

BOSTON UNIVERSITY
GRADUATE SCHOOL OF ARTS AND SCIENCES

Dissertation

**FROM PROSODIC STRUCTURE TO ACOUSTIC SALIENCY: AN fMRI
INVESTIGATION OF SPEECH RATE, CLARITY, AND EMPHASIS**

by

ELISA GOLFINOPOULOS

B.A., Hamilton College, 2003

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

Second Reader

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

Third Reader

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

Dedicated to my mother, Sara Dawn Ives Golfinopoulos,
and my father, Elias Golfinopoulos,
who gave me wings and showed me the stars.

ACKNOWLEDGMENTS

I am deeply indebted to my colleagues in the Speech Laboratory at Boston University — Jason Tourville, Jason Bohland, Jon Brumberg, Alfonso Nieto-Castañón, Satrajit Ghosh, Simon Overduin, Jenn Segawa, Kevin Reilly, Oren Civier, Maya Peeva, Deryk Beal, Rob Law, Sean Lorenz, and Misha Panko — as well as my colleagues in the Speech Communication Group at the Massachusetts Institute of Technology — Shanjing Cai, Julie Yoo, Carrie Niziolek — who have been teachers and confidants and to whom I owe the greatest thanks.

If there is one thing that I have learned from this degree it is that science requires perseverance and hard work and those of my teachers, especially Dr. Frank Guenther, Dr. Daniel Bullock, Dr. Joseph Perkell, and Dr. Isao Hayashi, who have dedicated their careers to the scientific pursuit deserve much credit and many thanks for their commitment to the field and to my own growth and understanding. They like my family — Yeia yeia, Antonis, Niki, Litsa, Niki, Anna, Theothoros, Theothoros, Eunice and Martin Ives, Sara, Elias and Ted Golfinopoulos, Danielle Pelletier, Marie-Ève, Antoine, and Jasmin Léveillé — challenge and inspire me to work hard and make meaningful contributions to the field.

This work was funded by the National Institute on Deafness and other Communication Disorders (R01 DC02852, F. Guenther PI) and the Center of Excellence for Learning in Education, Science and Technology (CELEST) an NSF Science of Learning Center (NSF SBE-0354378). Neuroimaging was performed at the Athinoula A. Martinos Imaging Center at McGovern Institute for Brain Research, MIT.

FROM PROSODIC STRUCTURE TO ACOUSTIC SALIENCY: AN fMRI INVESTIGATION OF SPEECH RATE, CLARITY, AND EMPHASIS

(Order No.)

ELISA GOLFINOPOULOS

Boston University Graduate School of Arts and Sciences, 2013

Major Professor: Frank H. Guenther, Professor of Speech, Language, & Hearing
Sciences

ABSTRACT

Acoustic variability in fluent speech can arise at many stages in speech production planning and execution. For example, at the phonological encoding stage, the grouping of phonemes into syllables determines which segments are coarticulated and, by consequence, segment-level acoustic variation. Likewise phonetic encoding, which determines the spatiotemporal extent of articulatory gestures, will affect the acoustic detail of segments.

Functional magnetic resonance imaging (fMRI) was used to measure brain activity of fluent adult speakers in four speaking conditions: fast, normal, clear, and emphatic (or stressed) speech. These speech manner changes typically result in acoustic variations that do not change the lexical or semantic identity of productions but do affect the acoustic saliency of phonemes, syllables and/or words. Acoustic responses recorded inside the scanner were assessed quantitatively using eight acoustic measures and sentence duration was used as a covariate of non-interest in the neuroimaging analysis.

Compared to normal speech, emphatic speech was characterized acoustically by a greater difference between stressed and unstressed vowels in intensity, duration, and fundamental frequency, and neurally by increased activity in right middle premotor cortex and supplementary motor area, and bilateral primary sensorimotor cortex. These findings are consistent with right-lateralized motor planning of prosodic variation in emphatic speech.

Clear speech involved an increase in average vowel and sentence durations and average vowel spacing, along with increased activity in left middle premotor cortex and bilateral primary sensorimotor cortex. These findings are consistent with an increased reliance on feedforward control, resulting in hyper-articulation, under clear as compared to normal speech.

Fast speech was characterized acoustically by reduced sentence duration and average vowel spacing, and neurally by increased activity in left anterior frontal operculum and posterior dorsal inferior frontal gyrus pars opercularis -- regions thought to be involved in sequencing and phrase-level structural processing.

Taken together these findings identify the acoustic and neural correlates of adjusting speech manner and underscore the different processing stages that can contribute to acoustic variability in fluent sentence production.

TABLE OF CONTENTS

Introduction	1
1.1 Organization of the Dissertation.....	12
Methods	14
2.1 Participants	14
2.2 Stimuli	14
2.3 Experimental Protocol	18
2.4 Practice Session	20
2.5 Neuroimaging Data Acquisition.....	21
2.6 Data Analyses	22
2.6.1 Acoustic Analysis.....	23
2.6.1.1 Segmentation	23
2.6.1.2 Pre-processing.....	25
2.6.1.3 Characterizing Acoustic Variation.....	26
2.6.1.4 Average Vowel Spacing (Mels)	32
2.6.1.5 Average Vowel Dispersion (Mels).....	33
2.6.1.6 Average Vowel F0 (Semitones)	33
2.6.1.7 Average Adjacent-vowel F0 Contrast (Semitones).....	34
2.6.1.8 Average Vowel Intensity (dB SPL)	35
2.6.1.9 Average Adjacent-vowel Intensity Contrast (dB SPL)	36
2.6.1.10 Average Vowel Duration (ms).....	37

2.6.1.11	<i>Average Adjacent-vowel Duration Contrast (ms)</i>	38
2.6.1.12	<i>Statistical Analysis of the Acoustic Data</i>	39
2.6.2	<i>Neuroimaging Analysis</i>	40
2.6.2.1	<i>Surface-based Analysis</i>	40
2.6.2.2	<i>Region of Interest Analysis</i>	42
2.6.2.3	<i>Partial Pearson's Correlation Analysis</i>	43
2.6.2.4	<i>Functional Connectivity Analysis</i>	45
Acoustic Analysis Results		46
3.1	Reaction Time	47
3.2	Average Sentence-level Duration.....	48
3.3	Average Vowel Dispersion.....	50
3.4	Average Vowel Spacing	52
3.5	Average Vowel F0.....	57
3.6	Average Adjacent-vowel F0 Contrast	58
3.7	Average Vowel Intensity	63
3.8	Average Adjacent-vowel Intensity Contrast	65
3.9	Average Vowel Duration.....	67
3.10	Average Adjacent-vowel Duration Contrast	70
3.11	Discussion.....	72
Neuroimaging Results		81
4.1	Speech – Baseline.....	81
4.2	Clear – Normal	85

4.3	Fast – Normal	88
4.4	Clear – Fast.....	91
4.5	Emphatic – Normal.....	93
4.6	Emphatic – Fast	96
4.7	Clear – Emphatic	98
4.8	Region of Interest Analysis	99
4.9	Partial Pearson Correlation Analysis	116
4.10	Functional Connectivity Analysis	118
4.11	Discussion.....	122
Conclusions and Future Directions.....		134
Bibliography.....		139
Curriculum Vitae.....		160

LIST OF TABLES

Table 1-1: Predictions of cortical responses associated with increased phonological and phonetic encoding of linguistic and prosodic contrasts.	12
Table 2-1: The 24 sentence stimuli used in the study.....	17
Table 2-2: The six control stimuli used in the study.	17
Table 2-3: A breakdown of trials removed across subjects.....	23
Table 2-4: Main predictions of the acoustic differences between each speech manner (fast, clear, and emphatic) and normal speech in terms of prosodic phrasing and vowel quality.	32
Table 3-1 Group-level RMANOVA results	46
Table 3-2: Summary of the group-level acoustic analysis results for the contrasts of each speech manner manipulation with normal sentence production.....	80
Table 4-1: Significant clusters associated with overt sentence production.	84
Table 4-2: Significant clusters for the Clear – Normal contrast after controlling for sentence duration.	87
Table 4-3: Significant clusters for the Fast – Normal contrast after controlling for sentence duration.	90

Table 4-4: Significant clusters for the Clear – Fast contrast after controlling for sentence duration.	92
Table 4-5: Significant clusters for the Emphatic – Normal contrast after controlling for sentence duration.....	95
Table 4-6: Significant clusters for the Emphatic – Fast contrast after controlling for sentence duration.	98
Table 4-7: Significant clusters for the Clear – Emphatic contrast after controlling for sentence duration.	99
Table 4-8: Group-level conditional correlation results for the 29 ROIs and six acoustic measures.	118
Table 4-9: Group-level conditional connectivity for the normal and fast conditions.....	120
Table 4-10: Pair-wise comparisons of conditional connectivity between the fast and normal conditions.	121
Table 4-11: Group-level conditional connectivity for the speech (all manners) and baseline conditions.....	121
Table 4-12: Pair-wise comparisons of conditional connectivity between the speech (all manners) and baseline conditions.....	121

Table 4-13: Hierarchical organization of distinct prefrontal and premotor brain regions inferred from the partial Pearson correlation analysis.	131
--	-----

LIST OF FIGURES

Figure 2-1: The predicted F1 (left) and F2 (right) midpoint values of three vowels (/I/, /a/, /U/) as a function of vowel duration and consonantal context.	15
Figure 2-2: Schematic of reduced inter-vowel acoustic distances when vowel duration decreases from 500 ms (grey) to 100 ms (black).....	16
Figure 2-3: Timeline of a single trial.....	19
Figure 2-4: Waveform, spectrogram, and onsets and offsets of the corner vowels for male subject 5's clear production of the sentence, "Duke got a deep bag."	27
Figure 2-5: Waveform, spectrogram, and onsets and offsets of the corner vowels for male subject 5's normal (left) and fast (right) productions of the sentence, "Duke got a deep bag."	27
Figure 2-6: Average vowel spacing (left) and vowel dispersion (right) for the clear (plotted in black) and fast (plotted in grey) speech conditions.....	28
Figure 2-7: Average vowel spacing (left) and vowel dispersion (right) for subject 5's emphatic (plotted in grey) and normal (plotted in black) conditions.	29
Figure 2-8: Plots of the F0 contour, intensity contour, and duration of the vowels in the sentence, "Duke got a deep back" for subject five's emphatic (plotted in grey) and normal (plotted in black) productions.	31

Figure 3-1: Average reaction time values for the group as a function of speech manner.	48
Figure 3-2: Average sentence-level duration values for the group as a function of speech manner.	50
Figure 3-3: Average vowel dispersion for the group as a function of speech manner.....	51
Figure 3-4: Average vowel spacing for the group as a function of speech manner.....	53
Figure 3-5: Average vowel spacing for the males and females as a function of speech manner.	56
Figure 3-6: Average vowel F0 for the group as a function of speech manner.	57
Figure 3-7: Average adjacent-vowel F0 contrast for the group as a function of speech manner.	60
Figure 3-8: Average adjacent-vowel F0 contrast for the males and females as a function of speech manner.....	61
Figure 3-9: Average vowel intensity for the group as a function of speech manner.....	64
Figure 3-10: Average adjacent-vowel intensity contrast for the group as a function of speech manner.	67

Figure 3-11: Average vowel duration for the group as a function of speech manner.....	70
Figure 3-12: Average adjacent-vowel duration contrast for the group as a function of speech manner.	72
Figure 3-13: Stages of prosodic reorganization for a sentence as discussed in Keating and Shattuck-Hufnagel (2002).....	76
Figure 4-1: Group maps for the Speech – Baseline contrast displaying the significance of vertices within significant clusters.	83
Figure 4-2: Group maps for the Clear – Normal contrast after controlling for sentence duration.	87
Figure 4-3: Group maps for the Fast – Normal contrast after controlling for sentence duration.	89
Figure 4-4: Left hemisphere group map for the Fast – Normal contrast when a less strict cluster-wise threshold of $p < 0.05$ for the hemisphere is used.....	91
Figure 4-5: Group maps for the Clear – Fast contrast after controlling for differences in sentence duration between the conditions.	92
Figure 4-6: Group maps for the Emphatic – Normal contrast after controlling for sentence duration.	96

Figure 4-7: Group maps for the Emphatic – Fast contrast after controlling for sentence duration.	97
Figure 4-8: Group maps for the Clear – Emphatic contrast after controlling for sentence duration.	99
Figure 4-9: Twenty-six regions of interest defined on the surface by creating a composite map of the six contrasts for the left and right hemispheres.	100
Figure 4-10: Mean effect sizes and 90% confidence interval around the mean for the left dorsal inferior frontal gyrus pars opercularis ROI as a function of contrast.	101
Figure 4-11: Mean effect sizes and 90% confidence interval around the mean for the left middle premotor cortex ROI as a function of contrast.	102
Figure 4-12: Mean effect sizes and 90% confidence interval around the mean for the left anterior insula as a function of contrast.	102
Figure 4-13: Mean effect sizes and 90% confidence interval around the mean for the right posterior insula ROI as a function of contrast.	103
Figure 4-14: Mean effect sizes and 90% confidence interval around the mean for the right planum polare ROI as a function of contrast.	103
Figure 4-15: Mean effect sizes and 90% confidence interval around the mean for the left anterior frontal operculum ROI as a function of contrast.	104

Figure 4-16: Mean effect sizes and 90% confidence interval around the mean for the left anterior central operculum ROI as a function of contrast.	104
Figure 4-17: Mean effect sizes and 90% confidence interval around the mean for the right supplementary motor area ROI as a function of contrast.	105
Figure 4-18: Mean effect sizes and 90% confidence interval around the mean for the right posterior dorsal superior temporal sulcus ROI as a function of contrast.	105
Figure 4-19: Mean effect sizes and 90% confidence interval around the mean for the left ventral motor cortex ROI as a function of contrast.	106
Figure 4-20: Mean effect sizes and 90% confidence interval around the mean for the right ventral premotor cortex ROI as a function of contrast.	106
Figure 4-21: Mean effect sizes and 90% confidence interval around the mean for the left anterior supramarginal gyrus ROI as a function of contrast.	107
Figure 4-22: Mean effect sizes and 90% confidence interval around the mean for the right anterior supramarginal gyrus ROI as a function of contrast.	107
Figure 4-23: Mean effect sizes and 90% confidence interval around the mean for the right posterior inferior frontal sulcus ROI as a function of contrast.	108
Figure 4-24: Mean effect sizes and 90% confidence interval around the mean for the right middle premotor cortex ROI as a function of contrast.	108

Figure 4-25: Mean effect sizes and 90% confidence interval around the mean for the right middle motor cortex spherical ROI as a function of contrast.	109
Figure 4-26: Mean effect sizes and 90% confidence interval around the mean for the right posterior central operculum ROI as a function of contrast.	109
Figure 4-27: Mean effect sizes and 90% confidence interval around the mean for the right ventral somatosensory cortex ROI as a function of contrast.....	110
Figure 4-28: Mean effect sizes and 90% confidence interval around the mean for the left dorsal motor cortex ROI as a function of contrast.	110
Figure 4-29: Mean effect sizes and 90% confidence interval around the mean for the left dorsal somatosensory cortex spherical ROI as a function of contrast.....	111
Figure 4-30: Mean effect sizes and 90% confidence interval around the mean for the left middle dorsal somatosensory cortex ROI as a function of contrast.	111
Figure 4-31: Mean effect sizes and 90% confidence interval around the mean for the left superior ventral somatosensory cortex ROI as a function of contrast.....	112
Figure 4-32: Mean effect sizes and 90% confidence interval around the mean for the left middle ventral somatosensory cortex ROI as a function of contrast.....	112
Figure 4-33: Mean effect sizes and 90% confidence interval around the mean for the right dorsal somatosensory cortex spherical ROI as a function of contrast.....	113

Figure 4-34: Mean effect sizes and 90% confidence interval around the mean for the left inferior ventral somatosensory cortex ROI as a function of contrast.	113
Figure 4-35: Mean effect sizes and 90% confidence interval around the mean for the right middle somatosensory cortex ROI as a function of contrast.	114
Figure 4-36: Mean effect sizes and 90% confidence interval around the mean for the left putamen as a function of contrast.	114
Figure 4-37: Mean effect sizes and 90% confidence interval around the mean for the left pallidum as a function of contrast.	115
Figure 4-38: Mean effect sizes and 90% confidence interval around the mean for the left thalamus as a function of contrast.	115

LIST OF ABBREVIATIONS

aCO	anterior central operculum
aFO	anterior frontal operculum
aINs	anterior insula
aPHg	anterior parahippocampal gyrus
ANOVA	analysis of variance
aSMg	anterior supramarginal gyrus
aSTg	anterior superior temporal gyrus
BOLD	blood oxygenation-level-dependent
CV	consonant vowel
CVC	consonant vowel consonant
dIFo	dorsal inferior frontal gyrus pars opercularis
DIVA	directions into velocities of articulators
dMC	dorsal primary motor cortex
dSC	dorsal somatosensory cortex
F0	fundamental frequency
F1	first formant
F2	second formant
FDR	false discovery rate

FIR	finite impulse response
fMRI	functional magnetic resonance imaging
FOV	field of view
H & H	hypo/hyperarticulation theory
Hemi	hemisphere
Hg	Heschl's gyrus
HRF	hemodynamic response function
inf	inferior
IQR	interquartile range
L	left
Lg	Lingual gyrus
LPC	linear predictive coding
med	medial
mid	middle
midCg	middle cingulate gyrus
midMC	middle motor cortex
midPMC	middle premotor cortex
midSC	middle somatosensory cortex
MNI	Montreal Neurological Institute

OC	occipital cortex
pal	pallidum
PCN	precuneus
pCO	posterior central operculum
pdPMC	posterior dorsal premotor cortex
pdSTs	posterior dorsal superior temporal sulcus
pFO	posterior frontal operculum
pIFs	posterior inferior frontal sulcus
pINs	posterior insula
PO	parietal operculum
preSMA	presupplementary motor area
pSTg	posterior superior temporal gyrus
PP	planum polare
PT	planum temporale
put	putamen
Q1	question one
Q2	question two
R	right
R1	response one

R2	response two
RMANOVA	repeated measures analysis of variance
RMS	root mean square
ROI	region of interest
SMA	supplementary motor area
SPL	sound pressure level
SPM	statistical parametric mapping
sup	superior
TE	echo time
thal	thalamus
TOF	temporal occipital fusiform gyrus
TR	repetition time
vIFo	ventral inferior frontal gyrus pars opercularis
vMC	ventral motor cortex
vPMC	ventral premotor cortex
vSC	ventral somatosensory cortex

CHAPTER 1

INTRODUCTION

Fluent adult speakers frequently manipulate the acoustic detail of their speech (Lindblom, 1963; Ménard et al., 2007; Moon and Lindblom, 1994; Picheny et al., 1986). Take, for instance, this example from Katz and Selkirk (2011) demonstrating how the degree of semantic redundancy in the discourse influences the production of the same words in the two different conversations (stress is marked by bold underlining and an acute accent, whereas optional stress is marked in italics and grave accents):

Q1: What happened today?

R1: **Elíza** *màiled* the **cáramels**.

Q2: Who mailed the caramels?

R2: **Elíza** mailed the caramels.

The different prominence patterns in R1 and R2 result in different acoustic realizations of the same words that do not change their lexical or semantic identity, but do change their acoustic saliency. In English, stress can involve relative increases in the duration of the word by 11% or more (Weismer and Ingrisano, 1979), relative increases in the intensity and fundamental frequency (F0) of stressed syllables as compared to neighboring non-stressed syllables (Cooper et al., 1985; Cooper et al., 1983; Eady and Cooper, 1986; Katz and Selkirk, 2011; Patel et al., 2011; Terken, 1991; Xu and Xu, 2005) and vowels that are more separable acoustically (Engstrand, 1988; Harris, 1978;

Lindblom et al., 2007). In R1, where all the constituents in the response are new to the listener, a speaker may respond with stress on all noun constituents, making the content words acoustically more salient and, as a result, potentially easier for the listener to segment and interpret. In R2 only the proper noun, which answers the wh-question receives stress. The fact that the speaker does not attempt to stress the semantically redundant word, “caramels,” indicates that the interpretability gained by making this constituent acoustically more salient is not worth the speaker’s time to plan and produce acoustically salient speech sounds. This is consistent with a widely held theory of speech production, referred to as hypo/hyper-articulation (H & H) theory, that characterizes speech production as a tradeoff between accommodating the listener’s comprehension and the speaker’s production effort (for a review see Lindblom, 1990). According to the H & H theory, speakers attempt to find a balance between conserving production effort (resulting, in extreme cases, in *hypoarticulated* speech production) and conveying an interpretable utterance to the listener (requiring, in extreme cases, *hyperarticulated* speech production).

Although the example above demonstrates how stress may facilitate the listener’s comprehension, it is perhaps less obvious that the production of stress comes at a physical cost to the speaker. Many researchers have studied the way in which the recruitment of the speech muscles (e.g., Sussman and MacNeilage, 1978; Tuller et al., 1982) and the articulatory movement patterns differ from normal with requirements of increased emphasis or clarity (e.g., de Jong, 1995; Kent and Netsell, 1971; Lindblom, 1964; Matthies et al., 2001; Wohlert and Hammen, 2000). Clear speech is similar to

emphasis in that it is produced with the goal of improving speech intelligibility (e.g., Krause and Braidá, 2004; Picheny et al., 1985). Clear speech is generally accompanied by a global decrease in speech rate (reducing to between 90 - 100 words per minute from more typical speeds of 160 - 205 words per minute; Picheny et al., 1986; although see Krause and Braidá, 2004), more frequent and/or longer pauses (Bradlow et al., 2003; Krause and Braidá, 2004; Picheny et al., 1986), increased energy in the 1000-3000 Hz range of long-term spectra (Krause and Braidá, 2004), an average overall increase in F0 by 1.5 semitones (Krause and Braidá, 2004), the lengthening of vowel and/or consonant durations (Krause and Braidá, 2004; Ménard et al., 2007; Moon and Lindblom, 1994; Picheny et al., 1986; Tasko and Greilick, 2010), and vowels that are more separable acoustically (Ménard et al., 2007; Moon and Lindblom, 1994; Picheny et al., 1986). Studies on articulatory kinematics demonstrate that as speech sounds become acoustically more distinct during clear and stressed speech production, functionally relevant articulatory movements increase in velocity (despite the global decrease in speech rate) and achieve more extreme spatial targets (stress: de Jong et al., 1993; de Jong et al., 1995; Kent and Netsell, 1971; Tuller et al., 1982; clear speech: Matthies et al., 2001; Perkell et al., 2002). These findings may be interpreted as indicating that while speakers are capable of increasing the extent and speed of articulatory movements, they do so only when the speaking context demands, because such movements require extra effort. Direct evidence favoring a theory of increased articulatory effort under clear conditions comes from a study by Perkell et al. (2002). Using the peak velocity of tongue blade opening movements as a measure of relative effort, three of seven speakers

required greater-than-normal effort to produce clear speech and three of the remaining speakers did not require less-than-normal effort (Perkell et al., 2002). These results support the H & H theory's prediction that for certain subjects as the speaking style becomes clearer, production effort increases.

The acoustic consequences of fast speech have also been used as evidence in support of the theory that the acoustic properties of speech sounds are not invariant during speech production and can be dependent on production effort (e.g., Lindblom, 1963). For example, fast speech tends to involve a reduction in pitch range (Fougeron and Jun, 1998), fewer and/or shorter pauses relative to normal speech (e.g., Fougeron and Jun, 1998; Turner and Weismer, 1993), and vowels that are acoustically less distinct (Lindblom, 1963; Stålhammar et al., 1973; but see Gay, 1978). Simplifications observed across speakers under fast speech conditions often occur at boundaries between monosyllabic words (e.g., Byrd and Tan, 1996; Hardcastle, 1985; Munhall and Löfqvist, 1992). For example, de Jong (2001) reported that the production of glottal stops that acoustically cue the boundary between one vowel consonant (VC) syllable and the start of another are eliminated under repeated VC syllable productions at fast rates. Since this acoustic cue disappears, fast repeated VC productions become acoustically "resyllabified" and sound more like repeated consonant vowel (CV) syllable productions (e.g., fast repeated productions of /ib/ 'eeb' become acoustically more similar to /bi/ 'bee'; de Jong, 2001). Similarly, a blending process occurs when two consonants that share the same articulator must be produced across lexical boundaries (so-called juncture geminates and exemplified by the utterance 'Kiss Ted'). Specifically, Munhall and

Löfqvist (1992) demonstrated that at slow rates two separate laryngeal movements are used when speakers attempt to produce these juncture geminates, whereas at fast rates, only one smooth laryngeal movement is used. The finding that speakers encode adjacent consonants within two lexically distinct syllables as a single combined movement under fast speech rates suggests that an underlying structural reorganization takes place when requirements for speed increase (cf., Fougeron and Jun, 1998). In particular, speakers may reduce boundaries respected under typical conditions and, by consequence, the total duration of utterances. Consistent with Lindblom's H & H theory, therefore, speakers simplify productions under fast speech, resulting in less distinctive speech sound contrasts. However, rather than reflecting physical limitations of the speaker as Lindblom initially postulated (cf., Lindblom, 1963), these simplifications may reflect a higher-level reorganization under fast speech rates that result in larger-than-normal, multi-syllabic sequences (cf., Fougeron and Jun, 1998).

Findings from acoustic and kinematic studies implicating separable, but potentially interacting, phonological and articulatory encoding stages in speech production are also supported by a large number of neuroimaging studies (for a review see Indefrey and Levelt, 2004). In particular, several studies implicate Broca's area and, more specifically, a posterior dorsal region in left inferior frontal gyrus pars opercularis in association with syllabification and order-related processing (e.g., Alario et al., 2006; Bohland and Guenther, 2006; Friederici, 2011; Indefrey and Levelt, 2004; Moser et al., 2009; Papoutsis et al., 2009; Riecker et al., 2008; Rogalsky and Hickok, 2011; Sato et al., 2004). For example, Gelfand and Bookheimer (2003) found that dorsal inferior frontal

gyrus pars opercularis contributes to sequence processing regardless of whether sequences involve phonological units. These results are corroborated by the functional magnetic resonance imaging (fMRI) study of Bohland and Guenther (2006), in which speakers were asked to produce multi-syllabic non-lexical stimuli that varied in syllable and sequence complexity. Both a sequence and a sequence-by-syllable interaction effect was found in posterior dorsal inferior frontal gyrus, indicating increased engagement of this region under conditions in which sequencing becomes more difficult as the sequence's syllabic constituents become more complex. In a meta-analysis, Indefrey and Levelt (2004) found that the left posterior inferior frontal gyrus was the only region that did not show sensitivity to whether speech was produced overtly or covertly, leading to the inference that it is the shared phonological processing properties of these tasks that modulate activity in this region. Together, these findings indicate that focal damage to posterior Broca's area would interfere with the structural organization of phrases. There is some evidence in support of this hypothesis. For example, one of the primary areas implicated in acquired apraxia of speech is posterior inferior frontal gyrus (Hillis et al., 2004; Rohrer et al., 2010) and dysprosody is one of the behavioral criteria used to distinguish apraxia of speech from other speech motor disorders (for a review see Jacks, 2009). Acquired apraxia of speech is commonly associated with slow speech rate (Kent and Rosenbek, 1982; Rohrer et al., 2010), lengthened segment durations in multisyllabic words or in words embedded in sentences (Collins et al., 1983; Haley and Overton, 2001; Kent and Rosenbek, 1982) and increased inter-syllabic pause time or equalized syllable durations (Kent and Rosenbek, 1982). Deger and Ziegler (2002) suggest that

such speech motor difficulties as scanning speech, disturbed intersyllabic coarticulation or pausing observed for individuals with apraxia of speech reflects an underlying difficulty in switching between different syllabic motor routines. In their study, when compared to fluent controls and nonapraxic aphasics, individuals with apraxia of speech required a longer time to plan and produce the second syllable in bi-syllabic utterances that differed only in terms of the initial phoneme of the second syllable (e.g., ‘data’ or ‘daba’ as compared to ‘dada’). Deger and Ziegler (2002) theorize that the individuals with apraxia of speech in their study could not properly assemble larger multi-syllabic sequences composed of different syllables, that, as noted above, is a potential encoding strategy used by fluent adult speakers to produce fast speech. Evidence suggesting a role for posterior Broca’s area in the dysprosody associated with apraxia of speech comes from a recent voxel-based morphometry study by Rohrer and colleagues (2010), who demonstrated that the reduced diadochokinetic rate (the rate at which speakers could repeat ‘pa-ta’, ‘ta-ka’, ‘pa-ta-ka’ and ‘pla-kra-ta’) of individuals with apraxia of speech is correlated with grey matter loss in the left inferior frontal gyrus pars opercularis. These findings indicate a direct relationship between the difficulties encoding rate experienced by individuals with apraxia of speech and focal damage to posterior Broca’s area.

In contrast, a number of neuroimaging studies have linked the left lateral premotor cortex with articulatory encoding (e.g., Dick et al., 2011; Hauk et al., 2004; Indefrey and Levelt, 2004; Papoutsi et al., 2009; Peeva et al., 2010; Tourville et al., 2008; Tremblay and Small, 2011; Wise et al., 1999). This functional distinction is often motivated by

reports of a somatotopic organization along the precentral gyrus with representations of the oro-facial articulators in ventral regions and representations of the intrinsic muscles of the larynx in more dorsal regions (Brown et al., 2009; Brown et al., 2008; Olthoff et al., 2008; Takai et al., 2010). However, recent fMRI findings by Chen and colleagues (2008), on rhythm perception and production during a finger tapping task, suggest that the selective engagement of subregions within premotor cortex may not simply be determined on the basis of the effectors controlled. In particular, the authors found that middle premotor cortex was the only subregion that was significantly active during listening to musical rhythms with anticipation of reproduction, during tapping in synchronization to musical rhythms, and during naïve passive perception in which sound movement associations had yet to be learned. Ventral premotor cortex, in contrast, was only significantly active during listening with anticipation and during tapping in synchronization. These results were interpreted as indicating that middle premotor cortex may uniquely contribute to tracking the metrical organization of sequences. A more recent fMRI study on oral motor mappings to percussive musical rhythms did not replicate the finding that middle premotor cortex is engaged under conditions in which subjects have not learned sound movement associations (Tsai et al., 2010). However, Tsai et al. (2010) did find that middle premotor cortex is engaged when there is a learned association between orofacial movements and metrically organized sound sequences. These results indicate that perceptual cues of familiar actions may be mapped directly to their corresponding motor representations with contributions from middle premotor cortex.

In addition to the hypothesis that the left prefrontal and premotor regions make distinct functional contributions to speech processing, another common claim is that the left and right hemispheres are specialized for processing linguistic (e.g., lexical tones, phonemes, syllables, and/or words) and prosodic (e.g., non-lexical changes in pitch, intensity, and/or duration) contrasts, respectively (for a review see Friederici, 2011). This hemispheric asymmetry was initially substantiated on the basis of studies on individuals with brain damage to the left or right hemispheres. For example, Van Lancker and Sidtis (1992) found that individuals with right hemisphere damage primarily relied on durational cues to make judgments about the emotional content of presented sentence stimuli, while those individuals with left hemisphere damage relied mostly on modulations in F0. These findings established the hypothesis that component properties within the speech signal are independently processed, a hypothesis that has since been elaborated (e.g., Hickok and Poeppel, 2007; Poeppel, 2003). In particular, it is thought that left non-primary auditory areas and prefrontal regions are preferentially engaged during the processing of rapid spectral changes (that occur within 20 – 40 ms and would be necessary, for example, for processing acoustic features associated with linguistic contrasts such as formant transitions in stop consonants) and right non-primary auditory areas and prefrontal regions are preferentially engaged during the processing of slower spectral changes (that occur within 150 – 250 ms and would be necessary, for example, for processing acoustic features associated with prosodic contrasts). Many studies on stress perception have supported this theory, demonstrating the contributions of right inferior frontal gyrus and/or frontal operculum to the processing of stress patterns in

sentences (e.g., Geiser et al., 2008; Klein et al., 2011; Perrone et al., 2010). However, since most previous studies on stress have involved perceptual decision making tasks, the extent to which this laterality effect applies to the production of stress is not yet known (although see Loevenbruck et al., 2005).




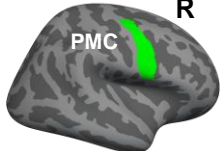
The separable engagement of Broca's area and middle premotor cortex in the phonological organization and in the selection of actions on the basis of sensorimotor associations, respectively, has also largely been based on studies involving perceptual decision making tasks or the production of multisyllabic word-like nonwords or single word stimuli. In the present fMRI study we examine the acoustic and neural correlates of manipulating the acoustic detail of prosodic and linguistic contrasts using real word sentence stimuli that all contained the same number of syllables across speech conditions. Furthermore, each sentence contained the same four corner vowels so that acoustic analysis could be performed on the vowels and acoustic manipulations could be compared between the conditions. Since a vowel's acoustic identity changes depending on its position in a sentence, each vowel appeared in all four positions analyzed in the sentence (see Table 2-1, for examples of the sentence materials). Production of these sentences involved minimal demand on working memory since the stimuli were orthographically presented at the time of production. In addition, all speakers took part in a 15 minute practice session prior to entering the scanner to familiarize themselves with the sentences and the production of the following four different speech manners: 1) **Normal** (Sentences were produced at a normal, conversational speaking rate); 2) **Fast** (Sentences were produced as quickly as possible without slurring any of the words

together); 3) **Clear** (Words in sentences were produced as distinctly as possible); and 4) **Emphatic** (Sentences were spoken at a normal rate with greater emphasis on the first and fourth word in the sentence). Differences between the conditions in terms of prosodic phrasing or vowel quality were assessed using eight acoustic measures (described in detail in Chapter 2). In order to ensure that only trials performed correctly were assessed, all trials in which subjects made a reading error or had an outlying reaction time or sentence-level acoustic production duration were removed from acoustic and neuroimaging analyses. In addition, sentence-level acoustic production duration, which could act as a confound, was controlled in the neuroimaging analysis using a covariate in the first-level general linear model. Significant regions selectively responsive to acoustic variation in prosodic phrasing or vowel quality were identified using subtraction and partial Pearson correlation analyses. Partial Pearson correlation analysis was also used to assess the connectivity between select regions of interest (e.g., Marrelec et al., 2007; Marrelec et al., 2008; Marrelec et al., 2006).

Based on previous findings we predicted that those speech manners that involved increased sequencing of linguistic contrasts would be associated with increased activity in left hemisphere prefrontal regions such as dorsal inferior frontal gyrus pars opercularis. Speech manners involving increased articulatory encoding of linguistic contrasts would be associated with increased activity in left hemisphere premotor regions such as middle premotor cortex. Speech manners that involved increased sequencing of prosodic contrasts would be associated with increased activity in right hemisphere prefrontal regions such as inferior frontal gyrus pars opercularis and frontal operculum.

While those speech manners involving increased articulatory encoding of prosodic contrasts would be associated with increased activity in right hemisphere premotor regions such as middle premotor cortex. Predictions of the neural responses expected in association with the phonological and phonetic encoding of linguistic and prosodic contrasts are summarized in Table 1-1.

Table 1-1: Predictions of cortical responses associated with increased phonological and phonetic encoding of linguistic and prosodic contrasts.

	Phonological Encoding of Linguistic Contrasts	Phonetic Encoding of Linguistic Contrasts	Phonological Encoding of Prosodic Contrasts	Phonetic Encoding of Prosodic Contrasts
Associated Cortical Brain Regions	L 	L 	 R	R 
Neuroimaging Studies Implicating the Cortical Brain Regions	Alario et al. (2006); Bohland and Guenther (2006); Friederici et al. (2006); Friederici et al. (2000); Friederici et al. (2003); Meyer et al. (2002); Papoutsis et al. (2009); Rieker et al. (2008)	Aziz-Zadeh et al. (2010); Indefrey and Levelt (2004); Mayer et al. (2002); Papoutsis et al. (2009); Peeva et al. (2010); Tremblay and Small, (2011); Wise et al. (1999)	Geiser et al. (2008); Klein et al. (2011); Meyer et al. (2002); Perrone et al. (2010)	

1.1 Organization of the Dissertation

The results of this study are described in the chapters to follow. The second chapter

describes the methods of the study (i.e., how the stimuli were constructed, how data were collected, and how they were analyzed). Chapter 3 and 4 describe the results of the acoustic and neuroimaging analyses, respectively. Chapter 5 concludes the dissertation with a discussion of possible directions for future related research.

CHAPTER 2

METHODS

2.1 Participants

Fourteen right-handed native speakers of English (balanced in gender; age range 18-35, mean age = 25) participated in this study after providing written informed consent in accordance with the guidelines of the Boston University Institutional Review Board and the Massachusetts Institute of Technology Human Research Committee. None of the study participants reported a neurological, speech, hearing, or voice impediment.

2.2 Stimuli

Many studies have shown that vowel quality changes as a consequence of rate, emphasis, and clarity (e.g., Fourakis, 1991; Gay, 1978; Lindblom, 1963; Ménard et al., 2007; Moon and Lindblom, 1994; Picheny et al., 1986). For example, Lindblom (1963) demonstrated that under conditions where rate is increased and stress is held constant there is a tendency for the midpoint values of vowel formant frequencies to fall short of their putative target values (although this deviation was more noticeable for F2 than F1 values in the data of Lindblom, 1963). Based on these findings, Lindblom (1963) developed the following model characterizing changes in F1 and F2 midpoint values as a function of the vowel's duration (τ) and its immediate consonantal frame:

$$F_{10} = \alpha_1 \cdot (375 - F_{1t}) \cdot e^{-\beta_1 \tau} + F_{1t}$$

$$F_{20} = \alpha_2 \bullet (F_{2i} - F_{2t}) \bullet e^{-\beta_2 \tau} + F_{2t}$$

The constants α and β are derived for each consonantal context independent of the vowel and $(F_{Ni} - F_{Nt})$ represents the difference between the initial formant frequency value of the vowel (i.e., “the locus”) and its putative “target” value. Figure 2-1 depicts the predicted F1 and F2 midpoint values for three vowels (/I/, /a/, and /U/) in three different consonantal frames as a function of the vowel’s duration.

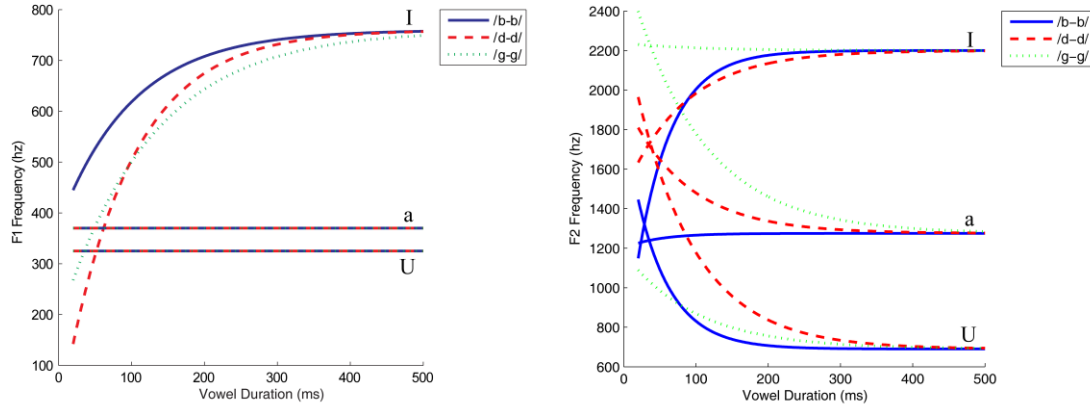


Figure 2-1: The predicted F1 (left) and F2 (right) midpoint values of three vowels (/I/, /a/, /U/) as a function of vowel duration and consonantal context.

As the vowel duration becomes shorter, the formant frequency values of the vowel become more similar to the surrounding consonants. According to the model of Lindblom (1963), the F1 midpoint values of /a/ and /U/ do not deviate from their putative target values. The predicted F1 and F2 values were determined by using the model and parameters described in the study of Lindblom (1963).

As illustrated in Figure 2-1, this model predicts that, as the vowel’s duration increases, the predicted formant frequency value asymptotically approaches its putative target value. For vowel durations that are less than 100 ms long, the formant frequency value for all three vowels within the same consonantal frame become acoustically similar, approaching a formant frequency value more characteristic of the consonantal frame. If the predicted F1 and F2 midpoint values of the three vowels (within a single consonantal

frame) for two different durations are plotted against each other, a reduced vowel triangle is observed. The displacement observed for vowel formant frequencies from more extreme positions in acoustic space to reduced positions is called “vowel reduction” and is demonstrated in Figure 2-2. This figure illustrates the way in which vowels can become acoustically less separable and therefore less distinct from each other as their duration decreases.

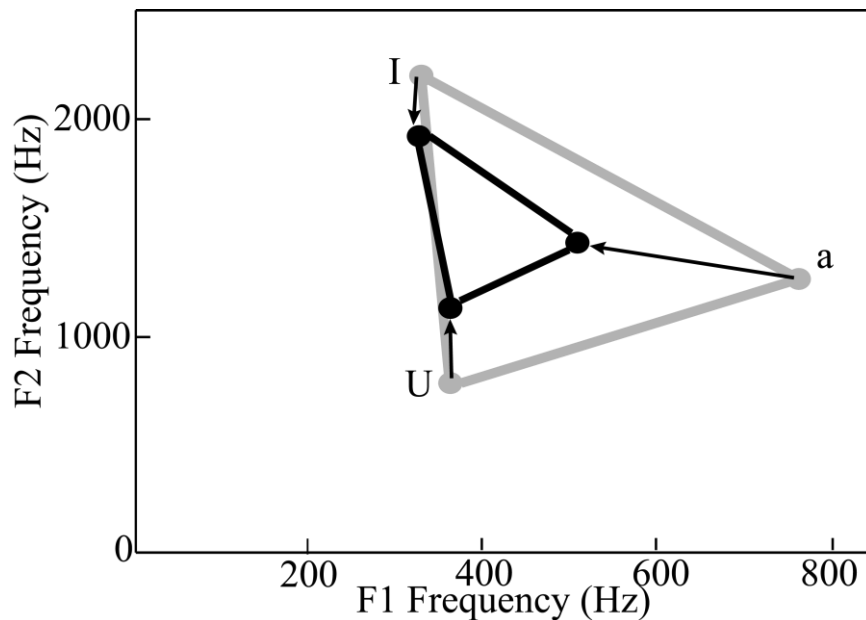


Figure 2-2: Schematic of reduced inter-vowel acoustic distances when vowel duration decreases from 500 ms (grey) to 100 ms (black).

Predicted F1 and F2 values for the three vowels (/I/, /a/, /U/) determined by using the model and parameters described in the study of Lindblom (1963) for vowel durations of 100 ms (black) and 500 ms (grey) and a /d-d/ context.

The speech material for the present study were designed to support the investigation of duration-induced vowel reduction by using sentence stimuli that each contained the four corner vowels of American English (/i/, /u/, /æ/, and /ɑ/). Since the acoustic properties of the vowels change depending on their position in the sentence each vowel appeared in

each of the four positions in the sentence analyzed, resulting in the 24 five syllable declarative sentence stimuli presented in Table 2-1.

Table 2-1: The 24 sentence stimuli used in the study.

For each sentence, the four corner vowels /i/, /u/, /æ/, and /ɑ/ were set within a consonant frame.

Pat got the soup beans.	Bob moves the back seat.	Dean had a juke box.	Jean got the boots back.
Dad chewed a hot beet.	Pat heats the tot's food.	Dad moves the deep pot.	Todd had the soup beans.
Scott pads the deep boots.	Jean scoots the pot back.	Duke got a deep bag.	Pete sued the bad cop.
Kat sees the juke box.	Jude packs the hot beets.	Scott keeps the suit tag.	Duke got the back seat.
Todd scoots the Jeep back.	Luke packs a deep box.	Bob keeps the bad boots.	Dean got the cat food.
Jude keeps the hat box.	Kat spots the deep boots.	Pete packs the hot food.	Luke keeps the tots back.

The six control stimuli used in the present study are presented in Table 2-2 and were designed to match the average length and visual density of the sentence stimuli while being completely devoid of lexical information. The characters used were box drawings from the unicode hexadecimal character set U250C-U256C from Microsoft Office Excel 2003. In order to sequence the box drawings, each alphabet letter was randomly assigned to a box drawing. Box drawings were then substituted as the letters in six of the 24 sentence stimuli.

Table 2-2: The six control stimuli used in the study.

The characters used were box drawings.

⌌⌌⌌ ⌌⌌⌌ ⌌ ⌌⌌⌌⌌ ⌌⌌⌌.	⌌⌌⌌ ⌌⌌⌌⌌ ⌌ ⌌⌌⌌ ⌌⌌⌌.
⌌⌌ ⌌⌌⌌⌌⌌ ⌌⌌ ⌌⌌⌌⌌ ⌌⌌⌌.	⌌⌌⌌ ⌌⌌⌌ ⌌⌌⌌ ⌌⌌⌌⌌⌌ ⌌⌌⌌.
⌌⌌⌌ ⌌⌌⌌⌌ ⌌ ⌌⌌⌌⌌ ⌌⌌⌌⌌.	⌌⌌⌌ ⌌⌌⌌⌌ ⌌ ⌌⌌⌌⌌ ⌌⌌⌌.

2.3 Experimental Protocol

During functional imaging, study participants performed an overt speech production task and an observational control task. The speakers could view the stimuli projected by a Hitachi (CP-X 1200 series) projector via an adjustable mirror positioned on the head coil. Figure 2-3 demonstrates a timeline for a single experimental trial. Each trial began with a 1 s orthographic presentation of a global speech manner (*normal*, *fast*, *clear*) or control (*watch*) cue. Immediately following this visual cue, the sentence or control stimulus was displayed on the screen for 3 s. Study participants were instructed to read the sentence stimulus out loud *as soon as it appeared* on the screen using the global speech manner indicated by the preceding visual cue. During this behavioral task period the scanner remained silent. Functional image acquisition began 1 s after the sentence disappeared from the screen and lasted 2.75 s. Image acquisition timing was chosen to include the putative peak in the delayed hemodynamic response associated with task performance; the hemodynamic response delay has been estimated to be between 4-7 s, depending on the brain region and task (Belin et al., 1999; Yang et al., 2000). A 7 s interval followed functional imaging before the onset of the next trial. This resulted in a total trial length of 14.75 s. Presentation Version 11 (www.neurobs.com) software was used to deliver the visual stimuli and trigger the scanner for image acquisition. Each experimental run contained 40 speech trials (10 trials for each speech manner) and 10 control trials. The trial order was pseudo-randomly permuted within each run so that no sentence stimulus could be presented more than once for a particular condition and no sentence could appear more than twice per run. Study participants were asked to perform

four, 12-minute runs. All speakers wore foam ear plugs, properly inserted into the ear canal during scanning.

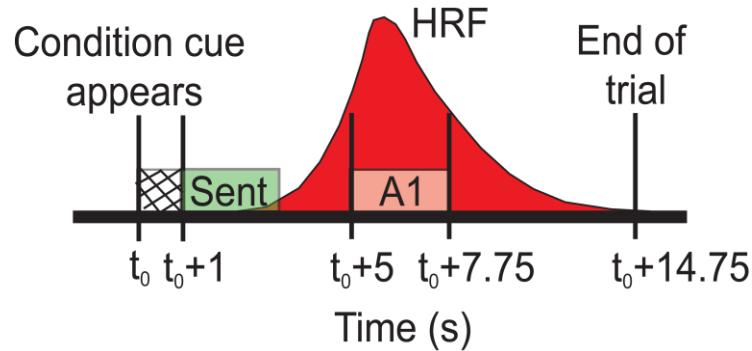


Figure 2-3: Timeline of a single trial.

At the onset of each trial, the speech manner/baseline visual cue appeared and remained on the screen for 1 s (rectangle with hash marks). Immediately following this visual cue, the sentence/control stimulus (Sent) was presented for 3 s. One second after speech/control task, one whole-brain volume was acquired (A1). Data acquisition was timed to cover the putative peak of the hemodynamic response to speech; the hemodynamic response function (HRF) is schematized in red. The next trial started 7 s after data acquisition was complete, resulting in a total trial length of 14.75 s.

In order to elicit similar speech manners across subjects, subjects were given explicit instructions about how to produce each speech manner. For the *normal* speech condition, speakers were instructed to read the sentences out loud at a conversational speech rate. In one half of all *normal* trials, the words in the first and fourth positions of the sentences were capitalized. Study participants were instructed to emphasize the capitalized words as though they were trying to answer a question (e.g., if the sentence were, “DUKE got a DEEP bag.” study participants were asked to respond with emphasis on the capitalized words as though they were answering the question, “*Who* got *which* bag?”), but otherwise produce the remaining words in these *emphatic* trials at a normal, conversational speech rate. For the *fast* speech trials, study participants were instructed to read the sentence out loud as rapidly as possible without slurring any of the words

together. Clear speech was elicited by asking the speakers to read the sentence out loud as though they were giving important instructions over a noisy telephone. For these *clear* speech trials, study participants were explicitly instructed not to increase their intensity, but to articulate the words as carefully as possible. For each *control* trial, the initial visual cue (*watch*) indicated that the speaker should remain silent and observe the control stimulus throughout the duration of the trial.

An MR-compatible microphone attached to the scanner bed transmitted study participants' acoustic productions to an IBM Thinkpad X60 inside the control room. The position of the microphone was individually adjusted for each subject in order to ensure the quality of the acoustic recordings. All recorded utterances were digitized at a 16 bit/22 kHz sampling rate and subsequently downsampled to 11 kHz.

2.4 Practice Session

Before entering the scanner, each speaker participated in a 15-minute (3 runs of 5 minutes) mock experiment that was designed to familiarize study participants with the sentence stimuli and conditions. During this practice session, the speech manner/control visual cue and sentence/baseline stimuli were the same as those used during functional imaging and appeared on the screen exactly as they would during the functional imaging experiment. The trial order was pseudo-randomly permuted within each of these mock sessions in the same way as the actual experiment. After completing each session, study participants received feedback about their productions to ensure consistent performance

during functional imaging (e.g., for clear speech conditions, the investigator would instruct the speaker to produce sentences before the sentence disappears.).

2.5 Neuroimaging Data Acquisition

Neuroimaging data were obtained from a 3.0 T Siemens Tim Trio scanner at the Athinoula A. Martinos Imaging Center at McGovern Institute for Brain Research, MIT. Images were acquired with a Siemens 12-channel head coil. Foam padding was packed around the subject's head to minimize head movement during the experiment. For each subject, a high resolution T1-weighted anatomical volume (128 slices in the sagittal plane, slice thickness = 1.33 mm, in-plane resolution = 1 mm^2 , TR = 2530 ms, TE = 3.39 ms, flip angle = 7° , matrix = 256×256 , FOV = 256 mm) was obtained prior to functional data acquisition. A sparse sampling method (Hall et al., 1999) was employed so that one whole-brain volume was acquired during each trial, just after each required speech or observational task had ended. This sparse fMRI design allowed the speaking task to be conducted in the absence of high frequency, high intensity image acquisition-related acoustic noise (Belin et al., 1999) and minimized image artifacts resulting from task-related movement (Birn et al., 1999). Functional volumes consisted of 45 axial slices, oriented along the bicommissural line, and covering the whole brain using a T2*-weighted gradient echo sequence (slice thickness = 3 mm, in plane resolution = 3.125 mm^2 , 10% slice gap, TR = 2750 ms, TE = 30 ms, flip angle = 90° , matrix = 64×64 , FOV = 200 mm).

2.6 Data Analyses

The analyses reported in this study focused on correct trials only. All trials in which sentences were incorrectly read were removed from acoustic analyses and regressed out in the first-level general linear model of the neuroimaging analyses. Production errors made during sentence reading included phonemic additions (“foods” instead of “food”), phonemic substitutions (“Dean” instead of “Jean”), and anticipatory errors (“hot box” instead of “hat box”). In total, 23 (out of 560) fast trials, 6 (out of 560) clear trials, 12 (out of 560) emphatic trials, and 13 (out of 560) normal trials were removed as a result of reading error. A recent study has shown that clear speech can be produced at a variety of speech rates (Krause and Braida, 2004). In order to ensure that the sentence-level durations of each condition were consistently produced by each speaker, each subject’s sentence-level duration distributions were plotted as boxplots and inspected by condition. Those duration values that fell outside the overall distribution pattern were considered outliers. These outlier trials were identified algorithmically by first calculating the range of duration values that comprised the middle 50% of values for each condition and subject (the interquartile range, IQR). Any duration value that was 1.5 times the IQR above 75% or below 25% of the duration values was considered an outlier. In total, 3 fast trials, 10 clear trials, 2 emphatic trials, and 5 normal trials were removed as a result of having sentence-level duration values outside the defined range. These outlier sentence-level duration values deviated by anywhere between 200 ms to 1 s from a subject’s mean value for a particular condition. In addition, in order to ensure that we only examined trials in which subjects followed task instructions and produced

sentences as soon as they were presented, we removed all trials with outlying reaction time values. The same outlier detection method employed to remove outlying sentence-level duration values was applied to remove trials with outlying reaction time values. In total, 9 fast trials, 1 clear trial, 2 emphatic trials, and 2 normal trials were removed as a result of having reaction time values that deviated substantially from either tail end of the distribution of reaction times for the condition and subject. These outlier reaction time values deviated by anywhere between 320 to 810 ms from a subject's mean value for a particular condition. Table 2-3 shows a breakdown of the trials removed by condition across the speakers as a percentage of the total number of trials.

Table 2-3: A breakdown of trials removed across subjects.

All trials in which sentences were incorrectly read or had outlying reaction time or sentence-level duration values were removed from acoustic analyses and regressed out in the first-level general linear model of the neuroimaging analyses.

	Total Number of Trials Included (of 560)	Reading Errors (% trials)	Reaction Time Outliers (% trials)	Sentence-Level Duration Outliers (% trials)
Fast	525	4.1	1.6	0.5
Normal	540	2.3	0.4	0.9
Emphatic	544	2.1	0.4	0.4
Clear	543	1.1	0.2	1.8

2.6.1 Acoustic Analysis

2.6.1.1 Segmentation

A custom built Matlab-based interactive display was used to mark the onsets and offsets of the corner vowels using both the spectrogram and RMS intensity contour as references. The vowel onset was identified as the time point where the local rise in the smoothed intensity contour was 50% of the local maximum and the vowel offset was

marked at the time point where the local fall in the smoothed intensity contour was 50% of the local minimum. These values were selected in order to minimize any effects on the formant values from neighboring consonants. The intensity contour was computed using a 15 ms moving Hamming window. Every identified vowel segment was inspected and manually corrected when algorithmic vowel boundary determination failed to identify the vowel segment entirely (such as when the syllable was partially or wholly devoiced) or when the preceding or subsequent consonant transitions were included within the vowel segment boundaries (such as for the word /muv/, where the preceding voiced nasal consonant transition of /m/ was often marked as part of the /u/ vowel segment). The interval between the labeled onset and the labeled offset of the vowel was defined as vowel duration.

Sentence-level onsets and offsets were identified using a custom-written script for PRAAT (v.5.1.41; Boersma, 2001). The sentence boundaries were identified using a smoothed intensity contour computed with a moving 80 ms Gaussian window. The onset of the sentence was identified as the time point where the rise in the smoothed intensity contour was more than 50% of the maximum intensity value for the sentence and remained above this threshold for a speaker-dependent period of time, and the sentence offset was marked as the time point where the smoothed intensity contour decreased to less than 50% of the maximum intensity for the sentence and remained below this threshold for a speaker-dependent period of time. Sentence-level segmentations were inspected during sentence boundary classification and manual edits were made as

needed. The interval between the labeled onset and offset of the sentence was defined as sentence-level duration.

2.6.1.2 Pre-processing

A linear predictive coding (LPC)-based tracking algorithm in PRAAT was used to track the first five formant frequency values. The LPC analysis employed a moving 25 ms Gaussian window with a +6 dB/octave pre-emphasis to boost the frequencies over 50 Hz. The frequency range analyzed for the LPC analysis was adjusted separately for males and females (upper bound = 5,000 Hz for males and 5,500 Hz for females). The first two formant frequencies (F1 and F2) of each corner vowel were then extracted at five equally spaced time points between the labeled vowel onset and offset. The formant values at each of the time points were averaged across the vowel segment and converted from Hertz to mels, a perceptually based frequency scale, using the formula in PRAAT ($MN = 550 * \ln(1 + FN/550)$ where FN is the Nth formant frequency in Hz), yielding a single M1 and M2 mel value for each vowel in the sentence.

Acoustic periodicity was identified with PRAAT using the autocorrelation method described in Boersma (1993). Fundamental frequency (F0) values were extracted at five equally spaced time points within the vowel segment. These five values were averaged and converted from Hz to semitones (st) using the formula $F0_{st} = 12 * \log_2(F0_{Hz})$, yielding a single F0 value for each of the four corner vowels in the sentence. Root-mean square sound pressure level values were extracted at five equally spaced time points within the vowel segment from a smoothed intensity contour computed with a moving 32 ms

Gaussian analysis window. These five values were averaged, yielding a single intensity value (in db SPL) for each corner vowel.

2.6.1.3 Characterizing Acoustic Variation

We predicted that when compared to normal conversational speech, fast speech production would be accompanied by shorter vowel durations, reduced inter-vowel acoustic distances, and greater vowel acoustic variability, while clear and emphatic speech production would be accompanied by longer vowel durations, larger inter-vowel acoustic distances, and less vowel acoustic variability. Figures 2-4 and 2-5, provide the waveform, spectrogram, and vowel segmentations of a male speaker producing the sentence, “Duke got a deep bag” under clear, normal, and fast conditions. The spectrogram demonstrates that the clear and fast speech manners fall on opposite ends of a production duration continuum, with clear speech being produced for a significantly longer duration than either normal or fast speech.

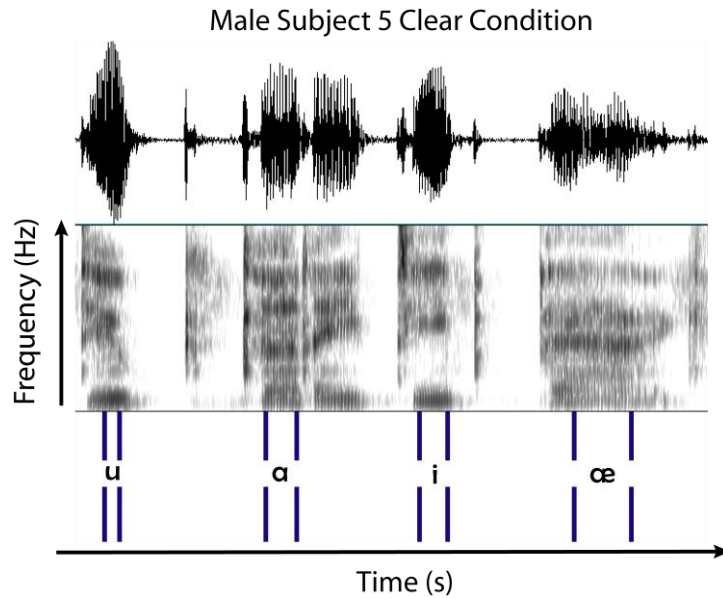


Figure 2-4: Waveform, spectrogram, and onsets and offsets of the corner vowels for male subject 5's clear production of the sentence, "Duke got a deep bag."

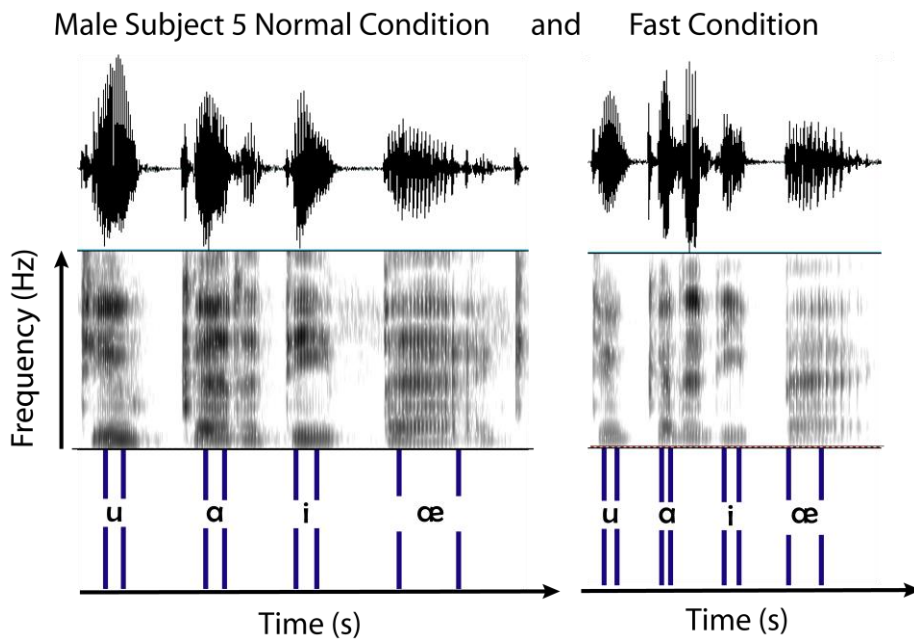


Figure 2-5: Waveform, spectrogram, and onsets and offsets of the corner vowels for male subject 5's normal (left) and fast (right) productions of the sentence, "Duke got a deep bag."

Figure 2-6 demonstrates acoustic differences in vowel productions for the two conditions that are on the most extreme ends of a sentence-level production duration continuum:

clear (plotted in black) and fast (plotted in grey) speech. For subject 5 and the sentence, “Duke got a deep bag” the average vowel spacing (left) shrinks as conditions change from clear to fast speech. These results indicate that there is a tendency for the vowels in the sentence to be less separable acoustically under fast speech conditions as compared to clear speech conditions. Figure 2-6 (right) presents ellipses that cover 68% of the variance in the first two formant frequencies of all of subject 5’s corner vowel productions for the clear (plotted in black) and fast (plotted in grey) speech conditions. This figure demonstrates that across all vowel productions under clear speech conditions vowel tokens tend to cluster more tightly around their mean value as compared to those produced under fast speech conditions. As a result, this subject’s clear vowel tokens are acoustically more consistent as compared to those produced under fast speech conditions.

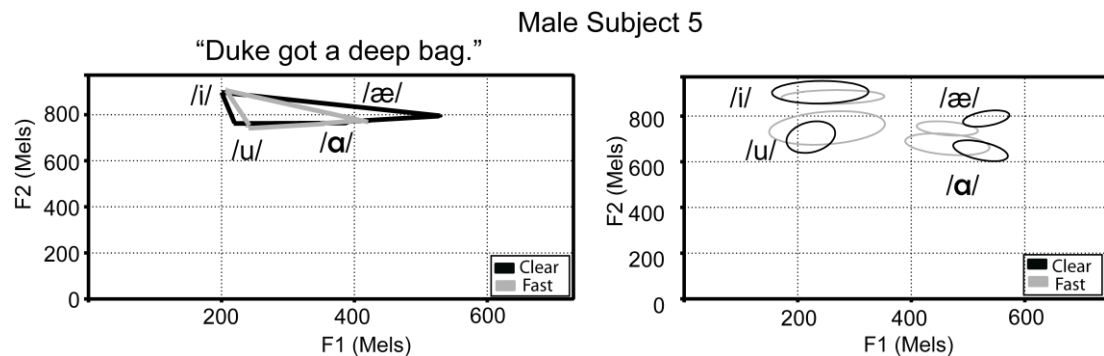


Figure 2-6: Average vowel spacing (left) and vowel dispersion (right) for the clear (plotted in black) and fast (plotted in grey) speech conditions.

Figure 2-7 presents the inter-vowel acoustic distances and acoustic vowel variability for subject 5’s emphatic (plotted in grey) and normal (plotted in black) productions. These plots demonstrate that for this subject except for the vowel /i/, acoustic differences in vowel quality between the normal and emphatic conditions are not significant.

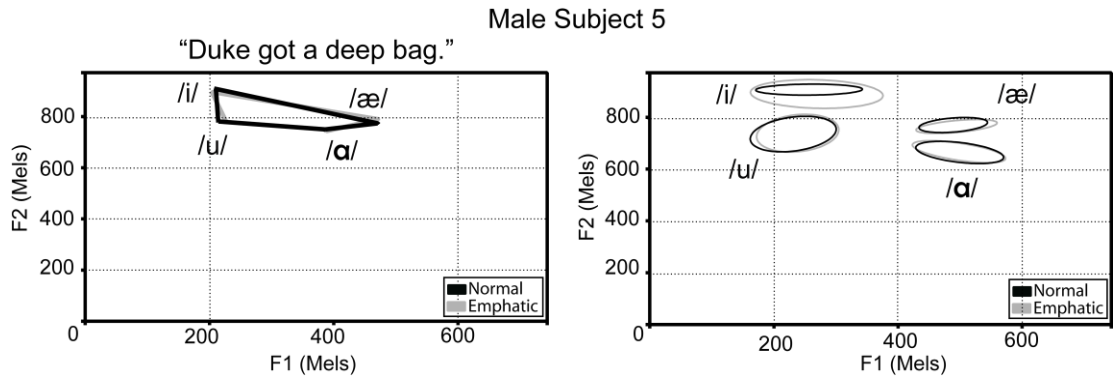


Figure 2-7: Average vowel spacing (left) and vowel dispersion (right) for subject 5’s emphatic (plotted in grey) and normal (plotted in black) conditions.

Relative to unstressed syllables, stressed syllables are usually higher in pitch, longer in duration, and greater in intensity (Kent and Read, 2002). These changes in prosodic contrasts make stressed syllables acoustically more salient relative to unstressed syllables. The primary locus for prosodic changes is the syllable nucleus, which in the present study is the vowel. Figure 2-8 displays the F0 contour, intensity contour, and duration of the vowels for the emphatic (plotted in grey) and normal (plotted in black) speech conditions. The F0 of the vowel /u/ in the word “Duke” does not differ significantly between the emphatic and normal conditions, despite the fact that under emphatic conditions this vowel is in a word that receives emphasis. Interestingly, the adjacent vowel, /ɑ/ in the word “got,” is lower in F0 under emphatic conditions as compared to normal conditions. The difference in F0 between adjacent stressed and unstressed vowels (hereafter referred to as *adjacent-vowel F0 contrast*) may have been intended by the speaker to make the vowel /u/, in the word, “Duke,” acoustically more salient relative to the adjacent vowel /ɑ/, in the word, “got,” under emphatic conditions as compared to normal conditions. Under emphatic conditions, we also find that the F0

of the vowel, /i/, in the emphasized word, “deep,” is higher than either the preceding vowel /ɑ/, in the word, “got,” or the following adjacent vowel, /æ/, in the word, “bag.” The effect of the increased F0 for the vowel /i/ relative to its neighbors is to divide the sentence into two intermediate intonational phrases as opposed to a single continuously descending F0 contour that is typical of declarative sentences and is observed under normal conditions for this subject and sentence. Therefore, for this sentence, subject 5 applies local changes to the prosodic contrasts under emphatic speech conditions. Subject 5 also applies phrase-level changes to the prosodic contrasts under emphatic speech conditions. For example, as can be seen in Figure 2-8 (right), subject 5 produced the four corner vowels with greater-than-normal intensity for the emphatic speech condition.

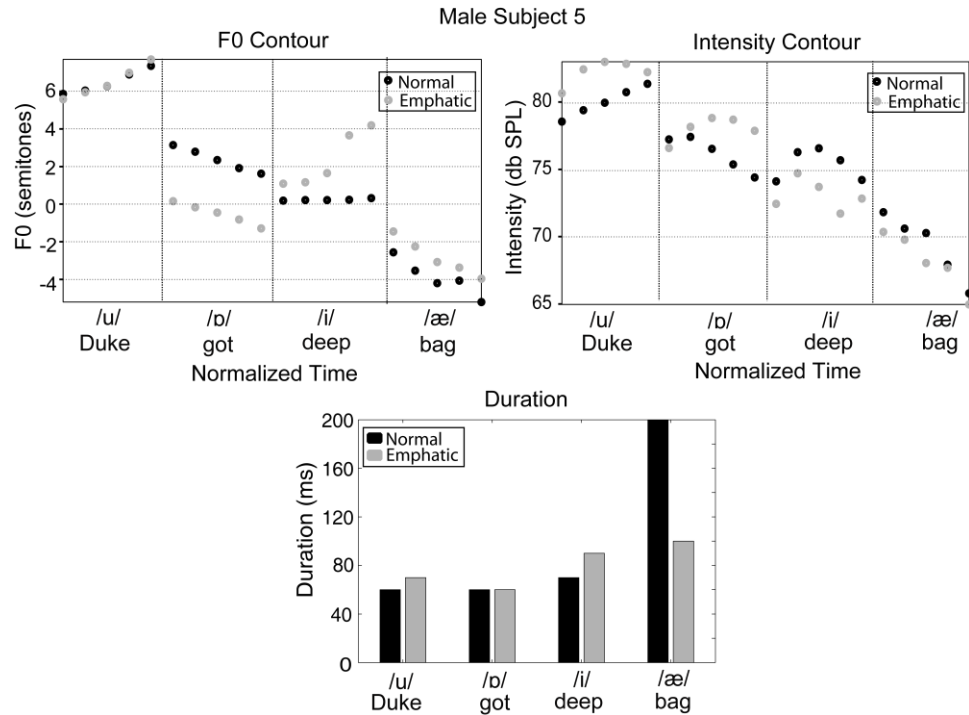


Figure 2-8: Plots of the F0 contour, intensity contour, and duration of the vowels in the sentence, “Duke got a deep back” for subject five’s emphatic (plotted in grey) and normal (plotted in black) productions.

The main predictions of the acoustic differences between each speech manner (fast, clear, and emphatic) and normal speech in terms of prosodic phrasing and vowel quality are summarized in Table 2-4. The < and > signs indicate cases where one of the speech manner manipulations will result in a value of the measure that is significantly less or greater than the normal condition.

Table 2-4: Main predictions of the acoustic differences between each speech manner (fast, clear, and emphatic) and normal speech in terms of prosodic phrasing and vowel quality.

The < and > signs indicate cases where one of the speech manner manipulations (fast, clear, emphatic) will result in a value of the measure that is significantly less than or greater than normal.

	Vowel Space	Vowel Dispersion	Vowel Duration	Vowel F0	Vowel Intensity	Adjacent-vowel F0 Contrast	Adjacent-vowel Intensity Contrast	Adjacent-vowel Duration Contrast
Fast – Normal	<	>	<					
Clear – Normal	>	<	>					
Emphatic – Normal	>	<	>	>	>	>	>	>

Provided in detail in the following subsections are mathematical descriptions of the eight different acoustic measures utilized in this study that characterize changes in the vowel quality (average vowel spacing and average vowel dispersion) and in the prosodic contrasts at the phrase-level (average vowel duration, average vowel F0, and average vowel intensity) and at the local-level (average adjacent-vowel duration contrast, average adjacent-vowel intensity contrast, average adjacent-vowel F0 contrast).

2.6.1.4 Average Vowel Spacing (Mels)

To quantify inter-vowel acoustic distances for each trial and subject, the acoustic Euclidean distance between all 6 pairs of vowels in the sentence (/i/-/u/, /i/-/æ/, /i/-/ɑ/, /u/-/æ/, /u/-/ɑ/, and /æ/-/ɑ/) were computed. These distances were averaged to obtain an average inter-vowel acoustic distance in Mels for a trial for each subject. The formula for average vowel spacing is provided on the next page:

$$AVS = \frac{2}{v(v-1)} \left\{ \sum_{i=1}^{v-1} \sum_{j=i+1}^v [(M1_i - M1_j)^2 + (M2_i - M2_j)^2]^{\frac{1}{2}} \right\}$$

where MN_i and MN_j are the Nth mel values of the vowel in the i th position and j th position of the sentence. The variable v is the total number of vowels analyzed in the sentence. For this study the number of vowels analyzed in the sentence equaled the number of vowel categories in the sentence.

2.6.1.5 Average Vowel Dispersion (Mels)

To assess acoustic precision of vowel production, the acoustic Euclidean distance between each vowel (/i/, /u/, /æ/, and /ɑ/) in a sentence and the average acoustic position of all tokens of that vowel were computed in Mels for each condition. This measure of dispersion was averaged across the vowels, yielding an average vowel dispersion value in Mels for each sentence. The formula for average vowel dispersion is provided below:

$$AVD = \frac{1}{v} \left\{ \sum_{i=1}^v [(M1_{i,c} - \overline{M1_c})^2 + (M2_{i,c} - \overline{M2_c})^2]^{\frac{1}{2}} \right\}$$

where $MN_{i,c}$ is the Nth mel value of the vowel in the i th position of the sentence within vowel category, c , and $\overline{MN_c}$ is the mean Nth mel value for the vowel token's corresponding vowel category. The variable v is the total number of vowels analyzed in the sentence.

2.6.1.6 Average Vowel F0 (Semitones)

The measure of vowel F0 in the present study took into account the fact that vowel categories differ in terms of their intrinsic F0. For example high vowels (/i/ and /u/) tend

to have a higher F0 than low vowels (/æ/ or /ɑ/) due to mechanical factors that determine their production (Whalen and Levitt, 1995). The measure of vowel F0 in the present study factors out these intrinsic differences in F0 so that differences in F0 across the vowels that are a direct consequence of changes in speech manner can be examined. To compute this average measure of vowel F0 across the sentence, the mean F0 value was computed for each vowel category (/i/, /u/, /æ/, and /ɑ/) produced under the normal speech condition. This mean F0 value for each vowel category was then subtracted from the F0 value of the vowel token in the sentence that belonged to the same vowel category (/i/, /u/, /æ/, and /ɑ/). This normalized measure of vowel F0 was then averaged across the four vowel categories, yielding a single average vowel F0 measure in semitones for the sentence. The formula is provided below:

$$AVF0 = \frac{1}{v} \left\{ \sum_{i=1}^v (F0_{i,c} - \overline{F0_c}) \right\}$$

where $F0_{i,c}$ is the F0 value of the vowel token produced in the i th position of the sentence within vowel category, c , and $\overline{F0_c}$ is the mean F0 value of the vowel token's corresponding vowel category for the normal (non-emphatic) speech condition. The variable, v , is the total number of vowels analyzed in the sentence.

2.6.1.7 Average Adjacent-vowel F0 Contrast (Semitones)

The average adjacent-vowel F0 contrast measure also accounted for differences in intrinsic F0 between vowel categories (/i/, /u/, /æ/, and /ɑ/) by first computing the mean F0 value for each vowel category produced under the normal speech condition. This mean F0 value was subtracted from the F0 value of the vowel token in the sentence that

belonged to the same vowel category (/i/, /u/, /æ/, and /ɑ/). The difference between the values of this normalized F0 measure for the vowels in the first and second position and for those in the fourth and fifth position of the sentence were computed and averaged to yield a single average adjacent-vowel F0 contrast measure in semitones for each sentence. The formula is provided below:

$$AAVF0 = \frac{1}{(v/2)} \left\{ \sum_{i=1,4} [(F0_{i,c} - \overline{F0_c})] - [(F0_{i+1,ac} - \overline{F0_{ac}})] \right\}$$

where $F0_{i,c}$ is the F0 value of the vowel token produced in the 1st and 4th positions of the sentence within vowel category, c , and $\overline{F0_c}$ is the mean F0 value of the vowel token's corresponding vowel category for the normal (non-emphatic) speech condition. $F0_{i+1,ac}$ is the F0 value of the adjacent vowel token within the adjacent vowel category, ac , and $\overline{F0_{ac}}$ is the mean F0 value of the adjacent vowel token's corresponding vowel category for the normal (non-emphatic) speech condition. The variable, v , is the total number of vowels analyzed in the sentence.

2.6.1.8 Average Vowel Intensity (dB SPL)

The measure of vowel intensity in the present study took into account the fact that vowel categories differ in terms of their intrinsic intensities. For example high vowels (/i/ and /u/) tend to have a lower intensity than low vowels (/æ/ or /ɑ/) due to mechanical factors that determine their production (Lehiste and Peterson, 1959). The measure of vowel intensity in the present study factors out these intrinsic differences in intensity so that differences in intensity across the vowels that are a direct consequence of the speech

manner can be examined. To compute this average measure of vowel intensity across the sentence, the mean intensity value was computed for each vowel category (/i/, /u/, /æ/, and /ɑ/) produced under the normal speech condition. This mean intensity value for each vowel category was then subtracted from the intensity value of the vowel token in the sentence that belonged to the same vowel category (/i/, /u/, /æ/, and /ɑ/). This normalized measure of vowel intensity was then averaged across the four vowel categories, yielding a single average vowel intensity measure in dB SPL for the sentence. The formula is provided below:

$$AVI = \frac{1}{v} \left\{ \sum_{i=1}^v (I_{i,c} - \overline{I}_c) \right\}$$

where $I_{i,c}$ is the intensity value of the vowel token produced in the i th position of the sentence within vowel category, c , and \overline{I}_c is the mean intensity value of the vowel token's corresponding vowel category for the normal (non-emphatic) speech condition. The variable, v , is the total number of vowels analyzed in the sentence.

2.6.1.9 Average Adjacent-vowel Intensity Contrast (dB SPL)

The average adjacent-vowel intensity contrast measure also accounted for differences in intrinsic intensity between vowel categories by first computing the mean intensity value for each vowel category (/i/, /u/, /æ/, and /ɑ/) produced under the normal speech condition. This mean intensity value was subtracted from the intensity value of the vowel token in each sentence that belonged to the same vowel category (/i/, /u/, /æ/, and /ɑ/). The difference between the values of this normalized intensity measure for the vowels in the first and second positions and for the vowels in the fourth and fifth

positions of the sentence were computed and averaged to yield a single average adjacent-vowel intensity contrast measure in dB SPL for each sentence. The formula is provided below:

$$AAVI = \frac{1}{(v/2)} \left\{ \sum_{i=1,4} [(I_{i,c} - \overline{I_c})] - [(I_{i+1,ac} - \overline{I_{ac}})] \right\}$$

where $I_{i,c}$ is the intensity value of the vowel token produced in the 1st and 4th positions of the sentence within vowel category, c , and $\overline{I_c}$ is the mean intensity value of the vowel token's corresponding vowel category for the normal (non-emphatic) speech condition. $I_{i+1,ac}$ is the intensity value of the adjacent vowel token and $\overline{I_{ac}}$ is the mean intensity value of the adjacent vowel token's corresponding vowel category for the normal (non-emphatic) speech condition. The variable, v , is the total number of vowels analyzed in the sentence.

2.6.1.10 Average Vowel Duration (ms)

The measure of vowel duration in the present study took into account the fact that vowel categories differ in terms of their intrinsic duration. For example high vowels (/i/ and /u/) tend to be shorter in duration than low vowels (/æ/ or /ɑ/) due to mechanical factors that determine their production (Peterson and Lehiste, 1960). The measure of vowel duration in the present study factors out these intrinsic differences in duration so that differences in duration across the vowels that are a direct consequence of changes in speech manner can be examined. To compute this average measure of vowel duration across the sentence, the mean duration value was computed for each vowel category (/i/, /u/, /æ/,

and /ɑ/) produced under the normal speech condition. This mean duration value for each vowel category was then subtracted from the duration value of the vowel token in the sentence that belonged to the same vowel category (/i/, /u/, /æ/, and /ɑ/). This normalized measure of vowel duration was then averaged across the four vowel categories, yielding a single average vowel duration measure in ms for the sentence. The formula is provided below:

$$AVD = \frac{1}{v} \left\{ \sum_{i=1}^v (D_{i,c} - \overline{D_c}) \right\}$$

where $D_{i,c}$ is the duration value of the vowel token produced in the i th position of the sentence within vowel category, c , and $\overline{D_c}$ is the mean duration value of the vowel token's corresponding vowel category for the normal (non-emphatic) speech condition. The variable, v , is the total number of vowels analyzed in the sentence.

2.6.1.11 Average Adjacent-vowel Duration Contrast (ms)

The average adjacent-vowel duration contrast measure also accounted for differences in intrinsic duration between vowel categories by first computing the mean duration value for each vowel category produced under the normal speech condition (/i/, /u/, /æ/, and /ɑ/). This mean duration value was subtracted from the duration value of the vowel token in the sentence that belonged to the same vowel category (/i/, /u/, /æ/, and /ɑ/). The difference between the values of this normalized duration measure for the vowels in the first and second positions of the sentence and between the vowels in the fourth and fifth positions of the sentence were then computed. These two values were then averaged to

yield a single average adjacent-vowel duration contrast measure in ms for each sentence.

The formula is provided on the next page:

$$AADC = \frac{1}{(v/2)} \left\{ \sum_{i=1,4} [(D_{i,c} - \overline{D_c})] - [(D_{i+1,ac} - \overline{D_{ac}})] \right\}$$

where $D_{i,c}$ is the duration value of the vowel token produced in the 1st and 4th positions of the sentence within vowel category, c , and $\overline{D_c}$ is the mean duration value of the vowel token's corresponding vowel category for the normal (non-emphatic) speech condition. $D_{i+1,ac}$ is the duration value of the adjacent vowel token and $\overline{D_{ac}}$ is the mean duration value of the adjacent vowel token's corresponding vowel category for the normal (non-emphatic) speech condition. The variable, v , is the total number of vowels analyzed in the sentence.

2.6.1.12 Statistical Analysis of the Acoustic Data

Statistical analysis of the acoustic data was performed using SAS Version 8.02 (SAS Institute Inc., <http://www.sas.com/>). Separate repeated measures analyses of variance (RMANOVA) were used to determine the influence of speech condition, speaker gender, and their interaction on each dependent variable (reaction time, sentence-level acoustic duration, average vowel spacing, average vowel dispersion, average vowel F0, average adjacent-vowel F0 contrast, average vowel intensity, average adjacent-vowel intensity contrast, average vowel duration, and average adjacent-vowel duration contrast). To estimate these effects, the group analysis was carried out using a mixed effect design in which each subject within each group is considered a random factor. For tests of the

main effects and their interaction a p -value < 0.05 was considered significant. For the 28 pair-wise contrasts a p -value < 0.05 FDR-corrected for multiple comparisons was used.

2.6.2 Neuroimaging Analysis

2.6.2.1 Surface-based Analysis

Brain reconstruction was generated using Freesurfer image analysis software on each subject's anatomical image (Dale et al., 1999; Fischl et al., 1999a). This software generates an inflated representation of the cortical surface after several processing stages, including intensity normalization, skull stripping, and white matter segmentation. Each subject's cortical surface is then registered to a spherical surface-based atlas (Fischl et al., 1999b). Functional neuroimaging data were processed using Nipype 0.4 (Gorgolewski et al., 2011). Preprocessing of fMRI data included realignment, image motion and intensity outlier identification, and registration. Motion correction was performed with SPM 8 and functional images were realigned to the mean echo planar image of the first session. Functional image volumes with 2 mm scan-to-scan deviations in the motion parameters and/or absolute intensity values that deviated by 3 standard deviations from the global mean signal were identified using the RapidArt toolbox (<http://www.nitrc.org/projects/rapidart/>). The mean echo planar image was then registered to the subject's T1-weighted anatomical image using Freesurfer. No smoothing was applied at the preprocessing stage of the analysis. First-level within-subject analyses were performed in the subject's native space using SPM 8. To reduce the effect of low-frequency noise, functional time-series data were temporally high-pass filtered at 128 s. Task-dependent blood oxygenation level dependent (BOLD) responses

were estimated for each subject using a general linear model. For each run, the onset and duration of each condition's image acquisition were modeled, creating five covariates of interest that were convolved with a finite impulse response (FIR) function. The following additional, subject-specific sources of variance were included as covariates of non-interest: (i) realignment parameters; (ii) image intensity and motion outliers; (iii) a linear detrending variable; and (iv) behavioral error trials. Despite the fact that neural populations can activate in under 10 ms, the hemodynamic response is slower requiring at least 1 second for blood flow to increase (Howseman et al., 1997) and therefore it is widely believed that subsecond temporal information cannot be resolved using fMRI time series. For this reason, we do not include reaction time as a covariate of non-interest in the first-level general linear model of the neuroimaging analysis since across subjects the mean difference in reaction times between the fastest and slowest conditions was only 117 ms. However, since the mean difference in the sentence duration of the fastest and slowest conditions was 1.03 s, sentence duration was added as a covariate of non-interest in the first-level general linear model when we assessed the six main contrasts of interest (Clear – Normal, Fast – Normal, Clear – Fast, Stress – Normal, Clear – Stress, Stress – Fast) for each subject.

For the second-level surface-based group analysis, each subject's volume-based contrast image was resampled to Freesurfer's fsaverage space, spatially smoothed with a 6 mm full-width/half-maximum (fwhm) Gaussian kernel, and entered into a two-sided one-sample t-test. A cluster is considered significant if all vertices inside the cluster meet threshold requirements ($p < 0.001$ uncorrected) and if the surface area of significant

contiguous vertices is larger than the maximum cluster size that could occur by chance 2.5% of the time for each hemisphere as determined by 10,000 Monte Carlo simulations. For visualization/reporting purposes, those significant vertices ($p < 0.001$) within clusters surpassing the cluster extent threshold ($p < 0.025$) for each hemisphere are plotted on Freesurfer's fsaverage inflated surface template with colors reflecting the significant p-values. Hot (red) colors represent significant vertices specific to the first condition of the contrast. Tables, provided in the results section, indicate the label, hemisphere, coordinate, and p-value of the peak vertex within the significant cluster as well as the p-value and surface area of the significant cluster for each contrast of interest. Talairach coordinates were converted into MNI coordinates using the tal2icbm transformation developed by Lancaster et al. (2007).

2.6.2.2 Region of Interest Analysis

Direct comparisons between condition-specific BOLD responses were also assessed for 26 functionally-defined cortical regions of interest (ROIs) and three anatomically-defined subcortical ROIs in a subsequent region of interest (ROI) analysis. Cortical ROIs were defined by creating a composite map of the six group-level surface-based contrasts for the left and right hemispheres. The location of the vertex with the most significant p-value inside each cluster was selected to be the point around which a circular ROI was defined. In cases where there was overlap between two clusters as in left anterior frontal operculum, left dorsal somatosensory cortex, and right posterior dorsal superior temporal cortex, only a single ROI was used and its location was taken to be the most significant vertex across the clusters in the region. A circular region of interest with an ~6 mm

radius was delineated on Freesurfer's fsaverage cortical surface around the location of the most significant vertex. These clusters were then mapped to each individual's space and projected into the volume. The three subcortical ROIs (putamen, pallidum, and thalamus) were parcellated according to a training set provided by Freesurfer (Fischl et al., 2002). Using a stand-alone MATLAB-based toolbox called REX (<http://www.nitrc.org/projects/rex/>) the mean BOLD signal of all voxels in each ROI was extracted and scaled by the mean BOLD signal across all voxels in the ROI across all scans. The mean regional responses for each stimulus event were modeled using a finite impulse response (FIR) basis function and the same set of regressors that were used in the first-level design matrix of the voxel-based analyses. Group-level effects were assessed by first computing regional contrasts for each subject. The regional contrasts were then pooled across subjects and individual ROIs were tested for significance at the group level using one-sample *t*-tests. A threshold of $p_{\text{FDR}} \leq 0.05$ was used to determine significant effects at the group level.

2.6.2.3 Partial Pearson's Correlation Analysis

Separate partial Pearson's correlation analyses were conducted to assess the relationships of significant cluster activity with the acoustic measures found to show statistically significant differences between the conditions in the acoustic analysis. The acoustic measures that were selected (average vowel intensity, average vowel duration, average vowel spacing, average adjacent-vowel F0 contrast, average adjacent-vowel intensity contrast, and average adjacent-vowel duration contrast) were considered to be the best representatives of changes in the vowel quality and prosodic phrasing observed with

changes in speech manner. Partial correlation is used to measure the association between two time series, such as a region and an acoustic measure, after the effects of other time series, such as the effects of other acoustic measures, are partialled out of both time series. Each of the six acoustic measure time series used for the partial Pearson correlation analysis were centered (with zero mean) for each participant to ensure that the conditional correlations reflect variability due to the experimental manipulations as opposed to between-subject differences in amplitude. Partial correlation coefficients were generated for each subject by computing the partial Pearson's correlation between the time-courses of the acoustic measure and the cluster-based ROIs while controlling for the effects of the remaining five acoustic measures and the conditions. To combine correlation coefficients across subjects and compute group-level statistical significance, correlation coefficients were converted to standard values using Fischer's r -to- z transform. For the second-level group analysis, two-sided one sample t -tests were performed and a threshold of $p_{FDR} \leq 0.05$ was used to determine significant partial correlations. Positive t -values indicate that an increase in the mean response of the cluster-based ROI is correlated with an increase in the acoustic measure, controlling for the remaining five acoustic measures and the five conditions. Negative t -values identify cluster-based ROIs in which significant decreases in mean BOLD response in the cluster-based ROI are correlated with increases in the acoustic measure, controlling for the remaining acoustic measures and the conditions.

2.6.2.4 Functional Connectivity Analysis

A partial Pearson correlation approach was also used to investigate the strength of connectivity between ROIs. Partial correlation provides an estimate of the degree of interaction between two ROIs after controlling for mutual dependencies of these regions with other regions in the network (Marrelec et al., 2007; Marrelec et al., 2008; Marrelec et al., 2006). Using a stand-alone MATLAB-based toolbox called REX (<http://www.nitrc.org/projects/rex/>) the mean BOLD signal of voxels in three subcortical anatomical regions of interest (left putamen, left pallidum, and left thalamus) and three spherical cortical regions of interest (left posterior dorsal inferior frontal gyrus, left middle premotor cortex, and left ventral motor cortex) was extracted and scaled by the mean BOLD signal across all voxels in the ROI across all scans. Partial Pearson correlation coefficients were then calculated for the separate conditions by computing the conditional correlation between the time series of each pair of ROIs for each subject, while controlling for the remaining ROIs. To combine partial correlation coefficients across subjects and compute group-level significance, correlation coefficients were converted to standard scores using Fischer's *r*-to-*z* transform. Significant differences in connectivity between conditions were tested using a paired *t*-test and a threshold of $p_{FDR} \leq 0.05$.

CHAPTER 3

ACOUSTIC ANALYSIS RESULTS

For each dependent variable (reaction time, sentence-level duration, average vowel dispersion, average vowel spacing, average vowel F0, average adjacent-vowel F0 contrast, average vowel intensity, average adjacent-vowel intensity contrast, average vowel duration, and average adjacent-vowel duration contrast) a separate, mixed effects design repeated measures ANOVA was performed with speech condition (clear, normal, fast, and emphatic) as a within-subject factor and speaker gender (male or female) as a between-subjects factor and each subject within each group considered to be a random factor. Table 3-1 provides a statistical summary of the main effects and the interactions for these analyses and the findings are described in detail in the sections that follow.

Table 3-1 Group-level RMANOVA results

DEPENDENT VARIABLE	CONDITION	GENDER	INTERACTION
Reaction Time	$F_{3, 36} = 27.3^*$	$F_{1, 12} = 1.0$	$F_{3, 36} = 1.1$
Sentence-level Duration	$F_{3, 36} = 107.6^*$	$F_{1, 12} = 0.8$	$F_{3, 36} = 0.5$
Average Vowel Dispersion	$F_{3, 36} = 2.3$	$F_{1, 12} = 0.6$	$F_{3, 36} = 1.6$
Average Vowel Spacing	$F_{3, 36} = 37.2^*$	$F_{1, 12} = 2.0$	$F_{3, 36} = 3.1^*$
Average Vowel Duration	$F_{3, 36} = 36.1^*$	$F_{1, 12} = 0.2$	$F_{3, 36} = 0.3$
Average Adjacent-vowel Duration Contrast	$F_{3, 36} = 18.3^*$	$F_{1, 12} = 1.3$	$F_{3, 36} = 1.0$
Average Vowel F0	$F_{3, 36} = 2.7$	$F_{1, 12} = 0.1$	$F_{3, 36} = 0.5$
Average Adjacent-vowel F0 Contrast	$F_{3, 36} = 31.3^*$	$F_{1, 12} = 6.6^*$	$F_{3, 36} = 4.0^*$
Average Vowel Intensity	$F_{3, 36} = 11.5^*$	$F_{1, 12} = 0.5$	$F_{3, 36} = 1.2$
Average Adjacent-vowel Intensity Contrast	$F_{3, 36} = 31.5^*$	$F_{1, 12} = 1.3$	$F_{3, 36} = 1.9$

* indicates significant $p < 0.05$

3.1 Reaction Time

Average reaction time values for the group are displayed as a function of speech condition in Figure 3-1. Table 3-1 indicates that the influence of condition on reaction time is statistically significant ($F_{3, 36} = 27.3$, $p < 0.0001$). Comparisons of conditional effects indicate that speakers are significantly faster to respond under fast speech conditions than when speaking normally ($F_{1, 36} = 40.9$, $p_{FDR} < 0.05$), with clarity ($F_{1, 36} = 10.7$, $p_{FDR} < 0.05$), or with emphasis ($F_{1, 36} = 71.3$, $p_{FDR} < 0.05$). In this study speakers did take significantly more time before initiating speech under normal ($F_{1, 36} = 9.7$, $p_{FDR} < 0.05$) and emphatic ($F_{1, 36} = 26.7$, $p_{FDR} < 0.05$) conditions as compared to clear conditions. The slower reaction time for the emphatic and normal conditions likely reflects an ambiguity in the initial speaking manner cue (which was *normal* for the emphatic and normal conditions) and did not explicitly indicate whether the forthcoming sentence would require emphasis. Based on the assumption that speed of production initiation reflects sentence production planning, these results suggest that knowledge of the prominence pattern is needed before sentence production initiation (cf., Keating and Shattuck-Hufnagel, 2002). The non-significant gender variable for reaction time confirms that males and females did not produce statistically different response latencies ($F_{1, 12} = 1.0$, $p = 0.35$). In addition, there is no evidence of a significant interaction effect between speaker gender and speech condition ($F_{3, 36} = 1.1$, $p = 0.36$). These reaction time results indicate that speakers required greater time to plan productions when requirements on clarity increase as compared to when requirements on speed increase.

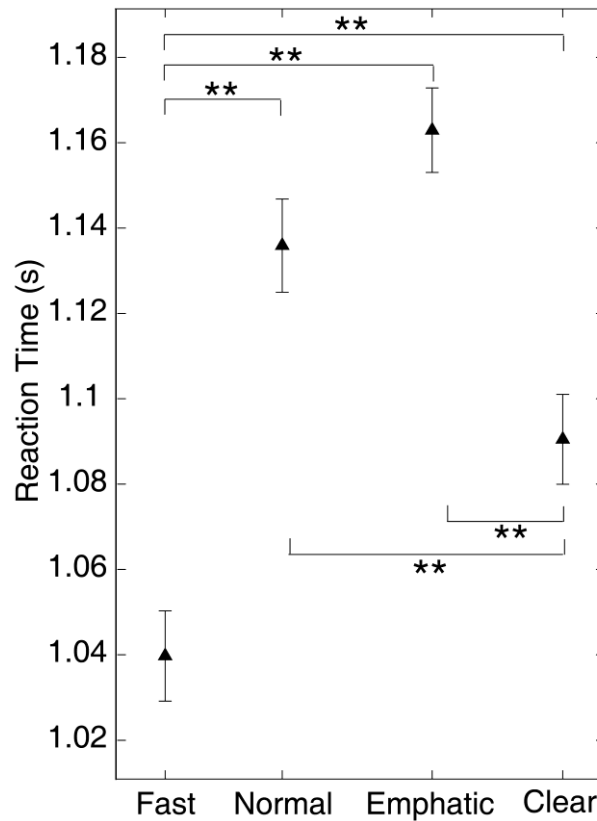


Figure 3-1: Average reaction time values for the group as a function of speech manner.

Study participants took significantly less time to plan sentences under fast speech conditions as compared to any other condition. Speakers required significantly more time before initiating speech under normal and emphatic conditions as compared to the clear condition. The slower reaction time for the emphatic and normal speech conditions likely reflects an ambiguity in the initial speaking manner cue that did not explicitly indicate whether the forthcoming sentence would require emphasis. These results suggest that knowledge of the stress pattern is necessary before sentence production initiation (cf., Keating and Shattuck-Hufnagel, 2002). ** indicates significance at $p_{FDR} < 0.05$.

3.2 Average Sentence-level Duration

Average sentence-level duration values (accounting for both the durations of syllables and the durations of pauses between syllables) are displayed for the group as a function of speech condition in Figure 3-2. The average sentence duration across subjects for clear speech is nearly twice as long as that for fast speech (clear: 2.1 s; fast: 1.1 s). On

average, subjects also took a longer time to produce sentences under emphatic conditions than under normal conditions (emphatic: 1.8 s; normal: 1.5 s). Table 3-1 indicates that the influence of speech condition on sentence-level duration is statistically significant ($F_{3, 36} = 107.6$, $p < 0.0001$). Comparisons of conditional effects indicate that sentence-level duration is significantly longer for the clear condition as compared to the normal (40% longer; $F_{1, 36} = 107.7$, $p_{FDR} < 0.05$), fast (91% longer; $F_{1, 36} = 298.3$, $p_{FDR} < 0.05$), and emphatic (17% longer; $F_{1, 36} = 30.5$, $p_{FDR} < 0.05$) conditions. These results are consistent with previous reports that speakers often reduce their speech rate significantly when producing clear speech with reductions in rate as much as one half that used during conversational speech (Picheny et al., 1986). Differences in rate between clear and conversational speech have been attributed to more frequent and/or longer pauses and the lengthening of segment durations (Bradlow et al., 2003; Picheny et al., 1986). Sentence-level duration is also significantly longer for the emphatic condition as compared to the normal ($F_{1, 36} = 23.6$, $p_{FDR} < 0.05$) and fast ($F_{1, 36} = 138.1$, $p_{FDR} < 0.05$) conditions. Speakers significantly reduced sentence-level durations when speaking quickly as compared to normal sentence production ($F_{1, 36} = 47.5$, $p_{FDR} < 0.05$), indicating that study participants followed instructions for the fast condition and reduced phrase-level production duration. The non-significant gender variable for sentence-level duration confirms that males and females produced sentences for a similar duration when averaged across speech conditions ($F_{1, 12} = 0.8$, $p = 0.40$). In addition, there is no statistical evidence for a significant interaction effect between speaker gender and speech condition ($F_{3, 36} = 0.5$, $p = 0.72$).

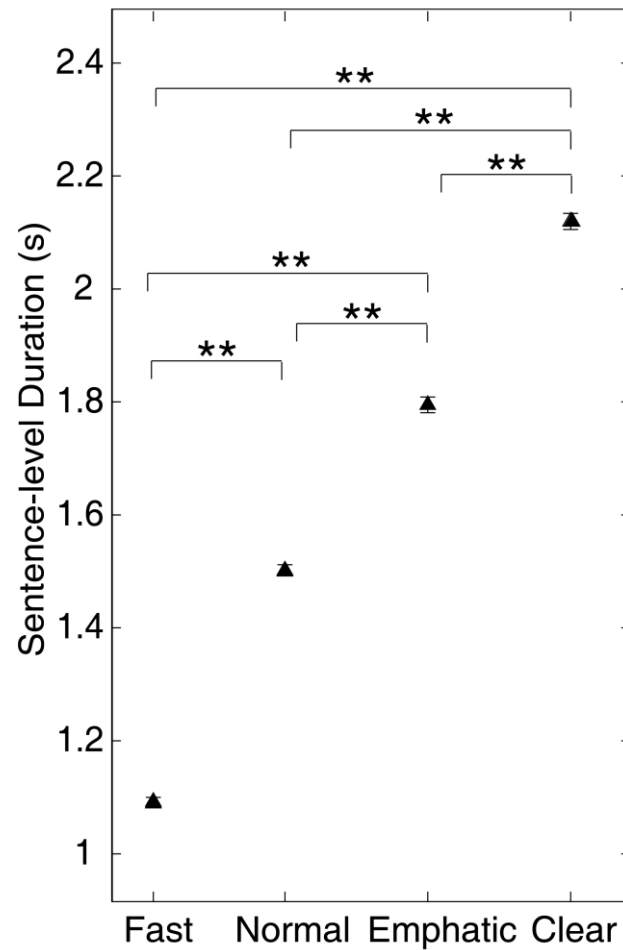


Figure 3-2: Average sentence-level duration values for the group as a function of speech manner.

The results demonstrate a monotonic increase in sentence-level duration as conditions change from fast to normal to emphatic to clear speech. Study participants took significantly less time to produce sentences in the fast condition as compared to any other condition. ** indicates significance at $p_{FDR} < 0.05$.

3.3 Average Vowel Dispersion

Group-level average distances of vowel tokens from their corresponding vowel-type means decrease by 2 mels from the fast condition to the emphatic condition, 2 mels from the emphatic condition to the normal condition, and 2 mels from the normal condition to

the clear condition. Figure 3-3 presents the average vowel dispersion values in mels for the group as a function of speech condition.

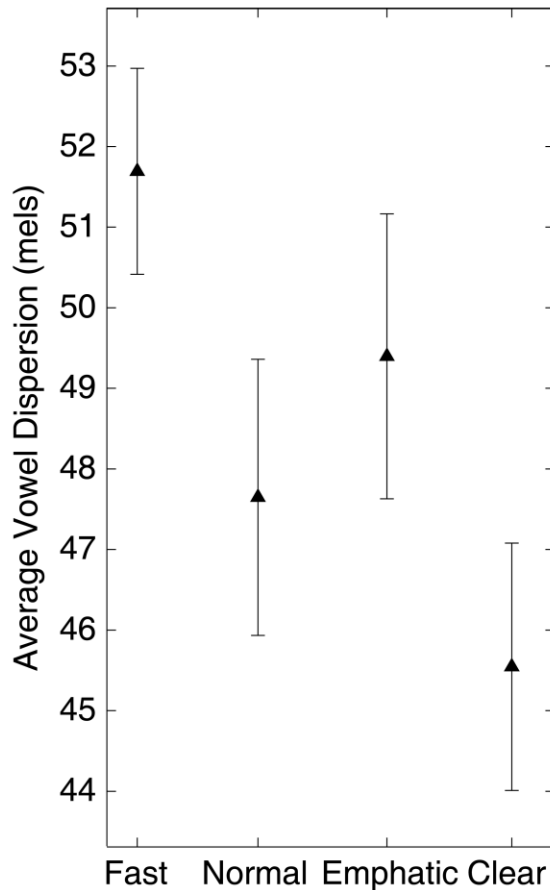


Figure 3-3: Average vowel dispersion for the group as a function of speech manner.

The changes in average vowel dispersion as a function of speech condition indicate that, as conditions change from clear to normal to fast, vowels tend to be more variable in acoustic space and therefore less precise acoustically. However, differences between conditions were not statistically significant upon correction for multiple comparisons.

Although these changes are consistent with a theory that the speech target is generally underspecified at the start of the movement and requires fine tuning towards the movement's end to produce more precise vowel segments under clear speech conditions (Guenther, 1995), this effect of speech condition on average vowel dispersion is not statistically significant ($F_{3, 36} = 2.3$, $p = 0.09$). Dispersion values pooled across

conditions do not differ significantly between the males and the females ($F_{1, 12} = 0.6$, $p = 0.44$) and there are no significant interactions between speaker gender and speech condition ($F_{3, 36} = 1.6$, $p = 0.21$). The results of acoustic studies that have examined the acoustic variability of vowel productions under requirements of increased clarity present mixed findings. While Chen (1980) demonstrated that tense vowels in CVC syllables have a tendency to cluster more tightly around the mean value of the vowel category in clearer-than-normal speech, more recently, this finding of reduced vowel dispersion for clear as compared to normal speech has not been reproduced (Krause and Braida, 2004; Ménard et al., 2007; Picheny et al., 1986). The results of the present study are more consistent with recent reports, indicating that when changing from fast to clear speech, speakers on average do not produce vowels with significantly reduced variability. One interpretation of these results is that highly overlearned motor skills like speech do not require the same sort of fine-tuning of targets when accuracy demands are increased (Ménard et al., 2007) and that, when instructed, speakers are as capable of maintaining vowel target productions under fast speech conditions as they are under normal speech conditions (cf., Gay, 1978).

3.4 Average Vowel Spacing

Figure 3-4 presents the average vowel spacing values in mels for the group as a function of speech condition. The figure demonstrates that across subjects inter-vowel acoustic distances reduce as the conditions change from clear to fast speech. The average inter-vowel acoustic distances decrease by 8 mels between the clear and emphatic conditions,

2 mels between the emphatic and normal conditions, and 7 mels between the normal and fast conditions. This main effect of speech condition on average vowel spacing is statistically significant ($F_{3, 36} = 37.2$, $p < 0.0001$).

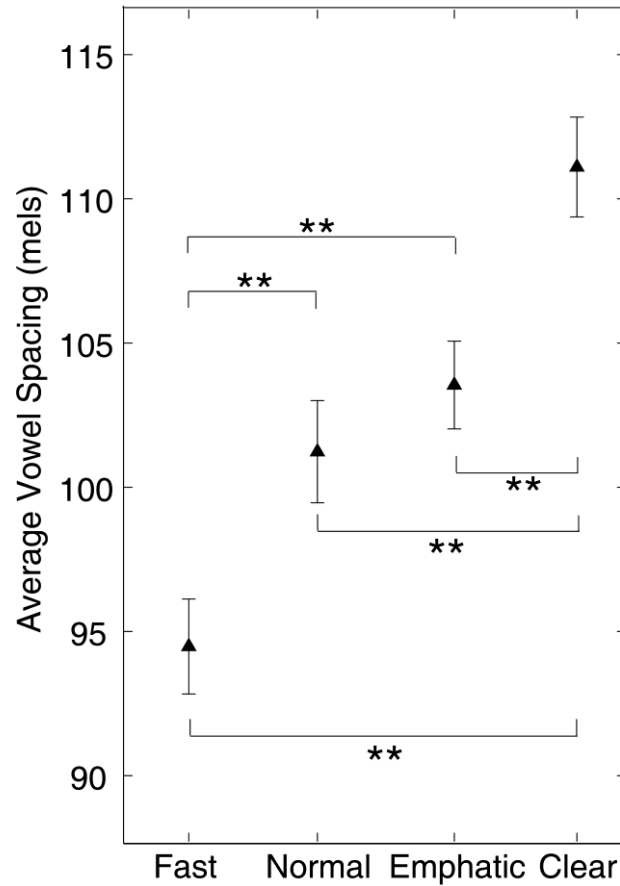


Figure 3-4: Average vowel spacing for the group as a function of speech manner.

The results demonstrate a monotonic increase in average vowel spacing as conditions change from fast to normal to emphatic to clear speech. Differences in average vowel spacing between the emphatic and normal conditions were not statistically significant. The statistically significant changes indicate that as conditions change from clear to normal to fast speech, vowels tend to be more spread a part in acoustic space. ** indicates significance at $p_{FDR} < 0.05$.

Pair-wise comparisons of conditions indicate that average vowel spacing values are significantly larger for the clear condition as compared to the normal ($F_{1, 36} = 38.6$, $p_{FDR} < 0.05$), fast ($F_{1, 36} = 109.5$, $p_{FDR} < 0.05$), or emphatic ($F_{1, 36} = 22.6$, $p_{FDR} < 0.05$)

conditions. Average vowel spacing is also significantly larger for normal as compared to fast speech ($F_{1,36} = 18.1$, $p_{FDR} < 0.05$) and for emphatic as compared to fast speech ($F_{1,36} = 32.6$, $p_{FDR} < 0.05$). Reduction in inter-vowel acoustic distances as speech manner changes from normal to fast speech has been reported in several previous studies (e.g., Fourakis, 1991; Lindblom, 1963; McRae et al., 2002; Ménard et al., 2007; Neel, 2008; Turner et al., 1995). It should be noted that reducing inter-vowel acoustic distances does not necessarily imply that vowels tend toward a more neutral position in acoustic space (cf., Fourakis, 1991). Fourakis (1991) demonstrated that phonetic vowel reduction, in which vowel formant frequencies move closer to a more central position in acoustic space, is minimal in American English and **not dependent** on tempo or stress, but rather on context (the vowel's frame). Fourakis (1991) did find that vowel space area was affected by stress and tempo, and not a product of contextual assimilation, suggesting that vowels tend to be less prototypical, but not more centralized under fast or non-stressed speech conditions.

Studies on disordered speech production often use acoustic vowel space as an index for articulatory range of movement (McRae et al., 2002; Turner et al., 1995). Reduced acoustic vowel space, accompanied by reduced articulatory working space area and slower production speeds has been reported for adult speakers with Down syndrome as compared to control speakers (Bunton and Leddy, 2011). Assuming that a relationship between acoustic vowel space and articulatory working space exists, we might conclude from the reduced inter-vowel acoustic distances associated with fast speech in the present study that a consequence of fast speech production is less extreme articulatory

movements for vowel targets than under normal conditions. However, previous research does not suggest such systematic changes in the kinematics of the articulators with changes in rate (e.g., Dromey and Ramig, 1998; Flege, 1988; Kuehn and Moll, 1976; Tuller et al., 1982). For example, Tuller and colleagues (1982) have demonstrated that under fast speech conditions muscle activity recorded from functionally relevant articulator muscles for the production of vowels changed in one of three ways: 1) a decrease in duration with no change in peak amplitude; 2) a decrease in duration with an increase in peak amplitude; or 3) no change in duration, but an increase in peak amplitude. These results provide evidence that a simple scaling rule is not sufficient to explain increases in speech rate.

The non-significant gender effect ($F_{1, 12} = 2.0$, $p = 0.19$) for average vowel spacing confirms that males and females produce similar acoustic distances between vowels when averaged across speech conditions. However, females and males do show differences in terms of the extent of change of average vowel spacing between certain speech conditions, as confirmed by the significant interaction effect ($F_{3, 36} = 3.1$, $p = 0.04$; see Figure 3-5). Pair-wise comparisons of the effects of the conditions indicate that the difference in average vowel spacing values between emphatic and normal speech conditions ($F_{1, 36} = 5.2$, $p_{FDR} < 0.05$) and between emphatic and fast speech conditions ($F_{1, 36} = 8.1$, $p_{FDR} < 0.05$) is significantly greater for males than females. In particular, while males demonstrate significantly greater average vowel spacing for emphatic than for normal speech ($F_{1, 36} = 7.0$, $p_{FDR} < 0.05$), there is no statistical evidence for a difference between these two conditions for females ($F_{1, 36} = 0.3$, $p = 0.56$). There is also

no statistical evidence for a difference in the average vowel spacing values between the emphatic and fast speech conditions for females ($F_{1,36} = 4.1$, $p = 0.05$), whereas the difference in average vowel spacing between the fast and emphatic conditions is significant for males ($F_{1,36} = 36.6$, $p_{\text{FDR}} < 0.05$). For all remaining comparisons of effects between the conditions (i.e., clear vs. normal, normal vs. fast, clear vs. fast, and clear vs. emphatic), the extent and direction of average vowel spacing change is similar between the genders.

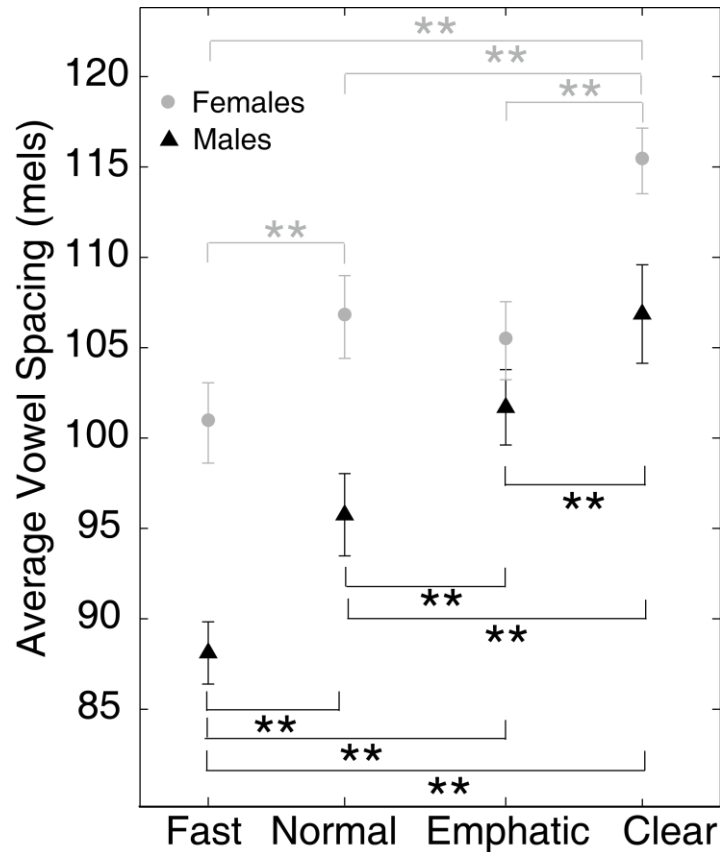


Figure 3-5: Average vowel spacing for the males and females as a function of speech manner.

Females (grey circles) do not demonstrate a significant difference in average vowel spacing between the emphatic and normal conditions. In contrast, the males (black triangles) demonstrate significantly greater average vowel spacing for emphatic than normal speech. ** indicates significance at $p_{\text{FDR}} < 0.05$.

3.5 Average Vowel F0

Group-level average vowel F0 values are displayed as a function of speech condition in Figure 3-6. The average vowel F0 for clear speech is half a semitone (st) higher than for normal speech (clear: 0.5 st; normal: 0.0 st). Vowels are also produced with a higher-than-normal F0 for fast and emphatic speech (fast: 0.5 st; emphatic: 0.4 st). Vowels are also produced with a higher-

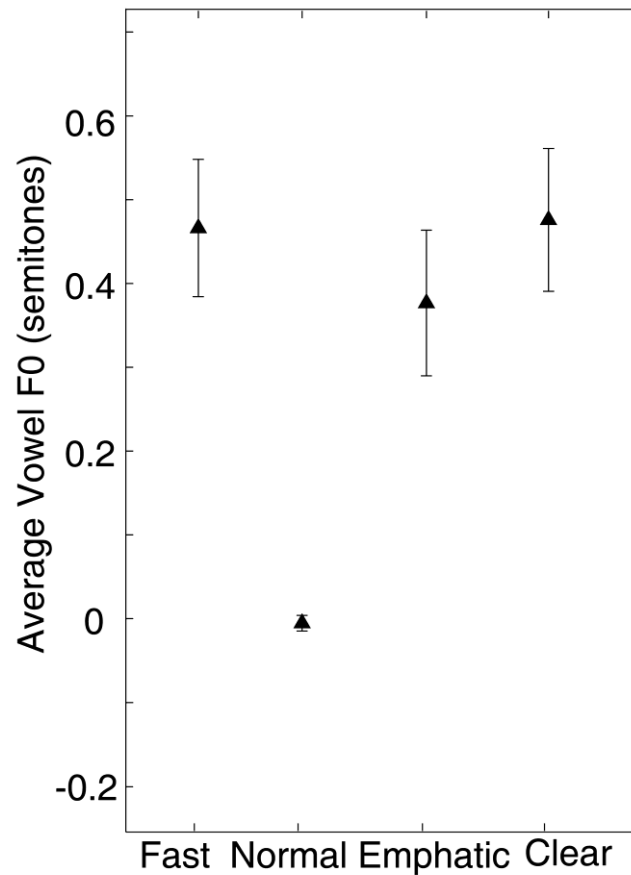


Figure 3-6: Average vowel F0 for the group as a function of speech manner.

Differences in average vowel F0 as a function of speech condition were small and not statistically significant in the repeated measures analysis of variance.

However, this effect of speech condition on average vowel F0 is not statistically significant ($F_{3, 36} = 2.7$, $p = 0.06$). Average vowel F0 values pooled across conditions do not differ significantly between the males and the females ($F_{1, 12} = 0.1$, $p = 0.79$) and there are no significant interactions between speaker gender and speech condition ($F_{3, 36} = 0.5$, $p = 0.66$). Although not significant, these results do reveal a tendency for average vowel F0 values to deviate more from normal under clear and fast speech conditions than under emphatic conditions. While it is possible that participants did not use F0 to convey emphasis it is also possible that the F0 changes associated with emphasis are local and should be evaluated with respect to the F0 values of neighboring nonemphasized syllables (cf., Cooper et al., 1983; Terken, 1991; Xu and Xu, 2005). An “adjacent-vowel F0 contrast” measure was developed to determine the relative F0 difference between vowels in syllables that received emphasis (those in the 1st and 4th positions of the sentence) as compared to those in syllables that did not (syllables in the 2nd and 5th positions of the sentence).

3.6 Average Adjacent-vowel F0 Contrast

Figure 3-7 displays the group average adjacent-vowel F0 contrast values as a function of speech condition. The figure demonstrates that across subjects the difference in F0 between adjacent vowels is approximately 2.5 st greater for emphatic than for fast, clear, and normal speech, while the values of this measure for fast, clear, and normal speech do not differ substantially from each other. Table 3-1 indicates that this effect of speech condition on the difference in F0 between adjacent vowels is statistically significant ($F_{3,$

$_{36} = 31.3$, $p < 0.0001$). Pair-wise comparisons of the effects of the conditions indicate that the average difference in F0 between adjacent vowels is significantly greater for emphatic than normal ($F_{1, 36} = 55.3$, $p_{FDR} < 0.05$), clear ($F_{1, 36} = 67.3$, $p_{FDR} < 0.05$), and fast ($F_{1, 36} = 64.5$, $p_{FDR} < 0.05$) speech. For all remaining comparisons, the difference in F0 between adjacent vowels is similar (normal vs. clear: $F_{1, 36} = 0.6$, $p_{FDR} > 0.05$; normal vs. fast: $F_{1, 36} = 0.4$, $p_{FDR} > 0.05$; clear vs. fast: $F_{1, 36} = 0.03$, $p_{FDR} > 0.05$). These results indicate that there is a relative increase in F0 for vowels in emphasized syllables compared to vowels of neighboring nonemphasized syllables under emphatic speech conditions. This relative increase in F0 for vowels is specific to the emphatic speech condition as vowels in the first and fourth positions of the sentence did not show the same significant relative increase in F0 compared to vowels in the second and fifth position of the sentence under normal, fast, or clear conditions.

Pooling across speech conditions, the significant gender variable reveals that the males and females differ in terms of the average adjacent-vowel F0 contrast ($F_{1, 36} = 6.6$, $p = 0.02$), with females showing greater adjacent-vowel F0 differences than males (see Figure 3-8; Females: 3.6 st; Males: 2.1 st). Pair-wise comparisons of the average adjacent-vowel F0 contrast measure between the males and females indicate significant differences in the values of this measure for the emphatic condition ($F_{1, 36} = 15.6$, $p_{FDR} < 0.05$). Under emphatic conditions, females demonstrated an average adjacent-vowel F0 contrast that was 3.3 st higher than that found for males (females_{emphatic}: 6.0 st; males_{emphatic}: 2.7 st). For the remaining conditions, there are no significant gender-based differences in this measure (normal: $F_{1, 36} = 0.9$, $p_{FDR} > 0.05$; clear: $F_{1, 36} = 4.5$, $p_{FDR} >$

0.05; fast: $F_{1, 36} = 2.1$, $p_{\text{FDR}} > 0.05$). While it is possible that gender-based differences in vocal tract size explain some of the between-group differences observed in this measure, it should be noted that a gender-based difference was not observed in the average vowel F0 measure, indicating that relative increases in F0 between vowels may be specific to female speakers.

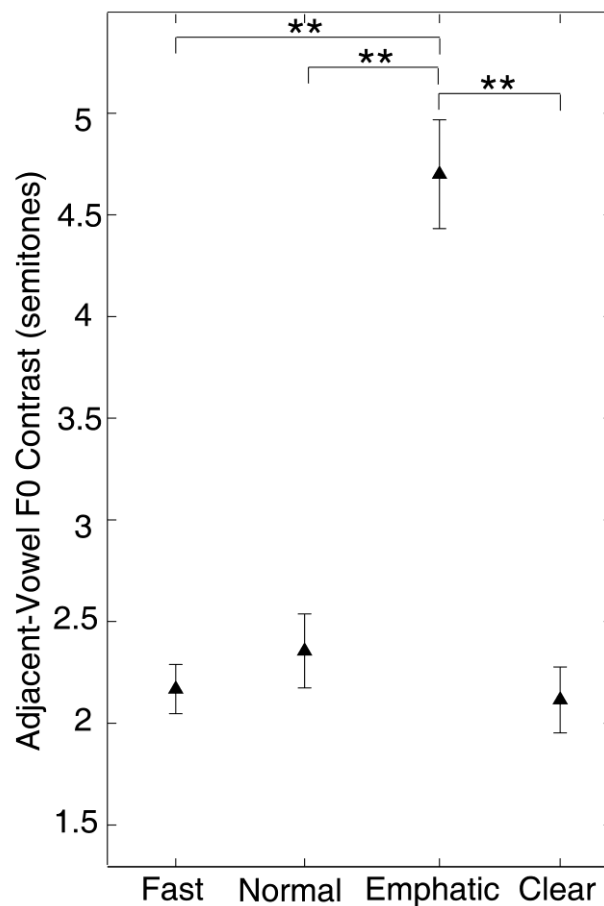


Figure 3-7: Average adjacent-vowel F0 contrast for the group as a function of speech manner.

Differences in F0 between adjacent (emphasized and nonemphasized) vowels were significantly larger for the emphatic condition than the fast, normal, and clear speech conditions. For all remaining comparisons of effects between the conditions (i.e., clear vs. fast, clear vs. normal, fast vs. normal), the difference in F0 between adjacent vowels is similar. ** indicates significance at $p_{\text{FDR}} < 0.05$.

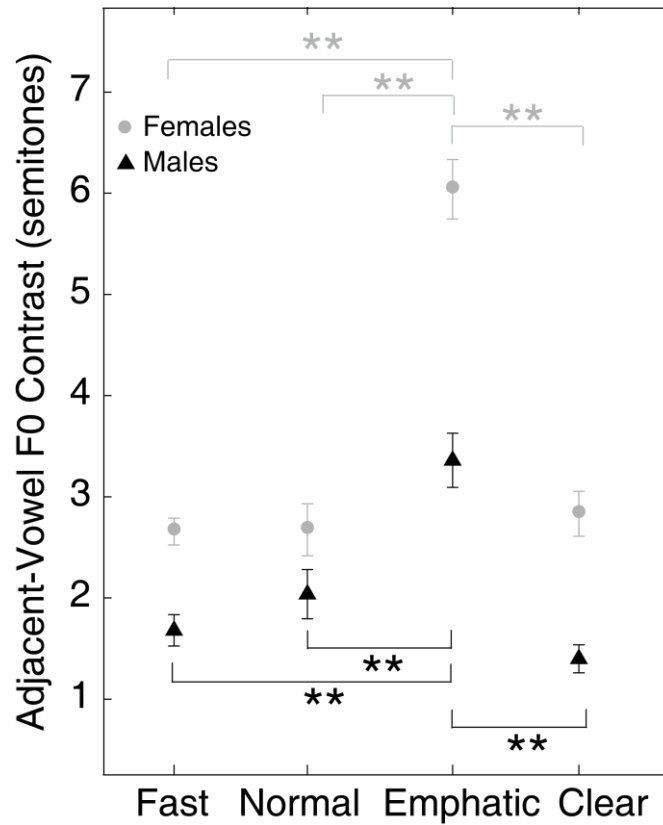


Figure 3-8: Average adjacent-vowel F0 contrast for the males and females as a function of speech manner.

While males (black triangles) demonstrate significantly greater-than-normal differences in F0 between adjacent vowels for the emphatic condition, the males do not show the same extent of difference as the females (grey circles). Females demonstrate much greater differences in average adjacent-vowel F0 contrast between the emphatic and normal conditions than the males. ** indicates significance at $p_{\text{FDR}} < 0.05$.

The significant interaction effect ($F_{3, 36} = 4.0$ $p = 0.01$) indicates that females and males do not vary F0 between adjacent vowels to the same extent between certain speech conditions (see Figure 3-8). While both males and females show significant differences in the average adjacent-vowel F0 contrast measure between the emphatic and normal conditions (males: $F_{1, 36} = 8.8$, $p_{\text{FDR}} < 0.05$; females: $F_{1, 36} = 57.0$, $p_{\text{FDR}} < 0.05$) and between the emphatic and fast conditions (males: $F_{1, 36} = 14.2$, $p_{\text{FDR}} < 0.05$; females: $F_{1, 36} = 57.6$, $p_{\text{FDR}} < 0.05$), the difference between the emphatic and normal conditions ($F_{1, 36}$

= 10.5, $p_{\text{FDR}} < 0.05$) and between the emphatic and fast conditions ($F_{1, 36} = 7.3$, $p_{\text{FDR}} < 0.05$) is significantly higher for females than males. Males and females do not demonstrate significant differences in this measure when comparing the normal and fast (males: $F_{1, 36} = 0.7$, $p_{\text{FDR}} > 0.05$; females: $F_{1, 36} = 0.0$, $p_{\text{FDR}} > 0.05$), normal and clear (males: $F_{1, 36} = 2.1$, $p_{\text{FDR}} > 0.05$; females: $F_{1, 36} = 0.1$, $p_{\text{FDR}} > 0.05$), and clear and fast (males: $F_{1, 36} = 0.5$, $p_{\text{FDR}} > 0.05$; females: $F_{1, 36} = 0.7$, $p_{\text{FDR}} > 0.05$) conditions.

The fact that increased average adjacent-vowel F0 contrast distinguishes emphasis from normal and fast speech for the group, suggests that this acoustic cue is a consequence of emphatic production in this study. The acoustic analysis demonstrated that females produced even larger changes in average adjacent-vowel F0 contrast than males. This gender-based difference may be linked to the earlier acoustic evidence demonstrating that females did not show significant differences in average vowel spacing between the emphatic and normal conditions and between the emphatic and fast conditions. Rather than change vowel quality and F0 height between adjacent vowels as the males had done, females may simply have relied on increased differences in F0 between adjacent vowels under emphatic speech conditions. Gender-based differences in the degree of F0 modulation has also been noted when adults address infants. Specifically, mothers expand their F0 range more when producing infant directed speech as compared to fathers (Fernald et al., 1989). Infant-directed speech is characterized by a slower speaking rate, a larger acoustic vowel space, and exaggerated pitch variation when compared with adult-directed speech (Fernald and Simon, 1984; Kuhl et al., 1997;

Liu et al., 2003). Relative F0 changes between adjacent vowels did not differentiate the remaining speech manners (i.e., clear, fast, or normal speech) from each other.

3.7 Average Vowel Intensity

Group-level average vowel intensity values accounting for differences in intensity between the vowel categories (see Chapter 2) are displayed as a function of speech condition in Figure 3-9. Differences in average vowel intensity were significantly larger-than-normal for the fast (1.3 db SPL), clear (1.5 db SPL), and emphatic (1.0 db SPL) speech conditions. The effect of speech condition on average vowel intensity is statistically significant ($F_{3, 36} = 11.5$, $p < 0.0001$). Pair-wise comparisons of the effects of the conditions indicate that average vowel intensity is significantly greater-than-normal for the clear ($F_{1, 36} = 28.9$, $p_{FDR} < 0.05$), fast ($F_{1, 36} = 22.6$, $p_{FDR} < 0.05$) and emphatic ($F_{1, 36} = 11.5$, $p_{FDR} < 0.05$) conditions. For all remaining comparisons of effects between the conditions, vowel intensity values do not differ significantly (clear vs. emphatic: $F_{1, 36} = 3.9$, $p_{FDR} > 0.05$; emphatic vs. fast: $F_{1, 36} = 1.8$, $p_{FDR} > 0.05$; clear vs. fast: $F_{1, 36} = 0.4$, $p_{FDR} > 0.05$).

Average vowel intensity values pooled across conditions do not differ significantly between the males and females ($F_{1, 12} = 0.5$, $p = 0.5$) and there are no significant interactions between speaker gender and speech condition ($F_{3, 36} = 1.2$, $p = 0.32$).

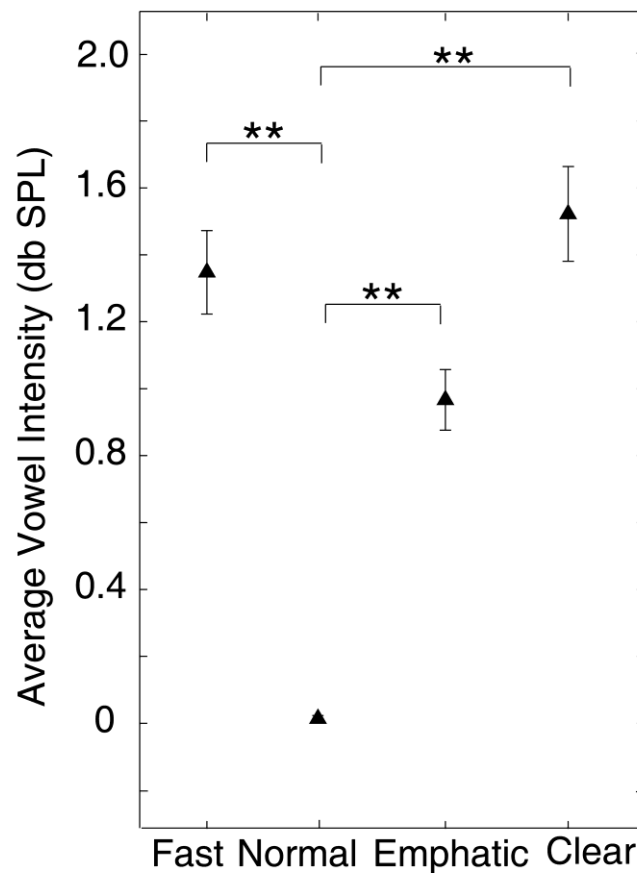


Figure 3-9: Average vowel intensity for the group as a function of speech manner. Differences in average vowel intensity were significantly larger-than-normal for the fast, clear, and emphatic conditions. ** indicates significance at $p_{FDR} < 0.05$.

Increases in vocal intensity involve changes in the typical control of respiratory, laryngeal, and articulatory systems. For example, as vocal intensity increases, there is a tendency for speakers to increase tracheal air pressure (the pressure that drives vocal folds apart) which can involve greater lung volume excursions and greater laryngeal resistance (Stathopoulos and Sapienza, 1993). While more recent evidence suggests a greater contribution from the respiratory as compared to the laryngeal system for increases in tracheal air pressure associated with increases in intensity (Finnegan et al.,

2000), systematic increases in stiffness at the level of the larynx have been observed with increases in rate (although this study did not examine whether intensity also increased; Ostry et al., 1983). Previous research also indicates a positive relationship between intensity and articulator speed and excursions (McClean and Tasko, 2002; Schulman, 1989). However, a relationship between increased speech rate and intensity remains tenuous. While McClean and Tasko (2003) report variable effects on intensity with increases in rate, Dromey and Ramig (1998) report an increase in mean SPL and decreases in lung volume excursion as rate increases. Gay (1978) implicated stress over rate as a major determinant of reduced vowel intensity, demonstrating that vowels in unstressed syllables have lower amplitudes as compared to vowels in syllables that are stressed and produced quickly. Although respiratory, laryngeal, and supralaryngeal kinematics were not recorded in the present study, the finding of greater-than-normal vowel intensity under fast, clear, and emphatic conditions is indicative of articulatory reorganization under all three speech conditions as compared to normal speech.

3.8 Average Adjacent-vowel Intensity Contrast

In order to determine the relative intensity difference between vowels in syllables that received emphasis (those in the 1st and 4th positions of the sentence) as compared to those in syllables that did not (syllables in the 2nd and 5th positions of the sentence), we developed an “adjacent-vowel intensity contrast” measure. The average adjacent-vowel intensity contrast displayed in Figure 3-10 as a function of speech condition reveals a similar pattern of results that was demonstrated for the adjacent-vowel F0 contrast

measure. In particular, the average adjacent-vowel contrast value is approximately 5 db SPL greater for the emphatic speech condition as compared to the fast, clear, and normal speech conditions, while the values of this measure for the fast, clear, and normal speech conditions do not differ substantially from each other. Table 3-1 confirms that the effect of speech condition on the average adjacent-vowel contrast measure is statistically significant ($F_{3, 36} = 31.5$, $p < 0.0001$). Direct comparisons between conditions indicate that the difference in intensity between adjacent vowels is significantly greater for the emphatic condition as compared to the normal ($F_{1, 36} = 60.9$, $p_{FDR} < 0.05$), clear ($F_{1, 36} = 66.3$, $p_{FDR} < 0.05$) and fast ($F_{1, 36} = 0.1$, $p_{FDR} < 0.05$) conditions. For all remaining comparisons of effects between the conditions, adjacent vowels show similar average adjacent-vowel intensity contrast values (normal vs. clear: $F_{1, 36} = 0.1$, $p_{FDR} > 0.05$; normal vs. fast: $F_{1, 36} = 0.0$, $p_{FDR} > 0.05$; clear vs. fast: $F_{1, 36} = 0.1$, $p_{FDR} > 0.05$). Pooling across conditions, average adjacent-vowel intensity contrast values are not significantly different between the males and females ($F_{1, 12} = 1.3$, $p = 0.28$) and males and females manipulated intensity between adjacent vowels to the same extent between conditions (no significant interaction effect between speech condition and speaker gender: $F_{3, 36} = 1.9$, $p = 0.15$). Like the influence of speech condition on the average adjacent-vowel F0 contrast values, the results of this analysis demonstrate that a common acoustic consequence of the expression of emphasis was greater changes in intensity between emphasized and nonemphasized syllables. Local intensity changes between adjacent vowels did not differentiate the remaining speech manners (i.e., clear, fast, or normal speech) from each other.

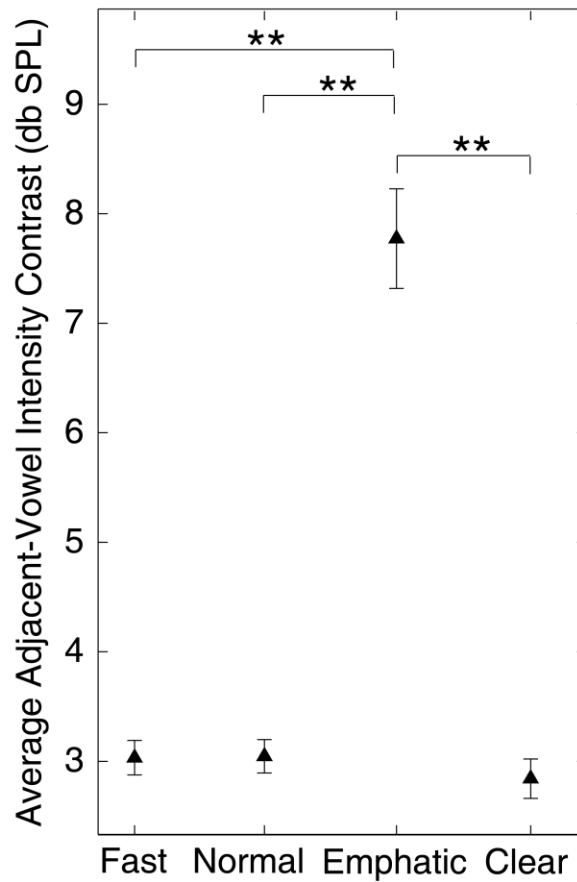


Figure 3-10: Average adjacent-vowel intensity contrast for the group as a function of speech manner.

Differences in intensity between adjacent (emphasized and nonemphasized) vowels were significantly larger for the emphatic condition than the fast, normal, and clear speech conditions. For all remaining comparisons of effects between the conditions (i.e., clear vs. fast, clear vs. normal, fast vs. normal), the difference in intensity between adjacent vowels is similar. ** indicates significance at $p_{\text{FDR}} < 0.05$.

3.9 Average Vowel Duration

Group-level average vowel duration values accounting for differences in duration between the vowel categories (see Chapter 2) are displayed as a function of speech condition in Figure 3-11. As can be seen in the figure, average vowel duration values increase by 4 ms from fast to normal speech, 6.5 ms from normal to emphatic speech,

and 7 ms from emphatic to clear speech. Table 3-1 indicates that this effect of speech condition on average vowel duration is statistically significant ($F_{3,36} = 36.1$, $p < 0.0001$). Direct comparisons between conditions indicate that the average vowel duration measure is significantly longer for clear speech as compared to normal ($F_{1,36} = 56.1$, $p_{FDR} < 0.05$), fast ($F_{1,36} = 93.7$, $p_{FDR} < 0.05$), and emphatic ($F_{1,36} = 13.2$, $p_{FDR} < 0.05$) speech. These results are consistent with previous reports that speakers often increase vowel durations when producing clear speech (e.g., Ménard et al., 2007; Moon and Lindblom, 1994; Perkell et al., 2002; Picheny et al., 1986). Speakers also significantly increased vowel durations for emphatic as compared to normal ($F_{1,36} = 13.2$, $p_{FDR} < 0.05$) and fast ($F_{1,36} = 33.9$, $p_{FDR} < 0.05$) speech conditions. These results are in agreement with earlier behavioral studies that indicate increases in vowel duration under stress speech conditions (e.g., Fourakis, 1991; Fry, 1955; Gay, 1978). Average vowel duration was not significantly shorter for the fast condition as compared to the normal condition under a strict FDR-corrected threshold ($F_{1,36} = 4.8$, $p_{FDR} > 0.05$). However, at a less conservative threshold (uncorrected for multiple comparisons) the effect of shorter-than-normal vowel durations under fast speech conditions was significant ($p = 0.04$). Many behavioral studies have reported that vowel duration decreases significantly as a consequence of increased speech rate (Fourakis, 1991; Gay, 1978; Lindblom, 1963; Ménard et al., 2007). Fourakis (1991) reported a dramatic shortening of vowel durations when conditions changed from slow-stressed to slow-unstressed speech or between slow-stressed and fast-unstressed, but less of a decrease in vowel duration between the fast-stressed and fast-unstressed conditions. This interaction between stress and tempo was attributed to

an “incompressibility effect” in which as vowel durations approach their minimal duration the effect of shortening factors, such as rate and removing stress from syllables, on vowel durations becomes less substantial (for a review see Klatt, 1976). This incompressibility effect may also explain our findings that the extent of vowel duration change for fast as compared to normal speech was not significant.

Average vowel duration values pooled across conditions do not differ significantly between the males and the females ($F_{1, 12} = 0.2$, $p = 0.7$) and there are no significant interactions between speaker gender and speech condition ($F_{3, 36} = 0.3$, $p = 0.84$). These results indicate that changes in vowel duration were affected to a greater extent by requirements for stress and clarity than for speed, consistent with the reports of previous studies (e.g., Fourakis, 1991; Gay, 1978).

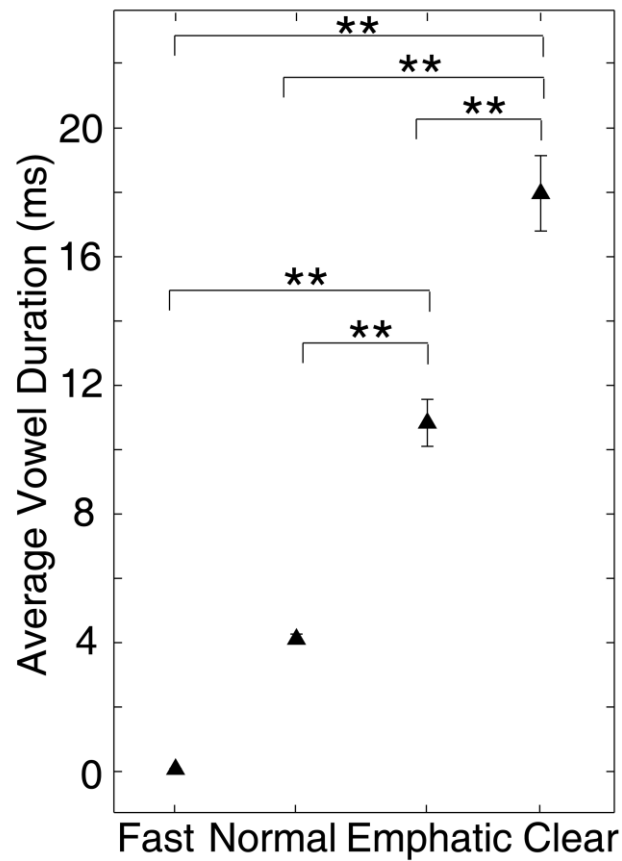


Figure 3-11: Average vowel duration for the group as a function of speech manner.

The results demonstrate a monotonic increase in average vowel duration as conditions change from fast to normal to emphatic to clear speech. Differences in vowel duration between the normal and fast conditions were not statistically significant at a strict threshold. These results may suggest that no additional timing effects are introduced under fast speech conditions after the mechanical effects of producing the vowels are accounted for. ** indicates significance at $p_{FDR} < 0.05$.

3.10 Average Adjacent-vowel Duration Contrast

In addition to examining the relative changes in F0 and intensity between adjacent emphasized and nonemphasized vowels, a quantitative measure was derived to examine differences in duration between adjacent emphasized and nonemphasized vowels and the results are presented as a function of speech condition in Figure 3-12. The figure

demonstrates that the mean difference in duration between adjacent vowels increases by approximately 2 ms from fast to normal speech, by approximately 3 ms from normal to clear speech, and by approximately 6 ms from clear to emphatic speech. Table 3-1 indicates that this effect of speech condition on the average adjacent-vowel duration contrast is statistically significant ($F_{3, 36} = 37.3$, $p < 0.0001$). Direct comparisons between conditions indicate that the average adjacent-vowel duration contrast is significantly greater for emphatic speech as compared to normal ($F_{1, 36} = 33.0$, $p_{FDR} < 0.05$), clear ($F_{1, 36} = 15.3$, $p_{FDR} < 0.05$), and fast ($F_{1, 36} = 47.7$, $p_{FDR} < 0.05$) speech and for clear speech as compared to fast speech ($F_{1, 36} = 9.0$, $p_{FDR} < 0.05$). There is no evidence for a significant difference in this measure between normal and clear speech ($F_{1, 36} = 3.4$, $p_{FDR} > 0.05$) and between normal and fast speech ($F_{1, 36} = 1.4$, $p_{FDR} > 0.05$). Pooling across conditions, the average adjacent-vowel duration contrast measure is not significantly different between the males and females ($F_{1, 12} = 1.3$, $p = 0.28$) and males and females varied duration between adjacent vowels to the same extent between conditions (no significant interaction effect between speech condition and speaker gender: $F_{3, 36} = 1.0$, $p = 0.42$). These analyses reveal significant effects of emphasis on adjacent-vowel durations and are consistent with those findings reported by Weismer and Ingrisano (1979) who observed that focused words have greater durations than adjacent unfocused words.

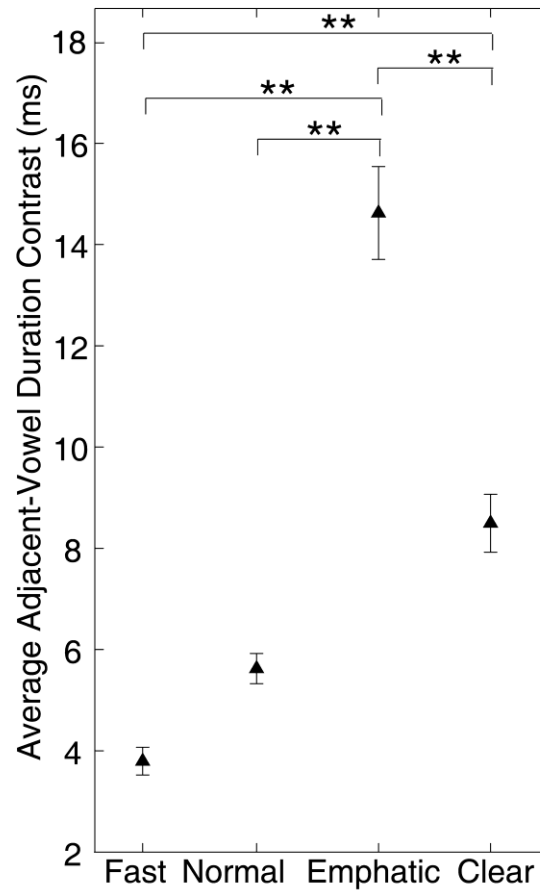


Figure 3-12: Average adjacent-vowel duration contrast for the group as a function of speech manner.

Differences in duration between adjacent (emphasized and nonemphasized) vowels were significantly larger for the emphatic condition than the fast, normal, and clear speech conditions. The clear speech condition also demonstrated significantly larger differences in duration between adjacent vowels than fast speech. For all remaining comparisons of effects between the conditions (i.e., clear vs. normal and fast vs. normal), the difference in duration between adjacent vowels is similar. ** indicates significance at $p_{FDR} < 0.05$.

3.11 Discussion

Many studies have demonstrated that the acoustic realizations of vowels differ significantly from normal as requirements for rate, clarity, and stress increase (e.g., Engstrand, 1988; Fougeron and Jun, 1998; Fourakis, 1991; Lindblom, 1963; Lindblom et

al., 2007; Ménard et al., 2007; Moon and Lindblom, 1994; Neel, 2008; Picheny et al., 1986). For example, as described in detail in Chapter 2, Lindblom (1963) demonstrated that under conditions in which vowel duration reduces and stress held constant there is a tendency for the midpoint values of vowel formant frequencies to exponentially fall short of their putative acoustic target values. In contrast, increased clarity is most often accompanied by increased acoustic distances between vowels as compared to normal speech (e.g., Ménard et al., 2007; Moon and Lindblom, 1994; Picheny et al., 1986). Therefore, consistent with previous acoustic studies the results of the present study demonstrates that both fast and clear speech conditions are accompanied by significant changes in vowel quality as compared to normal speech. Interestingly, significantly smaller-than-normal acoustic vowel spacing values observed under fast conditions in the present study were accompanied by relatively normal vowel durations. In other words, despite finding that, on average, vowel durations under normal and fast conditions were relatively similar, the inter-vowel acoustic distances was not similarly maintained under fast and normal conditions, even when reading errors had been removed from the analysis. These results are consistent with Lindblom's theory (Lindblom, 1963) that there is a nonlinear relationship between vowel duration and vowel quality such that beyond a certain point minimal reductions in vowel duration reduce the vowel's position in acoustic space significantly (also see Fourakis, 1991; Klatt, 1976). Given that, on average, vowel durations did not reduce significantly for fast speech and that vowels tend to reduce more than consonants (Gay, 1978), the results suggest that additional strategies may have been employed by speakers in order to produce sentences in

significantly shorter-than-normal durations under fast conditions.

One possible strategy to increase the rate of speech may be inferred from the electropalagraphy study of Hardcastle (1985). Hardcastle (1985) examined the degree of overlap of tongue tip movements for the consonant /l/ on tongue back movements for the consonant /k/ in sentence stimuli where the subject was required to produce the consonant /k/ immediately before the consonant /l/. If the /l/ is initiated before the release of the tongue back movement for the /k/, the release of the /k/ will be inaudible. Hardcastle (1985) demonstrated that the overlap of tongue tip movements on tongue back movements occurred to a similar extent under fast and slow/normal speech conditions except at clause/tone or sentence/tone boundaries (e.g., in sentences like, “*If it’s too weak, link it onto the other one.*”). Under these speaking contexts, three out of four subjects demonstrated coarticulation under fast speech conditions **only**. These findings suggest that in order to increase rate, speakers had assembled sentences as larger, multi-syllabic sequences as compared to normal/slow speech conditions. Further evidence in support of the theory that speakers encode speech as larger-than-normal multi-syllabic sequences comes from the findings of Fougeron and Jun (1998) that demonstrated a prosodic reorganization of sequential high and low tones under fast speech rates. In particular, Fougeron and Jun (1998) demonstrated that the boundary marking the end of a phrase and typically associated with a major change in tone and syllable length under normal conditions is reduced under fast rate conditions to a minor increase in tone and syllable duration. The effect of this prosodic reorganization is to make a sentence typically divided into multiple intonational phrases under normal

conditions reduce to a single intonational phrase under fast conditions. In the present study, pitch and intensity changes between adjacent vowels were similar between the fast and normal speech conditions. However, changes in duration between adjacent vowels for the fast speech condition were smaller than the normal condition and significantly reduced compared to the clear and emphatic speech conditions, suggesting that the prosodic structure may have changed under fast speech conditions, possibly as a consequence of using larger, multi-syllabic sequences to represent constituents within sentences.

This hypothesis would also be consistent with the model of Keating and Shattuck-Hufnagel (2002) that characterizes several stages of sentence-level encoding. Within their framework, higher-level prosodic representations are minimally represented by one intonational phrase and one prosodic word (see Figure 3-13 B). Keating and Shattuck-Hufnagel (2002) propose that speakers are sensitive to the length of intonational phrases, and length, among other factors, can determine whether intonational phrases will be divided into a number of smaller intermediate intonational phrases. Although it is not clear how length is determined (e.g., on the basis of number of words, syllables, phonetic duration, or some combination of these), speakers tend to divide intonational phrases into intermediate phrases of equal length (Keating and Shattuck-Hufnagel, 2002).

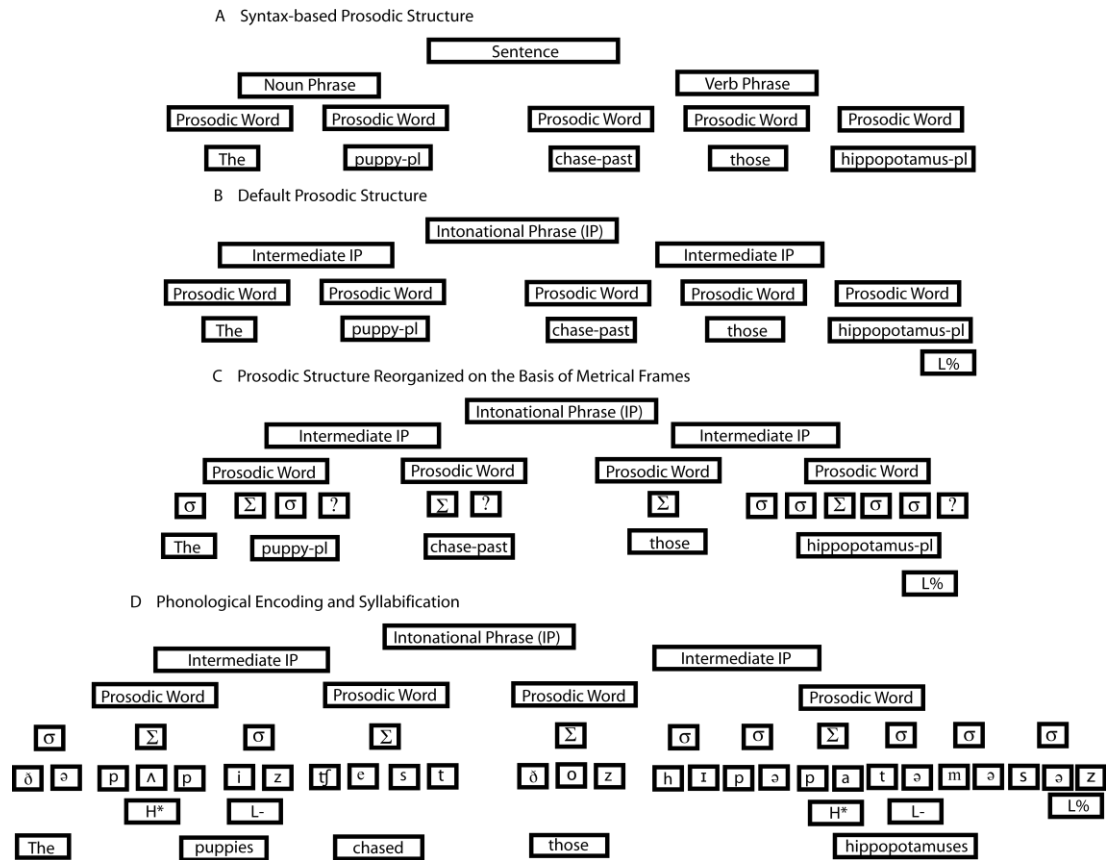


Figure 3-13: Stages of prosodic reorganization for a sentence as discussed in Keating and Shattuck-Hufnagel (2002).

(B) The default prosodic structure for the sentence, “The puppies chased those hippopotamuses” based on the syntactic organization demonstrated in (A). The number of prosodic words can decrease when information about the metrical structure becomes available (C). Resyllabification takes place at the final stage of phonological encoding, when sound segments are mapped to prosodic words (D).

Figure 3-13 demonstrates the encoding of prosodic structure for the sentence, “The puppies chased those hippopotamuses.” The default number of intermediate intonational phrases is based on syntax (Figure 3-13 A) so that the sentence is broken into a noun and verb phrase. However, since intonational phrases tend to divide into intermediate phrases of equal length, the intermediate intonational phrases are reorganized so that the word “chased” is shifted into the first intermediate intonational

phrase (Figure 3-13 C). In the model of Keating and Shattuck-Hufnagel (2002), the number of lexical words contained within one prosodic word can also change. For example, a typically stressless determiner such as ‘the,’ in the example above, will cliticize to its host ‘puppy.’ However, the stressless determiner, ‘the,’ would not cliticize (and therefore would constitute its own prosodic word), if a speaker emphatically stressed it (e.g., **The** puppies, not three.). Once the prosodic words are determined, sound segments can be serially ordered and the final syllable structure derived. It is posited that at this stage, empty slots within a syllable can be deleted and resyllabification can take place. Applying this model to the present study, prosodic restructuring under fast speech conditions could occur, minimally, when the size of the prosodic word is encoded. The use of larger, multi-syllabic prosodic words under fast speech conditions would permit less discretized and more continuous production of syllables. At the same time, larger prosodic words *may* require greater encoding to ensure that phonological segments are properly selected and temporally ordered.

It remains an open question how the size of prosodic constituents at fast speech rates affects encoding (for a discussion see Peramunage et al., 2011). Larger prosodic constituents could interfere with encoding as a result of having to select and temporally order a greater number of phonemes in a shorter period of time. If this is the case, one may expect significantly longer-than-normal reaction times under fast speech conditions as a consequence of the increased difficulty of selecting and temporally ordering phonemes among a growing number of competitors. Alternatively, a larger number of phonemes could facilitate the speed of production. In particular, Dell and Gordon (2003)

argue that in speech production the primary source of competition is from semantically related words. In the face of competing semantically-related candidates, a target word for production that has a high phonological neighborhood density (an index for the number of real words that sound the same as the target and determined by counting the number of real words that can be derived by changing one phoneme in the target word) will actually be produced faster than a word with a low phonological neighborhood density (Vitevitch, 2002). Dell and Gordon (2003) hypothesize that when a word is activated for production, phonological neighbors and the phonemes within the phonological neighbors also become activated. The phonemes shared by both the target and the phonological neighbors reinforce the target word through feedback connections and reduce reaction time latencies. Whether there is increased competition or increased facilitation, the use of larger prosodic constituents does allow for the possibility of resyllabification under fast speech rates.

Higher-level prosodic structure is also necessary for the realization of emphasis. In particular, in the present study speakers did not simply produce general increases in F0 across vowel segments, as we found no significant differences from normal in the average F0 across vowels under emphatic conditions. Under emphatic conditions, changes in F0 were *relative*. The acoustic realization of a vowel's F0 was dependent on the neighboring vowel's F0, and, presumably, on the overarching intonation contour of the sentence. This suggests that in order to shape relative differences in F0 between adjacent vowels speakers must minimally have had a representation of phrase-level boundary tone as well as a representation of the main stressed syllables that received a

pitch accent (see Figure 3-13 D). According to the framework of Keating and Shattuck-Hufnagel (2002), the encoding of pitch accents can be performed in parallel with phonological encoding since the location of pitch accents are based on metrical frame information about the location of the main stress in the lexical word. However, the interaction found between gender and emphatic speech production for average adjacent-vowel F0 contrast and for average vowel spacing, which was not found for either average adjacent-vowel intensity contrast or average adjacent-vowel duration contrast, suggests that, for this study, F0 may have been controlled independently from duration and intensity, but not from vowel quality. In particular, we found that men and women used average vowel spacing and average adjacent-vowel F0 contrast differently when distinguishing sentences produced under emphatic conditions from those produced under normal conditions. Females used large differences in F0 between emphatically stressed and non-stressed vowels and did not significantly increase average vowel spacing from normal under emphatic speech conditions. In contrast, males demonstrated significantly larger-than-normal acoustic vowel spacing values **and** adjacent-vowel F0 contrast values for emphatic speech conditions. These results lead to the inference that the acoustic realizations of vowel quality and of the local prosodic phrasing are not entirely independent from each other. More generally, when pooling across groups, the acoustic analysis demonstrated that the production of emphasis differed from normal sentence production primarily in terms of prosodic phrasing. In particular, under emphatic conditions sentence and vowel durations were significantly longer than normal and prosodic cues defined over adjacent vowels were significantly greater than normal.

The main results of the acoustic analyses are summarized in Table 3-2. The < and > signs indicate cases where one of the speech manner manipulations (fast, clear, emphatic) resulted in a value of the measure that was significantly less than or greater than the normal condition. The >> indicates cases of particularly large differences and the ≈ indicates no significant differences between the two conditions. Cells shaded in lavender indicate results that match the predictions described in Chapter 2. Cells shaded in red indicate those contrasts for which a significant increase or decrease was expected, but no significant difference was observed. As can be seen by the overwhelming number of lavender cells, the observed results largely confirm the expected changes.

Table 3-2: Summary of the group-level acoustic analysis results for the contrasts of each speech manner manipulation with normal sentence production.

The < and > signs indicate cases where one of the speech manner manipulations resulted in a value of the measure that was significantly less or greater than the normal condition. The >> indicates cases of particularly large differences and the ≈ indicates no significant differences. Cells shaded in (light) lavender indicate results that match predictions (See Table 2-5 in Chapter 2). Cells shaded in (dark) red indicate cases where the result was expected but not found.

	Vowel Space	Vowel Dispersion	Vowel Duration	Vowel F0	Vowel Intensity	Adjacent-vowel F0 Contrast	Adjacent-vowel Intensity Contrast	Adjacent-vowel Duration Contrast
Fast–Normal	<			≈	>	≈	≈	≈
Clear–Normal	>		>>	≈	>	≈	≈	≈
Emphatic–Normal	≈	≈	>	≈	>	>	>	>

CHAPTER 4

NEUROIMAGING RESULTS

4.1 Speech – Baseline

In order to determine the general effect of overt sentence production across subjects, we performed a random effects analysis and examined the Speech (all manners) – Baseline contrast using a positive-tailed one sample t-test. Figure 4-1 and Table 4-1 present the significant clusters that are specific to overt sentence production. In general, sentence production engaged both hemispheres similarly with extensive activity along the central sulcus. Of all the clusters in the left hemisphere the most significant vertex was located in middle motor cortex (MNI xyz = -48, -3, 38, cluster size = 6777 mm²) whereas of the clusters in the right hemisphere the most significant vertex was located in Heschl's gyrus (xyz = 54, -20, 3; cluster size = 1041 mm²). A cluster of significant activity particular to the left hemisphere was found in posterior inferior frontal sulcus (peak vertex xyz = -38, 24, 15; cluster size = 111 mm²). Aside from the large cluster along the central sulcus, peak vertices in three additional sentence production-specific clusters were found in left hemisphere: 1) dorsal motor cortex (xyz = -9, 12, 44, cluster size = 386 mm²); 2) dorsal somatosensory cortex (xyz = -23, -26, 65, cluster size = 58 mm²), and 3) pre-supplementary motor area (xyz = -8, 19, 43, cluster size = 386 mm²). Sentence production activity particular to the right hemisphere was observed in supplementary motor area (xyz = 7, 3, 56; cluster size = 598 mm²). The right hemisphere also demonstrated extensive activity in middle cingulate gyrus (xyz = 9, 10, 57; cluster size =

231 mm²) and planum temporale. Peak vertices in three additional speech-specific clusters were found in right hemisphere: 1) anterior insula (xyz = 35, 16, 3, cluster size = 173 mm²; and xyz = 40, 7, -15, cluster size = 210 mm²); and 2) dorsal motor cortex (peak xyz = 17, -18, 60; cluster size = 83 mm²).

The left and right hemisphere cortical maps of sentence production-specific changes in BOLD signal were used as a guide to derive anatomically-defined left and right hemisphere inclusive ROI masks that encompassed prefrontal, premotor, motor, auditory, and somatosensory regions. The anatomically defined cortical regions included in this mask were posterior inferior frontal sulcus (pIFs), dorsal inferior frontal gyrus pars opercularis (dIFo), ventral inferior frontal gyrus pars opercularis (vIFo), anterior and posterior frontal operculum (aFO/pFO), anterior and posterior insula (aINs/pINs), posterior dorsal, middle, and ventral premotor cortex (pdPMC/midPMC/vPMC), dorsal, middle, and ventral motor cortex (dMC/midMC/vMC), anterior and posterior central operculum (aCO/pCO), dorsal and ventral somatosensory cortex (dSC/vSC), anterior supramarginal gyrus (aSMg), parietal operculum (PO), planum polare (PP), anterior superior temporal gyrus (aSTg), Heschl's gyrus (Hg), planum temporale (PT), posterior superior temporal gyrus (pSTg), posterior dorsal superior temporal sulcus (pdSTs), pre-supplementary motor area (preSMA), supplementary motor area (SMA), and middle cingulate gyrus (midCg). These ROI masks are outlined in yellow on the Speech – Baseline contrast in Figure 4-1 and are used in the contrasts that follow to investigate task-related changes in BOLD signal in regions specific to overt sentence production.

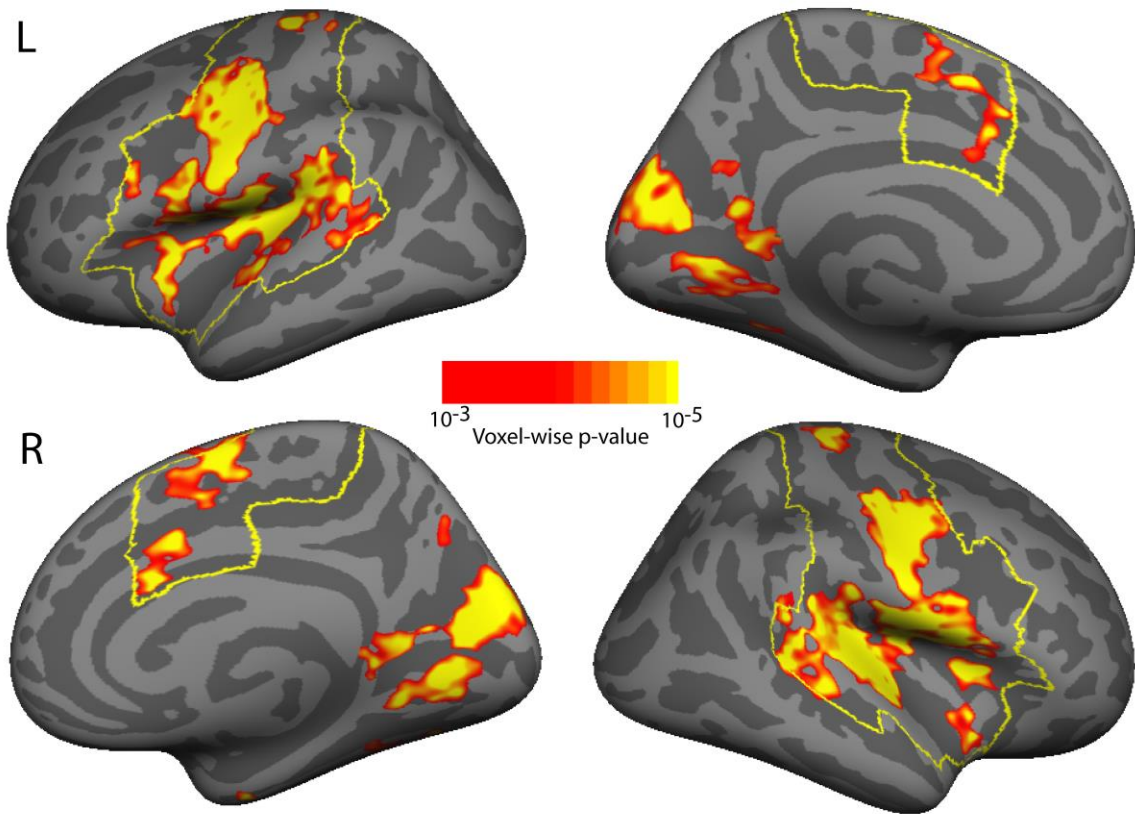


Figure 4-1: Group maps for the Speech – Baseline contrast displaying the significance of vertices within significant clusters.

Results are derived from a random effects analysis (14 subjects). A cluster is considered significant if all vertices inside the cluster meet threshold requirements ($p < 0.001$ uncorrected) and if the surface area of significant contiguous vertices is larger than the maximum cluster size that could occur by chance as determined by 10,000 Monte Carlo simulations. Clusters are overlaid on the left (L) and right (R) lateral and medial inflated surfaces of Freesurfer's fsaverage template. The left and right hemisphere cortical maps of sentence production-specific changes in BOLD signal were used as a guide to derive anatomically defined left and right hemisphere ROI inclusive masks that are overlaid in yellow on the Speech – Baseline contrast.

Table 4-1: Significant clusters associated with overt sentence production.

Stereotaxic coordinates are provided in both Montreal Neurological Institute (MNI) and Talairach spaces. The label, hemisphere (Hemi), coordinate, and p-value of the peak vertex within the significant cluster are provided as well as the p-value and surface area of the significant cluster. Abbreviations: aINs = anterior insula; aPHg = anterior parahippocampal gyrus; dMC = dorsal motor cortex; dSC = dorsal somatosensory cortex; Hg = Heschl's gyrus; Lg = lingual gyrus; midCg = middle cingulate gyrus; midMC = middle motor cortex; OC = occipital cortex; PCN = precuneus; pIFs = posterior inferior frontal sulcus; preSMA = pre-supplementary motor area; SMA = supplementary motor area; TOF = Temporal Occipital Fusiform gyrus. inf = inferior; med = medial. L indicates anatomical left and R indicates anatomical right.

Label	Hemi	Coordinate		Vertex p-value	Cluster-wise p-value	Size (mm ²)
		MNI	Talairach (x, y, z)			
Frontal Cortex						
pIFs	L	(-38, 24, 15)	(-37, 20, 19)	0.000003	0.000	111.0
preSMA	L	(-8, 19, 43)	(-9, 12, 44)	0.000000	0.000	385.8
midCg	L	(10, 22, 19)	(8, 17, 23)	0.000000	0.000	230.9
SMA	L	(9, 10, 57)	(7, 3, 56)	0.000000	0.000	597.5
Rolandic Cortex						
midMC	L	(-48, -3, 38)	(-46, -8, 37)	0.000000	0.000	6777.0
dMC	L	(-18, -21, 55)	(-19, -26, 52)	0.000000	0.000	251.1
dSC	L	(-23, -26, 65)	(-23, -31, 60)	0.000069	0.015	58.2
dMC	R	(17, -18, 60)	(14, -24, 56)	0.000000	0.000	327.7
Insular Cortex						
aINs	R	(35, 16, 3)	(31, 13, 9)	0.000000	0.000	172.7
aINs	R	(40, 7, -15)	(37, 6, -8)	0.000000	0.000	210.4
Temporal Cortex						
TOF	L	(-31, -54, -8)	(-29, -51, -9)	0.000036	0.002	76.2
Hg	R	(54, -20, 3)	(49, -21, 6)	0.000000	0.000	5993.5
TOF	R	(32, -69, -7)	(29, -66, -8)	0.000000	0.017	56.0
TOF	R	(34, -53, -13)	(31, -50, -12)	0.000230	0.000	135.7
aPHg	R	(35, -10, -35)	(32, -8, -28)	0.000005	0.002	79.2
Parietal Cortex						
PCN	L	(-19, -52, 8)	(-19, -51, 6)	0.000002	0.000	302.6
PCN	L	(-8, -63, 30)	(-9, -63, 26)	0.000230	0.006	68.0
PCN	R	(10, -65, 46)	(7, -67, 40)	0.000192	0.001	81.4
Occipital Cortex						
Cuneus	L	(-4, -69, 21)	(-5, -67, 16)	0.000000	0.000	1276.7
inf OC	L	(-24, -85, -1)	(-23, -81, -5)	0.000395	0.025	52.3
Lg	L	(-10, -65, 8)	(-10, -63, 5)	0.000000	0.000	731.8
Cuneus	R	(7, -76, 28)	(5, -75, 23)	0.000000	0.000	1745.9
med Lg	R	(11, -63, 10)	(9, -62, 8)	0.000000	0.000	914.6

4.2 Clear – Normal

In the present fMRI experiment, we kept the total number of syllables constant across conditions as opposed to fixing the total movement time as previous neuroimaging studies on speech rate have done (e.g., Paus et al., 1996; Riecker et al., 2006; Sidtis et al., 2003). The rationale for matching the number of syllables across conditions was that differences in the number of presentations influences the BOLD response in sensory cortical regions like the left posterior superior temporal gyrus (Dhankhar et al., 1997). As a result, the total movement time was not a controlled parameter across conditions in the present study. Specifically, we found that the mean difference in sentence duration between the fastest and slowest conditions was 1.03 s and that subjects were engaged for a significantly longer period of time under clear conditions as compared to fast conditions. Since the hemodynamic response likely resolves temporal differences greater than one second (Howseman et al., 1997) and longer production durations may place greater demands on attention, working memory and/or inhibitory control for the controlled release of speech movements, we performed a second random effects fMRI analysis in which we include the sentence-level duration values as a covariate of non-interest in the first-level general linear model. By including this covariate of non-interest we attempt to minimize the influence of sentence-level production durations on the contrasts (Sörös et al., 2010). At the same time, this design permits the mapping of those regions that contribute to the planning and encoding of sentences produced under clear, fast, normal, and emphatic conditions. All contrasts described in the sections that follow are based on this second neuroimaging analysis that includes sentence-level duration as a

covariate of non-interest in the first level general linear model. In order to examine both the positive and negative group-level responses associated with direct comparisons between conditions, random effects analyses were performed using a two-tailed one sample t-test.

If a region contributes to the processing associated with producing acoustically salient linguistic contrasts, we would expect activity in this region to be significantly modulated by clear speech in the Clear – Normal contrast, since the acoustic analysis demonstrated that the production of clear speech differed from normal largely in terms of vowel quality. We predicted that areas that showed a clear speech manner effect would primarily be those that have been implicated in phonetic encoding, such as left lateral premotor cortex, since it has been shown that under clear speech conditions speakers use more extreme articulatory gestures to produce speech (e.g., Matthies et al., 2001; Perkell et al., 2002; Wohlert and Hammen, 2000). The findings are largely consistent with this prediction (see Table 4-2 and Figure 4-2 for the coordinates and map of the significant clusters) with the most significant vertex of all clusters in the left hemisphere located in middle premotor cortex ($xyz = -51, 0, 47$; cluster size = 62 mm²). A significant main effect of clear speech production was also observed in left ventral motor and somatosensory cortices and right ventral premotor and middle motor cortices. In contrast to clear speech, a cluster of activity specific to normal sentence production was observed in right planum polare. Activation in the right planum polare has been observed for processing tones (Binder et al., 2000; Koelsch et al., 2002), generating melodies and non-melodic utterances (Brown et al., 2006; Callan et al., 2006), and processing syntax

(Friederici et al., 2000; Meyer et al., 2002), suggesting that this site may facilitate the processing of syntax and/or prosody.

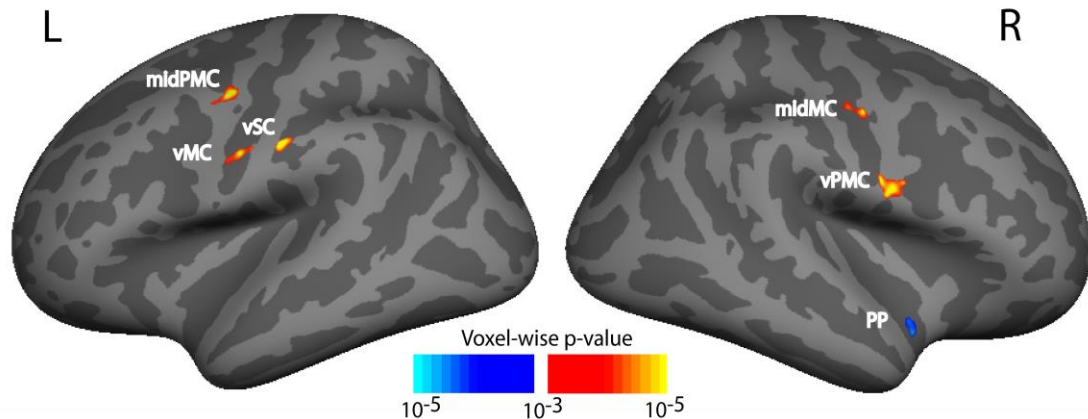


Figure 4-2: Group maps for the Clear – Normal contrast after controlling for sentence duration.

Relative increases in BOLD specific to clear speech production were more common in bilateral motor and premotor regions with a left hemisphere bias in middle premotor cortex and a right hemisphere bias in ventral premotor cortex. Normal speech production was associated with increased activity in right planum polare. See Table 4-2 for abbreviations of cluster labels.

Table 4-2: Significant clusters for the Clear – Normal contrast after controlling for sentence duration.

Abbreviations: midMC = middle motor cortex; midPMC = middle premotor cortex; PP = planum polare; vMC = ventral motor cortex; vPMC = ventral premotor cortex; vSC = ventral somatosensory cortex. L indicates anatomical left and R indicates anatomical right.

Label	Hemi	Coordinate		Vertex p-value	Cluster p-value	Size (mm ²)
		MNI (x, y, z)	Talairach (x, y, z)			
Rolandic Cortex						
midPMC	L	(-51,0,47)	(-48,-5,45)	0.000001	0.001	62.1
vMC	L	(-52,-6,21)	(-49,-9,22)	0.000006	0.006	40.3
vSC	L	(-59,-12,32)	(-56,-15,31)	0.000000	0.004	43.8
vPMC	R	(65,9,12)	(59,5,16)	0.000002	0.000	112.8
midMC	R	(47,-4,34)	(42,-9,35)	0.000007	0.003	45.3
Temporal Cortex						
PP	R	(49,-3,-24)	(45,-2,-17)	-0.000044	0.020	33.3

4.3 Fast – Normal

The Fast – Normal sentence production contrast was examined to identify regions associated with producing utterances under conditions in which speed is a priority. Less time to produce each speech sound, as compared to normal speech, places greater demands on the sequencing and timing of speech movements. We predicted that areas that showed a fast speech manner effect would be those that have been implicated in phonological encoding and, in particular, syllabification and sequencing. Consistent with this prediction, the comparison of fast with normal sentence production, controlling for differences in sentence duration between conditions, showed fast-speech-specific activity in left hemisphere prefrontal regions (see Table 4-3 and Figure 4-3 for a map and the coordinates of the significant clusters). In the left hemisphere, a small cluster was observed in dorsal inferior frontal gyrus pars opercularis and a larger, second cluster more inferior and anterior in frontal operculum. Additional clusters of activity specific to fast speech were observed in left dorsal motor cortex, left peri-Rolandic operculum, and right posterior dorsal superior temporal sulcus. A cluster of activity specific to normal sentence production was observed in right posterior inferior frontal sulcus. Several recent neuroimaging studies have shown that right posterior inferior frontal gyrus is involved in processing stress in sentences (e.g., Geiser et al., 2008; Klein et al., 2011; Perrone et al., 2010) or prosodic contrasts (e.g., Meyer et al., 2002; for a review see Friederici, 2011). It is possible that increased activation in this region is associated with the higher processing demands of encoding the hierarchical pattern of prosodic contrasts

under normal speech conditions, which may have been reduced to a single, larger intonational phrase under fast speech conditions (cf., Fougeron and Jun, 1998).

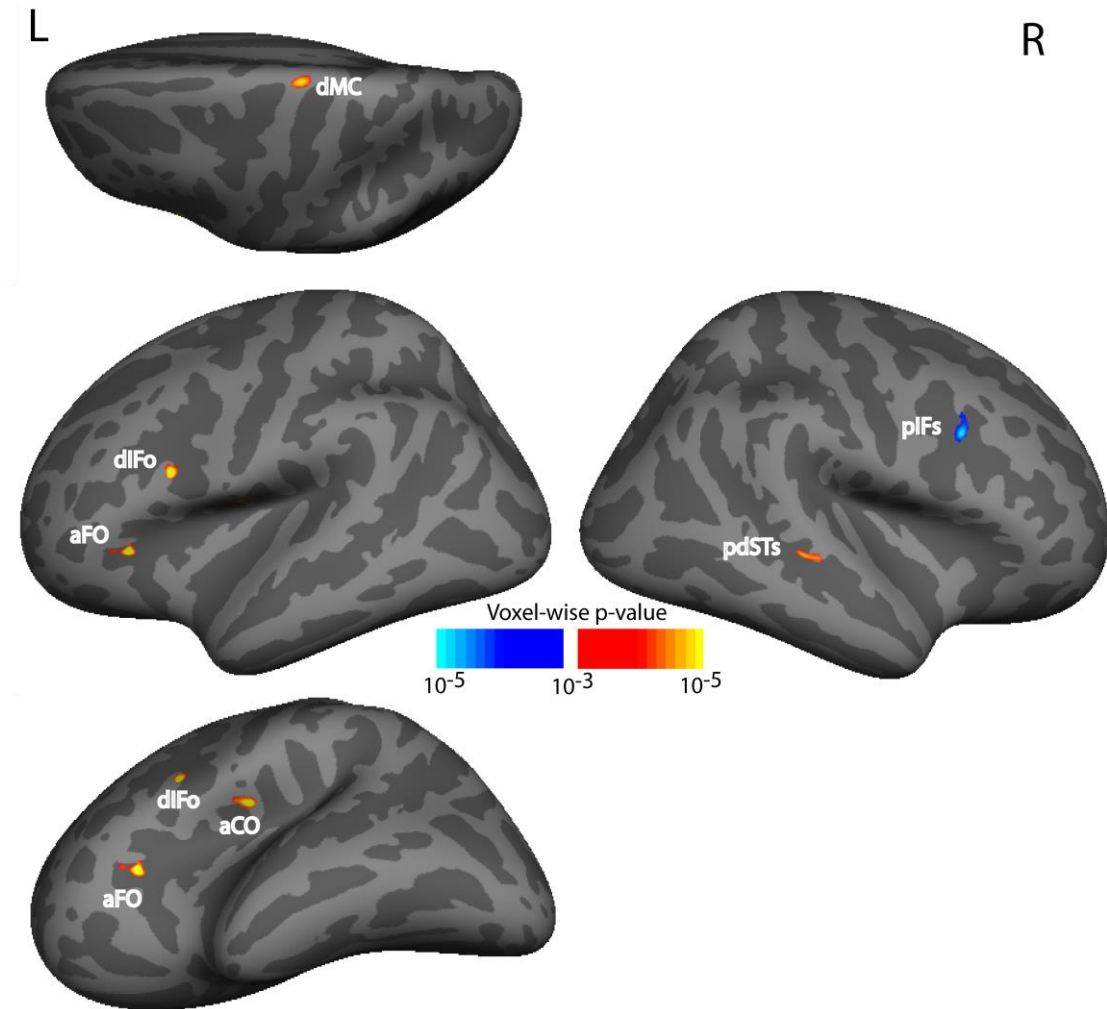


Figure 4-3: Group maps for the Fast – Normal contrast after controlling for sentence duration.

Many of the regions found to be specific to fast sentence production overlap with regions that have previously been implicated in sequencing and phrase-level structural processing. See Table 4-3 for abbreviations of vertex labels.

Table 4-3: Significant clusters for the Fast – Normal contrast after controlling for sentence duration.

Abbreviations: aCO = anterior central operculum; aFO = anterior frontal operculum; dIFo = dorsal inferior frontal gyrus pars opercularis; dMC = dorsal motor cortex; pdSTs = posterior dorsal superior temporal sulcus; pIFs = posterior inferior frontal sulcus. L indicates anatomical left and R indicates anatomical right. ** This cluster was only marginally significant using a strict cluster-wise threshold of $p < 0.025$ for each hemisphere.

Label	Hemi	Coordinate		Vertex p-value	Cluster p-value	Size (mm ²)
		MNI	Talairach			
		(x, y, z)				
Frontal Cortex						
aFO	L	(-37,34,-7)	(-35,31,0)	0.000000	0.013	35.2
dIFo	L	(-45,15,16)	(-43,11,19)	0.000000	0.018	33.7
pIFs	R	(36,17,23)	(32,12,26)	-0.000011	0.004	45.4
Rolandic Cortex						
midPMC	L	(-52, 0, 46)	(-50,-6,45)	-0.000004	0.031**	30.1
aCO	L	(-49,-4,8)	(-46,-6,10)	0.000001	0.003	45.3
dMC	L	(-9,-14,74)	(-11,-21,69)	0.000014	0.002	46.1
Temporal Cortex						
pdSTs	R	(52,-28,-1)	(47,-28,1)	0.000031	0.011	36.2

Since the acoustic analysis demonstrated that normal sentence production was produced with greater average vowel spacing than fast speech, we expected to find increased activation in left middle premotor cortex, like clear speech in the Clear – Normal contrast, for normal sentence production in the Fast – Normal contrast. Using a less conservative cluster-wise threshold for the left hemisphere ($p < 0.05$), an additional cluster of activity specific to normal sentence production was found in left middle premotor cortex (xyz = -52, 0, 46, cluster size = 30 mm²; cluster p-value = 0.03; see Figure 4-4). The location of the peak vertex within this cluster was within 2 mm of the location of the significant vertex in left middle premotor cortex in association with clear speech in the Clear – Normal contrast. A less conservative cluster-wise threshold for the right hemisphere ($p < 0.05$) did not reveal any additional clusters of activity.

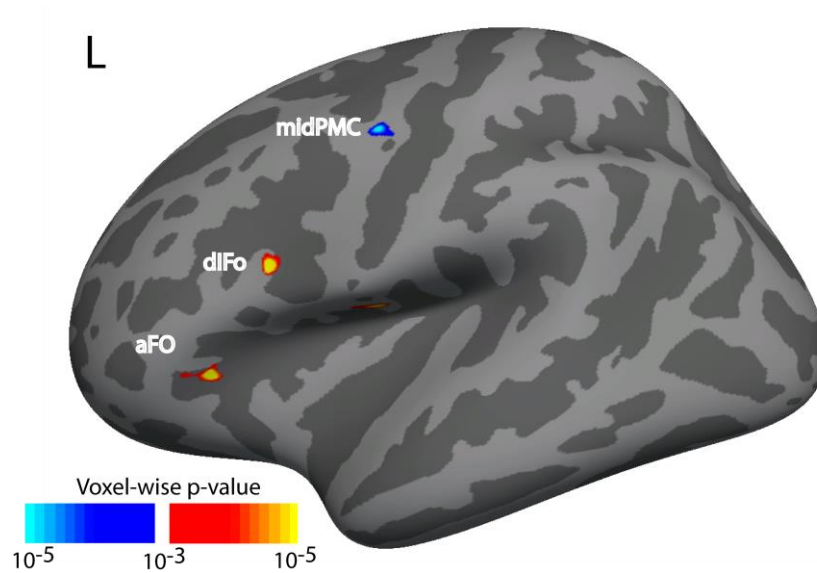


Figure 4-4: Left hemisphere group map for the Fast – Normal contrast when a less strict cluster-wise threshold of $p < 0.05$ for the hemisphere is used.

4.4 Clear – Fast

If a region contributes to the processing associated with producing acoustically salient linguistic contrasts, we would expect it to be engaged for clear speech in both the Clear – Normal and Clear – Fast contrasts, since the acoustic analysis demonstrated significantly larger average vowel spacing for clear speech as compared to normal and fast speech. Consistent with this line of reasoning, the locations of significant clusters associated with clear speech for the Clear – Fast contrast largely overlap those observed in association with clear speech for the Clear – Normal contrast. In particular, three clusters that were specific to clear sentence production were found in left middle premotor cortex, right ventral premotor cortex, and right middle motor cortex (see Figure 4-5 and Table 4-4 for a map and the coordinates of the significant clusters). The locations of peak vertices within two of these clusters (left middle premotor cortex and right ventral premotor

cortex) were within 5 mm of the locations of peak vertices specific to clear speech production in the Clear – Normal contrast. A cluster of activity specific to fast sentence production was observed in left dorsal somatosensory cortex.

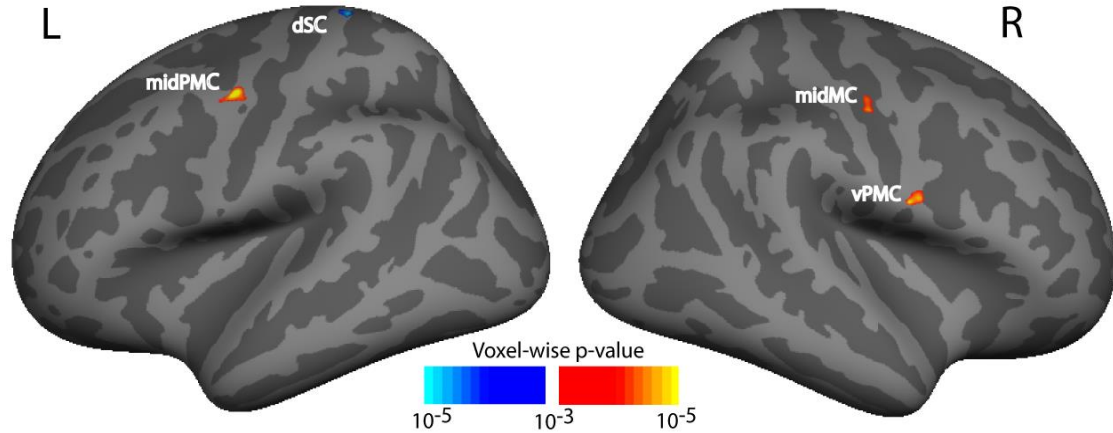


Figure 4-5: Group maps for the Clear – Fast contrast after controlling for differences in sentence duration between the conditions.

The locations of the peak vertices in left middle premotor cortex and right ventral premotor cortex specific to clear speech in this contrast are within 5 mm of those specific to clear speech production in the Clear – Normal contrast. See Table 4-4 for abbreviations of vertex labels.

Table 4-4: Significant clusters for the Clear – Fast contrast after controlling for sentence duration.

Abbreviations: dSC = dorsal somatosensory cortex; midMC =middle motor cortex; midPMC = middle premotor cortex; vPMC = ventral premotor cortex. L indicates anatomical left and R indicates anatomical right.

Label	Hemi	Coordinate		Vertex p-value	Cluster p-value	Size (mm ²)
		MNI (x, y, z)	Talairach			
Rolandic Cortex						
midPMC	L	(-51,1,46)	(-49,-5,45)	0.000001	0.001	54.4
dSC	L	(-20,-23,73)	(-21,-30,67)	-0.000009	0.021	32.3
vPMC	R	(65,7,9)	(59,3,14)	0.000011	0.003	45.1
midMC	R	(43,-11,31)	(38,-15,32)	0.000053	0.016	34.1

4.5 Emphatic – Normal

If a region contributes to the processing associated with producing acoustically salient prosodic contrasts, we would expect activity in this region to be significantly modulated by emphatic speech in the Emphatic – Normal contrast, since the acoustic analysis demonstrated that the production of emphasis differed from normal sentence production largely in terms of prosodic phrasing. The comparison of emphatic with normal sentence production revealed significant main effects of emphatic speech production primarily in right hemisphere cortical regions (see Figure 4-6 and Table 4-5 for a map and the coordinates of the significant clusters). Clusters specific to emphatic speech were observed in right hemisphere SMA, middle premotor and motor cortex, and ventral somatosensory cortex. In contrast, normal-speech-specific activity was almost exclusively found in the left hemisphere, with increased responses in left anterior insula, left supramarginal gyrus, and right dorsal somatosensory cortex.

In neuroimaging studies of speech, prosodic processing has previously been shown to preferentially engage the right hemisphere (Meyer et al., 2002; for a review see Friederici, 2011). As noted in the introduction, the extent to which this laterality effect extends to the production of stress is not yet known (although see Loevenbruck et al., 2005). The results from the present study seem to indicate that the right hemispheric asymmetry does extend to the production of prosody and, in particular, to the production of emphasis. Studies on the speed-accuracy tradeoff provide an alternative explanation for the asymmetric activation found in association with emphatic speech production. In particular, Seidler and colleagues (2004) demonstrated that activity in left motor cortex

and left peri-Rolandic operculum were associated with greater endpoint variability (as observed under fast responses) whereas activation in right ventral premotor cortex, right sensorimotor cortex, and right peri-Rolandic operculum were associated with greater endpoint accuracy. In the study of Seidler et al. (2004), subjects manipulated a joystick using discrete right hand and finger movements to hit targets of different sizes visually presented to them. Movements of the right hand, typically, engage the contralateral hemisphere. Yet, Seidler et al. (2004) demonstrated greater contributions from the ipsilateral hemisphere for more accurate productions. It was suggested that accurate movements may demand greater descending drive to successfully achieve targets and consistent with this interpretation increased recruitment of regions in the ipsilateral hemisphere have been demonstrated for tasks requiring greater force (e.g., Dai et al., 2001; Dettmers et al., 1995; van Duinen et al., 2007).

Table 4-5: Significant clusters for the Emphatic – Normal contrast after controlling for sentence duration.

Abbreviations: aINs = anterior insula; aSMg = anterior supramarginal gyrus; dSC = dorsal somatosensory cortex; midMC = middle motor cortex; midPMC = middle premotor cortex; pCO = posterior central operculum; SMA = supplementary motor area; vSC = ventral somatosensory cortex. L indicates anatomical left and R indicates anatomical right.

Label	Hemi	Coordinate		Vertex p-value	Cluster p-value	Size (mm ²)
		MNI (x, y, z)	Talairach			
Frontal Cortex						
SMA	R	(12,3,64)	(9,-5,62)	0.000002	0.012	36.4
Rolandic Cortex						
vSC	L	(-47,-21,35)	(-45,-24,33)	0.000037	0.014	34.7
dSC	L	(-24,-26,64)	(-24,-31,58)	0.000000	0.000	60.6
midPMC	R	(43,-3,51)	(38,-9,50)	0.000000	0.001	58.6
midMC	R	(39,-7,36)	(35,-12,37)	0.000000	0.000	122.9
pCO	R	(44,-7,16)	(39,-10,19)	0.000003	0.015	35.0
vSC	R	(51,-8,27)	(46,-12,28)	0.000001	0.004	43.6
dSC	R	(35,-21,61)	(30,-27,57)	0.000007	0.023	31.6
dSC	R	(49,-20,50)	(44,-25,48)	-0.000010	0.019	33.1
Insular Cortex						
aINs	L	(-29,28,-2)	(-27,25,4)	-0.000003	0.013	35.1
Parietal Cortex						
aSMg	L	(-59,-34,44)	(-56,-37,40)	-0.000003	0.021	32.5

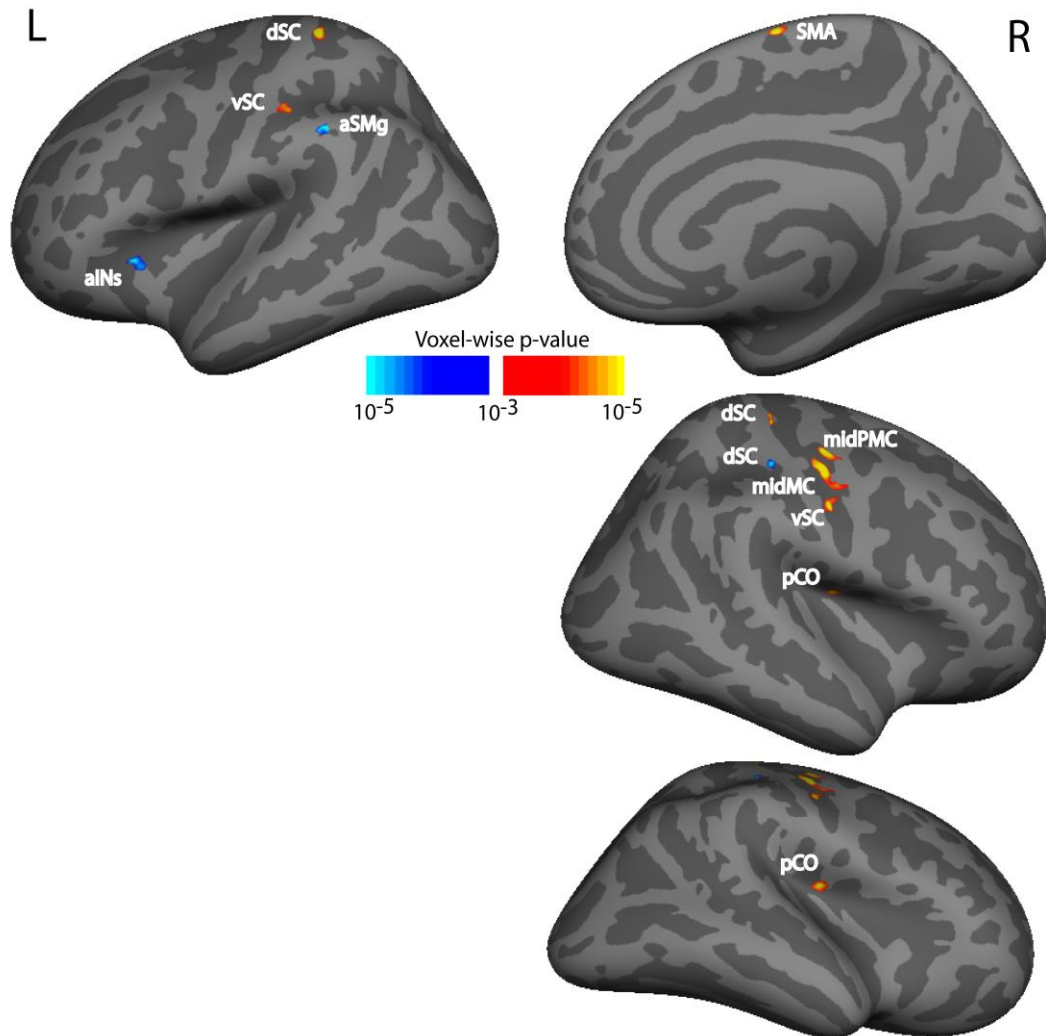


Figure 4-6: Group maps for the Emphatic – Normal contrast after controlling for sentence duration.

Clusters associated with emphatic production are primarily found in the right hemisphere, whereas clusters associated with normal sentence production are almost exclusively found in the left hemisphere. See Table 4-5 for abbreviations of vertex labels.

4.6 Emphatic – Fast

The acoustic analysis demonstrated that the acoustic differences between emphatic and fast sentence production were in terms of prosodic phrasing *and* vowel quality.

Therefore, we would expect that for the contrast of emphatic speech with fast speech, emphatic-speech-specific activity would be in regions that contribute either to the processing associated with producing acoustically more distinct linguistic contrasts *or* to the processing associated with producing acoustically more distinct prosodic contrasts. In general, the locations of significant clusters associated with emphasis in the Emphatic – Fast contrast largely overlap with those observed in association with clear speech for the Clear – Normal and Clear – Fast contrasts. Specifically, three clusters that were more active in association with emphatic speech production were found in left middle premotor cortex, right ventral premotor cortex, and right middle motor cortex (see Figure 4-7 and Table 4-6 for a map and the coordinates of the significant clusters). In contrast, fast-speech-specific clusters were observed in left dorsal somatosensory cortex, left ventral somatosensory cortex, right posterior dorsal superior temporal sulcus, left anterior frontal operculum, and right posterior insula.

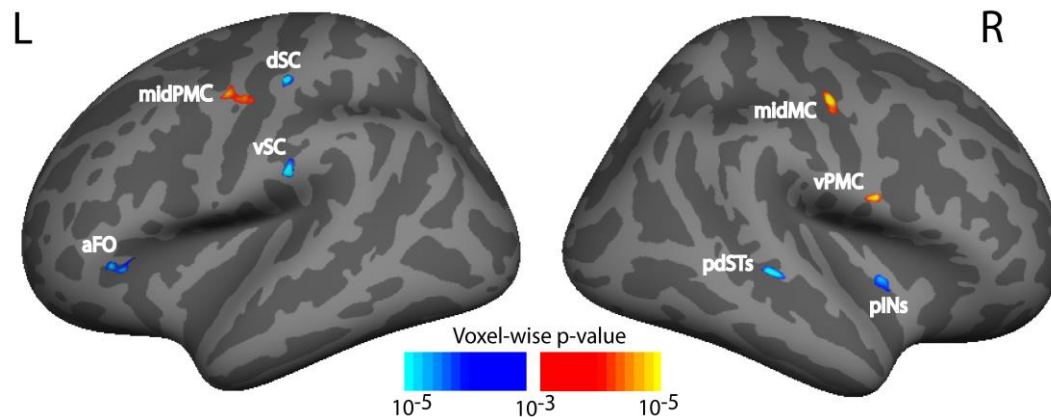


Figure 4-7: Group maps for the Emphatic – Fast contrast after controlling for sentence duration.

See Table 4-6 for abbreviations of vertex labels.

Table 4-6: Significant clusters for the Emphatic – Fast contrast after controlling for sentence duration.

Abbreviations: aFO = anterior frontal operculum; dSC = dorsal somatosensory cortex; midMC = middle motor cortex; midPMC = middle premotor cortex; pdSTs = posterior dorsal superior temporal sulcus; pINs = posterior insula; vPMC = ventral premotor cortex; vSC = ventral somatosensory cortex. L indicates anatomical left and R indicates anatomical right.

Label	Hemi	Coordinate		Vertex p-value	Cluster p-value	Size (mm ²)
		MNI (x, y, z)	Talairach			
Frontal Cortex						
aFO	L	(-43,34,-10)	(-40,31,-2)	-0.000013	0.001	49.3
Insular Cortex						
pINS	R	(41,-2,-6)	(37,-3,0)	-0.000033	0.009	36.8
Rolandic Cortex						
midMC	L	(-51,1,46)	(-49,-5,45)	0.000026	0.000	83.4
vSC	L	(-61,-21,24)	(-58,-23,22)	-0.000000	0.015	33.8
dSC	L	(-47,-12,54)	(-45,-17,50)	-0.000000	0.019	32.2
vPMC	R	(64,9,10)	(58,6,15)	0.000008	0.024	31.0
midMC	R	(39,-9,34)	(35,-13,35)	0.000001	0.000	58.3
Temporal Cortex						
pdSTs	R	(53,-29,0)	(48,-29,2)	-0.000001	0.005	40.7

4.7 Clear – Emphatic

To distinguish those regions that contribute to the processing associated with producing acoustically more distinct linguistic contrasts from those that contribute to the processing associated with producing acoustically more distinct prosodic contrasts we examined the Clear – Emphatic contrast. The only cluster that survived was more active in association with clear speech production and is located in right supramarginal gyrus (see Figure 4-8 and Table 4-7 for a map and the coordinates of the significant clusters).

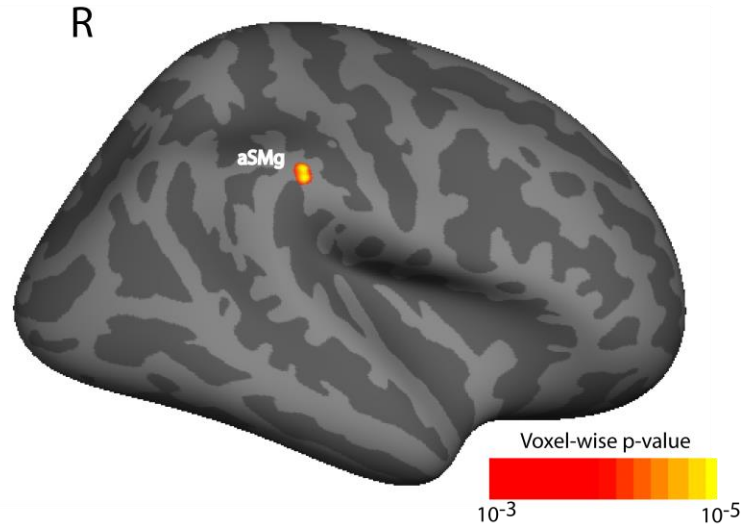


Figure 4-8: Group maps for the Clear – Emphatic contrast after controlling for sentence duration.

Table 4-7: Significant clusters for the Clear – Emphatic contrast after controlling for sentence duration.

Abbreviation: aSMg = anterior supramarginal gyrus. R indicates anatomical right.

Label	Hemi	Coordinate		Vertex p-value	Cluster p-value	Size (mm ²)
		MNI (x, y, z)	Talairach			
Parietal Cortex						
aSMg	R	(64,-20,40)	(58,-24,39)	0.000004	0.002	46.6

4.8 Region of Interest Analysis

In addition to performing direct comparisons between condition-specific brain activity at the level of the voxel we also performed direct comparisons between condition-specific brain activations for 26 spherical ROIs and three anatomically-defined subcortical ROIs (putamen, pallidum, and thalamus). As noted in Chapter 2, spherical ROIs were defined by creating a composite map of the six contrasts for the left and right hemispheres. The location of the vertex with the most significant p-value inside each cluster was selected

to be the point around which a circular ROI was delineated on the left and right lateral and medial inflated surfaces of Freesurfer's fsaverage template (see Figure 4-9).

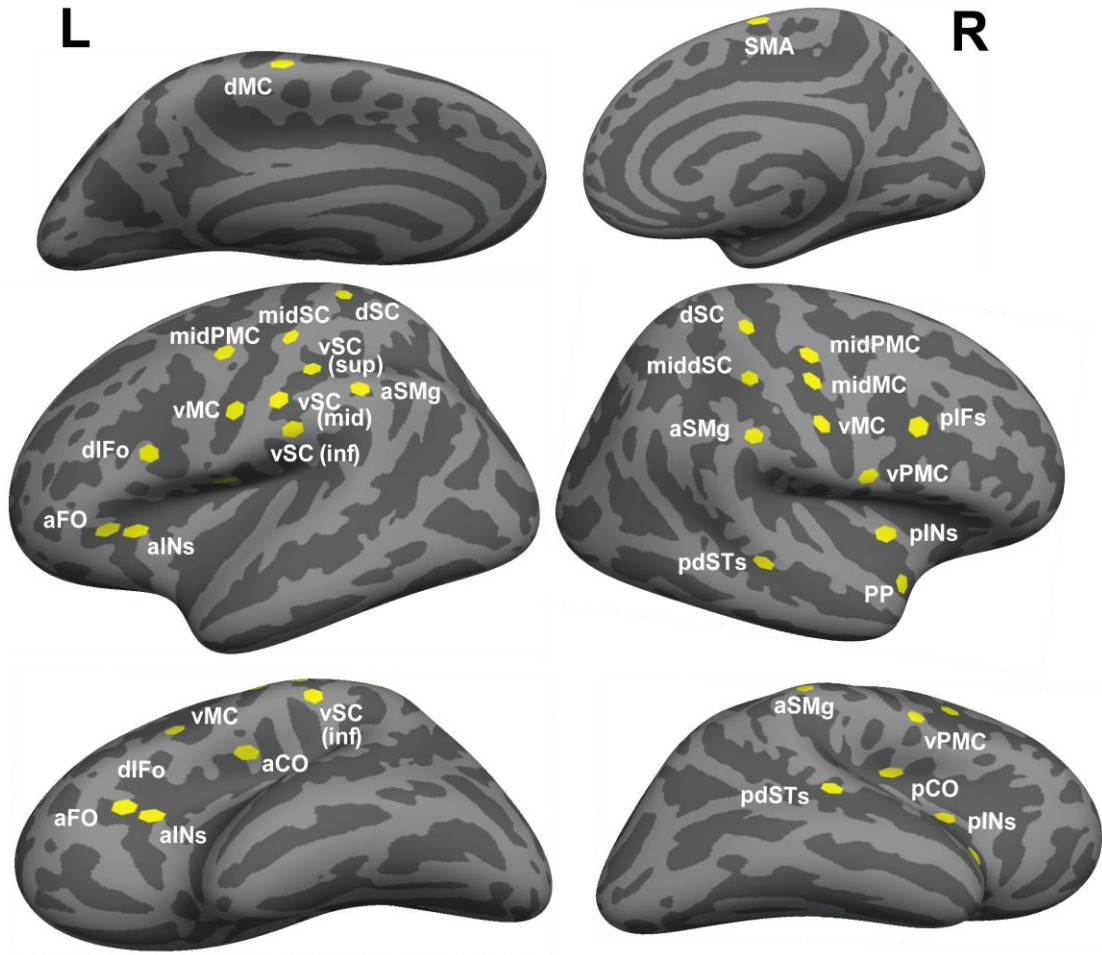


Figure 4-9: Twenty-six regions of interest defined on the surface by creating a composite map of the six contrasts for the left and right hemispheres.

A circular ROI was delineated on the left (L) and right (R) lateral and medial inflated surfaces of Freesurfer's fsaverage template. Abbreviations: aCO = anterior central operculum; aFO = anterior frontal operculum; aINs = anterior insula; aSMg = anterior supramarginal gyrus; dlFo = dorsal inferior frontal gyrus pars opercularis; dMC = dorsal motor cortex; dSC = dorsal somatosensory cortex; midMC = middle motor cortex; midPMC = middle premotor cortex; midSC = middle somatosensory cortex; pal = pallidum; pCO = posterior central operculum; pdSTs = posterior dorsal superior temporal sulcus; pIFs = posterior inferior frontal sulcus; pINs = posterior insula; PP = planum polare; put = putamen; SMA = supplementary motor area; thal = thalamus; vMC = ventral motor cortex; vPMC = ventral premotor cortex; vSC = ventral somatosensory cortex. inf = inferior; mid = middle; sup = superior.

As can be seen in Figure 4-10, relative to normal speech, the effects in left dorsal inferior frontal gyrus pars opercularis increase as conditions change from clear to emphatic to fast speech. However, the effects are not statistically significant for any of the contrasts.

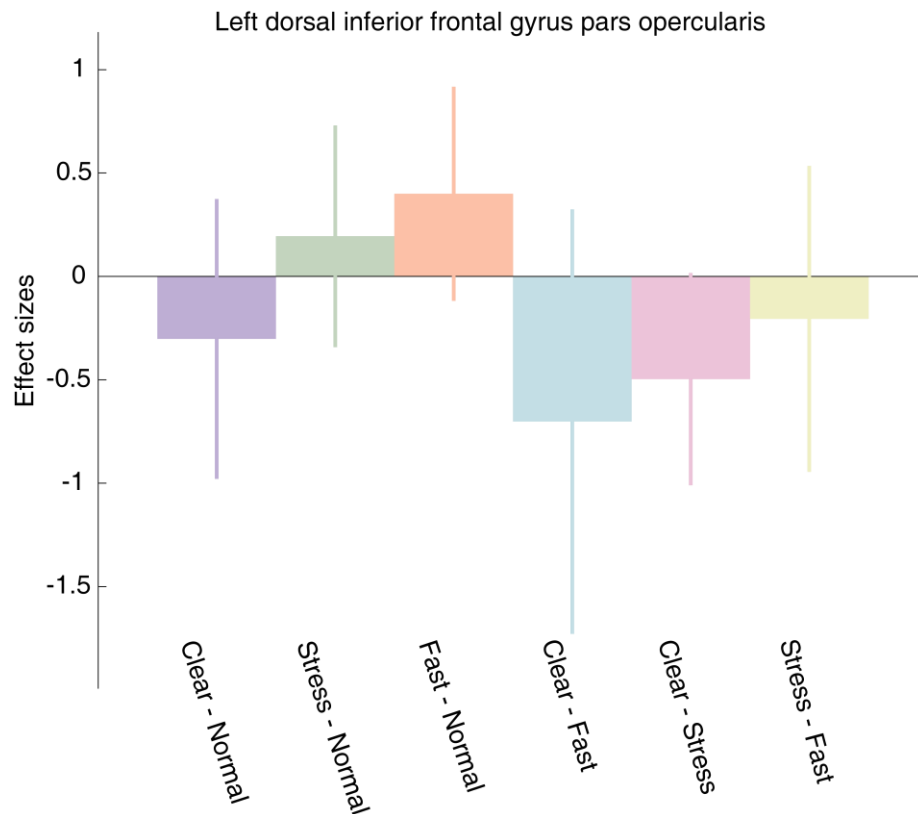


Figure 4-10: Mean effect sizes and 90% confidence interval around the mean for the left dorsal inferior frontal gyrus pars opercularis ROI as a function of contrast.

Relative to normal speech, the effects in left dorsal inferior frontal gyrus pars opercularis increase as conditions change from clear to emphatic to fast speech. However, none of the effects were statistically significant.

The figures that follow provide the results of the ROI analysis for the remaining regions of interest. In a majority of the cases the effects were not found to be statistically significant at a strict threshold of $p_{FDR} \leq 0.05$.

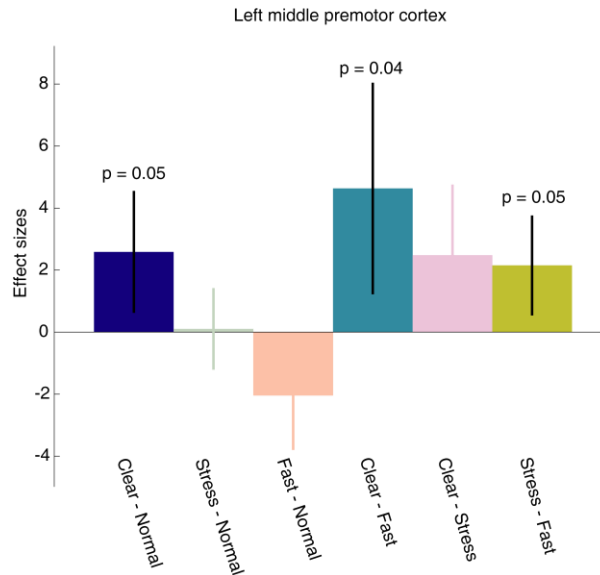


Figure 4-11: Mean effect sizes and 90% confidence interval around the mean for the left middle premotor cortex ROI as a function of contrast.

Consistent with the surface-based analysis this ROI was statistically significant for the clear condition when compared to the normal and fast conditions and for the emphatic condition when compared to the fast condition. Bars with dark lines indicate significant ROI-level effects ($p \leq 0.05$).

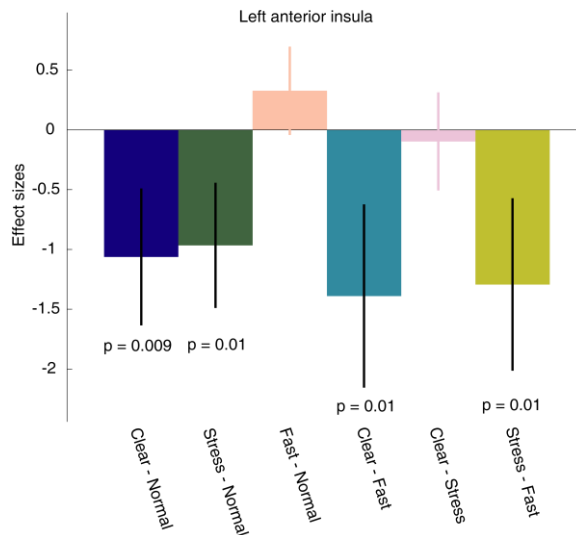


Figure 4-12: Mean effect sizes and 90% confidence interval around the mean for the left anterior insula as a function of contrast.

This ROI was statistically significant for the normal and fast conditions when compared to the clear and stress speech conditions. Bars with dark lines indicate significant ROI-level effects ($p \leq 0.05$).

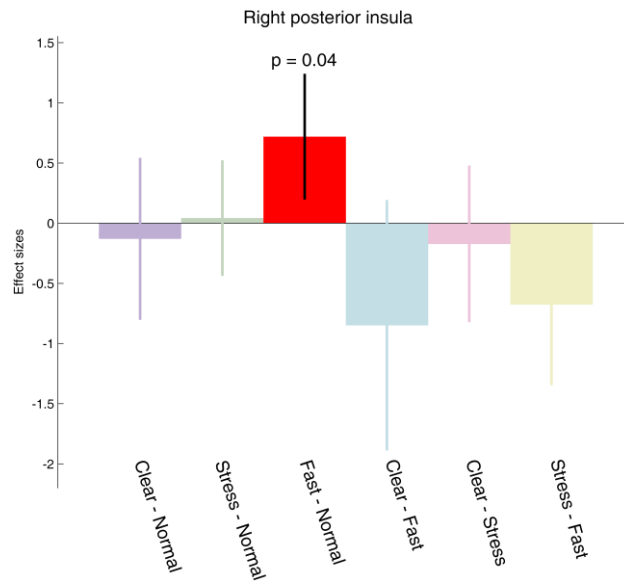


Figure 4-13: Mean effect sizes and 90% confidence interval around the mean for the right posterior insula ROI as a function of contrast.

This ROI was only statistically significant for the fast condition when compared to the normal condition. Bar with dark line indicates a significant ROI-level effect ($p \leq 0.05$).

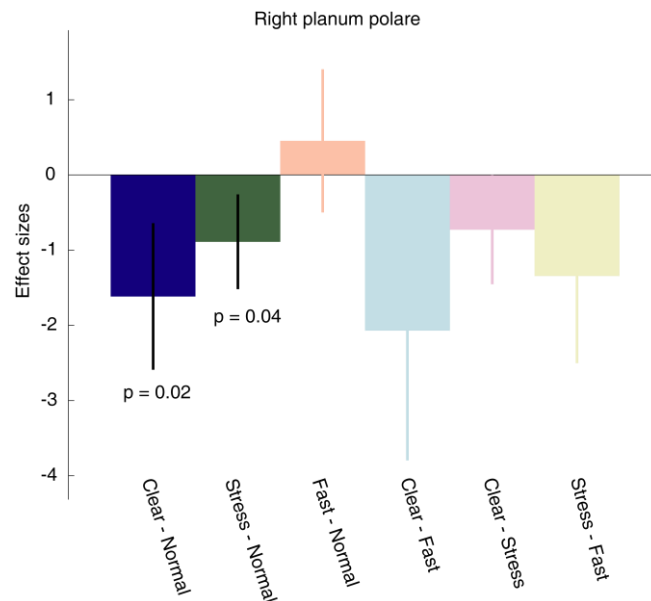


Figure 4-14: Mean effect sizes and 90% confidence interval around the mean for the right planum polare ROI as a function of contrast.

This ROI was statistically significant for the normal condition when compared to the clear and emphatic conditions. Bars with dark lines indicate significant ROI-level effects ($p \leq 0.05$).

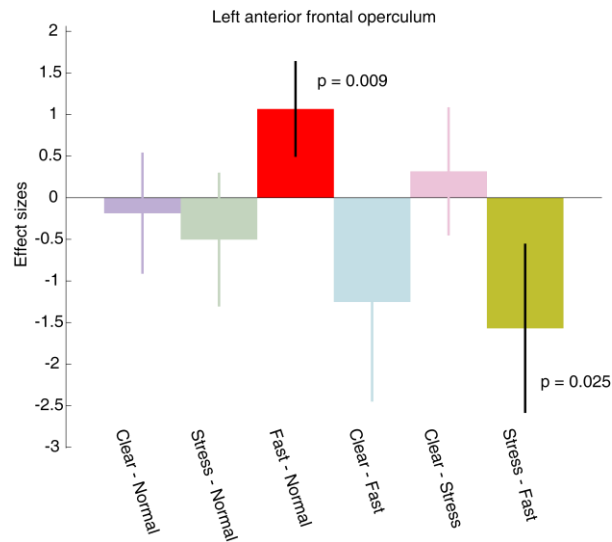


Figure 4-15: Mean effect sizes and 90% confidence interval around the mean for the left anterior frontal operculum ROI as a function of contrast.

This ROI was statistically significant for the fast condition when compared to the normal and emphatic conditions. Bars with dark lines indicate significant ROI-level effects ($p \leq 0.05$).

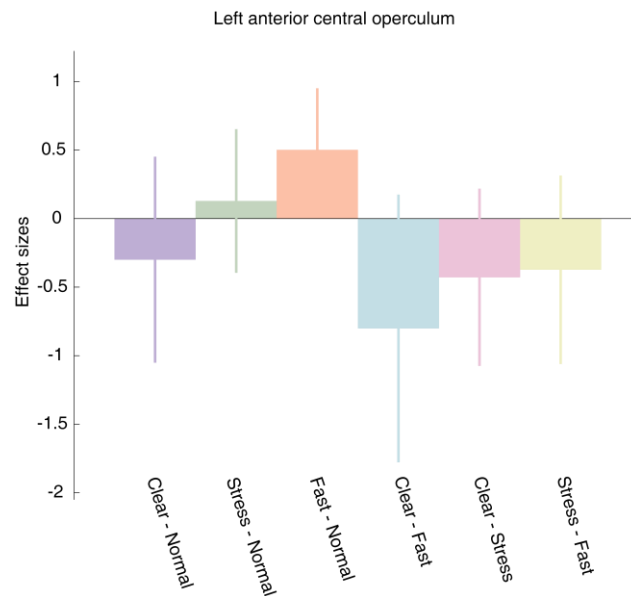


Figure 4-16: Mean effect sizes and 90% confidence interval around the mean for the left anterior central operculum ROI as a function of contrast.

Relative to normal speech, the effects in left central operculum increase as conditions change from clear to emphatic to fast speech. However, none of the effects were statistically significant.

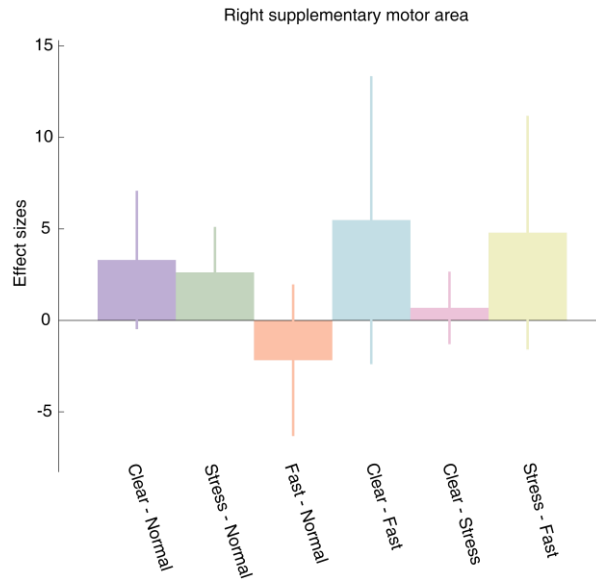


Figure 4-17: Mean effect sizes and 90% confidence interval around the mean for the right supplementary motor area ROI as a function of contrast.

Relative to normal speech, the effects in right supplementary motor area reduce as conditions change from clear to emphatic to fast speech. However, none of the effects were statistically significant.

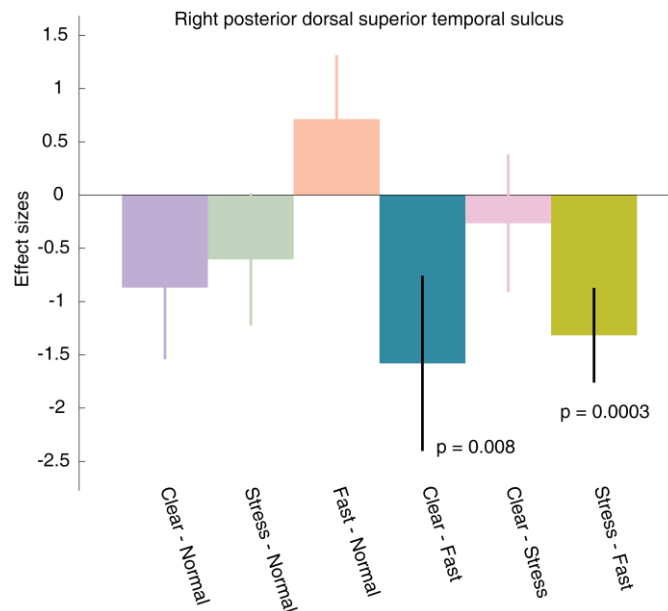


Figure 4-18: Mean effect sizes and 90% confidence interval around the mean for the right posterior dorsal superior temporal sulcus ROI as a function of contrast.

This ROI was statistically significant for the fast condition when compared to the clear and emphatic conditions. Bars with dark lines indicate significant ROI-level effects ($p_{FDR} \leq 0.05$).

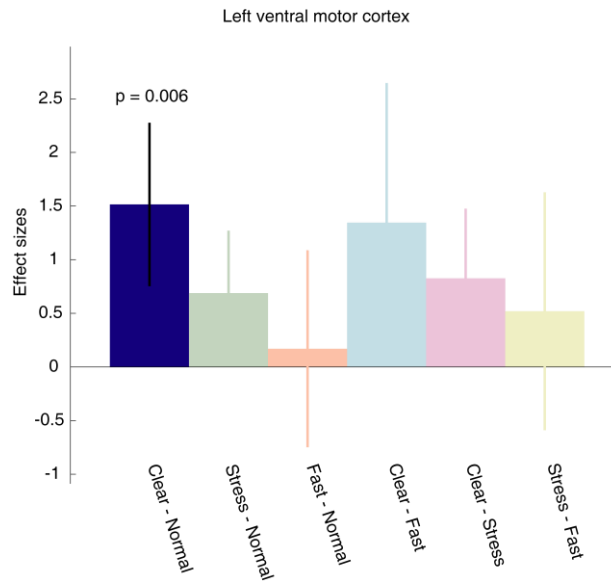


Figure 4-19: Mean effect sizes and 90% confidence interval around the mean for the left ventral motor cortex ROI as a function of contrast.

Consistent with the surface-based analysis, this ROI was statistically significant for the clear condition when compared to the normal condition. Bar with dark line indicates a significant ROI- level effect ($p_{FDR} \leq 0.05$).

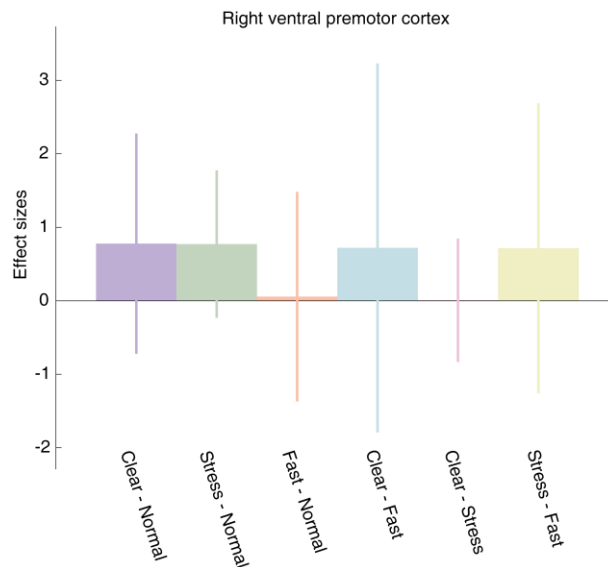


Figure 4-20: Mean effect sizes and 90% confidence interval around the mean for the right ventral premotor cortex ROI as a function of contrast.

Consistent with the surface-based analysis, this ROI demonstrated increased effects for the clear and emphatic conditions relative to normal and fast conditions. However, none of the effects were statistically significant.

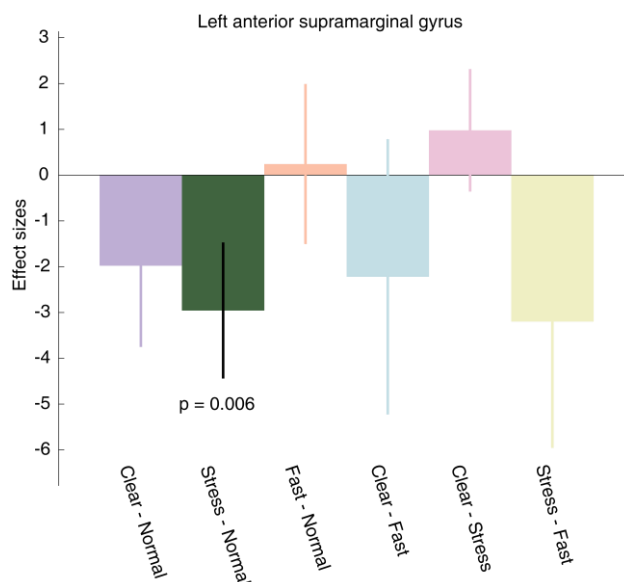


Figure 4-21: Mean effect sizes and 90% confidence interval around the mean for the left anterior supramarginal gyrus ROI as a function of contrast.

Consistent with the surface-based analysis, this ROI demonstrated increased effects for the normal condition when compared to the emphatic condition. This effect was significant ($p_{FDR} \leq 0.05$).

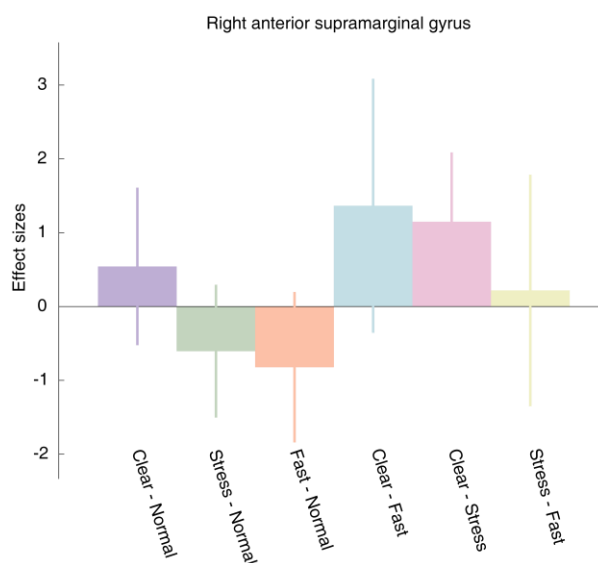


Figure 4-22: Mean effect sizes and 90% confidence interval around the mean for the right anterior supramarginal gyrus ROI as a function of contrast.

This ROI demonstrated increased effects for the clear condition when compared to normal, fast, and emphatic conditions. However, none of the effects were statistically significant.

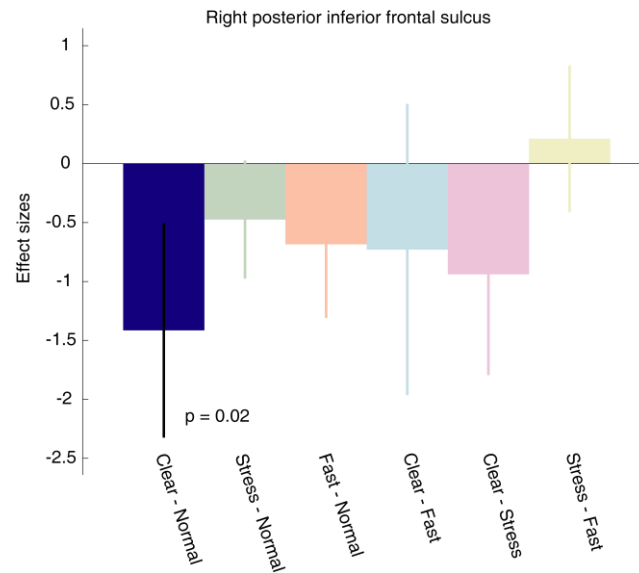


Figure 4-23: Mean effect sizes and 90% confidence interval around the mean for the right posterior inferior frontal sulcus ROI as a function of contrast.

This ROI was statistically significant for the normal condition when compared to the clear condition. Bar with dark line indicates a significant ROI-level effect ($p \leq 0.05$).

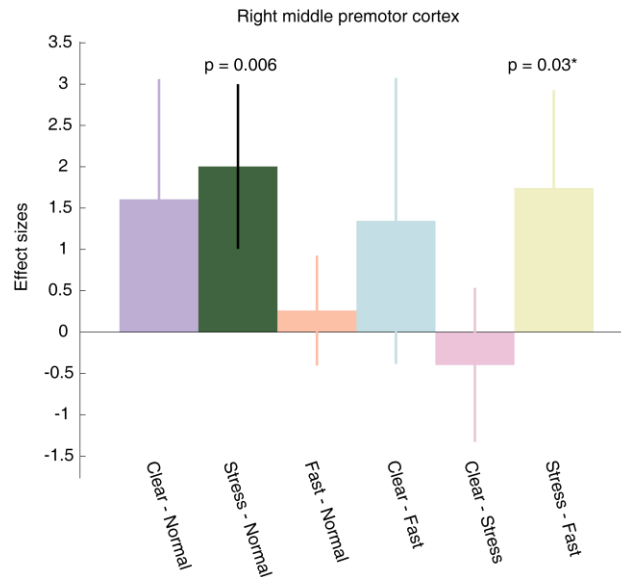


Figure 4-24: Mean effect sizes and 90% confidence interval around the mean for the right middle premotor cortex ROI as a function of contrast.

Consistent with the surface-based analysis, this ROI demonstrated increased effects for the emphatic condition when compared to the normal condition. Bar with dark line indicates a significant ROI-level effect using a strict threshold of $p_{FDR} \leq 0.05$. A significant effect using a less strict threshold of $p \leq 0.05$ was found for emphatic when compared to fast speech.

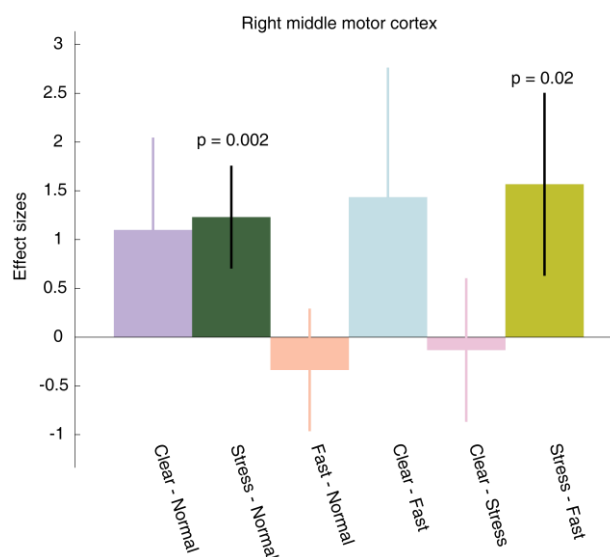


Figure 4-25: Mean effect sizes and 90% confidence interval around the mean for the right middle motor cortex spherical ROI as a function of contrast.

Consistent with the surface-based analysis, this ROI demonstrated increased effects for the emphatic condition when compared to normal and fast conditions. Bars with dark lines indicate significant ROI-level effects ($p_{FDR} \leq 0.05$).

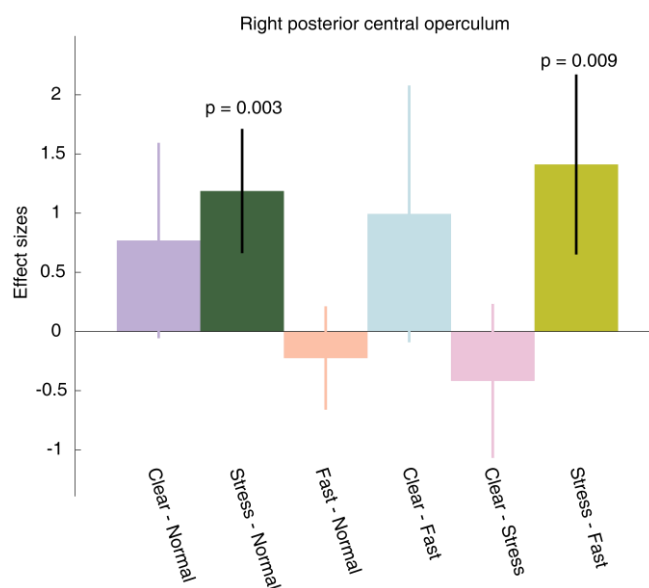


Figure 4-26: Mean effect sizes and 90% confidence interval around the mean for the right posterior central operculum ROI as a function of contrast.

This ROI demonstrated increased effects for the emphatic condition when compared to normal and fast conditions. Bars with dark lines indicate significant ROI-level effects ($p_{FDR} \leq 0.05$).

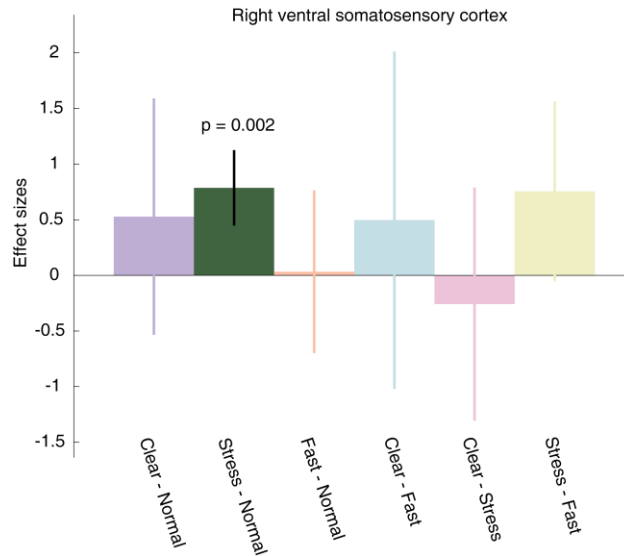


Figure 4-27: Mean effect sizes and 90% confidence interval around the mean for the right ventral somatosensory cortex ROI as a function of contrast.

This ROI demonstrated increased effects for the emphatic condition when compared to normal condition. Bar with dark line indicates a significant ROI-level effect ($p_{FDR} \leq 0.05$).

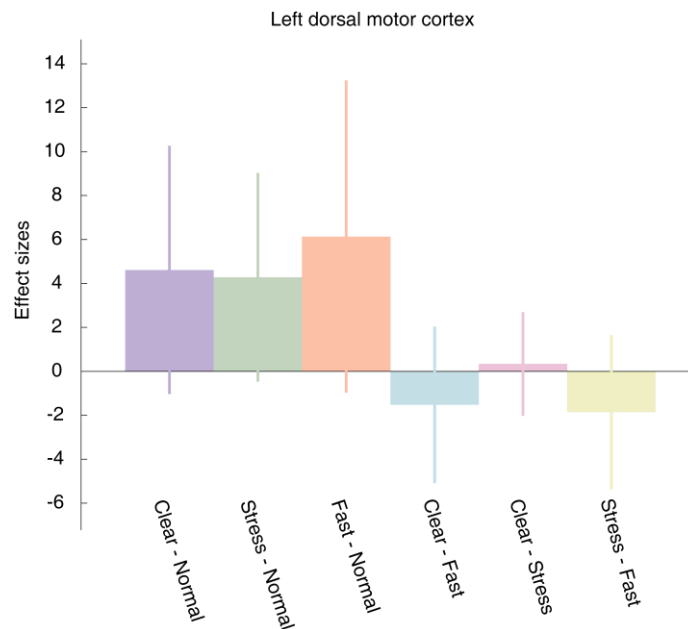


Figure 4-28: Mean effect sizes and 90% confidence interval around the mean for the left dorsal motor cortex ROI as a function of contrast.

This ROI demonstrated increased effects for all three speech manners as compared to normal. However, none of the effects were statistically significant.

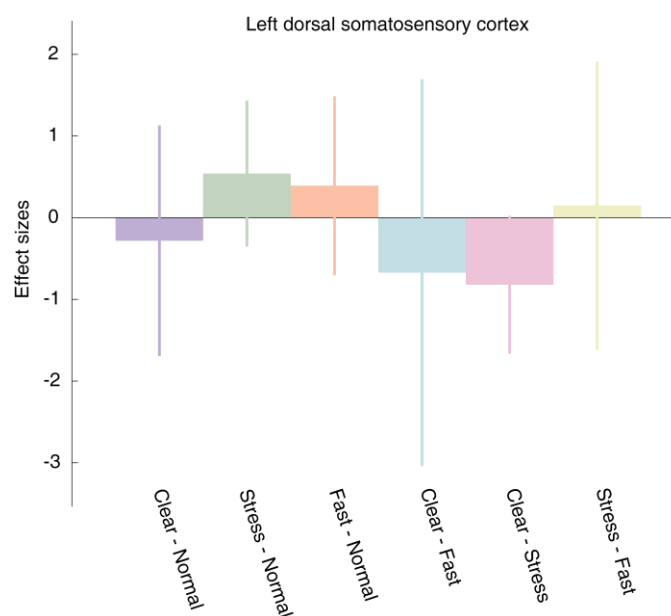


Figure 4-29: Mean effect sizes and 90% confidence interval around the mean for the left dorsal somatosensory cortex spherical ROI as a function of contrast.

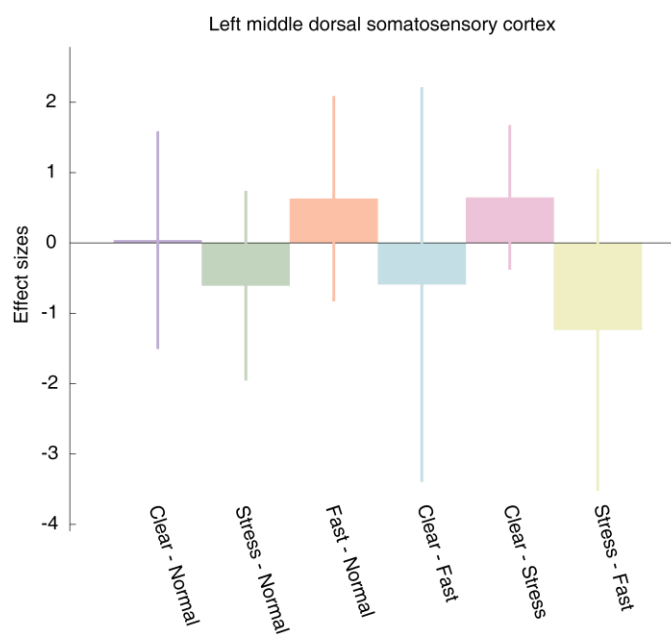


Figure 4-30: Mean effect sizes and 90% confidence interval around the mean for the left middle dorsal somatosensory cortex ROI as a function of contrast.

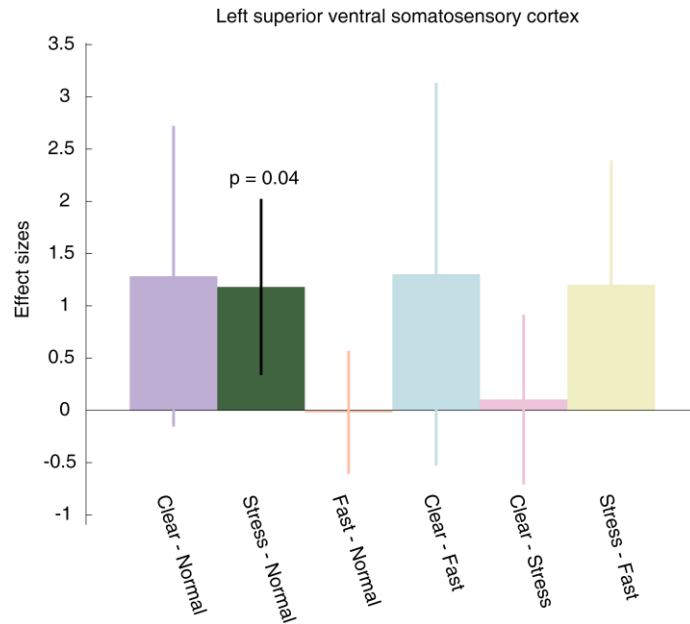


Figure 4-31: Mean effect sizes and 90% confidence interval around the mean for the left superior ventral somatosensory cortex ROI as a function of contrast.

This ROI demonstrated statistically significant increased effects for the emphatic condition when compared to normal ($p \leq 0.05$).

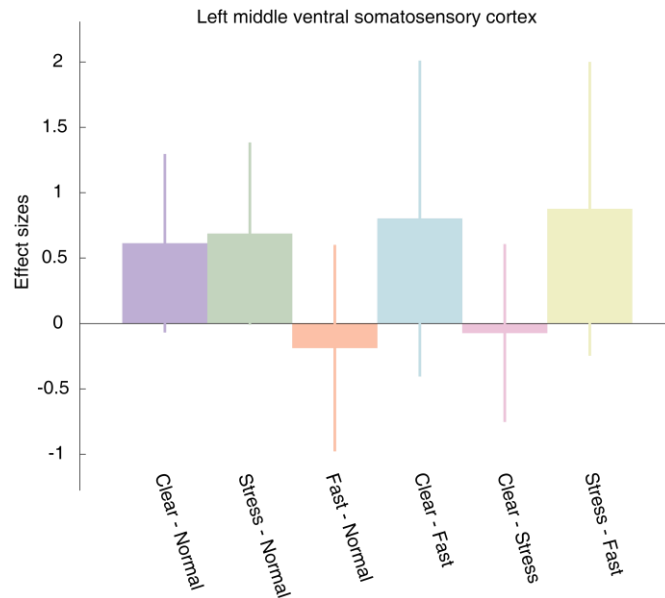


Figure 4-32: Mean effect sizes and 90% confidence interval around the mean for the left middle ventral somatosensory cortex ROI as a function of contrast.

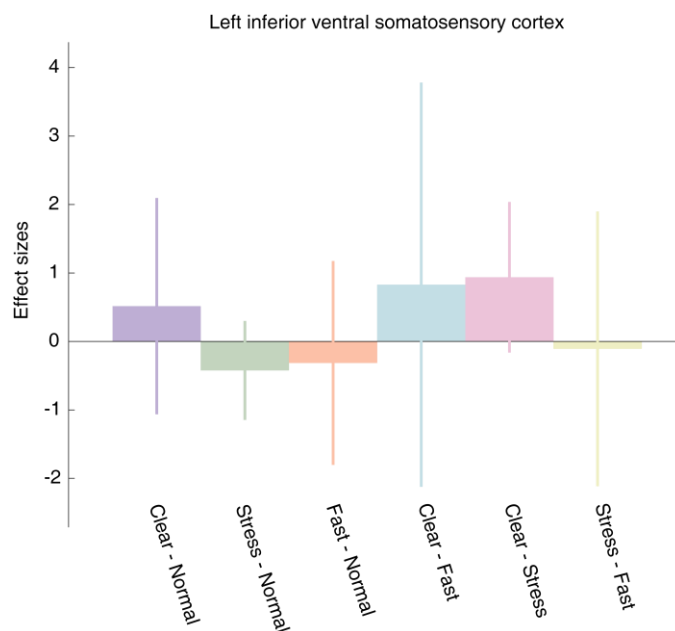


Figure 4-33: Mean effect sizes and 90% confidence interval around the mean for the right dorsal somatosensory cortex spherical ROI as a function of contrast.

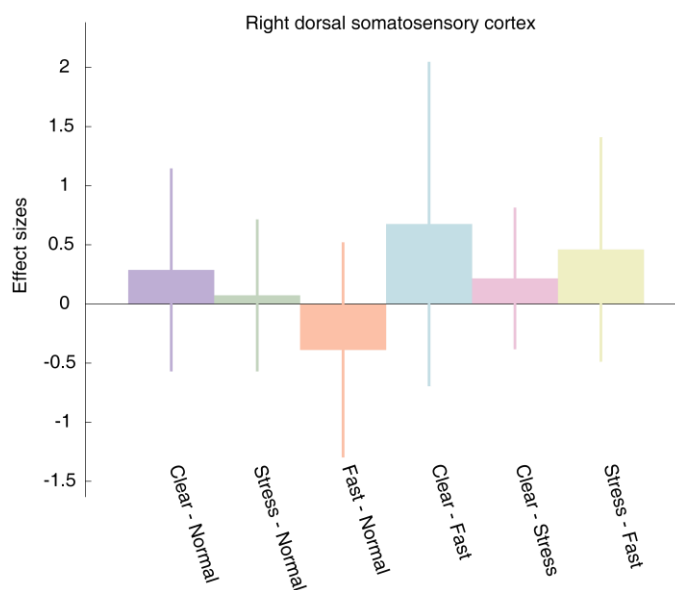


Figure 4-34: Mean effect sizes and 90% confidence interval around the mean for the left inferior ventral somatosensory cortex ROI as a function of contrast.

Relative to normal speech, the effects in right dorsal somatosensory cortex decrease as conditions change from clear to emphatic to fast speech. However, none of the effects were statistically significant.

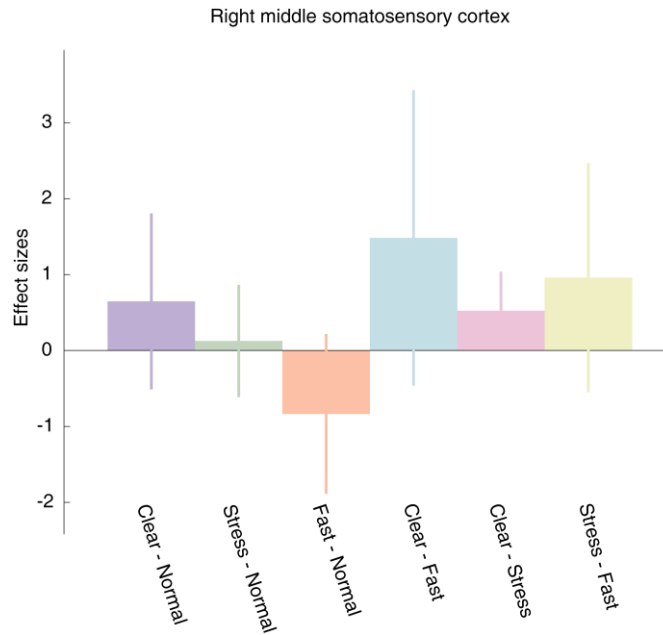


Figure 4-35: Mean effect sizes and 90% confidence interval around the mean for the right middle somatosensory cortex ROI as a function of contrast.

Relative to normal speech, the effects in right middle somatosensory cortex decrease as conditions change from clear to emphatic to fast speech. However, none of the effects were statistically significant.

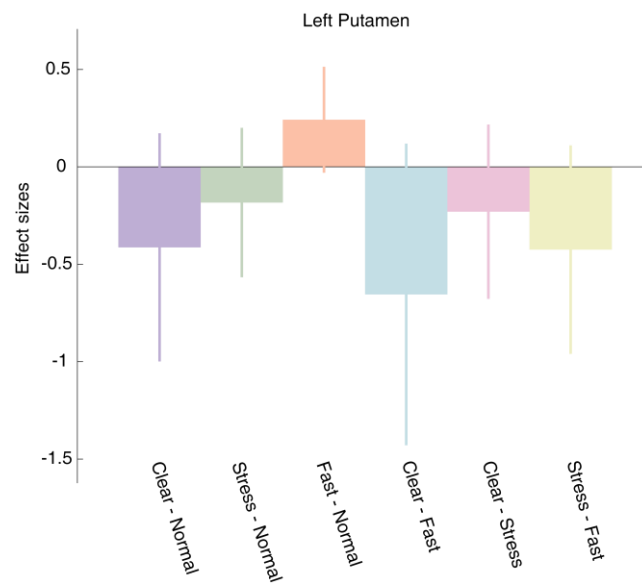


Figure 4-36: Mean effect sizes and 90% confidence interval around the mean for the left putamen as a function of contrast.

Relative to normal speech, the effects in left putamen increase as conditions change from clear to emphatic to fast speech. However, none of the effects were statistically significant.

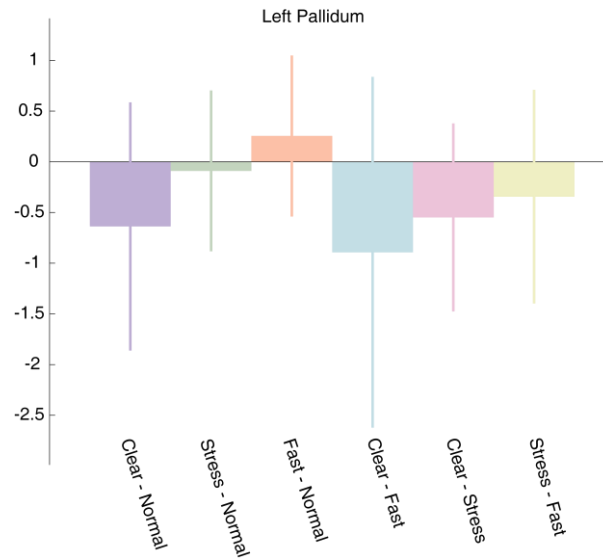


Figure 4-37: Mean effect sizes and 90% confidence interval around the mean for the left pallidum as a function of contrast.

Relative to normal speech, the effects in left pallidum show a similar trend as those in left putamen, increasing as conditions change from clear to emphatic to fast speech. However, none of the effects were statistically significant.

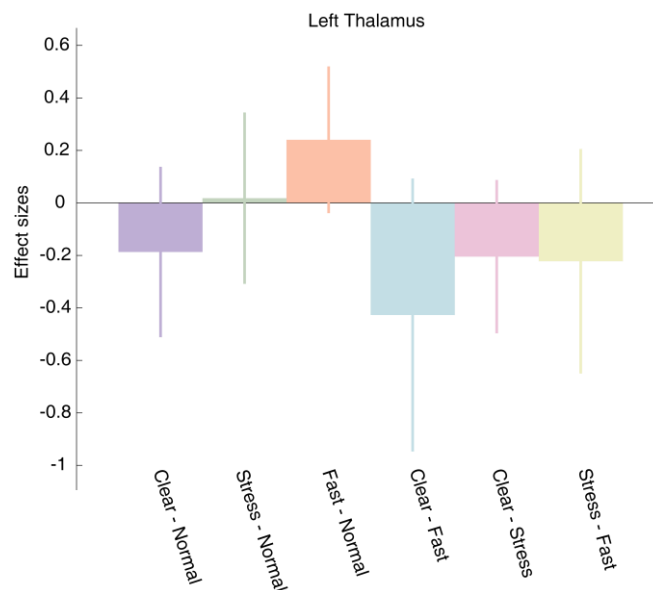


Figure 4-38: Mean effect sizes and 90% confidence interval around the mean for the left thalamus as a function of contrast.

Relative to normal speech, the effects in left thalamus show a similar trend as those in left putamen and pallidum, increasing as conditions change from clear to emphatic to fast speech. However, none of the effects were statistically significant.

4.9 Partial Pearson Correlation Analysis

In order to determine whether the engagement of the clusters revealed by the six main contrasts of interest is correlated with the acoustic properties of the speech sounds produced, separate partial Pearson's correlation analyses were performed on the 29 ROIs (26 spherical ROIs and 3 anatomically-defined subcortical ROIs) and six acoustic measures. A challenge facing any study that uses behavioral measures to understand the increased engagement of a brain region is dealing with reduced sensitivity as a result of the fact that the behavioral variable of interest is correlated with other behavioral variables and the shared properties of these variables modulate a region's activity (cf. Graves et al., 2007). In order to control for shared properties between the acoustic measures derived in this study, we used a *partial* Pearson correlation approach which is used to measure the association between two time series after partialling out the effects of other time series. By partialling out the five acoustic measures and the conditions, we can determine whether the processes engaging a particular ROI are specific to the acoustic measure that we are examining.

The partial correlation analysis revealed a significant positive conditional correlation between inter-vowel acoustic distances and the clusters in premotor and motor cortical areas typically associated with oro-laryngeal control (left middle premotor cortex, left ventral motor cortex, left anterior central operculum, right ventral premotor cortex, right middle motor cortex, right posterior central operculum, right dorsal somatosensory cortex, and right ventral somatosensory cortex; see Table 4-8 for t-values of the group-level tests). A significant positive conditional correlation between average

vowel spacing and the mean responses in sensorimotor cluster-based ROIs suggests a direct relationship between the processes associated with increasing the acoustic contrast of vowels and the engagement of these regions. The mean BOLD response in left middle premotor cortex was also positively correlated with average adjacent-vowel duration contrast and average vowel intensity. These findings indicate that the processes that contribute to the engagement of left middle premotor cortex are not only associated with determining the quality of the vowels but also certain aspects of the prosodic context. Left dorsal inferior frontal gyrus, left ventral motor cortex, and left anterior supramarginal gyrus were all significantly positively correlated with average adjacent-vowel intensity contrast, indicating that these regions may contribute to the processing associated with determining aspects of the prosodic organization or the precise intensity values that are an integral part of the hierarchical prosodic structure. The mean response in left anterior supramarginal gyrus was also negatively correlated with average vowel spacing, indicating that the engagement of this region increases when vowels are acoustically less separable in terms of intensity. Several regions including right supplementary motor area, right posterior inferior frontal sulcus, and left anterior frontal operculum were not significantly correlated with any of the acoustic measures. These results may be interpreted as indicating that the acoustic measures are not linearly related to the processes that engage these regions. Alternatively, it is possible that this analysis was not sensitive enough to determine whether a linear relationship exists.

Table 4-8: Group-level conditional correlation results for the 29 ROIs and six acoustic measures.

Positive and negative values indicate a positive and negative conditional correlation, respectively. The **bold** font and ** indicate that the conditional correlation survived a threshold of $p_{FDR} \leq 0.05$ whereas the * indicates that the conditional correlation survived a threshold of $p \leq 0.05$. See Figure 4-9 for abbreviations of ROIs.

ROI	T-values					
	Vowel Spacing	Vowel Duration	Adjacent-vowel Duration Contrast	Adjacent-vowel F0 Contrast	Vowel Intensity	Adjacent-vowel Intensity Contrast
L aFO	0.4	0.8	-0.1	0.2	1.3	0.5
L aINs	0.7	0.7	0.5	-0.7	0.1	2.9*
L dIFo	1.7	0.4	2.5*	-1.2	1.1	3.5**
L midPMC	4.6**	1.1	3.4**	-0.4	2.6**	2.1
L dMC	1.1	0.1	-0.7	0.7	2.5*	1.5
L vMC	3.4**	0.7	-0.2	1.6	2.5*	2.9**
L aCO	3.7**	2.4*	0.5	0.7	1.7	1.7
L sup vSC	-1.1	0.6	-1.4	-1.0	1.0	-0.3
L mid vSC	1.3	0.6	1.0	1.0	1.4	0.5
L inf vSC	1.4	2.5*	-0.8	-0.4	-0.9	-0.8
L midSC	1.2	0.2	-0.9	-0.8	1.4	1.2
L dSC	2.1	0.4	0.2	-0.6	2.3*	1.6
L put	3.5**	0.5	0.2	-1.7	1.1	1.3
L pal	3.5**	-0.1	-0.4	-0.5	0.7	3.2**
L thal	3.6**	0.6	-0.2	-0.9	4.0**	0.8
L aSMg	-3.3**	0.8	0.2	1.1	2.3*	2.9**
R pIFs	-0.4	-0.7	1.3	0.3	-1.1	1.9
R pINs	2.2*	0.5	-0.1	0.1	1.8	1.1
R SMA	1.8	0.5	-0.9	0.0	2.3*	0.6
R vPMC	3.4**	-0.3	-0.2	0.3	0.1	0.5
R midPMC	2.3*	1.1	1.4	0.6	0.5	0.0
R midMC	3.7**	1.4	1.5	0.8	2.1	0.9
R pCO	5.3**	-2.2*	1.9	1.0	0.3	-0.4
R dSC	3.1**	0.8	0.0	-1.5	0.3	1.6
R midSC	-0.5	1.5	0.2	-0.4	-0.7	0.4
R vSC	4.6**	0.8	1.5	0.0	2.2*	1.9
R aSMg	-0.3	0.7	-2.5*	1.3	-0.2	-1.1
R PP	-0.2	-0.5	-0.4	-2.0	-1.0	1.3
R pdSTs	2.9*	-0.6	-0.7	-2.7*	0.3	1.4

4.10 Functional Connectivity Analysis

A partial Pearson correlation analysis was also employed to evaluate the separable, but potentially interacting, functional contributions of left posterior dorsal inferior frontal

gyrus pars opercularis and left middle premotor cortex to phonological and phonetic encoding on the basis of their interactions with the putamen, classically considered to be a motor structure. Specifically, differences in the functional connectivity between the fast and normal conditions were assessed for a network of six ROIs that included left posterior dorsal inferior frontal gyrus, left middle premotor cortex, left motor cortex, left putamen, left pallidum, and left thalamus. The two conditions were selected for the functional connectivity analysis because they each showed increased activity in one of the two regions of interest (left posterior dorsal inferior frontal gyrus pars opercularis or left middle premotor cortex) in the Fast – Normal contrast of the surface-based analysis. Group-level t-values for conditional connectivity between ROIs for the normal and fast conditions are summarized in Table 4-9. Table 4-10 provides the t-values for pair-wise comparisons of the conditional connectivity between the fast and normal conditions. Pair-wise comparisons of the functional connectivity between the fast and normal conditions revealed that the negative conditional correlation between left middle premotor cortex and left pallidum was significantly more negative under fast than normal conditions. In other words, under fast speech conditions as left middle premotor cortex becomes increasingly engaged, activation in left pallidum reduces significantly. In addition, the positive correlation between left middle premotor cortex and left putamen under fast speech conditions showed a trend toward being significantly greater under fast than normal sentence production conditions. These results suggest that under fast speech conditions the putamen receives excitatory drive from the premotor cortex which in turn reduces the basal ganglia's inhibitory control over motor areas and facilitates more rapid

production of the motor programs underlying speech. However, since response selection at the level of the basal ganglia is made more rapidly than normal, the selection may not be influenced as much by inputs from regions processing sensory feedback that would otherwise help to refine selections. More generally, these findings indicate that there is a significant relationship between left middle premotor cortex and basal ganglia. In no cases was a significant relationship between left posterior dorsal inferior frontal gyrus and regions within the basal ganglia found. Together these results provide additional evidence for a distinction in the functional recruitment of left middle premotor cortex and left posterior dorsal inferior frontal gyrus.

Table 4-9: Group-level conditional connectivity for the normal and fast conditions.

The left and right tables provide t-values for conditional connectivity between ROIs for the normal and fast conditions, respectively. Positive values indicate a positive conditional correlation and negative values indicate a negative conditional correlation. The **bold** font and ** indicate that the conditional correlation survived a threshold of $p_{FDR} \leq 0.05$. Abbreviations: dIFo = dorsal inferior frontal gyrus pars opercularis; midPMC = middle premotor cortex; vMC = ventral motor cortex; put = putamen; pal = pallidum; thal = thalamus. L indicates anatomical left and R indicates anatomical right. The * indicates that the conditional correlation was only borderline significant with $p = 0.06$.

ROI	T-values for the Normal Condition						T-values for the Fast Condition					
	L midPMC	L vMC	L put	L pal	L thal		L midPMC	L vMC	L put	L pal	L thal	
L dIFo	2.51	1.19	0.74	0.02	2.22		1.43	1.61	1.01	-0.89	1.74	
L midPMC		0.97	-0.20	-0.12	2.45			0.46	2.08*	-3.06**	1.57	
L vMC			0.26	0.09	-0.54				0.44	-0.81	0.23	
L put				4.32**	4.73**					5.41**	5.54**	
L pal					1.46						3.01**	

Table 4-10: Pair-wise comparisons of conditional connectivity between the fast and normal conditions.

The * indicates that the conditional correlation was only borderline significant $p = 0.07$.

T-values for the Fast – Normal Contrast					
ROI	L midPMC	L vMC	L put	L pal	L thal
L dIFo	-0.96	0.27	0.55	-1.28	-0.75
L midPMC		-0.59	1.95*	-3.43**	-1.34
L vMC			0.32	-1.22	1.10
L put				-0.45	-0.01
L pal					0.38

Under baseline conditions the engagement of left middle premotor cortex is significantly positively correlated with the engagement of left posterior dorsal inferior frontal gyrus and left thalamus (See Table 4-11). This significant baseline connectivity does not differ significantly from speech production (all tasks; See Table 4-12).

Table 4-11: Group-level conditional connectivity for the speech (all manners) and baseline conditions.

ROI	T-values for the Baseline Condition					T-values for Speech				
	L midPMC	L vMC	L put	L pal	L thal	L midPMC	L vMC	L put	L pal	L thal
L dIFo	2.78**	1.69	1.00	-0.43	2.94**	2.07	2.08	0.79	-0.62	3.27**
L midPMC		1.00	0.91	-1.17	2.76*		0.57	1.45	-1.22	2.40
L vMC			1.02	0.22	-1.36			0.31	-0.23	-0.21
L put				3.59**	4.99**				6.07**	6.58**
L pal					3.05**					1.84

Table 4-12: Pair-wise comparisons of conditional connectivity between the speech (all manners) and baseline conditions.

T-values for the Speech – Baseline Contrast					
ROI	L midPMC	L vMC	L put	L pal	L thal
L dIFo	-1.80	-0.28	-0.40	-0.04	-0.36
L midPMC		-0.89	0.75	0.60	-0.27
L vMC			-1.53	-0.73	1.92
L put				0.39	0.63
L pal					-1.55

4.11 Discussion

Recent white matter fiber tracking studies have demonstrated two distinct routes between frontal and temporal areas that subserve speech and language processing (Saur et al., 2008; Wilson et al., 2011): 1) a dorsal route linking superior temporal areas to the posterior portion of inferior frontal gyrus pars opercularis and dorsal premotor cortex; and 2) a ventral route linking anterior and middle temporal regions to inferior frontal gyrus pars orbitalis, pars triangularis, and, to some extent, pars opercularis. The exact functional contributions of these pathways seem to be a matter of debate. For instance, it has been proposed that a pathways' involvement may be distinguished on the basis of the sensory stimulus inducing speech and language processing. More specifically, it has been suggested that speakers may use different computational processes depending on whether stimuli are semantically meaningful or not (e.g., Binder et al., 2003; Fiebach et al., 2002; Fiez et al., 1999; Hickok and Poeppel, 2004, 2007; McDermott et al., 2003; Poldrack et al., 1999; Price et al., 1997). Within this framework, it has been shown that the dorsal pathway is involved in grapheme-to-phoneme processing, assembling and reading out word-like non-words, while the ventral pathway supports the direct mapping between orthographic percepts and stored lexico-semantic representations (e.g., Fiebach et al., 2002; Fiez et al., 1999; McDermott et al., 2003; Saur et al., 2008). Consistent with a theory of a ventral semantic processing system, anterolateral superior temporal gyrus and sulcus have been associated with intelligible stimulus perception (e.g., Binder et al., 2000; Davis and Johnsrude, 2003; Obleser and Kotz, 2010; Okada et al., 2010; Scott et al., 2000) whereas planum temporale and supramarginal gyrus, regions implicated in the

dorsal route, have been reported in association with phonological assembly when lexical/semantic information is not available or directly applicable for task performance (e.g., Fiez et al., 1999; McDermott et al., 2003; Price et al., 1997). A related view suggests a dichotomy based on the involvement of verbal working memory (Jonides et al., 1998; Paulesu et al., 1993). Paulesu et al. (1993) have suggested that in working memory for letters, Broca's area acts as an articulatory mechanism for subvocal rehearsal while left supramarginal gyrus acts as a phonological storage site. Structurally it has been shown that Broca's area connects both directly and indirectly through the inferior parietal lobule with Wernicke's area (Catani and Jones, 2005), whereas the hippocampus, a region implicated in episodic memory encoding and potentially binding associations to lexical representations (Binder et al., 2003; Jackson and Schacter, 2004), has been shown to be among the regions that connect with orbitofrontal regions, including the inferior frontal gyrus, via the uncinate fasciculus (Catani et al., 2002). Alternatively, the dorsal route could contribute to higher-level processes such as organizing words and phrases to create meaningful sentences. Wilson et al. (2011) have shown that reduced white matter integrity in the left superior longitudinal fasciculus/arcuate fasciculus is correlated with syntax processing whereas there was a trend for the ventral route to be correlated with single word processing (picture naming and word comprehension).

The majority of these accounts do not directly address functional distinctions in the recruitment of left posterior dorsal inferior frontal gyrus pars opercularis and left lateral premotor cortex (although see Friederici, 2011), despite a growing number of neuroimaging studies that indicate that such a distinction exists (e.g., Alario et al., 2006;

Indefrey and Levelt, 2004; Papoutsis et al., 2009; Wise et al., 1999). In the present study we extend the earlier neuroimaging findings that have revealed differences in the functional contributions of left posterior dorsal inferior frontal gyrus and left lateral premotor cortex, by mapping activation sites associated with changing speech manner during sentence production. Changes in speech manner typically result in acoustic variations that do not change the lexical or semantic identity of utterances but do affect the acoustic saliency of linguistic and/or prosodic contrasts. The use of sentence stimuli afforded us the possibility of examining differences between the conditions in terms of phrase-level and local-level acoustic changes in vowel quality and prosodic phrasing.

Significant increased BOLD response in left middle premotor cortex was consistently associated with clear speech production, which demonstrated increased average vowel spacing in the acoustic analysis when compared to all the other speech conditions. Left middle premotor cortex was also significantly more active for emphatic sentence production when compared to fast sentence production. The acoustic analysis demonstrated that emphatic sentence production was associated with significantly greater acoustic vowel contrast than fast sentence production. It is unlikely that increased activation in left middle premotor cortex in association with clear speech, when compared to normal and fast speech, and in association with emphatic speech, when compared to fast speech, is attributed to task difficulty as a cluster in this region was also found for the, presumably, least effortful normal sentence production task when compared to fast speech. This finding is consistent with the interpretation that left middle premotor cortex contributes to the processing associated with increasing the acoustic

contrast of linguistic units (e.g., phonemes, syllables, lexical tones) as the acoustic analysis demonstrated that normal sentence production was associated with significantly greater average vowel spacing than fast speech. Further evidence in support of the interpretation that the engagement of left middle premotor cortex is sensitive to the phonetic details of phonological contrasts and may directly contribute to their physical realization is provided by the partial Pearson correlation analysis demonstrating that left middle premotor cortex is positively correlated with average vowel spacing. This conditional correlation result indicates that the engagement of left middle premotor cortex increases when vowels are acoustically more separable. The weight of evidence from the present study favors the interpretation that left middle premotor cortex is engaged when linguistic contrasts are relatively more acoustically distinct. While there is less evidence available from the present study supporting the engagement of the homologous right hemisphere region in determining the physical realization of prosodic contrasts, this region was significantly engaged for emphatic sentence production in the Emphatic – Normal contrast. The acoustic analysis demonstrated that the acoustic measures that differentiated emphatic from normal sentence production were primarily those that characterized the prosodic patterning of phrases, indicating that the homologous right hemisphere middle premotor region may contribute to processing that determines the prosodic organization of phrases or the precise prosodic values that constitute the hierarchical prosodic structure used under emphatic speech conditions.

Previous neuroimaging findings demonstrate relatively increased activation in left middle premotor cortex not only when tasks involve mapping phonological

representations to articulatory codes (Callan et al., 2006; Kircher et al., 2004), but also during tasks that involve coordinated laryngeal and supralaryngeal movements such as glottal stops and phonation (Brown et al., 2008) or smiling (Hennenlotter et al., 2005). Furthermore, this region is sensitive to external sensory cues that relate to the individual's own motor experiences performing the same or similar tasks. For example, a recent neuroimaging study found increased activation bilaterally in this region when both experienced musicians and actresses process stimuli performed by another individual that is most familiar to them given their expertise (i.e., violin music for violinists and dramatic monologues for actresses; Dick et al., 2011). This region has also been found to be relatively more active when study participants view others performing oro-facial actions as compared to viewing static face images (e.g., Buccino et al., 2001). In the study of Buccino et al. (2001) the findings were related to mirror neurons, found in monkey premotor cortex (area F5) and active both when a monkey observes another individual performing specific goal-directed movements (such as grasping, holding, and tearing) and when the monkey performs a similar task (di Pellegrino et al., 1992). In the study of Buccino and colleagues (2001) activation in premotor cortex was somatotopically arranged in response to observing actions involving different effectors, suggesting that perceptual cues of familiar actions are mapped to their corresponding motor representations. Taken together, the results of the present and earlier studies suggest that the engagement of middle premotor cortex is sensitive to the phonetic details of phonological contrasts and may directly contribute to their physical realization.

In the present study, we found an effect of rate in posterior dorsal inferior frontal gyrus pars opercularis when compared to normal speech. This finding is consistent with the reports of several neuroimaging studies for the increased engagement of posterior inferior frontal gyrus as rate increases (e.g., Sidtis et al., 2006; Sidtis et al., 2003). Previous neuroimaging studies indicate that posterior dorsal inferior frontal gyrus pars opercularis is engaged in processes that operate on phonological representations such as syllabification (Costafreda et al., 2006; Indefrey and Levelt, 2004; Moser et al., 2009; Papoutsis et al., 2009). In particular, Papoutsis et al. (2009) showed a significant main effect of pseudoword length (4 vs. two syllables), but no effect of phonotactic frequency in posterior dorsal inferior frontal gyrus, indicating that the number of syllables and not production familiarity significantly engages this region. In addition, studies have observed that posterior inferior frontal gyrus is relatively more active for tasks involving sequencing (Bohland and Guenther, 2006; Gelfand and Bookheimer, 2003), perceptual verbal transformation effects (Kondo and Kashino, 2009; Sato et al., 2004), and the temporal ordering of syllables (Moser et al., 2009). Therefore, tasks that engage this region often seem to involve processing sequential patterns either for the purposes of determining how constituents fit within the sequence or how the entire sequence is influenced by constituents contained within the sequence. The finding from the present study suggests that processing at the level of left posterior dorsal inferior frontal gyrus pars opercularis can be modulated by conditions that do not change the lexical or semantic identity of utterances but do reduce the acoustic saliency of phonological contrasts.

The surface-based analysis indicated that left anterior frontal operculum may also contribute to phonological encoding. In particular, significant increased BOLD response in left anterior frontal operculum was observed for fast sentence production relative to normal and emphatic sentence production. When compared to emphatic speech in the acoustic analysis, fast speech demonstrated a significantly reduced prominence pattern, with significantly reduced adjacent-vowel intensity, F0, and duration contrast. Therefore, it is possible that increased activation in anterior frontal operculum is associated with the increased processing demands of encoding a single, larger intonational phrase under fast speech conditions. Additional support for the hypothesis that the engagement of left anterior frontal operculum is driven by more abstract properties, as opposed to physical realizations, of utterances comes from the partial Pearson correlation analysis that demonstrated that anterior frontal operculum is not significantly correlated with any of the acoustic measures. Earlier neuroimaging studies have shown increased activity in this region in association with covert fast repeated monosyllabic word productions (Shergill et al., 2002) and fluent speech when compared to pauses (Kircher et al., 2004). Activation in this region has also been observed for generating lyrics and melodies (Kleber et al., 2007; Saito et al., 2006), and for processing syntax (Meyer et al., 2002; Ye and Zhou, 2009) and semantics (Poldrack et al., 1999; Rémy et al., 2003). The findings of the present study extend previous findings suggesting that left anterior frontal operculum may contribute to strengthening the links between sequential phonological units or to the size of the abstract phonological frame.




It is worth noting that while a number of neuroimaging studies have demonstrated increases in medial premotor regions in association with increases in rate (Forstmann et al., 2008; Ivanoff et al., 2008; Mechelli et al., 2000; Miura et al., 2003; Riecker et al., 2006; Riecker et al., 2005; van Veen et al., 2008), the present investigation did not replicate these earlier findings. Increased activity in right SMA was found in association with emphatic sentence production when compared to normal speech. It is well documented that SMA is part of a closed basal ganglia-thalamocortical loop sending projections to the putamen and receiving dense projections from the ventral lateral nucleus (Alexander et al., 1986; Haber and Calzavara, 2009) and that the basal ganglia contributes to the production prosody (e.g., Geiser et al., 2008; Hashimoto and Sakai, 2003; Kent and Rosenbek, 1982; Paus et al., 1996; Pell et al., 2006; Riecker et al., 2006; Riecker et al., 2005; Sidtis et al., 2003; Walsh and Smith, 2012). Interestingly, several neuroimaging studies have demonstrated a negative linear relationship with syllable repetition rate and activity in the putamen and caudate (Riecker et al., 2006; Riecker et al., 2005; Sidtis et al., 2003), leading to the inference that the recruitment of the striatum may contribute to syllable productions when there are long inter-stimulus intervals. Neuroimaging studies that have reported contributions from the caudate in speeded speech tasks (e.g., Hashimoto and Sakai, 2003) tend to involve more complex (i.e., non-repeated) syllable motor sequences.

The functional connectivity analysis demonstrated that the fast and normal speech conditions were significantly differentiable in terms of interactions between left middle premotor cortex and left pallidum. In particular, the significant negative correlation that

was found between left middle premotor cortex and left pallidum was greater under fast speech conditions than normal conditions. These results suggest that under fast speech conditions the premotor cortex reduces the basal ganglia's inhibitory control over motor areas and facilitates speech motor response selection earlier in time. More generally, the findings from the functional connectivity analysis did not indicate a significant coupling between left posterior dorsal inferior frontal gyrus pars opercularis and either the putamen or pallidum under fast speech conditions, providing additional evidence for a distinction in the functional recruitment of left middle premotor cortex and left posterior dorsal inferior frontal gyrus.

Central to the findings of the present study is the idea that the quality of speech motor output depends on multiple, potentially interacting, processing stages (cf., Graves et al., 2007). The results of the partial Pearson correlation analysis suggest hierarchically organized processing in the left prefrontal and premotor regions (summarized in Table 4 – 13). The progression of activity from left anterior frontal operculum to left posterior dorsal inferior frontal gyrus to left premotor cortex is correlated with an increasing number of the acoustic measures that characterize the acoustic realization of the sentences. These results indicate that distinct brain regions may contribute to different processing stages in speech production, an organization that may be essential for the hierarchical processing that is proposed to underlie speech production.

Table 4-13: Hierarchical organization of distinct prefrontal and premotor brain regions inferred from the partial Pearson correlation analysis.

	Left anterior frontal operculum	Left posterior dorsal inferior frontal gyrus	Left middle premotor cortex
Location of spherical ROI for the partial Pearson correlation analysis	L 	L 	L 
Summary of partial Pearson correlation results	No significant partial correlations with any of the acoustic measures tested	Positive partial correlation with average adjacent-vowel intensity contrast	Positive partial correlation with average vowel spacing, average vowel intensity, and average adjacent-vowel duration contrast.
Proposed processing stage	Phonological encoding of the abstract phonological frame	Phonological encoding of the links between phonological elements	Phonetic encoding of linguistic contrasts

In addition, the present study predicts that linguistic and prosodic variability may arise from difficulties in sequencing linguistic or prosodic contrasts or from difficulties encoding the spatio-temporal patterns of articulatory movements associated with linguistic or prosodic contrasts. These difficulties may present simultaneously, but they may also be entirely independent from one another (Buchwald and Miozzo, 2012). Critically, based on findings from the present study, the different stages of the speech production process would involve (at least) two distinct, but potentially interacting, brain regions: left posterior dorsal inferior frontal gyrus and left middle premotor cortex. The

implications of the present study resolve seemingly conflicting findings in the literature regarding the locus of cortical damage for individuals with acquired apraxia of speech (e.g., Dronkers, 1996; Hillis et al., 2004; Josephs et al., 2012; Josephs et al., 2006; Rohrer et al., 2010). For example, Josephs et al. (2012) argue that bilateral supplementary motor area and dorsolateral premotor cortex are the primary loci of cortical degeneration in individuals with acquired apraxia of speech, whereas others implicate left inferior frontal gyrus pars opercularis (Rohrer et al., 2010). The findings of the present study lead to the inference that the consequence of damage to left middle premotor cortex would result in less distinctive linguistic contrasts that arise from motor-based deficits in scaling articulatory gestures, whereas damage to posterior dorsal inferior frontal gyrus pars opercularis would result in less distinctive linguistic contrasts that arise from sequencing-based deficits in determining the phonological organization of phrases.

In summary, the results of the present study demonstrate that although both left posterior dorsal inferior frontal gyrus and middle premotor cortex contribute to the acoustic realization of phonological units, there is a critical functional distinction to be made between these two regions. Differential responses were demonstrated in the two regions depending on the degree of vowel quality and prosodic contrast produced. Having controlled for the effects of errors and movement time, we show that processing associated with making linguistic contrasts acoustically more distinct is associated with increased activity in left middle premotor cortex, whereas processing related to increasing the rate of speech is associated with increased activity in left posterior dorsal

inferior frontal gyrus pars opercularis. This distinction appears to generalize findings from recent fMRI studies of speech motor planning and production, suggesting that the differential prefrontal and premotor regions underlie a general distinction in speech processing in terms of phonological and phonetic encoding.

CHAPTER 5

CONCLUSIONS AND FUTURE DIRECTIONS

Central to most theories of speech production is the idea that the production of speech involves multiple, relatively distinct processing stages (cf., Binder et al., 2003) and while these stages may perform distinct functions, it is through their interaction that the acoustic realization of a speech sound is determined (Graves et al., 2007). At the early stages of encoding a phrase, fluent adult speakers must serially organize speech sounds on the basis of syntactic, prosodic, and semantic rules. Decisions within each of these categories affect the articulatory encoding of speech sounds. For example determining the speech tempo, like determining the phrasal stress pattern, affects the locations of clause and phrase boundaries, which in turn determine the extent to which a segment will be lengthened (Klatt, 1976). Semantic factors can also influence the duration of segments within an utterance. For example, segmental lengthening is observed for content words as compared to function words (MacKay, 1987). At the level of segmental encoding, in English, the decision to select one allophone over another is determined on the basis of the phoneme's location in the syllable and the word's stress (Fougeron, 1999). For instance, [p] and [p^h] are allophones of the phoneme /p/. The allophone [p^h] always occurs when the phoneme /p/ is in the syllable initial position and is followed by a stressed vowel (as in the word pin) whereas [p] is used in all other circumstances (as in the word spin). At the level of articulatory encoding, syllables that are high frequency (i.e., that have been encountered many times over a lifetime) will be produced faster than

low-frequency syllables, which in turn affects the global speech rate of utterances.

Collectively, the acoustic and neuroimaging results presented in Chapters 3 and 4 provide important advancements to our understanding of the way in which the speech sounds can be determined acoustically by different stages of speech planning and motor programming. The vast majority of neuroimaging studies on speech production have investigated the neural correlates of speech rate alone, without considering the changes in vowel quality or prosodic detail that accompanies increases in speech rate (e.g., Paus et al., 1996; Riecker et al., 2006; Riecker et al., 2005; Sidtis et al., 2003; Wildgruber et al., 2001). In addition, a large majority of these studies use non-lexical syllables (e.g., Riecker et al., 2006; Riecker et al., 2005; Wildgruber et al., 2001) with the number of syllables differing markedly across the rate conditions. In the present fMRI experiment, we controlled for the total number of speech syllables produced and used sentences containing real words.

However, several aspects of this research remain open to speculation and warrant further investigation. One of the difficulties in using real words as stimuli is the effect of phonological neighborhood density. A number of behavioral studies have demonstrated that increased phonological neighborhood density facilitates speech production (Gahl et al., 2012; Vitevitch, 2002). For example, Vitevitch (2002) used a picture naming paradigm and found that reaction times for naming pictures of words with high density neighborhoods were faster than those with low density neighborhoods. In a more recent study, Gahl et al. (2012) demonstrated that when produced in sentences, words with many phonological neighbors were produced with shorter durations and reduced vowel

spaces as compared to words with few phonological neighbors. Dell and Gordon (2003) account for the facilitation effect for words with high density neighborhoods by arguing that speech production and speech recognition differentially affect speech processing. For speech recognition, words with high phonological neighborhood densities activate a large number of competitors and so the target word's activation is relatively less differentiable from the activation of competing phonological neighbors. This effect contrasts with that of speech production, for which it is argued that the primary source of competition comes from semantically related words. In the face of competing semantically-related candidates, a word with a high phonological neighborhood density will be reinforced by the activation of the shared phonemes of phonological neighbors. Phonological neighborhood density was not a controlled variable in the present study and may have influenced our findings. If phonological neighborhood density affects the prosodic structure of utterances in the same way as fast speech does, we would expect that the effects of producing words with many phonological neighbors would be similar to the effects of producing speech quickly. Therefore, one possible avenue for further study is to compare the differential effects of phonological neighborhood density with the effects of clear and fast speech manners. In particular, speech stimuli could be divided in half, with one set being short sentences containing pseudowords and the other half containing real words. Since phonotactic frequency and phonological neighborhood density covary, pseudowords act as a reasonable control for the phonotactic frequency effects observed when using real words (Okada and Hickok, 2006). If the real word sentences and pseudo-word sentences are matched for phonotactic frequency, an effect

of phonological neighborhood density could be observed by comparing the production of words to the production of pseudowords. During the experiment speakers would be instructed to produce each sentence stimulus either clearly or quickly to examine how the effects of speech manner compare with the effects of phonological neighborhood density.

Further investigation of the neural correlates of prosodic words may also be worthwhile. For example, a study could be designed to localize regions that differentially process prosodic words, lexical words, and phonemes following the repetition suppression protocol of Peeva et al. (2010), which utilizes four experimental conditions that selectively repeat different features of the stimuli and result in different predicted patterns of neural activity across conditions. For example, in the *all-same* condition, subjects would repeat the same word across the trial (e.g., “auditor”). This would lead to the habituation of BOLD response in regions that are engaged for processing of phonemes, lexical words, and prosodic words. In the *different-lexical-and-prosodic-word* condition, stimuli that have the same phonemes, but differ in terms of the identity of the lexical and prosodic words (such as the words “an audit” (one prosodic word) vs. “audit ann” (two prosodic words)) would be produced repeatedly across the trial. In the *different-phoneme-and-prosodic-word* condition, stimuli with different phonemes and different prosodic words, but the same lexical words (e.g., “audit ‘em” (one prosodic word) vs. “audit them” (two prosodic words)) would be produced repeatedly across the trial. In the *all-different* condition, stimuli that differ in terms of the phonemes, lexical words, and prosodic words (e.g., “audit ann” and “audit ‘em”) would be produced

repeatedly across the trial. If a brain region supports processing of prosodic words, regardless of phonemic content, a region's response pattern would be most active during the *all-different* condition as well as during the *different-phoneme-and-prosodic-word* condition, but not during the *all-same* condition. If a brain region supports processing of lexical words a region's response pattern would be most active during the *different-lexical-and-prosodic-word* condition and during the *all-different* condition, but not during the *all same* or *different-phoneme-and-prosodic-word* conditions.

This dissertation provides one approach towards understanding the neural correlates associated with manipulating the acoustic detail of prosodic and linguistic contrasts. While it has helped shed some light on the differential recruitment of lateral premotor cortex and posterior dorsal inferior frontal gyrus pars opercularis, it has also raised additional questions regarding the underlying units of speech production and the relationship between phonological neighborhood density and speech manner. It is our hope that these questions inspire future related studies so that progress on understanding the different processing stages involved in speech production continues.

BIBLIOGRAPHY

- 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.
- Alexander, G.E., DeLong, M.R., Strick, P.L., 1986. Parallel organization of functionally segregated circuits linking basal ganglia and cortex. *Annual Reviews of Neuroscience* 9, 357-381.
- Aziz-Zadeh, L., Sheng, T., Gheyntanchi, A., 2010. Common premotor regions for the perception and production of prosody and correlations with empathy and prosodic ability. *PLoS ONE* 5, e8759.
- Belin, P., Zatorre, R.J., Hoge, R., Evans, A.C., Pike, B., 1999. Event-related fMRI of the auditory cortex. *NeuroImage* 10, 417-429.
- Binder, J.R., Frost, J.A., Hammeke, T.A., Bellgowan, P.S., Springer, J.A., Kaufman, J.N., Possing, E.T., 2000. Human temporal lobe activation by speech and nonspeech sounds. *Cerebral Cortex* 10, 512-528.
- Binder, J.R., McKiernan, K.A., Parsons, M.E., Westbury, C.F., Possing, E.T., Kaufman, J.N., Buchanan, L., 2003. Neural correlates of lexical access during visual word recognition. *Journal of Cognitive Neuroscience* 15, 372-393.
- Birn, R.M., Bandettini, P.A., Cox, R.W., Shaker, R., 1999. Event-related fMRI of tasks involving brief motion. *Human Brain Mapping* 7, 106-114.
- Boersma, P., 1993. Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. *Proceedings of the Institute of Phonetic Sciences, Amsterdam*, pp. 97-110.
- Boersma, P., 2001. Praat, a system for doing phonetics by computer. *Glott International* 5, 341-345.

- Bohland, J.W., Guenther, F.H., 2006. An fMRI investigation of syllable sequence production. *NeuroImage* 32, 821-841.
- Bradlow, A.R., Kraus, N., Hayes, E., 2003. Speaking clearly for children with learning disabilities: Sentence perception in noise. *Journal of Speech, Language, and Hearing Research* 46, 80-97.
- Brown, S., Laird, A.R., Pfordresher, P.Q., Thelen, S.M., Turkeltaub, P., Liotti, M., 2009. The somatotopy of speech: phonation and articulation in the human motor cortex. *Brain and Cognition* 70, 31-41.
- Brown, S., Martinez, M.J., Parsons, L.M., 2006. Music and language side by side in the brain: a PET study of the generation of melodies and sentences. *European Journal of Neuroscience* 23, 2791-2803.
- Brown, S., Ngan, E., Liotti, M., 2008. A larynx area in the human motor cortex. *Cerebral Cortex* 18, 837-845.
- Buccino, G., Binkofski, F., Fink, G.R., Fadiga, L., Fogassi, L., Gallese, V., Seitz, R.J., Zilles, K., Rizzolatti, G., Freund, H.J., 2001. Action observation activates premotor and parietal areas in a somatotopic manner: an fMRI study. *European Journal of Neuroscience* 13, 400-404.
- Buchwald, A., Miozzo, M., 2012. Phonological and motor errors in individuals with acquired sound production impairment. *Journal of Speech, Language, and Hearing Research* 55, S1573-S1586.
- Bunton, K., Leddy, M., 2011. An evaluation of articulatory working space area in vowel production of adults with Down syndrome. *Clinical Linguistics & Phonetics* 25, 321-334.
- Byrd, D., Tan, C.C., 1996. Saying consonant clusters quickly. *Journal of Phonetics* 24, 263-282.

- Callan, D.E., Tsytsarev, V., Hanakawa, T., Callan, A.M., Katsuhara, M., Fukuyama, H., Turner, R., 2006. Song and speech: brain regions involved with perception and covert production. *NeuroImage* 31, 1327-1342.
- Catani, M., Howard, R.J., Pajevic, S., Jones, D.K., 2002. Virtual in vivo interactive dissection of white matter fasciculi in the human brain. *NeuroImage* 17, 77-94.
- Catani, M., Jones, D.K., 2005. Perisylvian language networks of the human brain. *Annals of Neurology* 57, 8-16.
- Chen, F.R., 1980. Acoustic characteristics and intelligibility of clear and conversational speech at the segmental level. Massachusetts Institute of Technology.
- Chen, J.L., Penhune, V.B., Zatorre, R.J., 2008. Listening to musical rhythms recruits motor regions of the brain. *Cerebral Cortex* 18, 2844-2854.
- Collins, M., Rosenbek, J.C., Wertz, R.T., 1983. Spectrographic analysis of vowel and word duration in apraxia of speech. *Journal of Speech, Language, and Hearing Research* 26, 224-230.
- Cooper, W.E., Eady, S.J., Mueller, P.R., 1985. Acoustical aspects of contrastive stress in question-answer contexts. *The Journal of the Acoustical Society of America* 77, 2142-2156.
- Cooper, W.E., Soares, C., Ham, A., Damon, K., 1983. The influence of inter and intra speaker tempo on fundamental frequency and palatalization. *The Journal of the Acoustical Society of America* 73, 1723-1730.
- Costafreda, S.G., Fu, C.H.Y., Lee, L., Everitt, B., Brammer, M.J., David, A.S., 2006. A systematic review and quantitative appraisal of fMRI studies of verbal fluency: role of the left inferior frontal gyrus. *Human Brain Mapping* 27, 799-810.
- Dai, T.H., Liu, J.Z., Sahgal, V., Brown, R.W., Yue, G.H., 2001. Relationship between muscle output and functional MRI-measured brain activation. *Experimental Brain Research* 140, 290-300.

- Dale, A.M., Fischl, B., Sereno, M.I., 1999. Cortical surface-based analysis: I. Segmentation and surface reconstruction. *NeuroImage* 9, 179-194.
- Davis, M.H., Johnsrude, I.S., 2003. Hierarchical processing in spoken language comprehension. *The Journal of Neuroscience* 23, 3423-3431.
- de Jong, K.J., 1995. The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation. *The Journal of the Acoustical Society of America* 97, 491-504.
- de Jong, K.J., 2001. Rate-induced resyllabification revisited. *Language and Speech* 44, 197-216.
- de Jong, K.J., Beckman, M.E., Edwards, J., 1993. The interplay between prosodic structure and coarticulation. *Language and Speech* 36, 197-212.
- Deger, K., Ziegler, W., 2002. Speech motor programming in apraxia of speech. *Journal of Phonetics* 30, 321-335.
- Dell, G.S., Gordon, J.K., 2003. Neighbors in the lexicon: Friends or foes? *Phonetics and Phonology in Language Comprehension and Production: Differences and Similarities* 6, 9-37.
- Dettmers, C., Fink, G.R., Lemon, R.N., Stephan, K.M., Passingham, R.E., Silbersweig, D., Holmes, A., Ridding, M.C., Brooks, D.J., Frackowiak, R.S., 1995. Relation between cerebral activity and force in the motor areas of the human brain. *Journal of Neurophysiology* 74, 802-815.
- Dhankhar, A., Wexler, B.E., Fulbright, R.K., Halwes, T., Blamire, A.M., Shulman, R.G., 1997. Functional magnetic resonance imaging assessment of the human brain auditory cortex response to increasing word presentation rates. *Journal of Neurophysiology* 77, 476-483.

- di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., Rizzolatti, G., 1992. Understanding motor events: a neurophysiological study. *Experimental Brain Research* 91, 176-180.
- Dick, F., Lee, H.L., Nusbaum, H., Price, C.J., 2011. Auditory-motor expertise alters "speech selectivity" in professional musicians and actors. *Cerebral Cortex* 21, 938-948.
- Dromey, C., Ramig, L.O., 1998. Intentional changes in sound pressure level and rate: Their impact on measures of respiration, phonation, and articulation. *Journal of Speech, Language, and Hearing Research* 41, 1003-1018.
- Dronkers, N.F., 1996. A new brain region for coordinating speech articulation. *Nature* 384, 159-161.
- Eady, S.J., Cooper, W.E., 1986. Speech intonation and focus location in matched statements and questions. *The Journal of the Acoustical Society of America* 80, 402-415.
- Engstrand, O., 1988. Articulatory correlates of stress and speaking rate in Swedish VCV utterances. *The Journal of the Acoustical Society of America* 83, 1863-1875.
- Fernald, A., Simon, T., 1984. Expanded intonation contours in mothers' speech to newborns. *Developmental Psychology* 20, 104-113.
- Fernald, A., Taeschner, T., Dunn, J., Papousek, M., de Boysson-Bardies, B., Fukui, I., 1989. A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *Journal of Child Language* 16, 477-501.
- Fiebach, C.J., Friederici, A.D., Müller, K., von Cramon, D.Y., 2002. fMRI evidence for dual routes to the mental lexicon in visual word recognition. *Journal of Cognitive Neuroscience* 14, 11-23.

- Fiez, J.A., Balota, D.A., Raichle, M.E., Petersen, S.E., 1999. Effects of lexicality, frequency, and spelling-to-sound consistency on the functional anatomy of reading. *Neuron* 24, 205-218.
- Finnegan, E.M., Luschei, E.S., Hoffman, H.T., 2000. Modulations in respiratory and laryngeal activity associated with changes in vocal intensity during speech. *Journal of Speech, Language, and Hearing Research* 43, 934-950.
- Fischl, B., Salat, D.H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., van der Kouwe, A., Killiany, R., Kennedy, D., Klaveness, S., 2002. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron* 33, 341-355.
- 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.H., Dale, A.M., 1999b. High-resolution intersubject averaging and a coordinate system for the cortical surface. *Human Brain Mapping* 8, 272-284.
- Flege, J.E., 1988. Effects of speaking rate on tongue position and velocity of movement in vowel production. *The Journal of the Acoustical Society of America* 84, 901-916.
- Forstmann, B.U., Dutilh, G., Brown, S., Neumann, J., Von Cramon, D.Y., Ridderinkhof, K.R., Wagenmakers, E.J., 2008. Striatum and pre-SMA facilitate decision-making under time pressure. *Proceedings of the National Academy of Sciences of the United States of America* 105, 17538-17542.
- Fougeron, C., 1999. Prosodically conditioned articulatory variations: A review. *UCLA Working Papers in Phonetics*, 1-74.
- Fougeron, C., Jun, S.A., 1998. Rate effects on French intonation: Prosodic organization and phonetic realization. *Journal of Phonetics* 26, 45-69.

- Fourakis, M., 1991. Tempo, stress, and vowel reduction in American English. *The Journal of the Acoustical Society of America* 90, 1816-1827.
- Friederici, A.D., 2011. The brain basis of language processing: from structure to function. *Physiological Reviews* 91, 1357-1392.
- Friederici, A.D., Bahlmann, J., Heim, S., Schubotz, R.I., Anwander, A., 2006. The brain differentiates human and non-human grammars: functional localization and structural connectivity. *Proceedings of the National Academy of Sciences of the United States of America* 103, 2458-2463.
- Friederici, A.D., Meyer, M., von Cramon, D.Y., 2000. Auditory language comprehension: An event-related fMRI study on the processing of syntactic and lexical information. *Brain and Language* 74, 289-300.
- Friederici, A.D., Rüschemeyer, S.A., Hahne, A., Fiebach, C.J., 2003. The role of left inferior frontal and superior temporal cortex in sentence comprehension: localizing syntactic and semantic processes. *Cerebral Cortex* 13, 170-177.
- Fry, D.B., 1955. Duration and intensity as physical correlates of linguistic stress. *The Journal of the Acoustical Society of America* 27, 765-768.
- Gahl, S., Yao, Y., Johnson, K., 2012. Why reduce? Phonological neighborhood density and phonetic reduction in spontaneous speech. *Journal of Memory and Language* 66, 789-806.
- Gay, T., 1978. Effect of speaking rate on vowel formant movements. *The Journal of the Acoustical Society of America* 63, 223-230.
- Geiser, E., Zaehle, T., Jancke, L., Meyer, M., 2008. The neural correlate of speech rhythm as evidenced by metrical speech processing. *Journal of Cognitive Neuroscience* 20, 541-552.
- Gelfand, J.R., Bookheimer, S.Y., 2003. Dissociating neural mechanisms of temporal sequencing and processing phonemes. *Neuron* 38, 831-842.

- Gorgolewski, K., Burns, C.D., Madison, C., Clark, D., Halchenko, Y.O., Waskom, M.L., Ghosh, S.S., 2011. Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in python. *Frontiers in Neuroinformatics* 5, 1-5.
- 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, 617-631.
- Guenther, F.H., 1995. Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production. *Psychological Review* 102, 594-621.
- Haber, S.N., Calzavara, R., 2009. The cortico-basal ganglia integrative network: the role of the thalamus. *Brain Research Bulletin* 78, 69-74.
- Haley, K.L., Overton, H.B., 2001. Word length and vowel duration in apraxia of speech: The use of relative measures. *Brain and Language* 79, 397-406.
- Hall, D.A., Haggard, M.P., Akeroyd, M.A., Palmer, A.R., Summerfield, A.Q., Elliott, M.R., Gurney, E.M., Bowtell, R.W., 1999. "Sparse" temporal sampling in auditory fMRI. *Human Brain Mapping* 7, 213-223.
- Hardcastle, W.J., 1985. Some phonetic and syntactic constraints on lingual coarticulation during /kl/ sequences. *Speech Communication* 4, 247-263.
- Harris, K.S., 1978. Vowel duration change and its underlying physiological mechanisms. *Language and Speech* 21, 354-361.
- 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.
- Hauk, O., Johnsrude, I., Pulvermüller, F., 2004. Somatotopic representation of action words in human motor and premotor cortex. *Neuron* 41, 301-307.

- Hennenlotter, A., Schroeder, U., Erhard, P., Castrop, F., Haslinger, B., Stoecker, D., Lange, K.W., Ceballos-Baumann, A.O., 2005. A common neural basis for receptive and expressive communication of pleasant facial affect. *NeuroImage* 26, 581-591.
- Hickok, G., Poeppel, D., 2004. Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition* 92, 67-99.
- Hickok, G., Poeppel, D., 2007. The cortical organization of speech processing. *Nature Reviews Neuroscience* 8, 393-402.
- Hillis, A.E., Work, M., Barker, P.B., Jacobs, M.A., Breese, E.L., Maurer, K., 2004. Re-examining the brain regions crucial for orchestrating speech articulation. *Brain* 127, 1479-1487.
- Howseman, A.M., Josephs, O., Rees, G., Friston, K.J., 1997. Special issues in functional magnetic resonance imaging. SPM course, short course notes.
- Indefrey, P., Levelt, W.J., 2004. The spatial and temporal signatures of word production components. *Cognition* 92, 101-144.
- Ivanoff, J., Branning, P., Marois, R., 2008. fMRI evidence for a dual process account of the speed-accuracy tradeoff in decision-making. *PLoS ONE* 3, e2635.
- Jacks, A., 2009. Acoustic correlates of prosodic abnormality in acquired apraxia of speech. *Perspectives on Neurophysiology and Neurogenic Speech and Language Disorders* 19, 83-89.
- Jackson, O., Schacter, D.L., 2004. Encoding activity in anterior medial temporal lobe supports subsequent associative recognition. *NeuroImage* 21, 456-462.
- Jonides, J., Schumacher, E.H., Smith, E.E., Koeppe, R.A., Awh, E., Reuter-Lorenz, P.A., Marshuetz, C., Willis, C.R., 1998. The role of parietal cortex in verbal working memory. *Journal of Neuroscience* 18, 5026-5034.

- Josephs, K.A., Duffy, J.R., Strand, E.A., Machulda, M.M., Senjem, M.L., Master, A.V., Lowe, V.J., Jack Jr, C.R., Whitwell, J.L., 2012. Characterizing a neurodegenerative syndrome: primary progressive apraxia of speech. *Brain* 135, 1522-1536.
- Josephs, K.A., Duffy, J.R., Strand, E.A., Whitwell, J.L., Layton, K.F., Parisi, J.E., Hauser, M.F., Witte, R.J., Boeve, B.F., Knopman, D.S., 2006. Clinicopathological and imaging correlates of progressive aphasia and apraxia of speech. *Brain* 129, 1385-1398.
- Katz, J., Selkirk, E., 2011. Contrastive focus vs. discourse-new: Evidence from phonetic prominence in English. *Language* 87, 771-816.
- Keating, P., Shattuck-Hufnagel, S., 2002. A prosodic view of word form encoding for speech production. *UCLA Working Papers in Phonetics*, 112-156.
- Kent, R.D., Netsell, R., 1971. Effects of stress contrasts on certain articulatory parameters. *Phonetica* 24, 23-44.
- Kent, R.D., Read, C., 2002. *The acoustic analysis of speech*. Singular San Diego, CA.
- Kent, R.D., Rosenbek, J.C., 1982. Prosodic disturbance and neurologic lesion. *Brain and Language* 15, 259-291.
- Kircher, T.T.J., Brammer, M.J., Levelt, W., Bartels, M., McGuire, P.K., 2004. Pausing for thought: engagement of left temporal cortex during pauses in speech. *NeuroImage* 21, 84-90.
- Klatt, D.H., 1976. Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *The Journal of the Acoustical Society of America* 59, 1208-1221.
- Kleber, B., Birbaumer, N., Veit, R., Trevorrow, T., Lotze, M., 2007. Overt and imagined singing of an Italian aria. *NeuroImage* 36, 889-900.

- Klein, E., Domahs, U., Grande, M., Domahs, F., 2011. Neuro-cognitive foundations of word stress processing-evidence from fMRI. *Behavioral and Brain Functions* 7, 15-31.
- Koelsch, S., Gunter, T.C., Cramon, D.Y., Zysset, S., Lohmann, G., Friederici, A.D., 2002. Bach speaks: A cortical "language-network" serves the processing of music. *NeuroImage* 17, 956-966.
- Kondo, H.M., Kashino, M., 2009. Involvement of the thalamocortical loop in the spontaneous switching of percepts in auditory streaming. *The Journal of Neuroscience* 29, 12695-12701.
- Krause, J.C., Braida, L.D., 2004. Acoustic properties of naturally produced clear speech at normal speaking rates. *The Journal of the Acoustical Society of America* 115, 362-378.
- Kuehn, D.P., Moll, K.L., 1976. A cineradiographic study of VC and CV articulatory velocities. *Journal of Phonetics* 4, 303-320.
- Kuhl, P.K., Andruski, J.E., Chistovich, I.A., Chistovich, L.A., Kozhevnikova, E.V., Ryskina, V.L., Stolyarova, E.I., Sundberg, U., Lacerda, F., 1997. Cross-language analysis of phonetic units in language addressed to infants. *Science* 277, 684-686.
- Lancaster, J.L., Tordesillas-Gutierrez, D., Martinez, M., Salinas, F., Evans, A., Zilles, K., Mazziotta, J.C., Fox, P.T., 2007. Bias between MNI and Talairach coordinates analyzed using the ICBM-152 brain template. *Human Brain Mapping* 28, 1194-1205.
- Lehiste, I., Peterson, G.E., 1959. Vowel amplitude and phonemic stress in American English. *The Journal of the Acoustical Society of America* 31, 428-435.
- Lindblom, B., 1963. Spectrographic study of vowel reduction. *The Journal of the Acoustical Society of America* 35, 1773-1781.

- Lindblom, B., 1964. Articulatory activity in vowels. *The Journal of the Acoustical Society of America* 36, 1038-1038.
- Lindblom, B., 1990. Explaining phonetic variation: A sketch of the H&H theory. In: Hardcastle, W., Marchal, A. (Eds.), *Speech Production and Speech Modeling*. Kluwer, Dordrecht, pp. 403–439.
- Lindblom, B., Agwuele, A., Sussman, H.M., Cortes, E.E., 2007. The effect of emphatic stress on consonant vowel coarticulation. *The Journal of the Acoustical Society of America* 121, 3802-3813.
- Liu, H.M., Kuhl, P.K., Tsao, F.M., 2003. An association between mothers' speech clarity and infants' speech discrimination skills. *Developmental Science* 6, F1-F10.
- Loevenbruck, H., Baciú, M., Segebarth, C., Abry, C., 2005. The left inferior frontal gyrus under focus: an fMRI study of the production of deixis via syntactic extraction and prosodic focus. *Journal of Neurolinguistics* 18, 237-258.
- MacKay, D.G., 1987. *The organization of perception and action: A theory for language and other cognitive skills*. Springer-Verlag New York, Inc.
- Marrelec, G., Horwitz, B., Kim, J., Pélérini-Issac, M., Benali, H., Doyon, J., 2007. Using partial correlation to enhance structural equation modeling of functional MRI data. *Magnetic Resonance Imaging* 25, 1181-1189.
- Marrelec, G., Kim, J., Doyon, J., Horwitz, B., 2008. Large-scale neural model validation of partial correlation analysis for effective connectivity investigation in functional MRI. *Human Brain Mapping* 30, 941-950.
- Marrelec, G., Krainik, A., Duffau, H., Pélérini-Issac, M., Lehericy, S., Doyon, J., Benali, H., 2006. Partial correlation for functional brain interactivity investigation in functional MRI. *NeuroImage* 32, 228-237.

- Matthies, M., Perrier, P., Perkell, J.S., Zandipour, M., 2001. Variation in anticipatory coarticulation with changes in clarity and rate. *Journal of Speech, Language, and Hearing Research* 44, 340-353.
- Mayer, J., Wildgruber, D., Riecker, A., Dogil, G., Ackermann, H., Grodd, W., 2002. Prosody production and perception: converging evidence from fMRI studies. *Speech Prosody 2002, International Conference, Aix-en-Provence, France*, pp. 487-490.
- McClean, M.D., Tasko, S.M., 2002. Association of orofacial with laryngeal and respiratory motor output during speech. *Experimental Brain Research* 146, 481-489.
- McClean, M.D., Tasko, S.M., 2003. Association of orofacial muscle activity and movement during changes in speech rate and intensity. *Journal of Speech, Language, and Hearing Research* 46, 1387-1400.
- McDermott, K.B., Petersen, S.E., Watson, J.M., Ojemann, J.G., 2003. A procedure for identifying regions preferentially activated by attention to semantic and phonological relations using functional magnetic resonance imaging. *Neuropsychologia* 41, 293-303.
- McRae, P.A., Tjaden, K., Schoonings, B., 2002. Acoustic and perceptual consequences of articulatory rate change in Parkinson disease. *Journal of Speech, Language, and Hearing Research* 45, 35-50.
- Mechelli, A., Friston, K.J., Price, C.J., 2000. The effects of presentation rate during word and pseudoword reading: a comparison of PET and fMRI. *Journal of Cognitive Neuroscience* 12, 145-156.
- Ménard, L., Polak, M., Denny, M., Burton, E., Lane, H., Matthies, M.L., Marrone, N., Perkell, J.S., Tiede, M., Vick, J., 2007. Interactions of speaking condition and auditory feedback on vowel production in postlingually deaf adults with cochlear implants. *The Journal of the Acoustical Society of America* 121, 3790-3801.

- Meyer, M., Alter, K., Friederici, A.D., Lohmann, G., von Cramon, D.Y., 2002. FMRI reveals brain regions mediating slow prosodic modulations in spoken sentences. *Human Brain Mapping* 17, 73-88.
- Miura, N., Iwata, K., Watanabe, J., Sugiura, M., Akitsuki, Y., Sassa, Y., Ikuta, N., Okamoto, H., Watanabe, Y., Riera, J., Maeda, Y., Matsue, Y., Kawashima, R., 2003. Cortical activation during reading aloud of long sentences: fMRI study. *NeuroReport* 14, 1563-1566.
- Moon, S.J., Lindblom, B., 1994. Interaction between duration, context, and speaking style in English stressed vowels. *The Journal of the Acoustical Society of America* 96, 40-55.
- Moser, D., Baker, J.M., Sanchez, C.E., Rorden, C., Fridriksson, J., 2009. Temporal order processing of syllables in the left parietal lobe. *The Journal of Neuroscience* 29, 12568-12573.
- Munhall, K., Löfqvist, A., 1992. Gestural aggregation in speech: laryngeal gestures. *Journal of Phonetics* 20, 111-126.
- Neel, A.T., 2008. Vowel space characteristics and vowel identification accuracy. *Journal of Speech, Language, and Hearing Research* 51, 574-585.
- Obleser, J., Kotz, S.A., 2010. Expectancy constraints in degraded speech modulate the language comprehension network. *Cerebral Cortex* 20, 633-640.
- Okada, K., Hickok, G., 2006. Identification of lexical-phonological networks in the superior temporal sulcus using functional magnetic resonance imaging. *NeuroReport* 17, 1293-1296.
- Okada, K., Rong, F., Venezia, J., Matchin, W., Hsieh, I.H., Saberi, K., Serences, J.T., Hickok, G., 2010. Hierarchical organization of human auditory cortex: evidence from acoustic invariance in the response to intelligible speech. *Cerebral Cortex* 20, 2486-2495.

- Olthoff, A., Baudewig, J., Kruse, E., Dechent, P., 2008. Cortical sensorimotor control in vocalization: a functional magnetic resonance imaging study. *Laryngoscope* 118, 2091-2096.
- Ostry, D.J., Munhall, K., Parush, A., 1983. Control of rate and movement duration in speech. *The Journal of the Acoustical Society of America* 74, S116-S116.
- Papoutsis, M., De Zwart, J.A., Jansma, J.M., Pickering, M.J., Bednar, J.A., Horwitz, B., 2009. From phonemes to articulatory codes: an fMRI study of the role of Broca's area in speech production. *Cerebral Cortex* 19, 2156-2165.
- Patel, R., Niziolek, C., Reilly, K., Guenther, F.H., 2011. Prosodic adaptations to pitch perturbation in running speech. *Journal of Speech, Language, and Hearing Research* 54, 1051-1059.
- Paulesu, E., Frith, C.D., Frackowiak, R.S., 1993. The neural correlates of the verbal component of working memory. *Nature* 362, 342-345.
- Paus, T., Perry, D.W., Zatorre, R.J., Worsley, K.J., Evans, A.C., 1996. Modulation of cerebral blood flow in the human auditory cortex during speech: role of motor-to-sensory discharges. *European Journal of Neuroscience* 8, 2236-2246.
- Peeva, M.G., Guenther, F.H., Tourville, J.A., Nieto-Castañón, A., Anton, J.L., Nazarian, B., Alario, F.X., 2010. Distinct representations of phonemes, syllables, and supra-syllabic sequences in the speech production network. *NeuroImage* 50, 626-638.
- Pell, M.D., Cheang, H.S., Leonard, C.L., 2006. The impact of Parkinson's disease on vocal-prosodic communication from the perspective of listeners. *Brain and Language* 97, 123-134.
- Peramunage, D., Blumstein, S.E., Myers, E.B., Goldrick, M., Baese-Berk, M., 2011. Phonological neighborhood effects in spoken word production: An fMRI study. *Journal of Cognitive Neuroscience* 23, 593-603.

- Perkell, J.S., Zandipour, M., Matthies, M.L., Lane, H., 2002. Economy of effort in different speaking conditions. I. A preliminary study of intersubject differences and modeling issues. *The Journal of the Acoustical Society of America* 112, 1627-1641.
- Perrone, M., Dohen, M., Loevenbruck, H., Sato, M., Pichat, C., Yvert, G., Baci, M., 2010. An fMRI study of the perception of contrastive prosodic focus in French. *Speech Prosody*, Chicago, Illinois.
- Peterson, G.E., Lehiste, I., 1960. Duration of syllable nuclei in English. *Journal of the Acoustical Society of America* 32, 693-703.
- Picheny, M.A., Durlach, N.I., Braida, L.D., 1985. Speaking clearly for the hard of hearing I: Intelligibility differences between clear and conversational speech. *Journal of Speech and Hearing Research* 28, 96-103.
- Picheny, M.A., Durlach, N.I., Braida, L.D., 1986. Speaking clearly for the hard of hearing. II: Acoustic characteristics of clear and conversational speech. *Journal of Speech and Hearing Research* 29, 434-446.
- Poeppel, D., 2003. The analysis of speech in different temporal integration windows: cerebral lateralization as 'asymmetric sampling in time'. *Speech Communication* 41, 245-255.
- Poldrack, R.A., Wagner, A.D., Prull, M.W., Desmond, J.E., Glover, G.H., Gabrieli, J.D., 1999. Functional specialization for semantic and phonological processing in the left inferior prefrontal cortex. *NeuroImage* 10, 15-35.
- Price, C.J., Moore, C.J., Humphreys, G.W., Wise, R.J.S., 1997. Segregating semantic from phonological processes during reading. *Journal of Cognitive Neuroscience* 9, 727-733.
- Rémy, F., Frankenstein, U.N., Mincic, A., Tomanek, B., Stroman, P.W., 2003. Pain modulates cerebral activity during cognitive performance. *NeuroImage* 19, 655-664.

- Riecker, A., Brendel, B., Ziegler, W., Erb, M., Ackermann, H., 2008. The influence of syllable onset complexity and syllable frequency on speech motor control. *Brain and Language* 107, 102-113.
- Riecker, A., Kassubek, J., Groschel, K., Grodd, W., Ackermann, H., 2006. The cerebral control of speech tempo: opposite relationship between speaking rate and BOLD signal changes at striatal and cerebellar structures. *NeuroImage* 29, 46-53.
- Riecker, A., Mathiak, K., Wildgruber, D., Erb, M., Hertrich, I., Grodd, W., Ackermann, H., 2005. fMRI reveals two distinct cerebral networks subserving speech motor control. *Neurology* 64, 700-706.
- Rogalsky, C., Hickok, G., 2011. The role of Broca's area in sentence comprehension. *Journal of Cognitive Neuroscience* 23, 1664-1680.
- Rohrer, J.D., Rossor, M.N., Warren, J.D., 2010. Apraxia in progressive nonfluent aphasia. *Journal of Neurology* 257, 569-574.
- Saito, Y., Ishii, K., Yagi, K., Tatsumi, I.F., Mizusawa, H., 2006. Cerebral networks for spontaneous and synchronized singing and speaking. *NeuroReport* 17, 1893-1897.
- Sato, M., Baciú, M., Loevenbruck, H., Schwartz, J.L., Cathiard, M.A., Segebarth, C., Abry, C., 2004. Multistable representation of speech forms: A functional MRI study of verbal transformations. *NeuroImage* 23, 1143-1151.
- Saur, D., Kreher, B.W., Schnell, S., Kümmerer, D., Kellmeyer, P., Vry, M.S., Umarova, R., Musso, M., Glauche, V., Abel, S., 2008. Ventral and dorsal pathways for language. *Proceedings of the National Academy of Sciences* 105, 18035-18040.
- Schulman, R., 1989. Articulatory dynamics of loud and normal speech. *The Journal of the Acoustical Society of America* 85, 295-312.
- Scott, S.K., Blank, C.C., Rosen, S., Wise, R.J., 2000. Identification of a pathway for intelligible speech in the left temporal lobe. *Brain* 123 Pt 12, 2400-2406.

- Seidler, R.D., Noll, D.C., Thiers, G., 2004. Feedforward and feedback processes in motor control. *NeuroImage* 22, 1775-1783.
- Shergill, S.S., Brammer, M.J., Fukuda, R., Bullmore, E., Amaro, E., Jr., Murray, R.M., McGuire, P.K., 2002. Modulation of activity in temporal cortex during generation of inner speech. *Human Brain Mapping* 16, 219-227.
- Sidtis, J.J., Gomez, C., Groshong, A., Strother, S.C., Rottenberg, D.A., 2006. Mapping cerebral blood flow during speech production in hereditary ataxia. *NeuroImage* 31, 246-254.
- Sidtis, J.J., Strother, S.C., Rottenberg, D.A., 2003. Predicting performance from functional imaging data: methods matter. *NeuroImage* 20, 615-624.
- Sörös, P., MacIntosh, B.J., Tam, F., Graham, S.J., 2010. fMRI-compatible registration of jaw movements using a fiber-optic bend sensor. *Frontiers in Human Neuroscience* 4:24. doi: 10.3389/fnhum.2010.00024.
- Stålhammar, U., Karlsson, I., Fant, G., 1973. Contextual effects on vowel nuclei. *KTH Speech Transmission Laboratory Quarterly Progress and Status Report* 4, 1-18.
- Stathopoulos, E.T., Sapienza, C., 1993. Respiratory and laryngeal function of women and men during vocal intensity variation. *Journal of Speech, Language, and Hearing Research* 36, 64-75.
- Sussman, H.M., MacNeilage, P.F., 1978. Motor unit correlates of stress: Preliminary observations. *The Journal of the Acoustical Society of America* 64, 338-340.
- 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.

- Tasko, S.M., Greilick, K., 2010. Acoustic and articulatory features of diphthong production: a speech clarity study. *Journal of Speech, Language, and Hearing Research* 53, 84-99.
- Terken, J., 1991. Fundamental frequency and perceived prominence of accented syllables. *The Journal of the Acoustical Society of America* 89, 1768-1776.
- Tourville, J.A., Reilly, K.J., Guenther, F.H., 2008. Neural mechanisms underlying auditory feedback control of speech. *NeuroImage* 39, 1429-1443.
- Tremblay, P., Small, S.L., 2011. On the context-dependent nature of the contribution of the ventral premotor cortex to speech perception. *NeuroImage* 57, 1561-1571.
- Tsai, C.G., Chen, C.C., Chou, T.L., Chen, J.H., 2010. Neural mechanisms involved in the oral representation of percussion music: An fMRI study. *Brain and Cognition* 74, 123-131.
- Tuller, B., Harris, K.S., Kelso, J.A.S., 1982. Stress and rate: Differential transformations of articulation. *Journal of the Acoustical Society of America* 71, 1534-1543.
- Turner, G.S., Tjaden, K., Weismer, G., 1995. The influence of speaking rate on vowel space and speech intelligibility for individuals with amyotrophic lateral sclerosis. *Journal of Speech, Language, and Hearing Research* 38, 1001-1013.
- Turner, G.S., Weismer, G., 1993. Characteristics of speaking rate in the dysarthria associated with amyotrophic lateral sclerosis. *Journal of Speech, Language, and Hearing Research* 36, 1134-1344.
- van Duinen, H., Renken, R., Maurits, N.M., Zijdewind, I., 2007. Relation between muscle and brain activity during isometric contractions of the first dorsal interosseus muscle. *Human Brain Mapping* 29, 281-299.
- Van Lancker, D., Sidtis, J.J., 1992. The identification of affective-prosodic stimuli by left-and right-hemisphere-damaged subjects: all errors are not created equal. *Journal of Speech, Language, and Hearing Research* 35, 963-970.

- van Veen, V., Krug, M.K., Carter, C.S., 2008. The neural and computational basis of controlled speed-accuracy tradeoff during task performance. *Journal of Cognitive Neuroscience* 20, 1952-1965.
- Vitevitch, M.S., 2002. The influence of phonological similarity neighborhoods on speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 28, 735-747.
- Walsh, B., Smith, A., 2012. Basic parameters of articulatory movements and acoustics in individuals with Parkinson's disease. *Movement Disorders* 27, 843-850.
- Weismer, G., Ingrisano, D., 1979. Phrase-level timing patterns in English: Effects of emphatic stress location and speaking rate. *Journal of Speech, Language, and Hearing Research* 22, 516-533.
- Whalen, D.H., Levitt, A.G., 1995. The universality of intrinsic F0 of vowels. *Journal of Phonetics* 23, 349-366.
- Wildgruber, D., Ackermann, H., Grodd, W., 2001. Differential contributions of motor cortex, basal ganglia, and cerebellum to speech motor control: effects of syllable repetition rate evaluated by fMRI. *NeuroImage* 13, 101-109.
- Wilson, S.M., Galantucci, S., Tartaglia, M.C., Rising, K., Patterson, D.K., Henry, M.L., Ogar, J.M., DeLeon, J., Miller, B.L., Gorno-Tempini, M.L., 2011. Syntactic processing depends on dorsal language tracts. *Neuron* 72, 397-403.
- Wise, R.J., Greene, J., Büchel, C., Scott, S.K., 1999. Brain regions involved in articulation. *Lancet* 353, 1057-1061.
- Wohlert, A.B., Hammen, V.L., 2000. Lip muscle activity related to speech rate and loudness. *Journal of Speech, Language, and Hearing Research* 43, 1229-1239.
- Xu, Y., Xu, C.X., 2005. Phonetic realization of focus in English declarative intonation. *Journal of Phonetics* 33, 159-197.

- Yang, Y., Engelien, A., Engelien, W., Xu, S., Stern, E., Silbersweig, D.A., 2000. A silent event-related functional MRI technique for brain activation studies without interference of scanner acoustic noise. *Magnetic Resonance in Medicine* 43, 185-190.
- Ye, Z., Zhou, X., 2009. Conflict control during sentence comprehension: fMRI evidence. *NeuroImage* 48, 280-290.

CURRICULUM VITAE

Elisa Golfinopoulos

677 Beacon Street Boston, MA 02215

Tel: (617) 353-6181 (office); e-mail: egolfino@bu.edu

Date of Birth: April 17, 1981

EDUCATION

- | | |
|-------------|---|
| 2005 – 2013 | PhD. Cognitive and Neural Systems
Boston University, Boston, MA |
| 1999 – 2003 | B.A. with departmental honors in Mathematics
Hamilton College, Clinton, NY |

EXPERIENCE

- | | |
|-------------|---|
| 2005 – 2013 | Research Assistant for Dr. Frank H. Guenther
Cognitive & Neural Systems, Boston University, Boston, MA |
| Spring 2007 | Teaching Fellow for Dr. Daniel H. Bullock
Cognitive & Neural Systems, Boston University, Boston, MA |
| 2003 – 2005 | Research Analyst I
The Lewin Group, Falls Church, VA |
| Summer 2002 | Research Assistant for Dr. Larry Knop
Mathematics Department, Hamilton College, Clinton NY |
| 2001 – 2003 | Teaching Assistant for Dr. Sally Cockburn
Mathematics Department, Hamilton College, Clinton NY |

AWARDS

- 2005 – 2013 Boston University Graduate Research Fellowship
- 2003 *Phi Beta Kappa* National Honor Society
- 2003 Alexander Hamilton's List Recognizing achievements in writing
- 2002 Award for Most Outstanding Student Presenter at MathFest
- 1999 – 2003 Philadelphia University Alumnae Book Award

PUBLICATIONS

Golfinopoulos E, Tourville JA, Bohland JW, Ghosh SS, Guenther FH, (2011). fMRI investigation of unexpected somatosensory feedback perturbation during speech.

Golfinopoulos E, Tourville JA, Guenther FH, (2010). The integration of large-scale neural network modeling and functional magnetic resonance imaging in speech production. *NeuroImage* 52:862-874.

Golfinopoulos E, Guenther FH (2011). Right motor cortical regions mediate the production of prominent words in English Sentences. 2011 Human Brain Mapping Meeting in Quebec City, Canada

Golfinopoulos E, Guenther FH (2009). An fMRI investigation of the neural bases of suprasegmental timing in speech production. 2009 Society for Neuroscience Meeting in Chicago, Illinois.

PROFESSIONAL MEMBERSHIPS AND SERVICE

Student Member, The Society for Neuroscience