2021

Internal and external speech timing mechanisms in persistent developmental stuttering

BOSTON UNIVERSITY

SARGENT COLLEGE OF HEALTH AND REHABILITATION SCIENCES

Dissertation

INTERNAL AND EXTERNAL SPEECH TIMING MECHANISMS IN PERSISTENT DEVELOPMENTAL STUTTERING

by

SAUL ALEXANDER FRANKFORD

B.A., Northwestern University, 2013

Submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

2021

Approved by

First Reader

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

Second Reader

Cara E. Stepp, Ph.D.
Professor of Speech, Language, & Hearing Sciences
Professor of Biomedical Engineering
Professor of Otolaryngology – Head and Neck Surgery

Third Reader

Daniel Bullock, Ph.D.
Professor *Emeritus* of Psychological and Brain Sciences

Fourth Reader

Soo-Eun Chang, Ph.D.
Rosa Casco Solano-Lopez Research Professor of Child and Adolescent Psychiatry
Associate Professor of Psychiatry
University of Michigan

DEDICATION

I would like to dedicate this work to my wonderful wife and life-partner, Shira, and my beautiful son, Henry.

ACKNOWLEDGMENTS

This dissertation could not have been possible without the collaboration and support of so many. First, I would like to acknowledge the support and guidance of my Ph.D. mentor, Frank Guenther, who graciously allowed me to join his lab. I always relished the opportunity to dig deeper into his expansive knowledge of neural systems and theoretical ideas about the neural control of speech production.

Next, I must thank Jason Tourville and Alfonso Nieto-Castañón, from whom I learned so much about neuroimaging and statistical analysis, as well as the other collaborators on these studies: Shanqing Cai, Elizabeth Heller Murray, Matthew Masapollo, Ina Jessen-Groeschel, Lauren Jiang, Katalina Aguilar, Brittany Steinfeld, and Mona Tong. Many additional thanks to other members of the Guenther Lab: Jennifer Segawa, Ayoub Daliri, Matthias Heyne, Elaine Kearney, Megan Thompson, Hilary Miller, and Ricky Falsini. Special acknowledgement must also go to Barbara Holland, who works tirelessly "behind the scenes" so that everything in our lab runs smoothly.

I would also like to thank the staffs at the BU Cognitive Neuroimaging Center, including Jay Bohland and Shruthi Chakrapani, and the MGH Athinoula A. Martinos Center for Biomedical Imaging for their support with MRI data collection.

I am grateful to Diane Constantino for allowing me to participate in activities at the BU Center for Stuttering Therapy and for her help with recruitment.

Additional thanks go to my dissertation committee members — Soo-Eun Chang, Cara Stepp, and Dan Bullock — for their guidance throughout the dissertation process, and members of the Department of Speech, Language, & Hearing Sciences for their help

during my doctoral studies. Notable among them include Swathi Kiran, Melanie Matthies, Michelle Mentis, Ada Kwok, Mary McDonough, Katie Brown, and Julianne Leber.

I would like to acknowledge my family and friends for their constant support throughout this process. Special thanks go to Gabriel Cler for many caffeine-filled hours of moral and writing support.

Finally, I am eternally grateful for the love and support of my wonderful wife, Shira, and my son, Henry. There is no one else I would have rather had by my side throughout this whole process.

INTERNAL AND EXTERNAL SPEECH TIMING MECHANISMS IN PERSISTENT DEVELOPMENTAL STUTTERING

SAUL ALEXANDER FRANKFORD

Boston University Sargent College of Health and Rehabilitation Sciences, 2021

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

ABSTRACT

Stuttering is a developmental speech disorder characterized by interruptions of fluency. A large body of research suggests that stuttering occurs due to a reduced ability to generate timing signals in order to sequence speech sounds. One piece of supporting evidence for this is that when speaking along with an external timing source like a metronome, disfluencies suddenly and significantly decrease. The aim of this dissertation was to characterize the effects of using auditory cues to time speech on neural activation and auditory feedback processing, and how these effects may contribute to fluency in adults who stutter (AWS).

Two studies were carried out to examine these effects. In the first study, functional magnetic resonance imaging was used to measure brain activity while AWS and adults who do not stutter (ANS) read sentences aloud either using natural speech timing or aligning each syllable to the beat of a metronome. Consistent with previous literature, AWS produced fewer disfluent trials in the externally paced condition than in the normal condition. Collapsing across the AWS and ANS groups, participants had greater activation in the metronome-timed condition in regions associated with speech sequencing, sensory feedback control, and timing perception. AWS also demonstrated

increased functional connectivity among cerebellar regions during externally paced speech.

In the second study, responses to online spectral and timing perturbations of auditory feedback were measured while AWS and ANS read sentences with and without metronome pacing. Results indicated that AWS showed no responses to spectral perturbations during the non-paced condition and significant compensatory responses during the paced condition along with fewer disfluencies, while responses in ANS showed the opposite effect. For the timing perturbation, no significant differences were found between groups in either condition.

Together, these studies indicate that the deficit in stuttering is related to spectral processing rather than purely temporal processing, and that externally paced speech recruits compensatory neural regions that may help resolve this deficit.

TABLE OF CONTENTS

DEDICATION iv

ACKNOWLEDGMENTS v

ABSTRACT vii

TABLE OF CONTENTS ix

LIST OF TABLES xiii

LIST OF FIGURES xvi

LIST OF ABBREVIATIONS xxv

CHAPTER I: INTRODUCTION 1

Speech Motor Control 1

Investigations of Auditory Feedback in Speech Production 3

Vowel Formant Perturbations 4

The DIVA Model of Speech Production 7

Stuttering 9

General Background 9

Neural Correlates of Stuttering 10

Evidence for Disrupted Timing Abilities in Stuttering 12

Sensory-Motor Integration Disturbance in Stuttering 22

Summary of Dissertation 25

CHAPTER II: The Neural Circuitry Underlying the “Rhythm Effect” in Stuttering 27

Abstract 27

Introduction 29

Methods 32
Subjects 32
fMRI Paradigm 34
Data Acquisition 36
Behavioral Analysis 38
Task Activation fMRI Analysis 39
Functional Connectivity Analysis 44
Results 47
Behavioral Analysis 47
Task Activation fMRI Analysis 49
Brain-Behavior Correlation Analyses 52
Functional Connectivity Analyses 52
Discussion 60
A Possible Compensatory Role for the Cerebellum in AWS 61
Changes in Activation during Isochronous Speech 64
Correlation Between Activation and Severity 69
Limitations 72
Conclusion 75
Supplementary Materials 77
CHAPTER III: Responses to auditory feedback perturbations in persons who stutter during syllable-timed speech. 102
Introduction 102

Methods and Materials 106
Participants 106
Experimental Setup 107
Stimuli 107
Procedure 108
Focal Perturbations 110
Analyses 114
Results 121
Disfluency Rate 121
Speaking Rate and Rhythmicity 122
Formant Perturbation 123
Timing Perturbation 125
Discussion 128
Auditory feedback timing control in AWS 129
The effect of syllable-timed speech on auditory feedback control 132
Additional considerations 134
Conclusion 136
Supplementary Materials 137
CHAPTER IV: Conclusions 138
Future Directions 139
Extending This Work to CWS 139
Clarifying the Nature of the Rhythm Effect in Stuttering 140

Stuttering Subtypes 142

APPENDIX A 143

APPENDIX B 144

APPENDIX C 145

BIBLIOGRAPHY 147

CURRICULUM VITAE 169

LIST OF TABLES

Table 2.1. Demographic and stuttering severity data from adults who stutter. F = female; M = male; SSI-4 = Stuttering Severity Index – Fourth Edition. SSI-Mod = a modified version of the SSI-4 that does not include a subscore related to concomitant movements. Disfluency Rate = the percent of trials containing disfluencies during the *normal* speech condition. 33

Table 2.2. Descriptive and inferential statistics for speaking rate and CV-IVI. Error estimates indicate 95% confidence intervals. Significant effects are highlighted in bold. 48

Table 2.3. Cortical clusters with activation differences between the *rhythm* and *normal* conditions collapsed across groups (vertex-wise $p < .01$, cluster-wise $p_{FDR} < .05$). MNI = Montreal Neurological Institute, FDR = false discovery rate, adPMC = anterior dorsal premotor cortex, AG = angular gyrus, aSMg = supramarginal gyrus, dMC = dorsal primary motor cortex, mdPMC = middle dorsal premotor cortex, OC = occipital cortex, pMFg = posterior middle frontal gyrus, PO = parietal operculum, preSMA = presupplementary motor area, pSMg = posterior supramarginal gyrus, PT = planum temporale, SMA = supplementary motor area, SPL = superior parietal lobule. 51

Table 2.4. Functional connectivity analysis - condition and interaction effects. Roman numerals indicate cerebellar lobules. ROI = region-of-interest, MNI = Montreal Neurological Institute, FDR = false discovery rate, L = left, R = right, ACC = anterior cingulate cortex, AG = angular gyrus, aINS = anterior insula, aITG = anterior inferior temporal gyrus, aMFG = anterior middle frontal gyrus, Cbm = cerebellum, dCMA = dorsal cingulate motor area, Den = dentate nucleus, dSC = dorsal primary somatosensory cortex, FMC = fronto-medial cortex; FOC = fronto-orbital cortex, FP = frontal pole, H = Heschl's gyrus, Inter = interposed nucleus, ITO = inferior temporo-occipital cortex, vSC = ventral primary somatosensory cortex, preSMA = presupplementary motor area, LG = lingual gyrus, MC = primary motor cortex, OC = occipital cortex, VA = ventro-anterior regions of the thalamus, Ver = vermis, SPL = superior parietal lobule, OC = occipital cortex, PCC = posterior cingulate cortex, PCN = precuneus, preSMA = presupplementary motor area, SCC = subcallosal cortex, SFG = superior frontal gyrus, SMA = supplementary motor area, SPL = superior parietal lobule, TOF = temporo-occipital fusiform gyrus, vCMA = ventral cingulate motor area. 56

Table 2.5. Functional connectivity analysis – correlations with SSI-Mod and Disfluency Rate. Roman numerals indicate cerebellar lobules. Regions of the SpeechLabel atlas (Cai, Tourville, et al., 2014) containing at least 10 voxels of a given cluster are indicated in parentheses. ROI = region-of-interest, MNI = Montreal Neurological Institute, FDR = false discovery rate, L = left, R = right, aCG = anterior cingulate gyrus, aCO = anterior central operculum, AG = angular gyrus, aINS = anterior

insula, aSMg = anterior supramarginal gyrus, avSTs = anterior ventral superior temporal sulcus, Cbm = cerebellum, dPMC = dorsal premotor cortex, dMC = dorsal primary motor cortex, FMC = fronto-medial cortex, FOC = fronto-orbital cortex, FP = frontal pole, IFr = inferior frontal gyrus pars orbitalis, ITO = inferior temporo-occipital cortex, LG = lingual gyrus, MGN = medial geniculate nucleus of the thalamus, midMC = middle primary motor cortex, MTO = middle temporo-occipital cortex, PCN = precuneus, pdSTs = posterior dorsal superior temporal sulcus, pFO = posterior frontal operculum, pIFs = posterior inferior frontal sulcus, pINS = posterior insula, pITg = posterior inferior temporal gyrus, pMFg = posterior middle frontal gyrus, pMTg = posterior middle temporal gyrus, PO = parietal operculum, pPHg = posterior parahippocampal gyrus, pSMg = posterior supramarginal gyrus, PT = planum temporale, pTFg = posterior temporal fusiform gyrus, OC = occipital cortex, SFg = superior frontal gyrus, SMA = supplementary motor area, SN = substantia nigra, STh = subthalamic nucleus, TOF = temporo-occipital fusiform gyrus, VA = ventral anterior portion of the thalamus, Ver = vermis, VPM = ventral postero-medial portion of the thalamus, vSC = ventral primary somatosensory cortex. 59

Table 2.S1. Exploratory regions-of-interest with significant group effects in either *normal – baseline* or *rhythm – baseline* contrasts ($p < 0.05$). unc = uncorrected, MTO = middle temporo-occipital cortex, pMTg = posterior middle temporal gyrus, avSTs = anterior ventral superior temporal sulcus, Cbm = cerebellum, aFO = anterior frontal operculum, aMTg = anterior middle temporal gyrus..... 96

Table 2.S2. Exploratory regions-of-interest with significant task activation group x condition interactions ($p < 0.05$). See Figure 2.S2 for individual group and condition effects. unc = uncorrected, H = Heschl's gyrus, PP = planum polare, aIFt = anterior inferior frontal gyrus pars triangularis, FOC = fronto-orbital cortex, PT = planum temporale, avSTs = anterior ventral superior temporal sulcus, TP = temporal pole, Cbm = cerebellum, VPM = ventral postero-medial portion of the thalamus. 97

Table 2.S3. Exploratory regions-of-interest with activation differences between the *rhythm* and *normal* conditions for ANS and AWS ($p < 0.05$). * indicates regions that survive a significance threshold of $p_{FDR} < 0.05$ for their respective analyses, unc = uncorrected, SPL = superior parietal lobule, aSMg = anterior supramarginal gyrus, mdPMC = middle dorsal premotor cortex, PT = planum temporale, SMA = supplementary motor area, aINS = anterior insula, PP = planum polare, ITO = inferior temporo-occipital cortex, preSMA = presupplementary motor area, pdPMC = posterior dorsal premotor cortex, FMC = fronto-medial cortex, vPMC = ventral premotor cortex, PO = parietal operculum, pSMg = posterior supramarginal gyrus, aCG = anterior cingulate gyrus, aSTg = anterior superior temporal gyrus, pSTg = posterior superior temporal gyrus, dMC = dorsal primary motor cortex, pIFt = posterior inferior frontal gyrus pars triangularis, vIFo = ventral inferior frontal gyrus pars opercularis, dIFo = dorsal inferior frontal gyrus pars opercularis, pPHg =

posterior parahippocampal gyrus, aIFt = anterior inferior frontal gyrus pars triangularis, H = Heschl's gyrus, Cbm = cerebellum, GPe = external portion of the globus pallidus. 99

Table 2.S4. Exploratory regions-of-interest with significant correlations between severity measures and speech activation in AWS ($p < 0.05$). unc = uncorrested, mdPMC = middle dorsal premotor cortex, aFO = anterior frontal operculum, midPMC = middle premotor cortex, pdPMC = posterior dorsal premotor cortex, aSMg = anterior supramarginal gyrus, SPL = superior parietal lobule, adPMC = anterior dorsal premotor cortex, PO = parietal operculum, pFO = posterior frontal operculum, PCN = precuneus, SFg = superior frontal gyrus, aTFg = anterior temporal fusiform gyrus, aCG = anterior cingulate gyrus, aCO = anterior central operculum, pCO = posterior central operculum, aINS = anterior insula, aIFs = anterior inferior frontal sulcus, vSC = ventral primary somatosensory cortex, MTO = middle temporo-occipital cortex, PT = planum temporale, pvSTs = posterior ventral superior temporal sulcus, pdSTs = posterior dorsal superior temporal sulcus, H = Heschl's gyrus, pSMg = posterior supramarginal gyrus, pFO = posterior frontal operculum, pIFs = posterior inferior frontal sulcus, VA = ventral anterior portion of the thalamus, VL = ventral lateral portion of the thalamus, Cbm = cerebellum. ... 101

Table 3.1. Symbols used to denote sound/syllable boundaries in the present study. 120

Table 3.2. Descriptive and inferential statistics for speaking rate and CV-ISD. Error estimates indicate 95% confidence intervals. Significant effects are highlighted in bold. 123

Table 3.S1. Results from an alternative model for predicting formant perturbation responses, substituting Disfluency Rate for SSI-mod. *df* = degrees of freedom, * = $p < 0.05$ 137

Table 3.S2. Results from an alternative model for predicting formant perturbation responses, substituting Disfluency Rate for SSI-mod. *df* = degrees of freedom, * = $p < 0.05$, ** = $p < 0.01$ 137

LIST OF FIGURES

Figure 2.1. Schematic diagram illustrating the temporal structure of stimulus presentation during functional data acquisition. At the start of each trial, isochronous tone sequences were presented for 3.0 seconds. The visual stimulus then appeared and remained on screen for 4.6 seconds. 1.1 seconds after stimulus offset, a whole-brain volume was acquired. The next trial started 0.33 seconds after data acquisition was complete. TR = repetition time. 36

Figure 2.2. Comparison of dysfluencies between the *normal* and *rhythm* conditions for AWS. Circles represent individual participants. *$p < .05$. 49

Figure 2.3. Cortical clusters significantly more active during the *rhythm* condition than the *normal* condition collapsed across both groups and displayed on an inflated cortical surface (vertex-wise $p < .01$, cluster-wise $p_{FDR} < .05$). 1) Left supplementary motor area, 2) Left lateral superior parietal cortex, 3) posterior superior parietal cortex, 4) Right superior parietal Cortex, 5) Right posterior supramarginal gyrus, 6) Right dorsal premotor cortex. Black outlines indicate cortical regions-of-interest (ROIs) used in the exploratory analysis. FDR = false discovery rate. 51

Figure 2.4. A summary of functional connections that are significantly different between the *normal* and *rhythm* conditions in AWS. Seed regions for these connections are indicated on the left side on a transparent 3D rendering of the cerebellum (viewed posteriorly), and colors in the rest of the figure refer back to these seed regions. Two target clusters (representing two distinct connections) are displayed in the right portion of the figure. Target cluster 1 is projected onto an inflated surface of cerebral cortex (anterior view), along with the full cortical region-of-interest (ROI) parcellation of the SpeechLabel atlas described in Cai, Tourville, et al. (2014). Target cluster 2 is displayed on a transparent 3D rendering of the cerebellum (top view: superior; bottom view: posterior). The connectivity effect sizes in the *normal* and *rhythm* conditions for each connection are displayed below each cluster visualization. Roman numerals indicate cerebellar lobules. Error bars indicate 90% confidence intervals. N = normal, R = rhythm, Cbm = cerebellum. 61

Figure 2.5. Two notable correlations of cerebellar functional connectivity (*normal* > *baseline*) with stuttering severity. Seed regions for these connections are indicated in the left side of the figure on a transparent 3D rendering of the cerebellum viewed posteriorly. Colors in the rest of the figure refer back to these seed regions. Target clusters are either displayed on the same transparent rendering of the cerebellum or projected onto an inflated surface of cerebral cortex, along with the full cortical region-of-interest (ROI) parcellation of the SpeechLabel atlas described in Cai, Tourville, et al. (2014). The '+' and '-' indicate positive and negative correlations, respectively. The right portion of the figure plots the beta estimates of the PPI regressors from individual AWS against stuttering severity. Roman numerals indicate cerebellar lobules. Full results of this analysis can be found in Figures

2.S14 and 2.S15 and Table 2.5. L = left, R = right, Cbm = cerebellum, IFr = inferior frontal gyrus pars orbitalis. 67

Figure 2.6. Correlations of functional connectivity (*rhythm* > *baseline*) between seeds regions-of-interest (ROIs) and right temporo-parietal junction with stuttering severity. Seed regions for these connections are indicated in the left side of the figure either on an inflated surface of the left cerebral cortex or on a transparent 3D rendering of right subcortical structures viewed medially. Colors in the rest of the figure refer back to these seed regions. Target clusters are projected onto an inflated surface of the right cerebral cortex, along with the full cortical ROI parcellation of the SpeechLabel atlas described in Cai, Tourville, et al. (2014). The black dashed oval indicates a rough border of the right temporo-parietal junction. The right portion of the figure plots the beta estimates of the PPI regressors from individual AWS against stuttering severity for each functional connection. Full results of this analysis can be found in Figures 2.S16 and 2.S17 and Table 2.5. L = left, R = right, aSMg = anterior supramarginal gyrus, FOC = fronto-orbital cortex, ITO = inferior temporo-occipital junction, VA = ventral anterior portion of the thalamus. 68

Figure 2.7. Correlations of functional connectivity (*rhythm* > *baseline*) between right PT and medial premotor cortex with stuttering severity. The seed region is indicated in the left side of the figure on an inflated surface of the right cerebral cortex. One target cluster (strattling the midline) is projected onto an inflated surface of the cerebral cortex, along with the full cortical ROI parcellation of the SpeechLabel atlas described in Cai, Tourville, et al. (2014). Below, the beta estimates of the PPI regressors from individual AWS are plotted against stuttering severity. Full results of this analysis can be found in Figures 2.S16 and 2.S17 and Table 2.5. L = left, R = right, PT = planum temporale. 69

Figure 2.S1. Significant clusters for the (A) *normal - baseline* or (B) *rhythm - baseline* contrasts collapsed across both groups (vertex-wise $p < .01$, cluster-wise $p_{FDR} < .05$). Cortical results (top) are displayed on an inflated cortical surface, along with the full cortical ROI parcellation of the SpeechLabel atlas described in Cai, Tourville, et al. (2014). Subcortical and cerebellar results (bottom) are displayed on an axial slice series of a template brain. Color shading indicates t-values. L = left, R = right, Lat = lateral surface, Med = medial surface. 78

Figure 2.S2. Individual group and condition effects from the exploratory regions-of-interest that had a significant interaction between group and condition. See Table 2.S2 for statistics. PP = planum polare, H = Heschl's gyrus, PT = planum temporale, aIFt = anterior inferior frontal gyrus pars triangularis, FOC = fronto-orbital cortex, avSTs = anterior ventral superior temporal sulcus, TP = temporal pole, Cbm = cerebellum, VPM = ventral postero-medial portion of the thalamus, N = *Normal - Baseline* condition, R = *Rhythm - Baseline* condition, ANS = adults who do not stutter, AWS = adults who stutter. Error bars indicuate 90% confidence intervals. 79

Figure 2.S3. Exploratory regions-of-interest (ROIs) significantly more active during the *rhythm* condition than the *normal* condition for ANS and AWS combined in the exploratory analysis ($p < .05$) are highlighted in yellow and plotted on an inflated cortical surface. ROIs highlighted in red and labeled reached significance at a stricter threshold of $p_{FDR} < .05$. ROIs highlighted in purple were significantly more active during the *normal* condition than the *rhythm* condition ($p < .05$). Black outlines indicate cortical ROIs used in the exploratory analysis. FDR = false discovery rate, aCG = anterior cingulate gyrus, aSMg = anterior supramarginal gyrus, avSTs = anterior ventral superior temporal sulcus, Cbm = cerebellum, dMC = dorsal primary motor cortex, ITO = inferior temporo-occipital cortex, mdPMC = middle dorsal premotor cortex, midPMC = middle premotor cortex, pdPMC = posterior dorsal premotor cortex, pIFt = posterior inferior frontal gyrus pars triangularis, PO = parietal operculum, PP = planum polare, pSMg = posterior supramarginal gyrus, pSTg = posterior superior temporal gyrus, PT = planum temporale, SMA = supplementary motor area, SPL = superior parietal lobule, vIFo = ventral inferior frontal gyrus pars opercularis, vMC = ventral primary motor cortex, vPMC = ventral premotor cortex. 80

Figure 2.S4. Exploratory regions-of-interest (ROIs) significantly more active during the *rhythm* condition than the *normal* condition for AWS in the exploratory analysis ($p < 0.05$) are highlighted in yellow and plotted on an inflated cortical surface. ROIs highlighted in purple were significantly more active during the *normal* condition than the *rhythm* condition ($p < .05$). Black outlines indicate cortical ROIs used in the exploratory analysis. aIFt = anterior inferior frontal gyrus pars triangularis, dIFo = dorsal inferior frontal gyrus pars opercularis, Cbm = cerebellum, GPe = external portion of the globus pallidus, H = Heschl's gyrus, pPHg = posterior parahippocampal gyrus, pSMg = posterior supramarginal gyrus, vIFo = ventral inferior frontal gyrus pars opercularis. 81

Figure 2.S5. Exploratory regions-of-interest (ROIs) significantly more active during the *rhythm* condition than the *normal* condition for ANS in the exploratory analysis ($p < .05$) are highlighted in yellow and plotted on an inflated cortical surface. ROIs highlighted in red and labeled reached significance at a stricter threshold of $p_{FDR} < .05$. ROIs highlighted in purple were significantly more active during the *normal* condition than the *rhythm* condition ($p < .05$). Black outlines indicate cortical ROIs used in the exploratory analysis. FDR = false discovery rate, aCG = anterior cingulate gyrus, aINS = anterior insula, aSMg = anterior supramarginal gyrus, aSTg = anterior superior temporal gyrus, dMC = dorsal primary motor cortex, FMC = fronto-medial cortex, ITO = inferior temporo-occipital cortex, mdPMC = middle dorsal premotor cortex, pdPMC = posterior dorsal premotor cortex, pIFt = posterior inferior frontal gyrus pars triangularis, PP = planum polare, preSMA = presupplementary motor area, pSMg = posterior supramarginal gyrus, pSTg = posterior superior temporal gyrus, PT = planum temporale, SMA = supplementary motor area, SPL = superior parietal lobule, vPMC = ventral premotor cortex. 82

Figure 2.S6. Exploratory regions-of-interest (ROIs) with a positive correlation between *normal – baseline* activation and SSI-Mod in AWS ($p < .05$) are highlighted in yellow and plotted on an inflated cortical surface. Black outlines indicate cortical ROIs used in the exploratory analysis. adPMC = anterior dorsal premotor cortex, aFO = anterior frontal operculum, aSMg = anterior supramarginal gyrus, mdPMC = middle dorsal premotor cortex, PCN = precuneus, pdPMC = posterior dorsal premotor cortex, pFO = posterior frontal operculum, PO = parietal operculum. 83

Figure 2.S7. Exploratory regions-of-interest (ROIs) with a positive correlation between *rhythm – baseline* activation and SSI-Mod in AWS ($p < .05$) are highlighted in yellow and plotted on an inflated cortical surface. ROIs highlighted in purple were negatively correlated with SSI-Mod ($p < .05$). Black outlines indicate cortical ROIs used in the exploratory analysis. aCG = anterior cingulate gyrus, aFO = anterior frontal operculum, aTFg = anterior temporal fusiform gyrus, SFg = superior frontal gyrus. 84

Figure 2.S8. Exploratory regions-of-interest (ROIs) with a positive correlation between *normal – baseline* activation and Disfluency Rate in AWS ($p < .05$) are highlighted in yellow and plotted on an inflated cortical surface. Black outlines indicate cortical ROIs used in the exploratory analysis. aCG = anterior cingulate gyrus, aCO = anterior central operculum, aIFs = anterior inferior frontal sulcus, aINS = anterior insula, Cbm = cerebellum, H = Heschl's gyrus, MTO = middle temporo-occipital cortex, pCO = posterior central operculum, pdSTs = posterior dorsal superior temporal sulcus, pFO = posterior frontal operculum, pIFs = posterior inferior frontal sulcus, PO = parietal operculum, pSMg = posterior supramarginal gyrus, PT = planum temporale, pvSTs = posterior ventral superior temporal sulcus, VA = ventral anterior, VL = ventral lateral, vSC = ventral primary somatosensory cortex..... 85

Figure 2.S9. Across-subjects correlation between *normal – baseline* activation and Disfluency Rate for AWS in four highly significant exploratory regions-of-interest (ROIs; $p < .005$). Blue circles indicate individual AWS. L = left, R = right, VA = ventral anterior portion of the thalamus, VL = ventral lateral portion of the thalamus, aCO = anterior central operculum. 86

Figure 2.S10. Cortical regions-of-interest included as seed regions in the functional connectivity analyses. aIFt = anterior inferior frontal gyrus pars triangularis, pIFt = posterior inferior temporal gyrus pars triangularis, aFO = anterior frontal operculum, pFO = posterior frontal operculum, dIFo = dorsal inferior frontal gyrus pars opercularis, vIFo = ventral inferior frontal gyrus pars opercularis, aINS = anterior insula, pINS = posterior insula, vPMC = ventral premotor cortex, midPMC = mid premotor cortex, pdPMC = posterior dorsal premotor cortex, mdPMC = middle dorsal premotor cortex, vMC = ventral primary motor cortex, midMC = middle primary motor cortex, dMC = dorsal primary motor cortex, vSC = ventral primary somatosensory cortex, midSC = middle primary somatosensory cortex, aCO =

anterior central operculum, pCO = posterior central operculum, PO = parietal operculum, aSMg = anterior supramarginal gyrus, pSMg = posterior supramarginal gyrus, PT = planum temporale, H = Heschl's gyrus, PP = planum polare, aSTg = anterior superior temporal gyrus, pSTg = posterior superior temporal gyrus, adSTs = anterior dorsal superior temporal sulcus, avSTs = anterior ventral superior temporal sulcus, pdSTs = posterior dorsal superior temporal sulcus, pvSTs = posterior ventral superior temporal sulcus, TP = temporal pole, aITg = anterior inferior temporal gyrus, ITO = inferior temporo-occipital cortex, SPL = superior parietal lobule, OC = occipital cortex, LG = lingual gyrus, SMA = supplementary motor area, preSMA = presupplementary motor area, dCMA = dorsal cingulate motor area, vCMA = ventral cingulate motor area, aCG = anterior cingulate gyrus, IFr = inferior frontal gyrus pars orbitalis, FOC = fronto-orbital cortex, FMC = fronto-medial cortex. 86

Figure 2.S11. Subcortical regions-of-interest included as seed regions in the functional connectivity analyses. L = left, R = right, Cbm = cerebellum, VA = ventral anterior portion of the thalamus, VL = ventral lateral portion of the thalamus, VPM = ventral posteromedial thalamic nucleus, MGN = medial geniculate nucleus of the thalamus, GPe = external portion of the globus pallidus, GPi = internal portion of the globus pallidus, STh = subthalamic nucleus, SN = substantia nigra..... 87

Figure 2.S12. A summary of functional connections that show significant interactions between group and condition. Seed regions for these connections are indicated in the upper left panel on an inflated right hemisphere cortical surface (top; ROIs are as in Figure 2.S10) or on a transparent 3D rendering of the cerebellum and subcortical structures viewed posteriorly (bottom). Colors in the rest of the figure refer back to these seed regions. Three target clusters (representing three distinct connections) are displayed in the upper right portion of the figure. Target clusters are projected onto an inflated surface of cerebral cortex, along with the full cortical ROI parcellation of the SpeechLabel atlas described in Cai, Tourville, et al. (2014). The bottom portion of the figure shows the connectivity effect sizes for each connection in the *normal* and *rhythm* conditions, separately for each group. Error bars indicate 90% confidence intervals. N = normal, R = rhythm, aITg = anterior inferior temporal gyrus, Cbm = cerebellum, VA = ventral anterior portion of the thalamus..... 88

Figure 2.S13. A summary of functional connections that are significantly different between the *normal* and *rhythm* conditions in ANS. Seed regions for these connections are indicated in the upper left panel either on an inflated cortical surface (top; ROIs are as in Figure 2.S10) or on a transparent 3D rendering of the left hemisphere subcortical structures viewed from the right (bottom). Colors in the rest of the figure refer back to these seed regions. Six target clusters (representing 6 distinct connections) are displayed in the upper right portion of the figure. These clusters are projected onto an inflated surface of cerebral cortex, along with the full cortical ROI parcellation of the SpeechLabel atlas described in Cai, Tourville, et al. (2014). The bottom portion of the figure shows the connectivity effect sizes in the

normal and *rhythm* conditions for each connection. Error bars indicate 90% confidence intervals. N = normal, R = rhythm, ITO = inferior temporo-occipital cortex, H = Heschl's gyrus, Put = putamen, vSC = ventral primary somatosensory cortex, aINS = anterior insula, preSMA = presupplementary motor area. 89

Figure 2.S14. A summary of functional connectivity (*normal* - *baseline*) positively correlated with stuttering severity in AWS. Seed regions for these connections are indicated in the left panel either on an inflated cortical surface (top; ROIs are as in Figure 2.S10) or on a transparent 3D rendering of subcortical structures viewed posteriorly (bottom). Colors in the rest of the figure refer back to these seed regions. Five target clusters (representing five distinct connections) are displayed in the right portion of the figure. These clusters are either projected onto an inflated surface of the right hemisphere cerebral cortex, along with the cortical ROI parcellation of the SpeechLabel atlas described in Cai, Tourville, et al. (2014), or plotted on a 3D rendering of subcortical structures (viewed superiorly). L = left, R = right, pIFt = posterior inferior frontal gyrus pars triangularis, Cbm = cerebellum, vSC = ventral primary somatosensory cortex, VA = ventral anterior portion of the thalamus. 90

Figure 2.S15. A summary of functional connectivity (*normal* - *baseline*) negatively correlated with stuttering severity in AWS. Seed regions for these connections are indicated in the left panel either on an inflated cortical surface (top; ROIs are as in Figure 2.S10) or on a transparent 3D rendering of subcortical structures in each hemisphere viewed medially (bottom). Colors in the rest of the figure refer back to these seed regions. Thirteen target clusters (representing thirteen distinct connections) are displayed in the right portion of the figure. These clusters are either projected onto an inflated cortical surface, along with the cortical ROI parcellation of the SpeechLabel atlas described in Cai, Tourville et al. (2014), or plotted on a 3D rendering of subcortical structures viewed either posteriorly (left) or superiorly (right). L = left, R = right, aSMg = anterior supramarginal gyrus, aCO = anterior central operculum, OC = occipital cortex, Cbm = cerebellum, STh = subthalamic nucleus, midMC = middle primary motor cortex, VA = ventral anterior portion of the thalamus, vSC = ventral primary somatosensory cortex, pdSTs = posterior dorsal superior temporal sulcus. 91

Figure 2.S16. A summary of functional connectivity (*rhythm* - *baseline*) positively correlated with stuttering severity in AWS. The seed region for this connection is indicated in the left panel on an inflated left hemisphere cortical surface (ROIs are as in Figure 2.S10). Colors in the rest of the figure refer back to this region. One target cluster is displayed in the right portion of the figure, projected onto an inflated surface of the left hemisphere cerebral cortex, and viewed from the left (left) or inferiorly (right). L = left, A = anterior, P = posterior, aSMg = anterior supramarginal gyrus..... 92

Figure 2.S17. A summary of functional connectivity (*rhythm - baseline*) negatively correlated with stuttering severity in AWS. Seed regions for these connections are indicated in the left panel either on an inflated cortical surface (top; ROIs are as in Figure 2.S10) or on a transparent 3D rendering of subcortical structures in each hemisphere viewed medially (bottom). Colors in the rest of the figure refer back to these seed regions. Nineteen target clusters (representing nineteen distinct connections) are displayed in the right portion of the figure. These clusters are either projected onto an inflated cortical surface, along with the cortical ROI parcellation of the SpeechLabel atlas described in Cai, Tourville, et al. (2014), or plotted on a 3D rendering of the cerebellum viewed either posteriorly (left) or superiorly (right). L = left, R = right, aSMg = anterior supramarginal gyrus, FOC = fronto-orbital cortex, ITO = inferior temporo-occipital cortex, aCG = anterior cingulate gyrus, PT = planum temporale, TP = temporal pole, midMC = middle primary motor cortex, SN = substantia nigra, VA = ventral anterior portion of the thalamus, STh = subthalamic nucleus, Cbm = cerebellum. 93

Figure 2.S18. A summary of functional connectivity (*normal - baseline*) positively correlated with Disfluency Rate in AWS. The seed region for this connection is indicated in the left panel on a 3D rendering of the cerebellum viewed either posteriorly (top) or superiorly (bottom). Colors in the rest of the figure refer back to this region. One target cluster is displayed in the right portion of the figure, either plotted on a 3D rendering of subcortical structures viewed posteriorly (left) or projected onto an inflated surface of the right hemisphere cerebral cortex, and viewed from the left (right top) or inferiorly (right bottom). R = right, A = anterior, P = posterior, Cbm = cerebellum. 94

Figure 2.S19. A summary of functional connectivity (*normal - baseline*) negatively correlated with Disfluency Rate in AWS. Seed regions for these connections are indicated in the left panel either on an inflated left hemisphere cortical surface (top; ROIs are as in Figure 2.S10) or on a transparent 3D rendering of the cerebellum viewed posteriorly (bottom). Colors in the rest of the figure refer back to these seed regions. Eight target clusters (representing eight distinct connections) are displayed in the right portion of the figure. These clusters are either plotted on a 3D rendering of subcortical structures viewed posteriorly (top left), or projected onto an inflated cortical surface, along with the cortical ROI parcellation of the SpeechLabel atlas described in Cai, Tourville, et al. (2014). L = left, R = right, pdPMC = posterior dorsal premotor cortex, pCO = posterior central operculum, pdSTs = posterior dorsal superior temporal sulcus, Cbm = cerebellum. 95

Figure 3.1. A schematic diagram showing the setup for the experiment. Following the presentation of an orthographic stimulus sentence and a condition cue ("Normal" or "Rhythm"), participants read the sentence according to the cue. Participants' speech signal was recorded and fed to an experimental computer running Audapter. On perturbed trials, detection of the relevant speech cue (onset of /ɛ/ in "steady" in the

formant perturbation condition, onset of /s/ in “steady” in the timing perturbation condition) was used to initiate the pre-programmed auditory perturbation which was fed back to the participant via insert earphones. Both the perturbed and unperturbed signals were recorded for further analysis. 108

Figure 3.2: Examples of formant and timing perturbations. A. Example spectrograms of “The steady bat” during a formant perturbation trial generated from the recorded microphone signal (top) and headphone signal (bottom). The first and second formant traces during the word “steady” are displayed. In the headphone signal spectrogram, the original formant traces are displayed in black and the shifted formant traces are overlaid in blue. Note that only the first formant is perturbed. B. A similar set of example spectrograms of “The steady bat” from a timing perturbation trial. The dashed black line indicates the onset of the /s/ in “steady” in the microphone signal. The first dashed blue line indicates the offset of the /s/ in “steady” in the microphone signal, and the second dashed blue line indicates the offset of the /s/ in “steady” in the headphone signal. The dashed purple line indicates when auditory feedback is returned to normal. Phoneme boundaries and international phonetic alphabet symbols are indicated above the microphone signals. F1 = first formant, F2 = second formant, Hz = hertz..... 113

Figure 3.3. Comparison of dysfluencies between the *normal* and *rhythm* conditions for AWS. Circles represent individual participants..... 122

Figure 3.4. Time-normalized responses to the F1 perturbation during the /ɛ/ in “steady” in the *normal* (left) and *rhythm* (right) conditions. Solid curves indicate the average difference in F1 frequency between the perturbed and non-perturbed conditions for ANS (blue) and AWS (orange). Dashed lines indicate mean response +/- standard error of the mean. The blue bars indicate intervals of significant responses in ANS ($p < 0.05$, uncorrected). The orange bars represent intervals of significant responses in AWS ($p < 0.05$, uncorrected). The magenta bars indicate intervals of significant differences between ANS and AWS ($p < 0.05$, uncorrected). Duration of the non-normalized /ɛ/ response curves (averaged between perturbed and non-perturbed trials and across subjects) for each group and speaking condition are included to the right of the response curves for reference..... 123

Figure 3.5. Cumulative responses between the timing-perturbed and non-perturbed conditions at each of seven sound/syllable boundaries (see Table 3.1). A. Responses during the *normal* speaking condition. The blue and orange curves correspond to the ANS and AWS groups, respectively. Filled circles indicate responses that are significantly from 0 ($p < 0.05$, Bonferroni-corrected). B. Responses during the *rhythm* speaking condition. Colors represent the same as in A. 126

Figure 3.6. Scatterplots comparing disfluency rate during the *normal* condition with cumulative timing perturbation responses in AWS at each of seven sound/syllable boundaries (see Table 3.1). Circles indicate individual AWS for either the *normal*

(blue) or *rhythm* (green) conditions. Least squares lines for each condition are superimposed on the data..... 128

LIST OF ABBREVIATIONS

aCG ..... Anterior cingulate gyrus

adPMC ..... Anterior dorsal premotor cortex

aFO ..... Anterior frontal operculum

aINS ..... Anterior INS

aITg ..... Anterior inferior temporal gyrus

aMFg ..... Anterior middle frontal gyrus

aMTg ..... Anterior middle temporal gyrus

ANS ..... Adults who do not stutter

ART ..... Artifact Detection Tools

aSMg ..... Anterior SMg

ASR ..... Automatic speech recognition

avSTs ..... Anterior ventral superior temporal sulcus

AWS ..... Adults who stutter

BOLD ..... Blood oxygen-level-dependent

BU ..... Boston University

CNS ..... Children who do not stutter

CO ..... Central operculum

CSF ..... Cerebrospinal fluid

CV-ISD ..... Coefficient of variation of ISD

CV-IVI ..... Coefficient of variation of IVI

CWS ..... Children who stutter

DAF ..... Delayed Auditory Feedback

dB ..... Decibels

DIVA ..... Directions Into Velocities of Articulators

dPMC ..... Dorsal premotor cortex

f0 ..... Voice fundamental frequency

F1 ..... First formant

F2 ..... Second formant

FDR ..... False discovery rate

fMRI ..... Functional MRI

FOC ..... Fronto-orbital cortex

FWE ..... Familywise error

GLM ..... General Linear Model

GODIVA ..... Gradient order DIVA

gPPI ..... General PPI

H ..... Heschl's gyrus

HL ..... Hearing level

Hz ..... Hertz

IFG ..... Inferior frontal gyrus

IFo ..... Inferior frontal gyrus pars opercularis

IFr ..... Inferior frontal gyrus pars orbitalis

IFt ..... Inferior frontal gyrus pars triangularis

INS ..... Insula

IPs ..... Intraparietal sulcus

ISD ..... Inter-syllable duration

ITO ..... Inferior temporo-occipital cortex

IVI ..... Intervocalic interval

LG ..... Lingual gyrus

MC ..... Motor cortex

MGH ..... Massachusetts General Hospital

MNI ..... Montreal Neurological Institute

MRI ..... Magnetic Resonance Imaging

ms ..... Millisecond

MTO ..... Middle temporo-occipital cortex

OC ..... Occipital cortex

PCA ..... Principal components analysis

PCN ..... Precuneus

PET ..... Positron Emission Tomography

PMC ..... Premotor cortex

pMTg ..... Posterior middle temporal gyrus

PO ..... Parietal operculum

PPI ..... Psychophysiological interaction

preSMA ..... Presupplementary motor area

pSPL ..... Posterior SPL

PT ..... Planum temporale

PWS ..... People who stutter

RMS ..... Root mean square

ROI ..... Region of interest

SBC ..... Seed-based functional connectivity

SD ..... Standard deviation

SMA ..... Supplementary motor area

SMg ..... Supramarginal gyrus

SPL ..... Sound pressure level

SPL ..... Superior parietal lobule

SSI-4 ..... Stuttering Severity Index – 4th edition

SSI-Mod ..... Modified SSI

STFT ..... Short-time Fourier transform

TE ..... Echo time

TR ..... Repetition time

VA ..... Ventral anterior portion of the thalamus

VL ..... Ventral lateral portion of the thalamus

vMC ..... Ventral MC

vPMC ..... Ventral PMC

vSC ..... Ventral somatosensory cortex

CHAPTER I: INTRODUCTION

This dissertation comprises a set of experiments that aims to quantify and describe the role of speech motor timing in stuttering. This includes how internal timing mechanisms are disrupted and lead to stuttered speech and how external timing cues can help resolve these mechanisms. I start by providing a broad background on speech motor control, before giving an introduction to stuttering and its neural correlates. I then present a more detailed account of the data and theories that implicate disrupted motor timing abilities in stuttering. I also include background on an alternative theory of stuttering implicating disrupted or noisy sensory-motor transformations. Finally, I will provide an overview of the remainder of the dissertation, including the two studies carried out to investigate these roles in stuttering.

Speech Motor Control

Speech is a highly complex sensorimotor process that requires the coordination of over 100 muscles to produce meaningful communicative auditory signals that vary at a millisecond timescale. Despite this, humans are generally able to control speech with very high fidelity and minimal effort. While there are numerous control schemes proposed for how speech is carried out (cf. Parrell et al., 2019), most incorporate two basic forms of control — feedforward[1] and feedback control. In feedforward control, the

[1] Note that I use the term "feedforward" broadly to include any type of control that does not involve sensory feedback. Feedforward is often used to describe unidirectional motor commands that are released and not updated for a given utterance. Some models of speech production use a "model predictive" controller, where the sensory or motor consequences of a motor command are predicted using an internal model of the speech system and compared to the intended command to

speech articulators are set into action based on some pre-planned targets of speech output (either speech sounds as in the DIVA model of speech production [Guenther, 2016], or motor gestures as in the Task Dynamic model [Saltzman & Munhall, 1989]). This type of control is very stable when sensory information is not present (e.g., speaking in a noisy room where one is unable to hear themself speak). However, a purely feedforward system is problematic because it is not able to make corrections for speech errors or adjust in response to changes in the vocal tract that occur due to normal development, injury, or disease.

Feedback control is the process of using sensory information to guide movement. In speech production, feedback control is usually divided into two main groups: auditory feedback and somatosensory feedback. In auditory feedback control, acoustic speech output is processed by the speaker's auditory system and compared to some auditory goal or target (Guenther, 2016). If there is a discrepancy (e.g., a speaker means to say "bet," but it comes out closer to "bat"), an error signal will be generated to make a correction for subsequent speech. Somatosensory feedback control is similar but uses tactile, proprioceptive, and kinesthetic information from vocal tract structures like the tongue and lips to detect and correct errors. While feedback control allows for more flexibility in correcting errors and adjusting to changes in a speaker's vocal tract and environment, the drawback of a purely feedback control system for speech is the processing delay. This delay between when speech is produced, sensory information is

---

refine it in a shorter time than it would take to receive actual sensory feedback. For the sake of simplicity, in the present dissertation, feedforward subsumes the "model predictive" controller.

processed, and corrections can be implemented (approximately 100–150ms for auditory feedback, 22–75ms for somatosensory feedback [Guenther, 2016]) is often too slow to properly control the rapidly changing articulatory dynamics for individual phonemic (sub-syllabic) speech sounds at typical speaking rates (Greenberg et al., 2003; Perkell, 2012). By the time a phoneme has been produced and sensory information has been processed it is likely to have been completed. Thus, it is likely that in the fully developed adult speech system, speech is largely controlled using feedforward mechanisms, while feedback can be used to correct errors in speech produced at slower rates, maintain supra-syllabic (prosodic) goals, and adapt to changes in the vocal tract over time.

Investigations of Auditory Feedback in Speech Production

Over the past 25 years, numerous studies have been conducted to experimentally test how speakers use auditory feedback to correct for speech errors. These studies typically involve applying an artificial perturbation to a speaker's auditory feedback in real time, as if an error were occurring in production, and examining the vocal responses to these changes. This is frequently accomplished by perturbing an element of the acoustic signal, such as voice fundamental frequency (f0; e.g., Burnett et al., 1998; Chen et al., 2007) or vowel formants (e.g., Niziolek & Guenther, 2013; Purcell & Munhall, 2006), and measuring compensatory responses to the perturbed parameter. People typically compensate for these perturbations by adjusting their ongoing articulatory musculature so that this element of their speech changes in the opposite direction of the shift. For example, if a speaker's first formant frequency (F1) is increased in their

auditory feedback (making “bet” sound more like “bat”), they will generally respond by lowering the F1 of their actual speech output (sounding more like “bit”).

There are two general forms of this type of experiment. In the first form, which I will refer to as *perturbation* experiments, feedback is altered either at speech onset or soon thereafter, and responses are measured in the ensuing speech within the same utterance. In this type of experiment, perturbations are only applied on a random subset of trials and interspersed with unmodified trials so the effects of one perturbed trial do not carry forward into subsequent trials. In the second, which I will refer to as sensorimotor *adaptation* experiments, feedback alterations are presented and sustained across a block of many trials, allowing the speaker to adapt to the alteration and develop long-term changes in their production. Each of these forms provides unique information about the role of auditory feedback in speech productions. The *perturbation* studies investigate the role of auditory feedback in the control of ongoing speech, while *adaptation* studies investigate how auditory feedback errors are used to update auditory-motor mappings for feedforward control of future utterances. For conciseness, the following sections will discuss only the auditory perturbation literature.

Vowel Formant Perturbations

The first studies to carry out online perturbations examined control of voice fundamental frequency (f0) by briefly shifting the entire auditory signal during a sustained vocalization (e.g., Burnett et al., 1998). Following the development of signal processing hardware that could independently manipulate the vowel formant frequencies

from the underlying source signal, investigators began to examine how speakers respond to errors in their auditory feedback. The first studies to use this technique were adaptation studies (Houde & Jordan, 1998; Purcell & Munhall, 2006), which found that speakers responded to sustained formant shifts by making compensatory adjustments to their articulators to reduce error in their auditory feedback. This technique was soon translated into a perturbation paradigm (Purcell & Munhall, 2006). In this study, participants sustained the vowel /ɛ/ for 2.5 seconds, while their first formant frequency (F1) was perturbed either up or down to sound like the neighboring vowel (either /æ/ or /ɪ/, respectively). On average, participants opposed the perturbation with a magnitude of 11% (down) or 16% (up), demonstrating that at least for sustained vowels, auditory feedback is used to make online corrections for vowel formants. Following this, Tourville et al. (2008) shifted F1 either up or down by 30% while participants read /CɛC/ syllables in a prolonged manner and found similarly proportional responses. An advantage of this study was that, unlike Purcell and Munhall (2006) who introduced the perturbation gradually over 500ms, Tourville et al. (2008) perturbed the vowel at voicing onset and determined a more precise response latency of about 150ms, very similar to f0 perturbation responses. Subsequent studies using similar stimuli and parameters corroborated these findings (Cai et al., 2012; Parrell et al., 2017). Using a slightly more sophisticated perturbation, Niziolek and Guenther (2013) additionally found that responses had increased magnitude and reduced latency when perturbations were more likely to cross a phoneme boundary, demonstrating a sensitivity of the auditory feedback system to linguistic information.

At the same time, each of these studies applied the perturbation during an intentionally prolonged vowel. This made it difficult to determine whether the online feedback corrections demonstrated were indicative of the processes that occur during natural speech, where phonemes are often shorter in duration than the demonstrated response latencies (Greenberg et al., 2003). To address this, Cai et al. (2011) and Cai, Beal, et al. (2014) applied a formant perturbation during multisyllabic connected speech. Using a speech stimulus comprising only vowels and semivowels ("I owe you a yo-yo") such that the target formant (in this case the second formant, F2) was continuous throughout the phrase, these studies found that speakers do indeed make compensatory responses to auditory feedback errors during continuous speech. Responses occurred toward the end of the perturbed phoneme and into the productions of subsequent phonemes with a latency and proportional magnitude similar to that found in previous studies (~160ms, 10%–20% of perturbation).

While these studies (Cai et al., 2011; Cai, Beal, et al., 2014) did include speech stimuli that were multisyllabic and much closer to typical speaking contexts, they were somewhat contrived in order to maintain continuous formant trajectories. It is unclear, though, whether a brief gap in voicing, as occurs during stop consonants, would reduce the effect of a perturbation on subsequent phonemes or not. To better understand how these responses may differ when formant trajectories are not continuous throughout the utterance, the study described in Chapter III will examine responses to both spectral perturbations of the first formant (F1) and perturbations to auditory feedback timing in a group of adults who stutter (AWS) and adults who do not stutter (ANS) during more

ecologically valid speech stimuli, consisting of stop consonants, fricatives, and vowels.

The DIVA Model of Speech Production

In order to understand and interpret the results previously discussed, it is often helpful to have a cohesive framework that can bring together disparate lines of evidence and generate testable hypotheses. As previously mentioned, there are now numerous control schemes of speech articulation specified in computational models (Parrell et al., 2019). Here, I introduce an influential model that proposes neural mechanisms through which feedforward and feedback control are implemented.

The Directions Into Velocities of Articulators (DIVA) model is a biologically plausible, computational model of speech production developed in the Speech Lab at Boston University (Golfinopoulos et al., 2010; Guenther, 1994, 1995, 2006, 2016; Guenther et al., 1998, 2006; Guenther & Vladusich, 2012; Tourville & Guenther, 2011). It is defined at two levels: as a computationally explicit artificial neural network that is used to control an articulatory synthesizer, and as a description of the network of brain regions associated with speech motor control. It receives as inputs neural signals from higher-level linguistic/sequencing regions that cue the production of short chunks of speech, most commonly syllables, that contain their own set of well-learned motor commands, as well as associated auditory and somatosensory targets.

As a broad overview, when a node representing a given chunk in the Speech Sound Map located in left ventral premotor cortex (vPMC) is activated, it first waits for an initiation signal from the Initiation Map located in bilateral supplementary motor area

(SMA) to determine exactly when to start. This exact timing is mediated by a cortico-basal ganglia-thalamo-cortical loop that receives contextual information from various motor, sensory, and cognitive brain regions (see *Theories of Stuttering Involving Timing Control* below for more detail on this loop). After receiving the initiation signal, a set of stored feedforward motor commands that specify the timing of individual speech gestures are sent to an Articulator Map located in bilateral ventral primary motor cortex (vMC). These commands are sent both directly and by way of a cortico-cerebellar-thalamo-cortical loop and sent to the articulators (via brainstem motor nuclei) to be carried out. At the same time, stored expectations regarding the auditory and somatosensory consequences of that produced sound are sent from the Speech Sound Map to target maps in bilateral secondary auditory and somatosensory cortices, respectively. Sensory feedback from produced speech is then processed by the relevant sensory system in state maps (in bilateral primary and secondary sensory areas). Signals from the State Maps and Target Maps are sent to the sensory Error Maps (in bilateral secondary sensory cortex) to compare the sensory expectations and the actual sensory consequences. If there is a discrepancy between these two signals, an error signal is generated and sent to a Feedback Control Map in right vPMC, which then sends additional motor commands to the Articulator Map. These commands are then combined with the feedforward commands to generate the articulatory movements that correct the sensory discrepancies.

Stuttering

General Background

Stuttering is a speech disorder that is overtly characterized by speech disfluencies such as sound or syllable repetitions (e.g., "d- d- d- dog"), prolongations (e.g., "n---- ame"), and blocks — silent pauses that are often accompanied by tense articulatory postures (Max, 2004). As compared with other speech disorders like dysarthria, there is no underlying impairment of the speech neuromusculature, and unlike aphasia, language formulation and processing are largely intact[2]. While individuals often have certain words or sounds that they feel they stutter on more frequently (Bloodstein, 1995), individual disfluencies are largely unpredictable across situations[3] (Tichenor & Yaruss, 2021) and are accompanied by a feeling of loss of control (Tichenor & Yaruss, 2019). Thus, people who stutter (PWS) know what they want to say, but there is sometimes difficulty in translating that message into smooth and timely speech articulations.

Stuttering typically emerges early in childhood between the ages of 2 and 5 (Yairi & Ambrose, 2013). In this age range, up to 8% of children develop stuttering with approximately equal incidence in males and females (Yairi & Ambrose, 2013). In the course of development most of these children will recover, however stuttering persists

---

[2] There is a sizable body of literature examining phonological processing deficits in people who stutter (see Nippold, 2002) as well as comorbidities with other language and attention disorders (Healey & Reid, 2003). However, higher-level utterance formulation is generally not considered to be a key feature of stuttering (Nippold, 2012), and will not be discussed in this dissertation.

[3] People who stutter often feel that they can anticipate an upcoming disfluency (Jackson et al., 2015), and they may try to change what they say in order to avoid it. However, even if they are able to anticipate the overt stutter, it is not entirely predictable when the entire stuttering event (including the anticipation) will occur especially across varying situations (Bloodstein, 1995; Tichenor & Yaruss, 2021).

into adulthood for 1% of the population (Craig et al., 2009; Yairi & Ambrose, 1999), and is much more likely to persist in males (Yairi & Ambrose, 2013). As a child develops, they will often develop overt secondary behaviors such as eye-blinking and facial grimacing as behaviors used to escape the moment of stuttering (Guitar, 2014). Along with these more overt characteristics, PWS often develop covert behaviors to avoid a specific stuttering event or a communication situation more generally (Guitar, 2014). Moreover, persistent stuttering often has a severe psychological impact on those who experience it, including increased social anxiety and decreased self-confidence, emotional functioning, and overall mental health (Craig et al., 2009; Craig & Tran, 2006, 2014).

As a result of these clear negative impacts, a great deal of research has been carried out to find effective therapies that address both the overt speech characteristics and the covert consequences of stuttering. At the same time, there is still no consensus regarding the underlying mechanisms of stuttering. The experiments in this dissertation aim to help elucidate these mechanisms in order to improve outcomes for people who stutter.

Neural Correlates of Stuttering

Diverse brain imaging modalities have been used to examine how the brains of people who stutter differ from those who do not and how these measures change in different speaking scenarios or following therapy (see Etchell et al. [2018] for a complete literature review). Studies have consistently found that PWS show structural and

functional differences in the brain network pertaining to speech initiation and timing (cortico-thalamo-basal ganglia motor loop; Chang & Zhu, 2013; Giraud, 2008; Lu et al., 2010) and reduced structural integrity in speech planning areas (left ventral premotor cortex [vPMC] and inferior frontal gyrus [IFG]; Beal et al., 2013, 2015; Chang et al., 2008, 2011; Garnett et al., 2018; Kell et al., 2009; Lu et al., 2012). Functionally, previous work has indicated that during speech, adults who stutter (AWS) have reduced activation in left hemisphere auditory areas (Belyk et al., 2015; Braun et al., 1997; Chang et al., 2009; De Nil et al., 2000, 2008; Fox et al., 1996; Van Borsel et al., 2003) and overactivation in right hemisphere structures (Braun et al., 1997; De Nil et al., 2000; Fox et al., 1996, 2000; Ingham et al., 2000; Van Borsel et al., 2003), which are typically non-dominant for language processing. These studies suggest that stuttering occurs as the result of impaired speech timing, planning, and/or auditory processing, and that brain structures not normally involved in speech production are potentially recruited to compensate.

In addition to these task activation analyses, previous studies have examined task-based functional connectivity (i.e., activation coupling between multiple brain areas during a speaking task) differences between AWS and ANS. Some studies show reduced connectivity between left IFG and left precentral gyrus in AWS (Chang et al., 2011; Lu et al., 2009), which suggests an impairment in translating speech plans for motor execution (Guenther, 2016). Other studies show group differences in connectivity between auditory, motor, premotor, and subcortical areas (Chang et al., 2011; Kell et al., 2018; Lu, Chen, et al., 2010; Lu et al., 2009; Lu, Peng, et al., 2010). Results of these task-based connectivity

studies, as well as resting-state and structural connectivity studies (e.g., Chang & Zhu, 2013; Sitek et al., 2016), have made it apparent that stuttering behavior is not merely the result of disruptions to one or more separate brain regions but also involves differences in the ability for brain regions to communicate with one another during speech.

Evidence for Disrupted Timing Abilities in Stuttering

Numerous theories of stuttering have been proposed, addressing both the internal factors (disrupted speech, linguistic, emotional, or cognitive processes) that lead to overt disfluencies (Bloodstein, 1972; Guenther, 2016; Howell, 2010; Lieshout et al., 2014; M. D. Neilson & Neilson, 1987; Postma & Kolk, 1993; Vasiç & Wijnen, 2005; Webster, 1998) and the effects of a child's environment on persistence of stuttering (Lieshout et al., 2014; Smith, 1999; Starkweather & Gottwald, 1990). One of these theories suggests that individuals who stutter have an impaired ability to properly time the initiation and/or termination of speech segments, with both behavioral and neural studies supporting this idea (Alm, 2004; Etchell et al., 2014; Guenther, 2016; MacKay & MacDonald, 1984; Wingate, 2002; see Theories of Stuttering Involving Timing Control). Reducing speech rate improves fluency (Andrews et al., 1982), which may allow for more time to process initiation and termination of speech segments. In addition, PWS exhibit delayed reaction times and abnormal variability in motor coordination measures during both speech and non-speech motor tasks (Kleinow & Smith, 2000; Max et al., 2003; McClean & Runyan, 2000; Starkweather et al., 1984). More recent work has demonstrated similar motor timing differences in children who stutter (CWS) in the non-speech motor domain (Falk

et al., 2015; Howell et al., 1997; Olander et al., 2010; although see Hilger et al. [2016] for a counterexample), suggesting that these differences are related to primary characteristics of stuttering.

Externally Paced Speech

Additional evidence for a speech timing disruption in stuttering comes from speaking conditions that drastically reduce or eliminate disfluency (fluency-inducing conditions). AWS have a significant reduction in stuttering when speech timing signals come from outside an individual's own speech production network, such as when speaking in rhythm with a metronome (e.g., Andrews et al., 1982; Brady, 1969; Braun et al., 1997; Davidow, 2014; Stager et al., 2003; Toyomura et al., 2011), speaking along with another speaker ("choral reading" or "shadowing"; Alm, 2004; Andrews et al., 1982; Bloodstein, 1995; Toyomura et al., 2011), or singing (Alm, 2004; Andrews et al., 1982; Bloodstein, 1995; Stager et al., 2003). These effects suggest that using external timing cues to pace speech allow AWS to circumvent inefficient or impaired "internal" timing systems. The next section will describe in more detail one of these fluency-inducing conditions, syllable-timed (also metronome-timed or "rhythmic") speech, which is the primary way I am evaluating the role of external pacing in this dissertation.

Rhythmic Speech

Synchronizing speech with an isochronous pacing stimulus like a metronome is a technique extensively studied in the literature due to its robust fluency-inducing effects on speech in AWS (often dubbed the "rhythm effect"). It has been shown that this effect

works under a variety of conditions including with both auditory and visual pacing stimuli (Barber, 1940) and pacing from a beat stored in memory (Barber, 1940; Stager et al., 2003). Most studies examining this effect have used pacing stimuli that are much slower than typically produced speech (e.g., Stager et al., 2003; Toyomura et al., 2011), which presents a confound because reducing speech rate also induces fluency in AWS (Andrews et al., 1982). However, a few studies have confirmed that the technique is still effective at speaking rates that are comparable to typical speech (Davidow, 2014; Hanna & Morris, 1977), implying that there is something about either the external pace or the isochronicity that leads to greater fluency. In addition, rhythmic speech has been tested as a part of therapy protocols (Bothe et al., 2006) for both AWS (Öst et al., 1976; Toyomura et al., 2015) and CWS (Trajkovski et al., 2009, 2011) and fluency-inducing effects have been found to partially carry over to unpaced speech. However, the evidence base for this type of therapy is small and comparisons with other methods of therapy are sparing (Bothe et al., 2006; Neumann et al., 2017). More information regarding neural correlates of this effect, as well as acoustic and aerodynamic changes during this condition can be found in the introductions to Chapters II and III, respectively.

Neural Evidence

In the neural domain, both AWS and CWS show differences in the cortico-basal ganglia motor network (Chang & Zhu, 2013; Giraud, 2008; Lu, Peng, et al., 2010) and auditory sensory areas (Chang & Zhu, 2013; De Nil et al., 2000; Foundas et al., 2001, 2004; Fox et al., 2000; Yang et al., 2016) as compared to individuals who do not stutter.

The cortico-basal ganglia motor loop may be involved in timing the onsets and offsets of speech segments as this loop is implicated in selectively releasing motor programs for action (Alm, 2004; Mink, 1996). Specifically, the striatum in the basal ganglia receives input from large portions of sensory cortex, potentially aggregating sensory and motor information to guide the release of ongoing speech utterances (Alm, 2004; Guenther, 2016; Mink, 1996). Damage to this pathway has been associated with neurogenic stuttering (Ludlow et al., 1987; Theys et al., 2013), and there is evidence that modulation of dopamine receptors in the basal ganglia can lead to reduced disfluencies in AWS (Alm, 2004).

Further, the developmental trajectories of these dopamine receptors align well with the usual ages of stuttering onset and spontaneous resolution of symptoms. Proper motor initiation requires the ability to select a "desired" motor program and to inhibit other "undesired" motor programs, and these abilities are thought to be governed by two separate basal ganglia pathways, each of which responds differently to dopamine due to the presence of different receptors (Alm, 2004; Mink, 1996). A particularly low ratio between receptors that favor "desired" and "undesired" actions could lead to more competition between motor programs and greater stuttering. The lowest ratios of "desired" to "undesired" receptor density in neurotypical children occur around age 2, near the onset of stuttering. Greater decreases in "undesired" receptors, which start to decline around age 3, could lead to a more likely chance of recovery (Alm, 2004). Differences in dopamine receptors can also potentially explain differences in stuttering rates between males and females – females tend to have high ratios of "desired"

to “undesired” receptors during childhood which may make them more likely to recover (Alm, 2004). This is based solely on nonstuttering individuals, so future investigations examining children who stutter would be needed to confirm this hypothesis.

Auditory Feedback Timing Perturbations in Stuttering

Based on the evidence that stuttering is related to disruptions in the basal-ganglia network for timing speech gestures (Chang & Zhu, 2013; Giraud, 2008; Lu et al., 2010; see *Neural Evidence* above) as well as changes in auditory processing regions (Belyk et al., 2015; Braun et al., 1997; Chang et al., 2009; De Nil et al., 2000, 2008; Fox et al., 1996; Van Borsel et al., 2003), and the marked effect that delayed auditory feedback (DAF) has on fluency in people who stutter (e.g., Kalinowski et al., 1996), it may be that the ability to use timing information from auditory feedback to sequence speech is affected in stuttering. Recent studies have used software to modify the perceived timing of a self-produced speech gesture (Cai, Beal, et al., 2014; Cai et al., 2011; Floegel et al., 2020; Mitsuya et al., 2014; Ogane & Honda, 2014; Oschkinat & Hoole, 2020). This perturbation temporally stretches a small portion of the speech signal, so the boundary between two phonemes sounds delayed to the speaker. In contrast with the DAF paradigm, in which the delay is continuously maintained, this timing perturbation returns auditory feedback to normal relatively quickly so that it is imperceptible to most participants (Cai et al., 2011). Previous work has found that when auditory feedback timing is perturbed online, ANS delay the onset of a subsequent speech gesture (Cai, Beal, et al., 2014; Cai et al., 2011; Oschkinat & Hoole, 2020). In comparison, AWS’

responses were smaller and slightly delayed (Cai, Beal, et al., 2014). This result indicates that auditory feedback timing signals are either reduced (resulting in smaller responses) or delayed (resulting in delayed responses) in AWS. The study described in Chapter III will build on this finding by examining the effects of speaking rhythmically (a fluency-inducing condition; see *Rhythmic Speech* above) on responses to auditory feedback timing perturbations in AWS.

Theories of Stuttering Involving Timing Control

The DIVA model of speech production and the Gradient Order DIVA (GODIVA) model of speech sequencing propose that stuttering is the result of an impaired ability to properly time the onsets and offsets of either phonemic or gestural components of a speech sequence due to one or more disruptions in the left hemisphere cortico-basal ganglia-thalamo-cortical loop (Chang & Guenther, 2020; Civier et al., 2013; Guenther, 2016). In this loop, a wide range of contextual information from across the brain is aggregated in the striatum of the basal ganglia to resolve differences between competing actions, which then sends signals to the supplementary motor area to cue the next movement at the proper instant in time. The three main disfluency types are explained as follows: prolongations result from difficulty cuing the termination of a motor program; blocks result from difficulty cuing the initiation of a motor program; and repetitions result from repeated dropouts in initiation signals. This framework accounts for the large amount of variability in neural studies of stuttering by suggesting that any part of this loop — cortical, subcortical, or the connections between regions — can cause a

breakdown in the process, and model simulations corroborate that specific disruptions can indeed lead to stutters (Civier et al., 2013). The onset of stuttering is suggested to arise at the time/age when motor programs are first consolidated and the cuing mechanism changes from purely cortico-cortical connections to use of the basal ganglia loops.

Harrington (1988) proposed a model of normal speech production whereby the timing of speech is governed by 1) the coarticulatory intervals between consonants (onset/coda) and vowel (nucleus) within a syllable and 2) an utterance's rhythmic plan specified as time intervals between successive stressed syllables. He goes on to argue that hearing the auditory feedback of the first syllable provides an expected delay for the second syllable. He submits that when ANS speak with DAF, hearing the delay of feedback for the first syllable yields a prediction of when feedback for the second syllable will occur and this additional delay makes the speaker want to prolong the initial consonant of the subsequent syllable so that production and feedback occur at a normal delay. Within this framework, the author proposes that stuttering occurs due to an incorrect timing prediction of the auditory consequences of the first stressed syllable in an utterance. Specifically, AWS underestimate the delay in auditory feedback of a produced syllable (under normal feedback), which generates a false timing error that leads them to respond similarly to ANS under DAF. The speech system then uses somatosensory feedback to realign the vowel with the preceding consonant. Repetitions and prolongations occur when this process repeats until AWS stop using auditory feedback and rely on their rhythmic plan to time subsequent parts of the utterance. This

model accounts for the fact that fluency is increased in the absence of auditory feedback, and disfluencies primarily occur at the beginning of an utterance — once a syllable is produced fluently, subsequent syllables do not need to rely on the incorrect timing prediction.

The covert repair hypothesis (Postma & Kolk, 1993) is a theory that provides a mechanism for all types of verbal disfluencies, not just those that appear in stuttering. In essence, the covert repair hypothesis states that individuals have a monitoring mechanism that can be used to detect linguistic errors prior to the articulatory stage of speaking. When errors are detected, there is a complete stoppage of ongoing speech while the linguistic message is corrected. Thus, disfluencies are a side-effect of this error correction mechanism. Postma and Kolk (1993) based this pre-articulatory (or "internal") monitoring mechanism on a component of an influential model of speech production by Levelt (1989). In Levelt's (1989) model, the message of an upcoming utterance, the phonetic plan, and sensory feedback are all monitored using the speech comprehension system. While overt repairs can be made to an utterance, the speed with which corrections can be made and the absence of phonemic errors in the disfluent speech suggests that this specific mechanism is internal or covert (Postma & Kolk, 1993). Postma and Kolk further provide explanations for each type of disfluency encountered during speech. Non-stuttering disfluencies like phrase and whole-word repetitions occur when the error is at the semantic or lexical level since it is reasonable to restart the unit that needed to be corrected, while stuttering events like blocks, prolongations, and sub-syllabic repetitions occur at the phonemic level. These types of stuttering events are

determined by the strategy used to maintain the speaking turn – restarting the word in the case of repetitions, and holding or delaying speech in the case of prolongations and blocks. It should be emphasized that the only distinction the authors make between people who stutter and other non-stuttering speakers are the types of errors that lead to their disfluencies (mainly phonemic for stuttering disfluencies and lexical/semantic for others) and the frequency of disfluencies exhibited. Thus, this theory is dependent on the assertion that people who stutter have a deficit in phonological programming.

A related theory of stuttering is the vicious cycle hypothesis. The vicious cycle hypothesis (Vasiç & Wijnen, 2005) argues that PWS focus excessively on the timing aspect of their inner (pre-articulatory) or overt (articulatory) speech. In particular, due to the lack of clear evidence mentioned above for a phonological deficit in PWS, as well as positive evidence that the inner speech of PWS is not exceptionally errorful, Vasiç and Wijnen (2005) propose that excessive monitoring is the issue at hand rather than excessive errors. When PWS make corrections for perceived timing errors in their articulatory plan, they further interrupt the timing of their speech, creating a "vicious cycle."

In total, Vasiç and Wijnen (2005) make three predictions: disfluencies decrease when cognitive resources are taxed; disfluencies decrease when attention is directed away from the monitor mechanism; PWS have a lowered threshold for identifying disfluencies and will therefore detect disfluencies in other speakers' fluency more easily. Vasiç and Wijnen (2005) test the first two predictions using a dual-task paradigm and the results are able to support their hypothesis – PWS have reduced disfluencies during story telling

when they are distracted by an unrelated task (when cognitive resources are taxed) and when they monitored their output for a particular word (when the monitor mechanism is distracted). For the third prediction, another study by Russell et al. (2005), showed that AWS rated others' disfluencies more harshly than ANS, whether the speaker stuttered or not. As Brocklehurst and Corley (2011) points out, the vicious cycle hypothesis provides a much better account for the development of stuttering than how it occurs in the first place, though other studies have suggested a neural basis for hypervigilance in PWS (Arnstein et al., 2011). Thus, the vicious cycle hypothesis takes the theoretical framework of the covert repair hypothesis, and updates it to incorporate newer evidence regarding phonological processing and dual-task behavior in PWS.

Another theory for stuttering related to speech timing is EXPLAN. This theory has been described in a number of papers by Howell (e.g., Howell, 2004, 2010) and colleagues, and argues that disfluencies occur for both PWS and PNS when the speech planning system (PLAN) and the speech execution system (EX) are not well coordinated with one another. As ongoing speech requires simultaneous planning and execution, the two systems are presumed to function independently of one another with a crucial link between them. When speech is too fast or the utterance is too complex, the plan for a word will not be ready by the time the production of the previous word finishes. This can either lead to stalling behavior (whole-word repetition and pauses) so that the planning system has more time to finish, or advancing behavior (sound repetitions, prolongations, mid-word pauses) where the speaker repeatedly produces the incompletely planned utterance, only moving forward when the full plan is accessible. Howell proposes that

both CWS and children who do not stutter (CNS) start off using the stalling technique, but over time, CWS's method moves to the advancing pattern. It should be noted that similar to the previous two theories, stuttering is thought to be merely an extreme form of normal disfluencies, borne out of atypical connectivity between certain regions in the brain.

Sensory-Motor Integration Disturbance in Stuttering

As discussed in Speech Motor Control (above), the speech production system monitors auditory and somatosensory feedback to make online corrections to sensory errors and update motor commands in future utterances. Encapsulated in these processes is the ability to transform these errors from a sensory reference frame to a motor reference frame. Furthermore, in order to detect sensory errors, the speech system must be able to predict the sensory consequences of a motor action, implicating a transformation of motor information to sensory information. These transformations (often called "internal models") are fundamental to theories of motor control and have been a focus of the speech (and general) motor control literature.

As described above, one way to experimentally test the fidelity of these models is to alter auditory feedback (e.g., raising the first formant frequency) or somatosensory feedback (e.g., applying a mechanical load to the jaw). Research comparing responses to altered auditory feedback of AWS to those of ANS has found largely consistent results across vowel formant perturbation and voice fundamental frequency perturbations: AWS exhibit delayed (Bauer et al., 2007; Cai, Beal, et al., 2014; Loucks et al., 2012), reduced

(Cai et al., 2012; Loucks et al., 2012; Sares et al., 2018), and/or more temporally variable (Sares et al., 2018) responses compared to ANS. Furthermore, AWS show a reduced amount of sensorimotor adaptation in response to repeated formant perturbations (Cai et al., 2012; Daliri et al., 2018). Interestingly, Daliri et al. (2018) found that CWS show similar adaptation responses to CNS, suggesting that this reduced error correction (or updating of the internal model) is a secondary response to a lifetime of stuttering. In response to somatosensory feedback perturbations, AWS also exhibit decreased compensatory responses during speech (Caruso et al., 1987; Namasivayam & Van Lieshout, 2011) and non-speech jaw movement (Loucks & De Nil, 2006) tasks. Other tasks that require the use of somatosensory information to make non-speech oral motor actions (De Nil & Abbs, 1991; Howell et al., 1995) found discrepancies between people who do and do not stutter when visual feedback was not present that disappeared when feedback was restored. These studies support the idea that people who stutter have inefficient or disrupted somatosensory and auditory internal models. Noise masking auditory feedback and reducing speech rate may then induce fluency by circumventing these impacted systems or providing more time for processing to take place, respectively (Max et al., 2004).

Theories of Stuttering Involving Sensory-Motor Integration

A few different theories of stuttering are based on the idea that sensory-motor integration is disrupted for people who stutter. In an early version of this theory, Nielson and Nielson (1987) first set up a general computational model of speech production that

they call “adaptive model theory.” This theory addresses how intended auditory consequences are transformed into motor commands that can be utilized by the speech production system (i.e., through internal models). In addition, they propose computational and neural mechanisms for how these models can be adaptively updated based on sensory feedback (note that they do not address online compensation for feedback errors). Their prior work examining auditory-motor tracking tasks (M. D. Neilson, 1980) found that when PWS had to make a motor response to a changing auditory signal, they responded with greater time-delay than PNS (this was not the case when visual information was also present). Based on this, and comparing PWS’ responses with training-based tracking improvement for PNS (P. D. Neilson et al., 1985), they suggest that PWS have a deficiency in forming inverse auditory-motor models, and thus have a limited capacity for sensory-motor information processing. When the demands exceed this capacity, fluency breaks down (blocking, repeating the task) unless the task can be simplified or more time can be taken. Thus, disfluencies decrease with noise-masking (which frees up resources from auditory feedback processing) and decreased speech rate. In addition, higher linguistic processing demands, or extra tasks can compete with sensory-motor transformation resources and increase disfluencies.

With a plethora of additional behavioral and neural data, Max et al. (2004) detail a similar hypothesis regarding the inefficiency of internal models in stuttering. Specifically, they propose that while either inverse (sensory-to-motor) models or forward (motor-to-sensory) models could be affected, it is more likely that stuttering results from disrupted forward models in feedback control (i.e., inaccurate sensory predictions). This

disruption would lead to increased detected mismatches between predictions and afferent sensory information, even when production is accurate. These erroneous mismatches may lead to continually re-attempting the movement (repetitions) or sustaining the ongoing command (prolongations) until the mismatch is resolved. They suggest that findings of slower speech and non-speech movements in PWS indicate a strategy used to allow more time to process and integrate afferent inputs.

Hickok et al. (2011) provide an alternative account of the impact of disrupted internal models on stuttering within their state feedback control model of speech production. They suggest that stuttering occurs due to a noisy sensory-motor mapping in their model's sensory-motor interface, localized in the left temporo-parietal junction (which they refer to as area Stp). This noisy mapping, which can be modulated by temporal demands like increased speech rate or environmental stressors, allows for successful training of the internal models, but on a given utterance, generates erroneous predictions which lead to increased error signals as well as erroneous correction signals. While this specific account still has yet to be fleshed out in greater detail, it provides some testable hypotheses for future research.

Summary of Dissertation

The remainder of this dissertation comprises three chapters — two of which detail experimental work, and a final chapter that summarizes the results and makes suggestions for future research. Chapter II describes a neuroimaging study examining the neural correlates of the rhythm effect in stuttering. In it, AWS and ANS read sentences aloud in

two conditions: a *rhythm* condition, where subjects pace each syllable to the rate of an isochronous series of tones, and a *normal* condition, where they read the sentence using natural timing. Critically, and in contrast with previous work, 1) the tones are only presented *before* the sentence is read and are presented prior to all trials to avoid confounding neural processing of the tones with the speech itself, and 2) the tempo of the tones is set to match syllable rates across conditions and to match that of conversational English. Both standard univariate imaging analyses and psychophysiological interaction analyses were carried out to examine activation and functional connectivity changes between the conditions. Chapter III describes a purely behavioral experiment that examines the effects of external pacing ("rhythmic speech") on responses to spectral and timing perturbations of auditory feedback in AWS and ANS. Feedback perturbations are applied during multisyllabic utterances spoken in the *rhythm* and *normal* conditions described in Chapter II. Finally, Chapter IV brings together the results of the two studies to make general conclusions regarding 1) the role of internal speech timing mechanisms in leading to stuttering and 2) the mechanism by which external timing cues reduce stuttering, and discusses planned and potential follow-up research.

CHAPTER II: The Neural Circuitry Underlying the "Rhythm Effect" in Stuttering

Abstract

Purpose: Stuttering is characterized by intermittent speech disfluencies which are dramatically reduced when speakers synchronize their speech with a steady beat. The goal of this study was to characterize the neural underpinnings of this phenomenon using functional magnetic resonance imaging.

Method: Data were collected from 16 adults who stutter and 17 adults who do not stutter while they read sentences aloud either in a normal, self-paced fashion or paced by the beat of a series of isochronous tones ("rhythmic"). Task activation and task-based functional connectivity analyses were carried out to compare neural responses between speaking conditions and groups after controlling for speaking rate.

Results: Adults who stutter produced fewer disfluent trials in the rhythmic condition than in the normal condition. Adults who stutter did not have any significant changes in activation between the rhythmic condition and the normal condition, but when groups were collapsed, participants had greater activation in the rhythmic condition in regions associated with speech sequencing, sensory feedback control, and timing perception. Adults who stutter also demonstrated increased functional connectivity among cerebellar regions during rhythmic speech as compared to normal speech and decreased connectivity between left inferior cerebellum and left prefrontal cortex.

Conclusion: Modulation of connectivity in the cerebellum and prefrontal cortex during rhythmic speech suggests that this fluency-inducing technique activates a

compensatory timing system in the cerebellum and potentially modulates top-down motor control and attentional systems. These findings corroborate previous work associating the cerebellum with fluency in adults who stutter and indicate that the cerebellum may be targeted to enhance future therapeutic interventions.

Introduction

Stuttering is a speech disorder that impacts the production of smooth and timely articulations of planned utterances. Stuttering typically emerges early in childhood and persists over the lifespan for 1% of the population (Craig et al., 2009; Yairi & Ambrose, 1999). Speech of people who stutter (PWS) is characterized by perceptually salient repetitions and prolongations of individual phonemes, as well as abnormal silent pauses at the onset of syllables and words accompanied by tension in the articulatory musculature (Max, 2004). These disfluencies are often accompanied by other secondary behaviors such as eye-blinking and facial grimacing (Guitar, 2014). Along with these more overt characteristics, stuttering also has a severe impact on those who experience it, including increased social anxiety and decreased self-confidence, emotional functioning, and overall mental health (Craig et al., 2009; Craig & Tran, 2006, 2014). Gaining a better understanding of how and why stuttering occurs will help to lead to more targeted therapies and improve quality of life for PWS.

Considerable effort has been made to identify the core pathology underlying stuttering (for reviews, see Max, 2004; Max et al., 2004). Since the advent of neural imaging modalities such as positron emission tomography (PET) and magnetic resonance imaging (MRI), a great deal has been learned about the neural substrates of stuttering (Etchell et al., 2018; see Chapter I for a summary). Functional MRI (fMRI) is a technique that is used to investigate changes in neural activation (via changes in blood oxygenation) that occur when a person speaks. This technique provides a way to compare neural activation between two or more group and across multiple speaking

conditions (Poldrack, 2011). In addition, examining the correlations between activation in different areas of the brain — termed functional connectivity — can provide information on how networks of brain regions interact across multiple tasks (Friston, 2011). Both techniques have been used extensively to examine stuttering as a phenotype, speech planning and execution processes, developmental changes, and responses to therapy.

Beyond examining neural activation in AWS during typical speech, imaging studies have also looked at activation during conditions where AWS speak more fluently. One such condition that has been widely examined behaviorally is the rhythm effect in which stuttering disfluencies are dramatically reduced when speakers synchronize their speech movements with isochronous pacing stimuli (Azrin et al., 1968; Barber, 1940; Hutchinson & Norris, 1977; Stager et al., 1997; Toyomura et al., 2011). These fluency-enhancing effects are robust; they occur regardless of whether the pacing stimulus is presented in the acoustic or visual modalities (Barber, 1940), can be induced even by an imagined rhythm (Barber, 1940; Stager et al., 2003), and occur independently of speaking rate (Davidow, 2014; Hanna & Morris, 1977). Previous studies investigating changes in brain activation during the rhythm effect (Braun et al., 1997; Stager et al., 2003; Toyomura et al., 2011, 2015) have found that during isochronous speech, both AWS and ANS had increased activation in speech-related auditory and motor regions of cortex as well as parts of the basal ganglia. These activation increases were especially pronounced for AWS as compared to ANS. Toyomura et al. (2011) also demonstrated that these activation increases occurred in regions displaying under-activation during the unpaced speaking condition. This suggests that pacing speech along with a metronome improves

fluency by “normalizing” under-activation in speech production regions. In light of the functional connectivity studies mentioned previously, characterizing changes in brain connectivity between typical and isochronously paced speech could illuminate how external pacing leads to normalized activation in the speech network and, ultimately, fluency.

In the present study, we employed functional MRI during an overt isochronously paced sentence-reading task in AWS and ANS to characterize modulation of brain activation and functional connectivity related to the rhythm effect in stuttering. In addition, this study sought to address an important issue not previously accounted for in neuroimaging studies of the rhythm effect: a reduced speaking rate in the paced compared to the un-paced condition. Reduced speaking rate and paced speech can both induce fluency in AWS (Andrews et al., 1982), but the effects are dissociable — the rhythm effect increases fluency even when speaking rates are matched between speaking conditions (Davidow, 2014). Since brain activation is also modulated by speaking rate (Fox et al., 2000; Riecker et al., 2006), activation changes between paced and un-paced conditions may reflect either the planning/production features or the fluency-inducing effect of both unless rate is accounted for. Two prior studies (Braun et al., 1997; Stager et al., 2003) examined general differences between “fluent” and “dysfluent” speaking conditions, aiming to characterize the neural underpinnings of fluency without controlling for features that contributed (e.g., rate, speaking style, percent voicing). Toyomura et al. (2011) attempted to control for rate differences between the conditions by instructing participants to speak at similar rates during both conditions. However, they

still found a significantly reduced speaking rate in the metronome-paced condition that was not accounted for in their analyses. Separating out the effects of rate would help elucidate the neural underpinnings of the rhythm effect itself. In the present study, a combination of training and analysis procedures were used to accomplish this.

Methods

The current study complied with the principles of research involving human subjects as stipulated by the Boston University institutional review board (protocol 2421E) and the Massachusetts General Hospital human research committee, and participants gave informed consent before taking part. The entire experimental procedure took approximately 2 hours, and subjects received monetary compensation.

Subjects

Sixteen AWS (11 males/5 females, aged 18–58 years, mean age = 29.9 years, SD = 12.9 years) and seventeen ANS (11 males/6 females, aged 18–49 years, mean age = 28.7 years, SD = 8.1 years) from the greater Boston area were included in the final analyses. Age was not significantly different between groups (two-sample t-test; $t = 0.31$, $p = .756$). Subjects were native speakers of American English who reported normal (or corrected-to-normal) vision and no history of hearing, speech, language, or neurological disorders (aside from persistent developmental stuttering for the AWS). Handedness was measured with the Edinburgh Handedness Inventory (Oldfield, 1971). Using this metric, all AWS were found to be right-handed (scoring greater than 40), but there was more variability among ANS (13 right-handed, 1 left-handed, and 3 ambidextrous). There was

a significant difference in handedness score between groups (Wilcoxon rank-sum test; z = 2.29, p = .022); therefore, handedness score was included as a covariate in all group imaging comparisons. For each stuttering participant, stuttering severity was determined using the Stuttering Severity Instrument, Fourth Edition (Riley, 2008); mean score = 23.1, range: 9 to 42; see Table 2.1 for individual participants). Four additional subjects (3 AWS and 1 ANS) were also tested, but they were excluded during data inspection (described below in the *Behavioral Analysis* and *Task Activation fMRI Analysis* sections).

Subject ID	Age	Gender	SSI-4 Composite	SSI-Mod	Disfluency Rate
AWS01	19	F	28	19	0%
AWS02	22	F	31	26	3.03%
AWS03	31	F	30	22	3.03%
AWS04	21	M	9	7	1.92%
AWS05	58	M	14	11	0%
AWS06	23	M	42	29	0%
AWS07	53	M	27	22	0%
AWS08	44	M	20	16	0%
AWS09	20	M	18	15	1.52%
AWS10	22	M	27	18	3.02%
AWS11	21	M	19	16	6.06%
AWS12	20	M	24	14	1.52%
AWS13	18	F	14	11	0%
AWS14	35	M	30	19	0%
AWS15	42	M	22	17	1.52%
AWS16	29	M	14	12	0%

Table 2.1. Demographic and stuttering severity data from adults who stutter. F = female; M = male; SSI-4 = Stuttering Severity Index – Fourth Edition. SSI-Mod = a modified version of the SSI-4 that does not include a subscore related to concomitant movements. Disfluency Rate = the percent of trials containing disfluencies during the *normal* speech condition.

fMRI Paradigm

Sixteen eight-syllable sentences were selected from the Revised List of Phonetically Balanced Sentences (Harvard Sentences; *IEEE Recommended Practice for Speech Quality Measurements*, 1969; see Appendix A). These sentences, composed of one- and two-syllable words, contain a broad distribution of English speech sounds (e.g. "The juice of lemons makes fine punch"). During a functional brain-imaging session, subjects read aloud the stimulus sentences under two different speaking conditions, one in which individual syllables were paced by isochronous auditory beats (i.e., the *rhythm* condition), and one in which syllables were not paced (*i.e.,* the *normal* condition). For each trial, subjects were presented with eight isochronous tones (1000 Hz, 25ms duration) with a 270 ms interstimulus interval. This resulting rate of approximately 222 beats/min was chosen so that participants' speech would approximate the rate of the *normal* condition (based on previous estimates of mean speaking rate in English; Davidow, 2014; Pellegrino et al., 2004). Participants were instructed to refrain from using any part of their body (e.g., finger or foot) to tap to the rhythm.

To avoid confounding the auditory region blood oxygen-level-dependent (BOLD) response to the pace tone and speech auditory feedback, the pacing tones were terminated prior to the presentation of the orthographic stimulus. During a *rhythm* or *normal* trial, the orthography of a given sentence was presented with the corresponding trial identifier (i.e., "Rhythm" or "Normal") presented above the sentence. From this identifier, subjects were instructed to either read the sentence "in a rhythmic way" by aligning each syllable to a beat or in a natural way. Thus, on *rhythm* trials, subjects used the tones to pace their

forthcoming speech, while on *normal* trials, they read the stimuli at a normal speaking rate, rhythm and intonation (see Appendix B for detailed instructions). The font color was either blue for *rhythm* or green for *normal* or vice versa, and colors were counterbalanced across subjects. Subjects were instructed to begin reading aloud immediately after the sentence appeared on the screen. In the event that they made a mistake, they were asked to refrain from producing any corrections and remain silent until the next trial. Silent *baseline* trials were also included wherein subjects heard the tones, and saw a random series of typographical symbols (e.g., '+\^ &$/[|\ $=[[)*% /-@ \| -%-/') clustered into word-like groupings (matched to stimulus sentences); subjects refrained from speaking during these trials.

Subjects participated in a behavioral experiment (not reported here) prior to the imaging experiment that gave them experience with the speech stimuli and the task. The time between this prior exposure and the present experiment ranged from 0 to 424 days. Immediately prior to the imaging session, subjects practiced each sentence under both conditions until they demonstrated competence with the task and sentence production. Subjects also completed a set of six practice trials in the scanner prior to fMRI data collection. To control basic speech parameters across conditions and groups, subjects were provided with performance feedback on their overall speech rate and loudness during practice only. Following this practice set, subjects completed between two and four experimental runs of test trials depending on time constraints (14 ANS and 14 AWS completed four, 3 ANS and 1 AWS completed three, 1 AWS completed two). During the experimental session, verbal feedback was provided between runs if subjects consistently

performed outside of the specified speech rate (220 ms to 320 ms mean syllable duration). Each run consisted of 16 *rhythm* trials, 16 *normal* trials, and 16 *baseline* trials, pseudo-randomly interleaved within each run for each subject. All trials were audio-recorded for later processing.

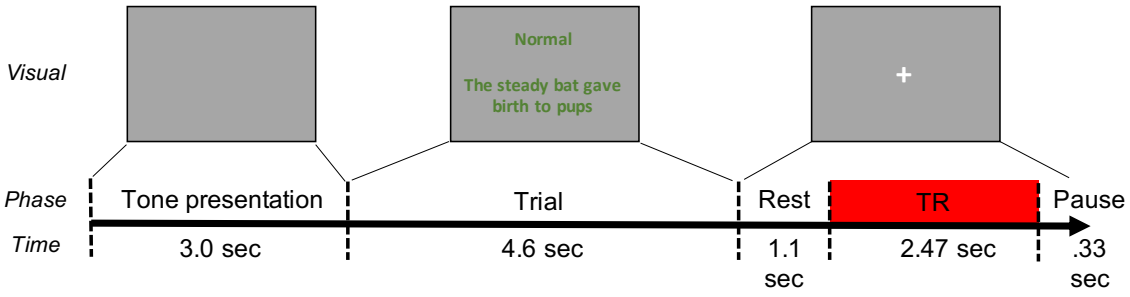

Figure 2.1. Schematic diagram illustrating the temporal structure of stimulus presentation during functional data acquisition. At the start of each trial, isochronous tone sequences were presented for 3.0 seconds. The visual stimulus then appeared and remained on screen for 4.6 seconds. 1.1 seconds after stimulus offset, a whole-brain volume was acquired. The next trial started 0.33 seconds after data acquisition was complete. TR = repetition time.

Data Acquisition

MRI data for this study were collected at two locations: the Athinoula A. Martinos Center for Biomedical Imaging at the Massachusetts General Hospital (MGH), Charlestown Campus (9 AWS, 9 ANS) and the Cognitive Neuroimaging Center at Boston University (BU; 8 AWS, 8 ANS). At MGH, images were acquired with a 3T Siemens Skyra scanner and a 32-channel head coil, while a 3T Siemens Prisma Scanner with a 64-channel head coil was used at BU. At each location, subjects lay supine in the scanner and functional volumes were collected using a gradient echo, echo planar imaging BOLD sequence (repetition time [TR] = 11.5 s, acquisition time = 2.47 s, echo

time [TE] = 30 ms, Flip Angle = 90°). Each functional volume covered the entire brain and was composed of 46 axial slices (64 x 64 matrix) acquired in interleaved order and accelerated using a simultaneous multislice factor of 3 with a 192 mm field of view. The in-plane resolution was 3.0 x 3.0 mm^2, and slice thickness was 3.0 mm with no gap. Additionally, a high-resolution T1-weighted whole-brain structural image was collected from each participant to anatomically localize the functional data (MPRAGE sequence, 256 x 256 x 176 mm^3 volume with a 1 mm isotropic resolution, TR = 2.53 s, inversion time = 1100 ms, echo time = 1.69 ms, flip angle = 7°).

Functional data were acquired using a sparse image acquisition paradigm (Eden et al., 1999; Hall et al., 1999) that allowed participants to produce the target sentences during silent intervals between volume acquisitions. Volumes were acquired 5.7–8.17 s after visual stimulus presentation to ensure a 4–6 second delay between the middle of sentence production (~2.3 seconds post-sentence presentation) and the middle of the acquisition (~6.9 seconds post-sentence presentation), aligning the acquisition to the peak of the canonical task-related BOLD response to the subject's production (Poldrack et al., 2011). Prior work has shown there is variation in the timing of this hemodynamic response across tasks, brain regions and participants (Handwerker et al., 2004; Janssen & Mendieta, 2020). However, since the functional volumes are acquired over 2.47 seconds, sentences are produced over the course of about 2 seconds, and there is a random amount of jitter between the start of the sentence production and the start of the acquisition at each trial, the single acquisition provides a broad sampling of the hemodynamic response across a range of different delay times. Furthermore, by scanning after speech production

has ended, this paradigm reduces head motion-induced scan artifacts, eliminates the influence of scanner noise on speaker performance, and allows subjects to perceive their own self-generated auditory feedback in the absence of scanner noise (e.g., Gracco et al., 2005). A schematic representation of the trial structure and timeline is shown in Figure 2.1.

Visual stimuli were projected onto a screen viewed from within the scanner via a mirror attached to the head coil. Auditory stimuli were delivered to both ears through Sensimetrics model S-14 MRI-compatible earphones using Matlab (The MathWorks, Natick, MA). Subjects' utterances were transduced with a Fibersound model FOM1-MR-30m fiber-optic microphone, sent to a laptop (Lenovo ThinkPad W540), and recorded using Matlab. Subjects took a short break after completing each run.

Behavioral Analysis

The open-source large-vocabulary continuous speech recognition engine *Julius* (Lee & Kawahara, 2009) was used in conjunction with the free *VoxForge* American English acoustic models (voxforge.org) to perform phoneme-level alignment on the sentence recordings. This resulted in phoneme boundary timing information for every trial. A researcher manually inspected each trial to ensure correct automatic detection of phoneme boundaries. Any trials in which the subject made a reading error, a condition error (i.e. spoke at an isochronous pace when they were cued to speak normally or vice versa), or a disfluency categorized as a stutter by a licensed speech-language pathologist were eliminated from further behavioral analysis. One ANS who made consistent

condition errors was eliminated from further analysis. One AWS was eliminated from further analysis due to an insufficient number of fluent trials during the *normal speech* condition (6/64 attempted). Neither were included in the total participant count in *Subjects*.

To evaluate whether there was a fluency-enhancing effect of isochronous pacing, the percentage of trials eliminated due to stuttering in the AWS group was compared between the two speaking conditions using a non-parametric Wilcoxon signed-rank test. Measures of the total sentence duration and intervocalic timing from each trial were also extracted to determine the rate and isochronicity of each production. Within a sentence, the average time between the centers of the eight successive vowels was calculated to determine the intervocalic interval (IVI). The reciprocal (1/IVI) was then calculated, resulting in a measure of speaking rate in units of IVIs per second. The coefficient of variation for intervocalic intervals (CV-IVIs) was also calculated by dividing the standard deviation of IVIs by the mean IVI. A higher CV-IVI indicates higher variability of IVI, while a CV-IVI of 0 reflects perfect isochronicity. Rate and CV-IVI were compared between groups and conditions using a mixed design ANOVA. A Bonferroni correction was applied across these two analyses to account for multiple testing.

Task Activation fMRI Analysis

Preprocessing: Following data collection, all images were processed through two preprocessing pipelines: a surface-based pipeline for cortical activation analyses and a volume-based pipeline for subcortical and cerebellar analyses. For both the surface- and

volume-based pipelines, functional images from each subject were simultaneously realigned to the mean subject image and unwarped (motion-by-inhomogeneity interactions) using SPM12's realign and unwarp procedure (Andersson et al., 2001). Outlier scans were detected with Artifact Detection Tools (ART; https://www.nitrc.org/projects/artifact_detect/) based on motion displacement (scan-to-scan motion threshold of 0.9 mm) and mean signal change (scan-to-scan signal change threshold of 5 standard deviations above the mean). For the surface-based pipeline, functional images from each subject were then coregistered with their high-resolution T1 structural images and resliced using SPM12's inter-modal registration procedure with a normalized mutual information objective function. The structural images were segmented into white matter, grey matter, and cerebrospinal fluid, and cortical surfaces were reconstructed using the FreeSurfer image analysis suite (freesurfer.net; Fischl et al., 1999). Functional data were then resampled at the location of the FreeSurfer fsaverage tessellation of each subject-specific cortical surface. For the vertex-level analyses (see *Second-Level Group Analyses* below), surfaces were additionally smoothed using iterative diffusion smoothing with 40 diffusion steps (equivalent to a 8 mm full-width half maximum smoothing kernel; Hagler et al., 2006).

For the volume-based pipeline, after the outlier detection step, functional volumes were then simultaneously segmented and normalized directly to Montreal Neurological Institute (MNI) space using SPM12's combined normalization and segmentation procedure (Ashburner & Friston, 2005). For the voxel-level analyses (see *Second-Level Group Analyses* below), volumes were also smoothed using an 8 mm full-width half

maximum smoothing kernel. A mask was then applied such that only voxels within the subcortical structures were submitted to subsequent analyses. The original T1 structural image from each subject was also centered, segmented and normalized using SPM12. Following preprocessing, two AWS (not included in the 16 described in *Subjects*) were eliminated from subsequent analyses; one due to excessive head motion in the scanner (>1.5mm average scan-to-scan motion) and one due to structural brain abnormalities.

First-level Analysis: After preprocessing, BOLD responses were estimated for each subject using a general linear model (GLM) in SPM12. Because images were collected in a sparse sequence with a relatively long TR, the BOLD response for each trial (event) was modeled as an individual epoch. The model included regressors for each of the conditions of interest: *normal*, *rhythm*, and *baseline*. Trials that contained reading errors, condition errors, or disfluencies were modeled as a single separate condition of non-interest. To control for differences in rate between the two conditions (see Results section), trial-by-trial mean IVIs were centered and added as a covariate of non-interest. These regressors were collapsed across runs to maximize power while controlling for potential differences in the number of trials produced without errors or disfluencies. For each run, regressors were added to remove linear effects of time (e.g., signal drift, adaptation) in addition to six motion covariates (taken from the realignment step) and a constant term, as well as outlier regressors (one regressor per identified outlier) to remove the effects of acquisitions with excessive scan-to-scan motion or global signal change (estimated from the artifact detection step, described above). The first-level General Linear Model regressor coefficients for the three conditions of interest were estimated at

each surface vertex and subcortical voxel. The mean *normal* speech and *rhythm* speech coefficients were then contrasted with the *baseline* condition to yield contrast effect-size values for the two contrasts of interest (*Normal – Baseline* and *Rhythm – Baseline*).

Region-of-Interest Definition: Cortical regions-of-interest (ROIs) were labeled according to a modified version of the SpeechLabel atlas previously described in (Cai, Tourville, et al., 2014); the atlas divides the cortex into macro-anatomically defined ROIs specifically tailored for studies of speech. Labels are applied by mapping the atlas from the FreeSurfer *fsaverage* cortical surface template to each individual surface reconstruction.

Subcortical and cerebellar ROIs were extracted from multiple atlases. Thalamic ROIs were extracted from the mean atlas of thalamic nuclei described by (Krauth et al., 2010). Basal ganglia ROIs were derived from the non-linear normalized probabilistic atlas of basal ganglia (ATAG) described by (Keuken et al., 2014). Each ROI was thresholded at a minimum probability threshold of 33% and combined in a single labeled volume in the atlas's native space (the MNI104 template). Cerebellar ROIs were derived from the SUIT 25% maximum probability atlas of cerebellar regions (Diedrichsen, 2006; Diedrichsen et al., 2009, 2011). Each atlas was non-linearly registered to the SPM12 MNI152 template and then combined into a single labeled volume.

Second-Level Group Analyses: Group activation differences were examined in the two speech conditions compared to baseline (*Normal – Baseline, Rhythm – Baseline)* as well as the *Group × Condition Interaction.* Additionally, differences between the two speech conditions (*Rhythm – Normal)* were examined in each group separately. All

group-level analyses were performed using a GLM with random effects across subjects. Group comparisons included the following four control covariates: a) subject motion (average framewise displacement score for each subject); b) acquisition site (MGH vs. BU); c) handedness (due to significant difference in handedness between the two groups; see Subjects section above), and d) stuttering severity, within the AWS group only. This severity covariate was a modification of the SSI-4 score, heretofore termed "SSI-Mod." SSI-Mod removes the secondary concomitants subscore from each subject's SSI-4 score, thus focusing the measure on speech-related function. The SSI-Mod and SSI-4 composite scores for each subject are included in Table 2.1. With 16 AWS and 17 ANS and four control covariates, power is sufficient (greater than 80%) to detect at a $p < .05$ false positive control level large between-group differences (Cohen's $d > 0.87$). It is not uncommon to find or expect such large effects in the context of voxel- or surface- level analyses, and these sample sizes are comparable to or larger than those of similar studies (Stager et al., 2003; Toyomura et al., 2011, 2015). Additional regression analyses were carried out to determine whether stuttering severity, measured by the SSI-Mod, or disfluencies occurring during the experiment were correlated with task activation. Because very few disfluencies occurred during the rhythm condition, we were only able to calculate the correlation between the percentage of disfluencies occurring during *normal* trials ("Disfluency Rate") and the *Normal - Baseline* activation. Note that because trials containing disfluencies were regressed out of the first-level effects, correlations with Disfluency Rate are capturing activation related to the *propensity* to stutter and not disfluent speech itself.

Two sets of group-level analyses were carried out to detect activation differences across groups and conditions: analyses at the level of the vertex (cortical) or voxel (subcortical), and exploratory ROI analyses. For the vertex/voxel analyses, the GLM was carried out on the smoothed data at each unit. Unit-wise statistics were first thresholded at a height threshold of $p < .01$ uncorrected. Cluster-level statistics were then estimated using a permutation/randomization analysis with 1000 simulations (Bullmore et al., 1999) and only clusters below $p_{FDR} < .05$ threshold are reported (topological False Discovery Rate, Chumbley et al., 2010). Additional ROI analyses were performed to determine if activation from other brain regions was also modulated by group or condition at a less strict threshold. First-level contrast effects calculated from non-smoothed data were averaged within each ROI. For each exploratory analysis, ROIs below a $p < .05$ uncorrected threshold are reported.

Functional Connectivity Analysis

Preprocessing and analysis: Seed-based functional connectivity analyses (SBC) were carried out using the CONN toolbox (Whitfield-Gabrieli & Nieto-Castanon, 2012). The same preprocessed data used for the task activation analysis were used for the functional connectivity analysis. The seeds for this analysis comprised a subset of the ROIs used in the exploratory task activation analysis, defined either in *fsaverage* surface (cortical) or MNI volume (subcortical) space. These included regions with significant positive activation (thresholded at *one-sided* $p < .05$, and corrected for multiple comparisons using a false discovery rate correction [FDR; Benjamini & Hochberg, 1995]

within each contrast) in the *Normal – Baseline* or *Rhythm – Baseline* contrasts, or significant *Rhythm – Normal* activation in either direction (thresholded at two-sided $p <$ .05, uncorrected) across all subjects. In addition, prior work has found that connectivity between left orbitofrontal regions and the cerebellum is both increased in adults who have spontaneously recovered from stuttering (Kell et al., 2018) and negatively associated with severity (Sitek et al., 2016), indicating a potential common substrate of fluency in AWS. To determine whether connectivity between these regions is also found in rhythm-induced fluency, three left orbitofrontal regions were added as seeds (see Figures 2.S10 and 2.S11 for a complete list).

The BOLD time series was first averaged within seed ROIs. To include connections between the speech production network and other regions that potentially have a moderating effect on this network, the target area in this analysis was extended to the whole brain. The target functional volume data were smoothed using an 8 mm full-width half maximum Gaussian smoothing kernel. Following preprocessing, an aCompCor (Behzadi et al., 2007) denoising procedure was used to eliminate extraneous motion, physiological, and artifactual effects from the BOLD signal in each subject. In each seed ROI and every voxel in the smoothed brain volume, denoising was carried out using a linear regression model (Nieto-Castañón, 2020) that included 5 white matter regressors, 5 cerebrospinal fluid (CSF) regressors, 6 subject-motion parameters plus their first-order temporal derivatives, scrubbing regressors to remove the effects of outlier scans (from artifact detection, described above), as well as separate regressors for each run/session (constant effects and first-order linear-trends), task condition (main and first-

order derivative terms), and error trials. No band-pass filter was applied in order to preserve high-frequency fluctuations in the residual data.

For each participant, a generalized PsychoPhysiological Interaction (gPPI; McLaren et al., 2012) analysis was implemented using a multiple regression model, predicting the signal in each target voxel with three sets of regressors: a) the BOLD time series in a seed ROI, characterizing baseline connectivity between a seed ROI and each target voxel; b) the main effects of each of the task conditions (*normal*, *rhythm*, and *baseline*), characterizing direct functional responses to each task in the target voxel; and c) their seed-time-series-by-task interactions (PPI terms) characterizing the relative changes in functional connectivity strength associated with each task. The implementation of PPI in CONN used in this paper (Nieto-Castañón, 2020) is based on the original Friston et al. (1997) formulation, where the interaction is modeled and estimated at the level of the BOLD signal directly. Among other potential benefits, this allows the direct application of PPI and gPPI to the analysis of sparse acquisition datasets. Second-level random effects analyses were then used to compare these interaction terms within and between groups and conditions, specifically the *Rhythm - Normal* contrast in AWS and ANS and the *Group* × *Condition* interaction. Additional analyses examining the correlation between *Normal – Baseline* and SSI-Mod, *Rhythm – Baseline* and SSI-Mod, and *Normal – Baseline* and Disfluency Rate in the normal condition were also carried out. All group-level analyses included the same four control covariates used in the task activation analyses. For each comparison, separate analyses were run from the 116 seed ROIs to the whole brain. Within each analysis, a two-step

thresholding procedure was used; voxels were thresholded at a $p < .001$ height threshold, followed by a cluster-size threshold of $p_{FDR} < .05$ estimated using Random Gaussian Field theory (Worsley et al., 1996). To control for family-wise error across the 116 separate seed-to-voxel analyses, a within-comparison Bonferroni correction was applied so that only significant clusters with $p_{FDR} < .00043$ (0.05/116) were reported.

Results

Behavioral Analysis

Stuttering occurred infrequently over the course of the experiment, with 7 out of 16 AWS producing no disfluencies. There was, however, a significantly lower percentage of disfluent trials in the *rhythm* condition (0.38%) compared to the *normal* condition (1.35%; $W = 42$, $p = .023$; see Figure 2.2). There was no group × condition interaction or group main effect on speaking rate but there was a significant main effect of condition with *normal* trials (3.773 IVI/sec) produced at a faster rate than *rhythm* trials (3.463 IVI/sec; $F(1,31) = 54.7$, $p_{FWE} < .001$). To examine whether this reduction in rate led to increased fluency rather than the isochronous pacing, we tested for a correlation between the change in speech rate and the reduction in disfluencies. These two measures were not significantly correlated ($r = -0.07$, $p = .80$). For isochronicity, there was no main effect of group or group × condition interaction. There was a significant main effect of condition, where subjects had a lower CV-IVI (greater isochronicity) in the *rhythm* condition (0.13) than the *normal* condition (0.25; $F(1,31) = 492.0$, $p_{FWE} < .001$). For complete results regarding speaking rate and CV-IVI, see Table 2.2.

Measure	*ANS*		*AWS*		Main effect of Group:	Main effect of Condition:	Interaction:
	Normal	*Rhythm*	*Normal*	*Rhythm*			
Speaking rate (IVI/sec)	3.797 ± 0.086	3.456 ± 0.080	3.748 ± 0.164	3.470 ± 0.173	$F(1,31) = 0.1$, $p_{FWE} = 1$	**$F(1,31) = 54.7$, $p_{FWE} < .001$**	$F(1,31) = 0.6$ $p_{FWE} = .92$
CV-IVI	0.259 ± 0.013	0.127 ± 0.006	0.251 ± 0.019	0.132 ± 0.007	$F(1,31) = 0.1$, $p_{FWE} = 1$	**$F(1,31) = 492.0$, $p_{FWE} < .001$**	$F(1,31) = 1.4$ $p_{FWE} = .48$

Table 2.2. Descriptive and inferential statistics for speaking rate and CV-IVI. Error estimates indicate 95% confidence intervals. Significant effects are highlighted in bold.

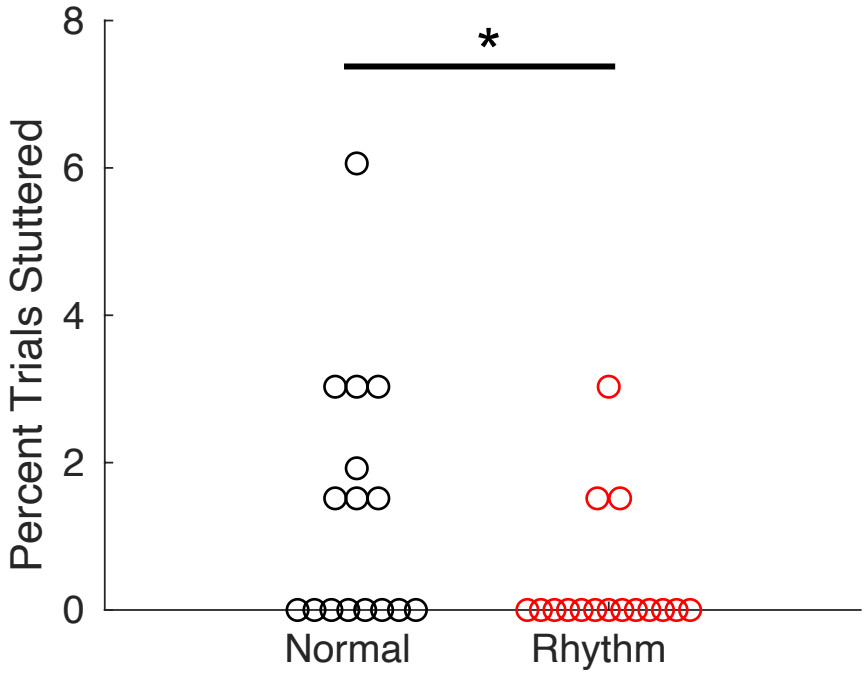

Figure 2.2. Comparison of dysfluencies between the *normal* and *rhythm* conditions for AWS. Circles represent individual participants. *$p < .05$.

Task Activation fMRI Analysis

For the vertex/voxel-wise analysis, no significant differences were found between groups for either *Normal - Baseline* or *Rhythm – Baseline* (vertex/voxel-level $p < .01$, cluster-level $p_{FDR} < .05$). Similarly, no clusters showed a significant interaction between groups and conditions. Within the AWS group, there were no significant differences between the two conditions. Because there were no significant group differences in either condition and no significant group × condition interactions, the *Rhythm – Normal* analysis was collapsed across groups to improve power. Clusters that had greater activation during the *rhythm* condition than the *normal* condition (vertex/voxel-level $p < .01$, cluster-level $p_{FDR} < .05$) are shown in Table 2.3 and Figure 2.3. These six clusters

include: left hemisphere cortex spanning posterior Sylvian fissure (planum temporale (PT) and parietal operculum (PO), supramarginal gyrus (SMg), and intraparietal sulcus (IPs); left posterior superior parietal lobule (pSPL); left supplementary motor area (SMA); right superior parietal lobule (SPL); right SMg; and right dorsal premotor cortex (dPMC). No regions in the cerebral cortex or subcortical structures were found to be more active during the *normal* condition than the *rhythm* condition.

In the exploratory ROI analysis, AWS had increased activation in left middle temporo-occipital cortex (MTO; $p = .004$), left posterior middle temporal gyrus (pMTg; $p = .010$), and left anterior ventral superior temporal sulcus (avSTs; $p = .042$) for the *Normal – Baseline* contrast compared to ANS, and decreased activation in cerebellar vermis X ($p = .049$; Table 2.S1). In the *Rhythm - Baseline* contrast, AWS had reduced activation in left anterior frontal operculum (aFO; $p < .009$), midline cerebellar vermis VIIIb ($p < .008$) and cerebellar vermis VIIIa ($p < .042$), and right anterior middle temporal gyrus (aMTg; $p < .040$) and cerebellar lobule X ($p < .046$) compared to ANS (Table 2.S1). Also in this exploratory analysis, interactions were found in a number of cortical and subcortical ROIs including bilateral auditory regions and left inferior cerebellum (see Table 2.S1 and Figure 2.S2 for complete results). In all cases, ANS had increased activation in the *Rhythm* condition compared to *Normal*, while AWS showed no change or a decrease. For complete exploratory ROI results for the *Rhythm – Normal* analysis in each group separately and combined, see Table 2.S3 and Figures 2.S3–2.S5.

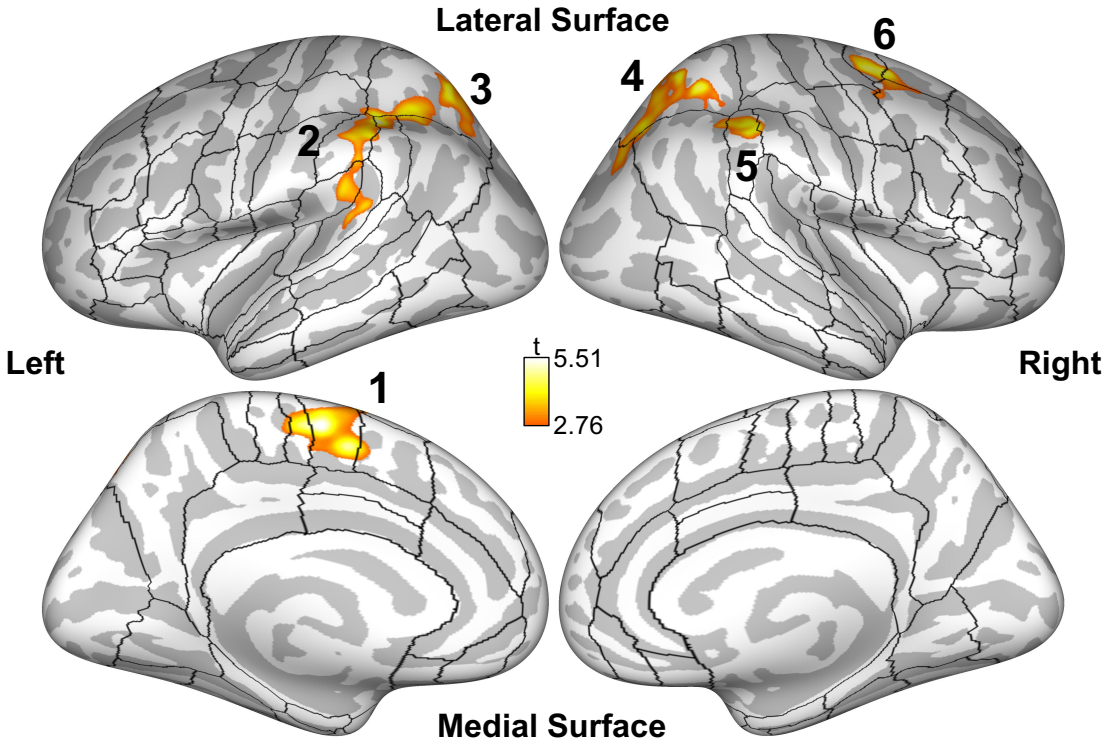

Figure 2.3. Cortical clusters significantly more active during the *rhythm* condition than the *normal* condition collapsed across both groups and displayed on an inflated cortical surface (vertex-wise $p < .01$, cluster-wise $p_{FDR} < .05$). 1) Left supplementary motor area, 2) Left lateral superior parietal cortex, 3) posterior superior parietal cortex, 4) Right superior parietal Cortex, 5) Right posterior supramarginal gyrus, 6) Right dorsal premotor cortex. Black outlines indicate cortical regions-of-interest (ROIs) used in the exploratory analysis. FDR = false discovery rate.

Cluster	Peak MNI Coordinates (x,y,z)			Cluster Mass	*p-FWE*
Combined Groups, Rhythm > Normal					
L Lateral Superior Parietal Cortex (aSMg, SPL, PO, PT, pSMg)	-42	-39	45	42456	.0080
R Superior Parietal Cortex (SPL, OC)	32	-51	55	29904	.0090
L Supplementary Motor Area (SMA, dMC, preSMA)	-09	-08	59	19658	.0166
L Posterior Superior Parietal Cortex (SPL)	-21	-68	59	13488	.0253
R Posterior Supramarginal Gyrus (pSMg, AG)	48	-32	46	12479	.0253
R Dorsal Premotor Cortex (mdPMC, adPMC, pMFg)	23	-04	53	12169	.0253

Table 2.3. Cortical clusters with activation differences between the *rhythm* and *normal* conditions collapsed across groups (vertex-wise $p < .01$, cluster-wise $p_{FDR} < .05$). MNI = Montreal Neurological Institute, FDR = false discovery rate, adPMC = anterior dorsal premotor cortex, AG = angular gyrus, aSMg = supramarginal gyrus, dMC = dorsal primary motor cortex, mdPMC = middle dorsal premotor cortex, OC = occipital cortex,

pMFg = posterior middle frontal gyrus, PO = parietal operculum, preSMA = presupplementary motor area, pSMg = posterior supramarginal gyrus, PT = planum temporale, SMA = supplementary motor area, SPL = superior parietal lobule.

Brain-Behavior Correlation Analyses

In our vertex/voxel-wise analysis, no significant clusters were found showing a correlation between SSI-Mod and *Normal - Baseline* or *Rhythm - Baseline*, or between Disfluency Rate and *Normal – Baseline*.

Exploratory results can be found in Table 2.S4 and Figures 2.S6–2.S9. Of note, positive correlations were found between SSI-Mod and activation in bilateral premotor and frontal opercular cortex and negative correlations were found in left medial prefrontal regions. In addition, positive correlations between Disfluency Rate and *Normal – Baseline* were found in right perisylvian regions, left putamen, and bilateral ventral anterior thalamus (VA)/ ventral lateral thalamus (VL) and inferior cerebellum.

Functional Connectivity Analyses

The set of 116 cortical and subcortical ROIs used as seed in the functional connectivity analyses is illustrated in Figures 2.S10 and 2.S11.

Within the AWS group, two connections were significantly different in the *rhythm* condition as compared to the *normal* condition ($p_{FDR} < .00043$), both involving the cerebellum (see Table 2.4 and Figure 2.4). The right dentate nucleus showed an increase in connectivity in the *rhythm* condition with a cluster covering right cerebellar lobule VI and Crus I, as well as vermis VI, while the left cerebellar lobule VIIIa

displayed reduced connectivity in the *rhythm* condition with a cluster in left anterior middle frontal gyrus (aMFg). To determine whether these differences were specific to AWS, a *post hoc* analysis found that these connections did not reach significance in the ANS group, even using an uncorrected alpha level of .05. Instead, ANS had different connections that were significantly different between conditions. Increased connectivity was found in the *rhythm* condition between right putamen and a cluster in anterior cingulate gyrus (aCG) straddling the midline, right anterior insula (aINS) and a cluster in left inferior frontal sulcus (IFS), and between left Heschl's gyrus (H), right presupplementary motor area (preSMA), and right ventral somatosensory cortex (vSC) seeds and clusters in left posterior superior parietal lobule abutting occipital cortex. There was also decrease in connectivity during the *rhythm* condition between left inferior temporo-occipital cortex (ITO) and left inferior frontal gyrus pars opercularis (IFo) and triangularis (IFt), and between right anterior dorsal premotor cortex (adPMC) and bilateral occipital cortex (OC; see Figure 2.S13).

There were three connections that showed a significant interaction between group and speech condition (*normal* and *rhythm*; Figure 2.S12). Connections that were lower in the *Rhythm* condition for AWS and greater in this condition for ANS included: left cerebellar lobules I–IV to left medial rolandic cortex and precuneus (PCN; result cluster labeled 1 in bottom-left panel of Figure 2.S12) and left VA to right lingual gyrus (LG) and OC (extending to right cerebellar lobule VI; cluster 2). A connection that was greater in the *Rhythm* condition for AWS and lesser in this condition for ANS was between right anterior inferior temporal gyrus (aITg) and a cluster covering parts of left central

operculum (CO), insula (INS), and surrounding regions. Simple effects from each group and condition are shown in the bottom panel of Figure 2.S12. Based on the results that showed increased connectivity for AWS between different parts of the cerebellum during isochronous speech, we performed a test comparing average pairwise connectivity among all 20 cerebellar ROIs active during speech. This test revealed that these ROIs show a significant group × condition interaction *(*$t = 2.73$*,* $p = .011$), driven by an increase in connectivity for AWS from *normal* to *rhythm* ($t = 2.68$, $p = .019$) and a non-significant decrease in connectivity for ANS ($t = -1.93$, $p = .073$).

For the AWS group, there were multiple functional connections that were significantly correlated with either SSI-Mod or Disfluency rate. Results are summarized in Table 2.5 and Figures 2.S14–2.S19. Of note, connectivity differences between the *normal* and *baseline* conditions were negatively correlated with SSI-Mod between cerebellar vermis crus II and bilateral cerebellum lobules IX and VIIIb (Figure 2.5). There was also a significant positive correlation between SSI-Mod and connections between left cerebellar lobule VIIIa and right inferior frontal gyrus pars orbitalis (IFr) in the *normal* condition compared to *baseline* (Figure 2.5). In addition, connectivity differences between the *rhythm* and *baseline* conditions were negatively correlated with SSI-Mod between the right temporoparietal junction and cluster in each of the left anterior supramarginal gyrus (aSMg), left IFr, left ITO, and right VA (Figure 2.6); and between right PT and a cluster in medial premotor cortex/SMA (Figure 2.7).

Seed ROI	Target Cluster Regions	Peak MNI Coordinates (x,y,z)			Cluster Size (# of Voxels)	*p-FDR*
AWS, Rhythm > Normal						
R Dentate Nucleus	Superior Cerebellum (R VI, Ver VI, R Crus I)	14	-72	-20	435	$< 1 \times 10^{-6}$
AWS, Normal > Rhythm						
L Cbm VIIIa	Left Anterior Middle Frontal Gyrus (L aMFg)	-28	34	30	170	.000207
ANS, Rhythm > Normal						
L H	Left Parieto-Occipital Cortex (L SPL)	-24	-70	38	186	.000195
R vSC	Left Parieto-Occipital Cortex (L SPL, L AG, L OC)	-24	-66	34	199	.000176
R aINS	Left Inferior Frontal Sulcus (L aIFs, L pIFs, L aMFg)	-50	26	26	243	.000032
R Putamen	Midline Cingulate Motor Cortex (L dCMA, R vCMA, L vCMA, R dCMA, L SMg, L aCG, R SFg, R aCG)	02	14	30	353	$< 1 \times 10^{-6}$
R preSMA	Left Parieto-Occipital Cortex (L SPL, L OC, L PCN, L AG)	-16	-70	34	224	.000100
ANS, Normal > Rhythm						
L ITO	Left Inferior Frontal Gyrus (L vIFo, L pIFt, L pFO)	-56	24	04	175	.000221
Group x Condition Interaction						
L Cbm I-IV	Medial Sensorimotor Cortex (L dSC, L pCG, L PCN, L dMC)	-16	-52	44	254	.000301
L VA	Right Inferior Occipital Cortex (R OC, R LG, R TOF, R VI, Ver VI)	08	-72	-08	365	.000023
R aITg	Left Fronto-Parietal Operculum (L aINS, L aCO, L pCO, L pINS, L pFO, L H, L PO, L vPMC)	-40	-12	14	507	$< 1 \times 10^{-6}$

Table 2.4. Functional connectivity analysis — condition and interaction effects. Roman numerals indicate cerebellar lobules. ROI = region-of-interest, MNI = Montreal Neurological Institute, FDR = false discovery rate, L = left, R = right, ACC = anterior cingulate cortex, AG = angular gyrus, aINS = anterior insula, aITG = anterior inferior temporal gyrus, aMFG = anterior middle frontal gyrus, Cbm = cerebellum, dCMA = dorsal cingulate motor area, Den = dentate nucleus, dSC = dorsal primary somatosensory cortex, FMC = fronto-medial cortex; FOC = fronto-orbital cortex, FP = frontal pole, H = Heschl's gyrus, Inter = interposed nucleus, ITO = inferior temporo-occipital cortex, vSC = ventral primary somatosensory cortex, preSMA = presupplementary motor area, LG = lingual gyrus, MC = primary motor cortex, OC = occipital cortex, VA = ventro-anterior regions of the thalamus, Ver = vermis, SPL = superior parietal lobule, OC = occipital cortex, PCC = posterior cingulate cortex, PCN = precuneus, preSMA = presupplementary motor area, SCC = subcallosal cortex, SFG = superior frontal gyrus, SMA = supplementary motor area, SPL = superior parietal lobule, TOF = temporo-occipital fusiform gyrus, vCMA = ventral cingulate motor area.

Seed ROI	Target Cluster Regions	Peak MNI Coordinates (x,y,z)			Cluster Size (# of Voxels)	*p-FDR*
AWS, Normal > Baseline Negative Correlation with SSI-Mod						
L aSMg	Left Parietal Operculum (L PO, L aSMg)	-32	-22	24	216	.000108
L aCO	Left Medial Prefrontal Cortex (L SFg, L aCG)	-12	38	30	227	.000031
L OC	Right Medial Posterior Temporal Lobe (R LG, R TOF, R pPHg, R pTFg)	22	-54	-12	320	.000001
L Cbm Crus I	Right Inferior Temporal Lobe (R LG, R TOF, R pPHg, R pTFg, R ITO, R avSTs, R adSTs, R VI, R V, R MGN, R SN, R VPM)	38	-42	-14	1672	$< 1 \times 10^{-6}$
	Left Inferior Medial Temporal Lobe (L LG, L TOF, L pTFg, L V, L VI, L I-IV, R I-IV)	-06	-52	-10	396	$< 1 \times 10^{-6}$
L STh	Left Occipital Cortex/Superior Cerebellum (L TOF, L OC, L VI, L Crus I, Ver VI)	-8	-78	-18	235	.000178
Cbm Vermis Crus II	Brainstem/Inferior Cerebellum (Brainstem, L IX, R VIIIb, R IX)	-4	-52	-60	174	.000364
R midMC	Left Medial Prefrontal Cortex (L SFg, L FP)	02	54	-08	231	.000109
R VA	Left Angular Gyrus (L AG)	-58	-64	28	264	.000026
	Right Middle Temporal-Occipital Cortex (R MTO)	60	-60	02	253	.000026
	Right Parietal Operculum (R PO, R PT, R pINS)	34	-24	16	240	.000029
R vSC	Left Medial Prefrontal Cortex (L SFg, L FP)	-04	52	18	353	.000001
R pdSTs	Midline Medial Prefrontal Cortex (L SFg, L FP, R SFg)	-12	40	26	184	.000247
AWS, Normal > Baseline Positive Correlation with SSI-Mod						
L Cbm VIIIa	Right Orbital Inferior Frontal Gyrus (R IFr, R FP, R FOC)	50	50	-12	234	.000061
L pIFt	Right Angular Gyrus (R AG)	36	-70	46	271	.000017
R VA	Right Inferior Temporal Sulcus (R pMTg, R pITg)	54	-38	-16	169	.000391
R vSC	Brainstem (Brainstem)	08	-14	-36	292	.000005
	Right Posterior Angular Gyrus (R AG, R OC)	36	-74	36	260	.000010

Seed ROI	Target Cluster Regions	Peak MNI Coordinates (x,y,z)			Cluster Size (# of Voxels)	*p-FDR*
AWS, Rhythm > Baseline Negative Correlation with SSI-Mod						
L aSMg	Right Parietal Operculum (R PO, R aSMg, R pINS)	38	-24	24	404	$< 1 \times 10^{-6}$
L FOC	Right Inferior Parietal Cortex (R aSMg, R PO, R vSC)	50	-22	40	374	$< 1 \times 10^{-6}$
L ITO	Right Dorsal Rolandic Cortex (R dMC, R dPMC, R vSC)	18	-10	70	358	$< 1 \times 10^{-6}$
	Right Frontal Operculum (R pFO, R aFO, R aINS)	32	14	10	299	.000002
	Right Temporo-Parietal Junction (R PT, R PO)	40	-30	30	290	.000002
L Cbm I-IV	Midline Occipital Cortex (R OC, L OC, R LG)	00	-70	06	401	$< 1 \times 10^{-6}$
L Cbm Crus I	Left Parieto-Occipital Fissure (L PCN, L OC)	-08	-64	28	259	.000073
R TP	R Middle Frontal Gyrus (R pMFg, R pIFS)	52	18	38	213	.000254
R PT	Midline Medial Precentral Gyrus (R dPMC, L dPMC, L dMC, L SMA)	10	-28	74	190	.000194
R aCG	Right Supramarginal Gyrus (R pSMg, R AG)	54	-40	46	350	.000001
R midMC	Midline Rostral Prefrontal Cortex (L aCG, L FP, R FP, R FMC)	-08	50	04	370	$< 1 \times 10^{-6}$
R STh	Right Anterior Insula (R aINS, R IFr)	34	22	12	368	.000001
R SN	Right Inferior Cerebellum (R IX, R X, L IX, Ver VIIIa, Ver VIIIb, L VIIb, R VIIIa, L VIIIa, L Dentate, R VIIb, Ver IX)	-10	-66	-40	404	$< 1 \times 10^{-6}$
R VA	Right Temporo-Parietal Junction (R PO, R PT, R aSMg, R vSC)	60	-22	20	559	$< 1 \times 10^{-6}$
	Right Middle Temporo-Occipital Cortex (R MTO)	60	-50	02	232	.000036
R Cbm I-IV	Midline Occipital Cortex (L OC, R OC, R PCN)	-08	-78	06	286	.000008
Cbm Vermis VI	Left Parieto-Occipital Fissure (L PCN, L OC, L LG)	-24	-62	24	622	$< 1 \times 10^{-6}$
	Right Parieto-Occipital Fissure (R PCN, R OC)	10	-64	16	382	$< 1 \times 10^{-6}$
	Right Anterior Cingulate Cortex (R aCG)	16	26	20	160	.000415
AWS, Rhythm > Baseline Positive Correlation with SSI-Mod						
L aSMg	Left Inferior Occipital Cortex (L OC, L TOF, L ITO)	-42	-74	-16	226	.000032
AWS, Normal > Baseline Negative Correlation with Disfluency Rate						
L pdPMC	Right Insula (R aINS, R aCO, R Putamen)	42	04	08	186	.000360
L pdSTs	Right Medial Temporal Cortex (R pPH, R pTFg)	34	-34	-20	191	.000059
	Right Lateral Occipital Cortex (R OC)	38	-64	-12	144	.000292
L pCO	Midline Rolandic Cortex (R dMC, R dPMC, R dSC, L dMC, L dSC)	14	-22	78		
L Dentate	Right Occipital Cortex (R OC)	12	-84	16	216	.000083
L Cbm Crus I	Right Dorsal Prefrontal Cortex (R FP, R SFg)	24	48	36	199	.000038
Cbm Vermis Crus I	Right Occipital Cortex (R OC, R PCN)	06	-80	28	1149	$< 1 \times 10^{-6}$
Cbm Vermis VIIIa	Right Occipital Cortex (R OC, R LG)	10	-62	06	184	.000223
AWS, Normal > Baseline Positive Correlation with Disfluency Rate						
R Cbm X	Inferior Temporal Cortex (R pITg, R pTFg, R VI)	46	-36	-28	356	.000001

Table 2.5. Functional connectivity analysis – correlations with SSI-Mod and Disfluency Rate. Roman numerals indicate cerebellar lobules. Regions of the SpeechLabel atlas (Cai, Tourville, et al., 2014) containing at least 10 voxels of a given cluster are indicated in parentheses. ROI = region-of-interest, MNI = Montreal Neurological Institute, FDR = false discovery rate, L = left, R = right, aCG = anterior cingulate gyrus, aCO = anterior central operculum, AG = angular gyrus, aINS = anterior insula, aSMg = anterior supramarginal gyrus, avSTs = anterior ventral superior temporal sulcus, Cbm = cerebellum, dPMC = dorsal premotor cortex, dMC = dorsal primary motor cortex, FMC = fronto-medial cortex, FOC = fronto-orbital cortex, FP = frontal pole, IFr = inferior frontal gyrus pars orbitalis, ITO = inferior temporo-occipital cortex, LG = lingual gyrus, MGN = medial geniculate nucleus of the thalamus, midMC = middle primary motor cortex, MTO = middle temporo-occipital cortex, PCN = precuneus, pdSTs = posterior dorsal superior temporal sulcus, pFO = posterior frontal operculum, pIFs = posterior inferior frontal sulcus, pINS = posterior insula, pITg = posterior inferior temporal gyrus, pMFg = posterior middle frontal gyrus, pMTg = posterior middle temporal gyrus, PO = parietal operculum, pPHg = posterior parahippocampal gyrus, pSMg = posterior supramarginal gyrus, PT = planum temporale, pTFg = posterior temporal fusiform gyrus, OC = occipital cortex, SFg = superior frontal gyrus, SMA = supplementary motor area, SN = substantia nigra, STh = subthalamic nucleus, TOF = temporo-occipital fusiform gyrus, VA = ventral anterior portion of the thalamus, Ver = vermis, VPM = ventral postero-medial portion of the thalamus, vSC = ventral primary somatosensory cortex.

Discussion

This study aimed to characterize the changes in functional activation and connectivity that occur when adults time their speech to an external metronomic beat and how these changes differ in AWS compared to ANS. Extending previous work, this paradigm was novel in that the metronome was paced at the typical rate of English speech. The rate and isochronicity of paced speech by AWS was also similar to that of ANS. Consistent with prior literature, AWS produced significantly fewer disfluencies during externally paced speech than during normal, internally paced speech (Figure 2.2). Controlling for speaking rate, participants exhibited greater activation during isochronously paced speech than internally paced speech in left hemisphere sensory association areas as well as bilateral attentional and premotor regions. AWS had greater functional connectivity during isochronous speech than internally paced speech within the cerebellum and reduced connectivity between left inferior cerebellum and left prefrontal cortex. Finally, there were significant correlations between SSI-Mod and functional connections within the cerebellum, between the cerebellum and orbitofrontal cortex, and between right temporoparietal junction and left hemisphere speech-related regions. The following sections discuss these results in relation to prior behavioral and neuroimaging literature.

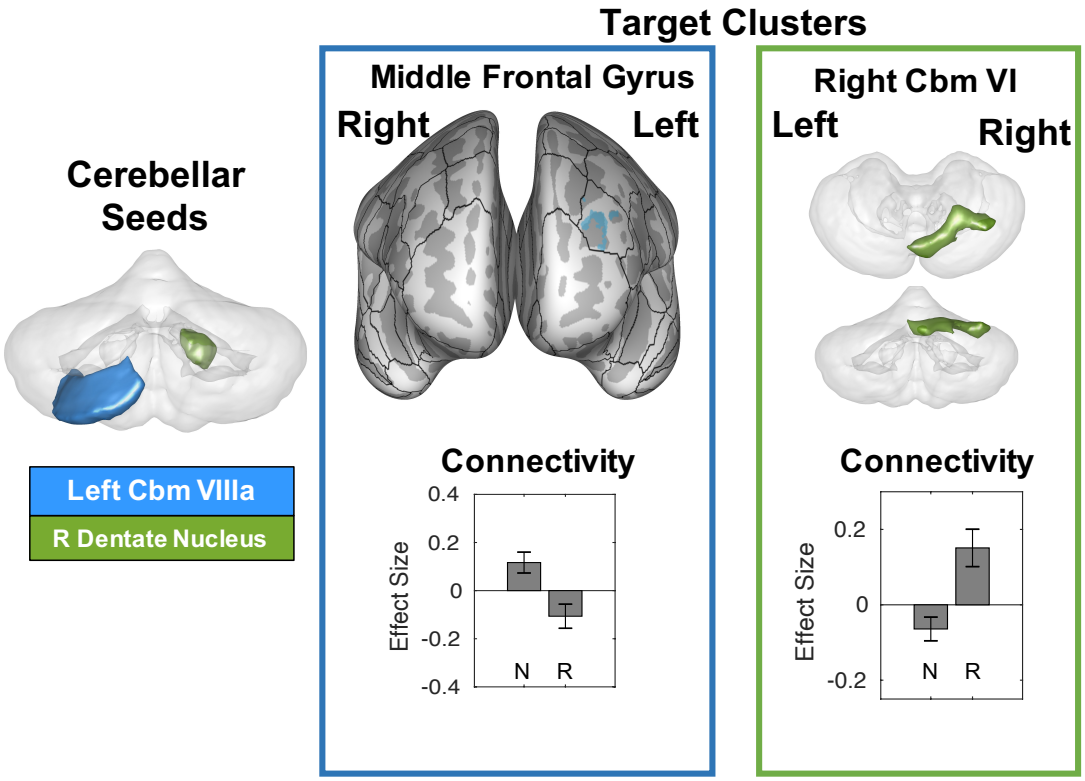

Figure 2.4. A summary of functional connections that are significantly different between the *normal* and *rhythm* conditions in AWS. Seed regions for these connections are indicated on the left side on a transparent 3D rendering of the cerebellum (viewed posteriorly), and colors in the rest of the figure refer back to these seed regions. Two target clusters (representing two distinct connections) are displayed in the right portion of the figure. Target cluster 1 is projected onto an inflated surface of cerebral cortex (anterior view), along with the full cortical region-of-interest (ROI) parcellation of the SpeechLabel atlas described in Cai, Tourville, et al. (2014). Target cluster 2 is displayed on a transparent 3D rendering of the cerebellum (top view: superior; bottom view: posterior). The connectivity effect sizes in the *normal* and *rhythm* conditions for each connection are displayed below each cluster visualization. Roman numerals indicate cerebellar lobules. Error bars indicate 90% confidence intervals. N = normal, R = rhythm, Cbm = cerebellum.

A Possible Compensatory Role for the Cerebellum in AWS

A role for the cerebellum in mediating speech timing is well-established (see Ackermann, 2008 for a review), and damage to this structure can lead to "scanning speech," where syllables are evenly paced (Duffy, 2013). Previous work posits that when

the basal-ganglia-SMA “internal” timing system is impaired in AWS, the cerebellum, along with lateral cortical premotor structures, forms part of an “external” timing system that is recruited (Alm, 2004; Etchell et al., 2014). Cerebellar lobule VI, in particular, is a common site of damage in ataxic dysarthria (Urban et al., 2013), and is found to be active during speech production and orofacial movement tasks (Guenther et al., 2016), supporting a role in speech execution. As such, cerebellar lobule VI would be able to provide additional contextual information to the cortico-basal ganglia-thalamo-cortical motor initiation loop for precisely timing speech segments. In support of this, numerous fMRI and PET studies demonstrate cerebellar overactivation and hyper-connectivity during normal speech production in AWS (Brown et al., 2005; Chang et al., 2009 [L Cbm VI]; Ingham et al., 2012 [R Cbm VI]; Lu, Peng, et al., 2010; Lu et al., 2012 [L Cbm VI]; Watkins et al., 2007 [L/R Cbm VIIIa]) that is reduced following therapy (De Nil et al., 2001; Lu et al., 2012 [L Cbm VI]; Neumann et al., 2003 [L/R Cbm VI]; Toyomura et al., 2015), a potential indication of an organic attempt at compensation. In the present study, the increased connectivity among speech-related regions of the cerebellum along with increased fluency during the rhythm condition may thus reflect similar neural processes.

In addition, the cerebellum and basal ganglia systems are thought to mediate distinct types of timing perception; the basal ganglia system is important for beat-based timing, while the cerebellar system is important for interval-based timing (Teki et al., 2011; Grube et al., 2010). Thus, with the detection of increased cerebellar involvement during metronome-timed speech in AWS but not ANS, AWS may be using this interval-based timing system to control syllable timing, while ANS use the more efficient beat-

based timing system mediated by the basal ganglia. From this, a difference in the isochronicity of speech during the *rhythm* condition between the two groups may be expected. However, this difference was not found. The lack of a difference may be due to the relatively simplistic measure of timing we used to characterize rhythmicity. While some previous work has shown that AWS and ANS compare comparably on isochronously-paced tasks (Max and Yudman, 2003), a more recent study using more sensitive circular statistics found a timing deficit in AWS when tapping along with a metronome (Sares et al., 2019). This difference will need to be investigated further in future studies of the rhythm effect in stuttering.

It should be noted that functional connectivity within the cerebellum does not reflect direct structural connectivity between a seed and target region. As suggested by Bernard et al. (2013), we interpret the result of increased within-cerebellar connectivity as reflecting an increase in synchrony among multiple cerebro-cerebellar loops. Thus, in AWS, areas of cerebral cortex may simultaneously impinge on distinct areas of cerebellum to utilize the cerebellum's temporal processing capabilities to ensure accurate speech timing during the *rhythm* condition.

The reduction in connectivity between left prefrontal cortex and inferior cerebellum may be an exception. Both regions are functionally connected during rest with areas of the ventral attention network including bilateral temporoparietal junction and inferior frontal gyrus (Buckner et al., 2011; Vossel et al., 2014; Yeo et al., 2011). This network is associated with modulating attention based on new or surprising stimuli (Vossel et al., 2014), is largely right-lateralized (Vossel et al., 2014), and overlaps with

regions involved in responding to sensory feedback errors during speech production (Golfinopoulos et al., 2011; Tourville et al., 2008). Indeed, cerebellar lobule VIII is also involved in sensory feedback control (Golfinopoulos et al., 2011; Tourville et al., 2008) as well as suprasyllabic speech sequencing (Bohland & Guenther, 2006). Thus, a reduction in connectivity between these two regions during the *rhythm* condition may reflect a decrease in reliance on this network in favor of more top-down control in AWS.

Changes in Activation during Isochronous Speech

Comparing neural activation between isochronously paced and normal speech showed that subjects had greater activation during isochronous speech in left hemisphere medial premotor and sensory association areas, bilateral parietal cortex, and right hemisphere dorsal premotor cortex. Activation in left temporo-parietal sensory association cortex (PT, aSMg) and right ventral premotor cortex (vPMC in the exploratory results) may be related to increased reliance on sensory feedback control during this novel speech condition. Previous studies have shown that sensory feedback errors (i.e., mismatches between the auditory signal expected from the current motor commands and the actual auditory signal) lead to increased activation in secondary auditory and somatosensory areas (Hashimoto & Sakai, 2003; Parkinson et al., 2012; Takaso et al., 2010; Tourville et al., 2008), whereas greater activation in right vPMC is thought to reflect the transformation of sensory errors into corrective motor responses (Elisa Golfinopoulos et al., 2011; Hashimoto & Sakai, 2003; Tourville et al., 2008). Temporo-parietal cortex may also play a more general role in audio-motor integration (Hickok et al., 2003); therefore, increased activation in this region may be indicative of

the need to hold the rhythmic auditory stimulus in working memory and translate it into a motoric response in the rhythm condition of the current study. This is supported by increased activity in bilateral intraparietal sulcus and posterior supramarginal gyrus, additional regions commonly recruited in working memory tasks (Rottschy et al., 2012).

There was also increased activation during isochronous speech in areas thought to be involved in speech planning and sequencing (left SMA; Bohland et al., 2010; Civier et al., 2013; Guenther, 2016), producing complex motor sequences (left SPL; Haslinger et al., 2002; Heim et al., 2012), producing novel sequences (left SPL; Jenkins et al., 1994; Segawa et al., 2015), attending to stimulus timing (left SPL; Coull, 2004), and controlled respiration (right dPMC; McKay et al., 2003). The rhythm condition requires participants to produce speech in an unfamiliar way. This change in their speech production results in speech becoming less automatic, and may require greater recruitment in these areas for timing the sequence of syllables (Alario et al., 2006; Bohland & Guenther, 2006; Schubotz & von Cramon, 2001). Bengtsson et al. (2004, 2005) found that for both finger tapping and simple repetition of "pa," more complex timing led to increased activation in SMA compared to simple patterns. The increased need to implement a timing pattern recruited the same structure that mediates temporal sequencing.

Unlike previous studies (Braun et al., 1997; Stager et al., 2003; Toyomura et al., 2011, 2015), AWS did not exhibit significantly increased activation in the *rhythm* condition compared to the *normal* condition. The most consistent finding from these studies was that both groups showed increased activation in bilateral auditory regions during isochronously paced speech and that AWS showed greater increases in the basal

ganglia. In the present study, the lack of clear between-condition effects within the AWS or between the AWS and ANS group may be due to more individual variability for AWS than ANS for this contrast. Future work is needed to determine whether this within-group variability is driving the null findings in the AWS group. Furthermore, Toyomura et al. (2011) found that while areas of the basal ganglia, left precentral gyrus, left SMA, left IFG, and left insula were less active in AWS during normal speech, activity in these areas increased to the level of ANS during isochronous speech. These results suggested that isochronously paced speech had a "normalizing" effect on activity in these regions, which differs from the present results.

There are methodological differences between the current work and similar studies that also could have impacted the results. In the current study, the rhythmic stimulus was presented prior to speaking regardless of the condition, unlike previous work in which the participant heard the stimulus while speaking and only during the *rhythm* condition (Toyomura et al. 2011). Thus, group effects reported by Toyomura and colleagues (2011) likely reflects auditory processing of the pacing stimulus in addition to any differences in speech motor processes. Second, our study sought to examine the rhythm effect when speech was produced at a conversational speaking rate. Previous studies used a metronome set at 92 – 100 beats per minute, considerably slower than the mean conversational rate in English (228 – 372 syllables per minute; Davidow, 2014; Pellegrino et al., 2011) and the rate observed in our study (approximately 207 syllables per minute). While Toyomura et al. (2011, 2015) instructed participants to speak at a similar rate during the *normal* condition (when previous studies had not), the slower

tempo overall may have led to increased auditory feedback processing. This could have modified the mechanisms by which ANS and AWS controlled their speech timing. Finally, only one of the previous studies accounted for disfluencies during the task in their imaging analysis (Stager et al., 2003), despite significant correlations with brain activation (Braun et al., 1997). However, given the small number of disfluencies in this and previous studies, this effect may have had a limited impact on the results.

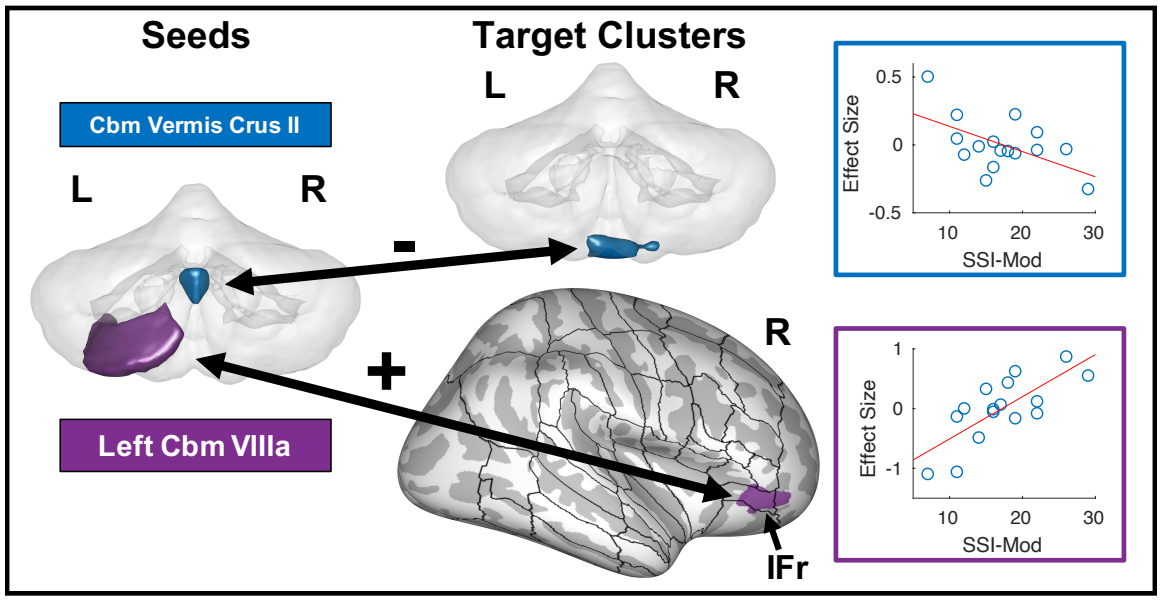

Figure 2.5. Two notable correlations of cerebellar functional connectivity (*normal* > *baseline*) with stuttering severity. Seed regions for these connections are indicated in the left side of the figure on a transparent 3D rendering of the cerebellum viewed posteriorly. Colors in the rest of the figure refer back to these seed regions. Target clusters are either displayed on the same transparent rendering of the cerebellum or projected onto an inflated surface of cerebral cortex, along with the full cortical region-of-interest (ROI) parcellation of the SpeechLabel atlas described in Cai, Tourville, et al. (2014). The '+' and '-' indicate positive and negative correlations, respectively. The right portion of the figure plots the beta estimates of the PPI regressors from individual AWS against stuttering severity. Roman numerals indicate cerebellar lobules. Full results of this analysis can be found in Figures 2.S14 and 2.S15 and Table 2.5. L = left, R = right, Cbm = cerebellum, IFr = inferior frontal gyrus pars orbitalis.

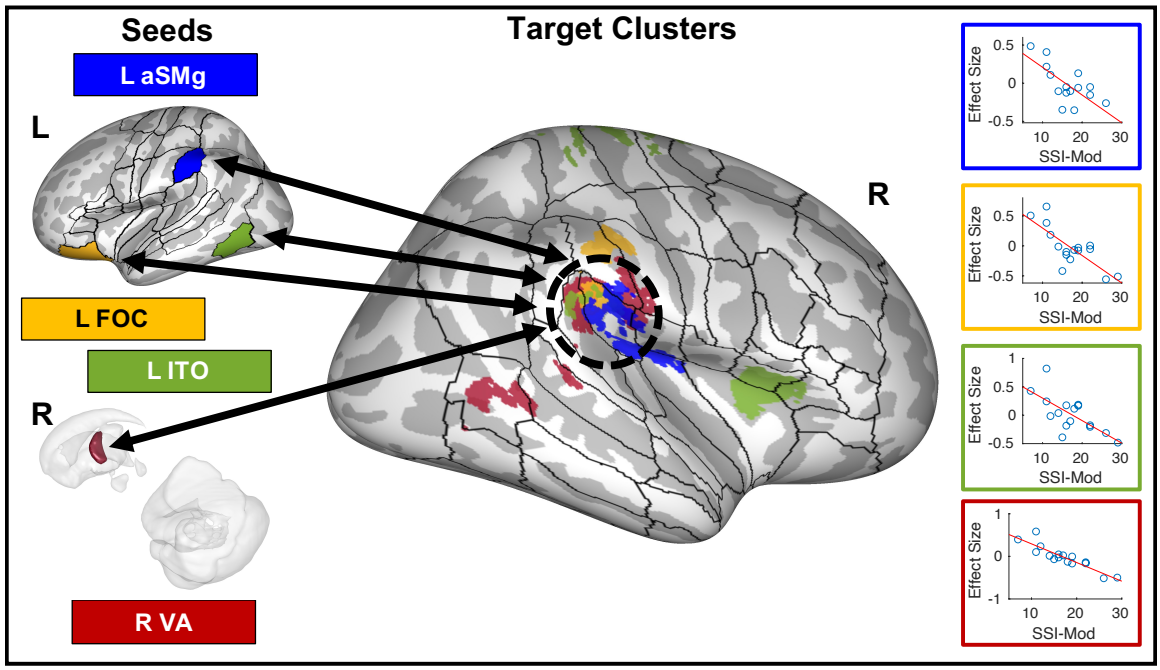

Figure 2.6. Correlations of functional connectivity (*rhythm* > *baseline*) between seeds regions-of-interest (ROIs) and right temporo-parietal junction with stuttering severity. Seed regions for these connections are indicated in the left side of the figure either on an inflated surface of the left cerebral cortex or on a transparent 3D rendering of right subcortical structures viewed medially. Colors in the rest of the figure refer back to these seed regions. Target clusters are projected onto an inflated surface of the right cerebral cortex, along with the full cortical ROI parcellation of the SpeechLabel atlas described in Cai, Tourville, et al. (2014). The black dashed oval indicates a rough border of the right temporo-parietal junction. The right portion of the figure plots the beta estimates of the PPI regressors from individual AWS against stuttering severity for each functional connection. Full results of this analysis can be found in Figures 2.S16 and 2.S17 and Table 2.5. L = left, R = right, aSMg = anterior supramarginal gyrus, FOC = fronto-orbital cortex, ITO = inferior temporo-occipital junction, VA = ventral anterior portion of the thalamus.

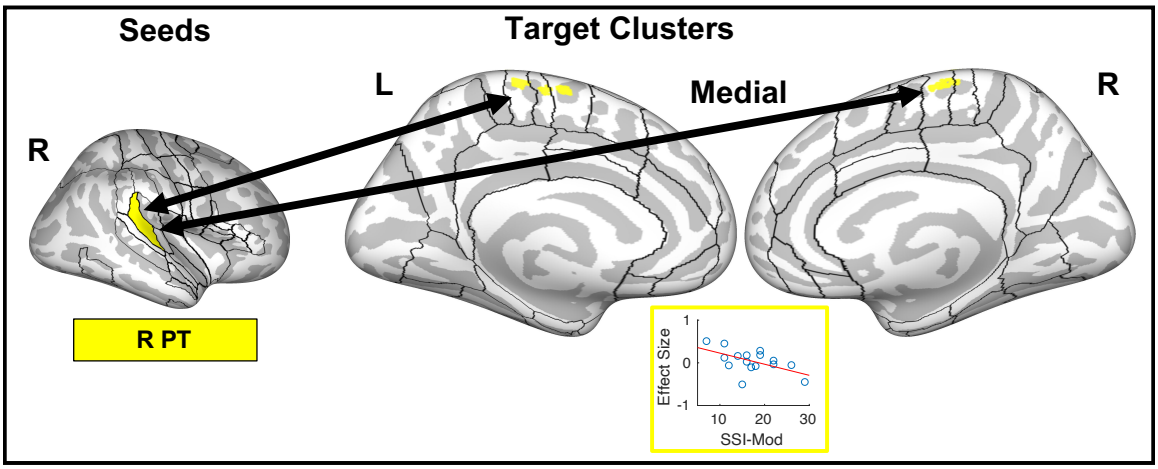

Figure 2.7. Correlations of functional connectivity (*rhythm* > *baseline*) between right PT and medial premotor cortex with stuttering severity. The seed region is indicated in the left side of the figure on an inflated surface of the right cerebral cortex. One target cluster (strattling the midline) is projected onto an inflated surface of the cerebral cortex, along with the full cortical ROI parcellation of the SpeechLabel atlas described in Cai, Tourville, et al. (2014). Below, the beta estimates of the PPI regressors from individual AWS are plotted against stuttering severity. Full results of this analysis can be found in Figures 2.S16 and 2.S17 and Table 2.5. L = left, R = right, PT = planum temporale.

Correlation Between Activation and Severity

The voxel/vertex-based analysis did not find significant correlations between Disfluency Rate and activation in the *Normal-Baseline* contrast. However, in the exploratory ROI analysis, activation in left VA thalamus and bilateral VL thalamus had among the strongest positive correlations with Disfluency Rate ($p < .005$; see Table 2.S4 and Figure 2.S9 for details). These nuclei are part of both the cortico-cerebellar and cortico-basal ganglia motor loops, and are structurally connected with premotor and primary motor areas (Barbas et al., 2013). As relays between subcortical structures and the cortex, increased activation for participants with a higher disfluency rate during the task may reflect greater reliance upon these modulatory pathways during speech. It is also

worth noting that with an exploratory threshold ($p < .05$, uncorrected), some ROIs follow similar patterns to previous literature. The propensity to stutter during the task, measured by Disfluency Rate, was associated with greater cortical activation in largely right hemisphere regions, and bilateral subcortical activation at uncorrected thresholds. The right-lateralized cortical associations in the present study may reflect increased compensatory activity in AWS (as in Braun et al., 1997; Cai, Tourville, et al., 2014; Kell et al., 2009; Preibisch et al., 2003; Salmelin et al., 2000). This is supported by the fact that fluency-inducing therapy lead to more left-lateralized activation (De Nil et al., 2003; Neumann et al., 2003, 2005), similar to that of neurotypical speakers. It should be noted that due to the low number of disfluencies exhibited during the task, determining a clear relationship between fluency and activation may not have been possible.

Functional connectivity between multiple seed ROIs and target clusters were significantly correlated with SSI-Mod. Given the large number of these significantly correlated connections, we focus here on what we consider to be the most salient findings; further detail regarding the full set of findings is provided in the supplementary materials.

When comparing the *normal* and *baseline* conditions, the negative association between SSI-Mod and the connection between cerebellar vermis crus II and midline inferior cerebellum indicates that less severe AWS have greater within-cerebellum connectivity. This fits conceptually with the result of increased connectivity within the cerebellum during the *rhythm* condition – both conditions associate the cerebellum with greater fluency. There was also a positive correlation between SSI-Mod and the

connection between L Cbm VIIIa and R IFr. The direction of this connection is surprising given previous work. For instance, Sitek et al. (2016) found a negative relationship between SSI scores and the connection between left cerebellum and IFr in resting state connectivity, and in Kell et al. (2018), there was hyperconnectivity between the cerebellum and left IFr in the comparison between overt and covert speech for recovered AWS. These findings suggested that greater fluency was associated with enhanced connections between these regions. However, the cerebellar regions involved in these connections were not as fine-grained as the ROI in the current study, and the specific tasks on which these connections were based were different than the *normal – baseline* comparison in the present study.

During the *rhythm* condition compared to *baseline*, multiple connections — between left fronto-orbital cortex (FOC), left aSMg, left ITO, and right VA seeds and overlapping clusters in right temporoparietal junction — were negatively correlated with SSI-Mod. Thus, more severe AWS had lower connection strengths compared to less severe AWS. In general, these connections support the idea that the right hemisphere is recruited to compensate for impaired left hemisphere processing (Braun et al., 1997; De Nil et al., 2000; Fox et al., 1996). Indeed, this temporo-parietal region was found to be hyperactive in a meta-analysis of stuttering neuroimaging studies (Belyk et al., 2015). The convergence of these connections specifically in the right temporo-parietal junction may imply association with this regions' role in responding to salient or unexpected events (Corbetta & Shulman, 2002). In the realm of speech production, these connections (especially with left aSMg) may reflect increased use of the somatosensory

feedback loop by less severe AWS to control speech during the *rhythm* condition (Elisa Golfinopoulos et al., 2011). One additional negative correlation worth mentioning for the *rhythm - baseline* contrast is the negative association between SSI-Mod and the connection between right PT and midline MC/PMC/SMA. In the auditory feedback loop as proposed by Tourville et al. (2008) and Guenther (2016), sensory state, target, and error maps send error signals to right premotor cortex to generate corrective motor commands. Connectivity between right PT and medial premotor regions may then reflect an interface between these sensory feedback loops and the SMA-basal ganglia "internal" timing system which is disrupted in stuttering (Chang & Guenther, 2020). More fluent speakers may use this connection to a greater extent in order to resolve conflicts between competing motor programs (Guenther, 2016).

Limitations

One potential limitation to this study was that trial types were pseudo-randomly presented within a given run. Since the sequence of tones was presented before every trial and the participants did not know the condition ahead of time, participants needed to refrain from speaking at the pace of the tone sequence during normal trials. This process of ignoring the tone sequence during production of their sentence may have recruited additional brain areas for the *normal* condition only, potentially confounding the neural response. However, presenting the tone sequence before every trial was done specifically to eliminate the confound of tone sequence auditory processing found it previous studies. Even if the *rhythm* trials and *normal* trials were presented in a blocked fashion such that

participants knew the condition ahead of time, they would still either have to ignore the tones on the *normal* trials or risk the confound of attending to the tones in one condition and not the other. As it is, there are a few indications that this contrast reflects the difference in speaking styles between conditions. First, the reduction in disfluencies in the "rhythm" condition compared to the "normal" condition shows that the fluency-enhancing effect took place. Thus, any neural changes between the conditions could plausibly reflect this effect. Second, the pattern of pacing tones that participants hear is quite simple and is the same throughout the experiment. Furthermore, the task is well-practiced by the participants from a similar behavioral experiment that they participated in prior to the fMRI task. Thus, listening to the tones before each trial is merely a reminder of the pace rather than something that requires significant attentional resources. Finally, all significant corrected results and most exploratory results from the *rhythm - normal* activation contrast demonstrated greater activation in the *rhythm* condition. If there were additional areas recruited for ignoring a rhythm when producing speech, they would have probably led to greater activation during the normal condition. That being said, if a given region mediated both isochronous speech production and ignoring an external pacing stimulus, the direction of activation change between conditions would be mixed which could potentially lead to false negative findings. It is also possible that the reduced connectivity between left anterior prefrontal cortex and left inferior cerebellum in the rhythm may reflect this additional "ignoring" process during the *normal* condition. Balancing the need to avoid the confounds of the auditory stimulus presentation and the process of ignoring the tones in the unpaced condition is a challenge that will need to be

addressed in future work.

In addition, as mentioned in the data acquisition section above, while the sparse sampling paradigm allows participants to hear themselves speak without addition scanner noise and decouples the functional acquisition from task-related motion, collecting a single data point per trial poses some challenges to interpretation of the results. One challenge is the assumption that the single acquisition captures the peak of the BOLD response, which has been shown to vary across brain regions and participants (Handwerker et al., 2004; Janssen & Mendieta, 2020). This is an issue common to many sparse sampling paradigms and implies that because the peak response of some brain regions may not be captured in this single acquisition, there is less power to detect significant results in these regions. For the present study, because of the prolonged duration of the sentence production (approximately 2 seconds) and the relatively slow acquisition of 2.47 seconds, the single acquisition would provide a broad sampling of the hemodynamic response across a range of different delay times. Furthermore, computing functional connectivity from sparsely sampled data has much less power and temporal resolution than for continuous data. This could negatively impact the detectability of significant connections that would otherwise be found with more scans and a greater sampling of timepoints that include BOLD response peaks from a broader range of regions and participants. Future studies investigating functional connectivity of speaking tasks that rely on auditory processing and speech production could be improved by acquiring more samples (see Perrachione & Ghosh [2013] for a discussion of these issues for task activation).

Finally, the current results are not consistent with a recent meta-analysis examining activation differences between AWS and ANS (Belyk et al., 2015, 2017) which found that AWS consistently had overactivation in right hemisphere cortical structures, and underactivation in left hemisphere structures, especially in motor and premotor areas. However, the present study's exploratory analysis suggested that AWS had decreased activation in left frontal operculum during the *rhythm* condition as compared to the ANS group. Previous work has shown gray matter and white matter anomalies in and near left IFG (Beal et al., 2013, 2015; Chang et al., 2008, 2011; Kell et al., 2009; Lu et al., 2012), which may be related to this under-activation. Based on the exploratory nature of these findings, future work as well as meta-analytic testing is needed to determine whether these are true population differences.

Conclusion

In this study, we examined brain activation patterns that co-occur with the introduction of an external pacing stimulus. We found that AWS showed an overall decrease in disfluencies during this condition, as well as functional connectivity changes both within the cerebellum and between the cerebellum and prefrontal cortex. Involvement of these structures suggests that isochronously paced speech activates compensatory timing systems and potentially modulates feedback control and attentional systems. This study provides greater insight into the network of brain areas that either support (or respond to) fluency in relation to the rhythm effect and its correspondence to longer-term fluency provided through natural compensation. It is our hope that in

conjunction with the large body of work already published on fluency-enhancing techniques and future studies with more focused analyses, the field will come to a better understanding of the pathophysiology of stuttering and fluency, and that this information will be used to provide more targeted treatments and, ultimately, improve quality of life for those who stutter.

Supplementary Materials

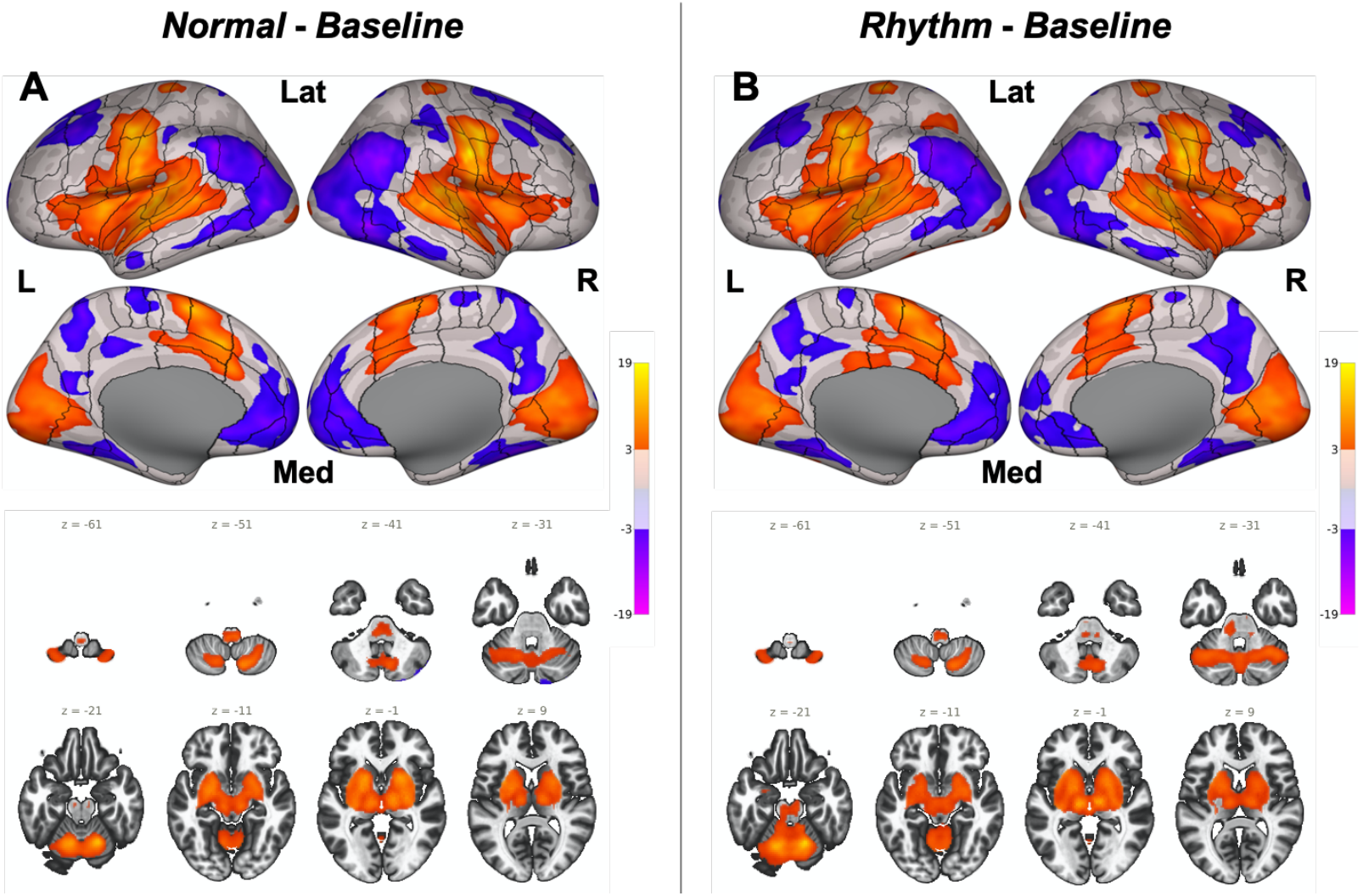

Figure 2.S1. Significant clusters for the (A) *normal - baseline* or (B) *rhythm - baseline* contrasts collapsed across both groups (vertex-wise $p < .01$, cluster-wise $p_{FDR} < .05$). Cortical results (top) are displayed on an inflated cortical surface, along with the full cortical ROI parcellation of the SpeechLabel atlas described in Cai, Tourville, et al. (2014). Subcortical and cerebellar results (bottom) are displayed on an axial slice series of a template brain. Color shading indicates t-values. L = left, R = right, Lat = lateral surface, Med = medial surface.

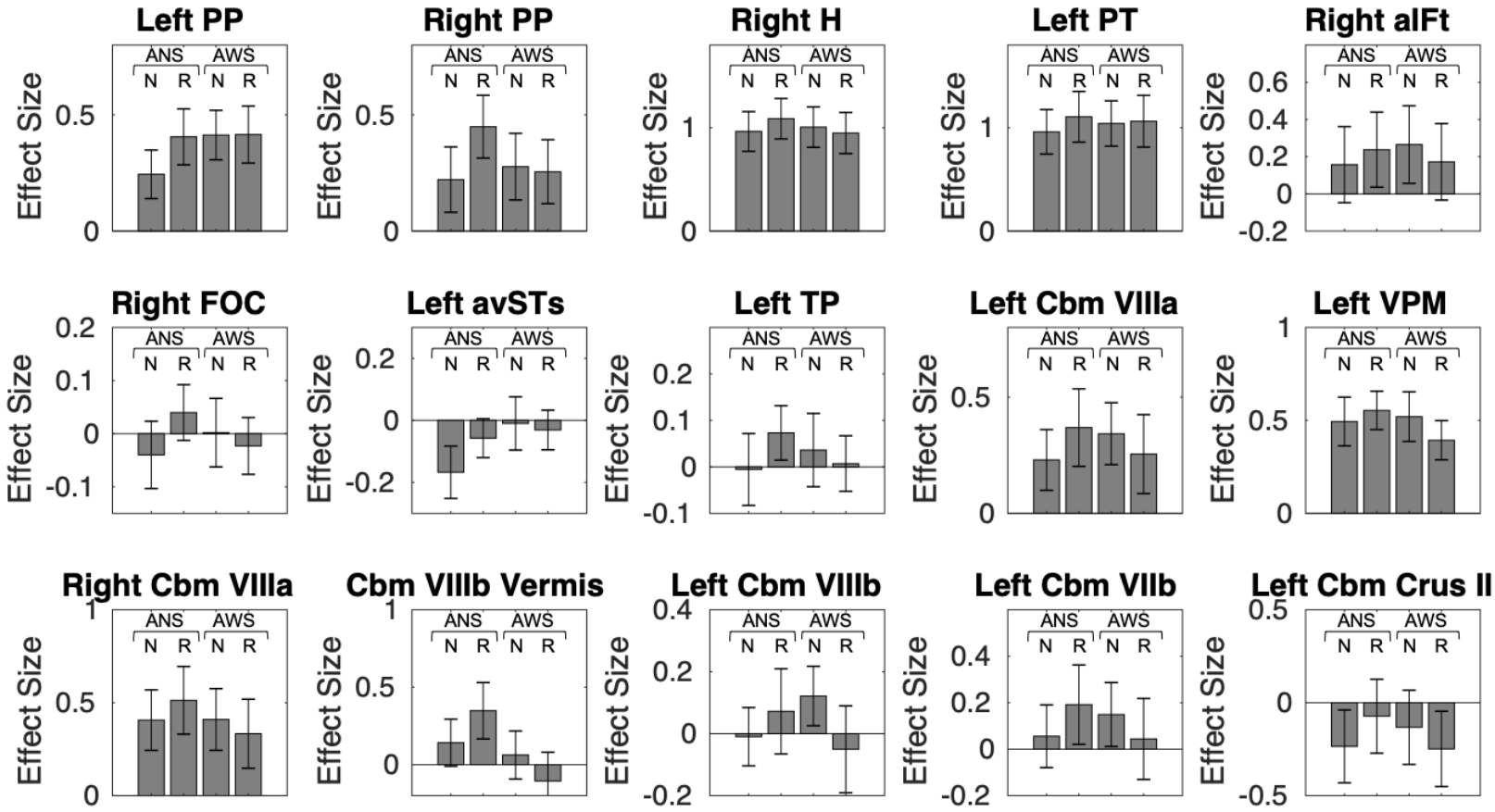

Figure 2.S2. Individual group and condition effects from the exploratory regions-of-interest that had a significant interaction between group and condition. See Table 2.S2 for statistics. PP = planum polare, H = Heschl's gyrus, PT = planum temporale, aIFt = anterior inferior frontal gyrus pars triangularis, FOC = fronto-orbital cortex, avSTs = anterior ventral superior temporal sulcus, TP = temporal pole, Cbm = cerebellum, VPM = ventral postero-medial portion of the thalamus, N = *Normal - Baseline* condition, R = *Rhythm - Baseline* condition, ANS = adults who do not stutter, AWS = adults who stutter. Error bars indicate 90% confidence intervals.

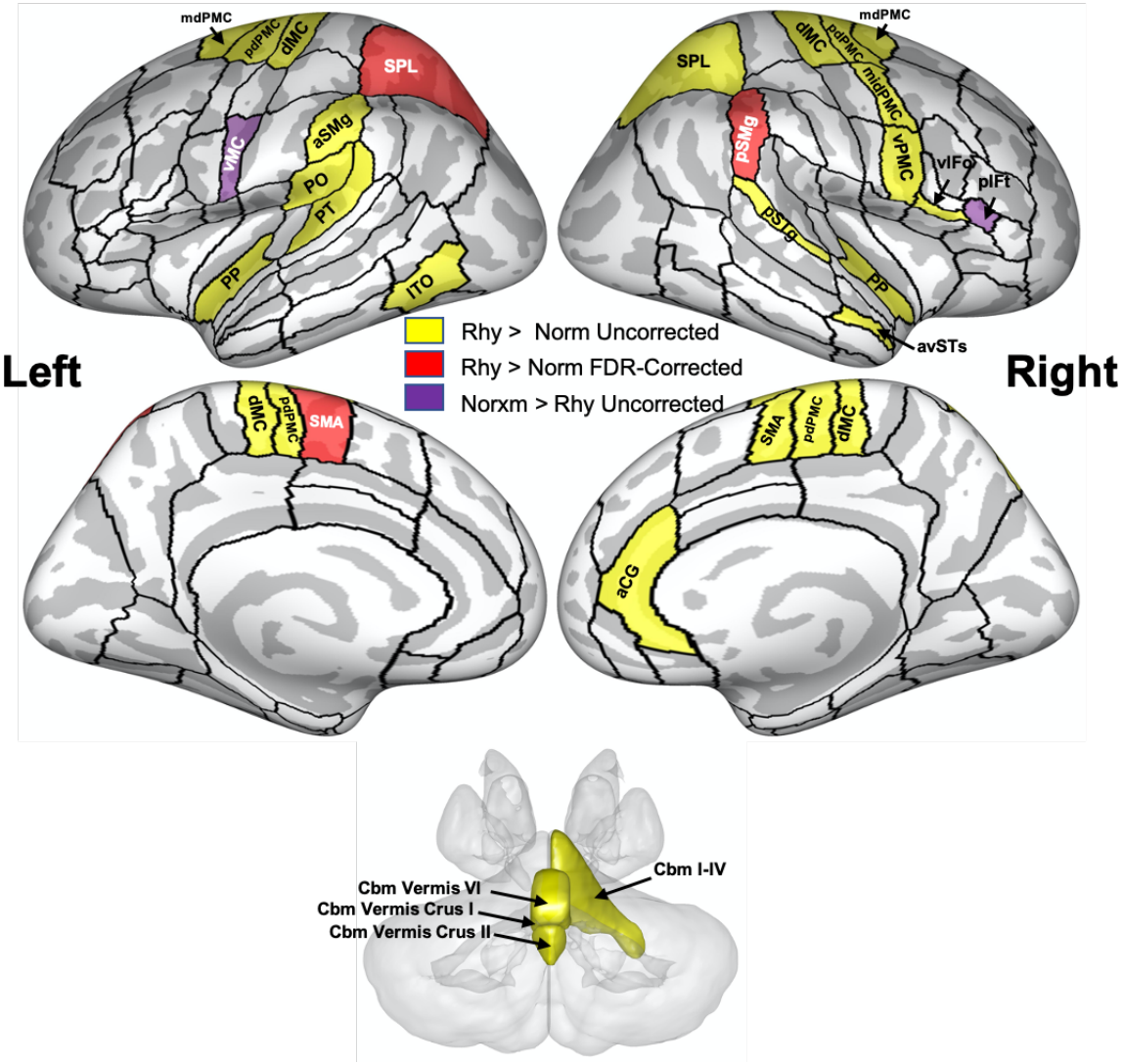

Figure 2.S3. Regions-of-interest (ROIs) significantly more active during the *rhythm* condition than the *normal* condition for ANS and AWS combined in the exploratory analysis ($p < .05$) are highlighted in yellow and plotted on an inflated cortical surface. ROIs highlighted in red and labeled reached significance at a stricter threshold of $p_{FDR} < .05$. ROIs highlighted in purple were significantly more active during the *normal* condition than the *rhythm* condition ($p < .05$). Black outlines indicate cortical ROIs used in the exploratory analysis. FDR = false discovery rate, aCG = anterior cingulate gyrus, aSMg = anterior supramarginal gyrus, avSTs = anterior ventral superior temporal sulcus, Cbm = cerebellum, dMC = dorsal primary motor cortex, ITO = inferior temporo-occipital cortex, mdPMC = middle dorsal premotor cortex, midPMC = middle premotor cortex, pdPMC = posterior dorsal premotor cortex, pIFt = posterior inferior frontal gyrus pars triangularis, PO = parietal operculum, PP = planum polare, pSMg = posterior supramarginal gyrus, pSTg = posterior superior temporal gyrus, PT = planum temporale, SMA = supplementary motor area, SPL = superior parietal lobule, vIFo = ventral inferior frontal gyrus pars opercularis, vMC = ventral primary motor cortex, vPMC = ventral premotor cortex.

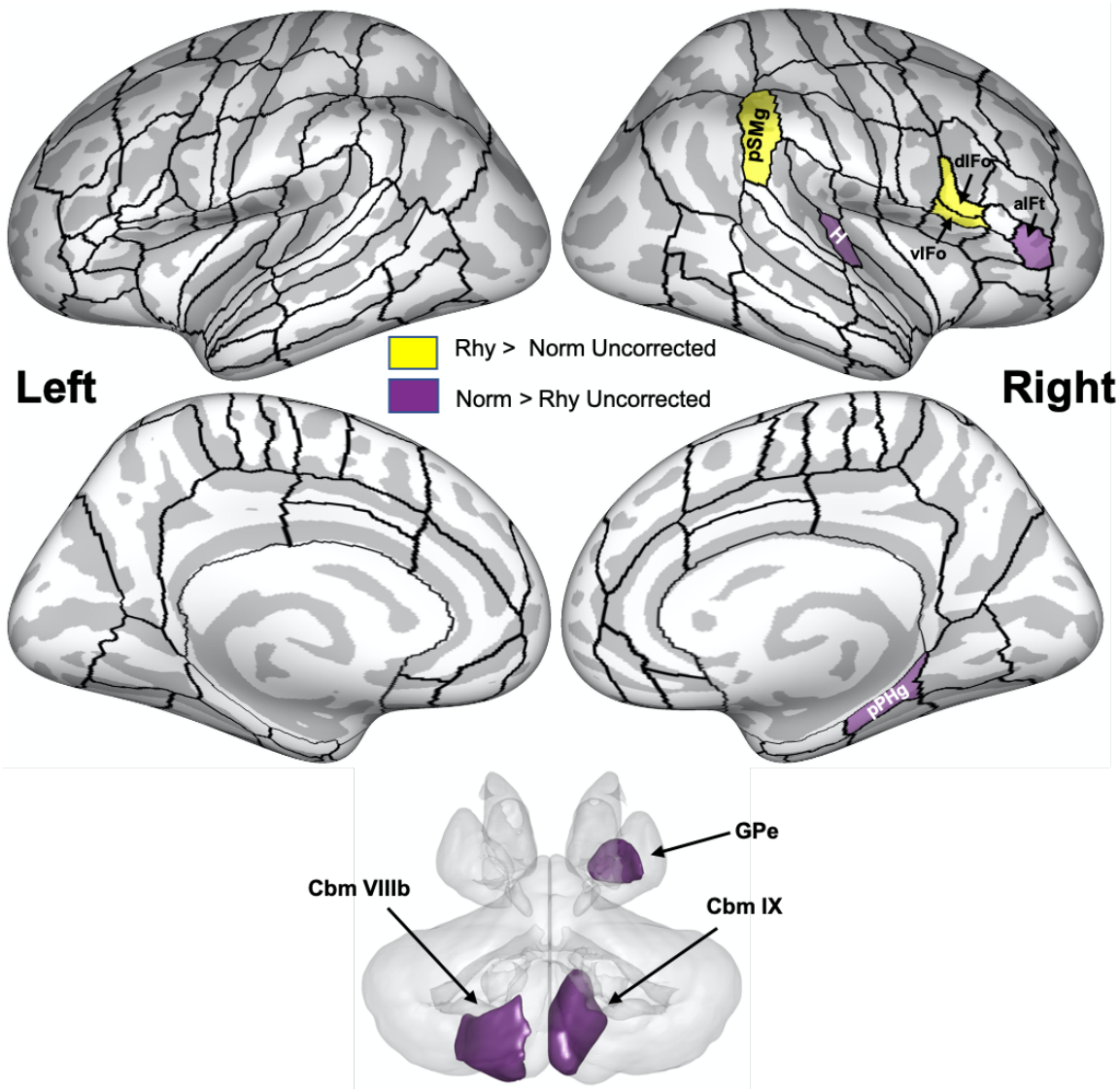

Figure 2.S4. Regions-of-interest (ROIs) significantly more active during the *rhythm* condition than the *normal* condition for AWS in the exploratory analysis ($p < 0.05$) are highlighted in yellow and plotted on an inflated cortical surface. ROIs highlighted in purple were significantly more active during the *normal* condition than the *rhythm* condition ($p < .05$). Black outlines indicate cortical ROIs used in the exploratory analysis. aIFt = anterior inferior frontal gyrus pars triangularis, dIFo = dorsal inferior frontal gyrus pars opercularis, Cbm = cerebellum, GPe = external portion of the globus pallidus, H = Heschl's gyrus, pPHg = posterior parahippocampal gyrus, pSMg = posterior supramarginal gyrus, vIFo = ventral inferior frontal gyrus pars opercularis.

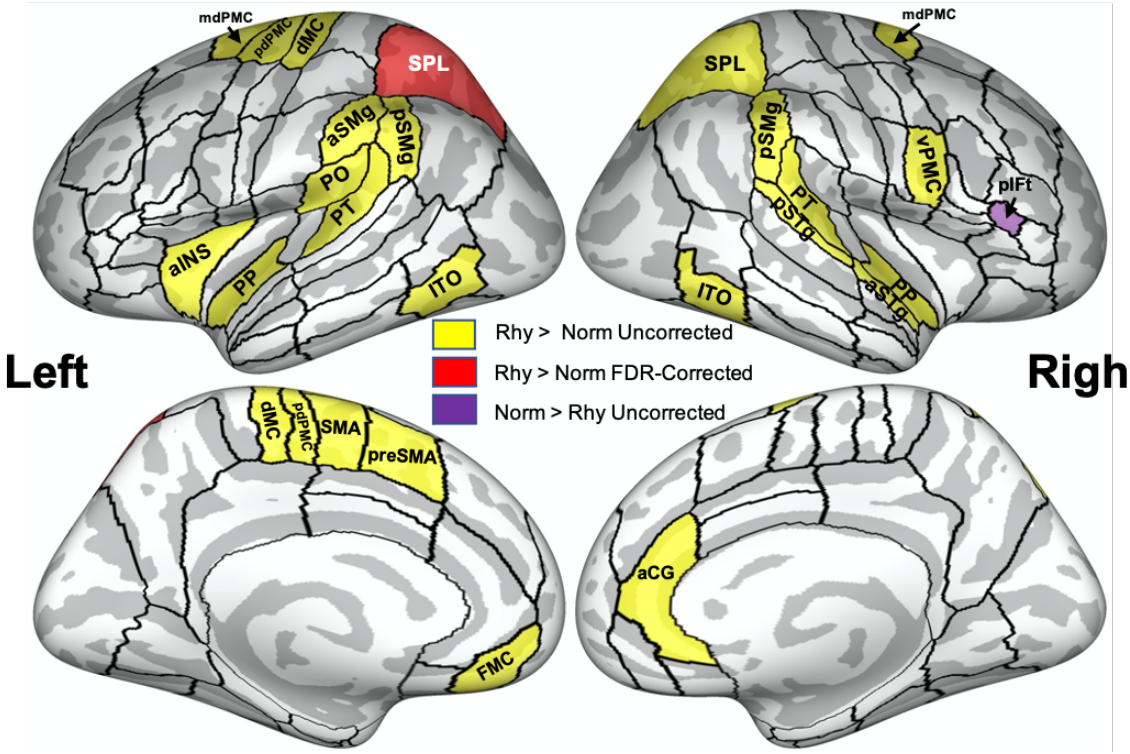

Figure 2.S5. Regions-of-interest (ROIs) significantly more active during the *rhythm* condition than the *normal* condition for ANS in the exploratory analysis ($p < .05$) are highlighted in yellow and plotted on an inflated cortical surface. ROIs highlighted in red and labeled reached significance at a stricter threshold of $p_{FDR} < .05$. ROIs highlighted in purple were significantly more active during the *normal* condition than the *rhythm* condition ($p < .05$). Black outlines indicate cortical ROIs used in the exploratory analysis. FDR = false discovery rate, aCG = anterior cingulate gyrus, aINS = anterior insula, aSMg = anterior supramarginal gyrus, aSTg = anterior superior temporal gyrus, dMC = dorsal primary motor cortex, FMC = fronto-medial cortex, ITO = inferior temporo-occipital cortex, mdPMC = middle dorsal premotor cortex, pdPMC = posterior dorsal premotor cortex, pIFt = posterior inferior frontal gyrus pars triangularis, PP = planum polare, preSMA = presupplementary motor area, pSMg = posterior supramarginal gyrus, pSTg = posterior superior temporal gyrus, PT = planum temporale, SMA = supplementary motor area, SPL = superior parietal lobule, vPMC = ventral premotor cortex.

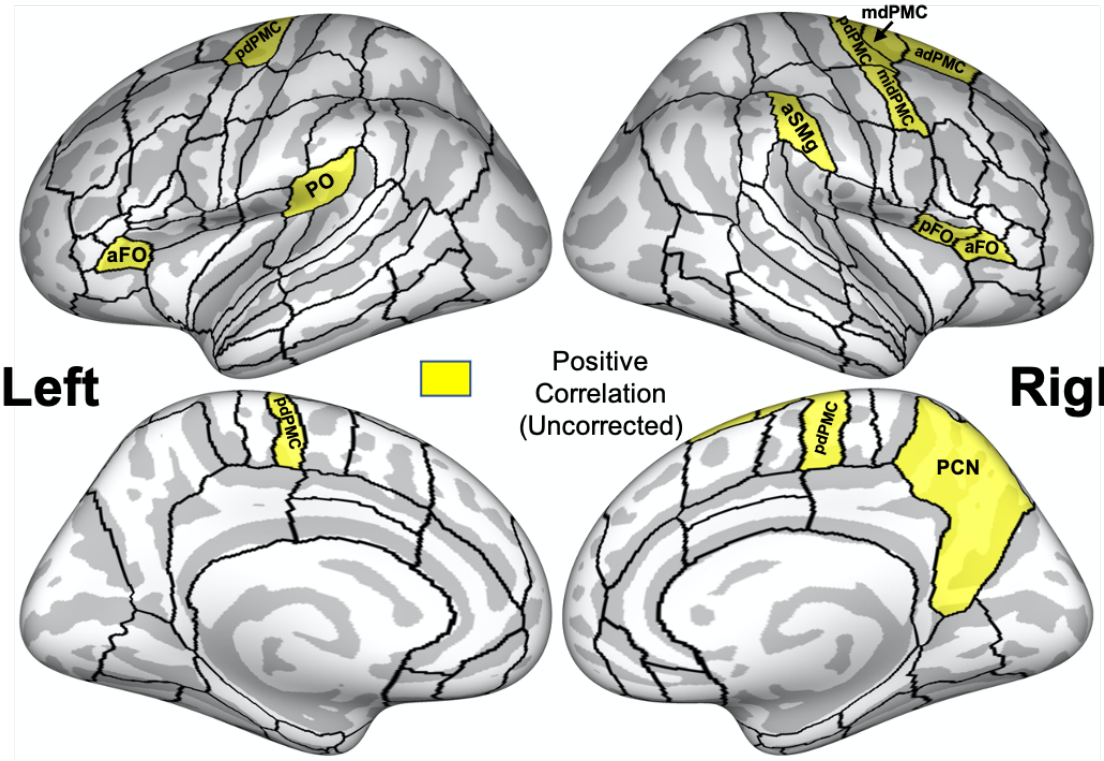

Figure 2.S6. Exploratory regions-of-interest (ROIs) with a positive correlation between *normal – baseline* activation and SSI-Mod in AWS ($p < .05$) are highlighted in yellow and plotted on an inflated cortical surface. Black outlines indicate cortical ROIs used in the exploratory analysis. adPMC = anterior dorsal premotor cortex, aFO = anterior frontal operculum, aSMg = anterior supramarginal gyrus, mdPMC = middle dorsal premotor cortex, PCN = precuneus, pdPMC = posterior dorsal premotor cortex, pFO = posterior frontal operculum, PO = parietal operculum.

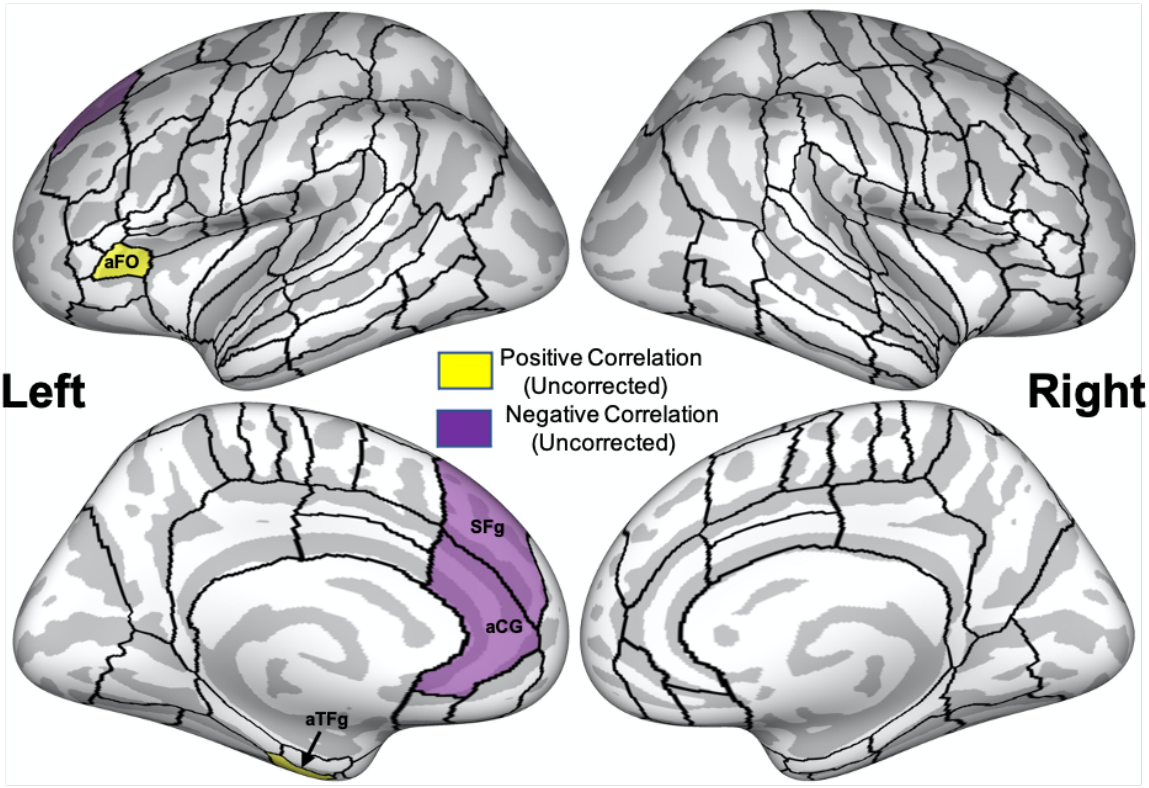

Figure 2.S7. Exploratory regions-of-interest (ROIs) with a positive correlation between *rhythm – baseline* activation and SSI-Mod in AWS ($p < .05$) are highlighted in yellow and plotted on an inflated cortical surface. ROIs highlighted in purple were negatively correlated with SSI-Mod ($p < .05$). Black outlines indicate cortical ROIs used in the exploratory analysis. aCG = anterior cingulate gyrus, aFO = anterior frontal operculum, aTFg = anterior temporal fusiform gyrus, SFg = superior frontal gyrus.

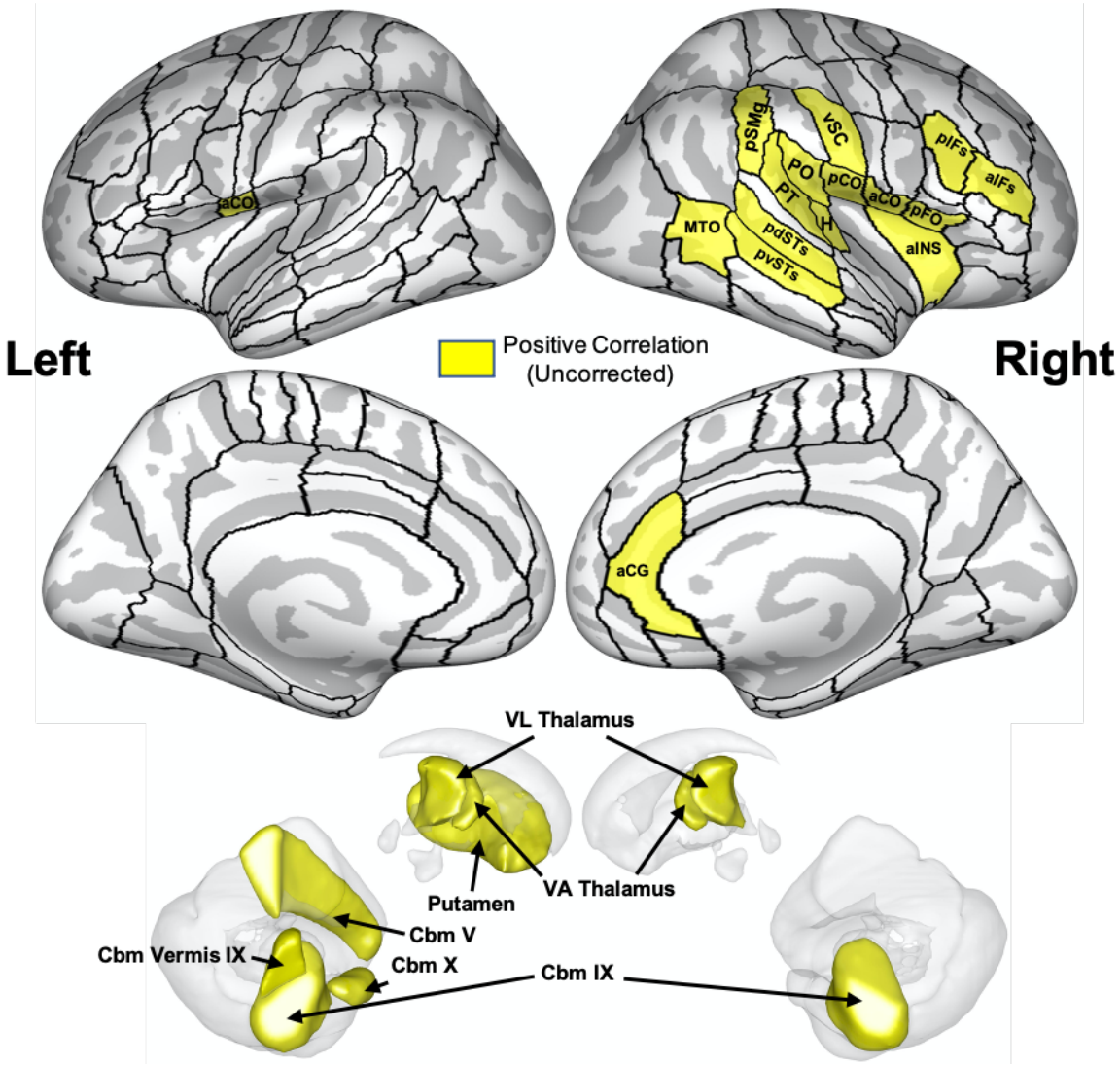

Figure 2.S8. Exploratory regions-of-interest (ROIs) with a positive correlation between *normal – baseline* activation and Disfluency Rate in AWS ($p < .05$) are highlighted in yellow and plotted on an inflated cortical surface. Black outlines indicate cortical ROIs used in the exploratory analysis. aCG = anterior cingulate gyrus, aCO = anterior central operculum, aIFs = anterior inferior frontal sulcus, aINS = anterior insula, Cbm = cerebellum, H = Heschl's gyrus, MTO = middle temporo-occipital cortex, pCO = posterior central operculum, pdSTs = posterior dorsal superior temporal sulcus, pFO = posterior frontal operculum, pIFs = posterior inferior frontal sulcus, PO = parietal operculum, pSMg = posterior supramarginal gyrus, PT = planum temporale, pvSTs = posterior ventral superior temporal sulcus, VA = ventral anterior, VL = ventral lateral, vSC = ventral primary somatosensory cortex.

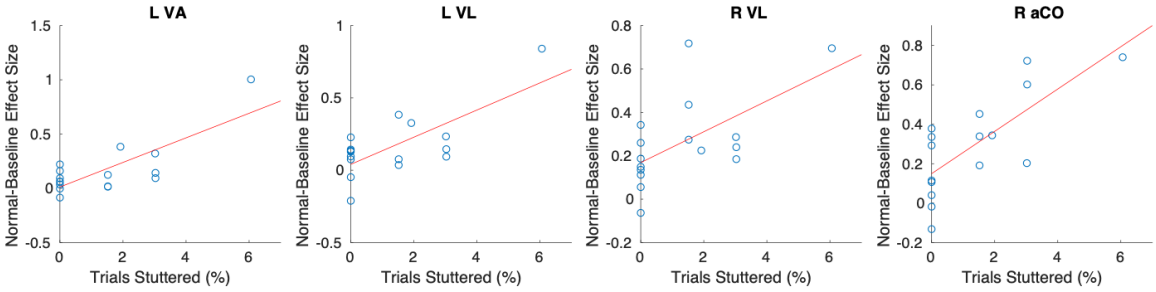

Figure 2.S9. Across-subjects correlation between *normal – baseline* activation and Disfluency Rate for AWS in four highly significant exploratory regions-of-interest (ROIs; *p* < .005). Blue circles indicate individual AWS. L = left, R = right, VA = ventral anterior portion of the thalamus, VL = ventral lateral portion of the thalamus, aCO = anterior central operculum.

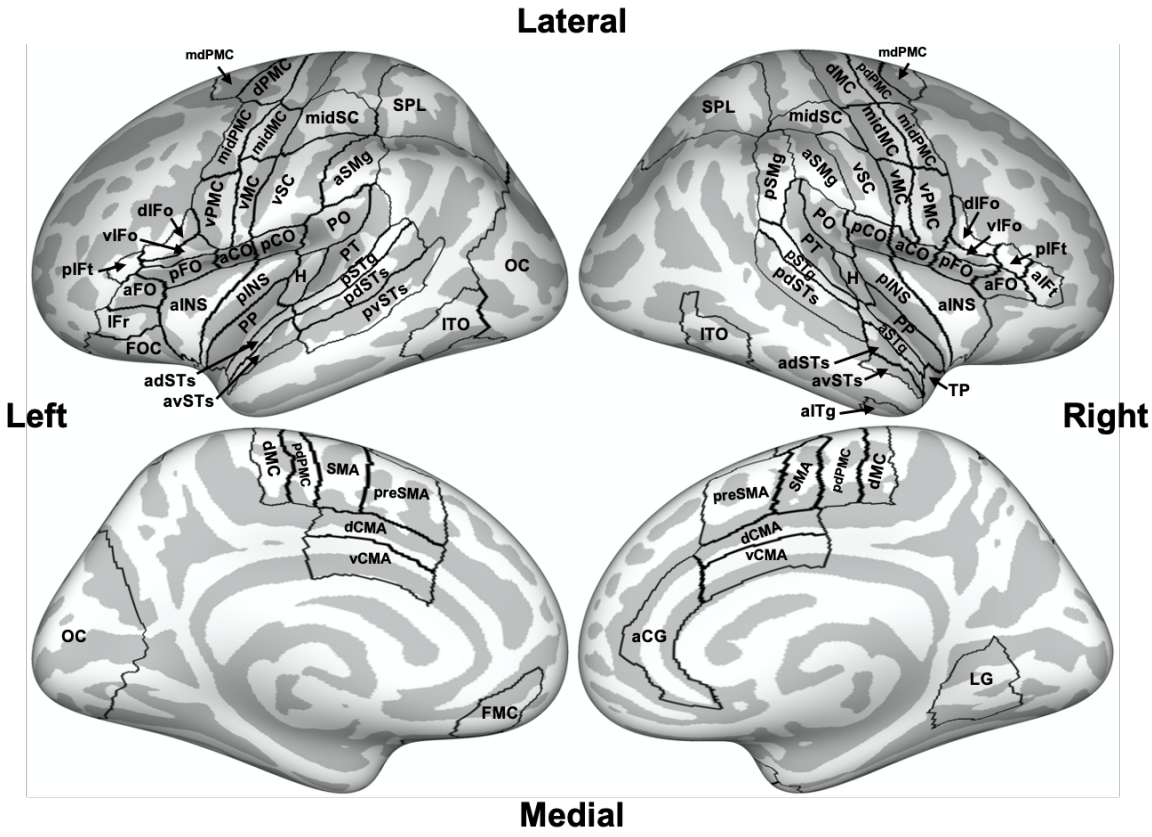

Figure 2.S10. Cortical regions-of-interest included as seed regions in the functional connectivity analyses. aIFt = anterior inferior frontal gyrus pars triangularis, pIFt = posterior inferior temporal gyrus pars triangularis, aFO = anterior frontal operculum, pFO = posterior frontal operculum, dIFo = dorsal inferior frontal gyrus pars opercularis, vIFo = ventral inferior frontal gyrus pars opercularis, aINS = anterior insula, pINS = posterior insula, vPMC = ventral premotor cortex, midPMC = mid premotor cortex, pdPMC =

posterior dorsal premotor cortex, mdPMC = middle dorsal premotor cortex, vMC = ventral primary motor cortex, midMC = middle primary motor cortex, dMC = dorsal primary motor cortex, vSC = ventral primary somatosensory cortex, midSC = middle primary somatosensory cortex, aCO = anterior central operculum, pCO = posterior central operculum, PO = parietal operculum, aSMg = anterior supramarginal gyrus, pSMg = posterior supramarginal gyrus, PT = planum temporale, H = Heschl's gyrus, PP = planum polare, aSTg = anterior superior temporal gyrus, pSTg = posterior superior temporal gyrus, adSTs = anterior dorsal superior temporal sulcus, avSTs = anterior ventral superior temporal sulcus, pdSTs = posterior dorsal superior temporal sulcus, pvSTs = posterior ventral superior temporal sulcus, TP = temporal pole, aITg = anterior inferior temporal gyrus, ITO = inferior temporo-occipital cortex, SPL = superior parietal lobule, OC = occipital cortex, LG = lingual gyrus, SMA = supplementary motor area, preSMA = presupplementary motor area, dCMA = dorsal cingulate motor area, vCMA = ventral cingulate motor area, aCG = anterior cingulate gyrus, IFr = inferior frontal gyrus pars orbitalis, FOC = fronto-orbital cortex, FMC = fronto-medial cortex.

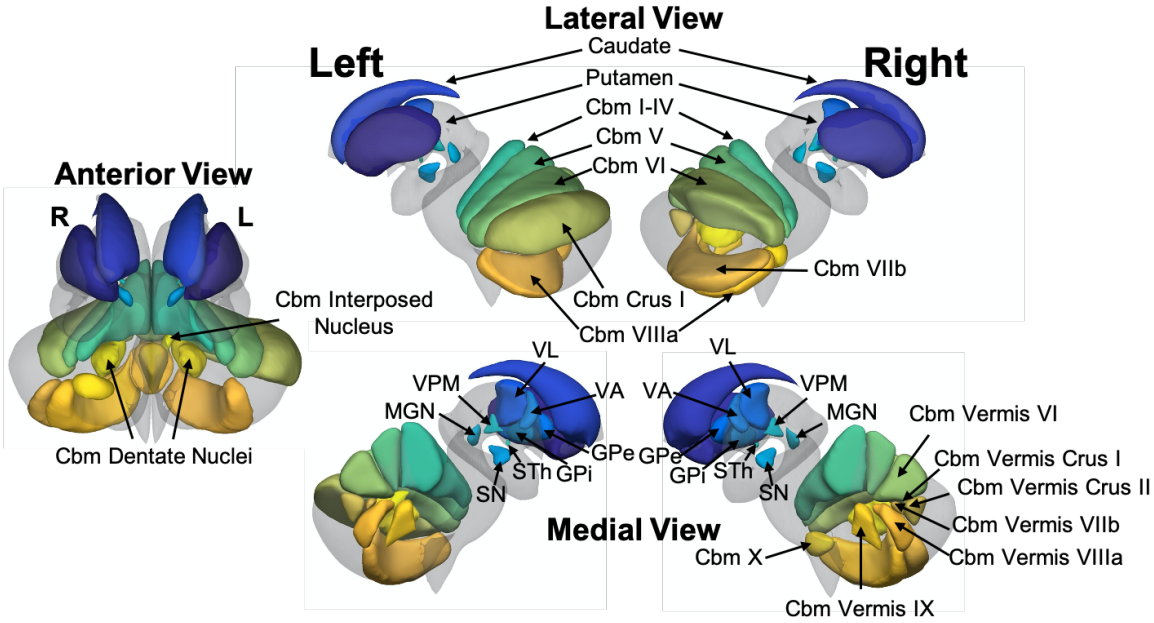

Figure 2.S11. Subcortical regions-of-interest included as seed regions in the functional connectivity analyses. L = left, R = right, Cbm = cerebellum, VA = ventral anterior portion of the thalamus, VL = ventral lateral portion of the thalamus, VPM = ventral posteromedial thalamic nucleus, MGN = medial geniculate nucleus of the thalamus, GPe = external portion of the globus pallidus, GPi = internal portion of the globus pallidus, STh = subthalamic nucleus, SN = substantia nigra.

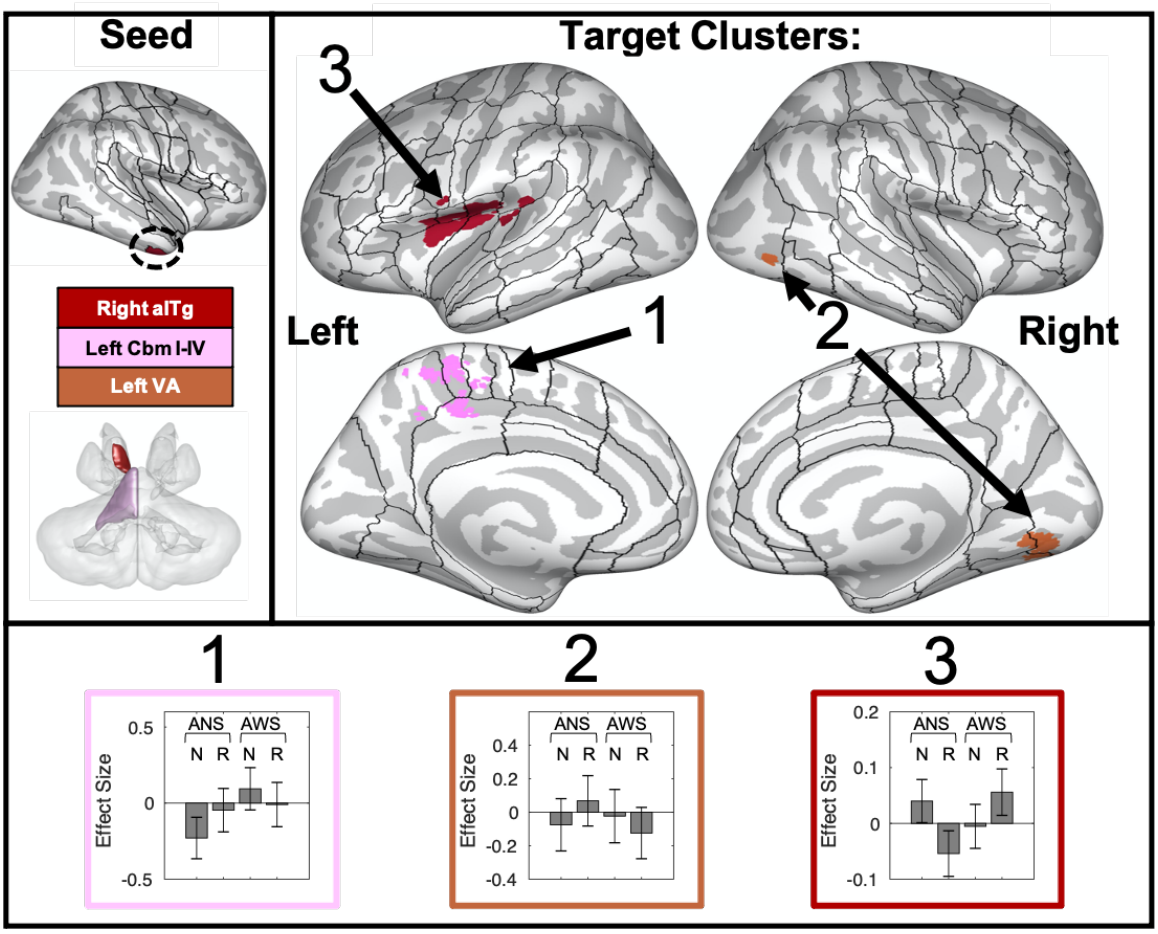

Figure 2.S12. A summary of functional connections that show significant interactions between group and condition. Seed regions for these connections are indicated in the upper left panel on an inflated right hemisphere cortical surface (top; ROIs are as in Figure 2.S10) or on a transparent 3D rendering of the cerebellum and subcortical structures viewed posteriorly (bottom). Colors in the rest of the figure refer back to these seed regions. Three target clusters (representing three distinct connections) are displayed in the upper right portion of the figure. Target clusters are projected onto an inflated surface of cerebral cortex, along with the full cortical ROI parcellation of the SpeechLabel atlas described in Cai, Tourville, et al. (2014). The bottom portion of the figure shows the connectivity effect sizes for each connection in the *normal* and *rhythm* conditions, separately for each group. Error bars indicate 90% confidence intervals. N = normal, R = rhythm, aITg = anterior inferior temporal gyrus, Cbm = cerebellum, VA = ventral anterior portion of the thalamus.

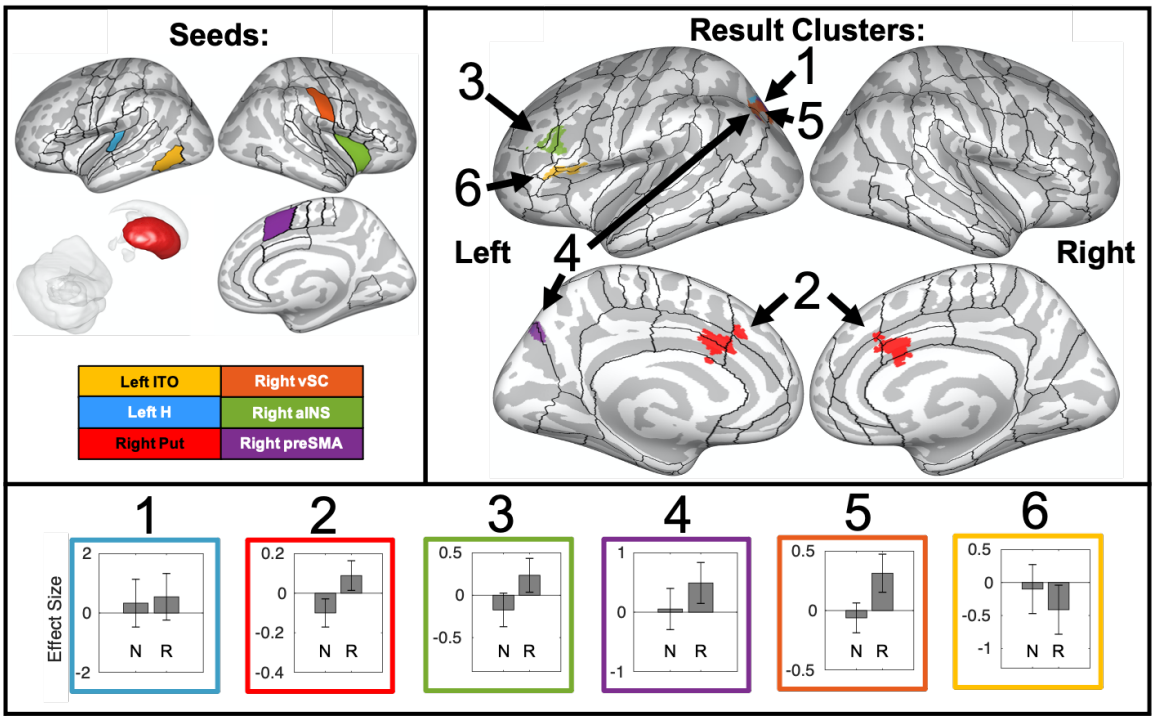

Figure 2.S13. A summary of functional connections that are significantly different between the *normal* and *rhythm* conditions in ANS. Seed regions for these connections are indicated in the upper left panel either on an inflated cortical surface (top; ROIs are as in Figure 2.S10) or on a transparent 3D rendering of the left hemisphere subcortical structures viewed from the right (bottom). Colors in the rest of the figure refer back to these seed regions. Six target clusters (representing 6 distinct connections) are displayed in the upper right portion of the figure. These clusters are projected onto an inflated surface of cerebral cortex, along with the full cortical ROI parcellation of the SpeechLabel atlas described in Cai, Tourville, et al. (2014). The bottom portion of the figure shows the connectivity effect sizes in the *normal* and *rhythm* conditions for each connection. Error bars indicate 90% confidence intervals. N = normal, R = rhythm, ITO = inferior temporo-occipital cortex, H = Heschl's gyrus, Put = putamen, vSC = ventral primary somatosensory cortex, aINS = anterior insula, preSMA = presupplementary motor area.

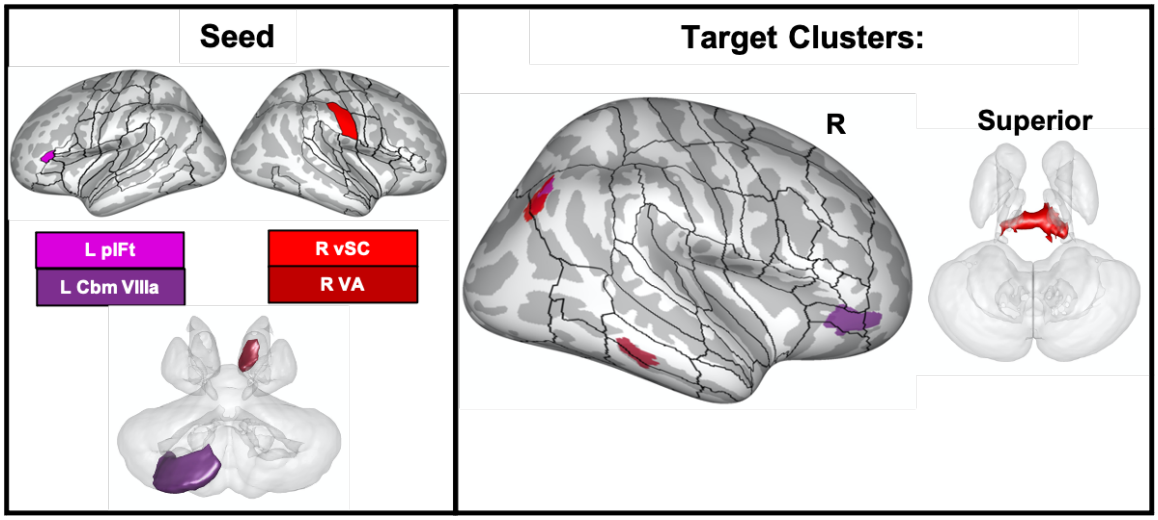

Figure 2.S14. A summary of functional connectivity (*normal* - *baseline*) positively correlated with stuttering severity in AWS. Seed regions for these connections are indicated in the left panel either on an inflated cortical surface (top; ROIs are as in Figure 2.S10) or on a transparent 3D rendering of subcortical structures viewed posteriorly (bottom). Colors in the rest of the figure refer back to these seed regions. Five target clusters (representing five distinct connections) are displayed in the right portion of the figure. These clusters are either projected onto an inflated surface of the right hemisphere cerebral cortex, along with the cortical ROI parcellation of the SpeechLabel atlas described in Cai, Tourville, et al. (2014), or plotted on a 3D rendering of subcortical structures (viewed superiorly). L = left, R = right, pIFt = posterior inferior frontal gyrus pars triangularis, Cbm = cerebellum, vSC = ventral primary somatosensory cortex, VA = ventral anterior portion of the thalamus.

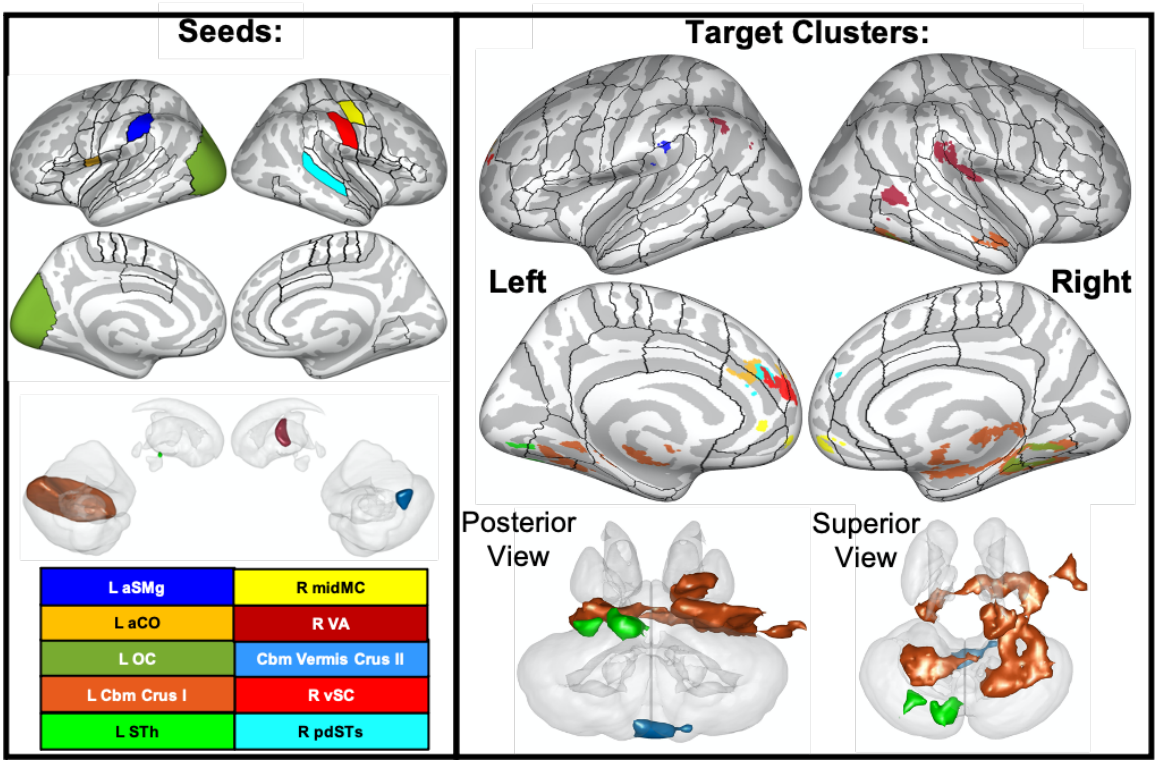

Figure 2.S15. A summary of functional connectivity (*normal* - *baseline*) negatively correlated with stuttering severity in AWS. Seed regions for these connections are indicated in the left panel either on an inflated cortical surface (top; ROIs are as in Figure 2.S10) or on a transparent 3D rendering of subcortical structures in each hemisphere viewed medially (bottom). Colors in the rest of the figure refer back to these seed regions. Thirteen target clusters (representing thirteen distinct connections) are displayed in the right portion of the figure. These clusters are either projected onto an inflated cortical surface, along with the cortical ROI parcellation of the SpeechLabel atlas described in Cai, Tourville et al. (2014), or plotted on a 3D rendering of subcortical structures viewed either posteriorly (left) or superiorly (right). L = left, R = right, aSMg = anterior supramarginal gyrus, aCO = anterior central operculum, OC = occipital cortex, Cbm = cerebellum, STh = subthalamic nucleus, midMC = middle primary motor cortex, VA = ventral anterior portion of the thalamus, vSC = ventral primary somatosensory cortex, pdSTs = posterior dorsal superior temporal sulcus.

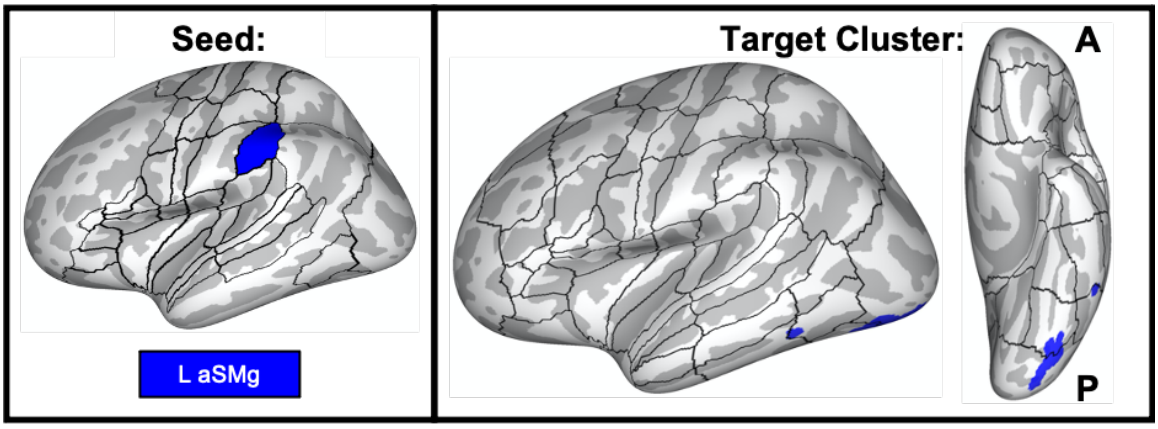

Figure 2.S16. A summary of functional connectivity (*rhythm* - *baseline*) positively correlated with stuttering severity in AWS. The seed region for this connection is indicated in the left panel on an inflated left hemisphere cortical surface (ROIs are as in Figure 2.S10). Colors in the rest of the figure refer back to this region. One target cluster is displayed in the right portion of the figure, projected onto an inflated surface of the left hemisphere cerebral cortex, and viewed from the left (left) or inferiorly (right). L = left, A = anterior, P = posterior, aSMg = anterior supramarginal gyrus.

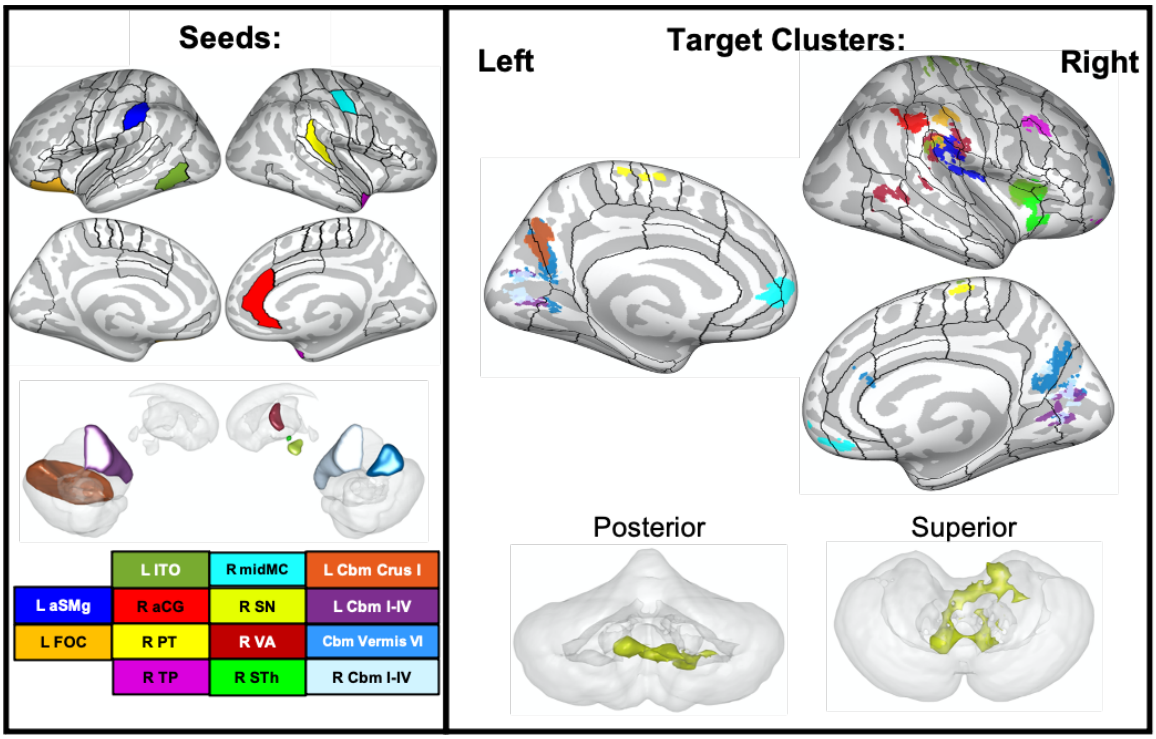

Figure 2.S17. A summary of functional connectivity (*rhythm* - *baseline*) negatively correlated with stuttering severity in AWS. Seed regions for these connections are indicated in the left panel either on an inflated cortical surface (top; ROIs are as in Figure 2.S10) or on a transparent 3D rendering of subcortical structures in each hemisphere viewed medially (bottom). Colors in the rest of the figure refer back to these seed regions. Nineteen target clusters (representing nineteen distinct connections) are displayed in the right portion of the figure. These clusters are either projected onto an inflated cortical surface, along with the cortical ROI parcellation of the SpeechLabel atlas described in Cai, Tourville, et al. (2014), or plotted on a 3D rendering of the cerebellum viewed either posteriorly (left) or superiorly (right). L = left, R = right, aSMg = anterior supramarginal gyrus, FOC = fronto-orbital cortex, ITO = inferior temporo-occipital cortex, aCG = anterior cingulate gyrus, PT = planum temporale, TP = temporal pole, midMC = middle primary motor cortex, SN = substantia nigra, VA = ventral anterior portion of the thalamus, STh = subthalamic nucleus, Cbm = cerebellum.

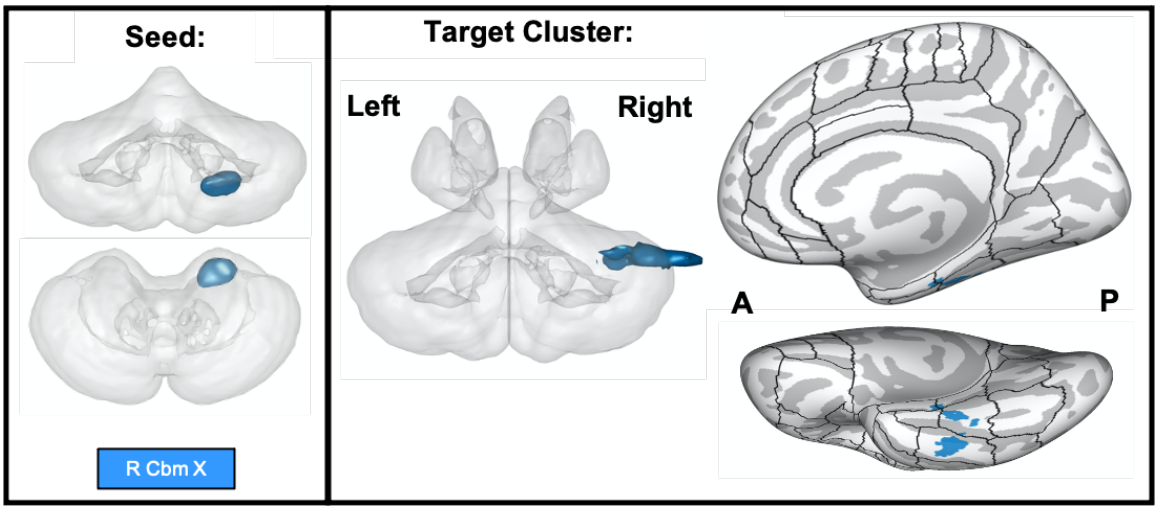

Figure 2.S18. A summary of functional connectivity (*normal* - *baseline*) positively correlated with Disfluency Rate in AWS. The seed region for this connection is indicated in the left panel on a 3D rendering of the cerebellum viewed either posteriorly (top) or superiorly (bottom). Colors in the rest of the figure refer back to this region. One target cluster is displayed in the right portion of the figure, either plotted on a 3D rendering of subcortical structures viewed posteriorly (left) or projected onto an inflated surface of the right hemisphere cerebral cortex, and viewed from the left (right top) or inferiorly (right bottom). R = right, A = anterior, P = posterior, Cbm = cerebellum.

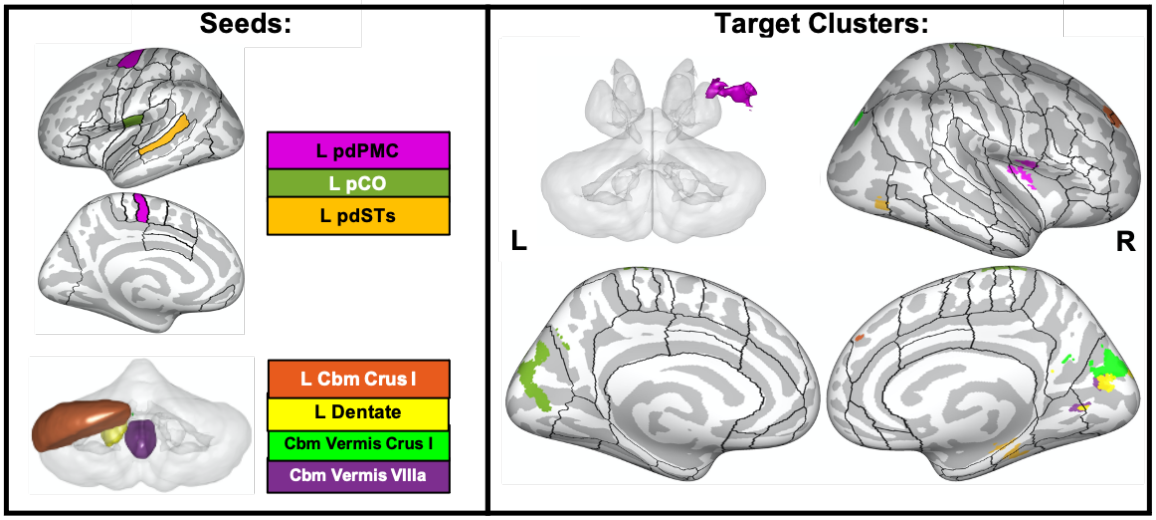

Figure 2.S19. A summary of functional connectivity (*normal* - *baseline*) negatively correlated with Disfluency Rate in AWS. Seed regions for these connections are indicated in the left panel either on an inflated left hemisphere cortical surface (top; ROIs are as in Figure 2.S10) or on a transparent 3D rendering of the cerebellum viewed posteriorly (bottom). Colors in the rest of the figure refer back to these seed regions. Eight target clusters (representing eight distinct connections) are displayed in the right portion of the figure. These clusters are either plotted on a 3D rendering of subcortical structures viewed posteriorly (top left), or projected onto an inflated cortical surface, along with the cortical ROI parcellation of the SpeechLabel atlas described in Cai, Tourville, et al. (2014). L = left, R = right, pdPMC = posterior dorsal premotor cortex, pCO = posterior central operculum, pdSTs = posterior dorsal superior temporal sulcus, Cbm = cerebellum.

ROI	Hemisphere	*t-value*	*p-unc*
AWS > ANS, Normal - Baseline			
MTO	Left	3.15	.0039
pMTg	Left	2.77	.0100
avSTs	Left	2.14	.0415
ANS > AWS, Normal - Baseline			
Cbm Vermis X	Midline	2.06	.0490
ANS > AWS, Rhythm - Baseline			
aFO	Left	2.84	.0086
aMTg	Right	2.16	.0400
Cbm Vermis VIIIb	Midline	2.86	.0082
Cbm Vermis VIIIa	Midline	2.14	.0417
Cbm X	Right	2.09	.0460

Table 2.S1. Exploratory regions-of-interest with significant group effects in either *normal – baseline* or *rhythm – baseline* contrasts ($p < 0.05$). unc = uncorrected, MTO = middle temporo-occipital cortex, pMTg = posterior middle temporal gyrus, avSTs = anterior ventral superior temporal sulcus, Cbm = cerebellum, aFO = anterior frontal operculum, aMTg = anterior middle temporal gyrus.

ROI	Hemisphere	*t-value*	*p-unc*
Group x Condition Interaction			
H	Right	-2.75	.0106
PP	Right	-2.70	.0118
	Left	-2.30	.0297
aIFt	Right	-2.66	.0131
FOC	Right	-2.40	.0235
PT	Left	-2.11	.0445
avSTs	Left	-2.10	.0448
TP	Left	-2.10	.0456
Cbm VIIIb	Left	-3.03	.005
Cbm VIIIa	Left	-2.75	.0106
	Right	-2.20	.0368
Cbm VIIb	Left	-2.52	.0181
Cbm Crus II	Left	-2.40	.0237
VPM	Left	-2.33	.0276
Cbm Vermis VIIIb	Midline	-2.26	.0323

Table 2.S2. Exploratory regions-of-interest with significant task activation group x condition interactions ($p < 0.05$). See Figure 2.S2 for individual group and condition effects. unc = uncorrected, H = Heschl's gyrus, PP = planum polare, aIFt = anterior inferior frontal gyrus pars triangularis, FOC = fronto-orbital cortex, PT = planum temporale, avSTs = anterior ventral superior temporal sulcus, TP = temporal pole, Cbm = cerebellum, VPM = ventral postero-medial portion of the thalamus.

ROI	Hemisphere	*t-value*	*p-unc*
ANS, Rhythm > Normal			
SPL	Left	4.72	.0003*
	Right	2.62	.0203
aSMg	Left	4.11	.0011
mdPMC	Right	3.98	.0014
	Left	2.25	.0418
PT	Left	3.82	.0019
	Right		
SMA	Left	3.36	.0046
aINS	Left	3.29	.0054
PP	Left	3.18	.0066
	Right	3.00	.0096
ITO	Left	3.06	.0085
	Right	2.17	.0480
preSMA	Left	2.79	.0145
pdPMC	Left	2.78	.0148
FMC	Left	2.76	.0152
vPMC	Right	2.59	.0213
PO	Left	2.53	.0242
pSMg	Right	2.43	.0293
	Left	2.24	.0418
aCG	Right	2.42	.0296
aSTg	Right	2.38	.0319
pSTg	Right	2.38	.0321
dMC	Left	2.33	.0355
ANS, Normal > Rhythm			
pIFt	Right	2.39	.0315
AWS, Rhythm > Normal			
vIFo	Right	3.06	.0092
pSMg	Right	2.51	.0260
dIFo	Right	2.27	.0411
AWS, Normal > Rhythm			
pPHg	Right	2.68	.0190
aIFt	Right	2.30	.0388
H	Right	2.20	.0466
Cbm VIIIb	Left	3.20	.0070
Cbm IX	Right	2.52	.0256
GPe	Right	2.22	.0449

Table 2.S3. Exploratory regions-of-interest with activation differences between the *rhythm* and *normal* conditions for ANS and AWS ($p < 0.05$). * indicates regions that survive a significance threshold of $p_{FDR} < 0.05$ for their respective analyses, unc = uncorrected, SPL = superior parietal lobule, aSMg = anterior supramarginal gyrus, mdPMC = middle dorsal premotor cortex, PT = planum temporale, SMA = supplementary motor area, aINS = anterior insula, PP = planum polare, ITO = inferior temporo-occipital cortex, preSMA = presupplementary motor area, pdPMC = posterior dorsal premotor cortex, FMC = fronto-medial cortex, vPMC = ventral premotor cortex, PO = parietal operculum, pSMg = posterior supramarginal gyrus, aCG = anterior cingulate gyrus, aSTg = anterior superior temporal gyrus, pSTg = posterior superior temporal gyrus, dMC = dorsal primary motor cortex, pIFt = posterior inferior frontal gyrus pars triangularis, vIFo = ventral inferior frontal gyrus pars opercularis, dIFo = dorsal inferior frontal gyrus pars opercularis, pPHg = posterior parahippocampal gyrus, aIFt = anterior inferior frontal gyrus pars triangularis, H = Heschl's gyrus, Cbm = cerebellum, GPe = external portion of the globus pallidus.

ROI	Hemisphere	*t-value*	*p-unc*
Normal-Baseline Correlation with SSI-Mod			
mdPMC	Right	3.05	.0101
aFO	Left	3.02	.0107
	Right	2.72	.0185
midPMC	Right	2.62	.0222
pdPMC	Left	2.49	.0284
	Right	2.37	.0355
aSMg	Right	2.43	.0315
SPL	Right	2.32	.0389
adPMC	Right	2.30	.0400
PO	Left	2.29	.0406
pFO	Right	2.21	.0475
PCN	Right	2.21	.0475
Rhythm-Baseline Correlation with SSI-Mod			
SFg	Left	-2.54	.0262
aTFg	Left	2.48	.0291
aCG	Left	-2.33	.0380
aFO	Left	2.20	.0484
Normal-Baseline Correlation with Disfluency Rate			
aCO	Right	3.96	.0019
	Left	2.47	.0298
pCO	Right	3.36	.0057
aINS	Right	3.28	.0065
PO	Right	3.22	.0073
aIFs	Right	3.17	.0081
vSC	Right	2.50	.0277
MTO	Right	2.49	.0285
PT	Right	2.46	.0300
pvSTs	Right	2.42	.0324
pdSTs	Right	2.41	.0329
H	Right	2.39	.0343
pSMg	Right	2.37	.0351
pFO	Right	2.35	.0364
pIFs	Right	2.30	.0405
aCG	Right	2.18	.0496
VA	Left	4.30	.0010
	Right	2.55	.0254
VL	Left	3.91	.0021
	Right	3.49	.0045
Cbm X	Left	2.73	.0184
Cbm IX	Left	2.54	.0258
	Right	2.42	.0326
Cbm V	Left	2.38	.0347
Putamen	Left	2.26	.0431
Cbm Vermis IX	Midline	2.22	.0462

Table 2.S4. Exploratory regions-of-interest with significant correlations between severity measures and speech activation in AWS ($p < 0.05$). unc = uncorrected, mdPMC = middle dorsal premotor cortex, aFO = anterior frontal operculum, midPMC = middle premotor cortex, pdPMC = posterior dorsal premotor cortex, aSMg = anterior supramarginal gyrus, SPL = superior parietal lobule, adPMC = anterior dorsal premotor cortex, PO = parietal operculum, pFO = posterior frontal operculum, PCN = precuneus, SFg = superior frontal gyrus, aTFg = anterior temporal fusiform gyrus, aCG = anterior cingulate gyrus, aCO = anterior central operculum, pCO = posterior central operculum, aINS = anterior insula, aIFs = anterior inferior frontal sulcus, vSC = ventral primary somatosensory cortex, MTO = middle temporo-occipital cortex, PT = planum temporale, pvSTs = posterior ventral superior temporal sulcus, pdSTs = posterior dorsal superior temporal sulcus, H = Heschl's gyrus, pSMg = posterior supramarginal gyrus, pFO = posterior frontal operculum, pIFs = posterior inferior frontal sulcus, VA = ventral anterior portion of the thalamus, VL = ventral lateral portion of the thalamus, Cbm = cerebellum.

CHAPTER III: Responses to auditory feedback perturbations in adults who stutter during syllable-timed speech.

Introduction

Persistent developmental stuttering is characterized by speech disfluencies such as sound repetitions, prolongations and blocks. It affects up to 8% of preschool-age children and persists into adulthood for 1% of the population (Yairi & Ambrose, 2013). Despite its prevalence and the expansive body of behavioral and neural stuttering research, the mechanisms underlying stuttering remain poorly understood. Numerous theories of stuttering have been proposed, addressing both the internal factors (disrupted speech, linguistic, emotional, or cognitive processes) that lead to overt disfluencies (e.g., Bloodstein, 1972; Guenther, 2016; Howell, 2010; Lieshout et al., 2014; Neilson & Neilson, 1987; Postma & Kolk, 1993; Vasiç & Wijnen, 2005; Webster, 1998) and the effects of a child's environment on persistence of stuttering (e.g., Lieshout et al., 2014; Smith, 1999; Starkweather & Gottwald, 1990). Two general categories of theories have been put forward that implicate disruptions in sensorimotor processing.

The first category suggests that stuttering occurs due to a reduced capacity for transforming sensory information to motor information (and vice versa; Hickok et al., 2011; Max et al., 2004; Neilson & Neilson, 1987). According to these theories, individual moments of stuttering occur when either a) insufficient mental resources are available to make this transformation and the whole system stops (M. D. Neilson & Neilson, 1987), or b) false errors are detected due to a disruption of the process that

generates a sensory prediction (Hickok et al., 2011; Max et al., 2004), leading either to delays until the mismatch is resolved (Max et al., 2004) or iterative (erroneous) correction signals (Hickok et al., 2011). Another set of theories posits that individuals who stutter have an impaired ability to properly time the initiation and/or termination of speech segments (Alm, 2004; Etchell et al., 2014; Guenther, 2016; MacKay & MacDonald, 1984; Wingate, 2002). This timing process also depends on the ability to combine sensory and motor information in order to determine the proper timing of these segments during an utterance.

A common experimental paradigm for testing interactions between sensory and motor processes examines online responses to perturbed sensory feedback. In studies using auditory feedback perturbations, one or more components of a participant's speech signal, such as voice fundamental frequency (f0; Burnett et al., 1998; Chen et al., 2007) or vowel formants (Niziolek & Guenther, 2013; Purcell & Munhall, 2006), are altered and fed back to them in real time and the ensuing responses are measured. Studies using this technique have demonstrated that adults who stutter (AWS) exhibit delayed (Bauer et al., 2007; Cai, Beal, et al., 2014; Loucks et al., 2012), reduced (Cai et al., 2012; Daliri et al., 2018; Daliri & Max, 2018; Loucks et al., 2012; Sares et al., 2018), and/or more temporally variable (Sares et al., 2018) responses compared to adults who do not stutter (ANS). In the somatosensory domain, AWS showed reduced speech compensatory responses to a bite block (Namasivayam & Van Lieshout, 2011) and reduced attenuation of (non-speech) jaw opening in response to masseter muscle vibration (Loucks & De Nil, 2006). These studies suggest that during a task involving sensory-motor interactions,

AWS exhibit processing delays and/or reduced scaling of corrective movements (often termed “gain” in the motor control systems literature [Guenther, 2016]), supporting both sets of theories described above.

Another type of auditory feedback perturbation more directly tests the role of sensorimotor integration for *timing* of ongoing speech by modifying the perceived timing of a self-produced speech gesture. This type of perturbation either temporally stretches or compresses a small portion of the speech signal (either in real time or using pre-recorded samples), so the duration of a phoneme sounds altered to the speaker. Previous work shows that in response to brief delays in auditory feedback, neurotypical speakers will delay the production of subsequent speech gestures and adapt by shortening the perturbed phoneme on future trials (Floegel et al., 2020; Mitsuya et al., 2014; Ogane & Honda, 2014; Oschkinat & Hoole, 2020).

The only study to evaluate AWS on this type of paradigm was Cai, Beal, et al. (2014). This study found that AWS show reduced responses to timing perturbations, suggesting impaired sensorimotor integration for timing control in stuttering. However, responding to the perturbation in this study required tracking formant changes in order to infer timing. As a result, it is unclear whether a pure timing perturbation that alters timing but not formant trajectories would show the same response impairment in AWS.

The first aim of the present study was to dissociate responses to a pure spectral perturbation and a pure temporal perturbation in order to understand the importance of each in stuttering. To achieve a pure temporal perturbation, the speech spectrum was stretched in time without altering spectral information such that the boundary between a

fricative and stop consonant was delayed in auditory feedback. The pure spectral perturbation involved an alteration in the frequency of a vowel's first formant without any adjustments to timing. If stuttering is related to an impairment in making motor adjustments to temporal auditory feedback cues (in addition to spectral cues as found in previous studies), AWS should show a response deficit (compared to ANS) to both perturbations. Conversely, if stuttering is related to an impairment in making adjustments to spectral cues and not pure temporal auditory feedback cues, AWS should show a response deficit to the formant perturbation and no response deficit to the timing perturbation.

In addition, one piece of support for an impaired speech timing system in stuttering is the well-documented phenomenon that speaking with an external timing cue like a metronome reduces disfluencies in PWS (e.g., Andrews et al., 1982; Brady, 1969; Braun et al., 1997; Davidow, 2014; Stager et al., 2003; Toyomura et al., 2011). It has been suggested that external cues allow PWS to rely less on inefficient or impaired "internal" timing mechanisms to sequence speech utterances (Alm, 2004; Etchell et al., 2014; Guenther, 2016). Prior neuroimaging work has indicated that externally paced speech may "normalize" activation in the brain and thus restore some speech timing function that can lead to fluent speech (Toyomura et al., 2011) and/or recruit alternative brain networks to a greater extent to circumvent or support impaired speech timing regions (Frankford et al., in press; Chapter II). Behaviorally, then, speaking in a manner that references an external stimulus may lead to normalized auditory motor integration for speech timing. Therefore, the second aim of the present study was to test whether

externally paced speech leads to fluency by helping resolve the disruptions implicated in either of the two theories discussed above. This was carried out by examining the effects of rhythmically paced speech on responses to altered auditory feedback in both the pure spectral and pure timing domains.

Methods and Materials

Participants

15 adults who stutter (AWS; 12 Males/3 Females, aged 18–44 years, mean age = 25.73 years [SD = 8.37]) and 16 adults who do not stutter (ANS; 12 Males/4 Females, aged 18–44 years, mean age = 26.69 years [SD = 6.79]) participated in this study. This unbalanced male-to-female ratio mirrors the prevalence of persistent developmental stuttering in the population (Bloodstein, 1995). All participants were native speakers of American English with no prior history of speech, language, or hearing disorders (other than stuttering for the AWS group), and all participants passed audiometric screenings with binaural pure-tone hearing thresholds of less than 25 decibels hearing level (dB HL) at 500, 1000, 2000, and 4000 Hz. AWS were video-recorded to collect a sample of speech during three tasks: 1) an in-person conversation, 2) a phone conversation, and 3) reading the Grandfather passage aloud. Based on these samples a trained speech-language pathologist evaluated stuttering severity using the Stuttering Severity Instrument – Fourth Edition (SSI-4; Riley, 2008). Individuals ranged from very mild (9) to very severe (42) with a mean score of 23. Participants provided informed written

consent and the study was approved by the Boston University Institutional Review Board.

Experimental Setup

Figure 3.1 provides a schematic diagram of the experimental setup. Participants were seated in front of a computer monitor in a sound-attenuating booth. They were fitted with ER1 headphones (Etymotic Research, Inc.) and an AT803 microphone (Audio-Technica) mounted to a headband via an adjustable metal arm. This arm was positioned such that the mouth to microphone distance was 10 cm for all subjects. The microphone signal was amplified and digitized using a MOTU Microbook external sound card and sent to a computer running MATLAB-based (MathWorks) custom experimental software. Auditory feedback was sent back through the Microbook and amplified using a Xenyx 802 (Behringer) analog mixer such that the signal played through the earphones sounded 4.5 dB louder than the microphone input. This amplification helped minimize participants' ability to hear their non-perturbed feedback through air or bone conduction.

Stimuli

Stimuli for this study mainly consisted of one "target" sentence ("The steady bat gave birth to pups") on which all experimental manipulations were applied. Fifteen "filler" sentences, selected from the Harvard sentence pool (i.e., the Revised List of Phonetically Balanced Sentences; *IEEE Recommended Practice for Speech Quality*

Measurements, 1969), were also included to reduce boredom and keep participants attending to the task. All sentences contained eight syllables.

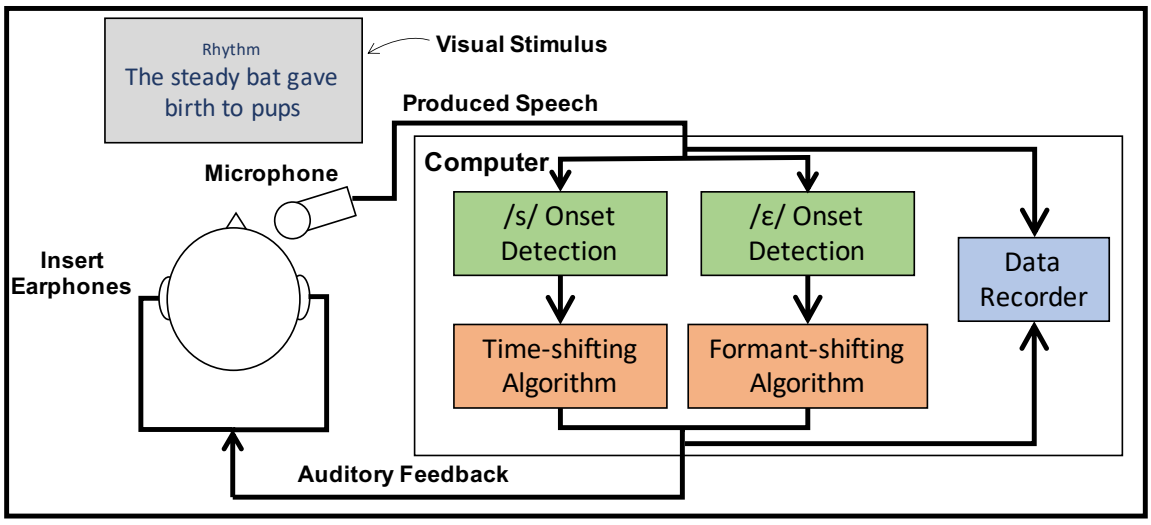

Figure 3.1. A schematic diagram showing the setup for the experiment. Following the presentation of an orthographic stimulus sentence and a condition cue ("Normal" or "Rhythm"), participants read the sentence according to the cue. Participants' speech signal was recorded and fed to an experimental computer running Audapter. On perturbed trials, detection of the relevant speech cue (onset of /ɛ/ in "steady" in the formant perturbation condition, onset of /s/ in "steady" in the timing perturbation condition) was used to initiate the pre-programmed auditory perturbation which was fed back to the participant via insert earphones. Both the perturbed and unperturbed signals were recorded for further analysis.

Procedure

Participants were instructed to read aloud sentences displayed on the computer monitor under two speaking conditions: either with each syllable evenly spaced (*rhythm speech* condition) or with a normal (unmodified) stress pattern (*normal speech* condition). At the beginning of every trial, participants viewed white crosshairs on a grey screen while eight isochronous tones (1000Hz pure tone, 25ms, 5ms ramped onset) were played with an inter-onset interval of 270ms. This resulting rate of approximately 222

beats/min was chosen so that participants' speech would approximate the rate of the *normal* condition (based on estimates of mean speaking rate in English; Davidow, 2014; Pellegrino et al., 2004). The trial type then appeared on the screen ("Normal" or "Rhythm") followed by a stimulus sentence. On *rhythm speech* trials, participants read the sentence with even stress at the rate of the tone stimuli aligning each syllable to a beat, while on *normal speech* trials, participants ignored the pacing tones, and spoke with normal rate and rhythm.

The experiment comprised three brief training runs followed by three experimental runs. During the first training run, participants received visual feedback on their loudness (a horizontal bar in relation to an upper and lower boundary) – males were trained to speak between 65 and 75 decibels sound pressure level (dB SPL), and females were trained to speak between 62.5 and 72.5 dB SPL. This difference was meant to account for natural differences in the speech sound intensity of males and females during conversational speech (Gelfer & Young, 1997). On the second training run, participants also were trained to speak with mean inter-syllable duration (ISD) between 220ms and 320ms (centered around the inter-onset interval of the tones). ISD was calculated as the time between the midpoints of successive vowels in the sentence. On the third training run, in addition to visual feedback on loudness and speaking rate, participants received feedback on the rhythmicity (isochronicity) of their speech. Rhythmicity was measured using the coefficient of variation (standard deviation/mean) of inter-syllable duration (CV-ISD), so lower coefficients of variance indicate greater rhythmicity. Using this feedback, subjects were trained to speak with a CV-ISD less than 0.25 for the *rhythm*

speech trials.

Each experimental run contained 80 speech trials, half *rhythm* and half *normal.* The target sentence (see "2.3. Stimuli") appeared in 80% of trials in each condition, while filler sentences comprised the remaining 20%. Half of these target trials contained experimental manipulations: one quarter included a perturbation of the first vowel formant (F1 perturbation) on the phoneme /ɛ/ in "steady" and another quarter included a brief delay in the transition between the /s/ and /t/ phonemes in the word "steady" (see Section 2.5. for details). The order of these trials in each run was pseudo-randomized such that every set of 10 trials contained two of each type of perturbation, one *rhythm* and one *normal*. In total, participants completed 240 trials, 192 of which contained target sentences. Of these, 96 were *rhythm* trials and 96 were *normal* trials, each containing 24 trials with an F1 perturbation, 24 trials with a timing perturbation, and 48 unperturbed trials. One subject only completed 200 trials (160 target trials) due to a technical error.

Focal Perturbations

In this study, two focal perturbations to auditory feedback were implemented mid-utterance using Audapter (Cai et al., 2011; Cai, Beal, et al., 2014; Tourville et al., 2013): an F1 perturbation and a timing perturbation. The latency in Audapter between the microphone signal and the processed headphone signal was 8 ms for the F1 perturbation trials and 16 ms for the timing perturbation trials as well as for the unperturbed trials. Recent work has shown that additional latency is incurred by the hardware used (Kim et al., 2020). While the total latency including hardware and software was not directly

measured, based on the hardware used and the estimates from Kim et al. (2020), the total latency was estimated as 23 ms for the F1 perturbation trials and 31 ms for the timing perturbation and unperturbed trials.

Formant Perturbation: The F1 perturbation was carried out as previously described in Cai et al. (2012). Briefly, the microphone signal was digitized at a sample frequency of 48000 Hz and downsampled by a factor of 3 to 16000 Hz for real-time processing. An autoregressive linear predictive coding algorithm, followed by a dynamic-programming tracking algorithm (Xia & Espy-Wilson, 2000), was used to estimate the formant frequencies in near real time. The tracked formant frequencies were then mapped to new, shifted values. In this experiment, fixed-ratio (+25%) shifting of F1 was used. Once the shifted formant frequencies were determined, a pole-substituting digital filter served to bring the formant resonance peaks from their original values to the new ones. The perturbation was applied to the vowel /ɛ/ in the word “steady” from the target sentence (Figure 3.2A) using an adaptive root mean square (RMS) threshold (based on previous trials; see Equation C1 in Appendix C for details) to determine the onset and offset. Perturbation onset occurred at the conclusion of the preceding /s/, when the ratio of the pre-emphasized (i.e., high-pass filtered) RMS and the unfiltered RMS went below an adaptive threshold (Equation C2). The perturbation was removed when the offset of the /ɛ/ in “steady” was detected using an adaptive RMS slope threshold (see Equations C3 and C4 in Appendix C for details).

Timing Perturbation: This study also used fine-scale temporal processing previously described in Tourville et al. (2013) to employ temporal dilation (slowing

down and speeding up) of auditory feedback using a phase vocoder (Bernsee, 1999). This focal timing perturbation was applied to the /s/ in "steady" and feedback was returned to normal by the end of the word (Figure 3.2B). To apply this perturbation at the desired time, Audapter relied on the online detection of the /s/, carried out in two steps. First, voicing onset (/ə/ in the preceding word, "the") was detected when the amplitude of the speech signal surpassed an RMS adaptive threshold (Equation C1) for at least 20 ms. Following this, the onset of the /s/ was detected when the ratio of the pre-emphasized (i.e., high-pass filtered) RMS and the unfiltered RMS exceeded an adaptive threshold (Equation C2 as in the formant perturbation) for at least 20 ms. See Appendix C for details on how these adaptive thresholds were calculated.

On a given timing-perturbed trial, once the /s/ was detected, Audapter downsampled the microphone signal as above and applied a short-time Fourier transform (STFT) on frames of 16ms (sliding by 4ms), saving the Fourier spectrum in memory. Through linear interpolation and inverse STFT re-synthesis, Audapter slowed down the auditory feedback to half speed for a subject-specific interval, equal to the average duration of the /s/ across non-perturbed target trials in previous runs using the forced-alignment speech recognition software Julius (Lee & Kawahara, 2009; see section 2.6.1. Data Processing) and carried out separately for the *rhythm* and *normal* conditions. The delayed feedback was maintained at normal speed for the same interval, then the feedback was accelerated to double speed until it realigned with the incoming microphone signal. This delayed the boundary between the /s/ and /t/ in the auditory signal by ~50 ms and returned feedback to normal by the end of the following syllable.

The gradual delay onset and offset assured feedback timing continuity that make the perturbed utterances sound qualitatively natural to the participant. Auditory feedback remained unperturbed for the rest of the trial.

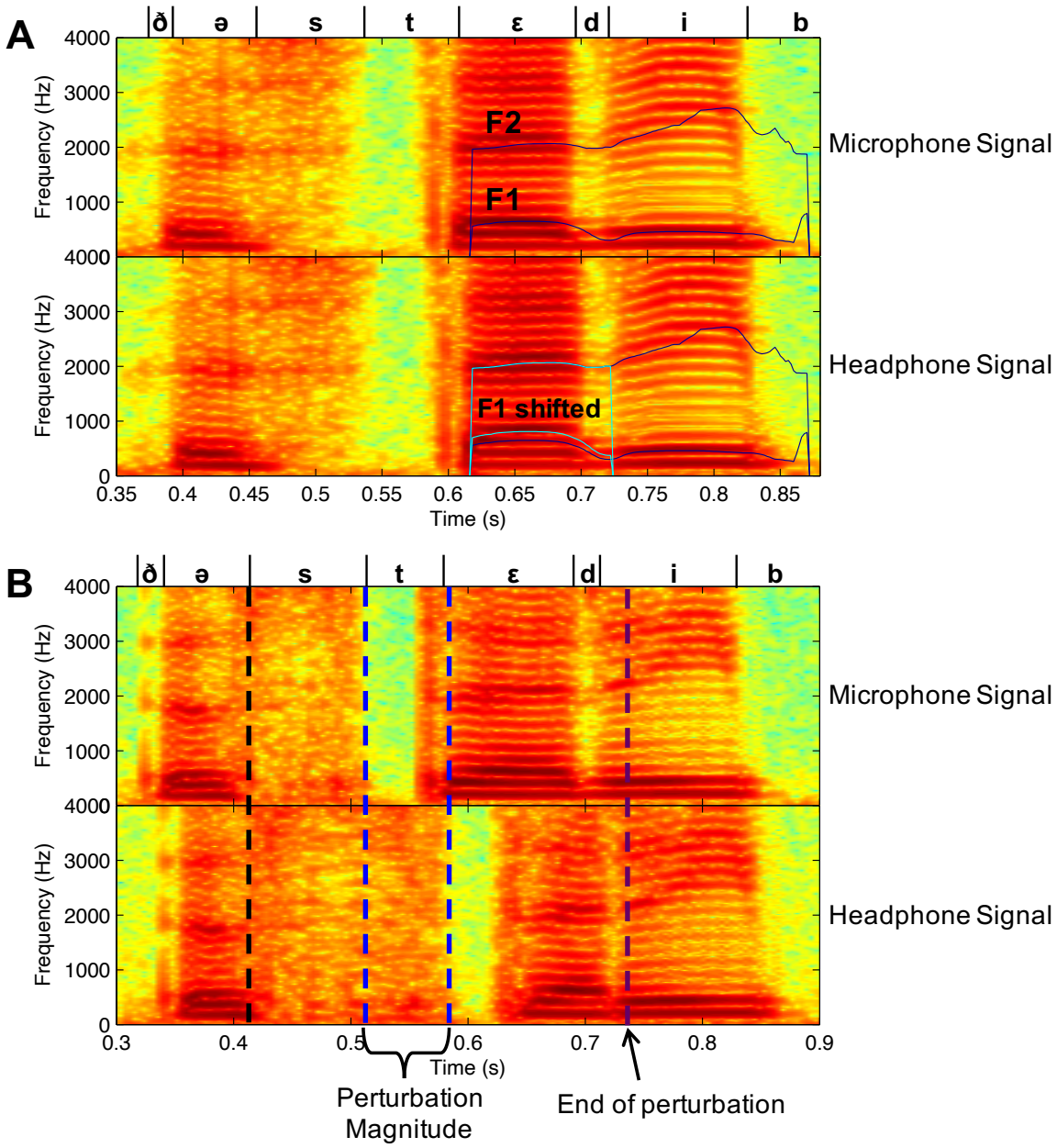

Figure 3.2: Examples of formant and timing perturbations. A. Example spectrograms of "The steady bat" during a formant perturbation trial generated from the recorded microphone signal (top) and headphone signal (bottom). The first and second formant

traces during the word "steady" are displayed. In the headphone signal spectrogram, the original formant traces are displayed in black and the shifted formant traces are overlaid in blue. Note that only the first formant is perturbed. B. A similar set of example spectrograms of "The steady bat" from a timing perturbation trial. The dashed black line indicates the onset of the /s/ in "steady" in the microphone signal. The first dashed blue line indicates the offset of the /s/ in "steady" in the microphone signal, and the second dashed blue line indicates the offset of the /s/ in "steady" in the headphone signal. The dashed purple line indicates when auditory feedback is returned to normal. Phoneme boundaries and international phonetic alphabet symbols are indicated above the microphone signals. F1 = first formant, F2 = second formant, Hz = hertz.

Analyses

Data Processing

An automatic speech recognition (ASR) engine, Julius [Lee & Kawahara, 2009], was used in conjunction with the free *VoxForge* American English acoustic models (voxforge.org) to determine phoneme boundary timing information for every trial. A researcher manually inspected each trial (presented in random order and blinded to condition) to ensure correct automatic detection of phoneme boundaries. Trials where there were gross ASR errors were removed. For each trial that included a perturbation, the same researcher compared the speech spectrogram from the microphone and the headphone signal and determined whether the perturbation occurred at the proper time within the utterance (i.e., during the /ɛ/ in "steady" for the formant perturbation or the /s/-/t/ boundary for the timing perturbation). Trials where this was not the case were discarded from further analysis. Any trials in which the subject made a reading error, a condition error (i.e., spoke rhythmically when they were cued to speak normally or vice versa), or a disfluency categorized as a stutter (determined by the experimenter and a speech-language therapist) were eliminated from further analysis. Because the timing

perturbation can lead to a stutter-like prolongation, only unambiguous disfluencies were eliminated. Finally, trials where subjects spoke outside of the trained ISD (220 ms – 320 ms) were also eliminated. In total, these procedures excluded an average of 13.8% of trials for ANS (SD: 6.5%) and 15.8% of trials for AWS (SD: 10.8%). This was not significantly different between groups ($t = 0.63$, $p = 0.53$).

To evaluate whether there was a fluency-enhancing effect of rhythmic pacing, the percentage of trials eliminated due to stuttering in the AWS group was compared between the two speaking conditions using a non-parametric Wilcoxon signed-rank test. Measures of the total sentence duration and intersyllable timing from each trial were also extracted to determine the rate and isochronicity of each production. Within a sentence, the average time between the centers of the eight successive vowels was calculated to determine the ISD. The reciprocal of the ISD from each sentence (1/ISD) was then calculated, resulting in a measure of speaking rate in units of syllables per second. Rate and CV-ISD were compared between groups and conditions using linear mixed effects models with group, condition, and group × condition interaction as fixed effects and subjects as random effects.

Formant Perturbation

Processing: Formant trajectories for the /ɛ/ in “steady” were extracted using a semi-automated process. First, the Audapter-determined formant trajectories were cropped within the ASR boundaries of the /ɛ/ and subsequent /d/. Only continuous segments surpassing an RMS intensity threshold were retained. This threshold was

initially set to the value used for detecting the onset of the perturbation described in Section 2.5. Adjustments to this threshold were made for vowel onset and offset in each subject to eliminate highly variable endpoints of the formant traces caused by pauses in voicing. Formant traces that were less than 30 frames (60 ms) were flagged for manual inspection and relabeling by a researcher. Traces were then time-normalized through linear interpolation for direct comparison across trials, and outliers (traces greater than 3 standard deviations away from the mean at any time point) were also flagged for manual inspection and relabeling. Manual relabeling was carried out blind to perturbation condition and consisted of labeling the onset and offset of a smooth, consistent first formant trace, often bounded by sudden changes in formant velocity due to changes in voicing status, and leaving a 2 ms buffer on either end. If a trace did not have a clear boundary, the entire trace was kept. This process of identifying outliers and relabeling trials was repeated until sudden changes in formant velocity were not identifiable at the beginning and end of each trace by visual inspection. Trials where a clear formant trajectory was not identifiable, where sudden large deviations occurred in an otherwise smooth formant trace, or where a vowel formant trace was shorter than 30ms were removed. During this process, 0.21% of total vowel traces were removed (0.23% from ANS, 0.19% from AWS). Using a Wilcoxon rank-sum test, this group difference was not found to be statistically significantly ($W = 250$, $p = 0.82$).

Compensation Analysis: Compared to previous formant perturbation studies, where perturbed vowels were produced in the context of lengthened single-syllable productions (e.g., Tourville et al., 2008) or continuous formant trajectories (Cai et al.,

2011; Cai, Beal, et al., 2014), each /ɛ/ vowel in the present study was relatively short in duration (~ 150 ms) and was followed immediately by the plosive /d/. In the context of consonants, vowel formants follow a trajectory with different values at the beginning, middle, and end. As in Cai et al. (2011) and Cai, Beal, et al. (2014), we wanted to make sure that when averaging formant traces across trials, the corresponding parts of the trajectory were directly compared. In order to account for variations in the duration of /ɛ/ across trials, traces were linearly interpolated to 100 evenly spaced points such that there was a single normalized time axis (see Figure 3.4). Because the actual duration of these traces could have an effect on responses (e.g., longer vowel duration could provide more time to process and respond to auditory feedback changes), vowel duration was compared between groups and conditions using linear mixed effects models with group, condition, and group × condition interaction as fixed effects and subjects as random effects.

Two analyses were then carried out to test for responses to the perturbations in each group and differences between the groups. In the first, formant traces were averaged across trials in the perturbed and non-perturbed conditions for each subject, and a deviation trace was created by subtracting the perturbed trace from the non-perturbed trace at each time point. To test for a response within groups in each speaking condition, deviation traces were compared to 0 using a one-sample t-test at each timepoint with an alpha criterion of 0.05. In addition, deviation traces were compared between groups using a two-sample t-test at each timepoint with an alpha criterion of 0.05.

Because this analysis required a test at each normalized timepoint, this introduced multiple comparisons that could have inflated the number of false positives. Since

adjacent timepoints would have highly correlated responses, however, a simple Bonferroni correction to correct for family-wise error would be overly conservative. A second analysis was therefore performed to determine whether there were significant responses and group differences overall. To do this, a principal components analysis (PCA) was carried out on all /ɛ/ formant traces across subjects and conditions to reduce 100 time points down to a small number of components that characterized 95% of the variance in individual formant traces. This variance was accounted for by the first 3 principal components, and the data were projected into this low-dimensional component space yielding three component scores for every trial. Similar to the first analysis, these scores were averaged across trials in the perturbed and non-perturbed conditions for each subject and speaking condition and subtracted to derive deviation scores.

The PCA scores were then used as the dependent variables in a multivariate general linear model (GLM) to determine whether responses were dependent on group, condition, or stuttering severity, while controlling for vowel duration. For group and severity analyses, the vowel duration regressor was averaged between the two conditions. For the condition, group x condition interaction, and severity x condition interaction analyses, the differences in vowel duration between conditions were included as regressors. For stuttering severity, two separate measures were used. The first was a modification of the SSI-4 score, heretofore termed "SSI-Mod." SSI-Mod removes the secondary concomitants subscore from each subject's SSI-4 score, thus focusing the measure on speech-related function. The second measure was the percentage of disfluent trials during the *normal* conditions (disfluency rate). Therefore, two separate models

were evaluated; one that included the SSI-Mod scores and a second that included the disfluency rates.

Timing Perturbation

Perturbation magnitudes for each time-perturbed trial were defined as the maximum difference in ASR phoneme boundaries between the microphone input signal and auditory feedback during the word "steady." For example, if the /s/-/t/ boundary was delayed by 50 ms in auditory feedback and the /t/-/I/ boundary was delayed by 60 ms, the perturbation magnitude for that trial would be 60 ms. One trial from each of three participants was removed due to ASR errors in the auditory feedback that led to erroneous perturbation magnitudes. Because the exact magnitude of the perturbation depended on a) the parameters that defined the time dilation, which were derived from previous productions, and b) the duration of the /s/ and /t/ phonemes in a given trial, it was difficult to ensure complete consistency across participants and trials. A linear mixed effects model with group, condition, and group x condition interaction as fixed effects and subjects as random effects was carried out to determine if any differences existed between groups and conditions. While there was not a significant effect of group ($F(1,29) = 0.93$, $p = 0.34$) or interaction between group and condition ($F(1,29) = 0.31$, $p = 0.58$), there was a main effect of condition ($F(1,29) = 22.15$, $p < 0.0001$) such that there were smaller perturbations in the *rhythm* condition than the *normal* condition (ANS *normal*: 59.4 ± 11.1 ms, ANS *rhythm*: 53.8 ± 11.0 ms, AWS *normal*: 56.5 ± 13.4 ms, AWS *rhythm*: 49.4 ± 8.9 ms). Because of this, perturbation magnitude was entered as a

covariate in subsequent analyses.

The durations from the onset of /s/ in "steady" to the offset of /s/ and to each of six subsequent syllable boundaries (see Table 3.1 for a description of these boundaries) in the ASR segmentation were calculated. To assess timing changes in response to the perturbation, these values were averaged across perturbed trials and across non-perturbed trials (separately within normal and rhythmic conditions) in each participant and the non-perturbed average was subtracted from the perturbed average to yield a cumulative response curve. This curve was compared to 0 at each time point for each speaking condition and group using a one-sample t-test and evaluated with a Bonferroni correction to correct for 7 time points.

Symbol	Landmark
/s/	Onset of "steady"
/t1/	s-t boundary in "steady"
/d/	Onset of /d/ in "steady"
/b1/	Onset of "bat"
/g/	Onset of "gave"
/b2/	Onset of "birth"
/t2/	Onset of "to"
/p/	Onset of "pups"

Table 3.1. Symbols used to denote sound/syllable boundaries in the present study.

A PCA was then performed on the response curves from all subjects to extract the components that characterize most of the variance in the responses. Similar to the formant perturbation analysis, the first three principal components were used for

subsequent analysis because they accounted for 95% of the variance in the data set. The responses of each subject were projected onto these principal components and used as the dependent variables in a multivariate GLM to determine whether responses were dependent on group, condition, or stuttering severity, while controlling for speaking rate and perturbation magnitude. For group and severity analyses, the speaking rate and perturbation magnitude regressors were averaged between the two conditions. For the condition, group x condition interaction, and severity x condition interaction analyses, the differences in rate and magnitude between conditions were included as regressors. As with the formant perturbation analysis, two separate models were evaluated; one that included the SSI-Mod scores and a second that included the disfluency rates. To test whether responses were correlated with rhythmicity in the *rhythm* condition, CV-ISD was added as a covariate to the original model.

Results

Disfluency Rate

For most subjects, stuttering occurred infrequently over the course of the experiment, with 6 out of 16 AWS producing no disfluencies. However, AWS produced significantly fewer disfluencies during the *rhythm* condition (1.2%) than in the *normal* condition (7.4%; $W = 52$, $p = 0.012$; Figure 3.3).

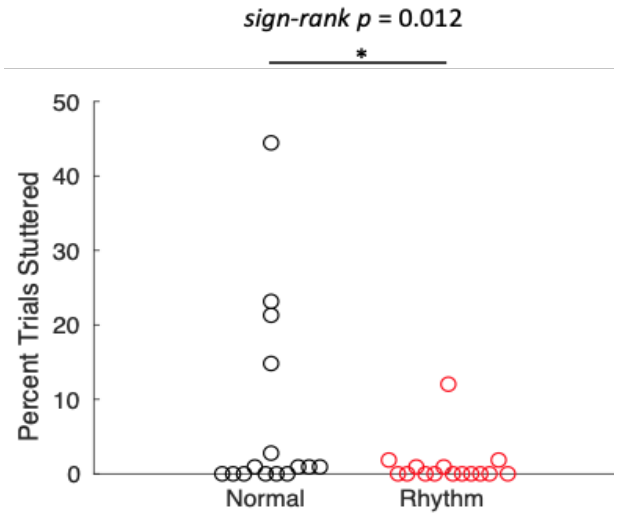

Figure 3.3. Comparison of dysfluencies between the *normal* and *rhythm* conditions for AWS. Circles represent individual participants.

Speaking Rate and Rhythmicity

For rate (Table 3.2), there was no significant effect of group ($F(1,29) = 0.19$, $p = 0.67$), but there was a significant effect of condition ($F(1,29) = 76.6$, $p < 0.0001$) that was modulated by a significant group × condition interaction ($F(1,29) = 6.15$, $p = 0.02$). In this case, participants in both groups spoke at a slower rate in the *rhythm* condition, but this difference was larger for ANS (ANS *normal*: 4.0 ± 0.2 syl/sec, ANS *rhythm*: 3.6 ± 0.1 syl/sec, AWS *normal*: 3.9 ± 0.2 syl/sec, AWS *rhythm*: 3.7 ± 0.1 syl/sec). Because of these significant effects, rate was included as a covariate in the timing perturbation analysis. To examine whether this reduction in rate led to increased fluency rather than the isochronous pacing, we tested for a correlation between the change in speech rate and the reduction in disfluencies. These two measures were not significantly correlated ($r = 0.10$, $p = 0.73$). As expected, there was a significant effect of condition on CV-ISD

($F(1,29) = 294.73$, $p < 0.0001$), but no effect of group ($F(1,29) = 0.53$, $p = 0.47$) or interaction ($F(1,29) = 0.41$, $p = 0.52$).

Measure	*ANS*		*AWS*		Main effect of Group:	Main effect of Condition:	Interaction:
	Normal	*Rhythm*	*Normal*	*Rhythm*			
Speaking rate (ISD/sec)	4.0 ± 0.2	3.6 ± 0.1	3.9 ± 0.2	3.7 ± 0.1	$F(1,29) = 0.19$, $p = 0.67$	**$F(1,29) = 76.60$, $p < 0.0001$**	**$F(1,29) = 6.15$ $p = 0.02$**
CV-ISD	0.27 ± 0.06	0.10 ± 0.02	0.26 ± 0.05	0.10 ± 0.02	$F(1,29) = 0.53$, $p = 0.47$	**$F(1,29) = 294.73$, $p < 0.0001$**	$F(1,29) = 0.41$ $p = 0.52$

Table 3.2. Descriptive and inferential statistics for speaking rate and CV-ISD. Error estimates indicate 95% confidence intervals. Significant effects are highlighted in bold.

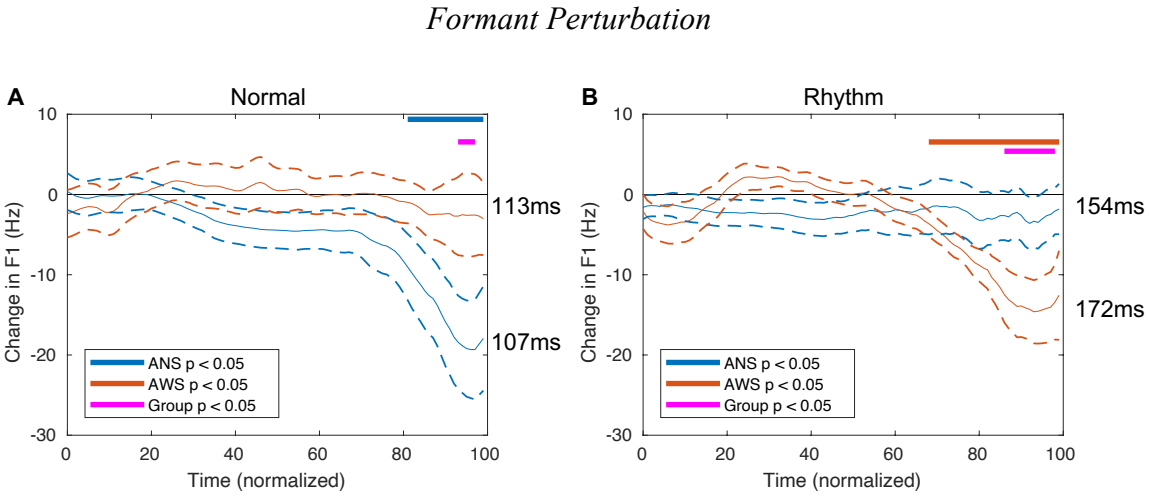

Figure 3.4. Time-normalized responses to the F1 perturbation during the /ɛ/ in "steady" in the *normal* (left) and *rhythm* (right) conditions. Solid curves indicate the average difference in F1 frequency between the perturbed and non-perturbed conditions for ANS (blue) and AWS (orange). Dashed lines indicate mean response +/- standard error of the mean. The blue bars indicate intervals of significant responses in ANS ($p < 0.05$, uncorrected). The orange bars represent intervals of significant responses in AWS ($p < 0.05$, uncorrected). The magenta bars indicate intervals of significant differences between ANS and AWS ($p < 0.05$, uncorrected). Duration of the non-normalized /ɛ/ response curves (averaged between perturbed and non-perturbed trials and across subjects) for each group and speaking condition are included to the right of the response curves for reference.

Figure 3.4 shows the time-normalized responses to the F1 perturbation during the /ɛ/ in "steady" for each condition with colored bars indicating when each group's response was different from 0, and when the groups were significantly different (uncorrected). For the *normal* condition (Figure 3.4A), ANS exhibit an opposing response that begins approximately 80% of the way through the vowel and continues until the transition to the /d/. AWS do not show a similar response prior to the end of the /ɛ/ vowel. There is a brief period during which the difference between the groups is significant using an uncorrected threshold of $p < 0.05$. Conversely, in the *rhythm* condition (Figure 3.4B) AWS respond to oppose the perturbation beginning at approximately 70% of the way through the vowel until the end, while ANS do not show any response. This difference in response is reflected as a group difference between approximately 85% and 95% through the vowel.

There was a significant main effect of condition on the duration of the /ɛ/ in "steady," where the vowel was longer in the rhythm condition than the normal condition ($F(1,29) = 105.9$, $p < 0.0001$), but no significant main effect of group ($F(1,29) = 3.3$, $p = 0.08$) or group by condition interaction ($F(1,29) = 1.3$, $p = 0.26$).

To determine the effects of group and condition on responses while controlling for vowel duration, a multivariate GLM was performed. There was no significant effect of group ($F(3, 25) = 0.75$, $p = 0.54$) or condition ($F(3, 25) = 0.03$, $p = 0.99$), but there was a significant group x condition interaction ($F(3, 25) = 3.18$, $p = 0.04$) on response magnitude. There was also no significant effect of either SSI-mod ($F(3, 25) = 0.22$, $p = 0.88$) or SSI-mod x condition ($F(3, 25) = 0.82$, $p = 0.50$). Finally, there was no significant

effect of vowel duration ($F(3, 25) = 1.22$, $p = 0.32$) or a significant vowel duration by condition interaction ($F(3,25) = 0.36$, $p = 0.78$) on response. After substituting in disfluency rate from the experimental session for SSI-Mod, we re-ran the model. There was no significant main effect of disfluency rate ($F(3, 25) = 0.81$, $p = 0.50$) or disfluency rate by condition interaction ($F(3, 25) = 1.60$, $p = 0.21$). For complete results of this new model, see Table 3.S1.

Timing Perturbation

Figure 3.5 shows the average cumulative timing responses to the timing perturbation across all time points in each group and condition. On average, both groups show a significant response to the perturbation in both conditions. In the *normal* condition, both groups first show this significant response at syllable boundary 3 (at the conclusion of the /di/ in "steady"), while in the *rhythm* condition, the groups exhibit significant differences after syllable boundary 2 (following the /ɛ/ in "steady"). This difference between conditions makes sense since in the *normal* condition, the /tɛ/ in "steady" is produced with a shorter duration (mean: 172 ms, SD: 22 ms) than in the *rhythm* condition (mean: 206 ms, SD: 41 ms). Thus, by the time the response begins, the sentence has already progressed to the next syllable in the *normal* condition but not in the *rhythm* condition.

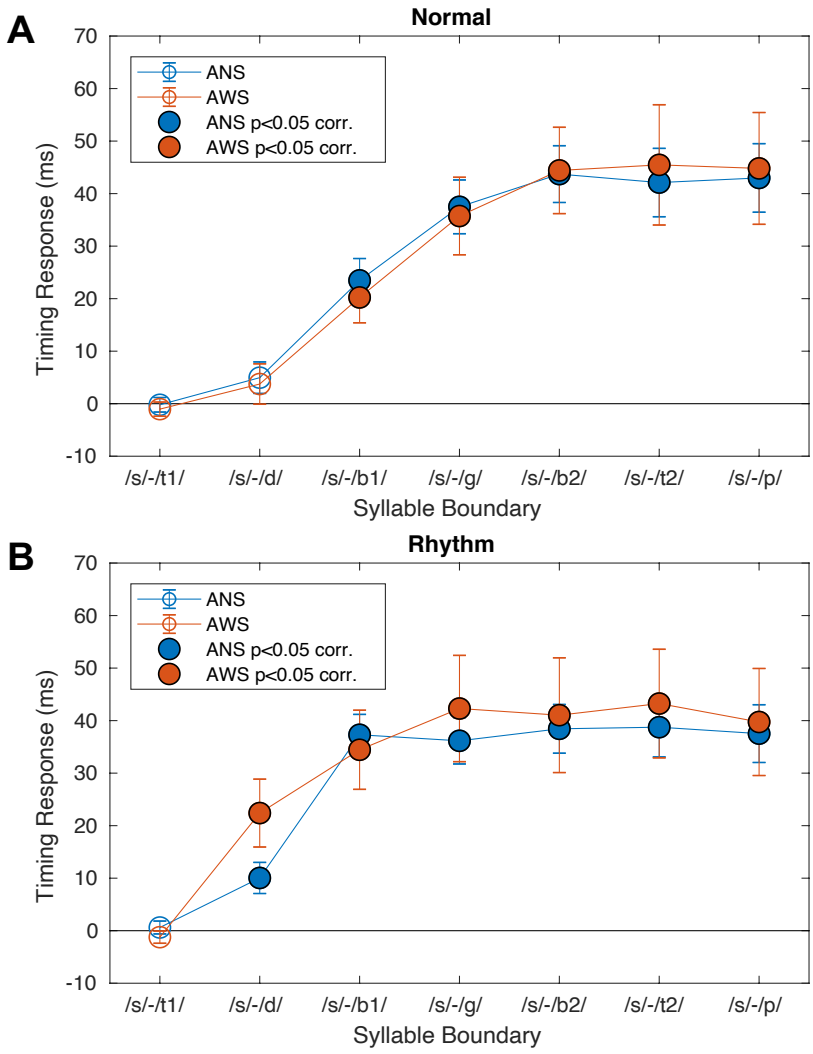

Figure 3.5. Cumulative responses between the timing-perturbed and non-perturbed conditions at each of seven sound/syllable boundaries (see Table 3.1). A. Responses during the *normal* speaking condition. The blue and orange curves correspond to the ANS and AWS groups, respectively. Filled circles indicate responses that differ significantly from 0 ($p < 0.05$, Bonferroni-corrected). B. Responses during the *rhythm* speaking condition. Colors represent the same as in A.

To determine the effects of other variables and control for factors like perturbation magnitude and rate, a multivariate GLM was performed. There was no significant effect of group ($F(3, 24) = 0.45$, $p = 0.72$) or group x condition interaction

($F(3, 24) = 0.91$, $p = 0.45$) on response magnitude, but there was a significant effect of condition ($F(3, 24) = 3.15$, $p = 0.04$). There was also no significant effect of either SSI-mod ($F(3, 24) = 2.39$, $p = 0.09$) or SSI-mod x condition ($F(3, 24) = 1.39$, $p = 0.27$). Perturbation magnitude had a significant effect ($F(3, 24) = 3.42$, $p = 0.03$) — larger perturbation magnitudes led to larger responses. Finally, there was no significant effect of mean rate on response ($F(3, 24) = 0.96$, $p = 0.43$). To follow up on the significant effect of condition, the effect sizes were projected back into syllable-boundary time. This analysis showed that the difference was mainly due to the earlier onset responses in the rhythm condition (with respect to syllable boundary in the sentence) rather than the total cumulative responses as measured at the end of the sentence. To confirm this, a GLM was performed where the independent variables were the same as above, but the dependent variable was the perturbation response at syllable boundary 7 (/s/-/p/). This analysis found no significant effect of condition ($F(1,26) = 0.001$, $p = 0.98$).

We then re-ran the model substituting in disfluency rate from the experimental session for SSI-Mod, and found a main effect of disfluency rate ($F(3, 24) = 5.22$, $p = 0.006$) that was significantly modulated by condition ($F(3, 24) = 3.48$, $p = 0.03$; see Figure 3.6; for complete results of this new model, see Table 3.S2). Accounting for all other variables, subjects with more disfluencies during the task had larger responses. Finally, to see if CV-ISD score was associated with the response, we added it into the original model and found that it was not a significant predictor of response ($F(3,23) = 1.35$, $p = 0.28$).

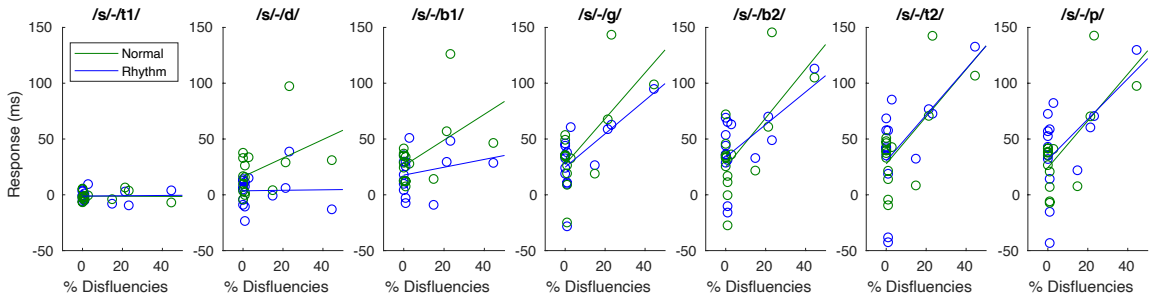

Figure 3.6. Scatterplots comparing disfluency rate during the *normal* condition with cumulative timing perturbation responses in AWS at each of seven sound/syllable boundaries (see Table 3.1). Circles indicate individual AWS for either the *normal* (blue) or *rhythm* (green) conditions. Least squares lines for each condition are superimposed on the data.

Discussion

The present experiment first aimed to examine whether a purely temporal auditory feedback timing perturbation exposes a response deficit in AWS as the spectro-temporal perturbation did in Cai, Beal, et al. (2014). As reported in previous studies (Cai et al., 2012; Loucks et al., 2012; Sares et al., 2018), AWS exhibited reduced or absent responses during a purely spectral perturbation compared to ANS in non-paced speech. For the timing perturbation, AWS did not exhibit the same response deficit as in Cai, Beal, et al. (2014); rather, their performance mirrored that of the non-stuttering controls, with both groups showing compensatory speech timing delays in response to focal delays in auditory feedback. In addition, this study examined whether pacing speech to an external stimulus (which is known to improve fluency in AWS) alters auditory feedback control processes in AWS. While AWS did produce significantly fewer disfluencies in the *rhythm* condition, changes in perturbation responses were more complex. AWS showed a significant compensatory response to the F1 perturbation in the *rhythm*

condition (while they did not in the *normal* condition). At the same time, ANS showed no significant response to the formant perturbation in the *rhythm* condition. For the timing perturbation, the *rhythm* condition did not change the overall responses in either group. These results are discussed in further detail below with respect to the stuttering theories mentioned in the Introduction as well as prior literature.

Auditory feedback timing control in AWS

The first paper to investigate auditory feedback-based speech timing control in AWS was Cai, Beal, et al. (2014). Taking advantage of the continuous formant trajectories in the carrier phrase "I owe you a yo-yo," the authors applied either an advancement (~45 ms) or delay (~24 ms) to these trajectories at the local minimum of the second formant (F2) during the first /o/ ("owe"), and measured changes in the timing of subsequent F2 landmarks compared to a non-perturbed condition. They found that while neither group responded to the advanced feedback, only ANS significantly responded to delayed feedback with a delay in subsequent landmarks. Furthermore, AWS' reduced responses were most pronounced earlier in the phrase. This study supported the theory that AWS exhibit an impairment in utilizing sensory cues for timing ongoing speech. Given these findings, the lack of a group difference in responses to the timing perturbation in the present study might seem surprising. However, Cai, Beal, et al. (2014) created a timing perturbation by applying an F1 and F2 perturbation that remapped the formants on a time lag to delay the F2 local minimum. Thus, the delayed timing signals were embedded in a task that required precise spectral tracking of the acoustic signal. In

contrast, the timing perturbation in the present study was applied to the middle of the /s/ and the occlusion of the /t/ in “steady” and involved a temporal prolongation of the entire speech signal with no modification of the spectral content. The present results, in combination with those of Cai, Beal, et al. (2014), suggest that AWS only have a deficit in their ability to use auditory feedback timing cues to sequence speech when those cues require tracking spectral features like formant frequencies.

This dichotomy between spectro-temporal and pure temporal perturbations can be thought of in terms of two motor timing theories described in the speech motor control literature: intrinsic (state) timing vs. extrinsic (clock) timing (Fowler, 1980; Kelso & Tuller, 1987). Extrinsic (clock) timing refers to the idea that the timing of subsequent speech segments in an utterance is planned in relation to an absolute timekeeper (e.g., in millisecond time). For intrinsic (state) timing, the planned temporal relations between adjacent speech segments are determined based on the relative progression of the speech system through a series of states (articulator positions and velocities, evident as formant trajectories in the acoustic signal). In the present study, extrinsic time (clock time) was perturbed, whereas Cai, Beal, et al. (2014) applied more of an intrinsic timing perturbation, i.e. changing the perceived state of the system (formants) to change the perception of time. The results of these two studies suggest that AWS have difficulty responding to intrinsic timing manipulations, but not external “clock time” manipulations.

An alternative possibility regarding the different responses found in the present study versus Cai, Beal, et al. (2014) involves the magnitude of the perturbation. While the

present study delayed the auditory feedback signal by 50 – 60 ms, the perturbation in Cai, Beal, et al. (2014) only introduced a delay of 20 – 25 ms. This could indicate that AWS have a more difficult time detecting and/or responding to more fine-grained temporal perturbations. Indeed, recent work indicates that AWS have less sensitivity to judging time intervals of various types (Devaraju et al., 2020; Schwartze & Kotz, 2020), although no studies have yet examined perceptual acuity for speech segment timing in AWS. Future studies would need to directly compare spectro-temporal and pure temporal perturbations of the same magnitude to confirm that magnitude differences did not lead to the differences found between Cai, Beal, et al. (2014) and the present study.

At the same time, while these interpretations are potentially compelling, care should be taken in relating the present results to the underlying mechanisms of stuttering. The presently discussed findings, along with most studies of auditory feedback control in stuttering, only include AWS who have had many years to develop adaptive behaviors that could influence these results. Indeed, the only auditory feedback manipulation study including both children who stutter (CWS) and AWS found that while AWS demonstrated diminished adaptation to a repeated formant perturbation, CWS responses were normal (Daliri et al., 2018). This may indicate that impaired sensory-motor transformation abilities found in spectral perturbations are secondary rather than core components of stuttering and may develop slowly over time. Future studies of spectral and temporal auditory feedback perturbations should examine responses in CWS as well as AWS in order to disentangle these possibilities.

The effect of syllable-timed speech on auditory feedback control

The presence of a compensatory response for AWS in response to the formant perturbation during the *rhythm* condition makes it tempting to suggest that speaking isochronously corrects the disturbance in auditory-motor transformation or makes this process more efficient. Previous work has shown that speaking along with a metronome reduces neural activation differences between AWS and ANS (Toyomura et al., 2011; although see Chapter II/Frankford et al. [in press] who did not find this). Furthermore, delayed auditory feedback, which also reduces disfluencies, causes pre-speech neural suppression in AWS to become more similar to ANS (Daliri & Max, 2018). Assuming that pre-speech neural suppression corresponds to speech motor control processes (Max & Daliri, 2019), this finding implicates a change in the forward modeling processes needed for error detection. A similar process could be taking place in the present study.

Alternatively, responding to feedback perturbations during the *rhythm* condition may reflect a sudden freeing of neural resources due to the timing of the sequence being preplanned. If the speech timing system in AWS requires more neural resources to generate phoneme initiation cues during normal speech, there may be reduced resources available to make corrections for spectral feedback errors. When suprasyllabic timing is predetermined as in the *rhythm* condition, these resources can be redirected to spectral auditory feedback control as well as fluent speech production.

However, there may be a simpler explanation for this finding: speaking isochronously led to an increase in the average duration of the /ɛ/ in “steady.” If AWS take longer to generate corrective movements to feedback errors due to inefficient inverse

models, as evidenced by previous feedback perturbation studies (Bauer et al., 2007; Cai, Beal, et al., 2014; Loucks et al., 2012), the correction may only be registered when the vowel is sufficiently long. Differences in vowel duration across conditions were at least partially accounted for by including vowel duration difference scores in the GLM, however this makes the assumption that the relationship between vowel duration and response is linear. In fact, the relationship is likely more complex; we would not expect any response if the vowel is shorter than the minimum auditory feedback processing delay. As such, this change in response between the *normal* and *rhythm* conditions may be a secondary effect of isochronous speech rather than directly related to the "rhythm effect."

An additional surprising result of this study was that ANS did not show any response to the formant perturbation during the *rhythm* condition. Because this is the first experiment to examine responses to formant perturbation during isochronous speech in ANS, this may reflect a general change in the balance of auditory feedback versus feedforward control for spectral cues during the *rhythm* condition, where ANS are more concerned with precise control of the timing aspects of the sentence (as evidenced by the clear responses to the timing perturbation). From this, it may be expected that during isochronous speech, ANS do not make spontaneous adjustments to formants (as in Niziolek et al. [2013]). Future research will need to examine this relationship further to determine why ANS respond to formant perturbations during normal (non-rhythmic) speech but not rhythmic speech, whereas AWS show the opposite pattern. In any case, these differences appear to be limited to the spectral perturbation; there were no response

differences between the AWS and ANS groups for the timing perturbation.

Finally, it was suggested in the previous section that AWS may have difficulty responding to intrinsic (state) timing perturbations (as in Cai, Beal, et al. [2014]), but not extrinsic (clock) timing perturbations (like the present study; see *Auditory feedback timing control in AWS*). Speaking isochronously may bias syllable timing away from an intrinsic timing mechanism and toward an extrinsic timing mechanism, circumventing the impaired intrinsic timing mechanism in AWS and leading to greater fluency. This hypothesis could be tested with an additional study investigating the effect of syllable-timed speech using the more intrinsic timing perturbation from Cai, Beal, et al. (2014).

Additional considerations

Previous studies of formant perturbations show that response magnitudes are not correlated with stuttering severity (Cai et al., 2012; Daliri et al., 2018) but response timing variability is (Sares et al., 2018), such that more severe AWS are more variable. In the previous timing perturbation study (Cai, Beal, et al., 2014), correlation of response to severity was not reported. Therefore, correlations between responses and stuttering severity were examined in the present study. Despite there being no group differences between AWS and ANS for the timing perturbation, there was a significant positive correlation between experimental disfluencies (during the *normal* condition) and the size of compensatory responses to the timing perturbation. This correlation indicates that those with a propensity to stutter during the normal speech task had a more sensitive response to extrinsic auditory feedback timing cues than those who stuttered less. The

fact that, unlike disfluency rate during the experiment, stuttering severity as measured by SSI-Mod did not correlate with responses may indicate that sensitivity to auditory feedback timing cues varies across time, so only the most local measure (i.e., within the same experimental session) of severity has a significant relationship. However, because there was not a significant group difference in response to the timing perturbation, it is difficult to determine how this within-group effect relates to responses of neurotypical speakers.

It should also be noted that this is only the third study to examine online auditory feedback control of formants and timing during a multisyllabic speech utterance and at syllable rates comparable to conversational English. Formant perturbations are generally applied during prolonged vowels (Purcell & Munhall, 2006) or consonant-vowel-consonant syllables (Niziolek & Guenther, 2013; Purcell & Munhall, 2006; Tourville et al., 2008), which is helpful for measuring responses because of the time it takes to process the speech signal and generate the appropriate corrective motor command (on the order of 100—150 ms; Guenther, 2016). At the same time, these studies are not as ecologically valid regarding the role of auditory feedback during normal speaking situations. Cai et al. (2011) and Cai, Beal, et al. (2014) achieved higher ecological validity by embedding perturbations during the course of a multisyllabic phrase. This is especially important for examining feedback control in AWS, as stuttering is more likely to occur during longer phrases compared to single words (e.g., Coalson et al., 2012; Soderberg, 1966). However, Cai et al. (2011) and Cai, Beal, et al. (2014) used stimuli comprised only of vowels and semivowels. The F1 perturbation results in this study

upheld the prior results within a phrase including a mixture of vowels, stop consonants, and fricatives, indicating that auditory feedback does indeed play a role in speech motor control during naturalistic utterances.

Conclusion

The present study demonstrated that a pure temporal auditory feedback perturbation does not elicit the same response deficits in AWS that was previously found in a spectro-temporal perturbation. In addition, syllable-timed speaking may alter the processes involved in responding to spectral auditory feedback perturbations but may not impact the auditory feedback control processes related to speech timing. Specifically, we found that while perturbations of F1 led to a compensatory response in ANS when speaking with natural timing, AWS did not show a significant response. However, during the syllable-timed speech condition, AWS did show a significant response while ANS do not. Additionally, we found that stuttering rate during the task was a significant predictor of responses to timing perturbations in AWS such that AWS who stuttered more had larger responses than those who stuttered less. As stuttering is a developmental speech disorder that generally emerges in early childhood and all participants included herein were adults, it is important to consider that the results of this study may reflect either primary characteristics of stuttering or secondary behaviors developed to compensate for or adapt to these primary characteristics. Future research examining the mechanisms of online auditory feedback processes in children who stutter will be necessary to clarify the roles of disrupted sensorimotor transformations and speech sequence timing in persistent developmental stuttering.

Supplementary Materials

Predictor	*df*	*F*	*p*
Group	3, 25	0.77	0.52
Vowel Duration	3, 25	1.65	0.20
Disfluency Rate	3, 25	0.81	0.50
Condition	3, 25	0.03	0.99
Group x Condition	3, 25	3.17	0.04*
Vowel Duration x Condition	3, 25	0.36	0.78
Disfluency Rate x Condition	3, 25	1.60	0.21

Table 3.S1. Results from an alternative model for predicting formant perturbation responses, substituting Disfluency Rate for SSI-mod. *df* = degrees of freedom, * = $p < 0.05$

Predictor	*df*	*F*	*p*
Group	3, 24	0.44	0.73
Perturbation Magnitude	3, 24	2.07	0.13
Mean Speaking Rate	3, 24	0.89	0.46
Disfluency Rate	3, 24	5.22	0.006**
Condition	3, 24	5.88	0.004**
Group x Condition	3, 24	0.79	0.51
Disfluency Rate x Condition	3, 24	3.48	0.03*

Table 3.S2. Results from an alternative model for predicting formant perturbation responses, substituting Disfluency Rate for SSI-mod. *df* = degrees of freedom, * = $p < 0.05$, ** = $p < 0.01$

CHAPTER IV: Conclusions

The overall aim of this dissertation was to probe the role of speech motor timing in stuttering, including how internal timing mechanisms are disrupted in people who stutter and how external timing cues can be used to increase fluency in these individuals.

In Chapter II, fMRI was used to characterize differences in neural activation and connectivity between externally paced and non-paced speech production in AWS and ANS. The results show that external auditory pacing activates speech timing and temporal processing regions to a greater extent than non-paced speech. In addition, they suggest external pacing recruits a compensatory timing network involving the cerebellum in AWS and potentially modulates top-down motor control and attentional systems and corroborate previous work associating the cerebellum with fluency in adults who stutter.

In Chapter III, formant and timing perturbations of auditory feedback were carried out during externally paced and non-paced speech. In contrast to previous experiments, these perturbations were applied during multi-syllabic utterances comprising both consonants and vowels. Furthermore, while a previous timing perturbation study modulated spectro-temporal feedback of vowel formants, the present work included a pure timing perturbation of phoneme duration and phoneme boundaries within a consonant cluster. The formant perturbation during the non-paced condition replicated the response deficit for AWS found in previous studies. In addition, there were five novel findings: 1) external pacing increased responses to formant perturbations in AWS, 2) external pacing decreased responses to formant perturbations in ANS, 3) in contrast with a spectro-temporal perturbation in a previous study, a pure temporal perturbation did not

lead to a response deficit in AWS, 4) external pacing did not alter responses to timing perturbations, and 5) there was a positive association between disfluency rate and responses to a timing perturbation for AWS. These results indicate that stuttering is related to a deficit in spectral processing rather than purely temporal processing and that externally timed speech can potentially modulate this processing. Together, these studies provide a clearer account of the effects of external timing cues on the neurobehavioral mechanisms underlying speech production in stuttering.

Future Directions

Extending This Work to CWS

The vast majority of research on stuttering and its underlying mechanism, including this dissertation, involves AWS. As stuttering usually emerges during early childhood, measuring neural and behavioral correlates of speech in AWS may either uncover primary markers of stuttering related to its underlying etiology, or secondary characteristics developed in response to stuttering for many years. In addition, many linguistic, cognitive, and physical changes occur during normal development which reduce generality from adults to children. Neural studies have suggested that AWS show right-hemisphere asymmetry, likely associated with compensatory mechanisms (Braun et al., 1997; Foundas et al., 2001; Fox et al., 2000; Preibisch et al., 2003), while CWS do not (Chang et al., 2008). In addition, people who stutter accrue secondary behaviors over time. These behaviors, both overt (e.g., muscle tension and escape behaviors) and covert (e.g. situational avoidance; Guitar, 2014), develop in response to stuttering and/or others'

reactions to their speech. At present, it is unclear whether the reduced responses to the spectral and spectro-temporal perturbations found in AWS in Cai, Beal, et al. (2014), and to the spectral perturbations in the present dissertation, reflect a primary characteristic of stuttering or secondary compensation. Daliri et al. (2018) found that while AWS demonstrated diminished adaptation to a repeated formant perturbation, CWS responses were normal. This may suggest that the diminished responses to spectral auditory feedback alterations are developed after childhood. It does not imply, however, that online speech timing control — a process with a different type of response and different putative brain mechanisms — would behave the same way. Therefore, in order to determine whether the results of this dissertation reflect primary or secondary characteristics of stuttering, future work should examine neural and behavioral correlates of externally paced speech and responses to auditory feedback manipulations in CWS.

Clarifying the Nature of the Rhythm Effect in Stuttering

Theories of stuttering often make certain assumptions regarding the nature and mechanism of the "rhythm effect." A common assumption in many of these theories, including this dissertation, is that the rhythm effect induces fluency by providing an external timing source to a speech system that has difficulty generating its own timing cues (Alm, 2004; Etchell et al., 2014; Guenther, 2016). This external timing hypothesis is supported by the myriad studies on impaired speech timing in stuttering discussed in Chapter I, and can also be used to explain other fluency-inducing conditions like choral speech. However, since AWS have increased fluency when speaking rhythmically, even

in the absence of a simultaneous metronome (as in the studies in this dissertation), it is unclear whether the effect truly reflects timing based on an "external" source. It is often proposed that "external" can be taken broadly to mean a timing pattern stored in memory rather than something that is outside of the speaker's control (Alm, 2004), though this is difficult to test directly.

An alternative suggestion is that the isochronicity of speech, rather than the externality of timing cues, is the important fluency inducing aspect of this effect. According to the variability model (Vmodel) and derived syllable initiation (SI) theory (Packman et al., 2007), the rhythm effect works by reducing variability in syllable stress which simplifies planning speech timing. Future work could test this hypothesis by having AWS speak with a pre-planned rhythm that varies between short and long segments and compare the number of disfluencies with both a unpaced condition and an isochronous condition as in the present work.

Another possibility is that speaking rhythmically diverts a speaker's attention away from monitoring their speech or other stressors in the environment (Vasiç and Wijnen, 2005). Dual-task paradigms have often shown that when an AWS' attention is drawn away from speech they become more fluent, although this depends on the secondary task used (Brocklehurst, 2008). This possibility could be tested by comparing the relative number of disfluencies in metronome-timed speech and speaking with a distractor task that captures equivalent attentional resources to metronome-timed speech.

Overall, clarifying the roles that each of these components of metronome-timed speech play in reducing disfluencies will be important for determining the attributes of

speech that lead to stuttering and the precise nature of internal vs. external timing, and refining the targets of therapeutic intervention.

Stuttering Subtypes

One significant challenge for understanding the mechanisms underlying stuttering is accounting for the large amount of individual and across-study variability found in the neural stuttering literature (Etchell et al., 2018; Wymbs et al., 2013). It has been suggested that this variability may be due to the presence of neural subtypes within the stuttering population (Alm, 2004; Chang & Guenther, 2020; Guenther, 2016). For example, if stuttering is conceptualized as a disruption of the cortico-basal ganglia network mediating motor program initiation, it may be that disruptions in either the cortex, basal ganglia, or the connections between these regions could lead to stuttering (Chang & Guenther, 2020). It is possible that different stuttering subtypes could be associated with different neural responses to metronome-timed speech, making group inferences difficult especially with relatively small sample sizes. Therefore, research specifically aimed at identifying stuttering subtypes using large samples of neural data and determining neural or behavioral markers that could distinguish between those subtypes in individuals is warranted. This research would help clarify the sources of variability in the stuttering literature, provide a clearer picture of intrinsic timing control in stuttering and how external timing sources lead to fluency, and potentially lead to more individualized therapy for people who stutter.

APPENDIX A

Stimulus sentences used in Chapter II and III. The target sentence from Chapter III is bolded.

1. Rice is often served in round bowls.
2. The juice of lemons makes fine punch.
3. The boy was there when the sun rose.
4. Her purse was full of useless trash.
5. Hoist the load to your left shoulder.
6. The young girl gave no clear response.
7. Sickness kept him home the third week.
8. Lift the square stone over the fence.
9. The friendly gang left the drug store.
10. The lease ran out in sixteen weeks.
11. **The steady bat gave birth to pups.**
12. There are more than two factors here.
13. The lawyer tried to lose his case.
14. The term ended late June that year.
15. The pipe began to rust while new.
16. Act on these orders with great speed.

APPENDIX B

Speaking instructions for participants for Chapters II and III

During the experimental session for the study in Chapter III, subjects were shown a PowerPoint presentation that included the following instructions:

"In this experiment, we will ask you to read aloud short sentences in two different ways:

- A Rhythmic way, paced by a regular beat in the earphones you will wear
- A Normal (non-rhythmic) way"

"At the beginning of each trial, before you start reading, you will hear eight beats. Those beats will always be regular."

"In trials of non-rhythmic (normal) speech, the font will be (green/blue) and there will be the word "Normal" above the sentence. Speak normally in these trials."

"In trials of rhythmic speech, the font will be (blue/green) and there will be the word "Rhythm" above the sentence. Speak rhythmically by aligning each syllable (vowel) to a beat."

Prior to the scanning session (Chapter II), they were told the following:

"In the second part of the study you will read sentences either in a rhythmic way or in a natural way. The crosshair (+) is your cue to stop reading. If you feel that you have said the sentence incorrectly, please do not "go back" and try to correct it. Always keep your head and body as still as possible even while reading the sentences. On some trials, instead of sentences, you will see characters you cannot read. During these trials, please look at the characters and keep your head and body as still as possible."

APPENDIX C

Chapter III equations for reference:

The onset of the /ə/ in the word “the” was detected using an adaptive RMS threshold determined by

$$thresh_{/ə/} = \frac{q_{90}(minRMS_{/ðə/}) + q_{10}(maxRMS_{/ðə/})}{2}, \qquad \text{(Equation C1)}$$

where $minRMS_{/ðə/}$ is the set of lowest RMS values from previous trials during production of “the”, $maxRMS_{/ðə/}$ is the set of the highest RMS values from previous trials during production of “the”, and q_x is the x^{th} quantile of the distribution of values.

The onset and offset of the /s/ in the word “steady” were detected using an adaptive RMS threshold determined by

$$thresh_{/s/} = \frac{q_{90}(minRAT_{/ə/}) + q_{10}(minRAT_{/st/})}{2}, \qquad \text{(Equation C2)}$$

where $minRAT_{/ə/}$ is the set of lowest values from previous trials of the ratio between pre-emphasized RMS and non-filtered RMS during production of the /ə/ (in the word “the”), $minRAT_{/st/}$ is the set of lowest values from previous trials of the ratio between pre-emphasized RMS and non-filtered RMS during production of /st/ (in the word “steady”), and q_x is the x^{th} quantile of the distribution of values.

The offset of the /ɛ/ in the word “steady” was detected when both 1) the RMS slope was negative for a duration determined by

$$dur_{/\varepsilon/off} = \frac{mean(maxDur_{negRMSslp})}{2} \qquad \text{(Equation C3)}$$

and 2) the sum of the RMS slope exceeded a value determined by

$$int_{/\varepsilon/off} = \frac{q_{90}(intRMSslp_{/\varepsilon d/}) + q_{10}(intRMSslp_{/di/})}{2} \qquad \text{(Equation C4)}$$

where $maxDur_{negRMSslp}$ is the set of the maximum duration of negative RMS slope intervals from previous trials during production of /di/ in "steady," $intRMSslp_{/\varepsilon d/}$ is the integral of the RMS slopes from the longest negative RMS slope intervals during /ɛd/ in "steady", $intRMSslp_{/di/}$ is the integral of the RMS slopes from the longest negative RMS slope intervals during /di/ in "steady," and q_x is the x^{th} quantile of the distribution of values.

BIBLIOGRAPHY

Ackermann, H. (2008). Cerebellar contributions to speech production and speech perception: Psycholinguistic and neurobiological perspectives. *Trends in Neurosciences*, *31*(6), 265–272. https://doi.org/10.1016/j.tins.2008.02.011

Alario, F.-X., Chainay, H., Lehericy, S., & Cohen, L. (2006). The role of the supplementary motor area (SMA) in word production. *Brain Research*, *1076*(1), 129–143. https://doi.org/10.1016/j.brainres.2005.11.104

Alm, P. A. (2004). Stuttering and the basal ganglia circuits: A critical review of possible relations. *Journal of Communication Disorders*, *37*(4), 325–369. https://doi.org/10.1016/j.jcomdis.2004.03.001

Andersson, J. L. R., Hutton, C., Ashburner, J., Turner, R., & Friston, K. (2001). Modeling Geometric Deformations in EPI Time Series. *NeuroImage*, *13*(5), 903–919. https://doi.org/10.1006/nimg.2001.0746

Andrews, G., Howie, P. M., Dozsa, M., & Guitar, B. E. (1982). Stuttering: Speech pattern characteristics under fluency-inducing conditions. *Journal of Speech and Hearing Research*, *25*(2), 208–216.

Arnstein, D., Lakey, B., Compton, R. J., & Kleinow, J. (2011). Preverbal error-monitoring in stutterers and fluent speakers. *Brain and Language*, *116*(3), 105–115. https://doi.org/10.1016/j.bandl.2010.12.005

Ashburner, J., & Friston, K. J. (2005). Unified segmentation. *NeuroImage*, *26*(3), 839–851. https://doi.org/10.1016/j.neuroimage.2005.02.018

Azrin, N., Jones, R. J., & Flye, B. (1968). A synchronization effect and its application to stuttering by a portable apparatus. *Journal of Applied Behavior Analysis*, *1*(4), 283–295. https://doi.org/10.1901/jaba.1968.1-283

Barbas, H., García-Cabezas, M. Á., & Zikopoulos, B. (2013). Frontal-thalamic circuits associated with language. *Brain and Language*, *126*(1), 49–61. https://doi.org/10.1016/j.bandl.2012.10.001

Barber, V. (1940). Studies in the Psychology of Stuttering, XVI: Rhythm as a Distraction in Stuttering. *Journal of Speech Disorders*, *5*(1), 29–42. https://doi.org/10.1044/jshd.0501.29

Bauer, J. J., Hubbard Seery, C., LaBonte, R., & Ruhnke, L. (2007). Voice F0 responses elicited by perturbations in pitch of auditory feedback in individuals that stutter and controls. *The Journal of the Acoustical Society of America*, *121*(5), 3201–3201. https://doi.org/10.1121/1.4782465

Beal, D. S., Gracco, V. L., Brettschneider, J., Kroll, R. M., & De Nil, L. F. (2013). A voxel-based morphometry (VBM) analysis of regional grey and white matter volume abnormalities within the speech production network of children who stutter. *Cortex*, *49*(8), 2151–2161. https://doi.org/10.1016/j.cortex.2012.08.013

Beal, D. S., Lerch, J. P., Cameron, B., Henderson, R., Gracco, V. L., & De Nil, L. F. (2015). The trajectory of gray matter development in Broca's area is abnormal in people who stutter. *Frontiers in Human Neuroscience*, *9*. https://doi.org/10.3389/fnhum.2015.00089

Behzadi, Y., Restom, K., Liau, J., & Liu, T. T. (2007). A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *NeuroImage*, *37*(1), 90–101. https://doi.org/10.1016/j.neuroimage.2007.04.042

Belyk, M., Kraft, S. J., & Brown, S. (2015). Stuttering as a trait or state—An ALE meta-analysis of neuroimaging studies. *European Journal of Neuroscience*, *41*(2), 275–284. https://doi.org/10.1111/ejn.12765

Belyk, M., Kraft, S. J., & Brown, S. (2017). Stuttering as a trait or a state revisited: Motor system involvement in persistent developmental stuttering. *European Journal of Neuroscience*, *45*(4), 622–624. https://doi.org/10.1111/ejn.13512

Bengtsson, S. L., Ehrsson, H. H., Forssberg, H., & Ullen, F. (2004). Dissociating brain regions controlling the temporal and ordinal structure of learned movement sequences. *European Journal of Neuroscience*, *19*(9), 2591–2602. https://doi.org/10.1111/j.0953-816X.2004.03269.x

Bengtsson, S. L., Ehrsson, H. H., Forssberg, H., & Ullén, F. (2005). Effector-independent voluntary timing: Behavioural and neuroimaging evidence. *European Journal of Neuroscience*, *22*(12), 3255–3265. https://doi.org/10.1111/j.1460-9568.2005.04517.x

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Bernard, J. A., Peltier, S. J., Wiggins, J. L., Jaeggi, S. M., Buschkuehl, M., Fling, B. W., Kwak, Y., Jonides, J., Monk, C. S., & Seidler, R. D. (2013). Disrupted cortico-cerebellar connectivity in older adults. *NeuroImage*, *83*, 103–119. https://doi.org/10.1016/j.neuroimage.2013.06.042

Bernsee, S. (1999, August 18). Time Stretching And Pitch Shifting of Audio Signals – An Overview. *Stephan Bernsee's Blog*. http://blogs.zynaptiq.com/bernsee/time-pitch-overview/

Bloodstein, O. (1972). The Anticipatory Struggle Hypothesis: Implications of Research on the Variability of Stuttering. *Journal of Speech and Hearing Research*, *15*(3), 487–499. https://doi.org/10.1044/jshr.1503.487

Bloodstein, O. (1995). *A handbook on stuttering* (5th ed). Singular Pub. Group.

Bohland, J. W., & Guenther, F. H. (2006). An fMRI investigation of syllable sequence production. *NeuroImage*, *32*(2), 821–841. https://doi.org/10.1016/j.neuroimage.2006.04.173

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

Bothe, A. K., Davidow, J. H., Bramlett, R. E., & Ingham, R. J. (2006). Stuttering Treatment Research 1970–2005: I. Systematic Review Incorporating Trial Quality Assessment of Behavioral, Cognitive, and Related Approaches. *American Journal of Speech-Language Pathology*, *15*(4), 321. https://doi.org/10.1044/1058-0360(2006/031)

Brady, J. P. (1969). Studies on the metronome effect on stuttering. *Behaviour Research and Therapy*, *7*(2), 197–204. https://doi.org/10.1016/0005-7967(69)90033-3

Braun, A. R., Varga, M., Stager, S., Schulz, G., Selbie, S., Maisog, J. M., Carson, R. E., & Ludlow, C. L. (1997). Altered patterns of cerebral activity during speech and language production in developmental stuttering. An H2 (15) O positron emission tomography study. *Brain*, *120*(5), 761–784.

Brocklehurst, P. H. (2008). A Review of Evidence for the Covert Repair Hypothesis of Stuttering. *Contemporary Issues in Communication Sciences and Disorders*, *35*, 25–43.

Brocklehurst, P. H., & Corley, M. (2011). Investigating the inner speech of people who stutter: Evidence for (and against) the Covert Repair Hypothesis. *Journal of Communication Disorders*, *44*(2), 246–260. https://doi.org/10.1016/j.jcomdis.2010.11.004

Brown, S., Ingham, R. J., Ingham, J. C., Laird, A. R., & Fox, P. T. (2005). Stuttered and fluent speech production: An ALE meta-analysis of functional neuroimaging studies. *Human Brain Mapping*, *25*(1), 105–117. https://doi.org/10.1002/hbm.20140

Buckner, R. L., Krienen, F. M., Castellanos, A., Diaz, J. C., & Yeo, B. T. T. (2011). The organization of the human cerebellum estimated by intrinsic functional connectivity. *Journal of Neurophysiology*, *106*(5), 2322–2345. https://doi.org/10.1152/jn.00339.2011

Bullmore, E. T., Suckling, J., Overmeyer, S., Rabe-Hesketh, S., Taylor, E., & Brammer, M. J. (1999). Global, voxel, and cluster tests, by theory and permutation, for a difference between two groups of structural MR images of the brain. *IEEE Transactions on Medical Imaging*, *18*(1), 32–42. https://doi.org/10.1109/42.750253

Burnett, T. A., Freedland, M. B., Larson, C. R., & Hain, T. C. (1998). Voice F0 responses to manipulations in pitch feedback. *The Journal of the Acoustical Society of America*, *103*(6), 3153–3161. https://doi.org/10.1121/1.423073

Cai, S., Beal, D. S., Ghosh, S. S., Guenther, F. H., & Perkell, J. S. (2014). Impaired timing adjustments in response to time-varying auditory perturbation during connected speech production in persons who stutter. *Brain and Language*, *129*, 24–29. https://doi.org/10.1016/j.bandl.2014.01.002

Cai, S., Beal, D. S., Ghosh, S. S., Tiede, M. K., Guenther, F. H., & Perkell, J. S. (2012). Weak responses to auditory feedback perturbation during articulation in persons who stutter: Evidence for abnormal auditory-motor transformation. *PLoS One*, *7*(7), e41830. https://doi.org/10.1371/journal.pone.0041830

Cai, S., Ghosh, S. S., Guenther, F. H., & Perkell, J. S. (2011). Focal Manipulations of Formant Trajectories Reveal a Role of Auditory Feedback in the Online Control of Both Within-Syllable and Between-Syllable Speech Timing. *Journal of Neuroscience*, *31*(45), 16483–16490. https://doi.org/10.1523/JNEUROSCI.3653-11.2011

Cai, S., Tourville, J. A., Beal, D. S., Perkell, J. S., Guenther, F. H., & Ghosh, S. S. (2014). Diffusion imaging of cerebral white matter in persons who stutter: Evidence for network-level anomalies. *Frontiers in Human Neuroscience*, *8*. https://doi.org/10.3389/fnhum.2014.00054

Caruso, A. J., Gracco, V. L., & Abbs, J. H. (1987). A Speech Motor Control Perspective on Stuttering: Preliminary Observations. In H. F. M. Peters & W. Hulstijn (Eds.), *Speech Motor Dynamics in Stuttering* (pp. 245–258). Springer Vienna. https://doi.org/10.1007/978-3-7091-6969-8_17

Chang, S.-E., & Guenther, F. H. (2020). Involvement of the Cortico-Basal Ganglia-Thalamocortical Loop in Developmental Stuttering. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.03088

Chang, S.-E., & Zhu, D. C. (2013). Neural network connectivity differences in children who stutter. *Brain*, *136*(12), 3709–3726. https://doi.org/10.1093/brain/awt275

Chang, S.-E., Erickson, K. I., Ambrose, N. G., Hasegawa-Johnson, M. A., & Ludlow, C. L. (2008). Brain anatomy differences in childhood stuttering. *NeuroImage*, *39*(3), 1333–1344. https://doi.org/10.1016/j.neuroimage.2007.09.067

Chang, S.-E., Horwitz, B., Ostuni, J., Reynolds, R., & Ludlow, C. L. (2011). Evidence of Left Inferior Frontal–Premotor Structural and Functional Connectivity Deficits in Adults Who Stutter. *Cerebral Cortex*, *21*(11), 2507–2518. https://doi.org/10.1093/cercor/bhr028

Chang, S.-E., Kenney, M. K., Loucks, T. M. J., & Ludlow, C. L. (2009). Brain activation abnormalities during speech and non-speech in stuttering speakers. *NeuroImage*, *46*(1), 201–212. https://doi.org/10.1016/j.neuroimage.2009.01.066

Chen, S. H., Liu, H., Xu, Y., & Larson, C. R. (2007). Voice F0 responses to pitch-shifted voice feedback during English speech. *The Journal of the Acoustical Society of America*, *121*(2), 1157–1163.

Chumbley, J., Worsley, K., Flandin, G., & Friston, K. (2010). Topological FDR for neuroimaging. *NeuroImage*, *49*(4), 3057–3064. https://doi.org/10.1016/j.neuroimage.2009.10.090

Civier, O., Bullock, D., Max, L., & Guenther, F. H. (2013). Computational modeling of stuttering caused by impairments in a basal ganglia thalamo-cortical circuit involved in syllable selection and initiation. *Brain and Language*, *126*(3), 263–278. https://doi.org/10.1016/j.bandl.2013.05.016

Coalson, G. A., Byrd, C. T., & Davis, B. L. (2012). The influence of phonetic complexity on stuttered speech. *Clinical Linguistics & Phonetics*, *26*(7), 646–659. https://doi.org/10.3109/02699206.2012.682696

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews. Neuroscience*, *3*(3), 201–215. https://doi.org/10.1038/nrn755

Coull, J. T. (2004). fMRI studies of temporal attention: Allocating attention within, or towards, time. *Cognitive Brain Research*, *21*(2), 216–226. https://doi.org/10.1016/j.cogbrainres.2004.02.011

Craig, A., & Tran, Y. (2006). Fear of speaking: Chronic anxiety and stammering. *Advances in Psychiatric Treatment*, *12*(1), 63–68. https://doi.org/10.1192/apt.12.1.63

Craig, A., & Tran, Y. (2014). Trait and social anxiety in adults with chronic stuttering: Conclusions following meta-analysis. *Journal of Fluency Disorders*, *40*, 35–43. https://doi.org/10.1016/j.jfludis.2014.01.001

Craig, A., Blumgart, E., & Tran, Y. (2009). The impact of stuttering on the quality of life in adults who stutter. *Journal of Fluency Disorders*, *34*(2), 61–71. https://doi.org/10.1016/j.jfludis.2009.05.002

Daliri, A., & Max, L. (2018). Stuttering adults' lack of pre-speech auditory modulation normalizes when speaking with delayed auditory feedback. *Cortex*, *99*, 55–68. https://doi.org/10.1016/j.cortex.2017.10.019

Daliri, A., Wieland, E. A., Cai, S., Guenther, F. H., & Chang, S.-E. (2018). Auditory-motor adaptation is reduced in adults who stutter but not in children who stutter. *Developmental Science*, *21*(2), e12521. https://doi.org/10.1111/desc.12521

Davidow, J. H. (2014). Systematic studies of modified vocalization: The effect of speech rate on speech production measures during metronome-paced speech in persons who stutter: Speech rate and speech production measures during metronome-paced speech in PWS. *International Journal of Language & Communication Disorders*, *49*(1), 100–112. https://doi.org/10.1111/1460-6984.12050

De Nil, L. F., Kroll, R. M., & Houle, S. (2001). Functional neuroimaging of cerebellar activation during single word reading and verb generation in stuttering and nonstuttering adults. *Neuroscience Letters*, *302*(2–3), 77–80. https://doi.org/10.1016/s0304-3940(01)01671-8

De Nil, L. F, Kroll, R. M., Lafaille, S. J., & Houle, S. (2003). A positron emission tomography study of short- and long-term treatment effects on functional brain activation in adults who stutter. *Journal of Fluency Disorders*, *28*(4), 357–380. https://doi.org/10.1016/j.jfludis.2003.07.002

De Nil, L. F., & Abbs, J. H. (1991). Kinaesthetic acuity of stutterers and non-stutterers for oral and non-oral movements. *Brain*, *114*(5), 2145–2158. https://doi.org/10.1093/brain/114.5.2145

De Nil, L. F., Beal, D. S., Lafaille, S. J., Kroll, R. M., Crawley, A. P., & Gracco, V. L. (2008). The effects of simulated stuttering and prolonged speech on the neural activation patterns of stuttering and nonstuttering adults. *Brain and Language*, *107*(2), 114–123. https://doi.org/10.1016/j.bandl.2008.07.003

De Nil, Luc F., Kroll, R. M., & Houle, S. (2001). Functional neuroimaging of cerebellar activation during single word reading and verb generation in stuttering and nonstuttering adults. *Neuroscience Letters*, *302*(2–3), 77–80. https://doi.org/10.1016/s0304-3940(01)01671-8

De Nil, L. F., Kroll, R. M., Kapur, S., & Houle, S. (2000). A Positron Emission Tomography Study of Silent and Oral Single Word Reading in Stuttering and Nonstuttering Adults. *Journal of Speech, Language, and Hearing Research*, *43*(4), 1038–1053. https://doi.org/10.1044/jslhr.4304.1038

Devaraju, D. S., Maruthy, S., & Kumar, A. U. (2020). Detection of Gap and Modulations: Auditory Temporal Resolution Deficits in Adults Who Stutter. *Folia Phoniatrica et Logopaedica*, *72*(1), 13–21. https://doi.org/10.1159/000499565

Diedrichsen, J. (2006). A spatially unbiased atlas template of the human cerebellum. *NeuroImage*, *33*(1), 127–138. https://doi.org/10.1016/j.neuroimage.2006.05.056

Diedrichsen, J., Balsters, J. H., Flavell, J., Cussans, E., & Ramnani, N. (2009). A probabilistic MR atlas of the human cerebellum. *NeuroImage*, *46*(1), 39–46. https://doi.org/10.1016/j.neuroimage.2009.01.045

Diedrichsen, J., Maderwald, S., Küper, M., Thürling, M., Rabe, K., Gizewski, E. R., Ladd, M. E., & Timmann, D. (2011). Imaging the deep cerebellar nuclei: A probabilistic atlas and normalization procedure. *NeuroImage*, *54*(3), 1786–1794. https://doi.org/10.1016/j.neuroimage.2010.10.035

Duffy, J. R. (2013). *Motor speech disorders: Substrates, differential diagnosis, and management* (Third edition). Elsevier.

Eden, G. F., Joseph, J. E., Brown, H. E., Brown, C. P., & Zeffiro, T. A. (1999). Utilizing hemodynamic delay and dispersion to detect fMRI signal change without auditory interference: The behavior interleaved gradients technique. *Magnetic Resonance in Medicine*, *41*(1), 13–20. https://doi.org/10.1002/(SICI)1522-2594(199901)41:1<13::AID-MRM4>3.0.CO;2-T

Etchell, A. C., Civier, O., Ballard, K. J., & Sowman, P. F. (2018). A systematic literature review of neuroimaging research on developmental stuttering between 1995 and 2016. *Journal of Fluency Disorders*, *55*, 6–45. https://doi.org/10.1016/j.jfludis.2017.03.007

Etchell, A. C., Johnson, B. W., & Sowman, P. F. (2014). Behavioral and multimodal neuroimaging evidence for a deficit in brain timing networks in stuttering: A hypothesis and theory. *Frontiers in Human Neuroscience*, *8*. https://doi.org/10.3389/fnhum.2014.00467

Falk, S., Müller, T., & Dalla Bella, S. (2015). Non-verbal sensorimotor timing deficits in children and adolescents who stutter. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00847

Fischl, B., Sereno, M. I., & Dale, A. M. (1999). Cortical Surface-Based Analysis. *NeuroImage*, *9*(2), 195–207. https://doi.org/10.1006/nimg.1998.0396

Floegel, M., Fuchs, S., & Kell, C. A. (2020). Differential contributions of the two cerebral hemispheres to temporal and spectral speech feedback control. *Nature Communications*, *11*(1), 2839. https://doi.org/10.1038/s41467-020-16743-2

Foundas, A. L., Bollich, A. M., Corey, D. M., Hurley, M., & Heilman, K. M. (2001). Anomalous anatomy of speech-language areas in adults with persistent developmental stuttering. *Neurology*, *57*(2), 207–215. https://doi.org/10.1212/WNL.57.2.207

Foundas, A. L., Bollich, A. M., Feldman, J., Corey, D. M., Hurley, M., Lemen, L. C., & Heilman, K. M. (2004). Aberrant auditory processing and atypical planum temporale in developmental stuttering. *Neurology*, *63*(9), 1640–1646. https://doi.org/10.1212/01.WNL.0000142993.33158.2A

Fowler, C. A. (1980). Coarticulation and theories of extrinsic timing. *Journal of Phonetics*, *8*(1), 113–133. https://doi.org/10.1016/S0095-4470(19)31446-9

Fox, P. T., Ingham, R. J., Ingham, J. C., Hirsch, T. B., Downs, J. H., Martin, C., Jerabek, P., Glass, T., & Lancaster, J. L. (1996). A PET study of the neural systems of stuttering. *Nature*, *382*(6587), 158–162. https://doi.org/10.1038/382158a0

Fox, P. T., Ingham, R. J., Ingham, J. C., Zamarripa, F., Xiong, J. H., & Lancaster, J. L. (2000). Brain correlates of stuttering and syllable production. A PET performance-correlation analysis. *Brain: A Journal of Neurology*, *123 (Pt 10)*, 1985–2004.

Frankford, S. A., Heller Murray, E. S., Masapollo, M., Cai, S., Tourville, J. A., Nieto-Castañón, A., & Guenther, F. H. (in press). The neural circuitry underlying the "rhythm effect" in stuttering. *Journal of Speech, Language, and Hearing Research*.

Friston, K. J., Buechel, C., Fink, G. R., Morris, J., Rolls, E., & Dolan, R. J. (1997). Psychophysiological and Modulatory Interactions in Neuroimaging. *NeuroImage*, *6*(3), 218–229. https://doi.org/10.1006/nimg.1997.0291

Garnett, E. O., Chow, H. M., Nieto-Castañón, A., Tourville, J. A., Guenther, F. H., & Chang, S.-E. (2018). Anomalous morphology in left hemisphere motor and premotor cortex of children who stutter. *Brain*, *141*(9), 2670–2684. https://doi.org/10.1093/brain/awy199

Gelfer, M. P., & Young, S. R. (1997). Comparisons of intensity measures and their stability in male and female speakers. *Journal of Voice*, *11*(2), 178–186. https://doi.org/10.1016/S0892-1997(97)80076-8

Giraud, A. (2008). Severity of dysfluency correlates with basal ganglia activity in persistent developmental stuttering. *Brain and Language*, *104*(2), 190–199. https://doi.org/10.1016/j.bandl.2007.04.005

Golfinopoulos, E., Tourville, J. A., & Guenther, F. H. (2010). The integration of large-scale neural network modeling and functional brain imaging in speech motor control. *NeuroImage*, *52*(3), 862–874. https://doi.org/10.1016/j.neuroimage.2009.10.023

Golfinopoulos, E., Tourville, J. A., Bohland, J. W., Ghosh, S. S., Nieto-Castanon, A., & Guenther, F. H. (2011). FMRI investigation of unexpected somatosensory feedback perturbation during speech. *NeuroImage*, *55*(3), 1324–1338. https://doi.org/10.1016/j.neuroimage.2010.12.065

Gracco, V. L., Tremblay, P., & Pike, B. (2005). Imaging speech production using fMRI. *NeuroImage*, *26*(1), 294–301. https://doi.org/10.1016/j.neuroimage.2005.01.033

Greenberg, S., Carvey, H., Hitchcock, L., & Chang, S. (2003). Temporal properties of spontaneous speech—A syllable-centric perspective. *Journal of Phonetics*, *31*(3–4), 465–485. https://doi.org/10.1016/j.wocn.2003.09.005

Grube, M., Cooper, F. E., Chinnery, P. F., & Griffiths, T. D. (2010). Dissociation of duration-based and beat-based auditory timing in cerebellar degeneration. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(25), 11597–11601. https://doi.org/10.1073/pnas.0910473107

Guenther, F. H. (1994). A neural network model of speech acquisition and motor equivalent speech production. *Biological Cybernetics*, *72*(1), 43–53. https://doi.org/10.1007/BF00206237

Guenther, F. H. (1995). Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production. *Psychological Review*, *102*(3), 594–621. https://doi.org/10.1037/0033-295X.102.3.594

Guenther, F. H. (2006). Cortical interactions underlying the production of speech sounds. *Journal of Communication Disorders*, *39*(5), 350–365. https://doi.org/10.1016/j.jcomdis.2006.06.013

Guenther, F. H. (2016). *Neural control of speech*. MIT Press.

Guenther, F. H., & Vladusich, T. (2012). A neural theory of speech acquisition and production. *Journal of Neurolinguistics*, *25*(5), 408–422. https://doi.org/10.1016/j.jneuroling.2009.08.006

Guenther, F. H., Ghosh, S. S., & Tourville, J. A. (2006). Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain and Language*, *96*(3), 280–301. https://doi.org/10.1016/j.bandl.2005.06.001

Guenther, F. H., Hampson, M., & Johnson, D. (1998). A theoretical investigation of reference frames for the planning of speech movements. *Psychological Review*, *105*(4), 611–633. https://doi.org/10.1037/0033-295X.105.4.611-633

Guitar, B. (2014). *Stuttering: An integrated approach to its nature and treatment* (4th ed). Wolters Kluwer Health/Lippincott Williams & Wilkins.

Hagler, D. J., Saygin, A. P., & Sereno, M. I. (2006). Smoothing and cluster thresholding for cortical surface-based group analysis of fMRI data. *NeuroImage*, *33*(4), 1093–1103. https://doi.org/10.1016/j.neuroimage.2006.07.036

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*(3), 213–223. https://doi.org/10.1002/(sici)1097-0193(1999)7:3<213::aid-hbm5>3.0.co;2-n

Handwerker, D. A., Ollinger, J. M., & D'Esposito, M. (2004). Variation of BOLD hemodynamic responses across subjects and brain regions and their effects on statistical analyses. *NeuroImage*, *21*(4), 1639–1651. https://doi.org/10.1016/j.neuroimage.2003.11.029

Hanna, R., & Morris, S. (1977). Stuttering, Speech Rate, and the Metronome Effect. *Perceptual and Motor Skills*, *44*(2), 452–454. https://doi.org/10.2466/pms.1977.44.2.452

Harrington, J. (1988). Stuttering, Delayed Auditory Feedback, and Linguistic Rhythm. *Journal of Speech, Language, and Hearing Research*, *31*(1), 36–47. https://doi.org/10.1044/jshr.3101.36

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*(1), 22–28. https://doi.org/10.1002/hbm.10119

Haslinger, B., Erhard, P., Weilke, F., Ceballos-Baumann, A. O., Bartenstein, P., Gräfin von Einsiedel, H., Schwaiger, M., Conrad, B., & Boecker, H. (2002). The role of lateral premotor–cerebellar–parietal circuits in motor sequence control: A parametric fMRI study. *Cognitive Brain Research*, *13*(2), 159–168. https://doi.org/10.1016/S0926-6410(01)00104-5

Healey, E. C., & Reid, R. (2003). ADHD and stuttering: A tutorial. *Journal of Fluency Disorders*, *28*(2), 79–93. https://doi.org/10.1016/S0094-730X(03)00021-4

Heim, S., Amunts, K., Hensel, T., Grande, M., Huber, W., Binkofski, F., & Eickhoff, S. B. (2012). The Role of Human Parietal Area 7A as a Link between Sequencing in Hand Actions and in Overt Speech Production. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00534

Hickok, G., Buchsbaum, B., Humphries, C., & Muftuler, T. (2003). Auditory–Motor Interaction Revealed by fMRI: Speech, Music, and Working Memory in Area Spt. *Journal of Cognitive Neuroscience*, *15*(5), 673–682. https://doi.org/10.1162/089892903322307393

Hickok, G., Houde, J., & Rong, F. (2011). Sensorimotor Integration in Speech Processing: Computational Basis and Neural Organization. *Neuron*, *69*(3), 407–422. https://doi.org/10.1016/j.neuron.2011.01.019

Hilger, A. I., Zelaznik, H., & Smith, A. (2016). Evidence That Bimanual Motor Timing Performance Is Not a Significant Factor in Developmental Stuttering. *Journal of Speech, Language, and Hearing Research: JSLHR*, *59*(4), 674–685. https://doi.org/10.1044/2016_JSLHR-S-15-0172

Houde, J. F., & Jordan, M. I. (1998). Sensorimotor Adaptation in Speech Production. *Science*, *279*(5354), 1213–1216. https://doi.org/10.1126/science.279.5354.1213

Howell, P. (2004). Effects of delayed auditory feedback and frequency-shifted feedback on speech control and some potentials for future development of prosthetic aids for stammering. *Stammering Research: An on-Line Journal Published by the British Stammering Association*, *1*(1), 31.

Howell, P. (2010). Behavioral effects arising from the neural substrates for atypical planning and execution of word production in stuttering. *Experimental Neurology*, *225*(1), 55–59. https://doi.org/10.1016/j.expneurol.2010.06.012

Howell, P., Au-Yeung, J., & Rustin, L. (1997). Clock and motor variances in lip-tracking: A comparison between children who stutter and those who do not. In W. Hulstijn, H. F. M. Peters, & P. Van Lieshout, *Speech production: Motor control, brain research and fluency disorders* (pp. 573–578). Elsevier Scientific.

Howell, P., Sackin, S., & Rustin, L. (1995). Comparison of speech motor development in stutterers and fluent speakers between 7 and 12 years old. *Journal of Fluency Disorders*, *20*(3), 243–255. https://doi.org/10.1016/0094-730X(94)00011-H

Hutchinson, J. M., & Norris, G. M. (1977). The differential effect of three auditory stimuli on the frequency of stuttering behaviors. *Journal of Fluency Disorders*, *2*(4), 283–293. https://doi.org/10.1016/0094-730X(77)90032-8

IEEE Recommended Practice for Speech Quality Measurements. (1969). *IEEE Transactions on Audio and Electroacoustics*, No. 17, 227–246. https://doi.org/10.1109/IEEESTD.1969.7405210

Ingham, R. J., Fox, P. T., Costello Ingham, J., & Zamarripa, F. (2000). Is Overt Stuttered Speech a Prerequisite for the Neural Activations Associated with Chronic Developmental Stuttering? *Brain and Language*, *75*(2), 163–194. https://doi.org/10.1006/brln.2000.2351

Ingham, R. J., Grafton, S. T., Bothe, A. K., & Ingham, J. C. (2012). Brain activity in adults who stutter: Similarities across speaking tasks and correlations with stuttering

frequency and speaking rate. *Brain and Language*, *122*(1), 11–24. https://doi.org/10.1016/j.bandl.2012.04.002

Jackson, E. S., Yaruss, J. S., Quesal, R. W., Terranova, V., & Whalen, D. H. (2015). Responses of adults who stutter to the anticipation of stuttering. *Journal of Fluency Disorders*, *45*, 38–51. https://doi.org/10.1016/j.jfludis.2015.05.002

Janssen, N., & Mendieta, C. C. R. (2020). The Dynamics of Speech Motor Control Revealed with Time-Resolved fMRI. *Cerebral Cortex*, *30*(1), 241–255. https://doi.org/10.1093/cercor/bhz084

Jenkins, I., Brooks, D., Nixon, P., Frackowiak, R., & Passingham, R. (1994). Motor sequence learning: A study with positron emission tomography. *The Journal of Neuroscience*, *14*(6), 3775–3790. https://doi.org/10.1523/JNEUROSCI.14-06-03775.1994

Kalinowski, J., Stuart, A., Sark, S., & Armson, J. (1996). Stuttering amelioration at various auditory feedback delays and speech rates. *International Journal of Language & Communication Disorders*, *31*(3), 259–269. https://doi.org/10.3109/13682829609033157

Kell, C. A., Neumann, K., Behrens, M., von Gudenberg, A. W., & Giraud, A.-L. (2018). Speaking-related changes in cortical functional connectivity associated with assisted and spontaneous recovery from developmental stuttering. *Journal of Fluency Disorders*, *55*, 135–144. https://doi.org/10.1016/j.jfludis.2017.02.001

Kell, C. A., Neumann, K., von Kriegstein, K., Posenenske, C., von Gudenberg, A. W., Euler, H., & Giraud, A.-L. (2009). How the brain repairs stuttering. *Brain*, *132*(10), 2747–2760. https://doi.org/10.1093/brain/awp185

Kelso, J. A. S., & Tuller, B. (1987). Intrinsic time in speech production: Theory, methodology, and preliminary observations. In E. Keller & M. Gopnik (Eds.), *Motor and sensory processes of language* (pp. 203–222). Erlbaum.

Keuken, M. C., Bazin, P.-L., Crown, L., Hootsmans, J., Laufer, A., Müller-Axt, C., Sier, R., van der Putten, E. J., Schäfer, A., Turner, R., & Forstmann, B. U. (2014). Quantifying inter-individual anatomical variability in the subcortex using 7 T structural MRI. *NeuroImage*, *94*, 40–46. https://doi.org/10.1016/j.neuroimage.2014.03.032

Kim, K. S., Wang, H., & Max, L. (2020). It's About Time: Minimizing Hardware and Software Latencies in Speech Research With Real-Time Auditory Feedback. *Journal of Speech, Language, and Hearing Research*, *63*(8), 2522–2534. https://doi.org/10.1044/2020_JSLHR-19-00419

Kleinow, J., & Smith, A. (2000). Influences of Length and Syntactic Complexity on the Speech Motor Stability of the Fluent Speech of Adults Who Stutter. *Journal of Speech Language and Hearing Research*, *43*(2), 548. https://doi.org/10.1044/jslhr.4302.548

Krauth, A., Blanc, R., Poveda, A., Jeanmonod, D., Morel, A., & Székely, G. (2010). A mean three-dimensional atlas of the human thalamus: Generation from multiple histological data. *NeuroImage*, *49*(3), 2053–2062. https://doi.org/10.1016/j.neuroimage.2009.10.042

Lee, A., & Kawahara, T. (2009, January). Recent Development of Open-Source Speech recognition Engine Julius. *Asia-Pacific Signal and Information Processing Association, 2009 Annual Summit and Conference*.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. MIT Press.

Lieshout, P. van, Ben-David, B., Lipski, M., & Namasivayam, A. (2014). The impact of threat and cognitive stress on speech motor control in people who stutter. *Journal of Fluency Disorders*, *40*, 93–109. https://doi.org/10.1016/j.jfludis.2014.02.003

Loucks, T., & De Nil, L. F. (2006). Anomalous sensorimotor integration in adults who stutter: A tendon vibration study. *Neuroscience Letters*, *402*(1–2), 195–200. https://doi.org/10.1016/j.neulet.2006.04.002

Loucks, T., Chon, H., & Han, W. (2012). Audiovocal integration in adults who stutter: Audiovocal integration in stuttering. *International Journal of Language & Communication Disorders*, *47*(4), 451–456. https://doi.org/10.1111/j.1460-6984.2011.00111.x

Lu, C., Chen, C., Ning, N., Ding, G., Guo, T., Peng, D., Yang, Y., Li, K., & Lin, C. (2010). The neural substrates for atypical planning and execution of word production in stuttering. *Experimental Neurology*, *221*(1), 146–156. https://doi.org/10.1016/j.expneurol.2009.10.016

Lu, C., Chen, C., Peng, D., You, W., Zhang, X., Ding, G., Deng, X., Yan, Q., & Howell, P. (2012). Neural anomaly and reorganization in speakers who stutter: A short-term intervention study. *Neurology*, *79*(7), 625–632. https://doi.org/10.1212/WNL.0b013e31826356d2

Lu, C., Ning, N., Peng, D., Ding, G., Li, K., Yang, Y., & Lin, C. (2009). The role of large-scale neural interactions for developmental stuttering. *Neuroscience*, *161*(4), 1008–1026. https://doi.org/10.1016/j.neuroscience.2009.04.020

Lu, C., Peng, D., Chen, C., Ning, N., Ding, G., Li, K., Yang, Y., & Lin, C. (2010). Altered effective connectivity and anomalous anatomy in the basal ganglia-

thalamocortical circuit of stuttering speakers. *Cortex*, *46*(1), 49–67. https://doi.org/10.1016/j.cortex.2009.02.017

Ludlow, C. L., Rosenberg, J., Salazar, A., Grafman, J., & Smutok, M. (1987). Site of penetrating brain lesions causing chronic acquired stuttering. *Annals of Neurology*, *22*(1), 60–66. https://doi.org/10.1002/ana.410220114

MacKay, D. G., & MacDonald, M. C. (1984). Stuttering as a Sequencing and Timing Disorder. In R. F. Curlee & W. H. Perkins (Eds.), *Nature and treatment of stuttering: New directions* (pp. 261–282). College-Hill Press.

Max, L. (2004). Stuttering and internal models for sensorimotor control: A theoretical perspective to generate testable hypotheses. In B. Maassen, R. Kent, P. Hermann, & P. Van Lieshout (Eds.), *Speech Motor Control: In Normal and Disordered Speech* (pp. 357–387). Oxford University Press.

Max, L., & Daliri, A. (2019). Limited Pre-Speech Auditory Modulation in Individuals Who Stutter: Data and Hypotheses. *Journal of Speech, Language, and Hearing Research*, *62*(8S), 3071–3084. https://doi.org/10.1044/2019_JSLHR-S-CSMC7-18-0358

Max, L., & Yudman, E. M. (2003). Accuracy and Variability of Isochronous Rhythmic Timing Across Motor Systems in Stuttering Versus Nonstuttering Individuals. *Journal of Speech, Language, and Hearing Research*, *46*(1), 146–163. https://doi.org/10.1044/1092-4388(2003/012)

Max, L., Caruso, A. J., & Gracco, V. L. (2003). Kinematic analyses of speech, orofacial nonspeech, and finger movements in stuttering and nonstuttering adults. *Journal of Speech, Language, and Hearing Research: JSLHR*, *46*(1), 215–232.

Max, L., Guenther, F. H., Gracco, V. L., Ghosh, S. S., & Wallace, M. E. (2004). Unstable or Insufficiently Activated Internal Models and Feedback-Biased Motor Control as Sources of Dysfluency: A Theoretical Model of Stuttering. *Contemporary Issues in Communication Science and Disorders*, *31*(Spring), 105–122. https://doi.org/10.1044/cicsd_31_S_105

McClean, M. D., & Runyan, C. M. (2000). Variations in the Relative Speeds of Orofacial Structures with Stuttering Severity. *Journal of Speech Language and Hearing Research*, *43*(6), 1524. https://doi.org/10.1044/jslhr.4306.1524

McKay, L. C., Evans, K. C., Frackowiak, R. S. J., & Corfield, D. R. (2003). Neural correlates of voluntary breathing in humans. *Journal of Applied Physiology*, *95*(3), 1170–1178. https://doi.org/10.1152/japplphysiol.00641.2002

McLaren, D. G., Ries, M. L., Xu, G., & Johnson, S. C. (2012). A generalized form of context-dependent psychophysiological interactions (gPPI): A comparison to standard approaches. *NeuroImage*, *61*(4), 1277–1286. https://doi.org/10.1016/j.neuroimage.2012.03.068

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

Mitsuya, T., MacDonald, E. N., & Munhall, K. G. (2014). Temporal control and compensation for perturbed voicing feedback. *The Journal of the Acoustical Society of America*, *135*(5), 2986–2994. https://doi.org/10.1121/1.4871359

Namasivayam, A. K., & Van Lieshout, P. (2011). Speech compensation in persons who stutter: Acoustic and perceptual data. *Canadian Acoustics*, *39*(3), 150–151.

Neilson, M. D. (1980). *Stuttering and the Control of Speech: A System Analysis Approach* [Ph.D. Thesis, University of New South Wales]. http://handle.unsw.edu.au/1959.4/64085

Neilson, M. D., & Neilson, P. D. (1987). Speech motor control and stuttering: A computational model of adaptive sensory-motor processing. *Speech Communication*, *6*(4), 325–333. https://doi.org/10.1016/0167-6393(87)90007-0

Neilson, P. D., Neilson, M. D., & O'Dwyer, N. J. (1985). Acquisition of motor skill in tracking tasks: Learning internal models. In D. G. Russell & B. Abernethy (Eds.), *Motor Memory and Control*. Human Performance Associates.

Neumann, K., Euler, H. A., Bosshardt, H.-G., Cook, S., Sandrieser, P., & Sommer, M. (2017). The Pathogenesis, Assessment and Treatment of Speech Fluency Disorders. *Deutsches Aerzteblatt Online*. https://doi.org/10.3238/arztebl.2017.0383

Neumann, K., Euler, H. A., Gudenberg, A. W. von, Giraud, A.-L., Lanfermann, H., Gall, V., & Preibisch, C. (2003). The nature and treatment of stuttering as revealed by fMRI. *Journal of Fluency Disorders*, *28*(4), 381–410. https://doi.org/10.1016/j.jfludis.2003.07.003

Neumann, K., Preibisch, C., Euler, H. A., Gudenberg, A. W. von, Lanfermann, H., Gall, V., & Giraud, A.-L. (2005). Cortical plasticity associated with stuttering therapy. *Journal of Fluency Disorders*, *30*(1), 23–39. https://doi.org/10.1016/j.jfludis.2004.12.002

Nieto-Castañón, A. (2020). *Handbook of functional connectivity Magnetic Resonance Imaging methods in CONN*. Hilbert Press.

Nippold, M. A. (2002). Stuttering and Phonology: Is There an Interaction? *American Journal of Speech-Language Pathology*, *11*(2), 99–110. https://doi.org/10.1044/1058-0360(2002/011)

Nippold, M. A. (2012). Stuttering and Language Ability in Children: Questioning the Connection. *American Journal of Speech-Language Pathology*, *21*(3), 183–196. https://doi.org/10.1044/1058-0360(2012/11-0078)

Niziolek, C. A., & Guenther, F. H. (2013). Vowel Category Boundaries Enhance Cortical and Behavioral Responses to Speech Feedback Alterations. *Journal of Neuroscience*, *33*(29), 12090–12098. https://doi.org/10.1523/JNEUROSCI.1008-13.2013

Ogane, R., & Honda, M. (2014). Speech Compensation for Time-Scale-Modified Auditory Feedback. *Journal of Speech, Language, and Hearing Research*, *57*(2). https://doi.org/10.1044/2014_JSLHR-S-12-0214

Olander, L., Smith, A., & Zelaznik, H. N. (2010). Evidence That a Motor Timing Deficit Is a Factor in the Development of Stuttering. *Journal of Speech Language and Hearing Research*, *53*(4), 876. https://doi.org/10.1044/1092-4388(2009/09-0007)

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, *9*(1), 97–113.

Oschkinat, M., & Hoole, P. (2020). Compensation to real-time temporal auditory feedback perturbation depends on syllable position. *The Journal of the Acoustical Society of America*, *148*(3), 1478–1495. https://doi.org/10.1121/10.0001765

Öst, L.-G., Götestam, K. G., & Melin, L. (1976). A controlled study of two behavioral methods in the treatment of stuttering. *Behavior Therapy*, *7*(5), 587–592. https://doi.org/10.1016/S0005-7894(76)80113-X

Packman, A., Code, C., & Onslow, M. (2007). On the cause of stuttering: Integrating theory with brain and behavioral research. *Journal of Neurolinguistics*, *20*(5), 353–362. https://doi.org/10.1016/j.jneuroling.2006.11.001

Parkinson, A. L., Flagmeier, S. G., Manes, J. L., Larson, C. R., Rogers, B., & Robin, D. A. (2012). Understanding the neural mechanisms involved in sensory control of voice production. *NeuroImage*, *61*(1), 314–322. https://doi.org/10.1016/j.neuroimage.2012.02.068

Parrell, B., Agnew, Z., Nagarajan, S., Houde, J., & Ivry, R. B. (2017). Impaired Feedforward Control and Enhanced Feedback Control of Speech in Patients with Cerebellar Degeneration. *The Journal of Neuroscience*, *37*(38), 9249–9258. https://doi.org/10.1523/JNEUROSCI.3363-16.2017

Parrell, B., Lammert, A. C., Ciccarelli, G., & Quatieri, T. F. (2019). Current models of speech motor control: A control-theoretic overview of architectures and properties. *The Journal of the Acoustical Society of America*, *145*(3), 1456–1481. https://doi.org/10.1121/1.5092807

Pellegrino, F., Coupé, C., & Marsico, E. (2011). Across-Language Perspective on Speech Information Rate. *Language*, *87*(3), 539–558. https://doi.org/10.1353/lan.2011.0057

Pellegrino, F., Farinas, J., & Rouas, J.-L. (2004). Automatic Estimation of Speaking Rate in Multilingual Spontaneous Speech. *Speech Prosody 2004*, 517–520.

Perkell, J. S. (2012). Movement goals and feedback and feedforward control mechanisms in speech production. *Journal of Neurolinguistics*, *25*(5), 382–407. https://doi.org/10.1016/j.jneuroling.2010.02.011

Perrachione, T. K., & Ghosh, S. S. (2013). Optimized Design and Analysis of Sparse-Sampling fMRI Experiments. *Frontiers in Neuroscience*, *7*. https://doi.org/10.3389/fnins.2013.00055

Poldrack, R. A., Nichols, T., & Mumford, J. (2011). *Handbook of Functional MRI Data Analysis*. Cambridge University Press. https://doi.org/10.1017/CBO9780511895029

Postma, A., & Kolk, H. (1993). The covert repair hypothesis: Prearticulatory repair processes in normal and stuttered disfluencies. *Journal of Speech and Hearing Research*, *36*(3), 472–487.

Preibisch, C., Neumann, K., Raab, P., Euler, H. A., von Gudenberg, A. W., Lanfermann, H., & Giraud, A.-L. (2003). Evidence for compensation for stuttering by the right frontal operculum. *NeuroImage*, *20*(2), 1356–1364. https://doi.org/10.1016/S1053-8119(03)00376-8

Purcell, D. W., & Munhall, K. G. (2006). Compensation following real-time manipulation of formants in isolated vowels. *The Journal of the Acoustical Society of America*, *119*(4), 2288. https://doi.org/10.1121/1.2173514

Riecker, A., Kassubek, J., Gröschel, 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*(1), 46–53. https://doi.org/10.1016/j.neuroimage.2005.03.046

Riley, G. D. (2008). *SSI-4, Stuttering severity instrument for children and adults* (4th ed.). Pro Ed.

Rottschy, C., Langner, R., Dogan, I., Reetz, K., Laird, A. R., Schulz, J. B., Fox, P. T., & Eickhoff, S. B. (2012). Modelling neural correlates of working memory: A

coordinate-based meta-analysis. *NeuroImage*, *60*(1), 830–846. https://doi.org/10.1016/j.neuroimage.2011.11.050

Russell, M., Corley, M., & Lickley, R. (2005). Magnitude estimation of disfluency by stutterers and nonstutterers. In *Phonological encoding and monitoring in normal and pathological speech* (pp. 248–260). Psychology Press.

Salmelin, R., Schnitzler, A., Schmitz, F., & Freund, H.-J. (2000). Single word reading in developmental stutterers and fluent speakers. *Brain*, *123*(6), 1184–1202. https://doi.org/10.1093/brain/123.6.1184

Saltzman, E. L., & Munhall, K. G. (1989). A Dynamical Approach to Gestural Patterning in Speech Production. *Ecological Psychology*, *1*(4), 333–382. https://doi.org/10.1207/s15326969eco0104_2

Sares, A. G., Deroche, M. L. D., Shiller, D. M., & Gracco, V. L. (2018). Timing variability of sensorimotor integration during vocalization in individuals who stutter. *Scientific Reports*, *8*(1), 16340. https://doi.org/10.1038/s41598-018-34517-1

Sares, A. G., Deroche, M. L. D., Shiller, D. M., & Gracco, V. L. (2019). Adults who stutter and metronome synchronization: Evidence for a nonspeech timing deficit. *Annals of the New York Academy of Sciences*, nyas.14117. https://doi.org/10.1111/nyas.14117

Schubotz, R. I., & von Cramon, D. Y. (2001). Interval and Ordinal Properties of Sequences Are Associated with Distinct Premotor Areas. *Cerebral Cortex*, *11*(3), 210–222. https://doi.org/10.1093/cercor/11.3.210

Schwartze, M., & Kotz, S. A. (2020). Decreased sensitivity to changing durational parameters of syllable sequences in people who stutter. *Language, Cognition and Neuroscience*, *35*(2), 179–187. https://doi.org/10.1080/23273798.2019.1642499

Segawa, J. A., Tourville, J. A., Beal, D. S., & Guenther, F. H. (2015). The Neural Correlates of Speech Motor Sequence Learning. *Journal of Cognitive Neuroscience*, *27*(4), 819–831. https://doi.org/10.1162/jocn_a_00737

Sitek, K. R., Cai, S., Beal, D. S., Perkell, J. S., Guenther, F. H., & Ghosh, S. S. (2016). Decreased Cerebellar-Orbitofrontal Connectivity Correlates with Stuttering Severity: Whole-Brain Functional and Structural Connectivity Associations with Persistent Developmental Stuttering. *Frontiers in Human Neuroscience*, *10*. https://doi.org/10.3389/fnhum.2016.00190

Smith, A. (1999). Stuttering: A unified approach to a multifactorial, dynamic disorder. In N. Bernstein Ratner & E. C. Healey (Eds.), *Stuttering research and practice: Bridging the gap* (pp. 27–44). Psychology Press.

Soderberg, G. A. (1966). The Relations of Stuttering to Word Length and Word Frequency. *Journal of Speech and Hearing Research*, *9*(4), 584–589. https://doi.org/10.1044/jshr.0904.584

Stager, S. V., Denman, D. W., & Ludlow, C. L. (1997). Modifications in Aerodynamic Variables by Persons Who Stutter Under Fluency-Evoking Conditions. *Journal of Speech, Language, and Hearing Research*, *40*(4), 832–847. https://doi.org/10.1044/jslhr.4004.832

Stager, S. V., Jeffries, K. J., & Braun, A. R. (2003). Common features of fluency-evoking conditions studied in stuttering subjects and controls: An PET study. *Journal of Fluency Disorders*, *28*(4), 319–336. https://doi.org/10.1016/j.jfludis.2003.08.004

Starkweather, C. W., & Gottwald, S. R. (1990). The demands and capacities model II: Clinical applications. *Journal of Fluency Disorders*, *15*(3), 143–157. https://doi.org/10.1016/0094-730X(90)90015-K

Starkweather, C. W., Franklin, S., & Smigo, T. M. (1984). Vocal and Finger Reaction Times in Stutterers and Nonstutterers: Differences and Correlations. *Journal of Speech, Language, and Hearing Research*, *27*(2), 193–196. https://doi.org/10.1044/jshr.2702.193

Strick, P. L., Dum, R. P., & Fiez, J. A. (2009). Cerebellum and Nonmotor Function. *Annual Review of Neuroscience*, *32*(1), 413–434. https://doi.org/10.1146/annurev.neuro.31.060407.125606

Takaso, H., Eisner, F., Wise, R. J. S., & Scott, S. K. (2010). The Effect of Delayed Auditory Feedback on Activity in the Temporal Lobe While Speaking: A Positron Emission Tomography Study. *Journal of Speech, Language, and Hearing Research*, *53*(2), 226–236. https://doi.org/10.1044/1092-4388(2009/09-0009)

Teki, S., Grube, M., Kumar, S., & Griffiths, T. D. (2011). Distinct Neural Substrates of Duration-Based and Beat-Based Auditory Timing. *Journal of Neuroscience*, *31*(10), 3805–3812. https://doi.org/10.1523/JNEUROSCI.5561-10.2011

Theys, C., De Nil, L., Thijs, V., van Wieringen, A., & Sunaert, S. (2013). A crucial role for the cortico-striato-cortical loop in the pathogenesis of stroke-related neurogenic stuttering: Neural Network of Neurogenic Stuttering. *Human Brain Mapping*, *34*(9), 2103–2112. https://doi.org/10.1002/hbm.22052

Tichenor, S. E., & Yaruss, J. S. (2019). Stuttering as Defined by Adults Who Stutter. *Journal of Speech, Language, and Hearing Research*, *62*(12), 4356–4369. https://doi.org/10.1044/2019_JSLHR-19-00137

Tichenor, S. E., & Yaruss, J. S. (2021). Variability of Stuttering: Behavior and Impact. *American Journal of Speech-Language Pathology*, *30*(1), 75–88. https://doi.org/10.1044/2020_AJSLP-20-00112

Tourville, J. A., & Guenther, F. H. (2011). The DIVA model: A neural theory of speech acquisition and production. *Language and Cognitive Processes*, *26*(7), 952–981. https://doi.org/10.1080/01690960903498424

Tourville, J. A., Cai, S., & Guenther, F. (2013). *Exploring auditory-motor interactions in normal and disordered speech*. 060180–060180. https://doi.org/10.1121/1.4800684

Tourville, J. A., Reilly, K. J., & Guenther, F. H. (2008). Neural mechanisms underlying auditory feedback control of speech. *NeuroImage*, *39*(3), 1429–1443. https://doi.org/10.1016/j.neuroimage.2007.09.054

Toyomura, A., Fujii, T., & Kuriki, S. (2011). Effect of external auditory pacing on the neural activity of stuttering speakers. *NeuroImage*, *57*(4), 1507–1516. https://doi.org/10.1016/j.neuroimage.2011.05.039

Toyomura, A., Fujii, T., & Kuriki, S. (2015). Effect of an 8-week practice of externally triggered speech on basal ganglia activity of stuttering and fluent speakers. *NeuroImage*, *109*, 458–468. https://doi.org/10.1016/j.neuroimage.2015.01.024

Trajkovski, N., Andrews, C., Onslow, M., O'Brian, S., Packman, A., & Menzies, R. (2011). A phase II trial of the Westmead Program: Syllable-timed speech treatment for pre-school children who stutter. *International Journal of Speech-Language Pathology*, *13*(6), 500–509. https://doi.org/10.3109/17549507.2011.578660

Trajkovski, N., Andrews, C., Onslow, M., Packman, A., O'Brian, S., & Menzies, R. (2009). Using syllable-timed speech to treat preschool children who stutter: A multiple baseline experiment. *Journal of Fluency Disorders*, *34*(1), 1–10. https://doi.org/10.1016/j.jfludis.2009.01.001

Urban, P. P., Marx, J., Hunsche, S., Gawehn, J., Vucurevic, G., Wicht, S., Massinger, C., Stoeter, P., & Hopf, H. C. (2003). Cerebellar Speech Representation: Lesion Topography in Dysarthria as Derived from Cerebellar Ischemia and Functional Magnetic Resonance Imaging. *Archives of Neurology*, *60*(7), 965. https://doi.org/10.1001/archneur.60.7.965

Van Borsel, J., Achten, E., Santens, P., Lahorte, P., & Voet, T. (2003). fMRI of developmental stuttering: A pilot study. *Brain and Language*, *85*(3), 369–376. https://doi.org/10.1016/S0093-934X(02)00588-6

Vasiç, N., & Wijnen, F. (2005). Stuttering as a monitoring deficit. In *Phonological encoding and monitoring in normal and pathological speech* (pp. 226–247). Psychology Press.

Vossel, S., Geng, J. J., & Fink, G. R. (2014). Dorsal and Ventral Attention Systems: Distinct Neural Circuits but Collaborative Roles. *The Neuroscientist*, *20*(2), 150–159. https://doi.org/10.1177/1073858413494269

Watkins, K. E., Smith, S. M., Davis, S., & Howell, P. (2007). Structural and functional abnormalities of the motor system in developmental stuttering. *Brain*, *131*(1), 50–59. https://doi.org/10.1093/brain/awm241

Webster, W. G. (1998). Brain models and the clinical management of stuttering. *Canadian Journal of Speech-Language Pathology and Audiology*, *22*(4), 220–231.

Whitfield-Gabrieli, S., & Nieto-Castanon, A. (2012). Conn: A Functional Connectivity Toolbox for Correlated and Anticorrelated Brain Networks. *Brain Connectivity*, *2*(3), 125–141. https://doi.org/10.1089/brain.2012.0073

Wingate, M. E. (2002). *Foundations of stuttering*. Academic Press.

Worsley, K. J., Marrett, S., Neelin, P., Vandal, A. C., Friston, K. J., & Evans, A. C. (1996). A unified statistical approach for determining significant signals in images of cerebral activation. *Human Brain Mapping*, *4*(1), 58–73. https://doi.org/10.1002/(SICI)1097-0193(1996)4:1<58::AID-HBM4>3.0.CO;2-O

Wymbs, N. F., Ingham, R. J., Ingham, J. C., Paolini, K. E., & Grafton, S. T. (2013). Individual differences in neural regions functionally related to real and imagined stuttering. *Brain and Language*, *124*(2), 153–164. https://doi.org/10.1016/j.bandl.2012.11.013

Xia, K., & Espy-Wilson, C. Y. (2000). *A new strategy of formant tracking based on dynamic programming*. Sixth International Conference on Spoken Language Processing.

Yairi, E., & Ambrose, N. (2013). Epidemiology of stuttering: 21st century advances. *Journal of Fluency Disorders*, *38*(2), 66–87. https://doi.org/10.1016/j.jfludis.2012.11.002

Yairi, E., & Ambrose, N. G. (1999). Early Childhood Stuttering I: Persistency and Recovery Rates. *Journal of Speech Language and Hearing Research*, *42*(5), 1097. https://doi.org/10.1044/jslhr.4205.1097

Yang, Y., Jia, F., Siok, W. T., & Tan, L. H. (2016). Altered functional connectivity in persistent developmental stuttering. *Scientific Reports*, *6*(1). https://doi.org/10.1038/srep19128

Yeo, B. T. T., Krienen, F. M., Sepulcre, J., Sabuncu, M. R., Lashkari, D., Hollinshead, M., Roffman, J. L., Smoller, J. W., Zöllei, L., Polimeni, J. R., Fischl, B., Liu, H., & Buckner, R. L. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of Neurophysiology*, *106*(3), 1125–1165. https://doi.org/10.1152/jn.00338.2011

CURRICULUM VITAE