately from the structural "slots" that represent the metrical timing and the output order of the selected phonemes. After selection, the phonological "filler" elements are inserted into the appropriate slots for production. An additional mechanism monitors which filler elements have already been used and which have yet to be placed into slots.

Later, Shattuck-Hufnagel refined the slot/filler theory and redefined both the slots and the fillers. As previously mentioned, she found that complex nuclei and onset clusters often functioned as a single error unit in spontaneous speech errors (Shattuck-Hufnagel, 1986). This suggests that the filler elements may not be individual phonemes as previously described, but may instead be SSCs. Shattuck-Hufnagel also refined the definition of the structural slots; previously it was unclear if slots defined positions within syllables, mora, or words. Based on a set of elicited speech error tasks that controlled for word position, syllable position, and lexical stress, she concluded that slots encode the filler elements' positions within a syllable.

1.3.2. Frame/content theory

MacNeilage (1998) offered a similar proposal to the slot/filler theory based on a different set of evidence. He claimed that speech evolved from mastication, from the repetitive jaw movements of chewing. His "frame/content" theory proposes that an open-close jaw cycle provides a syllable "frame" structure for utterances. Speech requires filling a syllable frame with phonological content, thus dissociating the syllable's structure from the specific sounds to be produced. By this theory, the motor representations used to execute speech movements are syllables, cycles of open-close jaw movements. Note that despite their different origins, the basic principles of the slot/filler theory and the frame/content theory are nearly identical, save the terminology used to describe the slots/frame and filler/content.

Like Shattuck-Hufnagel, MacNeilage also used speech errors to argue for the existence of separate frames and content. He also noted that the phonemes involved in exchange errors usually appear in the same position of the incorrect syllables, implying that the frames were correctly selected, and the content was inserted into the correct frame position, but in the incorrect syllable. Similarly, phonemes almost never exchange within a single syllable since the content cannot be inserted into a different position within the frame.

MacNeilage also argues that language development provides evidence for the frame/content theory. He hypothesizes that CV is the first syllable acquired and the only syllable that exists in all known languages because the close-open jaw movement is the basis of speech. Moreover, he noted that during the "reduplicated babbling" stage of

language development – in which children repeat CV syllables such as BA.BA.BA – cross-linguistically, syllables most often contain labial stop consonants and central vowels (MacNeilage & Davis, 2000). MacNeilage dubs these utterances “pure frames” and hypothesizes that they represent the acquisition of a CV frame, produced without regard to phonological content. Labial stop consonants result from a full jaw closure, and central vowels result from the jaw opening with the tongue in a relaxed position. The subsequent “variegated babbling” stage – in which children produce various CV syllables such as BA.DI.GU – represents the insertion of varying phonological content into the acquired frame. Thus, MacNeilage argues, children first learn syllabic frames and then learn to insert phonological content.

1.3.3. Comparison of slot/filler and frame/content theories

There are many similarities between the slot/filler theory and the frame/content theory. They both posit that phonological content is selected separately from the syllabic frames representing structure and timing and then combined at a later planning stage. They also both hypothesize that frames are syllable-sized. However, it should be noted that the slot/filler theory is a more nuanced proposal than the frame/content theory. Shattuck-Hufnagel has documented other influences on speech errors beyond syllabic frame position. For example, palatalization errors – /s/ mistakenly replaced by /š/ – are significantly more common than the opposite trend (Shattuck-Hufnagel, 1979). Vowel interactions in speech errors are strongly affected by the distinctive features of the items (Shattuck-Hufnagel, 1986). She also found that lexical stress affects speech errors (Shattuck-Hufnagel 1992). As previously mentioned, when two phonemes are part of

different syllables but in the same position within those syllables, they are more likely to interact in a speech error than if they were in different syllable positions. If those syllables have the same lexical stress, they are even more likely to interact. These additional effects on speech errors, however, do not detract from the slot/filler-frame/content theory. They may reflect the influences of other stages of speech production.

1.3.4. Syllabic representations of speech

Several computational models of speech have implemented the syllabic representations of speech proposed in the slot/filler and frame/content theories. In the spreading activation model, (Dell, 1986) syllables are not the units of motor output. They are, however, used during an intermediate step of phonological encoding. Once the morpheme or morphemes of an utterance have been chosen, the utterance is syllabified, activating the nodes corresponding to each syllable. The syllable nodes in turn activate subsyllabic constituent nodes. These activate phoneme nodes that activate the corresponding feature nodes – encoding place and manner – for motor output.

The oscillator-based associative recall (OSCAR) model (Vousden, et al., 2000) also uses syllables as an intermediate representation. Target words are first syllabified. Then the model uses oscillators to encode the syllabic positions of each phoneme of the syllable. Phonemes themselves are encoded as feature vectors.

Goldstein and colleagues' coupled oscillatory model of speech production uses syllables as the basic units of production (Goldstein et al., 2009; Nam et al., 2009). The authors propose that the timing of articulatory gestures, such as lip or tongue tip closures, are based on planning oscillators (like the close-open jaw cycles of the frame/content theory). The GO signal – to initiate the execution of a gesture – is triggered at a particular phase (usually 0) of that oscillation. Gestures synchronize based on a common phase relationship between their oscillators, with the syllable providing the timing basis.

The WEAVER model (Word-form encoding by activation and verification, Roelofs, 1997; Levelt, 1999; Levelt, et al., 1999) also uses the syllable as the primary unit of speech production. In the model, once all morphemes of a lexical item are concatenated and syllabified, the relevant phonetic "syllable scores" are activated. Similarly, both the DIVA model (most recently, Tourville & Guenther, 2011) and its syllable sequencing extension, GODIVA (Bohland, et al., 2010) – both of which will be discussed in more detail – use syllables as the primary unit of production. These last three models also use a "mental syllabary": WEAVER in the phonetic encoding stage, DIVA and GODIVA as the "speech sound map".

A mental syllabary stores the articulatory output codes for highly practiced syllables. Levelt and colleagues' (1999) argument for the syllabary is twofold. First, they suggest that there is greater coarticulation and "gestural dependence" within syllables compared to between them. Thus, storing syllable gestures saves the time and effort of assembling and interpolating between gestures of the syllable. Second, they suggest

that while the English language may contain a seemingly infinite number of possible syllables, speakers frequently use only a few hundred syllables. A mental syllabary takes advantage of this computational windfall and only stores the motor programs for these most frequent syllables. (Note that infrequent or new syllables that do not already exist in the syllabary must be composed from subsyllabic units.) Further support for the syllabary comes from evidence that high-frequency syllables – whose motor programs would be stored in the mental syllabary – are produced faster and with greater acoustic measures of coarticulation and greater acoustic consistency between productions than low-frequency syllables – that presumably require the concatenation of smaller motor units for production (Herrmann et al., 2008).

1.4. DIVA and GODIVA models

The Directions into Velocities of Articulators (DIVA) model of speech production is a biologically plausible, computational model of speech motor control (Figure 1.1, Guenther, 1994, 2006; Guenther, et al., 2006; Golfinopoulos, et al., 2011). Because each of its processing stages is also assigned a neural correlate, this and similar computational models are powerful tools as they provide theories of brain organization against which neuroimaging data can be tested.

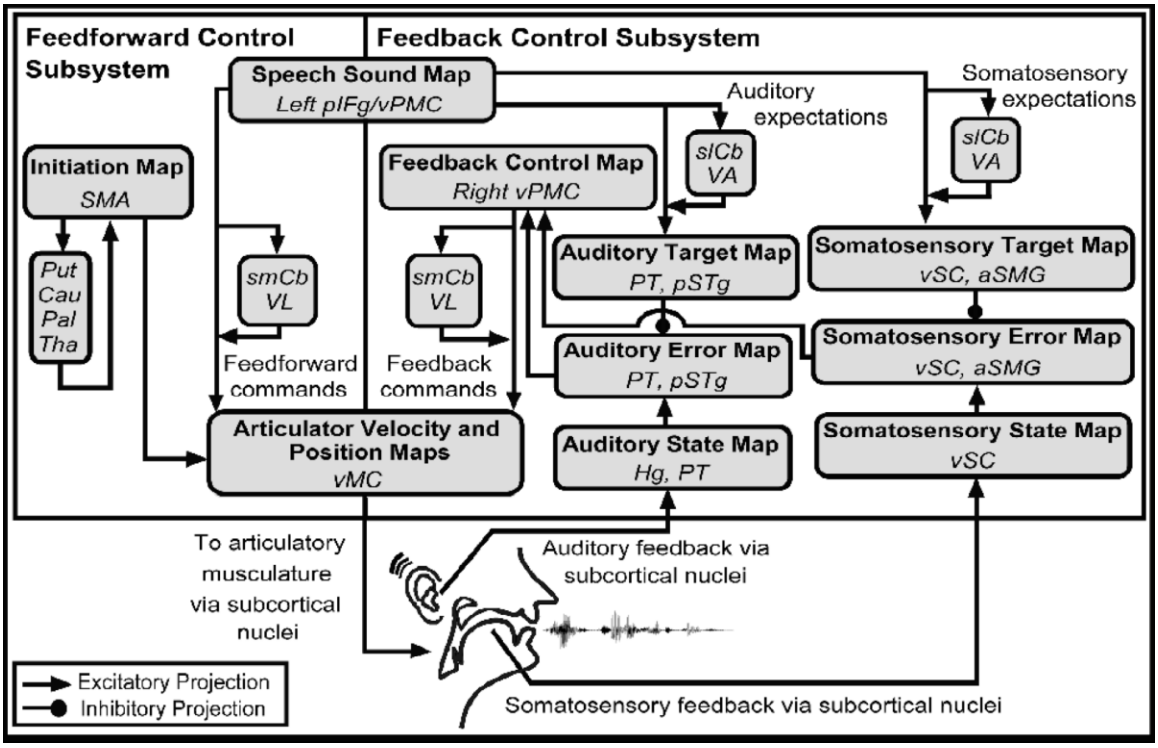


Figure 1.1. A schematic of the DIVA model (from Golfinopoulos, et al., 2010) Abbreviations: pIFg = posterior inferior frontal gyrus, vPMC = ventral premotor cortex, SMA = supplementary motor area, Put = putatmen, Cau = caudate, pal = Pallidum, Tha = thalamus, smCb = superior medial cerebellum,VL = ventral lateral nucleus of the thalamus, alCb = anterior lateral cerebellum, VA = ventral anterior nucleus of the thalamus, PT = planum temporale, pSTg = posterior superior temporal gyrus, Hg = Heschl's gyrus, vSC = Ventral somatosensory cortex, aSMG = anterior supramarginal gyrus, vMC = ventral motor cortex.

DIVA produces syllable-sized motor commands controlling articulator positions to achieve auditory speech sound targets. These feedforward commands each represent the movements to produce a syllable and are hypothesized to arise from the Speech Sound Map (SSM) in the left inferior frontal gyrus and ventral premotor cortex. The SSM projects to Articulator and Velocity Position Maps (modulated by a cortico-cerebellar

loop) that control the individual articulators and are hypothesized to lie in the ventral portion of the primary motor cortex.

Another important aspect of the DIVA model is its use of sensory feedback to tune the feedforward commands. In order to achieve the auditory targets for an utterance, the model monitors auditory and somatosensory input during speech and compares them to the expected sensory input. The expected sensory targets are stored in the Auditory Target Map, hypothesized to lie in the planum temporale (PT) and posterior superior temporal gyrus (pSTG), and the Somatosensory Target Map, in the ventral portion of the primary somatosensory cortex (vSC) and the anterior supramarginal gyrus (aSMG). Incoming sensory information is tracked in the Auditory State Map, in Heschl's gyrus and the PT, and the Somatosensory State Map, in vSC. Finally, the difference between the Target and State Maps are calculated in the Auditory Error map, in the PT and pSTG, and in the Somatosensory Error Map, in the vSC and aSMG. If any discrepancies are found, they are sent to the Feedback Control Map, in the right ventral premotor cortex, that transforms the error signal into corrective movements in the Articulator and Velocity Position Maps.

The Gradient Order DIVA (GODIVA) model extends the DIVA model to account for the planning of syllable sequences (Figure 1.2, Bohland, et al., 2010). Given a multi-syllabic target, the model organizes the selection and timing of the appropriate syllable-sized motor outputs. Like the DIVA model, each module of the GODIVA model has a hypothesized neural correlate, allowing researchers to test predictions about expected neural activity for a given speech sequencing task.

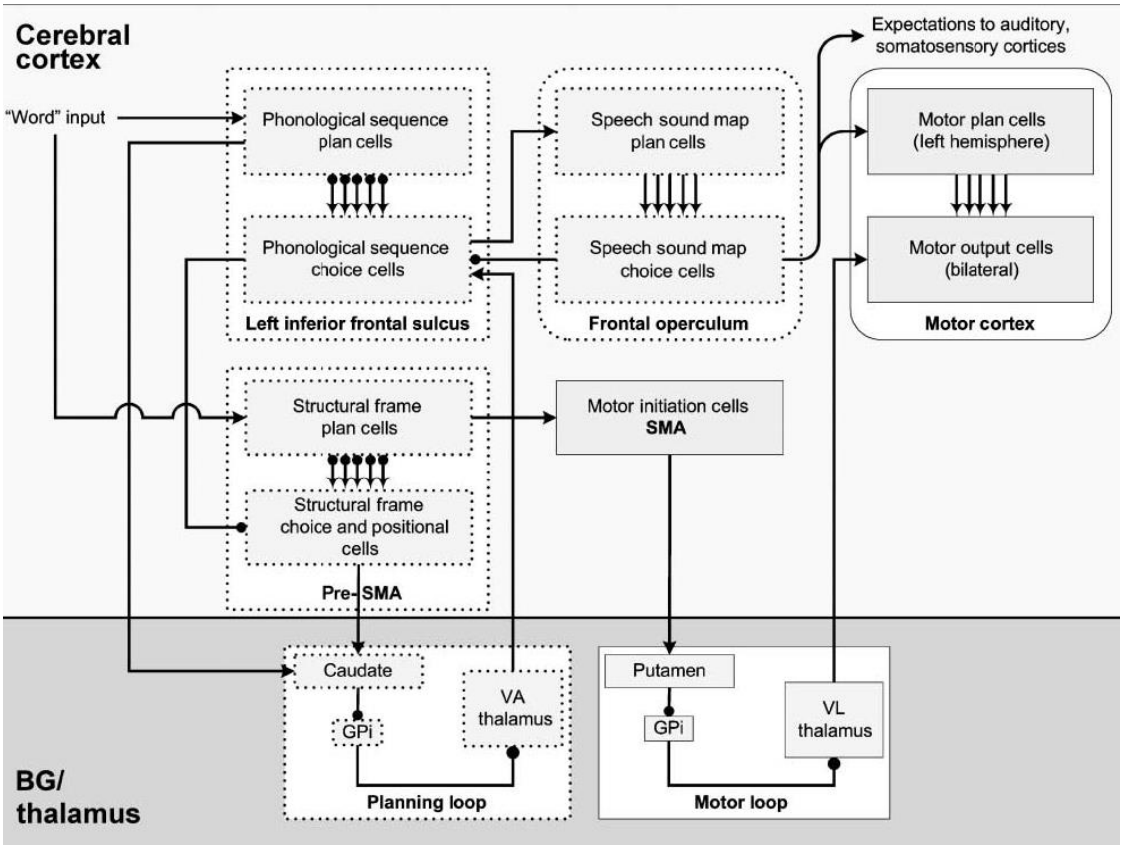


Figure 1.2. Schematic of the GODIVA model of speech sequencing. Abbreviations: pre-SMA = presupplementary motor area, GPi = internal segment of the globus pallidus, VA thalamus = ventral anterior nucleus of the thalamus, VL thalamus = ventral lateral nucleus of the thalamus.

The model incorporates several key ideas from speech motor research. First, the model uses competitive queuing (Bullock & Rhodes, 2003; Houghton, 1990) to produce a series of syllables in the intended order while using a biologically plausible representation (Averbeck et al., 2003). Competitive queuing models represent potential plans, each with their own level of activation. Using a winner-take-all method, the plan with the highest level of activation at a given time is chosen to be executed, and at

completion, its activity is suppressed to make way for the node with the next highest level of activation to be executed next.

In the GODIVA model, competitive queuing is implemented such that each region is represented by two layers, a planning layer and a choice layer. In a given layer, each potential item – e.g., all possible syllabic frame structures for an area selecting frames – is represented by a node. The node with the highest activation in the planning layer is chosen to be activated in the choice layer. Then the activity of this node is suppressed in the planning layer. To choose the next item, the planning node with the next highest activation is chosen by the choice layer, and so on.

GODIVA also implements the slot/filler frame/content theories. The GODIVA model receives the phonological content and syllabic frame input for an intended utterance in parallel. The phonological content representations are hypothesized to reside in the inferior frontal sulcus; using competitive queuing, phonemes for each syllable position are selected in the order in which they appear in a syllable sequence. Similarly, syllabic frame representations are hypothesized to reside in the presupplementary motor area, and frames for each syllable in a syllable sequence are selecting in the order of output. Once phonological content and frames are chosen (using a cortico-basal ganglia loop), the speech sound map – hypothesized to reside in the inferior frontal gyrus and ventral premotor cortex – generates the corresponding coarticulated syllable-sized motor program for a given syllable. The supplementary motor area (with another cortico-basal ganglia loop) regulates the movement initiation, and the primary motor cortex controls execution.

1.5. Organization of dissertation

The rest of this work is organized into 5 chapters; 4 chapters detail experimental work including literature reviews, and a final chapter summarizes the results with suggestions for further research. Chapter 2 describes a study of speech motor sequence learning and the motor units used in this task. Comparing behavioral measurements of utterances of sequences with novel consonant clusters (e.g. GVAZF) before and after practice provides evidence of speech motor sequence learning. Behavioral comparisons between novel sequences with and without the previously practiced consonant clusters suggest that consonant clusters were the motor units created during speech motor sequence learning.

Chapter 3 describes a study of the neural correlates of speech sequence learning. FMRI contrasts reveal increased activations for the production of novel sequences over previously learned sequences in areas of the brain implicated in the non-speech motor sequence learning literature. Increased activation for novel sequences also occurred in areas of the brain associated with speech learning based on sensory feedback. The importance of this latter set of regions to speech motor sequence learning is reinforced by correlations of behavioral measures of learning success with structural brain measures and with functional brain activity differences.

Chapter 4 details a neuroimaging study using fMRI-RS to seek neural representations of syllabic frame structure and phonological content in different parts of the brain. The

study also attempts to examine the nature of that phonological content, contrasting neural regions of phonemic and SSC representations.

Chapter 5 presents a neuroimaging study that attempts to use an fMRI-RS paradigm to illuminate neural representations of SSC separately from phonemic and syllabic representations. The results suggest that the different regions of auditory cortex process different representations of speech. However, a difficulty confound calls these results into question.

Finally, Chapter 6 summarizes the findings of the studies presented in this dissertation. Further research is suggested to build a more complete understanding of the representations used in the brain for speech production.

2. SUBSYLLABIC SPEECH MOTOR SEQUENCE LEARNING: BEHAVIORAL EVIDENCE OF LEARNING AND UNITS OF REPRESENTATION

2.1. Introduction

Motor sequence learning not only requires the acquisition of many movements, but also requires maintaining a precise serial order to those movements. In order to learn the serial order, it has been proposed that motor sequence learning occurs through hierarchical structuring of items in memory.

2.1.1. Models of motor sequence learning and performance

Lashley (1951) first argued against associative chain models of sequence memory and performance in which each item in a sequence is remembered through a pointer from the previous item. Lashley reasoned that exchange errors or Spoonerisms – the production of “queer old dean” instead of the intended “dear old queen” – would be impossible to explain with this model of memory. The exchanged segments are not linearly adjacent, and the speaker should not have access to any items other than those directly connected by the linear chaining. Another argument against associative chain models is that in well learned motor sequences, such as playing the piano, execution occurs very quickly. If an item must wait to be activated after the previous item is retrieved or executed, the sequence could probably not be performed as quickly. Lashley concluded that a non-linear structure is used in memory to store and retrieve sequences.

In his landmark paper, Miller (1956) proposed that this non-linear architecture is hierarchical. He suggested that in memory, adjacent items of a sequence are grouped together in “chunks” that can then be recalled a single item. He also suggested that human short-term memory capacity is limited by the number of chunks one can remember, but not the content within a chunk. Thus, a learned sequence can be stored using a hierarchical structure, as a sequence of chunks, each of which could be unpacked for retrieval. While some of the details of his proposal have been altered to accommodate new evidence (Cowan, 2000), the essence of this account remains in the current learning and memory literature.

While the proposals of both Lashley and Miller focused on the structure of short-term or working memory, the concept of chunking carries over to motor sequence learning and performance as well. Sakai and colleagues (2003) conducted a study using a 2x10 task in which subjects learn, by trial and error, the order in which to push two lit buttons on a 4-by-4 grid. These two movements make up a “set,” and 10 sets form a “hyperset.” In this task, each subject exhibits an individual pattern of chunking as demonstrated by certain sets with a long reaction time to the first movement and some sets with shorter reaction times. The onsets of sets with longer reaction times correspond to “chunk points”; the first element of a chunk requires more time to unpack the items within the chunk from memory. Shorter reaction times correspond to movements that have already been unpacked and need only be executed.

After learning, the authors presented the subjects with a hyperset using the same sets in a scrambled order. When this new hyperset maintained the adjacency and order of sets

within a chunk, the subjects used the same chunking points that they created during learning. (Since each subject had a unique set of learned chunking points, each of these new patterns was tailored to a subject's past performance.) However, when sets were shuffled so that items within a learned chunk were no longer adjacent, thus destroying the chunks, the subjects' productions were slower and had more errors. Thus, chunking can create a cohesive grouping of several adjacent movements, independent from other parts of the sequence.

2.1.2. Speech motor sequence learning

While there is a large literature on motor sequence learning of finger and arm movements, the literature on motor sequence learning of speech movements is sparse. Behavioral measures of learning have been shown in adults during practice of multisyllabic pseudowords: decreases in articulator movement duration and amplitude, and increases in accuracy (Namasivayam & van Lieshout, 2008; Rauschecker et al., 2008; Schultz, 2001). Decreased reaction times have been shown for object learning tasks, where subjects practice producing novel pseudowords that are associated with an new object (Breitenstein & Knecht, 2002; Cornelissen et al., 2004). Smits-Bandstra and colleagues (2006) not only demonstrated decreases in reaction time and utterance duration during practice of multisyllabic pseudowords, but also showed that both the learning curves and the retention time courses were parallel to those of a non-speech motor sequence learning task.

2.1.3. Present study

Here, we study speech motor sequence learning of phoneme sequences with novel consonant clusters (e.g., GVAZF). Our first goal was to demonstrate speech motor sequence learning with behavioral measures of speaker performance.

Our second goal was to understand the motor chunks used to represent the sequences. In non-speech motor sequence learning of random sequences, chunks consist of any adjacent items of the movement sequence, and this grouping is individual to each subject (Sakai et al., 2003). However, as previously discussed, phonotactic constraints limit allowable speech sequences. We hypothesize that speech differs from analogous motor sequencing tasks because speech chunks are shaped by syllabic structure. To test this, subjects produced novel sequences that retained subsyllabic constituents (SSCs) from previously learned sequences. We looked for behavioral advantages conferred by learned speech chunks, acquired during the practice of previously learned sequences.

2.2. Methods

2.2.1. Participants

8 right-handed speakers of American English (3 females) with normal or corrected-to-normal vision participated in this study. The subjects had no neurological conditions, hearing deficits, or speech perception or production deficits. Subjects were native speakers of American English and had no previous experience with any of the

languages used in stimuli creation. Informed consent was obtained according to the Boston University Institutional Review Board

2.2.2. Speech stimuli

Subjects produced single syllable pseudowords with a CCVCC syllable frame structure. In previous speech motor sequence learning work, subjects learned easy-to-produce syllable sequences such as FRACKERISTER (Rauschecker et al., 2008) or BAPABAPATA (Smits-Bandstra and De Nil, 2009). In contrast, we chose sequences with illegal consonant clusters to challenge the speech motor sequence learning system; not only is the whole sequence novel, but the subsequences – the consonant clusters – are also novel. Additionally, these sequences allowed us to probe the use of subsyllabic constituents (SSCs) as motor chunks.

In the *legal* condition, the biconsonantal onset and coda clusters occur readily in English; in the *illegal* condition, the clusters do not readily occur in English, but do occur in some other natural language. In order to rule out the possibility that subjects were learning an underlying rule or pattern of distinctive features (such as the allowable interval on the sonority scale), illegal consonant clusters were taken from a variety of language and language families including Hebrew, Leti and Taba, Romani, Polish, Lithuanian, Romanian, Georgian, Tepehua, Hungarian, and Pima. None of the pseudowords were an orthographic or a phonological word based on the MRC Psycholinguistic Database (Colheart, 1981). All illegal pseudowords had a neighborhood size of 0; no words could be created by adding, deleting, or substituting a single phoneme in any sequence.

In rapid speech in American English, a sonorant consonant can become the nucleus of its own syllable. For example, in rapid speech, the word LISTEN could be pronounced /lI.sn/ in which the /n/ is the nucleus of the second syllable. In order to prevent this type of syllabification and to maintain the integrity of the intended syllable structure, consonant clusters in the illegal pseudowords contain only non-sonorant consonants: stops, fricatives, and affricates. Vowels were limited to /I/, /æ/, /ɛ/, and /Λ/ in order to cover the entirety of the vowel space. Instances of each vowel were distributed equally within each condition and across all items.

Auditory stimuli were recorded with 32 bit sound at 4.41 kHz over a Samson C01U USB studio condenser microphone using Audacity software (http://audacity.sourceforge.net/). The speaker was a native speaker of American English who had previously practiced the sequences. Using the same software, selected pseudowords were normalized for intensity. Stimuli were chosen to maintain similar F0 across all sequences. The durations were adjusted to a constant length using PRAAT software that changes duration without changing F0 (Boersma & Weenink, 2007). Maximally, changes were < 10% of the total original length. Each auditory stimulus lasted 480 ms.

2.2.3. Experimental Paradigm

In the practice phase, subjects repeated a set of 4 *learned illegal* sequences and 4 *learned legal* sequences over 60 trials per pseudoword on two separate days (see Table 2.1 for a list of stimuli). The practice occurred over two consecutive days in order to

allow for overnight memory consolidation (Brashers-Krug et al., 1996; Davis, et al., 2009; Fenn et al., 2003; Stickgold, 2005). They performed 50 productions of each pseudoword on Day 1, and 10 productions on Day 2. Practice sessions occurred at least one and not more than two days prior to scanning to allow for memory consolidation (Brashers-Krug et al., 1996; Fenn et al., 2003; Stickgold, 2005; Davis, et al., 2009). Each syllable was produced 30 times per practice session. Syllables were presented in pseudorandom order.

Set 1	BLERK
	FLISK
	GRALVE
	PRUNGE
Set 2	DRALF
	FREMP
	PLIRTH
	TRULP

Set			**Set**
A	zdechb	bdechk	C
	shkizg	zkizf	
	fshapf	kshapk	
	gvusb	zvusch	
B	fshizg	kshizf	D
	gvuchb	zvuchk	
	shkepf	zkepk	
	zdasb	bdasch	
E	dzukf	vzukp	G
	tfeshch	gfeshp	
	shgatk	tgatp	
	kpimch	fpimsh	
F	shgekf	tgekp	H
	kpashch	fpashp	
	dzitk	vzitp	
	tfumch	gfumsh	

Table 2.1. *Left*: The legal stimuli. Half the subjects learned Set 1 of the legal tokens, and half learned Set 2. *Right:* The illegal stimuli. Each set of tokens was learned in the practice phase by one of the eight subjects. Sets to the left and right of each other constitute each other's *middle* tokens. Sets within the same quadrant (bounded by grey) constitute each other's *SSC* tokens. Sets from a diagonal quadrant were used for the *unrelated* condition.

Then on Day 2 in the testing phase, to compare the hierarchical learning of the learned pseudowords, subjects produced the *learned illegal* pseudowords as well as three types of novel illegal CCVCC-structured pseudowords using the same trial time course as in practice (Table 2.1, right). *Unrelated* pseudowords are composed of novel (illegal) onset

and coda clusters as well as novel middle CVC combinations. *Subsyllabic constituent* pseudowords (SSC) are composed of the onset and coda clusters from two different *learned illegal* pseudowords as well as a nucleus not associated with either cluster in those *learned illegal* pseudowords. *Middle* pseudowords are composed of the adjacent CVC elements from one *illegal learned* pseudoword, with consonant clusters that did not occur in any *learned illegal* pseudowords.

In order to balance inter-pseudoword differences in difficulty, 32 illegal pseudowords were used with equal frequency across subjects and conditions (Table, 2.1, right). The illegal stimuli were divided up into 8 sets of 4 pseudowords each (groups A-F). Each of the 8 subjects practiced a different stimulus set for the *learned illegal* condition; thus different sets of pseudowords constituted the *SSC*, *middle*, and *unrelated* conditions for that subject. The 8 *legal* pseudowords (Table 2.1, *left*) were also divided between the subjects so they practiced a different but balanced set of *learned legal* pseudowords.

During a single trial, the subject was presented with the orthographic representation of the pseudoword for 1450 ms. Then, 500 ms after the onset of this presentation, subjects heard the 480 ms auditory stimulus. The combination of the orthographic and auditory presentations is necessary as listeners have been shown to hear monosyllable illegal consonant clusters as epenthesized disyllabic words (Berent et al., 2007). For example, monosyllabic /lbɪf/ was perceptually identical to disyllabic /lə.bɪf/ to native speakers of English for whom /lb/ is an illegal onset cluster. Then the stimulus was removed and replaced by a fixation cross that remained on for the rest of the trial. After a random jittered pause of 500-1000 ms, a 50 ms tone acted as a GO signal for the subject to

repeat the pseudoword. Then, utterances were recorded for 1500 ms. Pseudowords were randomized across trials.

Before the practice session, subjects were first briefly introduced to the paradigm with 10 trials of two legal and two illegal pseudowords not used at any point in the rest of the study. They were given feedback on production and instructed to replicate the auditory stimulus, while making sure to produce all the sounds seen in the orthographic presentation. Subjects were also instructed to attempt to eliminate any schwa epenthesis, a common response when producing novel illegal consonant clusters. They were also asked to produce the pseudowords as quickly and accurately as possible.

2.2.4. Data analysis

To evaluate speech motor sequence learning, we measured changes in the following three *learning success indices* over the practice sessions: (i) error rate, (ii) reaction time and (iii) utterance duration. Error rate calculations were based on the percentage of trials with one or more errors, in which an error was defined as an utterance omission, repetition, or restart, or a phoneme addition, deletion, or substitution. Only the first 5 productions of each pseudoword during each practice phase day were used in the analysis to avoid confounds from subject fatigue over the course of the practice session. A single rater (JS) judged errors, reaction times, and utterance durations for all trials and was naïve to the condition of experiment conditions when possible[2].

[2] Due to the nature of the stimuli, the rater was not naïve to the experimental condition during the practice phase but was naïve to the experimental condition of the illegal stimuli during the testing phase.

Duration and reaction time measurements were based on the first 5 error-free trials of each pseudoword during each practice session day. Utterance onset and offset were automatically labeled based on sound pressure level thresholds individually chosen for each practice session, then hand-checked. These measures were used to calculate reaction time (time from GO signal to utterance onset) and utterance duration (time from onset to offset of the utterance).

To assess learning-related changes due to practice, we compared the error rate and accuracy changes from the first practice phase day to the second day with paired t-tests. Each behavioral measure was averaged within each condition and within each subject. We hypothesized that we would see greater learning in the illegal condition because those syllables included both novel pseudowords and novel consonant clusters whereas the legal condition includes novel pseudowords of familiar consonant clusters. Paired t-tests comparing the mean error rate and duration in the illegal and legal conditions were performed to test this hypothesis. T-tests were corrected for multiple comparisons using a false discovery rate threshold of < 0.05.

Similarly, to assess differences in production of learned and novel pseudowords during the testing phase, utterances were compared across the first five correct utterances for utterance duration and the first five utterances for error rate across the *learned*, *SSC*, *middle*, and *unrelated* conditions. One-way ANOVAs ($p_{FWE} < 0.05$, Bonferroni corrected) were employed to test for significant differences across the four conditions for each

behavioral measure. Bonferroni corrected paired t-tests ($p_{FWE} < 0.05$) were then used to test for behavioral differences between each pair of conditions.

2.3. Results

2.3.1. Practice phase

With practice, subjects showed behavioral improvements for illegal pseudowords, but not for legal pseudowords (Figure 2.1). Both utterance duration (mean change = 10.5%, s.d. = 8.8; $t_{(7)} = 7.12$, $p < 0.0001$) and percent error (mean change = 47.1%, s.d. = 24.8; $t_{(7)} = 11.23$, $p < 0.0001$) of illegal pseudowords decreased significantly from the first five trials to the last five trials of practice for illegal pseudowords. Neither measure significantly changed in legal pseudowords (duration: mean change = 1.1%, s.d. = 4.0; $t_{(7)} = -1.69$, $p = 0.1$, n.s; error: mean change = 7.1%, s.d. = 14.3; $t_{(7)} = 2.95$, $p = 0.055$,n.s.). The change in utterance duration (mean change = 37.0%, s.d. = 25.6; $t_{(7)} = 8.59$, $p < 0.0001$) and percent errors (mean change = 40.0%, s.d. = 25.5; $t_{(7)} = 9.41$, $p = 0. < 0.0001$) was also significantly greater for illegal pseudowords than legal pseudowords.

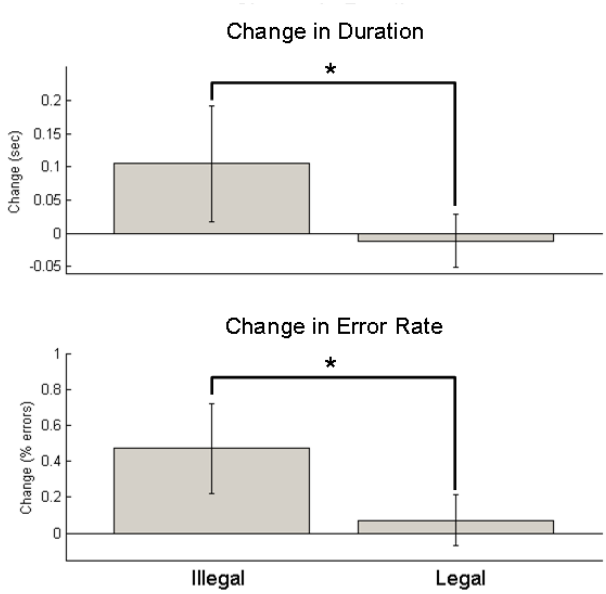


Figure 2.1 Comparison of first minus the last five trials of practice based on change in utterance duration (in milliseconds) and percentage of trials with at least one error.

Figure 2.2 and Figure 2.3 show the time course of these measures averaged across subjects and pseudowords over the 60 practice trials per pseudoword (50 per pseudoword on Day 1, 10 on Day 2).

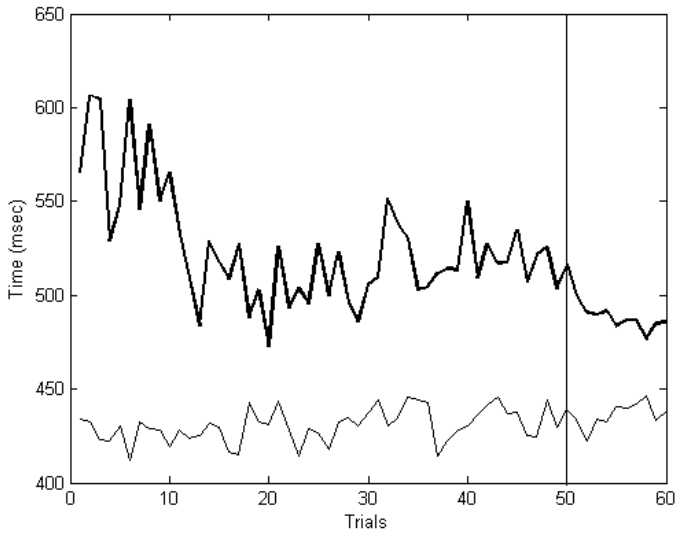


Figure 2.2. Time course of average utterance durations over the practice trials. The thick black line shows the results for learned illegal tokens. The thin grey line shows the results for learned legal tokens. The vertical line indicates the beginning of the second day of testing.

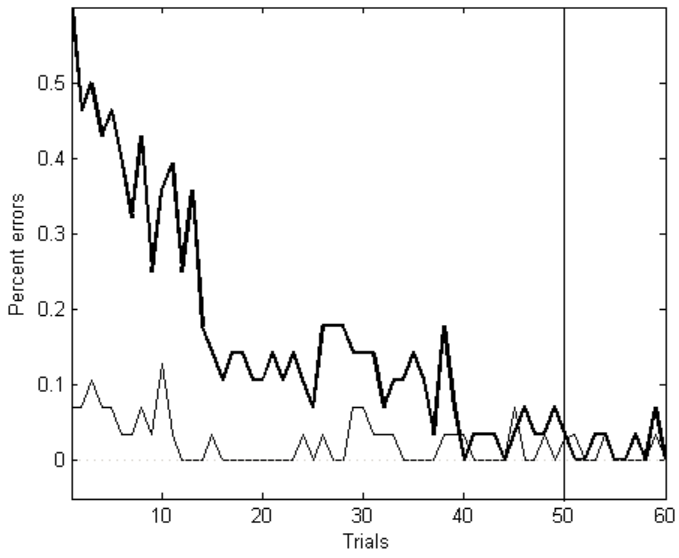


Figure 2.3. Time course of average percent of errors over the practice trials. The thick black line shows the results for learned illegal tokens. The thin grey line shows the results for learned legal tokens. The vertical line indicates the beginning of the second day of testing.

2.3.2. Testing phase

Significant differences in learning success indices across the four speaking conditions during the testing phase were noted (Figure 2.4). One-way ANOVAs tested for significant differences in learning indices across conditions; *post hoc* paired t-tests compared pairs of conditions. Utterance durations were significantly different between conditions ($F_{(3,4)} = 7.41$, $p < 0.0001$). The utterance duration for learned pseudowords was significantly shorter than both the *middle* (mean = 11.8%, s.d. = 10.9; $t_{(7)} = 6.42$, $p < 0.0001$) and *unrelated* (mean = 11.68%, s.d. = 10.8; $t_{(7)} = 6.40$, $p < 0.0001$) pseudowords in a paired t-test, corrected for multiple comparisons (Figure 2.4). Similarly, the utterance duration for *SSC* pseudowords was significantly shorter than

both the *middle* (mean = 10.5%, s.d. = 12.5; $t_{(7)} = 4.93$, $p < 0.0001$) and *unrelated* (mean = 10.3%, s.d. = 12.4; $t_{(7)} = 4.92$, $p < 0.0001$) pseudowords (Figure 2.4, top). The utterance durations for *learned* and *SSC* utterance durations were not significantly different from each other (mean = 1.38%, s.d. = 3.51; $t_{(7)} = 2.33$, $p = 0.03$, n.s.). The *middle* and *unrelated* durations were also not significantly different from each other (mean = 0.001%, s.d. = 0.12; $t_{(7)} = 0.06$, $p = 0.96$, n.s.).

The same pattern, with the same significance levels occurred in the error rates (Figure 2.4, bottom). Error rates were significantly different between conditions ($F_{(3,4)} = 28.47$, $p < 0.0001$). The error rate for learned pseudowords was significantly lower than both the *middle* (mean = 49.29%, s.d. = 41.78; $t_{(7)} = 6.98$, $p < 0.0001$) and *unrelated* (mean = 47.14%, s.d. = 34.71; $t_{(7)} = 8.04$, $p < 0.0001$) pseudowords in a paired t-test, corrected for multiple comparisons (Figure 2.4). Similarly, the error rate for *SSC* pseudowords was significantly lower than both the *middle* (mean = 39.78%, s.d. = 38.25; $t_{(7)} = 6.15$, $p < 0.0001$) and *unrelated* (mean = 37.62%, s.d. = 31.85; $t_{(7)} = 6.99$, $p < 0.0001$) pseudowords (Figure 2.4, top). The error rates for *learned* and *SSC* error rates were not significantly different from each other (mean = 9.52%, s.d. = 17.52; $t_{(7)} = 2.21$, $p = 0.04$, n.s.). The *middle* and *unrelated* error rates were also not significantly different from each other (mean = 2.14%, s.d. = 45.12; $t_{(7)} = 0.28$, $p = 0.78$, n.s.).

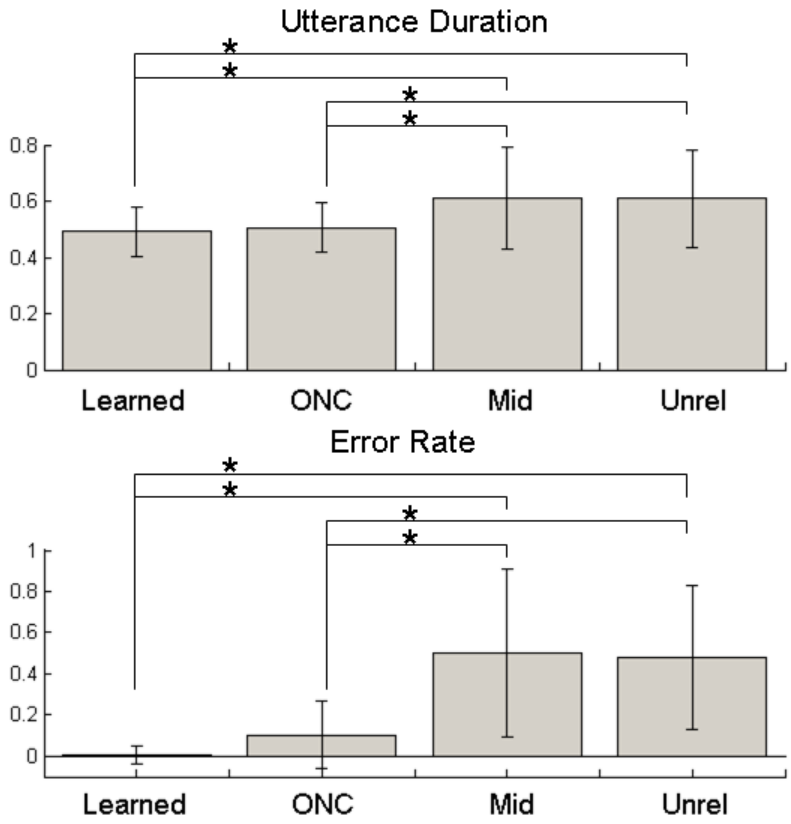


Figure 2.4. Comparison of the first five trials of testing across the learned, SSC (here, labeled "ONC"), mid and unrelated conditions. Asterisks indicate significance of $p > 0.001$ in a paired t-test, corrected for multiple comparisons. Non-asterisked pairs were not significant ($p > 0.05$, n.s.).

2.4. Discussion

We explored subsyllabic speech motor sequence learning in healthy adults. Over two days, subjects practiced novel speech sequences: pseudowords with legal (e.g., BLERK) or illegal (e.g., ZDECHB) consonant clusters. With practice, illegal pseudowords were produced more accurately and with fewer errors, indicating motor

sequence learning[3] occurred. Legal pseudowords saw no significant behavioral gains. Subjects were already at ceiling performance for the production of legal pseudowords; these contained familiar consonant clusters, and few behavioral gains could be made. In contrast, illegal pseudowords contained novel subsyllabic constituents (SSCs); speech motor sequence learning of the illegal consonant clusters allowed subjects to perform illegal pseudowords faster and more accurately with repeated practice.

This interpretation is also supported by direct exploration of the representations of the motor chunks used during subsyllabic speech motor sequence learning. After practice, subjects produced novel illegal pseudowords as fast and accurately as the *learned illegal* pseudowords only if they contained previously learned illegal consonant clusters. This suggests a hierarchical structure of learned speech sequences that follows syllable structure. These results further implicate SSCs – namely onset and coda consonant clusters – as the motor chunks used to learn these speech motor sequences.

These findings bring to light gaps in current models of speech production. Models such as WEAVER++ and DIVA and GODIVA (see Chapter 1.4 for more details on the DIVA and GODIVA models), use a mental syllabary, a repository for syllabic motor programs of frequently produced syllables (see Chapter 1.3.4 for more detail on the syllabary). While these qualitative descriptions of these models suggest alternative production processes for novel or low-frequency syllable production, neither computationally

[3] We are labeling the observed gains that come from repeated practice of the speech sequence as speech motor sequence learning. However, other kinds of learning might aid this process like phonological learning (see Section 3.4.4).

implements these processes. These results highlight the need for testable alternative processes.

It should be noted that there is a computational model of speech that does use SSCs. In Dell's (1986) spreading activation model (see Chapter 1.3.4), syllabic representations activate corresponding SSC representations. However, these are intermediate representations in order to ultimately identify the appropriate phonemes for articulation. This intermediate SSC stage was removed from later instantiations of the model (Dell & O'Seaghdha, 1992).

Moreover, the results of this study call into question the syllable as the primary motor representation of speech. First, does the brain store syllabic motor programs? These results showed no significant behavioral advantage for *learned illegal* pseudowords over novel *SSC* pseudowords even though the *learned illegal* pseudowords were practiced over 2 days and the *SSC* pseudowords were novel at the time of the testing phase. As previously discussed, these results implicate the SSC as the speech representations learned during speech motor learning, not syllables. Moreover, the *illegal learned* pseudowords were only presented in full during practice, and subjects never produced the illegal consonant clusters in any other context. If the goal of a syllabary is to use the most complete motor program available, why did the speech network store the smaller SSC units instead of larger syllabic units?

It is possible that two days of practice did not rank a syllable as frequent enough for storage in the syllabary. It is also possible that the behavioral results reflect an

intermediate stage of learning in which the SSCs are learned first, then syllables. If the subjects had continued practicing the *learned illegal* pseudowords for days, weeks, or months, the items might have been added to the syllabary, and we might have seen a behavioral advantage for *learned illegal* pseudowords over novel *SSC* pseudowords.

The results suggest that syllables are not the motor representations used for this speech motor sequence learning task (but that SSCs are). They also suggest that phoneme sequences that do not follow syllable structure organization are also not the motor representations used. Our results showed a behavioral advantage for the production of novel *SSC* pseudowords over novel *middle* pseudowords. If only the full *learned illegal* pseudowords are presented during training, naively, there should be no behavioral advantage to learning either the *middle* or *SSC* divisions. For example, if subjects learned the *learned illegal* pseudoword, TGEFSH, both the *SSC* chunking patterns {TG, E, FSH} and *middle* chunking patterns{T, GEF, SH} contain three items. Both require two transitions between sub-syllabic programs. Based solely on these merits, one would expect no behavioral advantage to learning consonant-consonant transitions compared with consonant-vowel or vowel-consonant transitions. However, the consonant-consonant transitions are the novel elements of the sequence, and learning them represents the greatest area of gain. Moreover, as previously discussed, the consonant-consonant transitions of a consonant cluster have greater articulatory cohesion than other transitions. It is more computationally efficient for the speech network to store representations of consonant clusters than other phoneme-to-phoneme transitions.

Note that we are not suggesting that motor sequence learning is equivalent to a change in phonotactic status of the learned consonant clusters. As previously discussed, the Sonority Sequencing Principle hypothesizes that the sonority contour determines which phoneme sequences within a syllable are legal for a given language, and, in English, word-edge consonants in a cluster must have a lower sonority ranking than the more medial consonants. Clusters which violate this contour are difficult to pronounce for a native speaker of English, and the more strongly a cluster violates this contour, the more difficult it usually is to pronounce; rising onset clusters are easier to pronounce than plateaus which, in turn, are easier than falling clusters (Davidson, 2006; Redford, 2008). However, it appears that SSP cannot account for all differences in difficulty producing non-native consonant clusters. For example, Smolensky and colleagues (2004) found that English-speaking subjects were less accurate to produce word-initial /vn/ than word-inital /zm/ even though the two clusters are equally unnatural, with the same sonority difference between the consonant components. This suggests that gaining fluency of a nonnative phoneme sequence results from the acquisition or modification of a new motor program, but not necessarily a change of the phonotactic status of the sequence.

Based on our findings, we theorize that consonant clusters are obligatorily produced as a single motor chunk. We have previously discussed the evidence for SSCs as a unit of speech representation based on data from speech errors, coarticulation, and language development (Chapter 1.2). Our data indicate that SSCs were the motor chunks used to learn the novel speech sequences, not syllables or other phoneme sequences. We suggest that the motor representation of an SSC is a single cohesive motor program. These results provide a window into production when a speaker produces an infrequent

or novel syllable, and emphasize that a motor unit smaller than the syllable is important to production.

3. NEURAL CORRELATES OF SPEECH MOTOR SEQUENCE LEARNING

3.1. Introduction

As anyone learning a new language can attest, mastering new speech sequences is difficult. Speech motor sequence learning requires coordinating and remembering complex sequences of articulator movements rapidly and accurately. However, little is known about the neural mechanisms that underlie this process.

3.1.1. Speech motor sequence learning

Behavioral data show that with practice, speakers produce novel speech sequences with shorter reaction times, decreased movement duration, increased accuracy, and increased movement stability (Namasivayam and van Lieshout 2008; Schultz, 2001; Smits-Bandstra et al., 2006). However, neuroimaging data is notably sparse. Rauschecker et al. (2008) not only demonstrated decreases in reaction time and utterance duration during practice of multisyllabic pseudowords, but also found activity decreases during covert repetitions of novel syllable sequences in the lateral and medial premotor cortices, superior temporal cortex, inferior frontal gyrus, and cerebellum. However, because movements were covert, they could not report behavioral evidence to demonstrate these changes were learning-related.

3.1.2. Neural correlates of non-speech motor sequence learning

In comparison, a large body of research (c.f. Hikosaka et al., 2002; Doyon et al., 2003) using human neuroimaging, single-unit recording, and pharmacological lesion studies has reliably established neural correlates to learning motor sequences of finger or eye movements. These brain areas include the cerebellum, basal ganglia, prefrontal cortex, and the supplementary motor area and presupplementary motor area.

3.1.2.1. Cerebellum

Cerebellar function has long been associated with motor learning and performance. Cerebellar ataxia occurs from damage to the cerebellar cortex, and is behaviorally identified by decreased movement accuracy and lack of fine control of voluntary movements including speech. Some patients with focal cerebellar lesions are impaired in learning motor sequences in an SRT task (Molinari et al., 1997). While this impairment is confounded by simple motor deficits classically associated with cerebellar damage, sequence learning impairment is still evident even after compensation for the motor deficits; it persists regardless of the hand used in the task and the lateralization of the lesion. Lateralized lesions only affect ipsilateral hand or arm performance for simple motor function, so this finding implies a higher-level sequence learning function in the cerebellum.

Monkeys with cerebellar lesions to both the cerebellar cortex and nuclei showed impaired performance on practiced SRT sequences compared with novel sequences, as well as impaired learning of new sequences compared to healthy monkeys (Nixon &

Passingham, 2000). In another focal lesion study, monkeys with muscimol-induced lesions in the dorsal and central parts of the dentate nucleus of the cerebellum were significantly impaired, as measured by number of errors, in performing learned sequences in a 2x5 task (Lu et al., 1998). (The 2x5 task is analogous to the 2x10 task, described in Chapter 2.1.1 but with only 5 sets per hyperset to accommodate the monkey's memory capacity.) However, there was no significant difference in performance of learning new sequences. Thus, the dentate nucleus of the cerebellum appears to be more involved in the performance of learned sequences, while the cerebellar cortex is more involved with the learning of new sequences. A large literature of imaging studies of the healthy human brain using PET or fMRI also supports this theory. Activity in the cerebellar cortex tends to be greater during the performance of new sequences compared to learned sequences, while activity in the dentate nucleus is greater for learned sequences. (Doyon, et al., 2002; Floyer-Lea & Matthews, 2004; Grafton, et al., 1994; Jenkins et al., 1994; Jueptner et al., 1997; Toni et al., 1998).

3.1.2.2. Basal Ganglia

Like the cerebellum, two distinct areas of the basal ganglia have been implicated in motor sequence learning and production. The "planning" region of the basal ganglia – the anterior caudate and anterior putamen (anterior to the anterior commissure) – is associated with planning motor sequences while the "motor" region – the middle and posterior putamen – is associated with execution. In a study using muscimol injections to temporarily and focally lesion monkeys trained on a 2x5 task, lesions to the planning region of the basal ganglia significantly impaired performance when acquiring a new

hyperset, as measured by the number of attempts needed to complete the hyperset (Miyachi et al., 1997). Lesions to the motor region resulted in significant behavioral deficits in performing hypersets learned previous to the injection. The planning region injections also impaired performance on previously learned hypersets, although the deficit was significantly smaller than for lesions to the motor region.

In a study of single cell recordings of monkey basal ganglia during the performance of a 2x5 task, cells were functionally classified as “new-preferring” if they were significantly more likely to fire during the acquisition of new hypersets than during the performance of previously learned hypersets, as “learning-preferring” if they were significantly more likely to fire during learned hypersets than new hypersets, or “non-selective” if the activity during learned and new hypersets were not significantly different (Miyachi et al., 2002). In the planning area of the basal ganglia, more recorded cells were classified as “new-preferring,” while in the motor area, more cells were classified as “learned-preferring”. Similarly, in neuroimaging studies of healthy subjects, activity in the planning region is greater for the production of novel sequences than for learned sequences, while activity in the motor region is greater during the production of previously learned sequences than for novel ones (Floyer-Lea & Matthews, 2004; Grafton, et al., 1994; Janowski et al., 2009; Jueptner et al., 1997).

Graybiel (1998) proposed that the basal ganglia stores and compresses incoming cortical motor and sensory information in order to recode these representations that are then performed as one unit (i.e. a chunk). Boyd and colleagues (2009) compared the chunking patterns of stroke patients with basal ganglia damage to those of healthy

controls during an SRT task with a learned sequence. They found that patients with basal ganglia damage performed the sequence with fewer chunks and fewer items within each chunk than healthy controls. Moreover, in single cell recordings in monkeys, cells recorded in the motor area of the basal ganglia tended to fire most consistently during a specific movement such as during the first button press of a set or during the reach for the second button of a set (Miyachi et al., 2002). Cells recorded in the planning area tended to have consistent firing during the delay period between the go signal and the onset of the first movement of the first set or over multiple sets. These results suggest that the anterior planning region could be more involved in the acquisition of chunks as well as the high-level performance of these chunks, and the posterior motor region could be more involved in the performance of learned chunks, more specifically in the performance of the individual motor actions within a learned chunk.

3.1.2.3. Prefrontal Cortex

In fMRI and PET studies of the human brain, the prefrontal cortex (PFC) tends to be more active during learning of new sequences than during the performance of previously learned sequences (Floyer-Lea & Matthews, 2004; Jenkins et al., 1994; Sakai et al., 1998; Sun et al., 2007). Averbeck and colleagues have shown that single cell recordings in the monkey PFC can predict their movements during eye-movement and drawing tasks that require the production of a series of discrete movements (Averbeck et al., 2002, 2003, 2006). However, while patients with PFC lesions are impaired in a SRT task – with longer reaction times than control subjects – they are able to chunk a

structured sequence with obvious chunking points exactly as healthy subjects do (Koch et al., 2006). This implies that the PFC is not involved in the chunking mechanism (at least in humans). It has been suggested that PFC is the site of working memory for sequencing tasks (for a review, see D'Esposito, 2000). Working memory is necessary to learning and producing a sequence because both tasks require not only sustaining a memory of the elements of a chunk as they are being produced, but also remembering cue-to-movement associations.

3.1.2.4. SMA and pre-SMA

The medial premotor cortex can be physiologically and functionally defined as two separate areas: the supplementary motor area (SMA) and the presupplementary motor area (pre-SMA) (Halsband et al., 1994; Matsuzaka et al., 1992). Pharmacological lesions to the pre-SMA impaired monkeys' ability to acquire new motor sequences in a 2x5 task (Nakamura et al., 1999). Performance on previously learned sequences was not significantly affected by the lesions. Temporary lesions of the SMA, however, did not affect the performance of new or learned sequences. Reaction time increased when lesions were applied to either the pre-SMA or SMA, but the effect was larger for SMA deactivations.

In a study of medial premotor cortex single cell recordings of monkeys, behavior of the neurons was classified as new-preferring, learned-preferring, or non-selective. The majority of the new-preferring neurons were recorded from the pre-SMA, while the majority of learned-preferring cells were recorded from the SMA (Nakamura et al., 1998).

Many new preferring cells showed decreased neuronal activity as learning progressed. Conversely, most learned-preferring cells increased their neuronal activity as learning progressed. The behavior of the pre-SMA is in agreement with some imaging studies of healthy human motor sequence learning. Activity in the pre-SMA is greater during the learning of new sequences than during performance of learned sequences (Floyer-Lea & Matthews, 2004; Sakai et al., 1998). Some studies found that SMA cells do not respond differentially to learned and novel sequences (Hikosaka et al., 1996, Nakamura et al., 1998). These studies hypothesize that the SMA is simply involved in the motor performance and not in learning per se. This is supported by a diffusion tensor imaging study (DTI) of possible striatal connections, where most of the seeds in the left SMA were connected to the motor area of the basal ganglia, while most of the seeds in the left pre-SMA were connected to the association area (Lehericy et al., 2004).

Single cell recordings from the pre-SMA not only show their preferential activity during learning of new sequences, but also a pattern of firing that implicates them in the initiation of chunks. Most of its new-preferring cells showed activity during the delay period between stimulus onset and the first button press (Nakamura et al., 1998). Even after a sequence was well learned and the cells' activity decreased drastically, the cell showed activity during this delay period and was quiet for the rest of the trial. The first set of the hyperset had the strongest delay period activity across cells, but later sets also showed delay period activity. Recall that in a hyperset there could be several hierarchical levels of chunking. Here, it is possible that the entire hyperset represents one chunk, but within that, each set is coded as a chunk containing the movements for the two button presses. Unpacking the hyperset chunk may require the strongest

neuronal activity during the first set's delay period. Unpacking the set chunks may require weaker neuronal activity during subsequent sets' delay periods.

While the SMA may be involved in more motoric aspects of motor sequence production, the pre-SMA may be necessary both during learning and during the production of learned sequences to initiate chunk production. In Kennerley et al (2004), subjects learned a 12-element sequence with bimanual button pushes. After locating each subject's chunking points based on reaction time data, the authors applied rTMS to the pre-SMA. When this disruption was applied immediately before the onset of a chunk, subjects were significantly impaired in performing the sequence compared to when no disruption was applied, as measured by reaction time. However, the application of rTMS during the sequence production at a non-chunking point had no effect on subjects' behavior.

3.1.3. Present study

In the present study, we combined a behavioral learning paradigm with functional and structural neural imaging to further our understanding of speech motor sequence learning. Our first goal was to demonstrate speech motor sequence learning with behavioral measures of speaker performance using novel sequences constructed to tax the speech motor learning system (e.g., GVAZF). Our second goal was to illuminate the neural circuitry responsible for speech motor sequence learning by using functional magnetic resonance imaging (fMRI) to compare brain activity during production of novel speech sequences that had been practiced compared to those that had not. Finally,

traditional motor sequence learning studies have demonstrated correlations between learning success and both brain response and anatomy (Gaser & Schlaug, 2003; Steele & Penhune, 2010; Tomassini et al., 2011), and studies in which novel speech sounds are learned have demonstrated analogous correlations (Golestani, et al., 2007; Golestani, and Pallier, 2007; Zhang, et al., 2009). We therefore explored whether individual differences in speech motor sequence learning success are correlated with measure of brain structure and function. To do so, we correlated subject performance with brain activity and with an estimate of white matter integrity that is derived from diffusion tensor imaging (DTI). Finally, in order to better understand the computational ramifications of the neuroimaging results, we compared simulations from the DIVA and GODIVA model to the fMRI results.

3.2. Methods

3.2.1. Participants

Eighteen right-handed native speakers (10 female, aged 20-43 years, mean 25.6 years) of American English participated. All subjects reported normal or corrected-to-normal vision and no history of hearing, speech, language, or neurological disorders. Informed consent was obtained according to the Boston University Institutional Review Board and the Massachusetts General Hospital Human Research Committee. Two subjects (1 female, ages 22 and 34 years) were removed from imaging analysis due to a large percentage of non-response errors (> 25%).

3.2.2. Speech stimuli

Subjects produced two types of monosyllabic speech sequences that contained bi- or tri-consonantal initial (onset) and final (coda) consonant clusters. *Legal* syllables (e.g., BLERK) contained consonant clusters that are legal in English, and *illegal* syllables (e.g., GVAZF) contained consonant clusters that are illegal in English, but legal in some other natural language. None of the subjects had prior experience with any languages in which these consonant clusters are legal. Each consonant cluster was used in only one syllable; no two syllables contained the same consonant cluster. The 24 legal and 30 illegal syllables are listed in Table 3.1.

Legal

BLERK	FREMP	KRENGTH	TRALP	GWEFTH	SPRIDTH
BRALK	GLANCH	PLARTH	THRIMF	TWERVE	SWARF
DRALF	GRALVE	PRENGE	DWILM	THWILB	SKELN
FLISK	KLELTH	SHRIDTH	KWANST	SPLERST	STISP

Illegal

FSEFK	VTHASHP	SHTAZG	BVIMPF	TVITP	PTACHST
FSHIKP	ZVEKCH	VBIMK	BZINSCH	BDANGT	TBASTF
FTHAMCH	FPESCH	VGAMSH	GVAZF	DKEDV	TGITK
FZICHB	FTEBSCH	ZBAPK	KVACHK	GBESB	TPIPF
VSEPSH	SHKEVT	ZDEBG	TFIPSHCH	KPESHCH	ZGEKF

Table 3.1. Orthographic representations of stimuli with legal consonant clusters in English (top) and illegal consonant clusters in English but legal in some other natural language (bottom).

The number of phonemes per syllable was balanced across conditions. None of the syllables formed an orthographic or a phonological word found in the MRC Psycholinguistic Database (Coltheart, 1981). All illegal syllables had a neighborhood size of 0 as no words could be created by adding, deleting, or substituting a single phoneme. In order to prevent syllabification of clusters with word-edge sonorant consonants and to maintain the integrity of the intended syllable structure, consonant clusters in the illegal syllables contained only non-sonorant consonants: stops, fricatives, and affricates. None of the syllables could be interpreted as more than one morpheme. Syllables contained the vowels: /ɛ, I, æ/, and vowels were distributed equally within each condition and across all items.

3.2.3. Practice Sessions

Prior to scanning, subjects completed two practice sessions over consecutive days in which they repeatedly produced 15 of the *legal* syllables and 15 of the *illegal* syllables. Subjects were divided into 4 groups, each of which practiced a different subset of the legal and illegal syllables. The illegal syllables that were not learned during the practice sessions by each group were used as novel illegal stimuli during functional imaging (see Section 1.5). Assignment of illegal syllables to the learned and novel categories was counterbalanced across subjects. Practice sessions occurred at least one and not more than two days prior to scanning to allow for memory consolidation (Brashers-Krug et al., 1996; Fenn et al., 2003; Stickgold, 2005; Davis, et al., 2009). Each syllable was produced 30 times per practice session. Syllables were presented in pseudorandom order.

During each practice trial, subjects simultaneously saw an orthographic representation of the syllable to be produced for 1450 ms and heard a 480 ms recording of the syllable; visual and auditory stimulus onsets were aligned. The orthographic representations had a consistent orthography-to-phonology mapping across syllables; for example, if E represented /ɛ/ in one syllable, it represented that sound in all syllables. Following stimulus presentation and a jittered pause of 500 to 1000 ms, a tone acted as a GO signal to cue subjects to produce the target syllable. Subject utterances were recorded at 44100 Hz for 1s with a Samson C01U USB studio condenser microphone. For four subjects, utterances were recorded for 2 s rather than of 1s, resulting in a longer inter-trial interval. Two-sample t-tests of learning-related behavioral measures and brain activity indicated no differences associated with this longer inter-trial interval.

Subjects were asked to produce the syllables as quickly and accurately as possible and to replicate the auditory stimulus while producing all the sounds seen in the orthographic cue. Subjects were also instructed to avoid schwa epenthesis - insertion of schwa sounds between phonemes – a common response when producing illegal consonant clusters. After instruction but prior to the practice sessions, subjects practiced 5 repetitions of two legal and two illegal syllables that were not used for the rest of the study. During these introductory trials, an experimenter provided feedback about production accuracy.

3.2.4. Behavioral data analysis

To evaluate speech motor sequence learning, we measured changes in the following three *learning success indices* over the practice sessions: (i) error rate, (ii) reaction time and (iii) utterance duration. Error rates were given by the percentage of the first 5 productions of each syllable during each practice session that were produced incorrectly. Error was based on the first 5 productions of each syllable to avoid confounds from subject fatigue over the course of the practice session. Errors were defined as phoneme additions, deletions, and substitutions, and utterance repetitions and restarts. A single rater judged errors for all trials[4]. A subset of recordings (including recordings from the fMRI session) were also rated for errors by a second rater; the inter-rater reliability, K, was 0.7708 (see Cohen, 1960). Duration and reaction time measurements were based on the first 5 error-free trials of each syllable during each session. Utterance onset and offset were automatically labeled based on sound pressure level thresholds individually chosen for each practice session, then hand-checked. These measures were used to calculate reaction time (time from GO signal to utterance onset) and utterance duration (time from onset to offset of the utterance).

To assess learning-related changes due to practice, we compared the error rate, accuracy, and reaction time changes from the first practice session to the second with paired t-tests. Each behavioral measure was averaged within each condition and within each subject. We hypothesized that we would see greater learning in the illegal condition because those syllables included both novel syllables and novel consonant clusters

[4] Due to obvious differences between the composition of illegal and legal stimuli, raters were not blind to the conditions of the practice session utterances they rated.

whereas the legal condition includes novel syllables of familiar consonant clusters. Paired t-tests comparing the mean error rate, duration, and reaction time in the illegal and legal conditions were performed to test this hypothesis. T-tests were corrected for multiple comparisons using a false discovery rate threshold of < 0.05.

3.2.5. fMRI paradigm

During functional imaging, subjects produced the 15 legal and 15 illegal syllables that they had learned during the practice sessions and the 15 novel illegal syllables that they had not been exposed to previously. Thus, there were three syllable production conditions: *learned legal*, *learned illegal*, and *novel illegal*. A *baseline* condition was also intermixed during imaging in which subjects viewed a series of asterisks on the screen instead of the orthographic stimulus and rested quietly instead of producing a syllable.

A *novel legal* condition was not included because pilot behavioral data showed no significant behavioral changes with repeated productions of the novel syllables[5]. This implied that subjects were already at ceiling performance for *legal* syllables when they are novel. Therefore, we expected brain activity to be similar for *novel legal* and *learned legal* productions. In order to keep the fMRI session to a manageable length for subjects, a *novel legal* condition was omitted.

We acquired fMRI data using a sparse sampling paradigm so that subjects heard the auditory cues and produced the target syllables between data acquisitions in the

[5] The data presented here showed significant improvements in accuracy for *legal* syllables, but not for utterance duration, but this improvement was marginal compared to that for the *illegal* syllables (Fig. 3.1).

absence of scanner noise (Hall et al., 1999). Subjects followed the same behavioral paradigm used during the practice session but with an additional pause after the syllable production to temporally align the image acquisition to the expected peak of the hemodynamic response (Belin et al., 1999; Yang et al., 2000) and to acquire the volume. A single trial lasted 10 s. Each run consisted of 40 trials and lasted 7 minutes. Subjects completed 8 runs, 80 trials per condition, and approximately 5 productions of each syllable[6]. Conditions were pseudorandomly distributed across the 8 runs with at least 8 instances of each condition appearing in each run.

Instructions and visual stimuli were projected onto a screen that subjects could view from within the scanner via a mirror attached to the head coil. Auditory stimuli were played over Sensimetrics MRI-compatible insert headphones model S-14. Subjects' productions were transduced by a Fibersound fiber optic microphone, model FOM1-MR-30m, to an IBM PC ThinkPad X61s and recorded using Matlab at 44.1kHz. The microphone was positioned approximately 3" from subjects' mouths.

Because the neuroimaging paradigm assumed learning of previously novel illegal syllables, only subjects that demonstrated significant reduction in two of the three

[6] Rauschecker et al. (2008) showed neural changes associated with only 5 repetitions of covertly produced novel pseudowords. While it would be ideal to present our stimuli only once during each scanning session so *novel* sequences remain so, the limited number of available legal and illegal consonant clusters that meet our stimulus construction restrictions (Section 3.2.2) constrains the number of possible stimuli. Thus, it is necessary for subjects to repeat syllables over the course of neuroimaging. Moreover, behavioral results with similar stimuli (Section 2.3.1) demonstrate that learning-related changes in accuracy and utterance duration continue well past the fifth repetition.

learning indices across the practice sessions were included in the neuroimaging analyses. All subjects included in the analysis met this criterion.

3.2.6. Image acquisition

MRI data were acquired using a 3 Tesla Siemens Trio Tim scanner with a 32 channel head coil. For each subject, a high-resolution T1-weighted volume was acquired (MPRAGE, voxel size: 1 mm^3, 256 sagittal images, TR: 2530 ms, TE: 3.44 ms, flip angle: 7°). Functional gradient echo EPI scans (41 horizontal slices, in plane resolution: 3.1 mm, slice thickness: 3 mm, gap: 25%, TR: 10 s, TA: 2.5 s, TE: 20 ms) were automatically registered to the AC-PC line and were collected sparsely with 10 s between scan onsets. Diffusion-weighted images were also acquired with a single-shot spin-echo echo-planar sequence (64 slices, voxel size: 2 mm^3, TR: 8020 ms, TE: 83 ms, GRAPPA parallel reconstruction). Diffusion weighting was performed along 60 independent directions with a b-value of 700 s/mm^2. A reference image with no diffusion weighting was also acquired.

3.2.7. fMRI behavioral data analysis

For each syllable production, reaction time, utterance duration, and error rate were calculated following the removal of noise associated with the scanner bore echo and peripheral equipment using a Wiener filter (Wiener, 1949). Raters were blind to the condition to which illegal utterances belonged. Each behavioral measure was averaged within each condition and within each subject. One-way ANOVAs ($p_{FWE} < 0.05$, Bonferroni corrected) were employed to test for significant differences across the three

conditions for each of the three behavioral measures. Bonferroni corrected paired t-tests ($p_{FWE} < 0.05$) were then used to test for behavioral differences between each pair of conditions.

3.2.8. fMRI data analysis

The Nipype (Ghosh et al., 2010) neuroimaging software interface was used to analyze imaging data that permitted the use of preferred processing routines from various neuroimaging analysis packages. Using SPM8 image processing tools (http://www.fil.ion.ucl.ac.uk/spm/software/spm8), functional images were motion-corrected and realigned to the subject's anatomical volume and high-pass filtered functional data with a standard 128 s cutoff frequency. Error trials, intensity-related outliers (> 3 standard deviations from subject mean), and motion-related outliers (> 2mm) were removed from the analysis; approximately 10% of all trials were removed due to these parameters. Blood oxygen level dependent (BOLD) responses were estimated using a general linear model (GLM), and the hemodynamic response function for each stimulus event was modeled as a finite impulse response. The model included 4 condition-specific variables – *learned illegal*, *novel illegal*, *learned legal*, and *baseline* – and additional covariates: utterance duration measures, linear detrending covariates, and motion parameters. The model was estimated for each subject. Model estimates for the *novel illegal* and *learned illegal* conditions were contrasted (*novel illegal - learned illegal*) at each voxel. Group statistics were then calculated separately for cortical and subcortical regions.

Surface-based analysis was used to assess group blood oxygen level dependent (BOLD) response differences in the *novel illegal* and *learned illegal* conditions in the cerebral cortex. T1 volume segmentation and cortical surface reconstruction for each subject were performed with the FreeSurfer image analysis suite (Dale et al., 1999; Fischl et al., 1999a; Fischl et al, 2002). The activity of cortical voxels in the *novel illegal - learned illegal* contrast volume for each subject was then mapped to that subject's cortical surface. Subject data were aligned by inflating each individual surface to a sphere and registering it to a template representing the average surface curvature of a set of neurologically normal adult brains (Fischl et al., 1999b). The surface-based contrast data were smoothed with a 6mm full-width half-maximum (FWHM) kernel and then averaged across subjects. Group-level t-statistics were calculated at each vertex. Vertex-wise statistics were first thresholded at $p < 0.001$ (uncorrected). Cluster-level significance thresholds were then estimated separately for each hemisphere using a Monte Carlo simulation with 10,000 iterations (Hayasaka and Nichols, 2003). Results were cluster-thresholded in each hemisphere at cluster-wise p (CWP) < 0.0167, to correct for surface-based tests in each hemisphere and one subcortical volume-based test.

Group differences in subcortical BOLD responses were assessed by normalizing and aligning individual T1 volumes to the MNI152 template using SPM8's DARTEL image registration toolbox (Ashburner, 2007; Klein, et al., 2009). Individual subject's voxel-based contrast data were smoothed using a 6 mm FWHM kernel and then averaged across subjects. Group-level t-statistics were calculated at each voxel and thresholded at $p < 0.001$ (uncorrected). After a subcortical mask was applied, the results were

thresholded at the cluster-level at CWP < 0.0167 (corrected) using a separate Monte Carlo simulation with 10,000 iterations.

In addition to mapping the brain regions that responded differently to *novel* and *learned illegal* syllables, the correlation between *learning success* and difference in BOLD response between the *novel illegal* and *learned illegal* conditions was calculated. The two significant measures of subjects' learning success were used: difference in the mean error rate and utterance duration for the novel and learned illegal syllables produced during fMRI normalized by the measures for the learned illegal syllables. For example, one of the measures of learning success was the percent increase in utterance duration of the *novel illegal* syllable over the *learned illegal* syllable.

A leave-one-out cross validation technique was used to avoid biases from non-independence of cluster selection and the BOLD-behavioral correlation measures (Esterman et al., 2010). For each subject, a set of significant clusters from the *novel illegal - learned illegal* contrast was calculated from a GLM as described above that excluded that subject's data. By not using a given subject's data to determine clusters from which the BOLD signal is extracted, the dependency between the voxel-selection procedure and the BOLD measures is removed. Each of the three learning success measures for each subject were then correlated with the mean beta coefficient within each significant cluster from the *novel illegal - learned illegal* contrast as determined by this leave-one-out method. Based on past evidence, we expected to find positive correlations between learning success and decreases in brain activity (Golestani and

Zatorre, 2004; Zhang, et al., 2009; Tomassini et al., 2011). We report significance values of $p_{FWE} < 0.05$ for a one-tailed (positive) correlation (Pearson's R).

3.2.9. DTI data analysis

Using the FMRIB Diffusion Toolbox (http://www.fmrib.ox.ac.uk/fsl), the diffusion-weighted raw data was first corrected for eddy-current distortions and motion artifacts. Diffusion tensors were then fitted at each voxel within a cortical mask (Behrens, et al., 2003). Data from two subjects were not included in this analysis due to excessive head motion during collection of the DTI volume that caused a failure in the DTI analysis software. DTI volumes were coregistered with subjects' anatomical T1-weighted volume using FreeSurfer. FreeSurfer was also used to identify white matter regions of interest (ROIs) that correspond to the significant cortical clusters identified in the *novel illegal - learned illegal* surface-based functional analysis. Each ROI consisted of the voxels that lie 2 mm below the gray-white surface vertices within a cluster (Kang, et al., 2012). The mean fractional anisotropy (FA) value within each white matter ROIs was then calculated. We correlated the mean FA of each ROI with measures of each subjects' learning success as described in Section 1.8. Based on past evidence, we expected to find positive correlations between learning success and brain structure integrity (Gaser & Schlaug, 2003; Golestani, et al., 2007; Golestani, and Pallier, 2007; Tomassini et al., 2011). We report significance values of $p_{FWE} < 0.05$ for a one-tailed (positive) correlation (Pearson's R).

3.2.10. GODIVA simulations

In order to quantitatively compare the fMRI results to the predictions made by the GODIVA model, two simulations mimicked the *novel illegal* and *learned illegal* conditions. In the study described in Chapter 2, subjects produced novel syllables as accurately and as quickly as practiced syllables only if they contained the previously learned SSCs, particularly consonant clusters. It was hypothesized that consonant clusters were the chunks used in sequence learning. Therefore, one simulation was run with the default parameters of the model to simulate a naïve native English speaker's available phonological representations in the *novel illegal* condition. It was assumed that as a naïve subject does not initially have representations for the illegal consonant clusters, they produced novel sequences as trisyllabic, schwa-epenthesized analogs; a subject's representation of /zbƐdb/ was really /zə.dƐb.gə/. A second simulation added a set of novel consonant clusters in the phoneme set – the onset /zb/ and the coda /db/ – to simulate the acquisition of the illegal consonant clusters in the *learned illegal* condition.

After simulation, activity in each planning and choice module was averaged across the activity time course. The scaling of the cell activations between modules is arbitrary; only the relative strength of activity between cells within a module is relevant. To compensate for this, the average activity values in the choice and planning modules of the IFS, preSMA, and SSM were first normalized by the maximum cell activation. Then, to simulate the utterance duration covariate added to the fMRI GLM, the activity measures were normalized by the duration of the activations. ROIs were created in volume space at the corresponding peak voxel locations of the *novel illegal – learned*

illegal cluster locations with the value of the sum of the planning and choice normalized activities for each module. This volume was then projected to an average brain's inflated surface and smoothed with a 6mm kernel for a quantitative comparison to the fMRI results.

3.3. Results

3.3.1. Behavioral measures of learning

Across-subject measures of error rate and utterance duration indicated significant improvement in performance between practice sessions for the learned illegal but not the learned legal syllables (Figure 3.1A). With practice, *learned illegal* syllables had a significantly lower error rate on the second practice session compared to the first (mean change = 25.8%, s.d. = 10.1; $t_{(15)} = -8.34$, $p < 0.0001$), as did *learned legal* syllables (mean change = 3.5%, s.d. = 3.8; $t_{(15)} = -3.51$,$p < 0.05$). Error rate decreased significantly more for the *learned illegal* syllables than *learned legal* syllables (mean change = 22.3%, s.d. = 11.4 $t_{(15)} = -7.39$, $p < 0.0001$). The duration of *learned illegal* syllables was significantly shorter during the second session compared to the first session (mean change = 85.4 ms, s.d. = 33.9, $t_{(15)} = 6.12$, $p < 0.0001$) but *learned legal* syllables showed no significant change during the second session (mean change = 25.0 ms, s.d. = 49.0, $t_{(15)} = -1.91$, $p = 0.025$, n.s.). Duration decreased significantly more for *learned illegal* than for *learned legal* syllables (mean change = 110.4 ms, S.D. = 59.9; $t_{(15)} = -7.45$, $p < 0.0001$). Reaction time did not significantly change from the first to the second practice session for either the *learned legal* (mean decrease = 8.3 ms, s.d. = 56.5, $t_{(15)} = 0.57$, $p = 0.02$, n.s.) or *learned illegal* syllables (mean decrease = 20.7 ms,

s.d. = 59.3, $t_{(15)}$ = 1.35, p = 0.05, n.s.) and changes in reaction time were not significantly different between conditions (mean = -12.4 ms, s.d. = 28.8, $t_{(15)}$ = -1.67, p = 0.12, n.s.).

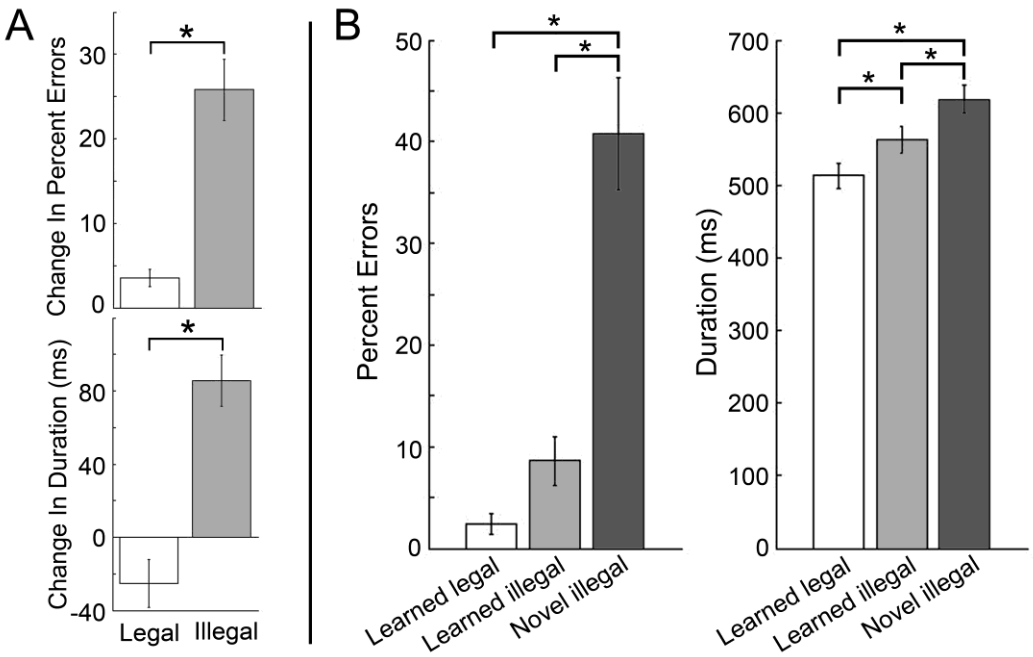


Figure 3.1. Behavioral results from sequence productions. The * symbol indicates significance of p < 0.0001. *A*. Practice session results comparing behavior on day 1 and day 2 of practice. The reduction in mean error rate (*top*) and utterance duration (*bottom*) was significantly greater for *learned illegal* syllables (gray) than *learned legal* syllables (white). *B*. Image session results. *Left*: Subjects produced *learned legal* (white) and *learned illegal* (light gray) syllables significantly more accurately than *novel illegal* syllables (dark gray). *Right*: Subjects produced *learned illegal* syllables significantly faster than *novel illegal* sequences, but slower than *learned legal* syllables. Bars indicate standard error.

Significant differences in learning success indices across the three speaking conditions during the fMRI session were noted (Figure 3.1B). One-way ANOVAs tested for significant differences in learning indices across conditions; *post hoc* paired t-tests compared pairs of conditions. Utterance durations were significantly different between conditions ($F_{(2,13)} = 8.39$, $p < 0.0001$). *Learned illegal* utterances were significantly shorter in duration than *novel illegal* utterances (mean difference = 55 ms, s.d. = 36.9, $t_{(15)} = 5.78$, $p < 0.0001$) and *learned legal* utterances were even shorter (mean difference = 50 ms, s.d. = 21.6, $t_{(15)} = 8.98$, $p < 0.0001$). Error rate (Figure 3.1B, *left*) was also significantly different across conditions ($F_{(2,13)} = 33.99$, $p < 0.0001$). No difference in accuracy was noted between *learned legal* and *learned illegal* utterances (mean difference = 6.1%, s.d. = 9.9, $t_{(15)} = 2.41$, $p = 0.03$, n.s.), but subjects committed more errors during *novel illegal* utterances compared to both the learned legal (mean difference = 38.3%, s.d. = 21.6, $t_{(15)} = 6.88$, $p < 0.0001$) and *learned illegal* (mean difference = 32.2%, s.d. = 17.8, $t_{(15)} = 7.01$, $p < 0.0001$). Reaction times were not significantly different between conditions ($F_{(2,13)} = 0.04$, $p = .96$, n.s.).[7]

3.3.2. FMRI analysis

Figure 3.2 and Table 3.2 shows the cortical brain regions that were significantly more active for *novel illegal* than *learned illegal* syllables (voxel-level $p < 0.001$, uncorrected;

[7] Note that minF'(2,58) = 16.13($p < 0.001$ ($F1_{(2,26)} = 26.05$, $F2_{(2,56)}$) was significant, based on utterance errors during scanning; however, we suggest that fixed-effects are not a "fallacy," but an inevitability across the subjects and items of the current task (Clark, 1973). Section 1.2.1 discusses the variation in performance of non-native consonant clusters; some are more difficult to pronounce than others, partially due to their sonority slope. Moreover, as will be discussed in Section 3.4.3, a significant correlation between individual subject's neuroanatomy and learning success implies that subjects will not perform as a homogenous group due to the inhomogeneity of their brain structure.

cluster-level $p < 0.0167$, corrected). The production of *novel illegal* syllables resulted in greater BOLD response in the frontal operculum and adjacent anterior insula cortex (referred to as the frontal operculum cluster hereafter) and superior parietal cortices bilaterally. In the left hemisphere, additional clusters were noted with peaks in the lateral premotor cortex (2 clusters, one in ventral lateral premotor cortex extending into the inferior frontal gyrus, pars opercularis, one in middle lateral premotor cortex extending into the inferior frontal sulcus), posterior superior temporal sulcus, planum temporale, inferior occipital-temporal cortex, and the globus pallidus. In the right hemisphere, the production of *novel illegal* syllables resulted in greater activity in the pre-supplementary motor area. No region was found to be significantly more active for the *learned illegal* than the *novel illegal* syllables.

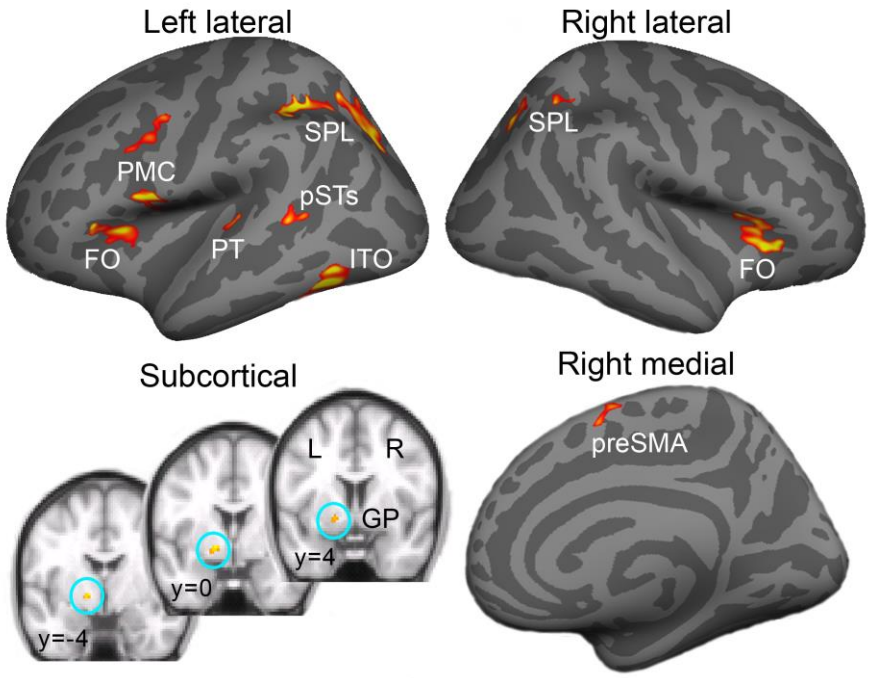


Figure 3.2. FMRI main effects of sequence learning (*novel illegal* sequences > *learned illegal* sequences). Significant clusters are shown on the left lateral (*upper left*), right lateral (*upper right*), and right medial (*lower right*) inflated surface representations of the FreeSurfer average template. Subcortical activity is shown on a series of coronal slices from the MNI305 template at the level of the pallidum (*bottom left*; y coordinate indicates mm distance from the anterior commissure in MNI space). Contrast volumes were first voxel thresholded at $p < 0.001$, then cluster thresholded at cluster-wise p (CWP) < 0.0167 to correct for three analyses: subcortical and 2 cortical hemispheres. CWP was calculated by separate Monte Carlo simulations for each of the three analyses. Abbreviations: FO = frontal operculum-anterior insula, PMC = premotor cortex, PT = planum temporale, pSTs = posterior superior temporal sulcus, ITO = inferior temporal-occipital cortex, SPL = superior parietal lobule, preSMA = presupplementary motor cortex, GP = globus pallidus.

Region name		MNI coordinates			t	Size	CWP
		x	y	z			
Left	PMC/IFS	-46.4	2.0	38.6	4.15	157 mm^2	0.0001
	PMC/IFo	-47.5	8.7	9.3	5.89	261 mm^2	0.0001
	FO	-44.4	26.2	3.3	6.14	212 mm^2	0.0001
	PT	-44.4	-27.6	3.6	4.12	57 mm^2	0.0131
	pSTS	-52.3	-40.5	7.4	4.68	100 mm^2	0.0004
	SPL	-35.9	-40.4	36.1	7.37	361 mm^2	0.0001
		-27.2	-62.7	26.4	5.34	445 mm^2	0.0001
	ITO	-48	-53.4	-6.6	8.77	308 mm^2	0.0001
	GP	-14.8	-38.1	-4.82	4.70	376 mm^3	0.0031
Right	Pre-SMA	7.5	7.4	53.4	4.14	87 mm^2	0.0009
	FO	29.3	26.9	0.9	5.70	329 mm^2	0.0001
	SPL	24.9	-56.3	36.9	5.16	136 mm^2	0.0001
		29.2	-49.8	40.3	4.68	92 mm^2	0.0003

Table 3.2. Summary of cortical and subcortical activation peaks for the main effect of learning (novel illegal > learned illegal contrast). From left to right, the columns show the region name, Talairach coordinates, T value, cluster size, and cluster-wise p (CWP). Abbreviations: FO = frontal operculum-anterior insula, PMC = premotor cortex, IFS = inferior frontal sulcus, IFo = inferior frontal gyrus, pars opercularis, pSTs = posterior superior temporal sulcus, ITO = inferior temporal-occipital cortex, SPL = superior parietal lobule, preSMA = presupplementary motor cortex, GP = globus pallidus.

3.3.3. Neural-behavioral correlation analysis

Learning success, as measured by the normalized difference in utterance duration between *novel illegal* and *learned illegal* syllables, was positively correlated with the mean response in the frontal operculum cluster ($r = 0.709$, $p = 0.0022$) identified in the *novel illegal* vs. *learned illegal* leave-one-out cross-validation contrast (Figure 3.3, *left*). No other significant correlations between other learning success measures and BOLD response were found in any of the significant clusters identified by the functional imaging analysis.

Difference in utterance duration between *novel illegal* and *learned illegal* syllables was also positively correlated with the FA values of the white matter under the cluster of posterior superior temporal sulcus activity noted in the functional imaging analysis ($r = 0.670$, $p = 0.0031$; Figure 3.3, *middle*). No other significant correlations between changes in learning indices and mean FA under the significant cortical clusters identified by the functional imaging analysis were noted.

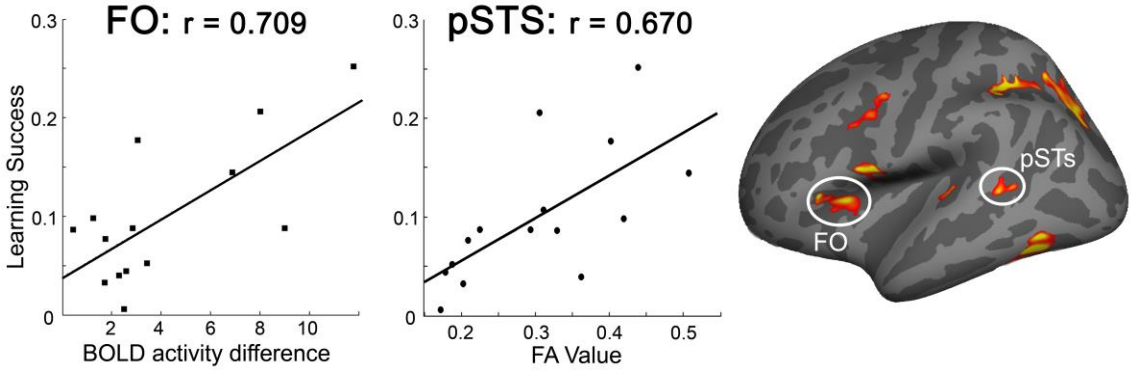


Figure 3.3. Neural markers of learning success as measured by the difference in utterance duration for novel and learned illegal syllables relative to the duration in the learned syllable condition *Left*. Significant correlations between individual learning success and the mean BOLD response given by the *novel illegal – learned illegal* syllable contrast within the left frontal operculum-anterior insula cluster (labeled FO in the brain image at the right; r = 0.709, p = 0.0022). *Middle.* Significant correlation between learning success and the mean fractional anisotropy underlying the left posterior superior temporal sulcus cluster (labeled pSTs in the brain image at the right) identified by the *novel illegal– learned illegal* syllable contrast (r = 0.670, p = 0.0031).

3.3.4. GODIVA simulations

Figure 3.4 displays the GODIVA *novel illegal* (*left colum*n) and *learned illegal* (*right column*) simulations. Figure 3.5 shows these simulations projected onto an averaged inflated brain.

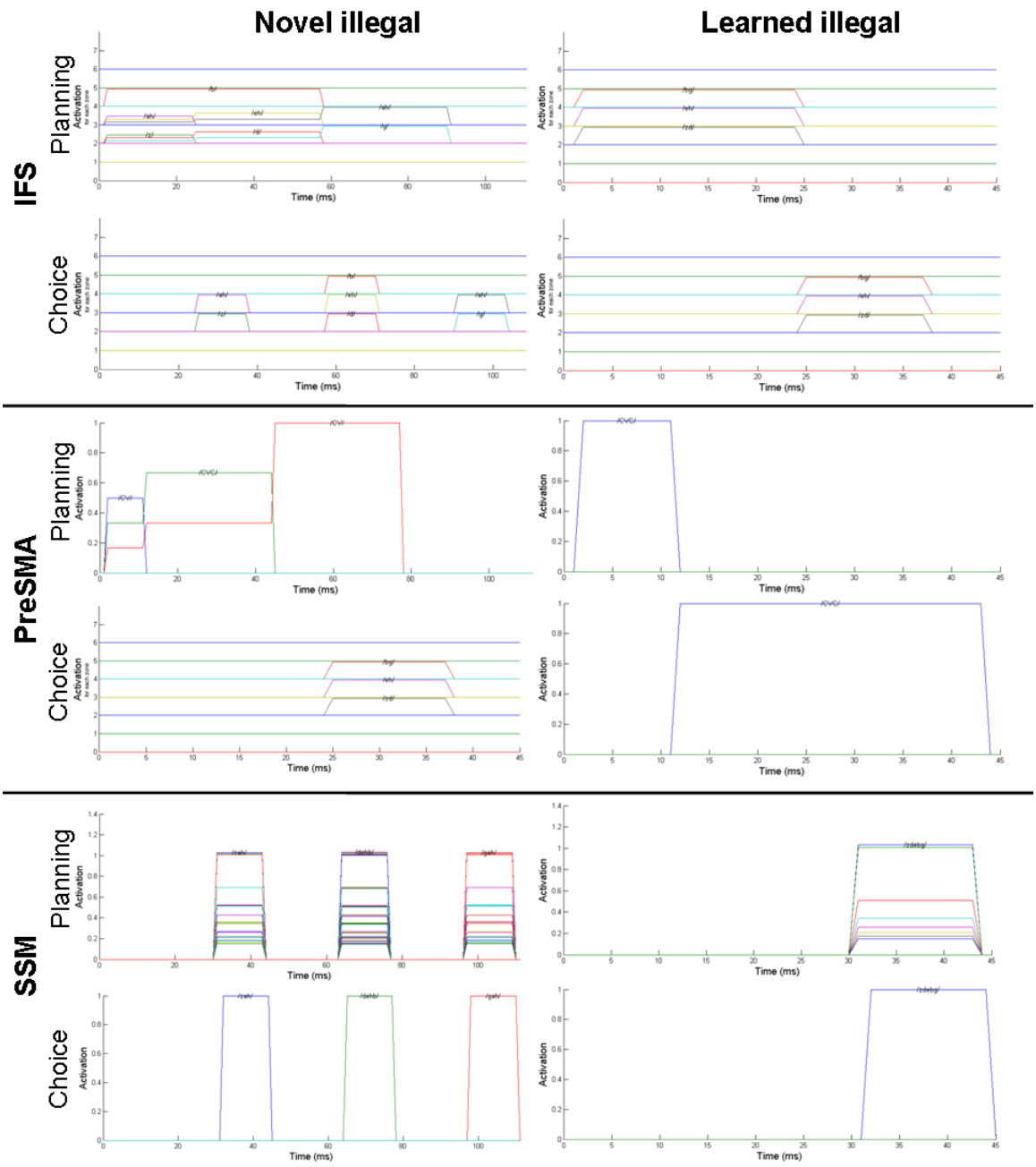


Figure 3.4. Simulation results using the GODIVA model showing cell activations (multiple colors) over time for (*left*) the utterance /zə.dɛb.gə/ to simulate the novel illegal condition and (*right*) the utterance /zbɛdb/ to simulate the learned illegal condition. Abbreviations: IFS = inferior frontal sulcus, preSMA = presupplementary motor area, SSM = speech sound map.

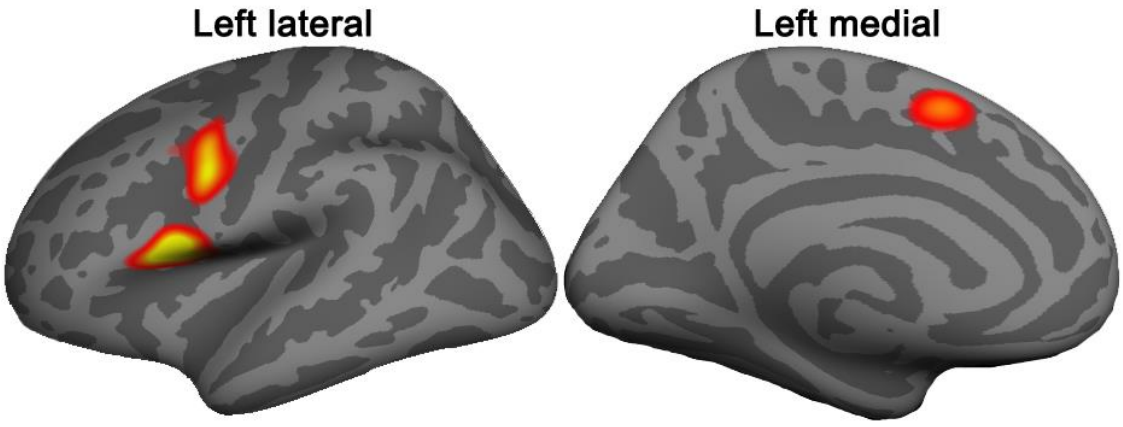


Figure 3.5. Contrast of GODIVA simulations for *illegal novel – illegal learned* utterances projected onto an inflated surface representation of the FreeSurfer average brain template.

3.4. Discussion

We explored behavioral and neural facets of speech motor sequence learning in healthy adults. Subjects practiced novel speech sequences: syllables with legal (e.g., BLERK) or illegal (e.g., GVAZF) consonant clusters. With practice, subjects produced both types of syllables more accurately and with shorter utterance durations, indicating sequence learning. Moreover, these gains were greater for the illegal syllables (Figure 3.1). The larger learning gain for the illegal syllables was likely due to the novelty of the consonant clusters. For legal syllables, the consonant clusters were familiar, so performance gains were primarily driven by associating the vowel and onset and coda clusters. Performance was further improved in the illegal syllables by motor sequence learning within the unfamiliar consonant clusters; in other words, subjects learned motor programs (and possibly phonological representations) for the new consonant clusters.

This assertion is supported by the work presented in the previous dissertation chapter and other neuroimaging findings.

Using fMRI, we compared brain activity during production of *learned* and *novel illegal* syllables (Figure 3.2). Subjects produced *learned illegal* syllables faster and more accurately than *novel illegal* syllables. BOLD responses during production of the *novel illegal* syllables were greater than for *learned illegal* syllables in a number of regions within the "minimal network" for speech motor control (Bohland & Guenther, 2006), including right presupplementary motor area, left planum temporale, posterior superior temporal sulcus (pSTs), lateral premotor cortex, globus pallidus, and the frontal operculum and anterior insula (FO) bilaterally. Moreover, significant correlations between subject's individual learning and functional or structural neural markers were noted in this network: learning success correlated with BOLD response reduction between the *novel* and *learned illegal* syllables in the left FO, and learning success correlated with white matter integrity under the left pSTs (Figure 3.3).

3.4.1. Lateral prefrontal cortex

The finding of greater activity for *novel* over *learned illegal* syllables in the lateral prefrontal cortex is loosely consistent with established neurocomputational models of speech production (Guenther et al., 2006; Bohland, et al., 2010; Golfinopoulos et al., 2010; Tourville and Guenther, 2011). The clusters of lateral premotor activity lie adjacent to motor representations of larynx and speech articulators (e.g., Takai, et al., 2010; Grabski, et al., 2012). We have postulated that motor programs for learned speech

sounds are represented in this region of the left hemisphere (Guenther et al., 2006). According to the Directions into Velocities of Articulators (DIVA) model of speech production (e.g., Guenther et al., 2006; Tourville and Guenther, 2011), each motor program encodes the movements required to produce a short speech sound sequence – e.g., a syllable or consonant cluster – rapidly and in a feedforward manner, without relying on auditory feedback. New motor programs are formed when novel speech sounds are learned or when existing motor programs are repeatedly produced in larger sequences. When a novel sequence is encountered, it is initially produced by accessing the motor programs of the individual phonemes that comprise the sequence. With practice, however, new motor programs, or chunks (Miller, 1956; Cowan, 2001), representing larger portions of the sequence, are learned. Our finding of greater activity in left lateral premotor areas for *novel* over *learned illegal* syllables is consistent with this view; *novel illegal* syllables require activation of more motor programs, and therefore more neurons in the lateral premotor cortex, than *learned illegal* syllables. Prior studies have identified an analogous reduction of activity in the more dorsal premotor region that encodes hand movements when novel hand movement sequences are learned (Jenkins et al., 1994; Jueptner et al., 1997; Honda et al., 1998; Toni et al., 1998; Doyon et al., 2002; Orban, et al., 2010).

Alternatively, the WEAVER model (Indfrey & Levelt, 2004) suggests that the lateral premotor cortex activation may be a result of the syllabification process. It is possible that subjects in our study may have incorrectly schwa-epenthesized the illegal consonant clusters of *novel illegal* syllables, requiring them to produce the syllable as three syllables. The model hypothesizes that this syllabification process takes place in

the inferior frontal gyrus[8], while the lateral premotor cortex in recruited to produce "articulatory scores," the motor programs for these syllables. Our finding of greater activity in a region encompassing both the inferior frontal gyrus and premotor cortex for *novel* over *learned illegal* syllables also agrees with these hypotheses. *Novel illegal* syllables may be syllabified into more units that require the articulation of three articulatory scores compared to monosyllabic *learned illegal* syllables.

The more ventral of the two lateral premotor cortex clusters borders on and infringes into the left inferior frontal sulcus (IFs). In speech production, IFs is implicated in phonological and verbal working memory (Chein et al., 2003; Nee, et al., 2013; Nixon et al., 2004; Poldrack et al., 1999; Smith & Jonides, 1999). It shows greater activity with greater complexity during the production of multi-syllabic utterances (Bohland, et al., 2009; Reicker, et al., 2008), but no significant difference in activity between overt and covert productions (Bohland, et al., 2009). The IFs is proposed to hold articulatory representations of the upcoming speech utterance in working memory (Baddeley, 2003; Bohland et al., 2009; Henson, 2001). Thus, learned illegal sequences required less IFs activity because phonological working memory stored fewer speech chunks than in novel illegal sequences. This function is also proposed to aid in non-speech motor sequence learning, and is proposed to reside in the dorsolateral prefrontal cortex (Barch

[8] The WEAVER model's neuroanatomical correlates are unclear due to the authors' use of vague anatomical terms. In the text, Indefrey and Levelt (2004) hypothesize that "Broca's area" – usually the inferior frontal gyrus – as the neural correlate to the syllabification process, but in a later figure, implicate both the inferior frontal gyrus and premotor cortex. Moreover, they later suggest that the lateral premotor cortex is involved in "articulation." This work references the neuroanatomical correlates provided in the text. Note, that no neural correlate has yet to be explicitly hypothesized for the syllabic speech motor programs – the so-called "articulatory scores" – hypothesized in the model.

et al., 1997; Curtis & D'Esposito, 2003; D'Esposito, et al., 2000; Grafton et al., 2002; Hazeltine et al., 1997). Our task requires verbal working memory and activates the left IFs, but the dorsal lateral prefrontal cortex is only active when task is visual or spatial (Courtney et al., 1998; Robertson, et al., 2001; Smith & Jonides, 1999; Wager & Smith, 2003). However, it is unknown whether this activation, which straddles the border on the IFs and lateral premotor cortex, represents activity in the IFs or if, due to anatomical differences between subjects and the averaged inflated brain template, appears to be there.

3.4.2. Presupplementary motor cortex and basal ganglia

Because novel syllables contain more motor "chunks", they require a greater computational effort by the motor sequencing network. Our findings of reduced activity for *learned* compared to *novel illegal* syllables in the right presupplementary motor area (preSMA) and the left globus pallidus – regions believed to form part of a basal ganglia-thalamo-cortical loop that initiates motor chunks (Contreras-Vidal, 1999; Eckert, et al., 2006; Haggard, 2008; Kotz and Schwartze, 2010) – are consistent with this view and mirror those of non-speech motor sequence learning studies (Hikosaka et al., 1996; Jueptner et al., 1997; Nakamura, et al., 1998; Sakai et al., 1998; Lehericy et al., 2005; Poldrack et al., 2005). We have proposed that this network interacts with the premotor cortex to sequentially execute a series of speech motor programs (Bohland et al., 2010).

More specifically, the subcortical loop with the preSMA is proposed to be involved in the preparatory processes for execution: selecting a desired response among multiple

possibilities and inhibiting those responses deemed inappropriate (Duque, et al., 2013; Mostofsky & Simmonds, 2008; Simmons, et al., 2008). This suggests that the increased preSMA activity for *novel* over *learned illegal* syllables may be due not only to executing more motor chunks, but also to inhibiting more inappropriate responses. For instance, subjects were specifically instructed to avoid schwa-epenthesis; when performing novel utterances, a more active effort may have been required to suppress this inclination.

Notably, the reduction in preSMA activity occurred in the right hemisphere, in contrast to prior studies associating bilateral preSMA with speech and non-speech motor sequence learning (e.g., Floyer-Lea & Matthews, 2004; Nakamura et al., 1998; Rauschecker, et al., 2008; Steele & Penhume, 2010).

3.4.3. Planum temporale and posterior superior temporal sulcus

Greater activity in left planum temporale and pSTs during *novel illegal* syllable production is consistent with proposals that these areas are involved in correcting or guiding speech movements (e.g., Guenther et al., 2006; Hickok, 2012). According to the DIVA model, these areas contain “auditory error maps” that become active when a mismatch is detected between expected and actual auditory feedback signals during speech (Guenther et al., 2006). The State Feedback Control model (SFC, Hickok, 2011) also hypothesizes that these regions are involved in auditory feedback control and that the planum temporale compares the expected and actual auditory feedback signals. However, the SFC and DIVA model diverge when describing the auditory state maps that represent the expected auditory signal. When DIVA hypothesizes that the expected

signal is a learned consequence of the feedforward motor program stored in the lateral premotor cortex, the SFC hypothesizes that the superior temporal cortex – the so-called "motor phonological system" – calculates the expected auditory output online, in real-time, based on the expected articulatory motor commands. Teasing apart these distinctions, however, is outside the scope of this research.

Based on both the DIVA and SFC models, reduced activity in the pSTs and planum temporale during the production of *learned illegal* syllables may be the consequence of detecting fewer (sub-phonemic[9]) auditory errors. Several neuroimaging studies support this interpretation. Greater activity in these areas has been found during speech when auditory feedback is altered (Toyomura et al., 2007; Tourville et al., 2008), and recently, using dynamic causal modeling, Parker Jones et al. (2013) found stronger endogenous connections from motor areas to auditory areas in nonnative (bilingual) speakers than native (monolinqual) speakers during an overt production task. In the DIVA model, error signals arising from these regions are used to fine-tune speech motor programs (Guenther et al., 2006; Tourville and Guenther, 2011). Thus, learning relies on the transmission of these signals to frontal regions involved in motor planning and execution. The correlation between learning success and white matter FA underlying pSTS supports this interpretation, revealing a potential physiological constraint on sensorimotor learning: reduced white matter integrity underlying pSTS may interfere with

[9] Although trials that involved phonemic errors were removed from the fMRI analysis, sub-phonemic variations (i.e., those that result in different auditory instantiations of the same phoneme) are likely to have resulted in some relatively poor pronunciations even in trials with no phonemic errors. Inter-trial variability has been shown to decrease with practice of speech movement sequences (Namasivayam & van Lieshout 2008); this is consistent with our finding of decreased activation with learning in auditory cortical areas thought to encode auditory errors.

the transmission of error signals to premotor regions, thereby hindering the formation and fine-tuning of speech motor programs.

3.4.4. Frontal operculum/anterior insula

Production of novel syllables also produced significantly greater activity in the frontal operculum and adjoining parts of the anterior insula (FO) bilaterally. FO has been implicated in monitoring auditory feedback during speech production (Hashimoto and Sakai, 2003; Cristoffels, et al., 2007) and is anatomically connected to both the posterior superior temporal cortex and lateral premotor cortex (Augustine, 1996; Saur et al., 2008; Axer et al., 2013; Cloutman, et al., 2012; Lemaire et al., 2012). Greater activity in this region during the production (Moser et al., 2009) and perception of novel speech sounds (Callan et al., 2004; Golestini and Zatorre, 2004; Raboyeau et al., 2004) compared to familiar speech sounds has been noted previously, consistent with the current findings. Furthermore, Golestini & Zatorre (2004) reported a correlation between activity in this region and the degree of success when learning novel phonetic contrasts, analogous to the correlation between FO activity and learning success found in the current study. Moreover, Golestani and Pallier (2007) found higher white matter density under the FO for speakers who were more successful at learning to produce a novel phoneme.

FO has also been associated with phonological processing, including translation of phonetic codes into articulatory scores (Dogil et al., 2002) and in phonological retrieval (Price, 1998). Combined, this previous work suggests that language-related sensory inputs are mapped to corresponding motor representations via phonological

representations encoded in FO. The novel sequences of speech sounds in the current study required new auditory-motor and orthography-to-motor mappings in the FO. The correlation between learning success and the reduction of activity in FO for the learned stimuli noted in this work suggests that speech motor sequence learning depends on the efficiently with which novel phonological representations and sensorimotor mappings are established.

3.4.5. Inferior temporal-occipital cortex and superior parietal lobules

Brain regions implicated in reading were also more active for *novel* than *learned illegal* syllables, including the bilateral superior parietal lobes (SPL) and left inferior temporal-occipital cortex (ITO). Both areas appear in neuroimaging studies of reading (Fiez & Peterson, 1998; Hagoort et al., 1999), and decreased activation in SPL and ITO is seen in both children and adults with dyslexia during reading compared to those with normal reading skills (Paulesu, et al., 2001; Peyrin et al., 2011; Richan, et al., 2009). However, these two brain areas appear to play different roles in the reading process

The SPL is implicated in shifting visual attention (Coull & Frith, 1998; Rushworth et al., 2001; Vandenberhe, et al., 2001) as well as shifting attention between auditory and visual stimuli (Shomstein & Yantis, 2004). The region has also been hypothesized to integrate sensory information across modalities. SPL shows activation for both visual and auditory stimuli and increased activity when these coincide (Calvert, et al., 2001; Degerman, et al., 2007). Intracranial recordings from SPL show that response to

concurrent auditory and visual stimuli was the algebraic sum of the responses to the unimodal stimuli presentations (Molholm, et al., 2006).

Our task involved both auditory and orthographic presentations of the syllable to be produced, and novel illegal syllables likely required prolonged attention to both auditory and orthographic stimuli as the cue-to-response association was probably weaker for *novel illegal* syllables than *learned illegal* syllables. Moreover, auditory presentation of illegal syllables has been shown to be perceptually identical to legalized, schwa-epenthesized pseudowords – e.g., LBIF perceived as LEBIF (Berent et al., 2006). Due to this reduced perceptual saliency of the acoustic presentation, subjects likely had to shift attention between the auditory and orthographic stimuli more for novel syllables due to increased reliance on the orthographic stimulus.

The ITO has also been implicated in reading, but with a different function. Patients with damage to this area can present with alexia, loss of reading ability (Leff, et al., 2001; Sakurai, et al., 2001). The region responds to visually presented words but not acoustically presented words (Cohen, et al., 2004), and responds more strongly to real letters compared to a false font (Price, et al., 1996). Repetition suppression occurs in the ITO even when an orthographic stimulus changes between upper- and lower-case (Dahaene, et al., 2001). This region is more active for reading pseudowords than words (Mechelli, et al., 2003). Because of these properties, many researchers have dubbed the ITO the “visual word form area” (Cohen & Daheane, 2004; McCandliss, et al., 2003). As in the SPL, we suggest that subjects in the current study had an increased reliance on the orthographic stimulus because of the reduced perceptual saliency of the acoustic

presentation. Moreover, given the novelty of the SSCs, subjects likely needed to "sound-out" the novel illegal visual stimuli as opposed to using a learned orthography-to-production relationship for learned illegal syllables. This difference is similar to the greater activation seen in the ITO for pseudowords than words (Mechelli, et al., 2003).

3.4.6. GODIVA simulations

Comparison of the GODIVA simulations with the fMRI *novel illegal – learned illegal* contrast demonstrates that the model correctly predicts increased activity in the preSMA, IFS, and vPMC for novel sequences. There are some activity differences not predicted by the GODIVA model, but these are outside of the scope of the model. For instance, the pSTS and PT activation differences are predicted by the DIVA model because of increased auditory errors for novel sequences. Additionally, ITO and SPL activation differences, which we hypothesize to reflect orthographic processing, are also outside the scope of the GODIVA model.

The increased activity for *novel illegal* productions in the bilateral FO is not predicted by either the GODIVA or DIVA models. We have hypothesized that the FO maps language-related sensory inputs to corresponding motor representations via phonological representations. This function is not in the scope of the either model; both begin planning an utterance after it has been chosen. The DIVA model does, however, include the auditory state and SSM modules, which we hypothesize to provide information about auditory and motor representations to the FO.

Beyond these omissions, there is another discrepancy between these data and the GODIVA model. The preSMA activation difference between *novel illegal – learned illegal* productions was right lateralized in our data, but left-lateralized in the GODIVA simulations. We have previously discussed the literature implicating bilateral pre-SMA activation for speech and non-speech motor sequence learning tasks. We suggest that the GODIVA model additionally assign the right preSMA to the module.

Importantly, these simulations quantitatively demonstrate that producing an utterance with fewer motor chunks results in less fMRI activity in keeping with the general theory that learned speech motor sequences can be produced more efficiently with chunking.

3.4.7. Summary

In summary, our results demonstrated behavioral improvements due to speech motor sequence learning and identified the network of brain regions involved in this process. Learning resulted in reduced activity in speech-specific frontal and posterior superior temporal cortex as well as brain regions known to be involved more generally in motor sequence planning and execution. Reduced activity throughout the motor sequence learning network supports the notion that motor sequence learning involves the merger of individual motor programs into larger units that allows the motor system to rely on fewer, larger motor programs. A significant correlation was found between learning success and activity in FO, supporting the view that motor sequence learning relies on mapping sensory representations of novel speech sound sequences to the motor system via phonological representations in FO. White matter FA underlying pSTS was also

significantly correlated with learning success, indicating that white matter integrity within the speech motor sequence learning network modulates learning by constraining the efficiency of sensory-to-motor signal transmission.

4. REPRESENTATION OF FRAME AND CONTENT IN THE BRAIN

4.1. Introduction

As previously discussed in Chapter 1.3, both the slot/filler theory (Shattuck-Hufnagel, 1983) and the frame/content theory (MacNeilage, 1998) propose that a speaker separately selects the phonological content and the syllabic frame structure of an intended utterance and only later merges these representations to produce a syllable. These models imply dissociable neural representations of a syllable, its frame structure, and its phonological content. Although strongly supported by the structured nature of speech errors (Shattuck-Hufnagel, 1983; Trieman & Danis, 1988) and advocated in several models of speech production (e.g., Bohland et al., 2010; Roelofs, 1997), this concept currently lacks direct neural evidence.

The slot/filler and frame/content theories posit three representations of speech in the brain: phonological, syllabic frame, and syllabic. There is direct neural evidence for syllabic representations in the ventral lateral premotor cortex (Peeva et al., 2010). There is also neural evidence for phonemic representations in the posterior superior temporal sulcus (Graves et al., 2008; Vaden et al., 2010), but it is unclear if phonological filler elements are phonemes or some other phonological representation. Computational models of speech that implement the slot/filler theory – the GODIVA model, the WEAVER++ model, and to some extent, the OSCAR model – use phonemes as the phonological filler elements (see Chapters 1.3.4 and 1.4). However, based on speech error data, Shattuck-Hufnagel (1986) suggested that subsyllabic constituents (SSCs) are the filler elements. Data from word games, language development, and from

phonotactic constraints also suggest that SSCs are important units of speech representation (see Chapter 1.2). However, no neural regions have yet been identified that process SSC representations.

To this date, there is also no direct evidence for syllabic frame structure representations in the brain. Several researchers have hypothesized that the presupplementary motor area (preSMA) is responsible for storing and selecting syllabic frame structures (Bohland et al., 2010; MacNeilage, 1998; Roelofs, 1997) based on neuroimaging and clinical evidence. FMRI reveals that the preSMA is more active during movement preparation than execution (Lee, et al., 1999), and more active for utterances containing more syllables (Ghosh, et al., 2008). In an fMRI study of syllable sequence production, activity in the pre-SMA was one of the only brain regions whose activity was modulated by syllable frame complexity (e.g., more active for STRA than RA; Bohland & Guenther, 2006)[10] although this manipulation was also confounded by syllable length. Moreover, when the preSMA is damaged by stroke, patients may spontaneously produce repetitive consonant-vowel (CV) syllable strings such as LALALA (Jonas, 1981) [11]. MacNeilage (1998) hypothesized that these types of utterances represent "pure frame" representations: syllabic frames produced without regard to phonological content (see Chapter 1.3.2 for more details). While there have been limited attempts to directly localize syllabic frame representations in the brain, it appears that the existing neural

[10] This activity pattern implies that larger, more complex frames require more wide-spread neural representations or longer, sustained activations. It is unclear if this is the behavior one would expect for the representations of syllabic frames. It is also possible that the selection of any frame, regardless of its complexity, would require the same amount of neural activity.
[11] Not all older studies distinguish between the more posterior supplementary motor area-proper and the preSMA.

evidence points to the preSMA as a brain region involved in frame storage and selection for speech sequence production.

The present study sought to identify specific brain regions involved in representing syllabic frame structures, phonemes, SSCs, and complete syllables using a functional magnetic resonance imaging repetition suppression (fMRI-RS) paradigm. Experimental conditions were constructed to vary by the amount of repetition of each speech representation of interest. We then defined expected cross-condition patterns of fMRI-RS for each hypothesized speech representation. For each anatomically-defined region of interest, we compared the cross-condition activity pattern to the predicted patterns in order to find regions that significantly matched expected patterns.

4.2. Methods

4.2.1. Participants

Seventeen right-handed native speakers (9 female, aged 20-43 years, mean 29.5 years) of American English participated. All subjects reported normal or corrected-to-normal vision and no history of hearing, speech, language, or neurological disorders. Informed consent was obtained according to the Boston University Institutional Review Board and the Massachusetts General Hospital Human Research Committee.

4.2.2. Stimuli

Stimuli were monosyllabic pseudowords with the frame structures CCV, CVC, CV, or VC. Stimuli were phonotactically legal in American English. The number of phonemes per pseudoword and the number of letters per orthographic stimulus were balanced across conditions. None of the syllables formed an orthographic or a phonological word found in the MRC Psycholinguistic Database (Coltheart, 1981). Syllable frequency, as reported in the MRC Psycholinguistic Database, was not significantly different between conditions (ANOVA, $F_{(3,92)} = 1.02$, $p = 0.39$, n.s.)

Auditory stimuli were recorded with 32 bit sound at 4.41 kHz over a Samson C01U USB studio condenser microphone using Audacity software (http://audacity.sourceforge.net/). The speaker was a native speaker of American English. Using the same software, selected sequences were normalized for intensity. Stimuli were chosen to maintain similar F0 across all sequences. The durations were adjusted to a constant length of 500 ms using PRAAT software that changes duration without changing F0 (Boersma & Weenink, 2007). Maximally, changes were < 10% of the total original length.

4.2.3. Paradigm

In each experimental block, subjects spoke pairs of pseudowords (Table 4.1). Blocks fell into four conditions that differed according to how often each type of speech representation – syllabic frame, phoneme, or complete syllable – was repeated between the pairs of pseudowords. In the *all same* condition, subjects produced the same pseudoword for all repetitions. In the *different phonemes only* condition, subjects

alternated between two pseudowords containing different phonemes but the same syllabic frames, (e.g., FAS and REEN, which both have CVC frames). In the *different frames only* condition, subjects alternated between two pseudowords containing the same phonemes but different frames, (e.g., GREE and REEG, which have CCV and CVC frames, respectively). In the *all different* condition, subjects alternated between pseudowords containing different phonemes and frames (e.g., DEEF and AP, which have CVC and VC frames respectively). For a full list of stimuli, see Table 4.2.

All same	**Different phonemes only**	**Different frames only**	**All different**
Same phonemes **Same** frames	**Different** phonemes **Same** frames	**Same** phonemes **Different** frames	**Different** phonemes **Different** frames
TWAI	**FAS**	**RAUD**	**DEEF**
TWAI	**REEN**	**DRAU**	**GLAI**
TWAI	**FAS**	**RAUD**	**DEEF**
TWAI	**REEN**	**DRAU**	**GLAI**
TWAI	**FAS**	**RAUD**	**DEEF**
TWAI	**REEN**	**DRAU**	**GLAI**

Table 4.1. The four experimental speaking conditions in the fMRI study. Each box represents the orthographic and auditory presentation of the pseudoword to the subject in the scanner. In the *all same* condition, subjects produced the same pseudoword for all repetitions. In the *different phonemes only* condition, subjects alternated between two pseudowords containing different phonemes but the same syllabic frames. In the *different frames only* condition, subjects alternated between two pseudowords containing the same phonemes but different frames. In the *all different* condition, subjects alternated between pseudowords containing different phonemes and frames**.**

All same	Different phonemes only		Different frames only		All different	
KUS	FAS	REEN	RAUD	DRAU	DEEF	GLAI
TWAI	KEID	ROS	REEG	GREE	PRAA	SMO
DAA	LEEZ	PAAK	LEEB	BLEE	WAA	AUN
AIT	GREE	SMU	WEES	SWEE	OING	ZO
BAUN	STAA	PLU	NAAS	SNAA	MAAT	FAU
PLAA	KLO	THRAI	FRAA	RAAF	NAI	KOS
SHAA	VU	NAA	ZEI	EIZ	SNEE	AANG
OWP	KEI	LAU	GEE	EEG	AUZ	TRAA
FEEM	ZAI	GOI	FU	UF	FLAU	ZOI
GLAU	EEN	AASH	VAA	AAV	VOI	SKU
LOI	EIN	AUD	KAA	AAK	VAAD	MOI
AAS	EEK	AAM	BAA	AAB	OS	LAAF
MAIZ						
KLEE						
TEI						
GRA						
POIN						
SLAU						
TAA						
AAM						
KAUN						
EES						
KO						

Table 4.2. Orthographic representations of stimuli used in all speaking conditions. Subjects repeated the pseudoword six times per block in the *all same* condition. Subjects alternated between the pairs of pseudowords three times, for a total of six utterances, per block for all other conditions.

The *all same* condition presented one pseudoword per block while all other speaking conditions presented pairs of pseudowords. In order to maintain equal novelty across

conditions, twice as many stimuli were used during the *all same* conditions compared to other conditions. In this way, the same number of pseudowords was presented in each speaking condition throughout the experiment

During a single trial, subjects saw an orthographic representation of the pseudoword for 900 ms and simultaneously heard the 500 ms auditory stimulus. Then, a white cross replaced the orthographic stimulus, cueing the subjects to produce the target pseudoword. A *baseline* condition was also intermixed in which subjects saw a series of asterisks on the screen instead of the orthographic stimulus and rested quietly instead of producing a pseudoword. Blocks lasted 15 s, and consisted of six 2 s trials followed by a 3 s pause so one block's effects would not confound the next. Runs consisted of eighteen blocks and lasted approximately 4.5 min. Pseudowords and conditions were randomized within runs. Each pseudoword or pseudoword pair was maximally used once per block and in 2-3 blocks throughout the experiment to maintain novelty. Each subject completed 6 runs that optimally allowed for approximately 27 blocks per condition per subject.

Instructions and visual stimuli were projected onto a screen that subjects could view from within the scanner via a mirror attached to the head coil. Auditory stimuli were played over Sensimetrics MRI-compatible insert headphones model S-14.

4.2.4. Image acquisition

MRI data were acquired using a 3 Tesla Siemens Trio Tim scanner with a 32 channel head coil. For each subject, a high-resolution T1-weighted volume was acquired (MPRAGE, voxel size: 1 mm^3, 256 sagittal images, TR: 2530 ms, TE: 3.44 ms, flip angle: 7°). Functional gradient echo EPI scans (41 horizontal slices, in plane resolution: 3.1 mm, slice thickness: 3 mm, gap: 25%, TR: 2.5s, TA: 2.5 s, TE: 20 ms) were automatically registered to the AC-PC line and were collected continuously.

4.2.5. FMRI data analysis

FMRI data preprocessing was conducted as described in Chapter 3.2.8. Blood oxygen level dependent (BOLD) responses were estimated using a general linear model (GLM), and the hemodynamic response function (HRF) for each stimulus block was modeled with the canonical HRF in SPM8.

Blocks consisted of 6 pseudoword repetitions. The first 3 repetitions (early component) and second 3 repetitions (late component) of each task block were modeled separately. Only the early component is presented in subsequent results. We found that when the early and late components were modeled together, the observed progression of brain activity over time did not fit well with a standard hemodynamic response (Fig. 4.1, “early + late”, A. Nieto-Castañón, personal communication, November 30, 2012). When the early component was modeled separately, the results not only had a better fit with a standard hemodynamic response function (Fig 4.1, “early only”), but also showed

greater statistical power as measured by a larger percent change from baseline (Fig 4.2).

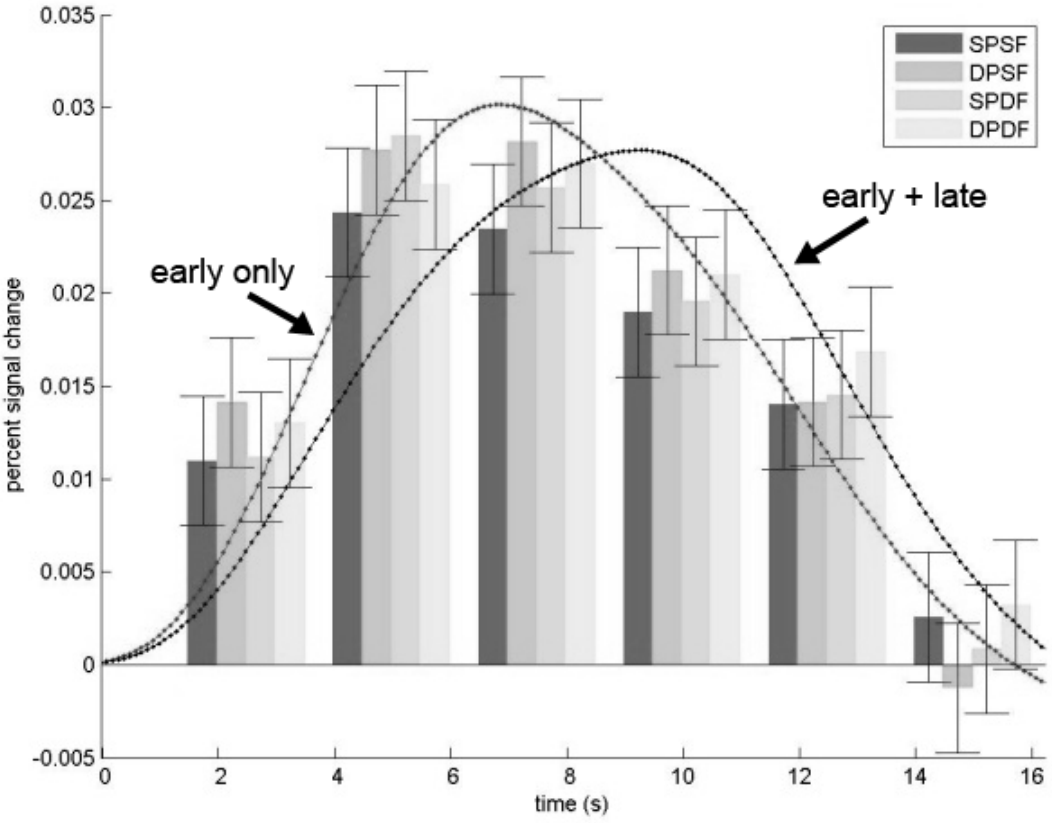


Figure 4.1. Bars show percent signal change for the first half of the task block averaged across all ROIs. Bars are grouped by volume's time of acquisition from stimulus onset (in seconds). The dotted line shows the modeled BOLD activity of whole block (fMRI activity convolved with the SPM's hemodynamic response function). The lines show modeled BOLD responses the "early only" and "early and late" components of the block. Abbreviations: SPSF = *all same* condition, DPSF = *different phonemes only* condition, SPDF = *different frames only* condition, DPDF = *all different* condition. (A. Nieto-Castañón, personal communication, November 30, 2012.)

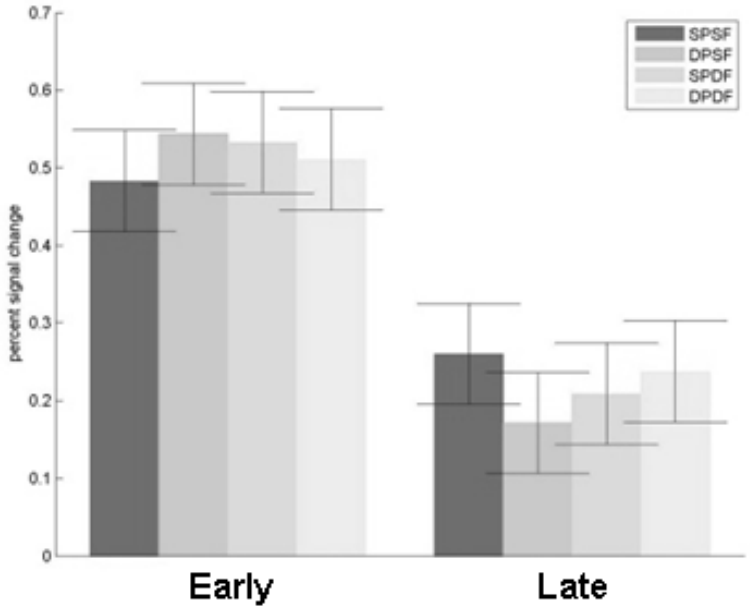


Figure 4.2. Percent signal change for the first half of the task block (early component) and second half (late component) averaged across all ROIs. See Figure 4.1 for abbreviations. (A. Nieto-Castañón, personal communication, November 30, 2012.)

The model included 10 condition-specific variables – the early and late components of the *all same*, *different frames only*, *different phonemes only*, *all different*, and *baseline* conditions – and additional covariates: linear detrending covariates and motion parameters. The model was estimated separately for each subject. Model estimates for each speaking condition were contrasted with the baseline condition. *All speaking* conditions were collapsed and also contrasted with baseline.

4.2.5.1. Voxel-wise analysis

Surface- and volume-based voxel-wise analysis was used to assess group blood oxygen level dependent (BOLD) response differences in the *all speaking - baseline* condition in the cerebral cortex using the same methods described in Chapter 3.2.8.

4.2.5.2. ROI-wise analysis

A region of interest (ROI) analysis was also performed for greater statistical power and better alignment of neuroanatomical regions across subjects (Nieto, et al., 2003). A set of ROIs were chosen that are reliably engaged during neuroimaging of speech tasks (Bohland & Guenther, 2006; Brown, et al., 2005; Indefrey & Levelt, 2004; Guenther et al., 2006; Turkeltaub, et al., 2002). These areas include the primary motor and somatosensory cortices, ventral and middle premotor cortex, inferior frontal cortex, superior temporal cortex, anterior insula, basal ganglia, thalamus, and cerebellum. ROIs parcellations were based on individual subjects' high-resolution anatomical MRI volumes, Cortical ROIs were parcellated using a FreeSurfer classifier (Fischl et al., 2004) trained on an atlas tailored to speech studies (Tourville & Guenther, 2003). Subcortical ROI were parcellated using a FreeSurfer classifier using the FreeSurfer subcortical training set (Fischl, et al., 2002). Cerebellar ROIs were parcellated with the SUIT atlas and software (Diedrichsen, 2006; Diedrichsen et al., 2009).

ROI-wise fMRI analyses were conducted using the REX toolbox (http://web.mit.edu/swg/rex/). Within each ROI, the data was whitened to compensate for temporal noise correlations. The GLM was estimated based on the same model

used in the voxel-wise analysis (Chapter 4.2.5.1), and the contrasts were calculated for the *all speaking - baseline* and for each speaking condition compared to *baseline* in the same manner as described in the voxel-wise analysis. Mean activity differences within each ROI were extracted for each contrast for each subject. Group-level t-statistics were calculated at each ROI for each contrast and thresholded at $p_{Fwe} < 0.05$ (corrected).

4.2.5.3. Across-condition activity pattern matching

Experimental conditions were constructed to differ by the amount of repetition of each speech representation. To quantitatively assess fMRI-RS across these conditions, we constructed models of the expected patterns of fMRI-RS for each hypothesized speech representation (Table 4.3). For ROIs encoding phonemic representations dissociated from syllable structure, we expected fMRI-RS to be greater (i.e. less fMRI activity compared to baseline) for the *all same* and *different frames only* conditions – which contain the same set of phonemes across pseudowords in a block – than the *different phonemes only* and *all differen*t conditions. For ROIs encoding SSC representations, we expected fMRI-RS to be greatest for the *all same* condition, weaker for the *different frames only* condition – which contain the same nuclei, but different onsets and codas – and weakest for the *different phonemes only* and *all different* conditions. For ROIs encoding syllabic frame representations dissociated from phonological content, we expected fMRI-RS to be greater for the *all same* and *different phonemes only* conditions – which contain the same syllabic frame in both repeated pseudowords – than the *different frames only* and *all different* conditions. For ROIs encoding representations of

full syllables, we expected fMRI-RS to be greater for the *all same* condition compared to all other speaking conditions. For ROIs that are insensitive to phonemic, SSC, syllabic frame, and syllabic representations, we expected equal fMRI-RS across all conditions

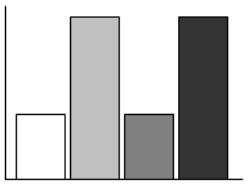
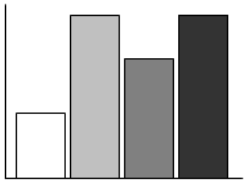
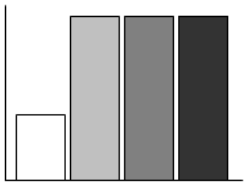
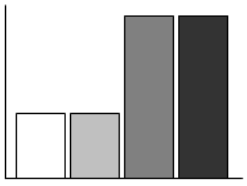
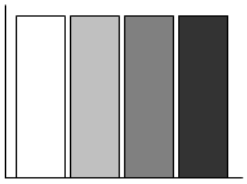
Speech representation	Expected pattern of fMRI-RS
Phoneme	
SSC	
Syllable	
Syllabic frame	
Phonologically insensitive	

Table 4.3. Predicted across-condition activity models for the phonemic, subsyllabic constituent (SSC), syllabic, syllabic frame, and phonologically insensitive speech representations. The bar plots represent the hypothesized relative fMRI activity for the (from left to right) *all same*, *different phonemes only*, *different frames only*, and *all different* conditions compared to baseline.

In a previous paper with similar methodology (Peeva, et al., 2010), the authors also included a *difficulty* model. However, this model is not used in the present analysis. Peeva et al. measured difficulty based on the number of repeated elements– phoneme, syllable, etc. –within a condition; more repeated elements corresponded to lower difficulty. However, for the current study, there is little direct neural evidence for representations of SSCs and syllabic frames. If these representations do not exist in the brain, it would be inaccurate to include them in constructing a difficulty model. However, if they do exist in the brain, it would be inaccurate to exclude them. Therefore, a difficulty model was not included in this analysis because there is no clear pattern with which to model this representation.

To quantify the strength of the match between the across-condition fMRI activity patterns and hypothesized models, we followed the method described by Peeva, et al. (2010). Each hypothesized model was characterized by six comparisons, one for each possible pairing between the four speaking conditions. For instance, in an area processing phonemic representations, we expected no statistical difference between the fMRI activity of the *all same* and *different frame only*, but we expected less activity in the *all same* activity than in the *all different* activity. For each ROI, paired t-tests compared the mean activity differences across subjects for each of the 6 comparisons. For the j-th comparison in the i-th model, p was defined as

$p_{ij} = p$ if A < B in the model,

$p_{ij} = 1 - p$, if A > B in the model, or

$$p_{ij} = 2 \cdot \min\left(p, 1-p\right)$$ if A = B in the model,

where p is the p-value from a one-sided t-test evaluating the fMRI activity of conditions A and B. Then, the fit of all comparisons to those in the i-th model was quantified as:

$$\lambda_i = \min_j p_{ij}$$

for each model. To calculate statistical thresholds, a Monte Carlo simulation approximated the expected distribution of λ values using Gaussian random noise with means equal to each model's hypothesized across-condition activity patterns. Noise means were scaled by a factor equal to the 95th percentile of the average observed BOLD data's between condition differences across all ROIs. Noise variances matched the between subject variance of BOLD responses across all ROIs and conditions. 10,000 Monte Carlo simulations were run. A model-level p-value, P_i, was calculated from this distribution such that

$$P_i = \frac{1}{m-1} \sum_{k \neq i} prob\left(\lambda_i \geq \bar{\lambda_i} \mid M_k\right)$$

where M_k ($i \neq k$) is one of m alternative models. ROIs that fit models with $P < 0.05$ are reported as significant.

4.3. Results

4.3.1. Voxel-wise analysis

Figure 4.3 and Table 4.4 show the cortical brain regions that were significantly more active for all speaking conditions than *baseline* is a voxel-wise surface-based analysis (voxel-level $p < 0.001$, uncorrected; cluster-level $p < 0.0167$, corrected). Speaking

resulted in bilateral clusters of greater BOLD response in the posterior superior temporal cortex including Heschl's gyrus, planum temporale, and planum polare, and extending into the anterior superior temporal gyrus. Bilateral clusters of greater activity for speech tasks as also seen in the ventral and middle portions of the pre- and post-central cortex extending into the middle premotor cortex. A cluster in the left medial premotor cortex included the presupplementary motor area and supplementary motor area, but an analogous cluster in the right hemisphere included only the presupplementary motor area. Clusters in the left inferior temporal-occipital lobe, superior parietal lobule, and anterior middle frontal gyrus were also active in the contrast, as well as, bilateral clusters in the inferior frontal gyrus.

Figure 4.3 and Table 4.4 also show the subcortical brain regions that were significantly more active for all speaking conditions than *baseline* in a voxel-wise volume-based analysis (voxel-level $p < 0.001$, uncorrected; cluster-level $p < 0.0167$, corrected). Speaking resulted in a subcortical cluster extending bilaterally into the basal ganglia – including the caudate, putamen, and pallidum – and thalamus. A second cluster extended bilaterally into the anterior and lateral aspects of the cerebellum and throughout the vermis.

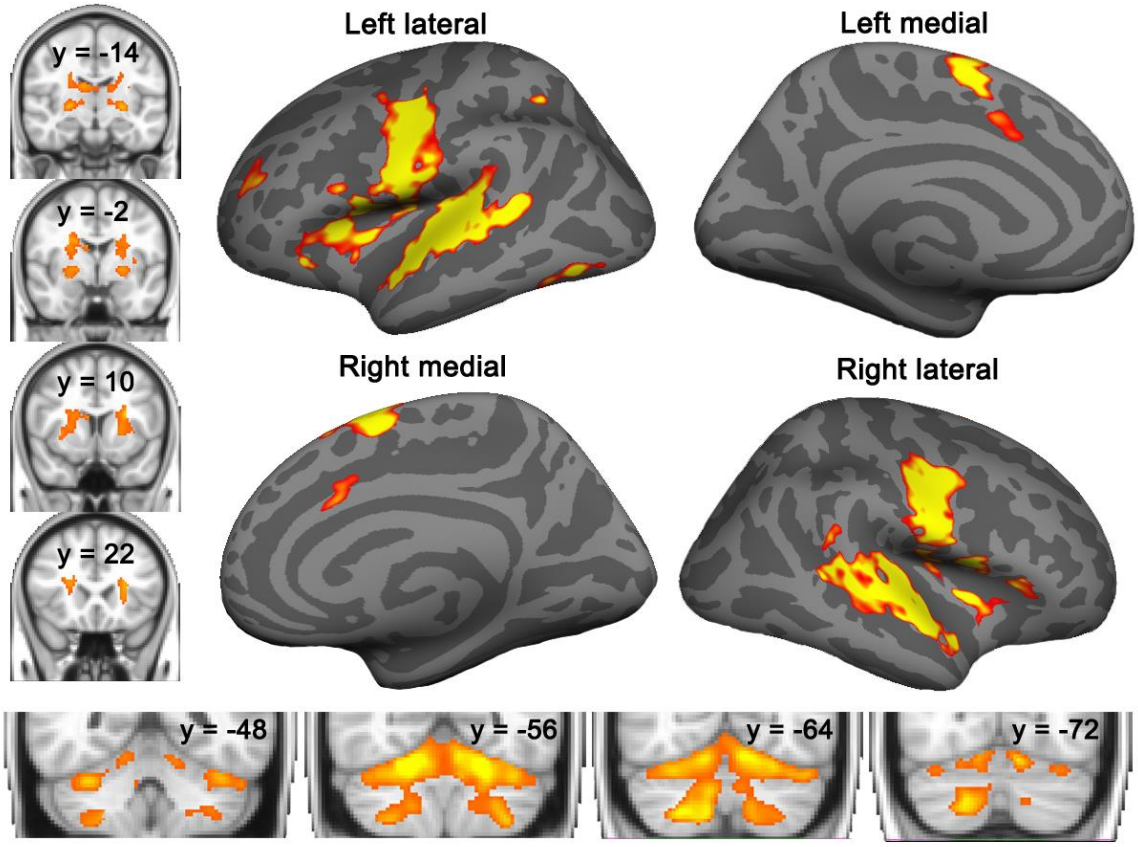


Figure 4.3. FMRI main effects of speaking (*all speaking* > *baseline*) in voxel-wise analysis. Significant clusters are shown on the left lateral (*upper left*), right lateral (*upper right*), and right medial (*lower right*) inflated surface representations of the FreeSurfer average template. Subcortical activity is shown on a series of coronal slices from the MNI305 template at the level of the subcortical nuclei (*left*) and cerebellum (*bottom*) y coordinate indicates mm distance from the anterior commissure in MNI space). Contrast volumes were first voxel thresholded at $p < 0.001$, then cluster thresholded at cluster-wise p (CWP) < 0.0167 to correct for three analyses: subcortical and 2 cortical hemispheres. CWP was calculated by separate Monte Carlo simulations for each of the three analyses.

Region name		Talairach coordinates			t	Size	CWP
		x	y	z			
Left	Planum temporale	-59.4	-16.5	2.9	14.3	2828 mm^2	0.0001
	Mid motor cortex	-43.3	-9.9	41.7	11.0	3633 mm^2	0.0001
	SMA	-6.3	-1.1	53.8	9.2	441 mm^2	0.0001
	aMFG	-36.3	43.1	17.4	4.6	293 mm^2	0.0001
	ITO	-40.3	-61.3	-1.5	8.2	225 mm^2	0.0131
	Superior parietal lobe	-27.2	-47.4	44.6	5.5	64 mm^2	0.0092
	IFo	-46.5	12.8	18.2	4.8	87 mm^2	0.0016
	pre-SMA	-10.6	11.3	36.5	4.1	99 mm^2	0.0006
Right	Heschl's gyrus	47.2	-20.5	7.7	11.5	2642 mm^2	0.0001
	Mid premotor cortex	52.2	0.5	41.0	10.2	1847 mm^2	0.0001
	Pre-SMA	6.6	7.2	58.1	8.9	497 mm^2	0.0001
	Anterior central operculum	47.1	6.7	3.5	9.6	318 mm^2	0.0001
	Anterior central operculum	44.3	-5.0	14.0	5.7	226 mm^2	0.0001
	Anterior insula	34.7	5.1	2.6	10.6	204 mm^2	0.0001
	Anterior frontal operculum	39.5	25.5	5.7	4.9	134 mm^2	0.0001
	Middle cingulate gyrus	6.2	17.1	29.2	4.2	114 mm^2	0.0002
	pSTG	57.6	-40.9	18.1	4.6	100 mm^2	0.0002
Bilat.	Cerebellar cortex	-22.8	-53.0	-46.9	9.7	47704 mm^3	0.0001
	Basal ganglia/thalamus	-26.7	-19.3	-17.5	6.5	27872 mm^3	0.0001

Table 4.4. Summary of cortical and subcortical activation peaks for the main effect of speaking (all speaking – baseline). Activation peaks for subcortical activation were both in the left hemisphere, but the clusters extended bilaterally. From left to right, the columns show the region name, Talairach coordinates, T value, cluster size, and cluster-wise p (CWP). Abbreviations: Bilat = Bilateral, SMA = supplementary motor cortex, IFo = inferior frontal gyrus, pars opercularis, aMFG = anterior middle frontal gyrus, preSMA = presupplementary motor cortex, pSTG = posterior superior temporal gyrus.

4.3.2. ROI-wise analysis

Table 4.5 and 4.6 show the cortical and subcortical brain regions that were significantly more active for all speaking conditions (*collapsed*, *all same*, *different phonemes only*, *different frames only*, and *all different*) than *baseline* in an ROI-wise analysis.

ROI Abbrev.		fMRI contrast compared to baseline				
		Collapsed	**All same**	**Diff phon**	**Diff frame**	**All diff**
Rolandic cortex						
Left	vPMC	X	X		X	X
	midPMC	X	X	X	X	X
	vMC	X	X	X	X	X
	SMA	X				
	preSMA	X				
Right	vPMC	X	X			
	midPMC	X	X	X	X	X
	vMC	X	X	X	X	X
	SMA	X				
	preSMA	X	X			
Frontal cortex						
Left	aIFs					
	pIFs	X				
	dIFo	X				
	vIFo	X	X			
	dIFt					
	vIFt				X	
	aFO	X				
	pFO	X	X			
Right	aIFs		X			
	pIFs					
	dIFo	X				
	vIFo					

	dIFt					
	vIFt					
	aFO					
	pFO	X				
Temporal cortex						
Left	aSTg	X	X	X	X	X
	pSTg	X	X	X	X	X
	adSTs					
	pdSTs	X	X	X	X	X
	H	X	X	X	X	X
	pCO	X	X	X	X	X
	PO	X	X		X	X
	PP	X	X		X	X
Right	aSTg	X	X	X	X	X
	pSTg	X	X	X	X	X
	adSTs	X				
	pdSTs	X	X	X	X	X
	H	X	X	X	X	X
	pCO	X				X
	PO					
	PP	X	X	X	X	X
	PT	X	X	X	X	X
Insular cortex						
Left	aINS	X				
Right	aINS					

Table 4.5. Cortical ROI-wise results of the all speaking conditions collapsed and individual speaking conditions (*all same*, *different phonemes only*, *different frame only*, and *all different*) compared to baseline. X indicates a significance of $p_{FDR} < 0.05$. Abbreviations: vPMC = ventral premotor cortex, midPMC = middle premotor cortex, vMC = ventral primary motor cortex, SMA = supplementary motor area, preSMA = presupplementary motor area, aIFs = anterior inferior frontal sulcus, pIFs = posterior inferior frontal sulcus, dIFo = dorsal inferior frontal gyrus, pars opercularis, vIFo = ventral inferior frontal gyrus, pars opercularis, dIFt = dorsal inferior frontal gyrus, pars triangularis, vIFt = ventral inferior frontal gyrus, pars triangularis, aFO = anterior frontal operculum, pFO = posterior frontal operculum, aSTG = anterior superior temporal gyrus, pSTg = posterior superior temporal gyrus, adSTs = anterior dorsal temporal sulcus, pdSTs = posterior dorsal temporal sulcus, H = Heschl's gyrus, pCO = posterior central operculum, PO = partietal operculum, PP = planum polare, PT = planum temporale, aINS = anterior insula.

ROI Abbrev.		fMRI contrast compared to baseline				
		Collapsed	**All same**	**Diff phon**	**Diff frame**	**All diff**
Subcortical nuclei						
Left	Thal	X				
	Caud	X				
	Put	X				
	Pall	X				X
Right	Thal	X				
	Caud	X				
	Put	X				
	Pall	X	X	X	X	X
Cerebellar cortex						
Left	I-IV	X	X	X	X	X
	V	X	X	X	X	X
	VI	X	X	X	X	X
	CrusI				X	
	CrusII	X	X			
	VIIb				X	
	VIIIa			X	X	X
	VIIIb					
	IX					
	X	X				
Right	I-IV					X
	V	X	X	X	X	X
	VI	X	X	X	X	X
	CrusI				X	

	CrusII	X				
	VIIb					
	VIIIa	X	X	X	X	X
	VIIIb					
	IX					
	X	X				
Vermis	VI	X	X	X	X	X
	CrusII	X	X	X	X	X
	VIIb	X				
	VIIIa	X			X	X
	VIIIb	X			X	
	IX	X				
	X					

Table 4.6. Subcortical ROI-wise results of the all speaking conditions collapsed and individual speaking conditions (*all same*, *different phonemes only*, *different frames only*, and *all different*) compared to baseline. X indicates a significance of $p_{FDR} < 0.05$. Cerebellar ROI abbreviations denote lobules. Abbreviations: Thal = thalamus, Caud = caudate, Put = putamen, Pall = pallidum.

4.3.3. Across condition activity pattern-matching analysis

Figure 4.4 shows cortical and subcortical regions that significantly ($P < 0.05$) matched a hypothesized speech representation model. The right posterior superior temporal sulcus, left supplementary motor area, and right anterior cerebellum (lobules I-IV and V) matched the phonemic representation model. The left posterior inferior frontal sulcus and dorsal inferior frontal gyrus, pars opercularis matched the SSC representation

model. The right cerebellar lobule VIIb matched the syllabic frame structure representation model. The left ventral and middle premotor cortex, bilateral cerebellar lobules VI and right lobule VIIIa matched the full syllable representation model. The left anterior frontal operculum and right supplementary motor area matched the phonologically insensitive representation model.

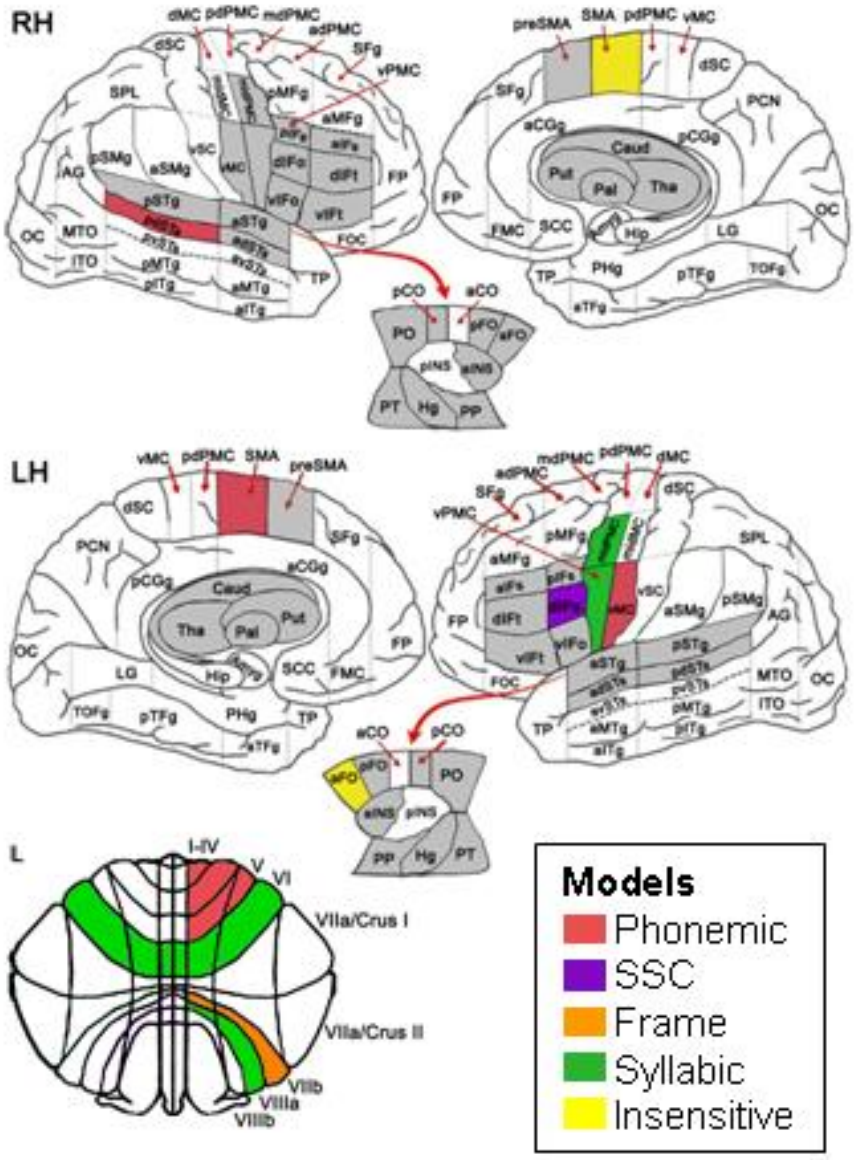


Figure 4.4. ROI brain map. Colors show significant ($P < 0.05$) matches to predicted speech representation models. Grey color indicates the ROI was included in the analysis, but did not significantly match a speech representation model. See tables 4.5 and 4.6 for definitions of the abbreviations.

4.4. Discussion

We compared the neural activity for speaking tasks to a silent baseline task and found greater activity for speech across both brain hemispheres in the primary motor and somatosensory cortices, the lateral and medial premotor cortices, superior temporal cortex, basal ganglia, and cerebellum. These regions have all been implicated in the "minimal network" needed for speech motor control (Bohland & Guenther, 2006). Other brain areas with greater activity for speech tasks include the left inferior temporal-occipital cortex and superior parietal lobules. These areas have previously been implicated in processing orthographic stimuli for phonological output and integrating cross-modal stimuli (see Chapter 3.4.5). Subjects in our study saw and heard representations of the target utterance, and it is not surprising for these activations to occur for this task. We also found that voxel- and ROI-wise measures were consistent. The measures diverge only in the sense that the activity within an ROI reflects the average activity across its voxels, but the voxel-wise analysis is not constrained in this way. For instance, a very small portion of a significant voxel-wise cluster verges into the left vIFt, but the vIFt is not significant in the ROI-wise results because most the voxels in the ROI are not significantly active in the contrast.

Relative patterns of across-condition fMRI activity were matched to hypothesized patterns of fMRI-RS for phonemic, subsyllabic constituent (SSC), syllabic frame, syllabic, and phonologically insensitive speech representations in a variety of speech-related ROIs.

4.4.1. Lateral prefrontal cortex

The lateral prefrontal cortex was more active for all speech tasks compared to *baseline*, and prefrontal regions matched two different models of speech representations. The left inferior frontal sulcus and inferior frontal gyrus, pars opercularis matched the subsyllabic constituent (SSC) representation model. The left ventral and middle premotor cortex activity matched the syllabic representation model.

This finding in the left premotor cortex is in accordance with the results of Peeva et al. (2010)[12]. According to the Directions into Velocities of Articulators (DIVA) model (Guenther et al., 2006; Tourville and Guenther, 2011), this area contains a speech sound map that stores feedforward motor programs for frequently used speech targets. The speech sound map is hypothesized to contain a syllabary that encodes these motor programs for frequently used syllables. Similarly, the WEAVER model (Indfrey & Levelt, 2004) suggests that the premotor cortex is involved in articulating these syllabic motor programs, although it is unclear if the motor programs are also stored in this area. Both the retrieval and articulation of syllabic motor programs would likely result in activating a syllabic representation of speech in the left ventral and middle premotor cortex.

In contrast to the syllabic representation in the lateral premotor cortex, the inferior frontal gyrus, pars opercularis (IFo) matched the SSC representation model. In the DIVA model, this region has also been proposed to also be a part of the speech sound map.

[12] Note that Peeva et al. found only the left ventral premotor cortex's activity matched a syllabic representation. They, however, used a 2-part dorsal-ventral partition of the lateral premotor cortex as opposed to the 3-part dorsal-middle-ventral partition used here.

We suggest that the IFo stores feedforward motor programs for SSC utterances, and propose two hypotheses for the use of these motor programs. First, SSCs may be an intermediate representation in the formation of syllabic motor programs[13]. In this hypothesis, infrequent syllables – which do not have stored syllabic motor programs in the lateral premotor syllabary – must be assembled into syllabic motor programs from smaller SSC programs before execution. We see syllabic and SSC representations in different areas of the speech sound map because the SSC representations are needed to synthesize the syllabic representations.

Alternatively, the speech sound map might directly output both SSC and syllabic motor programs. The speech network might use the largest available encoded articulatory program. Preferably, a syllabic motor program is used, but if it is not pre-stored in the syllabary in the lateral premotor cortex, SSC programs are articulated from programs stored in IFo. In the present study, the stimuli used represent a range of syllabic frequencies, and therefore some stimuli may have corresponding syllabic motor programs stored in the syllabary, but others may not. The SSC and syllabic representations in different areas of the prefrontal cortex may reflect the different strategies used for stimuli of varying syllabic frequencies.

While these results concur with the DIVA model, they contradict the WEAVER model, in part (Levelt & Roelofs, 2004). The WEAVER model hypothesizes that wordforms are

[13] The WEAVER model similarly hypothesizes that SSCs are an intermediate representation for assembling syllabic motor programs. However, this model is intended to explain the production of lexical items from memory, and the SSC representations are used to decompose wordforms stored in memory into phonemes. Moreover, the function is broadly assigned to “Wernicke’s area” that can cover many neuroanatomical regions in the frontal cortex.

syllabified in the inferior frontal gyrus (see footnote 1 of Chapter 3) and the corresponding syllabic motor programs are articulated using the lateral premotor cortex.[14] Our finding of syllabic representation in the lateral premotor cortex is in accordance with this model, but, assuming that syllabification processes would result in a syllabic representation, our finding of a SSC representation in the IFo appears to diverge from the function hypothesized by the WEAVER model.

The activity in the left inferior frontal sulcus (IFs) also matched the SSC representation model. As previously discussed, the IFs is implicated in phonological and verbal working memory (see Chapter 3.4.1). The GODIVA model, which implements the slot/filler and frame/content theories, receives the intended phonological content and syllabic frame for an intended utterance in parallel. The IFs is proposed to receive the phonological content input and select the appropriate phonemes of a target syllable for execution. These results suggest, however, that the phonological content of an utterance is not stored or selected by individual phonemes, but instead by SSC units. The "content" or "filler" items of an utterance can be multi-phonemic if the target syllable contains a complex onset, nucleus, or coda.

4.4.2. Posterior superior temporal sulcus

The right posterior superior temporal sulcus (pSTs) was significantly more active for all speech conditions than the *baseline* condition. In the cross-condition activity pattern

[14] As previously mentioned, the WEAVER model is unclear where the scores are stored in the brain. It does hypothesize that the just that the premotor cortex is involved in "phonetic encoding and articulation." If this function also involves storing the syllabic motor programs, this aspect would be identical to the hypotheses of the DIVA model.

matching analysis, activity in this region matched the phonemic representation model. This is consistent with the three repetition suppression studies presented in Chapter 1.1.2.2 and others concluding that the superior temporal cortex processes phonological representations for speech production and perception (Chang, et al., 2010; Graves, et al., 2008; Okada & Hickok, 2006; Peeva et al. 2010; Turkeltaub & Coslett, 2010; Vaden et al., 2010). Many studies have demonstrated the phonemic representation only in the left hemisphere or the subject's language dominant hemisphere (Chang, et al., 2010; Graves, et al., 2008; Peeva et al. 2010; Turkeltaub & Coslett, 2010). However, Vaden et al. (2010) found that bilateral pSTs showed greater repetition suppression for word lists with more phonological repetition. This finding is confirmed by an fMRI study that showed greater activity for words in high-density phonological neighborhoods compared to those in low-density neighborhoods in the bilateral pSTs (Okada & Hickok, 2006).

We hypothesize that the phonemic representation of speech in the pSTs could reflect two possible processes needed to perform the present task. During stimulus presentation, subjects must transform the auditory stimulus into a phonological representation of the target utterance (Klatt, 1979; McClellend & Elman, 1986; Norris, 1994). This phonological representation is needed in order to choose an appropriate speech motor program for production (Guenther, 1994; Hickok & Poeppel, 2007; Roelofs, 1997). Damage to bilateral pSTs can result in auditory verbal aphasia, also known as word deafness. Subjects present with auditory comprehension and word repetition deficits but intact reading and writing skills, spontaneous speech, and recognition of non-speech sounds (Stefanatos, 2008; Wolberg, et al., 1990). This suggests that subjects with pSTs damage are unable to transform auditory inputs to their

corresponding phonological representations. The phonemic representation in the pSTs could reflect the phonological processing needed to transform the auditory stimuli into corresponding motor outputs.

Additionally, the bilateral pSTs is implicated in auditory feedback control during speech production (McGuire, et al. 1996). FMRI activation is greater in this region when subjects can hear their auditory feedback compared to when it is blocked. Activity is also greater when there is a mismatch between the expected and actual feedback (Tourville, et al., 2008; Toyomura, et al., 2007; Zheng, et al., 2010). The phonemic representation in the pSTs could be indicative of the same auditory-to-phonology transformation discussed above, but used for self-monitoring of speech production.

4.4.3. Cerebellum

The bilateral anterior and lateral areas of the cerebellum were active across speech tasks. We found that a variety of speech representation models – phonemic, syllabic frame, and syllabic – matched activity across different regions of the cerebellum. The right anterior cerebellum (lobules I-IV and V) matched a phonemic representation model. This area is associated with somatosensory and sensorimotor function (Dobromyslin, et al., 2012; Habas et al., 2009; Krienen & Buckner, 2009; Stoodley et al., 2012; Stoodley & Schmahmann, 2009). It is functionally connected to the superior temporal cortex and to somatosensory and motor/premotor areas (O'Reilly, et al., 2010). In a meta-analysis of auditory neuroimaging studies, this region of the cerebellum was most likely to be active compared to baseline (Petacchi, et al., 2005). Combining these findings, we

hypothesize that right lobules I-IV and V aid the pSTs in translating an auditory stimulus into a phonological code that can be used to produce the corresponding target utterance or to self-monitor speech output.

Activity in the bilateral cerebellar lobules VI and VIIIa matched the syllabic representation model. These areas are also implicated in language and sensorimotor processes (Dobromyslin, et al., 2012; Habas et al., 2009; Krienen & Buckner, 2009; Stoodley et al., 2012; Stoodley & Schmahmann, 2009) with functional connections to the somatosensory and motor/premotor cortices (O'Reilly, et al., 2010). In an ALE analysis, these areas were more likely to show activation for motor than sensorimotor tasks (Stoodley & Schmahmann, 2009). They are significantly more active for overt than covert speech (Bohland & Guenther, 2008; Burisko & Fiez, 2010) and show increasing activation for increasingly complex speech utterances (Bohland & Guenther, 2008; Ghosh, et al., 2008). Simple tongue and lip movements activate this area (Grodd, et al., 2001; Nitschke, et al., 1996), and damage to lobule VI can result in dysarthria, a disorder characterized by difficulty with articulation. Given these findings, we hypothesize that these areas are involved in modulating and regulating the syllabic feedforward motor programs associated with the lateral premotor cortex. This is in agreement with hypotheses that the cerebellum modifies the performance of movements (Bastian, 2006).

Moreover, in a study perturbing auditory speech feedback, lobule VI was more active during perturbed speech in which the first formant was shifted compared to unperturbed speech (Tourville, et al., 2008). This suggests that this region is involved in auditory

feedback monitoring. We have hypothesized that auditory regions (pSTs) encode phonemic representations and motor regions (lateral middle premotor cortex) encode syllabic representations. Lobule VI may translate between auditory and motor representations, either by generating the inverse model to correct the syllabic motor program from auditory error signals or by generating the forward model from the corrected motor program to predict the expected auditory output from a speech motor program (see Wolpert et al., 1998).

The right lateral cerebellar lobule VIIb was the only ROI that matched the syllabic frame structure representation model. This area is implicated in language and working memory function (Ackermann, 2008; Chen & Desmond 2005; Stoodley, et al., 2012; Stoodley & Schmahmann, 2009). It is also implicated in timing and movement synchronization (Hazeltine, et al., 1998). In neuroimaging studies, lobule VIIb shows greater activation for auditory pacing over visual pacing (Jäncke, et al., 2000), random over fixed timing (Dreher & Grafman, 2002), and greater activation with greater rhythm complexity (Penhune, et al., 1998). Moreover, rTMS to the lateral cerebellum disrupts millisecond range timing (Koch, et al., 2007). We suggest that this region of the lateral cerebellum is involved in auditory processing of rhythm, which is the basis of syllabic frames.

The cerebellum is often thought to modulate cerebral cortical function with cortico-cerebellar loops; however our results did not reveal any cortical ROIs representing syllabic frame units. Does the representation of frames in lobule VIIb of the cerebellum reflect only part of a cortio-cerebellar loop, even though our conservative methodology

failed to reveal the analogous cortical region? Or is the cerebellum the sole neural correlate to syllabic frames? As previously discussed, some researchers have posited that the presupplementary motor area (preSMA) is the cortical locus of syllabic frame structure representations. This could be a cortical target of lobule VIIb to process frames. The two areas are co-activated during speech production (Adank, 2012; Turkeltaub, et al., 2002). Moreover, they are also co-activated during tasks relevant to representing syllabic frames such as pacing movements to internal or external triggers (Dreher & Grafman, 2002; Jantzen, et al., 2007) and coordinating the movements of multiple effectors (Blouin, et al., 2004)). Functional connectivity suggests that lobule VIIb is connected to the prefrontal cortex (Dobromyslin, et al., 2012; Habas et al., 2009; Stoodley & Schmahmann, 2009), and a study of musical improvisation found greater functional connectivity for rhythmic over melodic improvisation between the preSMA and a cluster of activation in the cerebellum with a peak in lobule VII (Manzano & Ullén, 2012).

The preSMA may not have significantly matched the frame model due to a limitation of ROI-wise analyses. This methodology assumes that the functional response of an ROI is consistent throughout the region. If only a small part of the ROI responds with a particular across-activity pattern, but the rest does not, the response of interest may be "averaged-out" by the pattern of the rest of the ROI. This may be an issue particularly in the preSMA. Neuroimaging studies have divided this region into several functionally separate areas (Fink, et al., 1997; Picard & Strick, 1996).

Another reason the frame model may not have matched preSMA activity may be because the ROI parcellation scheme used in this study – like most other cortical parcellations – uses the vertical commissure anterior line to delineate the SMA proper from the preSMA. Functional connectivity-based parcellations of individual subjects reveal that this is generally, but not absolutely, true across all human brains (Lee et al., 2010). Therefore, while there is evidence from other studies for a frame representation in the preSMA, methodological limitations may have limited our results from revealing this finding.

4.4.4. SMA

The bilateral supplementary motor area (SMA) was more active for *all speaking* compared to *baseline*. The left SMA matched the phonemic representation model, while the right SMA matched the phonologically insensitive representation model. The bilateral SMA are reliably activated for speech production tasks (Alario, et al., 2006; Bohland, et al., 2008; Ghosh et al., 2008; Indefrey & Levelt, 2004), and neuroimaging studies suggest that the SMA is involved in movement execution for speech utterances (see Chapter 3.1.2.4 for more on movement execution). The SMA shows greater activation for longer utterances and for overt over covert speech productions (Alario, et al., 2006; Bohland, et al., 2008; Ghosh et al., 2008; Palmer, et al., 2001).

However, we found functional differences between hemispheres that appear to be related to language-dominance between the cortical hemispheres. Mutism occurred in patients who underwent resection of the SMA within the dominant speech hemisphere,

but not for patients who underwent resections in the non-dominant hemisphere (Krainik, et al., 2003). Electrical stimulation of the SMA demonstrated that while both hemispheres were equally likely to elicit a motor response, left SMA stimulations were more likely to elicit vocalizations or speech arrest (Fried, et al., 1991). Our finding that the left SMA activity matched a phoneme representation pattern is in agreement with Peeva, et al. (2010). They proposed that the left SMA was involved in a cortico-basal ganglia-thalamo-cortico circuit to initiate phoneme-sized motor programs.

In contrast, in the present study, the right SMA activity matched a phonologically insensitive representation pattern. In a meta-analysis of speech neuroimaging studies, Indefrey and Levelt (2004) found that, the left and right SMA are equally reliably active across picture generation, word generation, and word reading studies, and that neither was reliably active for listening studies. However, they also found that the right SMA was not reliably active in pseudoword reading studies, but the left was[15]. The authors suggested that the right SMA is involved in phonological code retrieval of lexical items previously stored in memory. We hypothesize that in our study, subjects were aware of the repetition across pseudowords in a block. After two utterances, the upcoming target was predictable, and subjects would be able to retrieve phonological codes from memory for the rest of the block[16]. This suggests the right SMA matches a phonologically insensitive representation because all conditions relied equally on retrieval of pre-stored phonological codes to produce repeated words within a block.

---

[15] It appears that most, if not all, of the neuroimaging studies used in the analysis enrolled only right-handed individuals.

[16] The *all same* condition had greater repetition than the other speaking conditions. However, subjects could only recognize that they were in an *all same* block after they had processed the second pseudoword stimulus.

4.4.5. Frontal operculum

The left anterior frontal operculum (aFO) was more active for the collapsed *all – baseline* contrast, and matched the phonologically insensitive representation model. We suggest two possibilities for the neural function of the aFO in speech. First, we hypothesize that the insensitive representation may reflect the role of the aFO in articulatory movement coordination. The region shows greater fMRI activity for articulating a word than for either word retrieval (Kemeny, et al., 2006) or listening to words (Wise, et al., 1999). FMRI activity also increases with increased utterance complexity (Bohland, et al., 2006). We suggest that the FO may have exhibited a phonologically insensitive representation because roughly the same set of articulators is involved across all utterances.

The second hypothesis is that the phonologically insensitive representation found in the aFO reflects prosodic processing. Bilaterally, the aFO is more active when subjects listen to normal sentences compared to those with flattened prosodic contours (Meyer et al., 2004) and more active for non-speech sounds containing speech melody – F0 contour and amplitude envelope – than for normal or pseudoword sentences (Meyer et al., 2002). When constructing the auditory stimuli, we aimed to maintain similar F0 contours across all stimuli and conditions. Therefore, if the aFO processes prosodic information, all conditions would show the same amount of fMRI-RS, as seen in the phonologically insensitive model. It is also possible the aFO's pattern of activation is a result of both of these functions. Patients with damage to the region can present with

Foreign Accent Syndrome, a disorder characterized by both articulatory and prosodic abnormalities.

4.4.6. Summary

The present study used an fMRI-RS paradigm to elicit various patterns of activity across speaking conditions corresponding to various representations of speech – phonemic, SSC, syllabic frame, syllabic, and phonologically insensitive. We found phonemic representations in the right posterior superior temporal sulcus and right anterior cerebellum (lobules I-IV and V) suggesting that this cortico-cerebellar loop translates the target auditory stimulus into a phonological representation for motor output or for feedback control of ongoing utterances. We found an SSC representation in the left dorsal inferior frontal gyrus, pars opercularis, while we found a full syllabic representation in the left ventral and middle premotor cortex, bilateral cerebellar lobule VI and VIIIa. This suggests 2 possibilities: either SSC motor programs are assembled into syllabic motor programs, or both SSC and syllabic motor programs are used for feedforward motor control. We also found an SSC representation in the left inferior frontal sulcus suggesting that the phonological content/filler items in memory are not individual phonemes, but SSC units of representation. We found a syllabic frame structure representation in the right cerebellar lobule VIIb, suggesting that the timing and structure of syllables may be represented separately from its phonological content, modulated by cerebellar timing functions. We found a phonologically insensitive representation in the left anterior frontal operculum and right supplementary motor area

suggesting that these regions are not sensitive to phonemic, SSC, syllabic frame, or syllabic representations.

These results may provide neural evidence for the slot/filler and frame content theories. They suggest that syllabic frames are represented in the brain dissociated from phonological content, and that the phonological content is represented as subsyllabic units.

5. NEURAL CORRELATES OF CONSONANT CLUSTERS

5.1. Introduction

This study aims to discover the neural correlates of subsyllabic constituents (SSCs) in the phonemes of a consonant cluster are grouped into a single representation. Chapter 1.2 described evidence – from speech errors and word games, from language development, and from phonotactic constraints – that implies the importance of SSCs as a representation of speech. A couple of neuroimaging studies sought to compare the production of utterances with and without consonant clusters. Bohland et al. (2009) compared multi-syllabic speech sequence productions and found greater activation in bilateral pre-SMA, anterior insula-frontal operculum, and right superior cerebellum for stimuli with complex over simple onsets. Riecker et al. (2008) presented bisyllabic pseudowords and found greater activation in the left inferior frontal gyrus and anterior insula, and bilateral inferior cerebellum for complex over simple syllables. In both of these studies, syllabic complexity was only one of several factors, which may explain the discrepancy between findings.

Nonetheless, these traditional fMRI contrasts are limited by the interpretability of their results. It is unclear what functionality is implied in a region with greater fMRI activation for greater SSC complexity. If an area represents consonant clusters as a single unit, greater activation would not necessarily be expected for greater cluster complexity. Moreover, comparing the production onsets or codas with varying complexity is confounded by longer utterance duration and the increased articulatory load for complex SSCs.

The present study used a functional magnetic resonance imaging repetition suppression (fMRI-RS) paradigm to identify specific brain regions involved in representing phonemes, SSCs, and complete syllables. Experimental conditions were constructed to vary by the amount of repetition of each speech representation of interest. We then defined expected cross-condition patterns of fMRI-RS for each hypothesized speech representation. For each anatomically-defined region of interest, we compared the cross-condition activity pattern to the predicted patterns in order to find regions that significantly matched expected patterns.

5.2. Methods

5.2.1. Participants

Sixteen right-handed native speakers (8 female, aged 20-43 years, mean 29.9 years) of American English participated. All subjects reported normal or corrected-to-normal vision and no history of hearing, speech, language, or neurological disorders. Informed consent was obtained according to the Boston University Institutional Review Board and the Massachusetts General Hospital Human Research Committee.

5.2.2. Stimuli

Stimuli were bisyllabic pseudowords in which the syllables had the frame structure CCV and were phonotactically legal in American English. The number of phonemes per pseudoword and the number of letters per orthographic stimulus were balanced across

conditions. None of the syllables formed an orthographic or a phonological word found in the MRC Psycholinguistic Database (Coltheart, 1981). Syllable frequency, as reported in the MRC Database, was not significantly different between conditions (ANOVA, $F_{(3,116)} = 0.56$, $p = 0.64$, n.s.)

Auditory stimuli were recorded with 32 bit sound at 4.41 kHz over a Samson C01U USB studio condenser microphone using Audacity software (http://audacity.sourceforge.net/). The speaker was a native speaker of American English who had previously practiced the sequences. Using the same software, selected sequences were normalized for intensity. Stimuli were chosen to maintain similar F0 across all sequences. The durations were adjusted to a constant length using PRAAT software that changes duration without changing F0 (Boersma & Weenink, 2007). Maximally, changes were < 15% of the total original length. Each auditory stimulus lasted 700 ms.

In each experimental block, subjects spoke pairs of pseudowords (Table 5.1). Blocks fell into four conditions that differed according to how often each type of speech representation – phoneme, SSC, or complete syllable – was repeated between the pairs of pseudowords. In the *all same* condition, subjects produced the same bisyllabic pseudoword for all repetitions. In the *different syllables only* condition, subjects alternated between two pseudowords containing the same phonemes and SSCs, but different syllables, (e.g., GROI.SLEE and GREE.SLOI). In the *same phonemes only* condition, subjects alternated between two pseudowords containing the same phonemes, but different onset SSCs and syllables (e.g., FRA.GLAU and FLA.GRAU). In the *all different* condition, subjects alternated between pseudowords containing different

phonemes, onset SSCs, and syllables (e.g.,, KWAI.BLA and SMOO.KROI). The *all same* condition used only one pseudoword per block while all other speaking conditions alternated between pairs of pseudowords. To maintain similar novelty for stimuli throughout the experiment across conditions more pseudowords were used for the *all same* condition compared to the other speaking conditions. For a full list of stimuli, see Table 5.2.

All same	**Different syllables only**	**Same phonemes only**	**All different**
Same phonemes **Same** SSCs	**Different** phonemes **Same** SSCs	**Same** phonemes **Different** SSCs	**Different** phonemes **Different** SSCs
THRAI.SKOO	**GROI.SLEE**	**FRA.GLAU**	**KWAI.BLA**
THRAI.SKOO	**GREE.SLOI**	**FLA.GRAU**	**SMOO.KROI**
THRAI.SKOO	**GROI.SLEE**	**FRA.GLAU**	**KWAI.BLA**
THRAI.SKOO	**GREE.SLOI**	**FLA.GRAU**	**SMOO.KROI**
THRAI.SKOO	**GROI.SLEE**	**FRA.GLAU**	**KWAI.BLA**
THRAI.SKOO	**GREE.SLOI**	**FLA.GRAU**	**SMOO.KROI**

Table 5.1. The four speaking experimental conditions in the fMRI study. Each box represents the orthographic and auditory presentation of the pseudoword to the subject in the scanner. In the *all same* condition, subjects produced the same pseudowords for all repetitions. In the *different syllables only* condition, subjects alternated between two pseudowords containing the same phonemes and SSCs but different syllables. In the *same phonemes only* condition, subjects alternated between two pseudowords containing the same phonemes but different onset SSCs and syllables. In the *all different* condition, subjects alternated between pseudowords containing different phonemes, SSCs, and syllables.

All same	Different syllables only		Same phonemes only		All different	
blau.kloi	smee.gway	smay.gwee	gra.shwee	gwa.shree	kwoe.dra	spoi.blee
swoo.shroi	twoi.pree	twee.proi	klau.broi	krau.bloi	thray.ploe	gwoo.sna
thrai.skoo	glay.snoi	gloi.snay	fra.glau	fla.grau	pla.twai	snoo.shray
snau.froi	kla.shroe	kloe.shra	twa.kray	tra.kway	sma.froo	blai.tway
fwoe.swai	groi.slee	gree.sloi	kwee.sla	klee.swa	thrau.blay	twoo.spee
snee.flau	sta.throi	stoi.thra	slee.gwai	swee.glai	klai.proo	gwau.stee
grai.stau	drau.skoe	droe.skau	trau.shwai	twau.shrai	kwoo.frau	gla.snai
slau.gwoi	swoi.dree	swee.droi	twoe.kra	troe.kwa	kwai.bla	smoo.kroi
shrau.spoo						
smay.prau						
spoe.bray						
swoe.ploo						

Table 5.2. Orthographic representations of stimuli used in all speaking conditions. In the *all same* condition, subjects repeated the pseudoword six times per block in the *all same* condition. In all other speaking conditions, subjects alternated three times between the pairs of pseudowords, for a total of six utterances.

5.2.3. Experimental Paradigm

During a single trial, the subject saw an orthographic representation of the pseudoword for 1 s and simultaneously heard the 500 ms auditory stimulus. Then, a white cross replaced the orthographic stimulus, cueing subjects to produce the sequence. A *baseline* condition was also intermixed in which subjects saw a series of asterisks on the screen instead of the orthographic stimulus and rested quietly instead of producing a

pseudoword. Blocks lasted 20 s, and consisted of six 2.5 s trials followed by a 3 s pause so the effects of one block would not confound the following one. Runs consisted of fifteen blocks and lasted approximately 5 min. Pseudowords and conditions were randomized within runs. Each subject completed 7 runs that optimally allowed for approximately 21 blocks per condition per subject.

Instructions and visual stimuli were projected onto a screen that subjects could view from within the scanner via a mirror attached to the head coil. Auditory stimuli were played over Sensimetrics MRI-compatible insert headphones model S-14.

5.2.4. Image acquisition

Functional and anatomical volumes were collected using the same protocol described in Chapter 4.2.4.

5.2.5. fMRI data analysis

The model included 10 condition-specific variables – the early and late components of the *all same*, *different syllables only*, *same phonemes only*, *all different*, and *baseline* conditions – and additional covariates: linear detrending covariates and motion parameters. The model was estimated for each subject. Model estimates for each speaking condition were contrasted with the *baseline* condition as well as *all speaking* conditions contrasted with *baseline*. Group statistics were then calculated separately for cortical and subcortical regions.

5.2.5.1. fMRI analysis

Surface- and volume-based voxel-wise analysis was used to assess group blood oxygen level dependent (BOLD) response differences in the *all speaking* - *baseline* condition in the cerebral cortex using the same methods described in Chapter 3.2.8. ROI-wise analysis was used to assess the BOLD response differences in all four speaking conditions compared to *baseline* in the cerebral cortex using the same methods described in Chapter 4.2.5.2.

5.2.5.2. Across-condition activity pattern matching

Experiment conditions were constructed to differ by the amount of repetition of speech representations. To quantitatively assess fMRI-RS across these conditions, we constructed expected patterns of fMRI-RS for each speech representation (Table 5.3). For ROIs encoding representations of individual phonemes, we expected fMRI-RS to be greater for the *all same*, *different syllables only*, and *same phonemes only* conditions – which contain the same set of phonemes in both repeated pseudowords – than the *all differen*t condition. For ROIs encoding SSC representations, we expected fMRI-RS to be greatest for the *all same* and *different syllables only* conditions – which contain the same onset SSCs between the two pseudowords – and weakest for the *same phonemes only* and *all different* conditions. For ROIs encoding full syllable representations, we expected fMRI-RS to be greater for the *all same* condition – which contains the same syllables in both repeated pseudowords – than the *different syllables only, same phonemes only,* and *all different* conditions. For ROIs that are insensitive to

phonemic, SSC, and syllabic representations, we expected equal fMRI-RS across all conditions.

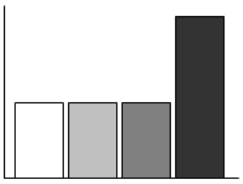
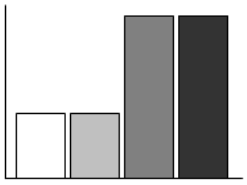
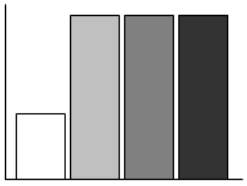
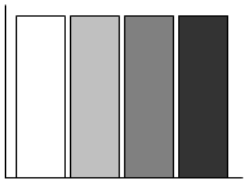
Speech representation	Expected pattern of fMRI-RS
Phoneme	
SSC	
Syllable	
Phonologically insensitive	

Table 5.3. Predicted across-condition activity for the phonemic, subsyllabic constituent (SSC), syllabic, and phonologically insensitive speech representations. The bar plots represent the hypothesized relative fMRI activity for the speech tasks (from left to right: *all same*, *different syllables only*, *same phonemes only*, *all different*) compared to baseline.

5.3. Results

5.3.1. Voxel-wise analysis

Figure 5.1 and Table 5.4 show the cortical brain regions that were significantly more active for all speaking conditions than *baseline* in a voxel-wise surface-based analysis (voxel-level $p < 0.001$, uncorrected; cluster-level $p < 0.0167$, corrected). Speaking resulted in bilateral clusters of greater BOLD response in the posterior superior temporal cortex including Heschl's gyrus, planum temporale, and planum polare. Bilateral clusters of greater activity for speech tasks were also seen in the ventral and middle portions of the pre- and post-central cortices extending into the middle premotor cortex. A cluster in the left medial premotor cortex included the presupplementary motor area and supplementary motor area, but an analogous cluster in the right hemisphere included only the presupplementary motor area. Clusters in the bilateral superior temporal lobule and occipital cortex, and left inferior temporal-occipital lobe were also active in the contrast, as were clusters in the right middle cingulate gyrus, left posterior frontal operculum into the ventral inferior frontal gyrus pars opercularis, and the right posterior central operculum.

Figure 5.1 and Table 5.4 also show the subcortical brain regions that were significantly more active for all speaking conditions than *baseline* in a voxel-wise volume-based analysis (voxel-level $p < 0.001$, uncorrected; cluster-level $p < 0.0167$, corrected). Speaking resulted in a subcortical cluster extending bilaterally into the basal ganglia – including the caudate, putamen, and pallidum – and thalamus. A second cluster extended bilaterally into the anterior and lateral aspects of the cerebellum.

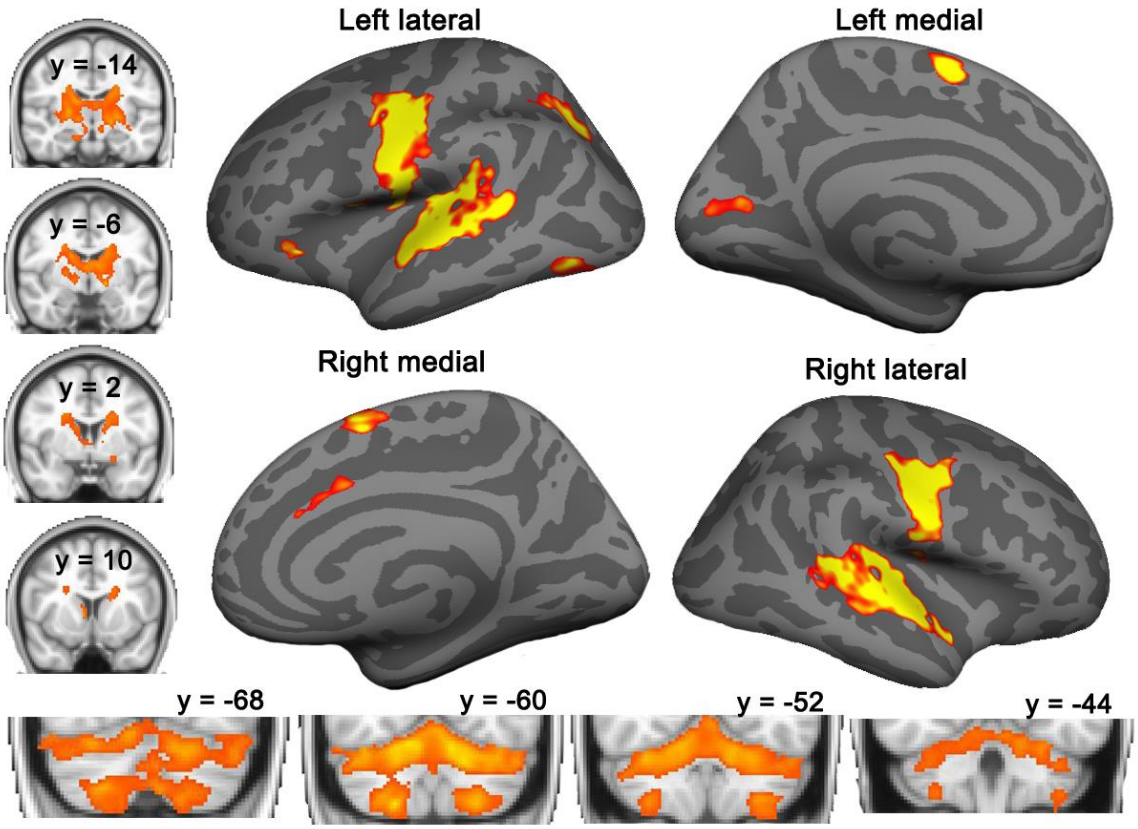


Figure 5.1. FMRI main effects of speaking (*all speaking* > *baseline*) in voxel-wise analysis. Significant clusters are shown on the left lateral (*upper left*), right lateral (*upper right*), and right medial (*lower right*) inflated surface representations of the FreeSurfer average template. Subcortical activity is shown on a series of coronal slices from the MNI305 template at the level of the subcortical nuclei (*left*) and cerebellum (*bottom*). Y coordinates indicate mm distance from the anterior commissure in MNI space). Contrast volumes were first voxel thresholded at $p < 0.001$, then cluster thresholded at cluster-wise p (CWP) < 0.0167 to correct for three analyses: subcortical and 2 cortical hemispheres. CWP was calculated by separate Monte Carlo simulations for each of the three analyses.

Region	name	Talairach coordinates x	y	z	t	Size	CWP
Left	Hechl's gyrus	-57.2	-15	3	12.5	2695 mm^2	0.0001
	Ventral motor cortex	-47.6	-5	26.7	13.1	2279 mm^2	0.0001
	Superior parietal lobule	-24.6	-57.3	37.6	6.3	516 mm^2	0.0001
	Occipital cortex	-16.1	-67.9	11.3	4.3	309 mm^2	0.0001
	ITO	-40.9	-64.6	-1.5	5.5	258 mm^2	0.0001
	SMA	-5.9	0	55.4	14.0	245 mm^2	0.0001
	Posterior frontal operculum	-44.3	6.9	5.1	5.9	119 mm^2	0.0001
	FOC	-31.7	28.8	-4.3	5.6	119 mm^2	0.0001
Right	vSC	50.9	-8.3	22.2	11.2	1713 mm^2	0.0001
	preSMA	7.5	7.7	52.1	6.1	287 mm^2	0.0001
	aSTg	59.8	-3.3	-1.8	13.0	268 mm^2	0.0001
	Occipital cortex	28.4	-91.1	-1.8	4.2	210 mm^2	0.0001
	Middle cingulated gyrus	8.7	11	32.9	4.2	136 mm^2	0.0001
	Superior parietal lobule	29.6	-55.4	45.2	5.1	110 mm^2	0.0002
	Posterior central operculum	46.3	-9.6	16.2	4.9	102 mm^2	0.0005
Bilat.	Cerebellar cortex	-20.8	-51.0	-48.6	9.8	74048 mm^3	0.0001
	Basal ganglia/thalamus	20.8	-17.5	20.3	4.4	4328 mm^3	0.0001

Table 5.4. Summary of cortical and subcortical activation peaks for the main effect of speaking (all speaking – *baseline*). From left to right, the columns show the region name, Talairach coordinates, T value, cluster size, and cluster-wise p (CWP). Abbreviations: ITO = inferior temporal-occipital cortex, SMA = supplementary motor cortex, FOC = orbito-frontal cortex, vSC = ventral somatosenory cortex, preSMA = presupplementary motor cortex, aSTG = anterior superior temporal gyrus, Bilat = Bilateral.

5.3.2. ROI-wise analysis

Table 5.5 and 5.6 show the cortical and subcortical brain regions that were significantly more active for all speaking conditions than *baseline*.

ROI Abbrev.		fMRI contrast compared to baseline				
		Collapsed	**All same**	**Diff syll**	**Same phon**	**All diff**
Rolandic cortex						
Left	vPMC	X	X	X	X	X
	midPMC	X	X	X	X	X
	vMC	X	X	X	X	X
	SMA	X	X	X	X	X
	preSMA	X	X	X	X	X
Right	vPMC	X			X	
	midPMC	X	X	X	X	X
	vMC	X	X	X	X	X
	SMA	X	X	X	X	X
	preSMA	X	X	X	X	X
Frontal cortex						
Left	aIFs	X		X	X	X
	pIFs	X	X	X	X	X
	dIFo	X	X	X	X	X
	vIFo				X	
	dIFt					
	vIFt				X	
	aFO			X	X	X
	pFO	X	X	X	X	X
Right	aIFs				X	X
	pIFs				X	X
	dIFo				X	
	vIFo				X	X
	dIFt					
	vIFt					
	aFO		X		X	X
	pFO		X		X	X
Temporal cortex						
Left	aSTg	X	X	X	X	X
	pSTg	X	X	X	X	X
	adSTs	X				

	pdSTs	X	X	X	X	X
	H	X	X	X	X	X
	pCO	X	X	X	X	X
	PO	X	X	X	X	X
	PP	X	X	X	X	X
Right	aSTg	X	X	X	X	X
	pSTg	X	X	X	X	X
	adSTs					
	pdSTs	X	X	X	X	X
	H	X	X	X	X	X
	pCO				X	X
	PO	X	X	X	X	X
	PP	X	X	X	X	X
	PT	X	X	X	X	X
Insular cortex						
Left	aINS	X				
Right	aINS					

Table 5.5. Cortical ROI-wise results of the all speaking conditions collapsed and individual speaking conditions (*all same*, *different syllables only*, *same phonemes only*, and *all different*) compared to baseline. X indicates a significance of $p_{FDR} < 0.05$. Abbreviations: vPMC = ventral premotor cortex, midPMC = middle premotor cortex, vMC = ventral primary motor cortex, SMA = supplementary motor area, preSMA = presupplementary motor area, aIFs = anterior inferior frontal sulcus, pIFs = posterior inferior frontal sulcus, dIFo = dorsal inferior frontal gyrus, pars opercularis, vIFo = ventral inferior frontal gyrus, pars opercularis, dIFt = dorsal inferior frontal gyrus, pars triangularis, vIFt = ventral inferior frontal gyrus, pars triangularis, aFO = anterior frontal operculum, pFO = posterior frontal operculum, aSTG = anterior superior temporal gyrus, pSTg = posterior superior temporal gyrus, adSTs = anterior dorsal temporal sulcus, pdSTs = posterior dorsal temporal sulcus, H = Heschl's gyrus, pCO = posterior central operculum, PO = parietal operculum, PP = planum polare, PT = planum temporale, aINS = anterior insula.

ROI Abbrev.		fMRI contrast compared to baseline				
		Collapsed	**All same**	**Diff syll**	**Same phon**	**All diff**
Subcortical nuclei						
Left	Thal	X	X	X	X	X
	Caud	X	X	X	X	X
	Put	X	X	X	X	X
	Pall	X	X	X	X	X
Right	Thal	X	X	X	X	X
	Caud	X	X	X	X	X
	Put	X	X	X	X	X
	Pall	X	X	X	X	X
Cerebellar cortex						
Left	I-IV	X	X	X	X	X
	V	X	X	X	X	X
	VI	X	X	X	X	X
	CrusI				X	X
	CrusII					
	VIIb				X	
	VIIIa	X	X	X	X	X
	VIIIb	X	X	X	X	X
	IX	X	X	X	X	X
	X	X	X	X	X	X
Right	I-IV	X	X	X	X	X
	V	X	X	X	X	X
	VI	X	X	X	X	X
	CrusI				X	
	CrusII					
	VIIb	X		X	X	X
	VIIIa	X	X	X	X	X
	VIIIb			X	X	X
	IX			X	X	X
	X	X	X	X	X	X
Vermis	VI	X	X	X	X	X
	CrusII	X	X	X	X	X

VIIb	X	X	X	X	X
VIIIa	X	X	X	X	X
VIIIb	X	X	X	X	X
IX	X	X	X	X	X
X		X	X	X	X

Table 5.6. Subcortical ROI-wise results of the all speaking conditions collapsed and individual speaking conditions (*all same*, *different syllables only*, *same phonemes only*, and *all different*) compared to baseline. X indicates a significance of $p_{FDR} < 0.05$. Cerebellar ROI abbreviations denote lobules. Abbreviations: Thal = thalamus, Caud = caudate, Put = putamen, Pall = pallidum.

Figure 5.2 shows the average activity for each condition across all cortical ROIs and subjects. Paired t-tests demonstrate that the fMRI activity – as quantified by the beta values – in the *same phonemes only* condition was significantly greater than the *all same* ($t_{(47)} = 4.47$, $p < 0.0001$), *different syllables only* ($t_{(47)} = 6.33$, $p < 0.0001$), and *all different* ($t_{(47)} = 4.13$, $p = 0.0002$) conditions. Activity was not significantly different (when corrected for multiple comparisons) between any other pairs of conditions (*all same*/*different syllables only:* $t_{(47)} = 0.45$, $p = 0.66$; *all same/all different*: $t_{(47)} = 2.9$, $p = 0.01$, *different syllables only/all different conditions*: $t_{(47)} = 2.75$, $p = 0.01$).

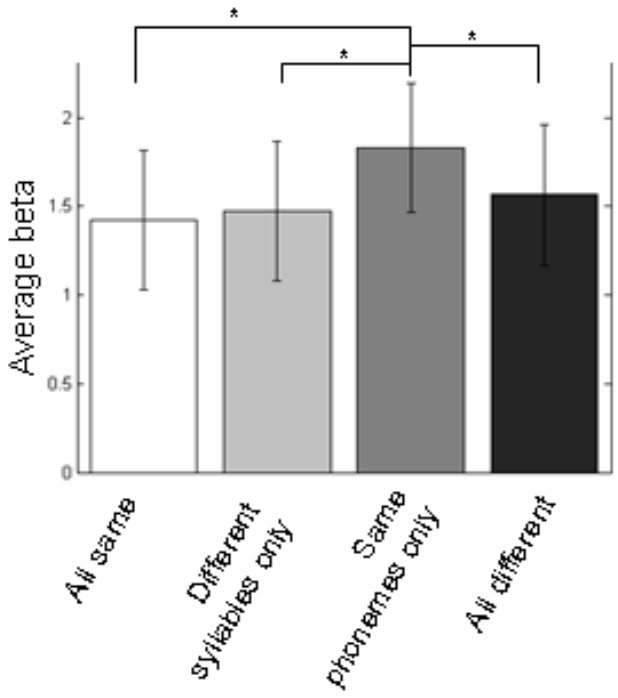


Figure 5.2. Average fMRI activity (beta values) for the four speaking conditions, averaged across subjects and ROIs.

5.3.3. Across condition activity pattern-matching analysis

Figure 5.3 shows cortical and subcortical regions that significantly ($P < 0.05$) matched a hypothesized pattern of speech representation. Bilateral posterior superior temporal gyrus and anterior dorsal superior temporal gyrus, and right posterior dorsal superior temporal sulcus matched a phonemic representation pattern of activation across conditions. The left posterior dorsal superior temporal sulcus and dorsal and ventral inferior temporal gyrus, pars triangularis matched an SSC representation pattern. The right anterior cerebellum (lobule VI) and bilateral lateral cerebellum (lobules VIIIa and VIIIb) also matched this representation pattern. The right lateral cerebellum (Crus II and

lobule VIIb) matched a syllabic representation pattern. No ROIs matched a phonologically insensitive representation pattern of activity.

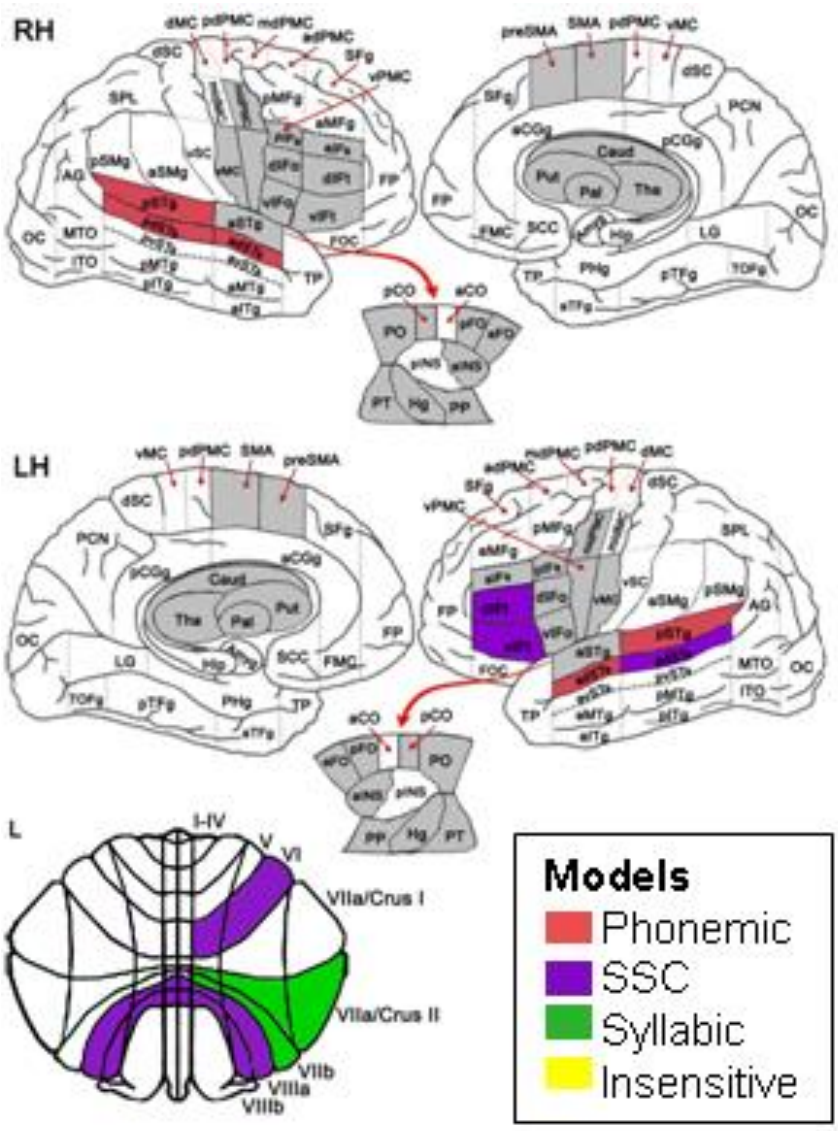


Figure 5.3. ROI brain map. Colors show significant ($P < 0.05$) matches to predicted speech representation models. Grey color indicates the ROI was included in the analysis, but did not significantly match a speech representation model. See tables 5.5 and 5.6 for definitions of the abbreviations.

5.4. Discussion

We compared the neural activity for speaking tasks to a silent baseline task and found greater activity for speech across both brain hemispheres in the primary motor and somatosensory cortices, the lateral and medial premotor cortices, superior temporal cortex, basal ganglia, and cerebellum. As previously discussed, these regions have all been implicated in the "minimal network" needed for speech motor control (Bohland & Guenther, 2006). Other brain areas with greater activity for speech tasks include the left inferior temporal-occipital cortex and superior parietal lobules that have both been implicated in processing orthographic stimuli (Chapter 3.4.). We also found that voxel- and ROI-wise measures were consistent. The measures diverge only in the sense that voxel-based contrast reveals sections of activation clusters that reside in a part of an ROI, but not enough to carry the average activation of that ROI.

Relative patterns of across-condition fMRI activity were matched to hypothesized models of fMRI-RS for phonemic, subsyllabic constituent (SSC), syllabic, and phonologically insensitive speech representations in a variety of speech-related ROIs.

5.4.1. Superior temporal cortex

In the cross-condition activity pattern matching analysis, activity in the bilateral superior temporal gyrus and anterior dorsal superior temporal sulcus and the right dorsal superior temporal sulcus (pdSTs) matched the phonemic representation model. This

representation in the superior temporal cortex is consistent with both past research and the results presented in Chapter 4 (see Chapter 4.4.2 for discussion).

In contrast to the phonemic representation in much of the superior temporal cortex, fMRI activity in the left pdSTs matched the SSC representation model. Several researchers have suggested that in the superior temporal cortex, primary auditory regions analyze the basic features of both speech and non-speech sounds (Binder, et al., 2000; Hickok & Poeppel, 2007; Liebenthal, et al., 2005). However, higher level auditory areas become more specialized to speech and are tuned to complex features in the signal. As a consistent extension of this theory, we hypothesize that there are SSC representations in the left pdSTs that process larger phonological units either for integration with orthographic stimuli or in preparation for motor execution.

5.4.2. Inferior frontal gyrus

Activity patterns in the left inferior frontal gyrus, pars triangularis (dorsal and ventral regions) matched the SSC representation model. This is slightly anterior to the region that matched this representation in Chapter 4 (see Chapter 4.4.1). This discrepancy could represent small differences in loci of the activation influencing these activity patterns.

5.4.3. Cerebellum

As in Chapter 4, we found multiple speech representations across different regions of the cerebellum. The right lateral cerebellum (Crus II and lobule VIIb) matched a syllabic

representation model. In Chapter 4, this area matched a frame representation model; in the present study, our syllabic model does not differentiate between full syllables and syllabic frame structures. The pattern of activation across conditions would be the same for an area processing either representation.

The activity in the right anterior cerebellum (lobule VI) and bilateral lateral cerebellum (lobules VIIIa and VIIIb) matched the SSC representation model. However, in the study presented in Chapter 4, both of these areas matched a syllabic representation model. In Chapter 4.4.3, we hypothesized that the syllabic representation in the cerebellum reflected a cortico-cerebellar loop connection with the middle premotor area for retrieving and articulating syllabic motor programs. We also hypothesized that SSC representations in the inferior frontal cortex reflect SSC motor programs used when syllabic motor programs are not available, and that these SSC motor programs are either intermediate programs used to build obligatory syllabic motor programs or an alternative unit of execution. The present finding that the motor area of the cerebellum can also process SSC representations seems to support the later hypothesis: SSC motor programs are an alternative unit of execution to syllabic motor programs. The stimuli in the present experiment used more complex syllables than those in Chapter 4; only CCV syllabic frames were used in the present study, whereas simpler CV, VC, CVC frames were used in Chapter 4. Therefore, the syllables in the present study were also, on average, less frequent. The average number of occurrences in the MRC Psychlinguistic Database for syllables in the present study was 29, but was 472 in Chapter 4. This suggests that very few utterances in the present study had pre-stored

motor programs in the syllabary and that all utterances required execution from SSC motor programs.

5.4.4. Difficulty confound

Despite these findings, these results appear to be confounded by differences in difficulty across speaking conditions. In particular, the *same phoneme only* condition required subjects to alternate between similar words with varying consonant clusters. This alternation is the basis of many "tongue twisters" such as "freshly fried flying fish" or "Which wristwatches are Swiss wristwatches?" These types of utterances are notoriously difficult to produce, and become more difficult with more repetitions, as in the task of the present study. Tongue twisters have longer utterance durations, slower speaking rate, and lower accuracy than for control utterances (Bowey, et al., 2005; Haber & Haber, 1982).

Figure 5.2 illustrates that this condition had significantly greater activity across ROIs than any other condition; we hypothesize this was due to the increased difficulty associated with this condition. The pattern matching analysis assumes that activation within an ROI is the same across conditions except for fMRI-RS-related reductions in activity. If other factors influence the relative cross-activity patterns, it is impossible to tease apart the contributions of the difficulty confound and the fMRI-RS-related differences in activity.

Moreover, in an fMRI study of covert tongue twister reading, the brain areas significantly more active for tongue twisters than control sentences were the left inferior frontal gyrus

and the superior temporal cortex (Keller, et al., 2008). In our study, these areas matched the SSC representation model. As pictured in Figure 5.3, the SSC model hypothesizes that activity for the *same phonemes only* condition is greater than the *all same* and *different syllable only* conditions. However, if the *same phonemes only* condition was inflated by the difficulty confound, it is possible that the superior temporal cortex and inferior frontal cortex ROIs would instead match the phoneme representation model. Because the current methodology cannot separate the contributions of difficulty and fMRI-RS, the current results are ambiguous at best.

5.4.5. Summary

This study used an fMRI-RS paradigm to elicit various patterns of activity across speaking conditions corresponding to different representations of speech – phonemic, SSC, syllabic, and phonologically insensitive. We found phoneme representations in the bilateral superior temporal cortex. We found SSC representations in the right anterior (lobule VI) and bilateral lateral (lobules VIIIa and VIIIb) cerebellum. While this is in conflict with the finding of syllabic representations in these regions in Chapter 4, we suggest that the discrepancy reflects the low frequency stimuli used in the present study that required greater use of SSC motor programs for execution. We found syllabic representations in the right lateral cerebellum (lobules VIIb and CrusII) that may reflect the syllabic frame representation seen in Chapter 4; activity in an area processing syllabic frames would match the syllable model used in the present study. However, these results are confounded with greater difficulty in the *same phonemes only* condition that requires the production of tongue-twister-like stimuli.

6. CONCLUSION

6.1. Summary of contributions

This dissertation presents a set of results – from behavioral data analyses, traditional fMRI contrasts, and fMRI-RS pattern matching – aimed at better understanding the types of representations of speech used throughout the brain.

In Chapter 2, we examined motor sequence learning of novel subsyllabic speech sequences with illegal or highly infrequent consonant clusters. These sequences allowed us to probe the behavioral aspects of speech motor sequence learning. We found subjects produced previously learned sequences faster and more accurately than similar novel sequences, demonstrating that speech motor sequence learning occurred with practice. We also found that subjects produced novel sequences as rapidly and accurately as learned sequences only if the novel sequences retained previously learned consonant clusters, suggesting that subjects produced the consonant clusters of novel sequences as individual phonemes, but with practice these are consolidated into a single articulatory program.

We also examined the neural correlates to speech motor sequence learning. Reductions in brain activity for learned sequences over novel sequences reinforce the hypothesis that the illegal consonant clusters of learned sequences have been consolidated to a single motor program, while novel illegal consonant clusters must be produced from multiple motor programs. Higher fMRI activity for novel than learned sequences in brain regions associated with speech motor planning and execution results

from the greater load of producing motor programs. Activity differences in the junction of the frontal operculum and anterior insula significantly correlated with learning success across subjects. This region is implicated in translating auditory and/or phonological representations of the speech stimulus to motor representations, and this finding implies that the efficacy of this translation predicts subject performance. Structural connectivity of the white matter under the left posterior superior temporal sulcus also significantly correlated with learning success, suggesting that the integrity of the auditory feedback network modulates subject performance of speech motor sequence learning.

Chapter 4 presented an fMRI-RS paradigm designed to examine speech representations – phonemic, subsyllabic constituent (SSC), syllabic frame, syllabic, and phonologically insensitive – used across cortical and subcortical regions implicated in speech production. These representations were chosen to tease apart the neural correlates of the slot/filler and frame/content models. We found dissociated syllabic frame representations in the lateral cerebellum, suggesting that the timing and structure of speech utterance is processed separately from its phonological content. We also found two types of phonological content representations. The phonemic representation model matched activity patterns in the superior temporal cortex, supplementary motor area, and anterior cerebellum, suggesting that these representations are used for execution, auditory feedback processing, and/or auditory target representations. The SSC representation models matched activity in the inferior frontal gyrus, suggesting that SSC motor programs are used for infrequent syllables when pre-stored syllabic motor programs (hypothesized to reside in the lateral premotor cortex) are not available.

Finally, Chapter 5 presented an fMRI-RS study aimed at further dissociating phonemic and SSC representations in the brain. We found that superior temporal cortex processed both types of representations, suggesting that this region is differentially involved in processing auditory stimuli and auditory feedback. Moreover, we found that in contrast to the syllabic representation in the lateral cerebellum found in Chapter 4, the lateral cerebellum may also process SSC representations. This finding, combined with the difference in syllabic frequency of the stimuli between studies, suggests that SSCs may be used as the output representation for low-frequency syllables. However, differences in difficulty between conditions confounded these findings.

6.2. Future directions

While the results of this dissertation provide important evidence for the representations of speech used in the brain, they also leave many unanswered questions. For instance, in Chapter 4 and 5, the results presented only use the early component of the functional block. A poor fit to a canonical hemodynamic response resulted from including the late component in the general linear model. When fMRI-RS is expected in some ROIs that process specific speech representations, we saw a dramatic decrease in fMRI activity as the block progressed that seemed to result from slower hemodynamic responses later in the block. This may have been due to a decrease in attention across the block that affected all ROIs activated by the speech task. However, these results cannot provide any definitive origin to this phenomenon.

The studies presented in Chapters 4 and 5 were also unable to fully clarify the relationship between SSC and syllabic representations in the frontal cortex. Are syllabic motor programs obligatory for executions? This suggests that SSCs are simply precursors for constructing syllabic programs. Or is the brain able to directly execute SSC motor programs for infrequent syllables? Causal network analyses such as structural equation modeling or Granger causality may be able to answer these questions, but are beyond the scope of the current research.

Further research into these kinds of questions will greatly help to clarify the neural processes of speech, an important goal not only for a basic understanding of the brain but also for understanding and alleviating speech disorders.

REFERENCES

Ackermann, H. (2008). Cerebellar contributions to speech production and speech perception: psycholinguistic and neurobiological perspectives. *Trends in Neuroscience*. 31, 6, 265-272.

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

Adank, P. (2012). The neural bases of difficult speech comprehension and speech production: two activation likelihood estimation (ALE) meta-analyses. *Brain and Language*. 122, 42-54.

Aguirre, G.K., Zarahn, E., & D'Esposito, M. (1998). The variability of human, BOLD hemodynamic responses. *NeuroImage*. 8, 360-369.

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

Anderson, B., Mruczek, R.E., Kawasaki, K., & Sheinberg, D. (2008). Effects of familiarity on neural activity in monkey inferior temporal lobe. *Cerebral Cortex*. 18, 11, 2549-2552.

Angenstein, F., Kammerer, E., & Scheich, H. (2009). The BOLD response in the rat hippocampus depends rather on local processing of signals than on the input or output activity. A combined functional MRI and electorphysiological study. *Journal of Neuroscience*. 29, 8, 2428-2439.

Ashburner, J. (2007). A fast diffeomorphic image registration algorithm. *NeuroImage*. 38, 1, 95-113.

Attwell, D., et al. (2010). Glial and neuronal control of brain blood flow. *Nature*. 468, 7321, 232-243.

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

Axer, H., Klingner, C.M., & Prescher, A. (2013). Fiber anatomy of dorsal and ventral language streams. *Brain and Language*. [Epub ahead of print retrieved February 26, 2013, from http://www.sciencedirect.com/science/article/pii/S0093934X12000879]

Baddeley, A. (2003). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience*. 4, 829-839.

Baker, C.I., Behrmann, M., & Olson, C.R. (2002). Impact of learning on representation of parts and wholes in monkey inferotemporal cortex. *Nature Neuroscience*. 5, 11, 1210-1216.

Barch, D.M., et al. (1997). Dissociating working memory from task difficulty in human prefrontal cortex. *Neuropsychologia*. 35, 10, 1373-1380.

Bastian, A.J. (2006). Learning to predict the future: the cerebellum adapts feedforward movement control. *Current Opinion in Neurobiology*. 16, 645-649.

Behrens, T.E., et al. (2003). Characterization and propagation of uncertainty in diffusion-weighted MR imaging. *Magnetic Resonance in Medicine*. 50, 5, 1077-1088.

Belin, P., Zatorre, R.J., Hoge, R., Evans, A.C., & Pike, B. (1999). Event-related fMRI of the auditory cortex. *NeuroImage*. 10, 417-419.

Berent, I., Steriade, D., Lennertz, T., & Vaknin, V. (2007). What we know about what we have never heard: Evidence from perceptual illusions. *Cognition*. 104, 591-630.

Binder, J.R., et al. (2000). Human temporal lobe activation by speech and nonspeech sounds. *Cerebral Cortex.* 10, 5, 512-528.

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

Blouin, J.-S., Bard, C., & Paillard, J. (2004). Contribution of the cerebellum to self-initiated synchronized movements: a PET study. *Experimental Brain Research*. 155, 63-68.

Bohland, J., Bullock, D., & Guenther, F.H. (2010). Neural representations and mechanisms for the performance of simple speech sequences. *Journal of Cognitive Neuroscience*. 22, 7,1504-1529.

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

Bolognini, N. & Ro, T. (2010). Transcranial magnetic stimulation: Disrupting neural activity to alter and assess brain function. *Journal of Neuroscience*. 30, 29, 9647-9650.

Bowey, J.A., McGuigan, M., Ruschena, A. (2005). On the association between serial naming speech for letters and digits and word-reading skill: towards a developmental account. *Journal of Research in Reading*. 28, 4, 400-422.

Brashers-Krug, T., Shadmehr, R., & Bizzi, E. (1996). Consolidation in human motor memory. *Nature*. 382, 252-255.

Browman, C.P. & Goldstein, L. (1988). Some notes on syllable structure in articulatory phonology. *Phonetica*. 45, 140-155.

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

Byrd, D. (1996). Influences on articulatory timing in consonant sequences. *Journal of Phonetics*. 24, 209-244.

Byrd, D. & Choi, S. (2006). At the juncture of prosody, phonology, and phonetics: The interaction of phrasal and syllable structure in shaping the timing of consonant gestures. *Proceedings of the 10th Conference on Laboratory Phonology*.

Byrd, D. & Tan, C.C. (1996). Saying consonant clusters quickly. *Journal of Phonetics.* 24, 263-282.

Callan, D.E., Jones, J.A., Callan, A.M., & Akahane-Yamada, R. (2004). Phonetic perceptual identification by native- and second-language speakers differentially activates brain regions involved with acoustic phonetic processing and those involved with articulator-auditory/orosensory internal models. *NeuroImage*. 22,1182-1194.

Calvert, G.A., Hansen, P.C., Iversen, S.D., & Brammer, M.J. (2001). Detection of audio-visual integration sites in humans by application of electrophysiological criteria to the BOLD effect. *NeuroImage*. 14, 427-438.

Chang, E.F., et al. (2010). Categorical speech representation in human superior temporal gyrus. *Nature Neuroscience*. 13, 11, 1428-1432.

Chein, J.M., Ravizza, S.M., & Fiez, J.A. (2003). Using neuroimaging to evaluate models of working memory and their implications for language processing. *Journal of Neurolinguistics*. 16, 315-339.

Chen, S.H.A. & Desmond, J.E. (2005). Temporal dynamics of cerebro-cerebellar network recruitment during a cognitive task. *Neuropsychologia*. 43, 1227-1237.

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

Clark, H.H. (1973). The language-as-fixed-effect fallacy: a critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*. 12, 335-359.

Clements, G.N. (1990). The role of the sonority cycle in core syllabification. In J. Kingston & M. Beckman (Eds.), *Papers in laboratory phonology 1: Between the grammar and physics of speech* (pp. 283-333). Cambridge: Cambridge University Press.

Cloutman, L.L., et al. (2012). The variation of function across the human insula mirrors its patterns of structural connectivity: Evidence from in vivo probabilistic tractography. *NeuroImage*, 59, 3514-3521.

Cohen, J. (1960). A coefficient for agreement for nominal scales. *Educational and Psychological Measurements*. 20, 37-46.

Cohen, L. & Dehaene, S. (2004). Specialization within the ventral stream: the case for the visual word form area. *NeuroImage*. 22, 466-476.

Cohen, L., Jobert, A., Le Bihan, D., & Dehaene, S. (2004). Distinct unimodel and multimodal regions for word processing int he left temporal cortex. *NeuroImage*. 23, 1256-1270.

Coltheart, M. (1981). The MRC Psycholinguistic Database. *Quarterly Journal of Experimental Psychology*. 33A, 497-505.

Contreras-Vidal, J.L. (1999). The gating functions of the basal ganglia in movement control. *Progress in Brain Research*. 121, 261-276.

Coull, J.T. & Frith, C.D. (1998). Differential activation of right superior parietal cortex and intraparietal sulcus by spatial and nonspatial attention. *NeuroImage*. 8, 176-187.

Courtney, S.M., Petit, L., Maisong, J.M., Ungerleider, L.G., & Haxby, J.V. (1998). An area specialized for spatial working memory in human frontal cortex. Science. 279, 1347-1351.

Curtis, C.E. & D'Esposito, M. (2003). Persistent activity in the prefrontal cortex during working memory. *Trends in Cognitive Science*. 7, 9, 415-423.

D'Esposito, M., Postle, B.R., & Rypma, B. (2000). Prefrontal cortical contributions to working memory: evidence from event-related fMRI studies. *Experimental Brain Research*. 133, 3-11.

da Silva, F.H.L., Wieringa, H.J., & Peters, M.J. (1991). Source localization of EEG versus MEG: Empirical comparison using visual evoked responses and theoretical considerations. *Brain Topography*. 4, 2, 133-142.

de Manzano, O. & Ullén, F. (2012). Activation and connectivity patterns of the presupplementary and dorsal premotor areas during free improvisation of melodies and rhythms. *NeuroImage*. 63, 272-280.

Dale, A.M., et al. (2000). Dynamic statistical parametric mapping : Combining fMRI and MEG for high-resolution imaging of cortical activity. *Neuron*. 26, 55-67.

Dale, A.M., Fischl, B., & Sereno, M.I. (1999). Cortical surface-based analysis. I. Segmentation and surface reconstruction. *Neuroimage*. 9, 179-194.

Dale, A.M. & Sereno, M.I. (1993). Improved localization of cortical activity by combining EEG and MEG with MRI cortical surface reconstruction: A linear approach. *Journal of Cognitive Neuroscience*. 5, 2, 162-176.

Davidson, L. (2006). Phonology, phonetics, or frequency: Influences on the production of non-native sequences. *Journal of Phonetics*. 34, 104-137.

Davis, M.H., Di Betta, A.M., Macdonald, M.J.E., & Gaskell, M.G. (2009) Learning and consolidation of novel spoken words. *Journal of Cognitive Neuroscience*. 21, 4, 803-820.

Dehaene, S., et al. (2001). Cerebral mechanisms of work masking and unconscious repetition priming. *Nature Neuroscience*. 4, 7, 752-758.

Dehaene-Lambertz, G., et al. (2006). Functional segregation of cortical language areas by sentence repetition. *Human Brain Mapping*. 27, 5, 360-371.

Dell, G.S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*. 93, 3, 283-321.

Dell, G.S. & O'Seaghdha, P.G. (1992). Stages of lexical access in language production. *Cognition.* 42, 287-314.

Desimone, R. (1996). Neural mechanisms for visual memory and their role in attention. *Proceedings of the National Academy of Science*. 93, 13494-13499.

Diedrichsen, J. (2006). A spatially unbiased atlas template of the human cerebellum. *NeuroImage*. 33, 1, 127-138.

Diedrichsen, J., Balsters, J.H., Flavell, J., Cussans, E., & Ramnani, N. (2009). A probabilistic MR atlas of the human cerebellum. *NeuroImage*. 46, 1, 39-46.

Dogil, G., et al. (2002), The speaking brain: a tutorial introduction of fMRI experiments in the production of speech, prosody and syntax. *Journal of Neurolinguistics*. 15, 59-90.

Donahue, M.J., Near, J., Blicher, J.U., & Jezzard, P. (2010). Baseline GABA concentration and fMRI response. *NeuroImage*. 53, 392-398.

Doyon, J. et al. (2002). Experience-dependent changes in cerebellar contributions to motor sequence learning. *Proceedings of the National Academy of Sciences*. 99, 2, 1017-1022.

Doyon, J., Penhune, V., Ungerleider, & L.G. (2003). Distinct contribution of the cortico-striatal and cortico-cerebellar systems to motor skill learning. *Neuropsychologia*. 41, 252-262.

Dreher, J.-C. & Grafman, J. (2002). The roles of the cerebellum and basal ganglia in timing and error prediction. *European Journal of Neuroscience*. 16, 1609-1619.

Duque, J., Olivier, E., & Rushworth, M. (in press). Top-down inhibitory control exerted by the medial frontal cortex during action selection under conflict. *Journal of Cognitive Neuroscience.*

Eckert, T., Preschel, T., Heinze, H.J., & Rotte, M. (2006). Increased pre-SMA activation in early PD patients during simple self-initiated hand movements. *Journal of Neurology*. 253, 2, 199-207.

Ekstrom, A. (2010). How and when the fMRI BOLD signal relates to underlying neural activity: The danger in dissociation. *Brain Research*. 62, 233-244.

Ekstrom, A., et al. (2009). Correlation between BOLD fMRI and theta-band local field potentials in the human hippocampal area. *Journal of Neurophysiology*. 101, 2668-2678.

Esterman, M., Tamber-Rosenau, B.J., Chiu, Y.C., & Yantis, S. (2010). Avoiding non-independence in fMRI data analysis: Leave one subject out. *NeuroImage*. 50, 572-576.

Ewbank, M.P., et al. (2011). Changing in “top down” connectivity underlie repetition suppression in the ventral visual pathway. *Journal of Neuroscience*. 31, 15, 5635-5642.

Fenn, K.M., Nusbaum, H.C., & Margoliash, D. (2003). Consolidation during sleep of perceptual leaning of spoken language. *Nature*. 425, 614-616.

Ferree, T.C., Clay, M.T., & Tucker, D.M. (2001). The spatial resolution of scalp EEG. *Neurocomputing*. 38-40, 1209-1216.

Fiez, J.A. & Peterson, S.E. (1998). Neuroimaging studies of word reading. *Proceedings of the National Academy of Sciences*. 95, 914-921.

Fink, G.R., Frackowiak, R.S.J., Pietrzyk, U., & Passingham, R.E. (1997). Multiple nonprimary motor areas in the human cortex. *Journal of Neurophysiology*. 77, 2164-2174.

Fischl, B., et al. (2002). Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron*. 33, 341-355.

Fischl, B., et al. (2004). Automatically parcellating the human cerebral cortex. *Cerebral Cortex*. 14, 11-22.

Fischl, B., Sereno, M.I., & Dale, A.M. (1999a). Cortical surface-based analysis. II: Inflation, flattening, and a surface-based coordinate system. *Neuroimage*. 9, 195-207.

Fischl, B., Sereno, M.I., Tootell, R.B., & Dale, A.M. (1999b). High-resolution intersubject averaging and a coordinate system for the cortical surface. *Human Brain Mapping*. 8,272-284.

Fowler, C.A., Treiman, R., & Gross, J. (1993). The structure of English syllables and polysyllables. *Journal of Memory and Language.* 32, 115-140.

Freedman, D.J., Riesenhuber, M., Poggio, T., & Miller, E.K. (2006). Experience-dependent sharpening of visual shape selectivity in inferior temporal cortex. *Cerebral Cortex*. 16, 1631-1644.

Fried, I., et al. (1991). Functional organization of human supplementary motor cortex studies by electrical stimulation. *Journal of Neuroscience*. 11, 11, 3656-3666.

Fromkin, V.A. (1971). The non-anomalous nature of anomalous utterances. *Language*. 47, 1, 27-52.

Fudge, E.C. (1969). Syllables. *Journal of Linguistics*. 5, 2, 253-286.

Gaser, C. & Schlaug, G. (2003). Brain structures differ between musicians and non-musicians. *Journal of Neuroscience*. 23, 27, 9240-9245.

Ghosh, S., et al. (2010). Nipype: Opensource platform for unified and replicable interaction with existing neuroimaging tools. *16th Annual Meeting of OHBM*. Barcelona, Spain.

Ghosh, S.S., Tourville, J.A., & Guenther, F.H. (2008). A neuroimaging study of premotor lateralization and cerebellar involvement in the production of phonemes and syllables. *Journal of Speech, Language, and Hearing Research*. 51, 5, 1183-1202.

Ghuman, A.S., Bar, M., Dobbins, I.G., & Schnyer, D.M. (2008). The effects of priming of rontal-temporal communication. *Proceedings of the National Academy of Sciences*. 105, 24, 8405-8409.

Gilbert, J.R., Gotts, S.J., Carver, F.W., & Martin, A. (2010). Object repetition leads to local increases in the temporal coordination of neural responses. *Frontiers in Human Neuroscience*. 4, 30, 1-12.

Goldstein, L., Nam, H., Saltzman, E., & Chitoran, I. (2009). Coupled oscillator planning model of speech timing and syllable structure. In G. Frant, H. Fujisaki, & J. Shen (Eds.), *Frontiers in Phonetics and Speech Science* (pp. 239-250). Beijing: The Commercial Press.

Golestani, N., et al. (2007). Brain structure predicts the learning of foreign speech sounds. *Cerebral Cortex*. 17, 3, 575-582.

Golestani, N. & Pallier, C. (2007). Anatomical correlates of foreign speech sound production. *Cerebral Cortex*. 17, 4, 929-934.

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

Golfinopoulos, E., Tourville, J.A., & Guenther, F.H. (2010). The integration of large-scale neural network modeling and functional brain imaging in speech motor control. *NeuroImage*, 52, 3, 862-874.

Gotts, S.J., Chow, C.C., & Martin, A. (2012). Repetition priming and repetition suppression: A case for enhanced efficiency through neural synchronization. *Cognitive Neuroscience*. 3, 3-4, 227-259.

Gow, D.W. (2012). The cortical organization of lexical knowledge: a dual lexicon model of spoken language. *Brain and Language.* 121, 273-288.

Grabski, K., et al. (2012). Functional MRI assessment of orofacial articulators: Neural correlates of lip, jaw, larynx, and tongue movements. *Human Brain Mapping*. 33, 10, 2306-2321.

Grafton, S.T., Hazeltine, E., & Ivry, R.B. (2002). Motor sequence learning with the nondominant left hand. *Experimental Brain Research*. 146, 369-378.

Graves, W.W., Grabowski, T.J., Mehta, S., & Gordon, J.K. (2007). A neural signature of phonological access: Distinguishing the effects of word frequency from familiarity and length in overt picture naming. *Journal of Cognitive Neuroscience*. 19, 4, 617-631.

Graves, W.W., Grabowski, T.J., Mehta, S., & Gupta, P. (2008). Left posterior superior temporal gyrus participates specifically in accessing lexical phonology. *Journal of Cognitive Neuroscience*. 20, 9, 1698-1710.

Grodd, W., Hülsmann, E., Lotze, M., Wildgruber, D., & Erb, M. (2001). Sensorimotor mapping of the human cerebellum: fMRI evidence of somatotopic organization. *Human Brain Mapping*. 13, 55-73.

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

Guenther, F.H. (2006). Cortical interactions underlying the production of speech sounds. *Journal of Communication Disorders*. 39, 5, 350-365.

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

Habar, L.R. & Haber, R.N. (1982). Does silent reading involve articulation? Evidence from tongue twisters. *The American Journal of Psychology*. 95, 3, 409-419.

Haggard, P. (2008). Human volition: towards a neuroscience of will. *Nature Reviews Neuroscience*. 9, 934-946.

Hagoort, P., et al. (1999). The neural circuitry involved in the reading of German words and pseudowords: A PET study. *Journal of Cognitive Neuroscience*. 11, 4, 383-398.

Hall, D.A., et al. (1999). “Sparse” temporal sampling in auditory fMRI. *Human Brain Mapping*. 7, 213-233.

Hamalainen, M. & Ilmoniemi, R. (1984). Interpreting measured magnetic fields of the brain: estimates of current distributions. Technical Report, Helsinki University of Technology, TKK-F-A559.

Hamilton, A.F. & Grafton, S.T. (2009). Repetition suppression for performed hand actions revealed by fMRI. *Human Brain Mapping*. 30, 9, 2898-2906.

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

Hasson, U., Nusbaum, H.C., & Small, S.L. (2006). Repetition suppression for spoken sentences and the effect of task demands. *Journal of Cognitive Neuroscience*. 18, 12, 2013-2029.

Hayasaka, S. & Nichols, T.E. (2003). Validating cluster size inference: random field and permutation methods. *NeuroImage*. 20, 2343-2356.

Hazeltine, E., Grafton, S.T., & Ivry, R.B. (1997). Attention and stimulus characteristics determine the locus of motor-sequence encoding: A PET study. *Brain*. 120, 123-140.

Hazeltine, E., Helmuth, L.L., & Ivry, R.B. (1998). Neural mechanisms of timing. *Trends in Cognitive Sciences*. 1, 3, 163-169.

Henson, R.N.A. (2003). Neuroimaging studies of priming. *Progress in Neurobiology*. 70, 53-81.

Henson, R.N.A., Shallice, T., Gorno-Tempini, M.L., & Dolan, R.J. (2002). Face repetition effects in implicit and explicit memory tests as measured by fMRI. *Cerebral Cortex*. 12, 2, 178-186.

Herrmann, F., Whiteside, S. P., & Cunningham, S. (2008). *An acoustic investigation into coarticulation and speech motor control: high vs. low frequency syllables*. Presented at Acoustical Society of America, Paris.

Hickok, G. (2012). Computational neuroanatomy of speech production. *Nature Reviews Neuroscience*. 13, 135-145.

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

Hikosaka, O., et al. (1996). Activation of human presupplementary motor area in learning of sequential procedures: a functional MRI study. *Journal of Neurophysiology*. 76, 617-621.

Hikosaka, O., Nakanumra, K., Sakai, K., & Nakahara, H. (2002). Central mechanisms of motor skill learning. *Current Opinion in Neurobiology*. 12, 217-222.

Hoeft, F., et al. (2007). Functional and morphometric brain dissociation between dyslexia and reading ability. *Proceedings of the National Academy of Sciences*. 104, 10, 4234-4239.

Honda, M., et al. (1998). Dynamic cortical involvement in implicit and explicit motor sequence learning: a PET study. *Brain*. 121, 2159-2173.

Houghton, G. (1990). The problem with serial order: A neural network model of sequence learning and recall. In R.Dale, C. Mellish, & M.Zock. (Eds.), *Current research in natural language generation* (pp. 287-319). London: Academic Press.

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

Ishai, A., Pessoa, L., Bikle, P.C., & Ungerleider, L.G. (2004). Repetition suppression of faces is modulated by emotion. *PNAS*. 101, 26, 9827-9832.

James, T.W. & Gauthier, I. (2006). Repetition-induced changes in BOLD response reflect accumulation of neural activity. *Human Brain Mapping*. 27, 37-46.

Jantzen, K.J., Oullier, O., Marshall, M., Steinberg, F.L., & Kelso, J.A.S. (2007). A parametric fMRI investigation of context effects in sensorimotor timing and cooridation. *Neuropschologia.* 45, 4, 673-684.

Jenkins, I.H., Brooks, D.J., Nixon, P.D., Frackowiak, R.S.J., & Passingham, R.E. (1994). Motor sequence learning: A study with positron emission tomography. *Journal of Neuroscience*. 14, 6, 3775-3790.

James, T.W., Humphrey, G.K., Gati, J.S., Menon, R.S., & Goodale, M.A. (1999). Repetition priming and the time course of object recognition: an fMRI study. *NeuroReport*. 10, 1019-1023.

Jäncke, L., Loose, R., Lutz, K., Specht, K., & Shah, N.J. (2000). Cortical activations during paced finger-tapping applying visual and auditory pacing stimuli. *Cognitive Brain Research*. 10, 51-66.

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

Jueptner, M., Frith, C.D., Brooks, D.J., Frackowiak, R.S.J., & Passingham, R.E. (1997). Anatomy of motor learning. II. Subcortical structures and learning by trial and error. *Journal of Neurophysiology*. 77, 1325-1337.

Kang, X., Herron, T.J., Turken, A.U., & Woods, D.L. (2012). Diffusion properties of cortical and pericortical tissue: regional variations reliability and methodological issues. *Magnetic Resonance Imaging*. 30, 8, 1111-1122.

Keller, T.A., Carpenter, P.A., Just, M.A. (2003). Brain imaging of tongue-twister sentence comprehension: Twisting the tongue and the brain. *Brain and Language*. 84, 189-203.

Kemeny, S., et al. (2006). Temporal dissociation of early lexical access and articulation using a delayed naming task – An fMRI study. *Cerebral Cortex*. 16, 4, 587-595.

Kemeny, S., Ye, F.Q., Birn, R., & Braun, A.R. (2005). Comparison of continuous overt speech fMRI using BOLD and arterial spin labeling. *Human Brain Mapping*. 24, 3, 173-183.

Kida, I., et al. (2006). Lamotrigine suppresses neurophysiological responses to somatosensory stimulation in the rodent. *NeuroImage*. 29, 216-224.

Kim, J.-H., et al. (2010). Defining functional SMA and pre-SMA subregions in human MFC using resting state fMRI: Functional connectivity-based parcellation method. *NeuroImage*. 49, 2375-2386.

Kim, S.G., Richter, W., & Uğurbil, K. (1997). Limitations of temporal resolution in functional MRI. *Magnetic Resonance in Medicine*. 37, 4, 631-636.

Klatt, D.H. (1979). Speech perception: a model of acoustic-phonetic analysis and lexical access. *Journal of Phonetics*. 7, 279-312.

Klein, A., et al. (2009). Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. *NeuroImage*. 46, 3, 786-802.

Koch, G., et al. (2007). Repetitive TMS of cerebellum interferes with millisecond time processing. *Experimental Brain Research*. 179, 291-299.

Kotz, S.A., & Schwartze, M. (2010). Cortical speech processing unplugged: a timely subcortico-cortical framework. *Trends in Cognitive Sciences*. 14, 9, 392-399.

Krainik, A., et al. (2003). Postoperative speech disorder after medial frontal surgery: Role of the supplementary motor area. *Neurology*. 60, 587-594.

Lee, K.-M., Chang, K.-H., & Roh, J.-K. (1999). Subregions within the supplementary motor area activate at different stages of movement preparation and execution. *NeuroImage*. 9, 117-123.

Lee Y. & Goldrick, M. (2008). The emergence of sub-syllabic representations. *Journal of Memory and Language*. 59, 155-168.

Leff, A.P., et al. (2001). The functional anatomy of single-word reading in patients with hemianopic and pure alexia. *Brain*. 124, 510-521.

Lehericy, S., et al. (2005). Distinct basal ganglia territories are engaged in early and advanced motor sequence learning. *Proceedings of the National Academy of Sciences*. 102, 35, 12566-12571.

Lemaire, J.J., et al. (2012). Extended Broca's area in the functional connectome of language in adults : Combined cortical and subcortical single-subject analysis using fMRI and DTI tractography. *Brain Topography*. [Epub ahead of print retrieved February 26, 2013, from http://link.springer.com/article/10.1007%2Fs10548-012-0257-7]

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

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

Li, L., Miller, E.K., & Desimone, R. (1993). The representation of stimulus familiarity in anterior inferior temporal cortex. *Jouranl of Neurophysiology*. 69, 6, 1918-1929.

Liebenthal, E., Binder, J.R., Spitzer, S.M., Possing, E.T., & Medler, D.A. (2005). Neural substrates of phonemic perception. *Cerebral Cortex*. 15, 10, 1621-1631.

Lin, F.-H., et al. (2006). Assessing and improving the spatial accuracy in MEG source localization by depth-weighted minimum-norm estates. *NeuroImage*. 31, 160-171.

Liu, A.K., Dale, A.M., & Belliveau, J.W. (2002). Monte Carlo simulation studies of EEG and MEG localization accurazy. *Human Brain Mapping*. 16, 47-62.

Logothetis, N.K. (2008). What we can do and what we cannot do with fMRI. *Nature*. 869-878.

Loevenbruck, H., Collins, M.J., Beckman, M.E., Krishnamurth, A.K., & Ahalt, S.C. (1999). Temporal coordination of articulatory gestures in consonant clusters and sequences of consonants. In O. Fujimura, B.D. Joseph, & B. Palek (Eds.) *Proceedings of Linguistics Phonetics 1998* (pp. 547-573). Prague: The Karolinum Press.

Lohmann, G., Erfurth, K., Müller, K., & Turner, R. (2012). Critical comments on dynamic causal modeling. *NeuroImage*. 59, 3, 2322-2329.

Lu, X., Hikosaka, O., & Miachi, S. (1998). Role of monkey cerebellar nuclei in skill for sequential movement. *Journal of Neurophysiology*. 79, 2245-2254.

McCandliss, B.D., Cohen, L., & Dehaene, S. (2003). The visual word form area: expertise for reading in the fusiform gyrus. *Trends in Cognitive Sciences*. 7, 7, 293-299.

McClelland, J.L. & Elman, J.L. (1986). The TRACE model of speech perception. *Cognitive Psychology*. 18, 1-86.

McLeod, S., Van Boorn, J., & Reed, V.A. (2001). Normal acquisition of consonant clusters. *American Journal of Speech-Language Pathology*. 10,2, 99-110.

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

MacMahon, D.B.T. & Olson, C.R. (2007). Repetition suppression in monkey inferotemporal cortex: Relation to behavioral priming. *Journal of Neurophysiology*. 97, 3532-3543.

MacNeilage, P.F. (1998). The frame/content theory of evolution of speech production. *Behavioral and Brain Sciences*. 21, 499-546.

MacNeilage, P.F. & Davis, B.L. (2000). On the origin of internal structure of word forms. *Science*. 288, 527-531.

Majdandžić, J., Bekkering, H., van Schie, H.T., & Toni, I. (2009). Movement-specific repetition suppression in ventral and dorsal premotor cortex during action observation. *Cerebral Cortex*. 19, 11, 2736-2745.Mechelli, A., Gorno-Tempini, M.L., & Price, C.J. (2003). Neuroimaging studies of word and pseudoword reading: Consistencies, inconsistencies, and limitations. *Journal of Cognitive Neuroscience*. 15, 2, 260-271.

Menanti, L., Gierhan, S.M.E., Segaert, K., & Hagoort, P. (2011). Shared language: Overlap and segregation of the neuronal infrastructure for speaking and listening revealed by functional MRI. *Psychological Science*. 22, 1173-1182.

Miller, E.K., Li, L., & Desimone, R. (1993). Activity of neurons in anterior inferior temporal cortex during a short-term memory task. *Journal of Neuroscience*. 13, 4, 1460-1478.

Miller, G.A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*. 63, 81-97.

Molholm, S. et al. (2006). Audio-visual multisensory integration in superior parietal lobule revealed by human intracranial recordings. *Journal of Neurophysiology*. 96, 2, 721-729.

Molins, A., Stufflebeam, S.M., Brown, E.N., Hämäläinen, M.S. (2008). Quantification of the benefit from integrating MEG and EEG data in minimum l_2-norm estimation. *NeuroImage*. 42, 1069-1077.

Moser, D., et al. (2009). Neural recruitment for the production of native and novel speech sounds. *NeuroImage*. 46, 2, 549-557.

Mostofsky, S.H. & Simmonds, D.J. (2008). Response inhibition and response selection: two sides of the same coin. *Journal of Cognitive Neuroscience*. 20, 5, 751-761.

Mukamel, R., et al. (2005). Coupling between neuronal firing, field potentials, and fMRI in human auditory cortex. *Science*. 309, 951-954.

Muthukumaraswamy, S.D., Evans, C.J., Edden, R.A.E., Wise, R.G., & Singh, K.D. (2012). Individual variability in the shape and amplitude of the BOLD-HRF correlates with endogenous GABAergic inhibition. *Human Brain Mapping*. 33, 2, 455-465.

Nakamura, K., Sakai, K., & Hikosaka, O. (1998). Neuronal activity in medial frontal cortex during learning of sequential procedures. *Journal of Neurophysiology*. 80, 2671-2687.

Nam, H., Goldstein, L., & Saltzman, E. (2009). Self-organization of syllable structure: a coupled oscillator model. In F. Pellegrino, E. Marisco, & I. Chitoran (Eds.)

Approaches to phonological complexity (pp. 299-328). Berlin, New York: Mouton de Gruyter.

Namasivayam, A.K. & van Lieshout, P. (2008) Investigating speech motor practice and learning in people who stutter. *Journal of Fluency Disorders*. 33:32-51.

Nee, D.E., et al. (2012). A meta-analysis of executive components of working memory. *Cerebral Cortex*. 23, 2, 264-282.

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

Nir, Y., et al. (2007). Coupling between neuronal firing rate, gamma LFP, and BOLD fMRI is related to interneuronal correlations. *Current Biology*. 17, 1275-1285.

Nitschke, M.F., Kleinschmidt, A., Wessel, K., & Frahm, J. (1996). Somatotopic motor representation in the human anterior cerebellum: A high-resolution functional MRI study. *Brain*. 119, 1023-1029.

Nixon, P., Lazarova, J., Hodinott-Hill, I., Gough, P., & Passingham, R. (2004). The inferior frontal gyrus and phonological processing: An investigation using rTMS. *Journal of Cognitive Neuroscience*. 16, 2, 289-300.

Nixon, P.D. & Passingham, R.E. (2000). The cerebellum and cognition: Cerebellar lesions impair sequence learning but not conditional visuomotor learning in monkeys. *Neuropsychologia*. 38, 1054-1072.

Norris, D. (1994). Shortlist: a connectionist model of continuous speech recognition. *Cognition*. 52, 189-234.

O'Reilly, J.X., Beckmann, C.F., Tomassini, V., Ramnani, N., & Johansen-Berg, H. (2010). Distinct and overlapping functional zones in the cerebellum defined by resting state functional connectivity. *Cerebral Cortex*. 20, 4, 953-965.

Okada, K. & Hickok, G. (2006). Identification of lexical-phonological networks in the superior temporal sulcus using functional magnetic resonance imaging. *NeuroReport*. 17, 1293-1296.

Orban, P., et al. (2010). The multifaceted nature of the relationship between performance and brain activity in motor sequence learning. *NeuroImage*. 49, 694-702.

Orfanidou, E., Marslen-Wison, W.D., & Davis, M.H. (2006). Neural response suppression predicts repetition priming of spoken words. *Journal of Cognitive Neuroscience*. 18, 8, 1237-1252.

Parker Jones, O., et al. (2013). Auditory-motor interactions for the production of native and non-native speech. *Journal of Neuroscience*. 33, 6, 2376-2387.

Parkes, L.M., et al. (2005). Quantifying the spatial resolution of the gradient echo and spin echo BOLD response at 3 Tesla. *Magnetic Resonance in Medicine*. 54, 1465-1472.

Paulesu, E., et al. (2001). Dyslexia: cultural diversity and biological unity. *Science*. 291, 2165-2167.

Pedreira, C., et al. (2010). Responses of human medial temporal lobe neurons are modulated by stimulus repetition. *Journal of Neurophysiology*. 103, 1, 97, 107.

Peeva, M. G. et al. (2010). Distinct representations of phonemes, syllables, and supra-syllabic sequences in the speech production network. *NeuroImage*. 50, 626-638.

Penhume, V.B., Zatorre, R.J., & Evans, A.C. (1998). Cerebellar contributions to motor timing: A PET study of auditory and visual rhythm reproduction. *Journal of Cognitive Neuroscience*. 10, 6, 752-765.

Petacchi, A., Laird, A.R., Fox, P.T., & Bower, .M. (2005). Cerebellum and auditory function : An ALE meta-analysis of functional neuroimagin studies. *Human Brain Mapping*. 25, 118-128.

Peyrin, C., Démonet, J.F., N'Guyen-Morel, M.A., Le Bas, J.F., & Valdois, S. (2011). Superior parietal lobule dysfunction in a homogeneous group of dyslexic children with visual attention span disorder. *Brain and Language*. 118:128-138.

Picard, N. & Strick, P.L. (1996). Motor areas of the medial wall: A review of their location and functional activation. *Cerebral Cortex*. 6, 342-353.

Price, C.J., Wise, R.J.S., & Frackowiak, R.S.J. (1996). Demonstrating the implicit processing of visually presented words and pseudowords. *Cerebral Cortex*. 6, 62, 62-70.

Poldrack, R.A., et al. (1999). Functional specialization for semantic and phonological processing in the left inferior prefrontal cortex. *NeuroImage*. 10, 15-35.

Poldrack, R.A., et al. (2005). The neural correlates of motor skill automaticity. *Journal of Neuroscience*. 32, 26, 5356-5364.

Price, C.J. (1998). The functional anatomy of word comprehension and production. Trends in Cognitive Sciences. 2, 8, 281-288.

Raboyeau, G., et al. (2004). Lexical learning of the English language: a PET study in healthy French subjects. *NeuroImage*. 22:1808-1818.

Rauschecker, A.M., Pringle, A., & Watkins, K.E. (2008). Changes in neural activity associated with learning to articulate novel auditory pseudowords by covert repetition. *Human Brain Mapping*. 29, 1231-1242.

Redford, M.A. (2008). Production constraints on learning novel onset phonotactics. *Cognition*. 107, 785-816.

Richan, F., Kronbichler, M., & Wimmer, H. (2009). Functional abnormalities in the dyslexic brain: A quantitative meta-analysis of neuroimaging studies. *Human Brain Mapping*. 30, 10, 3299-3308.

Robertson, E.M., Tormos, J.M., Maeda, F., & Pascual-Leone, A. (2001). The role of the dorsolateral prefrontal cortex during sequence learning is specific for spatial information. *Cerebral Cortex*. 11, 628-635.

Roebroeck, A., Formisano, E., & Goebel, R. (2011). The identification of interacting networks in the brain using fMRI: Model selection, causality and deconvolution. *NeuroImage*. 58, 2, 296-302.

Roelofs, A. (1997). The WEAVER model of word-form encoding in speech perception. *Cognition*. 249-284.

St. Heim, Opitz, B., Müller, K., & Frederici, A.D. (2003). Phonological processing during language production: fMRI evidence for a shared production-comprehension network. *Cognitive Brain Research*. 16, 285-296.

Sakai, K., et al. (1998). Transition of brain activation from frontal to parietal areas in visuomotor sequence learning. *Journal of Neuroscience*. 18, 5, 1827-1840.

Sakurai, Y., et al. (2000). Alexia caused by a fusiform or posterior inferior temporal lesion. *Journal of Neurological Sciences*. 178, 42-51.

Saur, D., et al. (2008). Ventral and dorsal pathways for language. *Proceedings of the National Academy of Sciences*. 105, 46, 18035-18040.

Selkirk, E. (1984). On the major class features and syllable theory. In M. Aronoff & R.T. Oehrle (Eds.) *Language sound structure* (pp. 107-136). : Cambridge, MA: MIT Press.

Shalom, D.B. & Poeppel, D. (2008). Functional anatomic models of language: Assembling the pieces. *Neuroscientist.* 14, 1, 119-127.

Schultz, K., et al. (2012). Simultaneous BOLD fMRI and fiber-optic calcium recording in rat neocortex. *Nature Methods*. 9, 6, 597-602.

Shattuck-Hufnagel, S. (1986). The representation of phonological information during speech production planning: Evidence from vowel errors in spontaneous speech. *Phonology Yearbook*. 3, 117-149.

Shattuck-Hufnagel, S. (1992). The role of word structure in segmental serial ordering. *Cognition*. 42, 213-259.

Shattuck-Hufnagel, S. & Klatt, D.H. (1979). The limited use of distinctive features and markedness in speech production: Evidence from speech error data. *Journal of Verbal Learning and Verbal Behavior*. 18, 41-55.

Shomstein, S. & Yantis, S. (2004). Control of attention shifts between vision and audition in human cortex. *Journal of Neuroscience*. 24, 47, 10702-10706.

Simmonds, D.J., Pekar, J.J., & Mostofsky, S.H. (2008). Meta-analysis of go/no-go tasks demonstrating that fMRI activation associated with response inhibition is task-dependent. *Neuropsychologia.* 46, 224-232.

Singh, K.D. (2012). Which “neural activity” do you mean? fMRI, MEG, oscillations and neurotransmitters. *NeuroImage*. 62, 1121-1130.

Smith, E.E. & Jonides, J. (1999). Storage and executive processes in the frontal lobes. *Science*. 283, 1657-1661.

Smits-Bandstra, S., De Nil, L.F., & Saint-Cyr, J.A. (2006). Speech and nonspeech sequence skill learning in adults who stutter. *Journal of Fluency Disorders*. 31, 116-136.

Steele, C.J. & Penhune, V.B. (2010). Specific increases within global decreases: A functional magnetic resonance imaging investigation of five days of motor sequence learning. *Journal of Neuroscience*. 30, 24, 8332-8341.

Stefanatos, G.A. (2008). Speech perceived through a damaged temporal window: Lessons from word deafness and aphasia. *Seminars in Speech and Language*. 29, 3, 239-252.

Stickgold, R. (2005). Sleep-dependent memory consolidation. *Nature*. 437, 1272-1278.

Stoodley, C.J. & Schmahmann, J.D. (2009). Functional topography in the human cerebellum: a meta-analysis of neuroimaging studies. *NeuroImage*. 15, 44, 489-501.

Stoodley, C.J., Valera, E.M., & Schmahmann, J.D. (2012). Functional topography of the cerebellum for motor and cognitive tasks: An fMRI study. *NeuroImage*. 59, 1560-1570.

Takai, O., Brown, S., & Liotti, M. (2010). Representation of the speech effectors in the human motor cortex: Somatotopy or overlap? *Brain and Language*. 113, 39-44.

Tomassini, V., et al. (2011). Structural and functional bases for individual differences in motor learning. *Human Brain Mapping*. 32, 494-508.

Toni, I., Krams, M., Turner, R., & Passingham, R.E. (1998). The time course of changes during motor sequence learning: A whole-brain fMRI study. *NeuroImage*. 8, 50-61.

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

Tourville, J.A. & Guenther, F.H. (2003). A cortical and cerebellar parcellation system for speech studies. *Language and Cognitive Processes*. 25, 7, 952-981.

Tourville, J.A. & Guenther, F.H. (2011). The DIVA model: A neural theory of speech acquisition and production. *Language and Cognitive Processes*. 25, 952-981.

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

Toyomura, A., et al. (2007). Neural correlates of auditory feedback control in human. *Neuroscience*. 146, 2, 499-503.

Treiman, R. & Danis, C. (1988). Short-term memory errors for spoken syllables are affected by the linguistic structure of the syllables. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 14, 1, 145-152.

Treiman, R., Fowler, C.A., Gross, J., Berch, D., & Weatherston, S. (1995). Syllable structure or word structure? Evidence for onset and rime units with disyllabic and trisyllabic stimuli. *Journal of Memory and Language.*

Treiman, R., Kessler, B., Bick, S. (2003). Influence of consonantal context on the pronunciation of vowels: A comparison of human readers and computational models. Cognition. 88, 49-78.

Turkeltaub, P.E. & Coslett, H.B. (2010). Localization of sublexical speech perception components. *Brain and Language*. 114, 1-15.

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

Vaden, Jr., K.I., Muftuler, L.T., & Hickok, G. (2010). Phonological repetition-suppression in bilateral superior temporal sulci. *NeuroImage*. 49, 1, 1018-1023.

Vanenberghe, R., Gitelman, D.R., Parrish, T.B., & Mesulam, M.M. (2001). Functional specificity of superior parietal mediation of spatial shifting. NeuroImage. 14, 661-673.

Vousden, J.I., Brown, G.D.A., & Harley, T.A. (2000). Serial control of phonology in speech production: A hierarchical model. *Cognitive Psychology*. 41, 101-175.

Wager, T.D. & Smith, E.E. (2003). Neuroimaging studies of working memory: a meta-analysis. Cognitive, Affective, and Behavioral Neuroscience. 3, 4, 255-274.

Wiener N (1949) *Extrapolation, interpolation, and smoothing of stationary time series*. Cambridge MA: MIT Press.

Wiggs, C.L. & Martin, A. (1998). Properties and mechanisms of perceptual priming. *Current Opinion in Neurobiology*. 8, 227-233.

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

Wolberg, S.C., Temlett, J.A., & Fritz, V.U. (1990). Pure word deafness. *South African Medical Journal*. 78, 668.

Yang, Y., et al. (2000). A silent event-related functional MRI technique for brain activation studies without interference of scanner acoustic noise. *Magnetic Resonance in Medicine*. 43, 185-190.

Zhang, Y., et al. (2009). Neural signatures of phonetic learning in adulthood: A magnetoencephalography study. *NeuroImage*. 46, 1, 226-240.

Zheng, Z.Z., Munhall, K.G., & Johnsrude, I.S. (2010). Functional overlap between regions involved in speech perception and in monitoring one's own voice during speech production. *Journal of Cognitive Neuroscience*. 22, 8, 1770-1781.

CURRICULUM VITAE

JENNIFER A. SEGAWA

33 Highland Street, Unit 2, Swampscott, MA 01907

Email: jsegawa@bu.edu

(443) 474 5399

Birth year: 1982

EDUCATION

2008 – Present **Ph.D. Candidate in Neuroscience**, Boston University

Specialization in Computational Neuroscience

Supervisor: Frank H. Guenther

2000 – 2004 **B.A. in Cognitive Science**, Johns Hopkins University

Concentrations in Linguistics, Computational Approaches

Supervisors: Luigi Burzio, William Badecker

WORK EXPERIENCE

2008 – Present **Research Assistant**

Speech Lab, Boston University, Boston, MA

Under the direction of Frank H. Guenther

2010 **Teaching Fellow**

Dept. of Cognitive and Neural Systems, Boston University

CN210/NE204 Introduction to Computational Models of Brain and Behavior

2004 – 2008	**Research Assistant**
	Dept. of Neurology, Massachusetts General Hospital, Boston, MA
	Under the direction of David W. Gow
2004	**Teaching Assistant**
	Dept. of Cognitive Science, Johns Hopkins University, Baltimore, MD
	050.240 The World of Language
2004	**Research Assistant**
	Dept. of Cognitive Science, Johns Hopkins University, Baltimore, MD
	Under the direction of Geraldine Legendre, Isabelle Barriere

AWARDS

2011	Northern Digital, Inc. Excellence Award, 9th International Seminar on Speech Production
2010	BU Teaching Fellow Award
2008	Boston University Dean/Presidential Fellowship

PUBLICATIONS

Segawa, J.A., Tourville, J.A., Beal, D.S., & Guenther, F.H. (submitted). Neural correlates of subsyllabic speech motor sequence learning.

Gow, D.W. & Segawa, J.A. (2009). Articulatory mediation of speech perception: A causal analysis of multi-model imaging data. *Cognition*. 110, 2, 222-236.

Gow, D.W., Segawa, J.A., Alfhors, S., & Lin, F.H. (2008). Lexical influences on speech perception: A Granger causality analysis of MEG and EEG source estimates. *NeuroImage*. 43, 3, 614-623.

PRESENTATIONS AND POSTERS

Segawa, J.A., Tourville, J.A., Beal, D.S., & Guenther, F.H. (2013, June). The representation of syllabic frame structures and phonological content in the brain. Meeting of the Organization on Human Brain Mapping, Seattle.

Segawa, J.A., Tourville, J.A., Beal, D.S., & Guenther, F.H. (2012, October). Dissociated neural representations of phonological content and syllabic frame structure. Society for Neuroscience, New Orleans.

Beal., D.S., Segawa, J.A., Tourville, J.A., Cai, S., & Guenther, F.H. (2012, October). Speech motor sequence learning difficulties in persistent developmental stuttering: An fMRI study. Society for Neuroscience, New Orleans.

Segawa, J.A., Tourville, J.A., & Guenther, F.H. (2011, June). Neuroimaging evidence for changes in phonological and structural frame representations in subsyllabic speech motor sequence learning. 9th International Seminar on Speech Production, Montreal, Canada.

Segawa, J.A., Tourville, J.A., & Guenther, F.H. (2011, June). Neural correlates of subsyllabic speak motor sequence learning. 17th Annual Meeting of the Organization on Human Brain Mapping, Quebec City, Canada.

Gow, D.W. & Segawa, J.A. (2007, November). Prosodic influences on segmental context effects: An analysis of neural dynamics. Annual Meeting of the Psychonomics Society, Long Beach.

Gow, D.W. & Segawa, J.A. (2007, May). Spatiotemporal imaging of assimilation context effects. Cognitive Neuroscience Society, New York.

Gow, D.W., Segawa, J.A., Meng, N., & Ho, M. (2006, August). Lexical influences on speech perception in time and space: Evidence from multimodal imaging. Annual Conference on Architectures and Mechanisms for Language Processing, Nijmegen.

Gow, D.W., Segawa, J.A., Meng, N., & Ho, M. (2006, May). Multimodal imaging of the Ganong effect: Spatiotemporal markers of top-down processing in speech perception. Conference on Cognitive and Neural Systems, Boston.