RHYTHM PERCEPTION AND NEURAL ACTIVATION DIFFERENCES BETWEEN ADULTS WHO DO AND DO NOT STUTTER

By

Elizabeth Ann Wieland

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Communicative Sciences and Disorders—Doctor of Philosophy

2020

ABSTRACT

RHYTHM PERCEPTION AND NEURAL ACTIVATION DIFFERENCES BETWEEN ADULTS WHO DO AND DO NOT STUTTER

By

Elizabeth Ann Wieland

Stuttering is a communicative disorder that involves disruptions to fluent speech, which are characterized by frequent repetition or prolongation of syllables or words, and/or by frequent hesitations or pauses. Prior research has identified a number of hallmarks and deficits associated with stuttering, including generalized timing deficits, and reduced functional connectivity in the rhythm network previously identified to be involved in the perception of musical meter. Building on assumptions of (1) shared neurocognitive resources exist for metrical structure-building for perception and production of auditory patterns in both music and language, and (2) functional similarity exists in processes involved in predictive action-preparation from metrical structure in auditory information, it was predicted that a core deficit in stuttering involves deficiencies in integrating candidate metrical structures with sensory evidence that would support them. To test this prediction, an experiment was designed using a same/different rhythm discrimination task. Critically, half the stimuli provided greater support for the induction of a beat/meter ("simple rhythms"), whereas the other half were matched in interval types but provided less signal-based statistical support for the induction of a beat/meter ("complex rhythms"). Participants were 36 adults who do and do not stutter, and the rhythm discrimination task was done while undergoing functional magnetic resonance imaging. For the behavioral results, statistical analyses using analysis of variance/covariance did not show any significant effects or interactions; however, a linear mixed effects model which accounted for multiple sources of variance revealed poorer performance on complex rhythm discrimination by adults who stutter compared with those that do

not stutter. For the neural results, activation in the core rhythm network during the rhythm discrimination task was observed for both groups (bilateral insula, bilateral STG, bilateral SMA, and bilateral premotor area) for both simple and complex rhythms. However, adults who stutter additionally showed activation in the bilateral putamen and bilateral IFG, suggesting that one or both of these areas may perform a compensatory function in rhythm perception and predictive action-preparation. These results can be interpreted with respect to predictive coding processes in the brain supporting perception, action, and cognition, as well as recent conceptual extensions to auditory processing of music, language, and speech, which propose that linguistic perception and production are "two sides of the same coin." Specifically, it is proposed that listeners attempt to build top-down metrical representations for structured auditory sequences, and during language processing, these top-down metrical representations must be merged with representations of other structures in language to give rise to a coherent overall linguistic representation. It is proposed that a core deficit in stuttering involves deficient processes for integration of top-down metrical/prosodic structure and/or bottom-up sensory indices of dynamic sensorimotor states, toward construction of a coherent overall linguistic representation. Evidence for this proposal comes from findings that: (1) distal context rate and rhythm cues in speech influence metrical/prosodic structures heard across identical acoustic material, thereby influencing goodness-of-fit evaluations of alternative top-down candidate representations of lexico-syntax; (2) a hallmark of stuttering is anomalous white matter connectivity and reduced functional organization of rhythm networks in the brain. This study is the first to investigate non-speech rhythm perception in adults who stutter, and the findings suggest new hypotheses regarding how dynamic connections among brain structures (e.g., basal ganglia, STG) instantiate computations toward the imputation of timing and meter from acoustically variable auditory signals.

ACKNOWLEDGEMENTS

Many people have helped me reach my academic goals during my time at Michigan State University. First, I would like to express my deepest gratitude to my advisor, Dr. Laura Dilley, for her guidance, support, patience, and confidence in my abilities over so many years. I would also like to thank each of my other committee members for their unique contributions to my research and overall education: Dr. Soo-Eun Chang, Dr. J. Devin McAuley, and Dr. J. Scott Yaruss. I additionally wish to thank all the other faculty members of the Communicative Sciences and Disorders Department and Psychology Department who have helped me navigate the challenges of my graduate journey.

I would like to acknowledge the many funding sources that helped me through graduate school and this dissertation. The National Institute on Deafness and other Communication Disorders (NIDCD) grant R01DC011277 (PI: Chang); the National Science Foundation (NSF) grants 0847653 (PI: Dilley) and 1431063 (PI: Dilley); the National Institutes of Health (NIH) grants 5R01HD061458-05 (PI: Redford; Co-I: Dilley), 5R01DC008581-05 (PI: Bergeson-Dana; Co-I: Dilley), and 5R01DC008581-10 (PI: Houston; Co-I: Dilley); and the GRAMMY Foundation (PI: McAuley). Additional resources were given by Michigan State University: Research in Autism, Intellectual and Neurodevelopmental Disabilities (PI: McAuley), Cognitive Science Program, College of Communication Arts & Sciences, Department of Communicative Sciences & Disorders, and the Graduate School.

I would also like to take the opportunity to thank the many lab members I have worked with and become close to in the Speech Perception & Production Lab, the Speech Neurophysiology Lab, and the Timing, Attention, and Perception Lab. I want to especially thank those that helped

me with data collection, recruitment, and overall support: Scarlett Doyle, Saralyn Rubsam, Gregory Spray, Kristin Hicks, Kaitlyn Ayres, Danielle Spannagel, Evamarie Burhnam, and Christopher Heffner.

I owe my deepest gratitude to my loving friends and devoted family who have tirelessly supported me through this arduous journey to a doctorate. Your love and encouragement are truly what made this accomplishment in my life possible. I must thank my parents and sister for instilling in me many of the qualities needed to persevere and accomplish this goliath task, and for their unconditional love and support. Halie and Paul, you have been by my side and cheering me on since grade school; you already know just how much our friendship means and how impossible this task would have been without you. Robert, you have always been my rock, my partner, and my love; thank you for never giving up on me. Aidan and Olivia, you may have extended this journey for me, but you brought me smiles and laughter every step of the way.

TABLE OF CONTENTS

LIST OF TABLES viii

LIST OF FIGURES ix

CHAPTER 1: INTRODUCTION: BACKGROUND AND LITERATURE REVIEW 1
1.1 Brief introduction and structural overview 1
1.2 Rationale for the study: The importance of accuracy in timing perception for action-production, including in speech 3
1.3 Language, speech, and embedded hierarchical structure 6
1.4 Linguistic frameworks for prosody in perception and production 11
1.5 DIVA, GODIVA, and feedforward vs. feedback control in language production 16
1.6 Predictive coding accounts of perception, action, and cognition 22
1.7 Neuroscience of language 25
1.8 Neurodevelopmental disorder of stuttering 30
1.8.1 Stuttering as a multifactorial disorder 30
1.8.2 Proposals for causal factors in stuttering related to feedforward and feedback representations 32
1.8.3 Evidence for deficient sensorimotor feedback or auditory-motor integration mechanisms 35
1.8.4 Linguistic factors in stuttered utterances 39
1.8.5 Neuroscience of language in people who stutter 45
1.9 Production and perception of non-linguistic auditory sequences 48
1.9.1 Auditory-motor connections in rhythm 49
1.9.2 Research with people who stutter 53
1.9.3 Additional neuroanatomical and behavioral observations motivating an fMRI study of rhythm discrimination in people who stutter 60
1.10 Summary and Research Questions 67
1.10.1 Synopsis of evidence and considerations motivating a study of rhythm perception in people who stutter 67
1.10.2 Research Questions 70
1.10.3 Theoretical Framework 71

CHAPTER 2: METHODS 75
2.1 Participants 75
2.2 Speech, language, and cognitive evaluation 76
2.3 Apparatus 77
2.4 Stimuli 78
2.5 Procedure 80
2.6 Data Analysis 81
2.6.1 Behavioral 81
2.6.2 Imaging 86

CHAPTER 3: RESULTS 88
3.1 Behavioral 88

3.1.1 Analysis of Variance (ANOVA)/Analysis of Covariance (ANCOVA) 88
3.1.2 Linear Mixed-Effects Modeling 91
3.2 Imaging 93
3.2.1 Brain activity associated with simple and complex rhythm discrimination 93
3.2.2 Group-specific finding 96
3.2.3 Group Contrasts 99

CHAPTER 4: DISCUSSION 101
4.1 Prior motivation and abbreviated summary 101
4.2 Recap of findings of the research 102
4.3 A predictive coding account of metrical structure-building 105
4.3.1 Introduction 105
4.3.2 Applications of predictive coding to language and speech processing 106
4.3.3 Comparisons of structure and affordance in language and music 109
4.3.4 Predictive coding and building metrical, prosodic, and phonological structures 112
4.3.5 Neural communication across spatially distributed and specialized brain regions 116
4.4 Extending the predictive coding account of metrical structure-building to a hypothesized explanation for stuttering 119
4.4.1 Brief review of core deficits and hallmarks of stuttering 119
4.4.2 Predictive coding framework: Conceptual overview 121
4.4.3 Stimulus properties and recent modeling: Some further preliminaries 122
4.4.4 A conceptual predictive coding account of behavioral results 125
4.4.5 A conceptual predictive coding account of neural results 127
4.4.6 Summary 131
4.6 Study limitations 132
4.7 Clinical implications 134
4.8 Future directions 139
4.9 Conclusions 143

APPENDICES 147
APPENDIX A: Participants who do not stutter table 148
APPENDIX B: Participants who do stutter table 149
APPENDIX C: Advertisement to recruit participants 150
APPENDIX D: Google form questions to filter participants 151
APPENDIX E: Hearing screening form 153
APPENDIX F: Handedness form 154
APPENDIX G: Participant history form 155
APPENDIX H: Participant background form 157
APPENDIX I: Participant strategies form 159

REFERENCES 162

LIST OF TABLES

Table 1: Average (SD) scores for the measures collected relating to background, disfluencies, speech, language and cognitive tests. 77

Table 2: Table of simple and complex rhythm sequences used split by interval. 79

Table 3: Table of possible trial categories in signal detection theory. 82

Table 4: Regions activated during the rhythm discrimination task for simple rhythms. A voxel-wise threshold at p = 0.01 and cluster threshold at 1350 voxels resulted in a corrected p = .05. 94

Table 5: Regions activated during the rhythm discrimination task for complex rhythms. A voxel-wise threshold at p = 0.01 and cluster threshold at 1350 voxels resulted in a corrected p = .05. 95

Table 6: Regions activated during the rhythm discrimination task for the simple-complex rhythm contrasts. A voxel-wise threshold at p = 0.01 and cluster threshold at 1350 voxels resulted in a corrected p = .05. 95

Table 7: Regions activated during the rhythm discrimination task for the adults who stutter during simple and complex rhythms. A voxel-wise threshold at p = 0.01 and cluster threshold at 1350 voxels resulted in a corrected p = .05 ($^{+}$voxel threshold at 900). 97

Table 8: Regions activated during the rhythm discrimination task for the adults who do not stutter during simple and complex rhythms. A voxel-wise threshold at p = 0.01 and cluster threshold at 1350 voxels resulted in a corrected p = .05 ($^{+}$voxel threshold at 900). 98

Table 9: Regions activated during the rhythm discrimination task for Rhythm Type simple (adults who stutter – adults who do not stutter) and complex (adults who stutter – adults who do not stutter) rhythms, and Group adults who do not stutter (simple – complex), and the overall group contrast (adults who stutter – adults who do not stutter). A voxel-wise threshold at p = 0.01 and cluster threshold at 1350 voxels resulted in a corrected p = .05. 100

Table 10: Survey, speech, language, and cognitive evaluation information for participants who do not stutter. 148

Table 11: Survey, speech, language, and cognitive evaluation information for participants who do stutter. 149

LIST OF FIGURES

Figure 1: Schematic of the DIVA model (reproduced from Guenther, 2016, Chapter 3).....18

Figure 2: Schematic of the GODIVA model (reproduced from Guenther, 2016, Chapter 8).....19

Figure 3: A schematic example of the types of rhythmic sequence stimuli used. The numbers represent the relative length of intervals in each sequence with 1 = 220–270 msec (value chosen at random on each trial), in steps of 10 msec. 79

Figure 4: Stimuli used in the experiment. The standard rhythm was played twice, then the comparison rhythm was either the same or different, and then the participant answered "same" or "different".....81

Figure 5: Internal response probability of occurrence curves for noise-alone and signal-plus-noise trials. Since the curves overlap, the internal response for a noise-alone trial may exceed the internal response for a signal-plus-noise trial. Vertical lines correspond to the criterion response. 82

Figure 6: Mean *c* score for the adults who do and do not stutter for simple and complex rhythm types. Error bars show mean ± 1 standard error of the mean (SEM). 88

Figure 7: Mean d′ score for the adults who do and do not stutter for simple and complex rhythm types. Error bars show mean ± 1 SEM. 89

Figure 8: Mean d′ scores for the adults who do and do not stutter for simple and complex rhythm types split by stuttering severity sub-groups. Error bars show mean ± 1 SEM.....91

Figure 9: Proportion of "different" responses for adults who do and do not stutter for simple and complex rhythm types. Error bars show mean ± 1 SEM.....93

Figure 10: Contrasts for simple, complex, and simple – complex rhythms across both groups. Areas with significant activation are labeled. 94

Figure 11: Contrasts for each Group (adults who stutter and adults who do not stutter) at each Rhythm Type (simple and complex). Areas with significant activation are labeled.....96

CHAPTER 1: INTRODUCTION: BACKGROUND AND LITERATURE REVIEW

1.1 Brief introduction and structural overview

Stuttering is a neurodevelopmental disorder often characterized by frequent repetition or prolongation of sounds, syllables, or words, and/or by frequent hesitations or pauses (Ambrose & Yairi, 1999; World-Health-Organization, 2010). Stuttering can affect quality of life; people who stutter may experience feelings of fear, embarrassment, anger, and helplessness while speaking (Yaruss, 2010). A frequent experience for people who stutter is knowing what they want to say but not being able to get the words out (Tichenor & Yaruss, 2018). Much remains to be understood regarding underlying deficits in coordination and timing in people who stutter and how to mitigate the negative impacts (Chang et al., 2016; Chang & Zhu, 2013; Etchell et al., 2014; Wieland et al., 2014), even while the reasons for effectiveness of certain therapies remain poorly understood (Armson & Kiefte, 2008; Bothe et al., 2006; Davidow, 2014; Humeniuk & Tarkowski, 2017; O'Donnell et al., 2008).

The foundation of understanding stuttering as a speech disorder, including how to ameliorate its negative life impacts, rests on rigorous scientific study of both typical and atypical speech processing mechanisms for production and perception. Additionally, it is pertinent to develop detailed characterizations of the ways that neural structures and networks both support these functionalities and may go awry. This dissertation aims to contribute to reducing the knowledge gap by conducting a rigorous experiment collecting behavioral and neural data to test a hypothesis regarding the existence of a predicted rhythm perception deficit in adults who stutter compared to those who do not stutter. The results of this experiment are argued to bear relevance

to understanding how the brain infers timing, structure, prosody/meter and meaning from environmental sound input in order to prepare appropriate actions, such that under certain conditions, convergences of causal mechanisms produce speech disorder.

The introduction and broader dissertation are organized as follows. First, in Section 1.2 I consider some arguments for why studying perceptual processing, including under conditions in which the input is not necessarily natural speech, are valid and often enlightening avenues for scientific investigation. Next, Section 1.2 reviews theoretical and empirical findings with respect to typical speech and language. These findings include different components of hierarchical linguistic representations under conditions of typical speech planning and production, with a particular focus on those aspects thought to be relevant for understanding stuttering (e.g., prosody and phonology). Next, Section 1.3 reviews typical neuroanatomy and neural processing, focusing on issues of how linguistic representations are built in the brain, including highlighting a recent synthesis regarding prosodic structure-building. Next, Section 1.4 reviews hallmarks of the neurodevelopmental disorder of stuttering, including theories and models of causal processes in planning/production of fluent and disfluent speech. The section focuses on a review of prosody- and timing-related literature judged relevant to motivating the experimental paradigm for the study of rhythm processing in people who stutter. Section 1.5 reviews experiments that involve perception/production of auditory information that is non-linguistic in nature, including influential work by Grahn, Patel, Iversen and colleagues (Grahn & Brett, 2007, 2009; Grahn & McAuley, 2009; Grahn & Rowe, 2009, 2013; Patel & Iversen, 2014; Ross et al., 2018) on neural bases of rhythm processing and auditory-motor interactions. Finally, Section 1.6 presents the research questions and specific hypotheses tested in the dissertation experiment and briefly sketches broader recent theoretical developments in neuroscience that motivate these ideas.

The remaining chapters of this dissertation are organized as follows. Chapter 2 presents the methods for a perception experiment which compares rhythm discrimination in adults who stutter with a control group of participants who do not stutter while undergoing functional magnetic resonance imaging (fMRI). Chapter 3 presents the behavioral and neural results for both groups along with rigorous statistical analyses reflecting current and standard best practices. Chapter 4 presents a discussion contextualizing and interpreting results for both groups which draws on diverse, interdisciplinary literatures in language and speech processing, music perception, cognitive neuroscience, linguistics, and stuttering. In particular, the outline of a new account of core deficiencies in stuttering is presented grounded in current research of predictive coding in the brain (cf. Clark, 2013), along with a new hypothesis about how neural structures for the so-called rhythm network (Grahn & Brett, 2007) coordinate to accomplish perception of metrical and prosodic structure. Finally, Chapter 5 highlights the main contributions of this dissertation, including consideration of implications for theoretical understanding, clinical practice and therapy, and future research.

1.2 Rationale for the study: The importance of accuracy in timing perception for action-production, including in speech

The experiment at the center of this dissertation investigates putative timing deficits in people who stutter, compared to those who do not, during a perceptual rhythm discrimination task comprised of tone sequences. The ability to perceive and produce rhythmic patterns in the environment is paramount to language processing and coordinating precise actions for a dynamic response (Dilley & McAuley, 2008; Large & Jones, 1999; McAuley & Jones, 2003; Patel, 2006).

The design of this experiment builds upon multiple published papers implicating the basal ganglia-thalamocortical network structures in disorders of timing, such as stuttering (e.g., Chang et al., 2016; Chang & Zhu, 2013; Wieland et al., 2015) and Parkinson's Disease (PD) (e.g., Grahn & Brett, 2009). Additionally the study utilizes an established rhythm discrimination paradigm and stimuli designed by Grahn and colleagues (2007) that has been used with patients who have PD (Grahn & Brett, 2009), and which was adapted for use with child participants who stutter (Chang et al., 2016; Wieland et al., 2015).

It may be questioned why a task involving perceptual discrimination of tone sequences would be justifiable as a method for testing possible causes of stuttering, which is primarily thought to be a disorder of speech production. Several arguments can be cited in support of this approach. First, numerous studies suggest that stuttering involves a deficit in coordination of the speech system (Chang, 2011; Hulstijn et al., 1992), a general motor or prediction deficit (Daliri & Max, 2015; Max et al., 2003; Zelaznik et al., 1997), and/or a more generalized timing deficit (Boutsen et al., 2000; Chang et al., 2016; Etchell et al., 2017; Etchell et al., 2014; Wieland et al., 2015). This suggests that under the appropriate conditions, timing deficits found in people who stutter, which most obviously manifest during production tasks, might also be observable during perception tasks.

A further argument for the validity of studying perception in people who stutter is that certain common neural pathways exist for transmitting auditory sensory input that unfolds over time from the cochlea (Blecher et al., 2016; Bohland et al., 2010; Dick et al., 2014; Glennon et al., 2020; Guenther et al., 2006; Guenther & Hickok, 2015; Hickok & Poeppel, 2004). These neural structures form a core, common pathway, regardless of whether high-level language centers in the cortex eventually determine those signals to be speech or nonspeech (Falk et al., 2014; Sammler,

2020; Tierney et al., 2018a, 2018b). Indeed, a fine line often exists between perceiving auditory input as garbled, unintelligible noises/tones, or instead as intelligible speech messages, as revealed by the cases of sine wave speech or noise-vocoded speech (Ahissar et al., 2001; Davis et al., 2005; Friederici et al., 2010; Peelle et al., 2013; Pitt et al., 2016; Remez et al., 2011; Remez et al., 1981). The fact that subtle knowledge-based and/or top-down mechanisms can distinguish auditory signals to be intelligible messages or non-speech content, underscores how extant shared neural pathways carry auditory signals, which do not need to first be judged as speech or nonspeech (Blecher et al., 2016; Dick et al., 2014; Guenther et al., 2006; Ravignani et al., 2019).

Further, evidence exists that neural pathways for timing perception form a common basis for responsive planning and/or preparation of precise motor actions across action and perception for both speech and music, though inter-relationships remain poorly understood (Falk & Dalla Bella, 2016; Falk et al., *2017;* Ravignani et al., 2019). Many situations require accurate discernment of event timing for initiating a proper response, regardless of whether that motor response involves articulatory motions or a non-speech gesture. Crucially, however, speech needs to be timed precisely due to the typically fast sequencing and coarticulatory specificity of motor actions involved (Godino-Llorente et al., 2017; Guenther et al., 1999; Perkell et al., 2000; Smith et al., 1993; Ćavar & Lulich, 2020). The typical speaking rate of American English is about 9-14 phonemes per second, and transitions must be well-timed and smooth to ensure speech is perceived as appropriate and fluent (Crystal & House, 1988). Such accurate timing further relies on neural mechanisms that must hold aspects of prior computed speech in memory while facets of upcoming speech are determined and computed on-the-fly. The production of fluent speech is thus an intricate process which involves conveying information acoustically over time by planning phonological structures, including segments, syllables (and stresses), as well as intonation phrases

on multiply dependent timescales within rhythmic hierarchies (Cummins & Port, 1998; Tilsen, 2013; Wagner & Watson, 2010). Nevertheless, speakers often show remarkable abilities to produce and/or adapt speech using a diversity of non-normative neural and/or physical production mechanisms (Garellek, 2020; Harper et al., 2020; Lane et al., 2005; Perkell et al., 1997; Perrachione et al., 2009, 2014; Proctor et al., 2020; Toutios et al., 2020), including in stuttering (Etchell et al., 2014; O'Donnell et al., 2008). In sum, appropriate response timing – whether subsequent actions involve speaking or some other motor act – first requires accurate perception of timing in moments leading up to the response, with the human perception and production system demonstrating remarkable flexibility to source/filter variation and production constraints. These considerations collectively highlight the importance of accuracy in general timing abilities for appropriate and accurate action planning and execution, and further support the premise of the experiment to be described.

1.3 Language, speech, and embedded hierarchical structure

Studies of stuttering and other communicative disorders using non-speech or artificially "constructed" stimuli must consider the complexity of language and speech, especially its hierarchical, embedded nature, in justifying the proposed approach; that is, they must consider the issue of validity (Schiavetti et al., 2010; Stanovich, 2012). Consideration of the validity of links to language may be particularly important, to the extent that a simple perceptual task may be entertained to study the communicative disorder. In this section, I consider issues of the complexity of language, including its sequential and hierarchical structured nature, on the way to further

rationalizing the proposed rhythm perception study investigating core deficits in people who stutter.

Language planning and production is complex, because expressing concepts and thoughts as fluent spoken utterances requires high-level linguistic planning across many layers of representation. These layers include syntax (i.e., the arrangement of words/phrases in language), semantics (i.e., the meanings of lexical units in language), the lexicon (i.e., the stored word units in language), phonology (i.e., the sound patterns of a language), and pragmatics (i.e., situational conventions for appropriate usage components) (e.g., Chomsky & Halle, 1968; Clark, 1996). Additionally, spoken language entails complex dependencies among word selection, word ordering, and the motor execution of speech. Multiple approaches hold that successful linguistic communication requires conceptual preparation from semantic intent, where this preparation ultimately must result in activation and selection of lexical items that are expressed fluently and intelligibly as speech articulation patterns (Bock, 1982; Dell, 1986; Fromkin, 1971; Garrett, 1980; Levelt, 1989; Shattuck-Hufnagel, 1979, 1987). In this section I will discuss some approaches to characterizing language structure, specifically focusing on language production, including both classic and contemporary models prominent in the field.

In the well-known model by Levelt and colleagues (WEAVER++; Levelt et al., 1999), production of words is viewed as carried out in stages that lead from conceptual preparation to the initiation of articulation. Building on Levelt (1989) and Roelofs (1992), conceptual preparation is viewed as the initial stage in production where communicative intentions activate lexical concepts. This leads to the retrieval of a word (i.e., *lemma*) from the mental lexicon. An active lexical concept spreads its activation to its lemma node; lemma selection is viewed as probabilistic and favors selection of the highest activated lemma (Levelt et al., 1999). Upon selection of a lemma, Levelt

et al. (1999) propose that its syntax becomes available for further grammatical encoding, thereby creating the appropriate syntactic environment for the word. Features such as verb tense are viewed as important and/or obligatory and are treated as diacritic parameters to the lemmas.

Upon selection of the syntactic word or lemma, it is proposed that the speaker begins preparation of the appropriate articulatory gestures for the selected word in its prosodic context. The first step is the retrieval of the word's phonological shape from the mental lexicon. According to WEAVER++, accessing the word form entails activation of three kinds of information: the word's morphological composition, its metrical shape, and its segmental composition. For instance, for the verb *escort* in progressive tense, the morphemes <*escort*> and <*ing*> will be accessed. The metrical and segmental properties of these morphemes are then spelled out. This metrical spell-out will include that the verb *escort* has weak-strong stress, and that the bound suffix -*ing* must be metrically weak. Segmental spell-out will include mapping the phonemes in the morphemic content without syllabification or allophonic detail, a step described in the next paragraph.

Levelt et al.'s (1999) model assumes that syllabification is next accomplished, where the morpheme's segments become simultaneously available with labeled links to indicate their correct ordering. The spelled-out segments are inserted into a metrical template successively, leading to a very limited sort of prosodic instantiation (i.e., syllabification). According to Levelt et al.'s (1999) limited prosodification scheme, syllabification of segments is assigned according to largely universal principles, such as onset maximization and sonority gradation; this syllabification process may result, for instance, in a phoneme from the end of one lexical item constituting part of the onset of the following lexical item (e.g., *escort us* might be spoken as *e-scor-tus*).

A near-final step in Levelt et al.'s (1999) approach is phonetic encoding. The theory endeavors only to explain how a phonological word's gestural score is computed, rather than a complete phonetic encoding of the dynamic articulatory gestures. However, the gestural score is assumed to give rise to articulatory gestures involving temporal dynamics specified on different articulatory tiers, such as the glottal tier or nasal tier in articulatory phonology terms (cf. Browman & Goldstein, 1992). In the final step, the gestural score is executed by the articulatory system, the details of which are considered by Levelt et al. (1999) to be beyond the scope of their devised theory.

Note that Levelt et al. (1999) assumed that speakers carryout some degree of "internal speech" monitoring. They considered the extent to which experimental evidence available at that time might support a view of internal speech monitoring as affecting the initial spell-out of segments, the "prosodified" representation, and/or the gestural score of a word. They reviewed experimental evidence supporting the idea that self-monitoring is sensitive to syllable structure, but they left open whether the monitoring process operates left-to-right or instead scans a whole structure. It seems clear that the internal monitoring process they envisioned did not entail registration and adjustment of minute phonetic details of segmental or suprasegmental pronunciation, leading to the view of internal monitoring in their theory being limited or coarse-grained in scope.

A range of other psycholinguistic theories have been proposed to explain the principles that underlie lexical selection toward production (e.g., Chen & Mirman, 2012; Guhe, 2020; Levelt et al., 1999; Oppenheim et al., 2010; Rommers et al., 2020; Tsuboi et al., 2020), including via recent extensions of concepts outlined in Levelt et al. (1999). Notably, these theories specify lexico-semantic interactions (i.e., how speakers may select among semantically-related lexical

alternatives). The theories vary in the nature of their principles and the level of detail they propose, for instance in the extent to which they specify morphological or phonological interactions. In general, these approaches aim to account for lexical choice patterns and data such as reaction times (RTs), and some involve proposals of a neural network model.

A notable approach to modeling lexical selection in speech production comes from Anders et al. (2015). Anders and colleagues (2015) proposed an evidence accumulation model of lexical selection according to which RT data is used to model lexical choice among a number of alternatives (cf. Ratcliff & McKoon, 2008). This model gives rise to a racing evidence accumulation design using multiple accumulators to account for lexical selection among competing alternatives, as an important stage in production. Anders et al.'s (2015) approach allows not only a more accurate and comprehensive modeling of RT data, but it better accords with, and generalizes, neural data involving judgment and decision-making among multiple alternatives in lexical selection toward production. Together with approaches that focus on perception/production links (Bohland et al., 2010), the work of Anders et al. (2015) may a bridge between theories that have focused on evidence accumulation during general behavioral and neural processes of decision-making (Ratcliff et al., 2016) and evidence that the brain is sensitive to gradient degrees of signal-based support for alternative prosodic structures (Dilley & McAuley, 2008; Heffner et al., 2013; Morrill, Dilley, & McAuley, 2014; Morrill, Dilley, McAuley, et al., 2014).

The importance and complexity of prosody for both production and perception of language has not always been appreciated under various dominant paradigms (e.g., Halle, 1983; Kazanina et al., 2018; Marslen-Wilson & Welsh, 1978), although an ever-growing body of work shows that this complexity is increasingly recognized in psycholinguistics (Mitterer et al., 2019; Reinisch et al., 2013; Watson, Arnold, et al., 2008; Watson, Tanenhaus, et al., 2008; Wheeldon & Waksler,

2004). The generation of speech requires the combination of segmental with prosodic information, which must be aligned prior to speech initiation (Bock, 1982; Dell, 1986; Levelt, 1999; Shattuck-Hufnagel, 1987; Shattuck-Hufnagel & Klatt, 1979). In the next section I therefore provide an overview of theories and models of prosody specifically to situate subsequent consideration of how prosody has been dealt with in proposals about language production.

1.4 Linguistic frameworks for prosody in perception and production

Prosody is an essential component of language that lends semantic value, phonological content, and emotional expression (Dilley et al., 2013), in some cases greatly altering the articulatory plan of utterances (Clopper et al., 2018). As such, prosody potentially poses considerable challenges for fluent speech. A closer look at prosody is therefore essential as an important preliminary to any phonetically-informed study of stuttering as a fluency disorder. In this section I will review concepts and notions of prosody and elaborate on some views of how the prosodic component of language is structured, including how prosody fits into select language production models.

Prosody is a term given to the pitch, timing, and rhythm of speech (Ladd, 2008; Lehiste, 1970); some definitions further include voice quality variation (Dilley et al., 1996; Gordon & Ladefoged, 2001; Redi & Shattuck-Hufnagel, 2001). The prosodic components of speech are important for effective (and affective) speech production requiring appropriate timing and pitch variation for emotional-pragmatic expression (Hellbernd & Sammler, 2016). Prosody is often described as the “music of speech” (Lehiste, 1970; Patel et al., 2008), an appellation that appears

to have been increasingly appreciated over time (Falk et al., 2014; Sammler, 2020; Tierney et al., 2018a, 2018b).

Within linguistic frameworks, prosody is most often discussed in terms of two kinds of constructs: *prominence* (sometimes termed 'emphasis' or 'accent') and *phrasing* (sometimes referred to as grouping) (Breen et al., 2012; Dilley, 2005; Dilley & Breen, in press; Dilley & Brown, 2005; Dilley et al., 2005; Ladd, 2008; Lehiste, 1970). Linguistic frameworks for prosody are also generally in agreement that languages have inventories of 'tones' and 'tunes' that correspond to learned patterns of pitch variation that can be used for prominence, grouping via phrasal boundaries, and overall sentence-level prosodic variation (Dilley, 2005; Ladd, 2008; Pierrehumbert, 1980). Regarding prominence, the variation in pitch and duration for individual stressed syllables, together with that due to overall tempo and sentence-level prosodic features, gives rise to *patterns* of perceptual prominence over time which create the metrical *rhythm* for spoken utterances (Brown et al., 2015; Dilley, 2005; Dilley & Breen, in press; Dilley & Pitt, 2010; Hayes, 1995; Lehiste, 1970; Liberman, 1995; Warren, 2000). These suprasegmental components of speech are associated with intonation patterns which can modulate and/or enhance the meaning of, or attention to, the speech signal (Darwin, 1975; Grossman et al., 2010).

A prominent theoretical framework for prosody within linguistics that builds on the idea of a metrical hierarchy is the autosegmental-metrical framework developed by Pierrehumbert and her colleagues (e.g., Beckman & Pierrehumbert, 1986; Dilley, 2005; Dilley & Breen, in press; Ladd, 2008; Pierrehumbert & Beckman, 1988; Pierrehumbert, 1980). Within this framework, prosodic variation in pitch is modeled as abstract H or L tones that combine into a set of stored tunes, the timing of which is precisely coordinated with phonological and syntactic "anchor points" within the linguistic signal corresponding to stressed/prominent syllables or phrasal edges (e.g.,

Dilley et al., 2005) according to a language-specific inventory of tunes (Grice et al., 2000). In the case of prominences, H and L tones can be "starred" (given with "*", e.g., H*), indicating temporal alignment with a stressed syllable, or unstarred, preceding or following a stressed syllable (indicated by "+" before or after the starred tone) but coordinated with the starred tone (e.g., L+H*). In the case of phrasal boundaries, H and L tones may occur singly or in combination at phrase edges, with annotations like "%" to indicate the boundary status, which phonetically is often accompanied by durational lengthening or voice quality change (e.g., glottalization). For instance, LH% would indicate a complex fall-rise pitch movement occurring over the last word(s) at the end of a major intonation phrase.

This dissertation focuses on the perception of metrical structure and beat, which are constructs thought to be similar across language and music (Liberman, 1975; Patel, 2010). In particular, both language and music are thought to be comprised of a hierarchical structure of auditory events differing in relative strength (Hayes, 1995; Patel, 2010). Indeed, a broad consensus exists among linguists and music theorists that all levels of metrical prominence patterns are organized hierarchically (cf. Hayes, 1995; Ladd, 2008; Patel et al., 2008). Further, sequences of metrical prominences in both music and language entail pitch-time dependencies among stressed elements which draw on the same, or similar, shared neural resources and mechanisms across both music and language (Dilley & McAuley, 2008; Jones, 1976; Ladd, 2008; Liberman, 1975; Patel, 2010).

At the word level, polysyllabic words have an internal metrical structure that renders certain syllables more prominent than other syllables, while monosyllabic words have a default level of prominence according to their status as content words (stressed by default) or function words (unstressed by default) (Ladd, 2008). At the phrasal level, rhythms of utterances involve

combining word-level prominence patterns with higher-level aspects of language, including syntax and semantic and pragmatic meanings, to shape but not fully determine the rhythmic structure of language. In terms of hierarchical rhythm, the lowest level is built on the word-level, with more prominence on stressed than unstressed syllables, while higher levels of prominence are built on phrase-level phonological constraints, such as language-specific rules that specify that the degree of phase prominence is greater than word prominence, as well as higher level aspects of language, such as its pragmatics, syntax, and/or semantic content.

Notably, the linguistic "focus" (i.e., the sentence-level intended pragmatic emphasis) must be combined with the stress and phrasing, giving rise to an overall prosodic plan according to which the word must be executed properly. The information structure of a sentence may pragmatically determine which word is under focus, resulting in different patterns of emphasis on words. For example, the spoken phrase "I would like the red *PEN*" – with focus on *PEN* - distinguishes between a pen and other writing instruments and entails a narrow focus on the word 'pen' (or a broad focus on the whole sentence), while "I would like the *RED* pen" pragmatically distinguishes between a red pen vs. other colored pens and entails a focus on the word 'red'. Representations for linguistic stress and phrase-level emphasis or accent often consist of grids of 'x's for syllables indicating relative prominence, with various degrees of metrical prominence or stress attested. This alternation of prominent and non-prominent syllables in speech is a common way of representing the fundamentally rhythmic nature of utterances cross-linguistically (Hayes, 1995; Lehiste, 1970; Liberman, 1995).

Multiple production challenges come with the ongoing need to generate rhythm and phrase-level prosodic structures in connected speech. For instance, executing the lexical stress pattern of a word correctly is made challenging in part by the fact that the acoustic properties of

stress and unstressed syllables involve detailed arrays of pitch, timing, and articulatory variables (Lehiste, 1970). During continuous speech production, individually selected words may have very different metrical patterns (e.g., a two-syllable word consisting of a weak-strong metrical pattern like *cre<u>ate</u>* vs. a three-syllable word with a strong-weak-weak pattern such as *<u>ca</u>nopy*), requiring adjustments on-the-fly to integrate these items continuously within fluent utterances. Further, the production of lexical stress patterns for words interacts with syntactic planning and morphology. For example, the word "record" can be a noun or verb based on the location of the stress (e.g., *<u>re</u>cord* or *re<u>cord</u>*, respectively). Additionally, the planning of an utterance requires grouping words and adding breaks in the necessary locations, which requires planning articulatory trajectories, timing, and pitch. Conveying the intended meaning for a sentence requires the correct words to be emphasized. For example, the phrase "I'll take the eggs benedict" is a specific egg dish vs. "I'll take the EGGS, Benedict" is a statement to a person named Benedict. These different dimensions require precise timing at multiple coordinated timescales of prosodic and syntactic structure including the intonation phrase, the foot, and the syllable.

Other challenges of retrieving and executing linguistic content under normal production demands entail that content must be organized at multiple hierarchically-related timescales (e.g., Sammler, 2018). The lexical stress properties of individually retrieved words (i.e., a faster timescale) must be integrated with sentence-level prosody (i.e., a slower timescale). This integration process can only occur after selection from competing words potentially differing in prosodic properties of stress and/or length (e.g., number of syllables and phonemes) (Brown et al., in press). In general, the complexity of language production planning entails important connections to systems for timing and temporal organization (Cummins & Port, 1998; Lehiste, 1970) which are integral to processes for language perception and production (e.g., Sammler, 2018). All of these

prosodic modifications must further be carried out rapidly in the context of ongoing word retrieval and fluent articulation, potentially posing problems for a person who stutters (Bohland et al., 2010; Civier et al., 2013).

1.5 DIVA, GODIVA, and feedforward vs. feedback control in language production

Having reviewed hierarchical components of linguistic communication and proposals for how these components relate to one another, I will review accounts of the role of auditory feedback in monitoring one's own utterances, as well as feed-forward models or prediction in language. Speech production involves monitoring auditory perceptual feedback of one's own voice, as well as integrating sensorimotor input with one's own ongoing utterances (Ackermann & Riecker, 2004; Levelt, 1999). During production, articulation entails transformation of a phonetic plan for articulator movements into the accurate production of sequences of produced speech sounds (Cai, 2011).

In terms of conceptualizing that speech perception and speech production are coordinated, the terms *feedback* and *feedforward* are relevant. *Feedback* refers to systems for delivering auditory indices of the speaker's speech signals (auditory) and movements (motor) during speech to the brain for potential error correction. Further, *feedforward* refers to linguistic representations and action plans which map the speaker's speech sound system (auditory) and learned motor programs (motor) to articulatory gestures (Cai et al., 2014; Guenther et al., 2006). Further, the term "loop" implies a mutual dependence of one entity on the other. I will use the term *perception-production loop* to refer to the dynamic interplay between perceptual processes that register sensory input from the auditory modality, as well as other modalities (e.g., proprioception)

influencing speech production, and the reverse state of affairs (i.e., the dynamic coordination of feedforward plans with feedback). The perception-production loop is at work, even if it is not part of our conscious awareness or experience (Guenther et al., 2006).

During speech production, a person is receiving both auditory feedback about sound structures being produced, as well as concurrent somatosensory and proprioceptive feedback about articulator position, including the lips and tongue. The Directions Into Velocities of Articulators (DIVA) model is a prominent model of the perception-production loop within the field of communicative sciences and disorders that was developed as a neural network model of speech motor acquisition and speech production (Guenther et al., 2006; see Figure 1). According to the DIVA model, speech movements are controlled and planned primarily in the auditory perceptual domain; further, they are implemented through both feedforward and feedback control (Guenther et al., 2006). The DIVA model posits that errors in speech are corrected based on overt sensory feedback (Golfinopoulos et al., 2010; Guenther et al., 2006; Tourville & Guenther, 2011). A syllable production starts with the activation of a speech sound map, which specifies the goal of speech planning toward the "correct" articulatory movements.

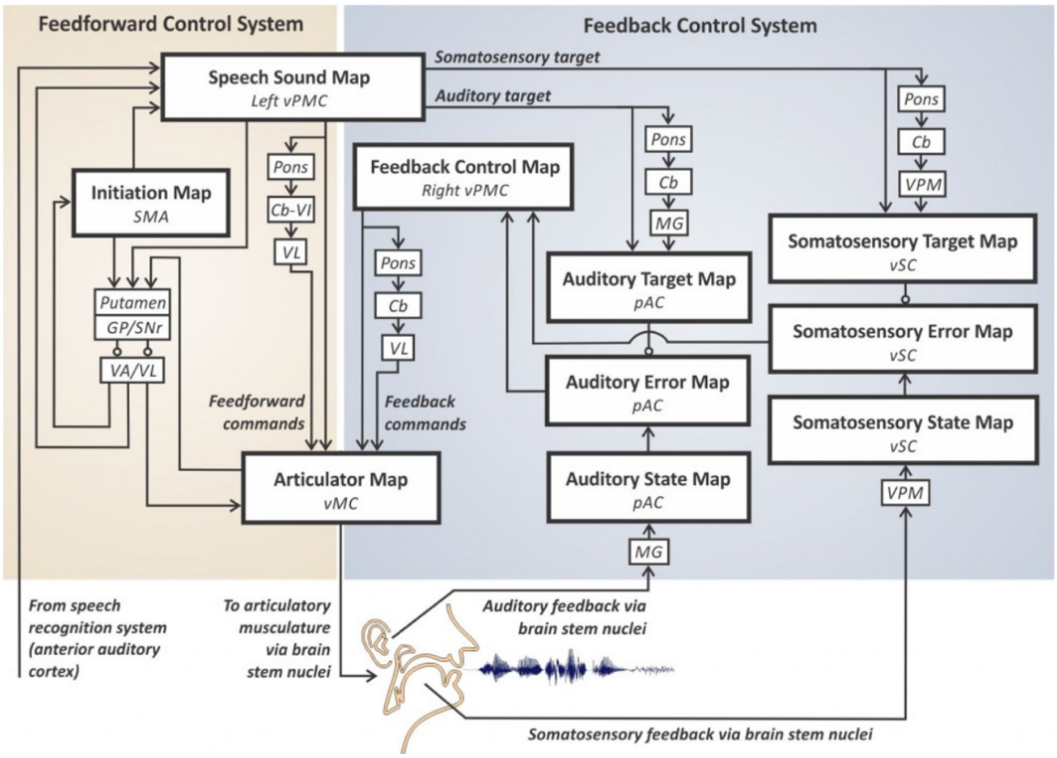

Figure 1: Schematic of the DIVA model (reproduced from Guenther, 2016, Chapter 3)

The related Gradient Order Directions Into Velocities of Articulators (GODIVA) model incorporates phonology in a limited way into the basic DIVA model. GODIVA differs from DIVA by incorporating a notion of syllables in an utterance, such as [stʌ] and [tɚ] in the utterance "stutter" into the proposed sequencing of the neural representations (Bohland et al., 2010; see Figure 2). Both DIVA and GODIVA models specify particular neural structures that are proposed to be involved in the perception-production loop during speech. According to both models, when the speech sound map has been activated, the primary motor cortex receives motor commands from the feedforward and feedback systems, which process the sensory and somatosensory input. The auditory error map then compares the inhibitory signals of the auditory target map and the auditory state map to identify mismatches. The auditory error map is then activated, and corrective signals are sent to the feedback control map. Somatosensory feedback is used to improve the

position of articulators as syllables are attempted to be produced (Civier et al., 2010). The models specify that as the magnitude of the error decreases, the feedforward command relies less on auditory feedback, where this auditory feedback is implemented by a cortical-subcortical-cortical loop between the basal ganglia and the supplementary motor area (SMA). The SMA is responsible for the initiation of speech, and the basal ganglia determine when to launch feedforward commands. GODIVA further theorizes that the pre-SMA contains a representation of the syllable frames (e.g., CCV, CV), while the inferior frontal sulcus contains a representation of the phonemes (e.g., [s], [t], [ʌ], etc.).

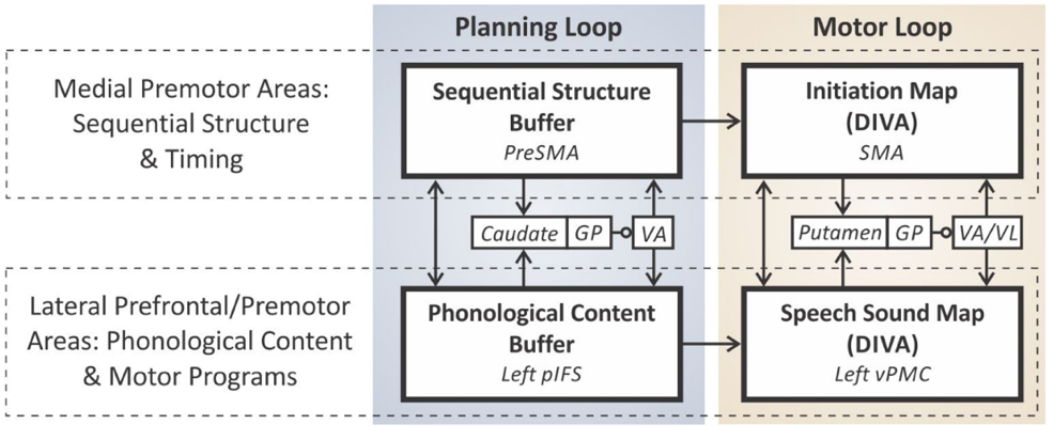

Figure 2: Schematic of the GODIVA model (reproduced from Guenther, 2016, Chapter 8)

DIVA and GODIVA additionally propose that as part of ongoing feedforward planning and proactive error correction, an efferent copy of the articulatory plan is sent to the effector organs. The incoming sensory input (e.g., auditory perceptual and proprioceptive information) is compared with the forward/efferent copy, and dynamic adjustment of motor signals controlling speech or other actions then take place. GODIVA, with its more elaborated treatment of syllable structures, further proposes that a motor command initiates an utterance, and an efference copy is

then sent to the basal ganglia to trigger the transition from one syllable to the next. In this way, the models assume that adjustments to ongoing action sequences can be made to minimize the error between incoming sensory information and the top-down predictions that were instantiated in efferent copies. DIVA and GODIVA build on work by Lashley (1951), who proposed that inherently parallel neural representations underlie serial action, an idea that has been increasingly supported by experimental evidence. Further relevant to speech motor planning, according to (GO)DIVA/ syllables targeted for production are based on high-level language planning via conceptual preparation and lexical selection.

The importance of error monitoring in both the DIVA and GODIVA models raises the question of what kinds of information speakers may monitor in auditory feedback. Recent findings support the possibility that feedback monitoring may include evaluation of the ongoing prosodic structure (e.g., timing and stress). Specifically, it is now well-established that hearing speech induces expectations about the rhythm of upcoming speech (Dilley & Pitt, 2010; Morrill, Dilley, & McAuley, 2014). These expectations include information about whether syllables will be stressed or not (Brown et al., 2015). Prosodic expectations can be so strong that hearing different "distal" prosodic contexts (e.g., inducing a weak-strong-weak vs. strong-weak-strong pattern across identical words) at the beginning of an utterance can induce, with large effect sizes, identical later-occurring "proximal" acoustic material to be heard as comprised of different words with different patterns of stress and segmentation (e.g., *timer derby* vs. *tie murder bee*) (Dilley et al., 2010; Dilley & McAuley, 2008). Likewise, under appropriate conditions manipulations to distal context speech rate can induce powerful prosodic expectations about durations of upcoming syllables and words contained in subsequent speech, causing identical "proximal" acoustic material to be heard as containing more or fewer syllables and/or words (e.g., hearing *leisure or*

time in a normal-rate context vs. hearing *leisure time* in a slowed-rate context, for an utterance produced as *Deena didn't have any leisure or time)* (Baese-Berk et al., 2019; Dilley & Pitt, 2010; Morrill, Dilley, & McAuley, 2014).

The robust, large effects of prosodic context on predictions about the prosodic structure and overall organization of upcoming speech – as demonstrated by experiments showing identical acoustic material being heard as comprised of different words with different prosodic structures and imputed segmentations relative to signals (e.g., Dilley & McAuley, 2008; Dilley & Pitt, 2010) – raises questions about the role of prosodic prediction in ongoing speech monitoring, for both fluent and disfluent utterances. The observation that prosodic expectations generate strong predictions capable of transforming perception of subsequent speech is consistent, however, with the view that ongoing spoken language production involves generation of feedforward expectations about prosodic characteristics like rhythm, timing, and metrical stress. On such a view, speakers generate expectations about the prosodic characteristics of to-be-spoken portions of material during production as a matter of course. Under conditions of presumptive mismatch between feedforward prosodic predictions and imputed sensorimotor feedback, this would be expected to trigger error corrective processes more typically associated with segmental feedback monitoring (Cai et al., 2014; Guenther et al., 2006).

It is notable that neither DIVA nor GODIVA have endeavored to include a prosodic predictive component as part of feedforward or feedback processes. Bohland and colleagues (2010) nevertheless highlight an awareness of this as a weakness of DIVA and GODIVA, stating in concluding sentences, "future instantiations of the GODIVA model should strive to explain how prosody and stress can be encoded at the phonological and phonetic levels" (p. 1529). As such, it appears that a more fully specified approach to dynamic integration of feedforward planning and

feedback as it relates to prosodic structure may be an important contribution to communicative science.

1.6 Predictive coding accounts of perception, action, and cognition

The DIVA and GODIVA models highlight the importance of forward modeling and prediction for both producing and perceiving speech phonemes and syllables. Prediction has become a central theme to much empirical and theoretical work on brain structure and organization in the past decade, which has culminated in a growing consensus around a framework known as *predictive coding*. Predictive coding represents an emergent, empirically-grounded approach to understanding how the brain accomplishes perception, action, and cognition. According to predictive coding, the brain is a "prediction engine"; the brain seeks to predict, and recapitulate, representations that best match external stimuli and sources (Arnal, 2012; Clark, 2013; Friston, 2008; Pickering & Garrod, 2013). On this view, alternative top-down representations "compete", via internal generation or recapitulation of past experiences, to become the best match to (sensory experiences of) external stimuli and sources (Clark, 2013). The search to identify a good match to sensory input is claimed to result in the identification of a "best match" that minimizes the "prediction error" between the top-down representation and the bottom-up input; this identification of the best-match representation in turn results in the conscious experience of "perceiving" that representation (Clark, 2013). Importantly, these predictive and (re)constructive processes underlie both perception and action-related motor planning (Clark, 2013; Pickering & Garrod, 2013).

An important component of predictive coding frameworks is the notion of hierarchical embedding of top-down models which iteratively "explain" bottom-up sensory information at successively "lower" levels. Clark (2013) puts it this way:

> "…one of the brain's key tricks…is to implement dumb processes that correct [error]… Such errors look to be corrected within a cascade …in which higher-level systems attempt to predict the inputs to lower-level ones on the basis of…emerging models of the causal structure of the world (i.e., the signal source). Errors in predicting lower level inputs cause higher-level models to adapt... Such a process…yields a brain that encodes…information about the source of the signals that regularly perturb it…" (p. 182).

Predictive coding approaches have recently begun to be applied to language, which has brought forward understanding of the importance of prediction for all aspects of the language system – including for syntax, semantics, phonotactics, prosody, and other components – across both perception and production (e.g., Dell & Chang, 2014; Pickering & Garrod, 2013). For instance, Pickering and Garrod (2013) advance a predictive coding-inspired approach to production and perception within social communication situations. Their framework highlights the relationship between perception and production of language, and the role of dyadic communication between a speaker and a listener in dynamic programming of production in response to perceived social and linguistic cues. In their view, production is just the implementation of predictive action modeling which occurs during perception. Importantly, Pickering and Garrod's (2013) proposal posits that covert imitative modeling of production, which takes place in an ongoing fashion during perception, involves prediction at *all* levels of linguistic structure in the context of continuous monitoring of the social context.

Another more recent predictive-coding inspired approach for production/planning is Dell and Chang's (2014) P-chain model. Dell and Chang's (2014) starting point is the observation that production is under-studied in psycholinguistics; as such, they argue for developing a framework that emphasizes hypothesized interrelations among psycholinguistic concepts. Rising to the occasion, they posit the following psycholinguistic interrelations: *(1) Processing involves Prediction*. According to this tenet, comprehension entails generation of expectations about upcoming material at multiple linguistic levels. These levels include, but are not limited to, dependencies in predicting the syntax and semantic content of utterances. *(2) Prediction is Production*. According to this idea, production is a top-down process akin to comprehension, since processing flows from intended meaning – the message to levels that encode linguistic forms. This point rests on the notion that prediction and production are related, in that being able to predict language entails being able to use that knowledge to produce language. *(3) Prediction leads to Prediction error*. According to this tenet, expectation meets reality when the words we predict we will hear are not the same as what we actually heard, and this mismatch creates an error signal. *(4) Prediction error creates Priming*. According to this idea, the error signal leads to changes in the prediction system which prime the system for future improvements in accuracy. *(5) Priming is imPlicit learning*. According to this tenet, the priming effect becomes a learned response to the error signal as it has been shown to persist undiminished over time and over unrelated sentences in studies. (*6) imPlicit learning is the mechanism for acquisition / adaptation of Processing, Prediction, and Production*. According to this idea, imPlicit learning becomes the means by which individuals adapt to understand specific contexts and speakers. This is also proposed to be how children acquire the underlying prediction system via a feedback loop that fine tunes the predictions over time via production settings. (7) *Production provides the input for training*

Processing. The final proposition puts a loop in the chain from production back to processing, demonstrating the central influence of production.

In summary, predictive coding represents an empirically-informed development in neuroscience and psycholinguistics that provides a framework for explaining and recasting numerous prior findings. Among them, it provides a new way of thinking about widespread predictive effects in language production and planning, as well as dynamic integration between ongoing input and perception, cognition, and action planning/execution. This framework bears many similarities to notions of feedforward mapping and feedback correction as instantiated in DIVA/GODIVA and other approaches, but suggests that top-down representations operate over general multisensory integration (including visual input), which DIVA/GODIVA does not actively take into account. Overall, predictive coding approaches represent an exciting development in neuroscience that will potentially lead to new insights in understanding the basis of language perception and production in the brain, including the role of prosodic variables (e.g., timing) in fluency disorders like stuttering.

1.7 Neuroscience of language

Arguments about potential core deficits in people who stutter must necessarily be informed by an understanding of neuroanatomy and the brain-behavior relationships that underlie both typical linguistic communication and stuttering. Further, an experiment investigating possible timing deficits in people who stutter should be cognizant of theories of neural processing of language. A particular area for motivational development in this dissertation is how neural areas that support general temporal processing relate to general theories of language organization and

social-cognitive processing in the brain. The current section of this dissertation therefore aims to provide an overview of these topics toward establishing groundwork necessary to motivate the experiment at the core of this research.

It is relatively uncontroversial that language areas responsible for activating, selecting, and implementing top-down concepts as strings of lexical items must effectively communicate with (sub-cortical) brain areas involved in timing and motor control, such as basal ganglia (Chang & Zhu, 2013; Civier et al., 2013). In order to perceive and produce rhythmic patterns in language (or music), a person must attend to the temporal information in their environment and coordinate their actions (Dilley & McAuley, 2008; Large & Jones, 1999; McAuley & Jones, 2003; Patel, 2006). The perception and production of these auditory rhythms in both language and music engage a network of sub-cortical and cortical brain areas that include the basal ganglia, SMA, premotor cortices, auditory cortex, and cerebellum (Bengtsson et al., 2009; Chen et al., 2008; Chen et al., 2006; Grahn & Brett, 2007; Karabanov et al., 2009; Lewis et al., 2004; Mayville et al., 2002; Schubotz et al., 2000; Schwartze & Kotz, 2016). Auditory-motor integration takes place in cortical structures like the inferior frontal gyrus (IFG), ventral motor cortex, ventral premotor cortex, and posterior superior temporal gyrus (STG). During speech production, speech planning and articulatory plans based on sensory feedback are processed in the left IFG (Guenther & Hickok, 2016). An essential point to a better understanding how language is typically organized in the brain and is processed across distributed areas of the brain further requires consideration of connections among these spatially disparate areas (e.g., cortical vs. subcortical structures).

One of the earliest models of language connectivity was proposed by Geschwind (1970), which was then expanded into the Broca-Wernicke-Lichtheim model (reviewed in Dick et al., 2014). The model consisted of the anterior Broca's area, posterior Wernicke's area, and the arcuate

fasciculus connecting the two regions (Dick et al., 2014). Due to the increasingly high spatial and temporal resolution of neuroimaging studies, the classical models are now known to have been overly simplistic, and these are now considered obsolete (Poeppel et al., 2012; Poeppel & Hickok, 2004). As acknowledged by Dick et al. (2014), the older Broca-Wernicke-Geschwind language model has given way to models which acknowledge that language is processed within a distributed cortical and subcortical system. These models emphasize the importance of communications *between* high-level language planning and social monitoring areas and subcortical motor control areas is consistent with a paradigm shift that has been underway in understanding the neurobiology of language.

This theoretical shift in language processing models has led to a new and growing focus on connections between areas involved in language processing has been accompanied by the related paradigm shift in conceptualization of how those connections operate. Specifically, certain pathways – including the arcuate fasciculus, classically assumed to be a language-specific pathway connecting Broca's and Wernicke's areas (Geschwind, 1970) – have been shown to be domain-general (Dick & Tremblay, 2012), rather than being simply dedicated pathways for linguistic information.

Among the empirical methodologies available for neuroscientific study in the modern era, diffusion tensor imaging permits the study of connections between brain areas *directly* by enabling investigation of differential morphology of these white matter fiber tracts in the brain. Due to such advances in neuroimaging, it has now been revealed that speech-language processing diverges into two streams once linguistic information has been processed within auditory regions known as the dual stream model. In particular, the ventral language pathway (i.e., extreme capsule fiber system and uncinate fasciculus) connects the superior temporal sulcus and the posterior inferior temporal

lobe (Hickok & Poeppel, 2004; Saur et al., 2008). This ventral language pathway corresponds to Levelt's (2001; 1989) lemma level of representation. Further, the ventral stream maps auditory speech sounds to meaning and is thought to process less complex syntactic structure (Dick et al., 2014; Hickok, 2009; Hickok & Poeppel, 2000, 2004, 2007; Rauschecker, 2011; Rauschecker & Scott, 2009; Rauschecker & Tian, 2000).

The complementary pathway – namely, the dorsal language stream, consisting of the arcuate fasciculus and superior longitudinal fasciculus – connects temporal lobe regions to the premotor cortex and the IFG (Perani et al., 2011). Empirical findings and theory hold that the dorsal stream is responsible for sensory-to-motor mapping (Hickok & Poeppel, 2004; Saur et al., 2008). Further, the dorsal language pathway has been reported to support phonological awareness (Friederici & Gierhan, 2013; Hickok & Poeppel, 2007; Yeatman et al., 2011) as well as syntax (Friederici, 2011; Friederici & Gierhan, 2013; Skeide & Friederici, 2016). Additional research has suggested that the left superior longitudinal fasciculus and arcuate fasciculus have important roles in combining phonemes into sequences to produce words and phrases (Ries et al., 2019), as well as in supporting speech repetition (Hickok, 2012; Ueno et al., 2011). Within the dorsal stream, activity can be observed in the same brain regions during both speech planning and execution, on the one hand, and speech perception, on the other hand (Callan et al., 2004; Hickok & Poeppel, 2007; Pulvermüller et al., 2006; Vigneau et al., 2006).

This dual stream model also posits that the dorsal pathway for language plays an important role in acquisition of speech. The dorsal stream's function of auditory-motor integration provides for a means of storing sensory representations in speech which can then be compared against articulatory attempts; any resulting mismatch can be used to correct future or ongoing articulation (Hickok & Poeppel, 2004, 2007). The DIVA model similarly holds that acquisition of spoken

language relies on usage of both the auditory and motor systems (Guenther, 1994; Guenther & Perkell, 2004; Guenther & Vladusich, 2012; Tourville & Guenther, 2011). The tight mapping between articulation and acoustic realizations, supported by the dorsal language pathway, thus facilitates children's acquisition of speech and strengthens connections between the auditory and motor systems (Guenther, 1994; Guenther, 1995; Guenther & Vladusich, 2012; Tourville & Guenther, 2011).

Rauschecker and Scott (2009) indicate that the same internal model structures are used for both spatial processing and speech processing, where this dual processing pathway also integrates the spatial (dorsal) pathway with research findings from speech processing as well as music. Additionally, Rauschecker (2011) discussed an updated conceptualization of dorsal stream pathway functions; his characterization of dorsal stream function was as follows:

> "This expanded concept of the dorsal stream not only unifies sensorimotor aspects of space and speech within the auditory domain; it also generalizes dorsal-stream function between vision and audition. In doing so, the revised concept turns some of the conventional wisdom about the dorsal stream on its head: it transforms it from a purely sensory or afferent pathway into an equally efferent pathway, in which predictive motor signals modify activity in sensory structures" (p. 17).

Rauschecker continues by pointing out that this enhanced view of the dorsal stream obviates the need to postulate a third pathway, since aspects of language function in the brain otherwise omitted from description are, via this change, incorporated into the dual pathway concept.

1.8 Neurodevelopmental disorder of stuttering

1.8.1 Stuttering as a multifactorial disorder

Research on stuttering has been invaluable in shedding light on mechanisms of the brain that support communication via spoken language. In particular, the finding that auditory feedback produces different speech behaviors during auditory processing for people who stutter compared to those who do not has been important for constraining and developing theories of speech perception and production (Guenther et al., 2006; Guenther & Hickok, 2016; Houde & Jordan, 2002; Kalinowski et al., 1993; Kalinowski et al., 1996; Natke et al., 2001; Natke et al., 2002; Stuart et al., 2004). Distinct behaviors by people who do and do not stutter have highlighted the importance of the perception-production loop for successful speech production, as well as speech perception.

As discussed in this section, it has been known that certain types of auditory input – providing external rhythmic or metronomic input, for instance – is beneficial in reducing stuttering, yet the reasons why these types of input induce fluency are not fully understood (Greenberg, 1970). Bringing evidence about stuttering to bear on theory generation and hypothesis testing will continue to advance understanding of how the brain supports fluent, spoken linguistic communication. In this section, I review some basic facts about stuttering, including its prevalence, behaviors, and neural hallmarks, before turning to the consideration of fluency-enhancing conditions later on.

Developmental stuttering occurs across all cultures and affects 1% of the adult population (Bloodstein & Bernstein Ratner, 2008; Yairi & Ambrose, 2013). Disfluencies associated with

stuttering typically begin between 2.5 years of age – the age at when children are first putting words together in short phrases – and four years of age. Seventy-five percent of people who stutter will end up recovering spontaneously or with therapy (Bloodstein & Bernstein Ratner, 2008; Yairi & Ambrose, 1999; Yairi & Ambrose, 2013). It is widely held that the cause of stuttering is multifactorial; some accounts further posit that stuttering is a dynamic and complex disorder that involves a combination of motoric, linguistic, and emotional factors (Smith & Weber, 2017; Yaruss, 2010). Although many theoretical models have attempted to account for the possible underlying mechanisms of stuttering, none to date have been able to provide a complete explanation of this complicated disorder (Packman et al., 2007).

The disorder of stuttering is often quantified with reference to specific kinds of disfluent speech productions and adjunct behaviors. Stuttering may be characterized by frequent repetition or prolongation of sounds, syllables, or words, and/or by frequent hesitations or pauses, which disrupt the rhythmic flow of the speech (Ambrose & Yairi, 1999; World-Health-Organization, 2010). Further, stuttering may be measured by exploring primary measurable dimensions, such as frequency, degree, and severity, but other aspects of stuttered speech can also be measured such as rate, pitch, loudness, articulation (Bloodstein, 1995). Although disfluent speech can be identified in both people who do and do not stutter, stuttering-like disfluencies and typical disfluencies are not the same. The term "other disfluencies" will be used in direct contrast with stuttering-like disfluencies (SLDs); other disfluencies represent occasional and natural disruptions of fluent speech (i.e., phrase repetitions, revisions, and interjections), whereas SLDs are abnormal disruptions that only occur as part of disordered speech (Ambrose & Yairi, 1999; Wingate, 1984). SLDs can be generally grouped into three types: repetitions of part words (e.g. *g-g-go green*), repetitions of single-syllable words (e.g., *go-go-go green*), and dysrhythmic phonations (blocks or

prolongations; e.g. …*go green* or *go-o-o green* respectively). Additionally, stuttering can be associated with secondary behaviors including forced effort, eye blinks, quick exhalation of breath, distortion of mouth muscles, frowning, or grimaces (Conture & Kelly, 1991; Wingate, 2002). In adults who stutter, secondary physical concomitants have been shown to increase in frequency with severity of the disorder (Archibald & De Nil, 1999). These secondary behaviors may be viewed as compensatory adaptations, but their co-occurrence with SLDs could also be an indication that stuttering may reflect a more general motor or timing deficit outside of the speech domain (Boutsen et al., 2000; Max et al., 2003; Zelaznik et al., 1997).

1.8.2 Proposals for causal factors in stuttering related to feedforward and feedback representations

This section further elaborates on some of the many factors that have been proposed to cause stuttering. Recognizing that language itself is a kind of feedforward representation, a remarkable number of proposals have asserted that stuttering is, in some sense, a language impairment. Since speakers must hear themselves to produce finely-tuned speech (e.g., Guenther et al., 2006), a further broad range of theories have identified difficulty with stuttering as lying at the juncture of integration of feedforward and feedback representations. For space and efficiency reasons, I will briefly review representative proposals that arguably focus on a language component (e.g., covert repair hypothesis, vicious circle hypothesis, and EXPLAN). I will then turn to consideration of empirical research supporting why stuttering involves unusual and compelling production and perception relationships which highlight feedforward and feedback representation integration as a critical area for scientific research. Because issues of integration of

feedforward and feedback are in fact at the heart of this dissertation, certain topics – such as the neural bases of timing in relation to feedforward and feedback representations – are considered in more depth in later sections.

A number of proposals have honed in on one or more specific levels of feedforward linguistic representation as causal origins of fluency difficulties in stuttering. One such proposal which can be said to exist in this category is the *covert repair hypothesis*, which proposes that breakdowns in fluency of people who stutter reflect attempts to repair phoneme selection errors during speech production (Postma & Kolk, 1993). According to this proposal, such a strategy results in a less efficient speech production system and/or slower repair process than for typical speakers.

The *vicious circle hypothesis* can be considered another approach that situates the causal origins of stuttering with factors lying at the interface of feedforward linguistic plans and feedback corrective processes. This hypothesis states that disfluencies are a result of attempts to internally correct errors (Vasic & Wijnen, 2005). This view holds that people who stutter are prone to be overly sensitive to errors by virtue of imposing excessively strict temporal constraints on the production plan, resulting in heightened likelihood of disfluency. Vasic and Wijnen (2005) further suggest that presentation of external timing cues to persons who stutter can function as a distraction to hyper-vigilant speech monitoring processes, so long as the person monitoring is required to attend to an external rhythm instead of linguistically salient components of a typical speech plan that normally cause distraction.

Further, the EXPLAN hypothesis implicates feedforward processes as causal in stuttering by claiming that errors are a result of an incomplete speech plan at the time of execution (Howell & Au-Yeung, 2002). According to this proposal, the feedforward speech plan formulated by

people who stutter is too slow to keep up with the output speed required for execution. The proposal posits that the slow speech plan results in stalling behaviors, like prolongations/blocks, and advancing behaviors, like repetitions, which occur when the speaker attempts to execute the speech plan despite it being incomplete. Similarly, the *multifactorial dynamic pathways (MDP) theory of stuttering* suggests that increasing linguistic complexity may lead to breakdowns in speech fluency by creating additional stress on neural systems as they interact with speech motor networks functioning in an aberrant way (Smith & Weber, 2017).

A more recent proposal for stuttering known as *Speech And Monitoring Interaction (SAMI)* focuses on the integration between the speech production and the speech monitoring systems as causally influencing stuttering and contextual variability in initiation of speech motor plans (Arenas, 2017). The SAMI framework suggests that fluctuations in vigilance of the monitoring system (i.e., attention) are the main source of contextual variability of stuttering. Like other proposals that hone in on issues of integration of top-down, feedforward linguistic planning and representations, and bottom-up, sensorimotor feedback, the SAMI framework posits that a hyper-vigilant monitoring system hinders the efficiency of the speech production system, resulting in an increased chance of a stuttering-like disfluency.

An even larger number of theories than these have invoked the juncture of feedforward, predictive (top-down) representations and ongoing (bottom-up) sensorimotor feedback as problematic in stuttering (Civier et al., 2010; Hickok et al., 2011; Max et al., 2004; Neilson & Neilson, 1987). Because some of these crucially rely on neural mechanisms, further discussion will be considered after issues related to general neuroanatomy and functional integration of linguistic representations have been considered. In the following, I turn to consideration of evidence supporting the scientific significance of questions regarding how feedforward processes

are integrated with feedback both for understanding stuttering as a disorder, as well as for understanding the organization of language and cognition in the brains of people who do and do not stutter.

1.8.3 Evidence for deficient sensorimotor feedback or auditory-motor integration mechanisms

Stuttering as a disorder has contributed much to understanding the role of predictive processes during speech production (Guenther, 2016). Part of the reason for this is the compelling ways in which modified auditory feedback presented to a person who stutters differentially changes communication behavior, relative to a person who does not. Other significant findings with clinical importance relate to ways in which external pacing signals reduce disfluency in people who stutter. These basic scientific and clinical findings form an important body of evidence implicating causal processes in stuttering, at least in part, to the juncture of feedforward predictive and feedback integrative processes during vocal control (Behroozmand et al., 2020; Behroozmand et al., 2018; Burnett et al., 1998; Daliri et al., 2018; Houde & Jordan, 1998; Liu et al., 2010; Perkell et al., 2000; Villacorta et al., 2007).

One of the most well-known fluency-inducing conditions in the field of stuttering involves real auditory feedback which has been digitally altered in real time to change its acoustic properties. Auditory feedback is interesting and valuable for research because it is the most accessible and easy-to-manipulate part of the speech chain, and manipulation of auditory feedback establishes causal relationships between different events and components of the speech process (Cai, 2011). Research has found that the influence of feedback on speech production in people who stutter relates to the fact that for people who stutter, presented with auditory feedback of their

own voices acoustically manipulated to have a time delay, often induces fluency (e.g., Cai et al., 2014). By contrast, for people who do not stutter, such time-delayed auditory feedback usually induces disfluencies (e.g., Jones & Striemer, 2007; Van Borsel et al., 2005). Moreover, people who stutter who are presented with pitch-shifted auditory feedback of their own voices likewise experience induced fluency (e.g., Loucks et al., 2012; Natke et al., 2001).

Experimental lab-based findings of the ameliorative effects of delayed auditory feedback on fluency were a compelling result that contributed to the development of a commercialized device known as the SpeechEasy in the early 2000's (Armson & Kiefte, 2008; O'Donnell et al., 2008). Disappointing however, were the findings that the fluency-inducing effects gained with utilizing this device reduced over a period of months (Foundas et al., 2013; Pollard et al., 2009). The fluency-inducing effects that were observed in a randomized clinical trial were commensurate with relative to standard behavioral techniques (e.g., fluency shaping), but not significantly greater (Ritto et al., 2016).

One of the most well-known fluency-inducing conditions for people who stutter involves speaking in time with an external pacing signal such as a metronome (Bothe et al., 2006; Davidow, 2014; Greenberg, 1970; Humeniuk & Tarkowski, 2017). The external pacing signal of a metronome or similar device can produce a synchronization effect, allowing people who stutter to entrain their motor responses to the beat of ongoing auditory events and speak with greater fluency (Azrin et al., 1968). Related to this is speaking rhythmically (Law et al., 2018). An external pacing signal, such as a metronome, or speaking rhythmically have both been argued to reduce demands on people who stutter to compute prosodic representations; however, surprisingly few details have been offered to support these proposals.

The basal ganglia have been implicated as an important neural mechanism in understanding the basis of effectiveness of external pacing signal presentation as a fluency-enhancing condition, in that the basal ganglia in people who stutter respond strongly to external rhythmic pacing signals (Toyomura et al., 2011, 2015). In fact, in the presence of a metronome the basal ganglia activity becomes normalized; by contrast, without this fluency-enhancing condition, neural activity is often reduced in the basal ganglia, compared to adults who do not stutter (Toyomura et al., 2011, 2015). Such findings are noteworthy in light of the fact that the basal ganglia are also important neural regions for timing of self-paced movements (Grahn & Brett, 2007; Grahn & Rowe, 2009), including the timing and sequencing of speech (Alm, 2004; Civier et al., 2013; Fujii & Wan, 2014; Jin & Costa, 2015; Schirmer, 2004).

The above findings are suggestive of further work implicating basal ganglia in fluency vs. disfluency in people who stutter. Specifically, in children who stutter, the basal ganglia exhibit structural abnormalities (Beal et al., 2013) as well as functional abnormalities (Chang et al., 2016; Chang & Zhu, 2013). Further, children who stutter are significantly less accurate at producing (Falk et al., 2015; Howell et al., 1997; Olander et al., 2010) and discriminating rhythms (Wieland et al., 2015) than children who do not stutter. Rhythm-based treatments are quite effective in children who stutter; usage of a syllable-timed speech technique resulted in a mean stuttering reduction of 96% in most preschool aged children who stutter (Trajkovski et al., 2011). These findings suggest that investigations into rhythm perception and production may provide insight into underlying causes of stuttering.

It has frequently been proposed that people who stutter rely more heavily on a feedback-based motor control strategy than people who do not, due to the former group having an impaired feedforward control system (Civier et al., 2010; De Nil et al., 2001; Kalveram & Jäncke, 1989;

Zimmermann, 1980). For instance, Max and colleagues (2004) suggest that people who stutter may have failed to acquire correct or stable connections between motor programs and sensory outcomes. According to these researchers, such unstable connections may result in difficulty by people who stutter to actually utilize these mappings for sensorimotor control of their speech output.

DIVA and GODIVA, as well-developed models of the auditory-motor interface during spoken communication, further contextualize understanding of feedforward and feedback processes in stuttering in a way that motivates the present study. In particular, Guenther and colleagues have argued that an over-reliance on auditory feedback and/or insufficiently robust feed-forward commands for generating upcoming rhythmic structures. In support of this proposition, they demonstrated through modeling that an over-reliance on auditory feedback could potentially lead to a variety of stuttering behaviors, including sound-syllable repetitions (Civier et al., 2010) and blocks (Civier et al., 2013). The DIVA model can simulate stuttering-like disfluencies by utilizing a self-repair controller that resets the current speech motor plan during an excessive amount of accumulated error signals (Civier et al., 2010). This model can simulate speech motor system deficits evidenced by people who stutter in terms of the sending of inaccurate predictions to the auditory system used for comparison in feedback. These inaccurate predictions may be a possible outcome of a malfunction in ventral primary motor cortex which results in "impairing" the feedforward motor system and amplifying the resulting error signals (Civier et al., 2013).

Further, the GODIVA model, which models speech motor control at the level of the syllable, was further used in modelling work by Civier and colleagues to investigate hypotheses regarding causal deficits in stuttering (Civier et al., 2013). For instance, extant evidence suggests

that delays in the speech motor plan initiation may be due to too much dopamine in the striatum (a region in the basal ganglia) in a manner that leads to abnormal integration of feedforward and feedback information associated with a deficit in connectivity between the primary motor cortex and the basal ganglia (Civier et al., 2013; Maguire et al., 2004). Civier et al. (2013) demonstrated how GODIVA could generate stuttering-like disfluencies under these conditions, such that stuttering-like disfluencies may occur when a speaker attempts to produce a transition from one syllable to the next at the wrong time.

The above review of theoretical and empirical work critically points to atypical integration of feedforward commands for vocal and linguistic control with sensorimotor feedback integrative processes. This review supports the contention that a better understanding of how these feedforward versus feedback integrative processes relate to language and general audition has the potential to inform not only the understanding of effectiveness of certain fluency-enhancing conditions and the potential for developing better treatments for stuttering, but also for the general understanding of predictive processes for language and communication in the brains of people who do not stutter. In the following, I consider how components of language representations come into play, specifically taking up the topic of how linguistic factors related to prosody have been shown to affect people who stutter.

1.8.4 Linguistic factors in stuttered utterances

The prior section reviewed proposals and evidence implicating language factors and feedforward-feedback integration issues and their roles in stuttering. While linguistic stress and complexity are both features of normal spoken language, certain proposals focus attention on these

components of language as potential triggers for stuttering-like disfluencies in speakers whose neural mechanisms for processing speech are compromised (e.g., through inefficient transmission in connective white fibers) or have attenuated functional connectivity in relevant pathways (Chang et al., 2011; Cykowski et al., 2010). In this section I briefly review hypotheses and proposals for how prosodic factors in language have been claimed to influence stuttering, providing short descriptions for space and efficiency reasons. Unlike the prior proposals, the below approaches provide little to no direct emphasis on issues of integration of feedforward and feedback information for communication processes.

The *Variability model (Vmodel)* was initially proposed to explain why prolonged speech was such an effective fluency-inducing technique. When examining prolonged speech more closely, the researchers noticed this type of speech reduced variation in stress across syllables (Packman et al., 1996). This reduction in stress variation is also found in rhythmic speech techniques, which suggests these speech patterns simplify speech production, leading to the assertion that reducing motor task demands assists a possibly unstable or overloaded speech production system (Packman et al., 1996).

Other theories point to a deficit in integration of segmental (linguistic) and paralinguistic (prosodic) information as a possible cause in stuttering, a general theme which will be significant for developing our eventual account of findings regarding differences between people who do and do not stutter in the central experiment of this dissertation. Before delving into these further, recall that prosody plays an important role in the production and perception of fluent speech (Bergmann, 1986; Besozzi & Adams, 1969; Hall et al., 1999; Packman et al., 2007; Packman et al., 1996; Wingate, 1966). In order to produce fluent speech, a speaker must integrate both the linguistic message consisting of stored lexical items, and the appropriate prosodic variation.

Kent (1984) highlighted the issue of integration of lexical, syntactic, and prosodic components in order to form an overall coherent linguistic representation. He proposed that successful integration of linguistic and prosodic information requires a central processor that combines the two temporal processing streams; a deficit in this central processor, he suggested, may be a cause of stuttering. Kent's (1984) original proposal became what is now known as the *neuropsycholinguistic theory of stuttering*; this theory identifies stuttering as a breakdown in the integration of segmental and prosodic information prior to speech initiation (Perkins et al., 1991). Other theoretical perspectives have likewise identified connections between stuttering and prosodic characteristics of speech. For instance, Arbisi-Kelm (2006) suggested that disfluency is triggered by semantic and syntactic factors in intonational structure, such that a theoretically-informed view of prosody in perception and production planning is viewed as allowing for optimal investigation into both word-level and sentence level-factors, as well as their interactions.

A variety of linguistic factors have been tested for their relevance to stuttering. In examining lexico-prosodic factors, Dayalu and colleagues (2002) found that content words are more often disfluent than function words. Additionally, Au-Yeung and colleagues (1998) found that function words are more disfluent when they are preceding rather than following content words. Sentence structural factors are also relevant; a number of studies have shown that longer utterances are more likely to be disfluent (Maner et al., 2000; Tornick & Bloodstein, 1976; Yaruss, 1999; but see Kleinow & Smith, 2000). Further, complex utterances are more likely to be disfluent (Kadi-Hanifi & Howell, 1992; Kleinow & Smith, 2000; Melnick & Conture, 2000; but see Maner et al., 2000). Further, the position of words within sentences or utterances influences the likelihood of stuttering. Studies have variably shown the first three words (Prins et al., 1991), the first two words (Au-Yeung et al., 1998; Koopmans et al., 1992), or the first word to be the most likely to

be disfluent (Brown, 1938; Tornick & Bloodstein, 1976). Strikingly, however, Kaasin and Bjerkan (1982) found the final word to be the most disfluent. These studies collectively highlight the role of theoretic constructs related to prosodic phrase structure and grouping –structural elements common to both language and music – in explaining variability in stuttering (Ladd, 2008; Patel, 2010; Pierrehumbert, 1980).

Multiple phonological factors have also been highlighted in relation to variability in stuttering. For instance, word position is a significant factor in stuttering. Word-initial sounds are found to be the most likely to be disfluent and/or most severely stuttered (Brown, 1938, 1945; Hahn, 1942; Hubbard, 1998; Natke et al., 2002; Soderberg, 1962; Taylor, 1966; Weiner, 1984). At the segmental level, consonants have been found to be more often disfluent than vowels (Brown, 1938, 1945; Hahn, 1942; Taylor, 1966; but see Soderberg, 1962). Lexico-prosodic factors are also significant; for instance, words that are produced with a phrase-level prominence on the lexically-stressed syllable have been found to be more likely to show disfluency and/or to show more severe stuttering behaviors, compared with words without a phrase-level prominence (Bergmann, 1986; Brown, 1938; Natke et al., 2002; Prins et al., 1991; Weiner, 1984; Wingate, 2012; but see Hubbard, 1998).

Utterance-level rhythm is another linguistic factor that is relevant to variable stuttering behaviors. Given that a hallmark of stuttering is a dysrhythmic flow of syllables in speech motor production, it has also been proposed that stuttering may be associated with underlying deficits in processing rhythm and timing used in speech (Alm, 2004). People tend to stutter more often on words that bear lexical or phrasal stress than on words that do not (Arbisi-Kelm, 2006; Bergmann, 1986; Prins et al., 1991; Wingate, 1984). Relatedly, previous researchers have viewed stuttering as essentially a *timing disorder* which manifests either in the speech motor control system (Kent,

1984) or during integration of segmental and prosodic components of speech prior to initiation of speech production (Perkins et al., 1991). People who stutter show not only deficits in speech fluency, but have also demonstrated differences in speech rate, pitch, loudness, and articulation (Bloodstein & Bernstein Ratner, 2008). Even during fluent speech, researchers have noted people who stutter show centralization of formant frequencies in vowel production (Klich & May, 1982) and slower overall speech rate (Bloodstein, 1944). Some researchers argue that stuttering is not solely a deficit in speech motor execution, but is a linguistically-conditioned disorder reflecting a failure of the sentence production process (Bernstein Ratner, 1997).

A number of studies have indicated prosodic elements to be tied to instances of stuttering. For instance, studies have shown that speech production of people who stutter can be influenced by the perception of prosodic elements of their own speech (Hargrave et al., 1994; Wingate, 1966). Previous research has reported group differences in prosodic characteristics (e.g., intonational patterns) during fluent and disfluent speech of adults who stutter (Arbisi-Kelm, 2010). Further research pointing to abnormal planning of prosody in adults who stutter was shown by Wendahl and Cole (1961). Wendahl and Cole (1961) modified recordings of adults who do and do not stutter to remove disfluencies and then asked participants to evaluate the speech on measures such as rate (i.e., normal tempo) and rhythm. Their results demonstrated that even during fluent productions, adults who stutter had a less typical rate of speech and used less rhythmical speech patterns than adults who do not stutter. Likewise, DiSimoni (1974) used a task in which adults who stutter were asked to repeat bisyllabic pseudowords consisting of a vowel-consonant-vowel sequence differing in the final vowel (e.g., /asa/, /asi/). DiSimoni (1974) showed that the timing of fluent productions of people who stutter differed from those of controls; relative to adults who do not stutter, adults who stutter produced longer and more variable durations for both the consonants and vowels.

Further, previous findings showing that external pacing helps to improve fluency supports a possible deficit in the rhythmic component of prosody for people who stutter (Glover et al., 1996; Ingham & Carroll, 1977; Wingate, 2002). It has been suggested that since fluency is often induced with rhythmically predictable meter, prosodic breakdowns in people who stutter might be fundamentally timing-based (Arbisi-Kelm, 2006).

The above studies have focused on adults who stutter. Few studies to date have examined prosodic characteristics of fluent speech in children who stutter, yet deficits in planning and/or executing prosodic structures have been proposed to be a proximal cause of stuttering (e.g., Karniol, 1995; Packman et al., 2007). Studies examining speech rate and duration have reported little difference between children who stutter and typically-developing children (Hall et al., 1999; Healey & Adams, 1981; Kelly & Conture, 1992; Ryan, 1992, 2000). Additionally, researchers have examined fluent vs. disfluent utterances and found no difference between these types of utterances (Chon et al., 2012; Logan & Conture, 1995). A recent study by Wieland, Dilley, Burnham, and Chang (in preparation) investigated prosodic components of speech produced by children who do and do not stutter using a well-validated system for linguistic coding of prosodic prominences and phrasal boundaries. No significant differences were found for any fundamental frequency-related, prominence-related, or phrasal boundary-related measures. Similarly, most measures were not significantly different between the fluent and disfluent phrases for the children who stutter except for the utterance duration. However, the results showed that utterance duration was longer and contained more syllables for phrases with stuttering-like disfluencies than those without.

In summary, proposals regarding the causes of stuttering differ considerably in relative emphases of the various hypothesized causal factors that are involved in chronic and/or variable

disfluency. Together with the research reviewed in Section 1.8.3, the findings reviewed support the contention that stuttering involves, on the one hand, an imbalance of sorts between feedforward production and feedback corrective or monitoring processes, and on the other hand, difficulties in integration or prediction of prosodic information with respect to segmental information during linguistic structure generation. Together, these convergent issues in the stuttering literature provide crucial threads motivating the focal area of investigation in this dissertation, namely neural processes of rhythm perception. In the next section I continue development of the motivation and background for the proposed research by reviewing structural and functional neuroimaging studies that have provided convergent evidence for atypical connectivity between speech motor and auditory regions in people who stutter.

1.8.5 Neuroscience of language in people who stutter

Neural auditory and motor regions have been a focus of a considerable number of neuroimaging studies aimed at investigating abnormal brain activation, and possible neural deficits, in adults who stutter (Chang et al., 2011; Chang et al., 2009; Fox et al., 1996; Watkins et al., 2008; Wymbs et al., 2013; Yang et al., 2016). Part of the considerable interest by the neuroscience community in these areas has come from convergent evidence that has shown that during speech tasks, adults who stutter have increased activity in the right premotor areas, as well as abnormally reduced activity in auditory association areas (Belyk et al., 2015; Braun et al., 1997; Brown et al., 2005; Budde et al., 2014; Chang et al., 2009; Fox et al., 2000). In this section I highlight some core research findings reflective of neural areas which studies have supported as showing abnormal activity in auditory and/or motor regions. In light of proposals regarding the

importance of neural auditory/motor interactions in the brain for rhythmic predictive processes (e.g., Patel & Iversen, 2014), the current section thus further serves the purpose of continued development of motivating the rhythm perception experiment with people who stutter at the core of this dissertation.

Multiple researchers have found subtle functional differences in regions supporting auditory-motor integration (Braun et al., 1997; Chang et al., 2009; Chang & Zhu, 2013; Fox et al., 1996; Watkins et al., 2008). More specifically, it was found that adults who stutter have reduced activity between motor and auditory areas (Braun et al., 1997; Fox et al., 1996) or else reduced activity between the left hemisphere motor and auditory areas (Neef et al., 2016; Watkins et al., 2008). Other studies have reported functional differences in brain regions supporting auditory-motor integration in children who stutter compared to controls (Chang et al., 2018; Chang et al., 2016; Chang & Zhu, 2013); areas showing different patterns of activation in children who stutter include left premotor areas, motor areas, and auditory cortical areas. Further, compared with children who do not stutter, children who stutter have been shown to have decreased white matter integrity along the dorsal language pathway in the left hemisphere, which is paramount for language formulation and speech production (Chang et al., 2019).

A finding that often accompanies studies showing neural evidence of less efficient and/or disrupted auditory-motor integration processes in people who stutter is the finding of apparent reliance on right-lateralized structures by people who stutter in possible compensation. To wit, multiple studies have shown not only the pattern of relatively reduced activation in people who stutter in left hemisphere structures, compared to controls, but also comparative *over*activation in *right* hemisphere structures in people who stutter (Brown et al., 2005; Budde et al., 2014; Fox et al., 1996; Preibisch et al., 2003); implicated right hemisphere structures include the primary motor

cortex, premotor cortex, pre-SMA, the inferior frontal gyrus (IFG), the insula, and the frontal and Rolandic operculum (Neef et al., 2016). Proposals that observed relative overactivation in right hemisphere structures in people who stutter may reflect compensatory mechanisms that act to offset the relative left-lateralization of activation observed in controls is supported by research with children who do not show the same overactivation which might develop as stuttering persists into adulthood (Chang et al., 2019).

Evidence also exists of adults who stutter having anomalous grey and white matter volume differences in speech planning and auditory-motor integration structures, as compared to adults who do not stutter (Beal et al., 2015; Chang et al., 2009; Cykowski et al., 2010; Cykowski et al., 2008; Fox et al., 2000; Kell et al., 2009; Sommer et al., 2002). In children who stutter compared to those that do not, decreased grey matter volume in critical speech areas such as the left IFG, motor, and pre-motor cortices have been found (Beal et al., 2013; Chang et al., 2008; Chang & Zhu, 2013). Of note, the nature of differences in gray matter between adults who do and do not stutter have varied across reports, with some studies reporting larger values (Beal et al., 2007; Lu et al., 2010; Song et al., 2007), others reporting smaller values (Kell et al., 2009; Lu et al., 2010) and still others reporting no differences (Jäncke et al., 2004) for gray matter volume for adults who stutter compared with those who do not. Collectively, these findings from neuroimaging studies converge in suggesting a well-established finding that auditory/motor integration areas show abnormal, reduced functionality in people who stutter. A further finding is that people who stutter instead appear to compensate by relying on structures in the right hemisphere differentially more than control participants under comparable conditions. In the following section, I initiate connecting these apparent deficiencies in auditory/motor integration in people who stutter and other topics to two critical constructs at the heart of my dissertation experiment: *meter* and *beat*.

1.9 Production and perception of non-linguistic auditory sequences

Arguments and evidence thus far support the contention that shared neural and computational resources are likely to be engaged in certain kinds of perception-action activities, whether the action involves communication (e.g., initiating speaking) or not (e.g., initiating a hand gesture for emphasis or a drum beat). That is, certain kinds of preparatory and/or motor activities require an individual to listen to auditory sequences and internally synchronize with or otherwise "time" them before producing a motor action (Jones, 2018; McAuley & Jones, 2003). Multiple lines of evidence support the relevance of perception and action preparation within rhythmic contexts as having relevance for both speech communicative and non-speech activities (Jantzen et al., 2016; Ladányi et al., 2020; Patel & Iversen, 2014; Ravignani et al., 2019).

Successful speech perception and production requires the ability to generate a periodic timing signal (i.e., an internal beat) (Senft et al., 2016; Spencer & Rogers, 2005; Whitfield et al., 2018). The basal ganglia thalamocortical network is engaged during rhythm perception and when temporal structure provides rhythmic expectations for efficient prediction of upcoming events, whether they are speech or non-speech (Dilley & Pitt, 2010; Kotz & Schmidt-Kassow, 2015; Kotz & Schwartze, 2010; Morrill, Dilley, & McAuley, 2014). This suggests that connectivity between the basal ganglia thalamocortical network provides a structure for the internal timing of sound sequences which are necessary for both speech perception and speech production (Chang et al., 2016).

The above arguments are consistent with the view that deficits in initiating and/or sustaining natural, fluent timing and rhythm during feedforward production and feedback speech control may underlie or causally contribute to stuttering as a disorder. Such an observation invokes

connections to the notion of *sustained rhythms*, as suggested colloquially by phrases like *The rhythm of her speech…* or *This music's got rhythm!* (or *He's got rhythm!*). Rhythms – auditory patterns that involve alternations of metrically prominent and non-prominent elements – involve structures among prominences, which are widely observed to be similarly hierarchically structured in both language and music (Ladd, 2008; Patel, 2010). Rhythms occur for both speech and non-speech (e.g., music, tones, or sometimes even unintelligible speech signals) (Jones, 2018). An important construct involves a situation when an auditory stimulus gives rise to the perception that the metrically prominent elements occur at quasi-predictable points in time, which I will refer to as a *beat*. In the following, I review evidence supporting a core idea in this dissertation, namely that beats – whether in auditory stimuli ostensibly heard as intelligible speech or as non-speech – are *temporal points of integration* between feedforward auditory perceptual signals and an ongoing, sensorimotor action plan.

1.9.1 Auditory-motor connections in rhythm

Among theoretic frameworks for perception, action, and cognition that consider both feedforward planning and feedback control, only a handful of approaches consider how the notion of a "beat" may be relevant for *both* speech and nonspeech motor planning and integration (e.g., Kotz et al., 2018; Patel, 2010; Zoefel et al., 2018). Instead, the vast majority of work on rhythm considers issues specific to *either* speech *or* music (e.g., Kotz et al., 2018). This seems surprising, since the notion of a "beat" is relevant for both non-speech perception (e.g., music), but also for speech, as revealed by the well-known "p-center" phenomenon (Hoequist, 1983; Marcus, 1981; Morton et al., 1976; Patel, 2010). Here, I build on cross-disciplinary work on the notion of a *beat*

to elaborate on how this construct is not only relevant to both speech and nonspeech, but also how it may be a critical concept for understanding the relationship between feedforward planning and feedback integration, including in stuttering. The next sections therefore review the putative role of beats in auditory-motor integration in rhythm across domains.

Across both music and speech, the notion of "beat" is often considered an "affordance", a kind of psychological construct or property which affords the ability to temporally synchronize one's actions to that external stimulus with respect to an unfolding signal – one that is *usually* – but not always – auditory (Kotz et al., 2018; Miller et al., 2013; Pasinski et al., 2016). Before proceeding, it is worth addressing one of the presumptively core challenges with cross-disciplinary studies of rhythm that appear to stem from to disciplinary assumptions about the role acoustic isochronicity across domains. On the one hand, scholars whose primary disciplinary affiliation lies with linguistics or speech communication often appear to start from disciplinary knowledge that speech does not consist of acoustically isochronous intervals (Patel, 2010; Shen & G, 1962), even though it may sound as if it does (Lehiste, 1977; White, 2002). Certainly, an account that points to notions of beat, rhythm, and metrical predictability as crucial for speech and language must hold insight for sequences that are neither acoustically *nor perceptually* regular or isochronous in any way, nor must they be predicated on notions that utterances are necessarily comprised of either perceptually or acoustically isochronous intervals (Brown et al., in press; Dilley & Pitt, 2010; Ding et al., 2017). On the other hand, scholars whose primary disciplinary affiliation lies with music often deal in models positing abstract metrical structures that have isochronous underlying intervals. However, empirical musicologists are well-familiar with core disciplinary findings that performed musical intervals are actually far from isochronous (Repp, 1998).

Having considered these preliminaries, I turn to reviewing research that links the movement "affordances" of metrical beats across domains with movement, action preparation, and auditory-motor integration. Notions of auditory-motor interactions relative to rhythm and beats have been carefully and thoughtfully developed within the work of Patel and Iversen (2014). Patel and Iversen's Auditory Simulation for Auditory Prediction (ASAP) posits the core assertion that beats are important *moments of integration* relative to feedforward planning and feedback control. The ASAP hypothesis specifically was developed relative to neural structure interactions for rhythm in the brain; according to the ASAP hypothesis, perception involves temporally precise two-way communication between auditory regions and motor planning regions. Under this hypothesis, connections between motor and auditory signals play a causal role in beat perception and movement "affordance", specifically by supporting temporal predictions for upcoming beats which might be physically acted upon at some later point in time.

Evidence in support of the ASAP hypothesis comes notably from studies that involve strictly perceptual imagery of a musical beat; notably, such visual imagery of musical beats reliably engages neural motor systems, even in absence of any motor action by participants. Patel and Iversen (2014) explained these beat imagery findings by proposing that motor planning uses simulation of body movement to entrain neural activity patterns to a beat period, such that communication proceeds from motor planning to auditory regions which function as predictive signals for timing upcoming beats. Flexible action-preparation is possible through appeal to beats, because predictive rhythm perception adjusts to tempo to account for changes in overall rate (McAuley et al., 2006; Patel & Iversen, 2014; Ross et al., 2018). Further, the motor system is able to make hierarchical predictions about timing at different timescales (Patel & Iversen, 2014). Although the ASAP hypothesis focuses ostensibly on rhythm *perception*, this proposal is also

applicable to rhythm *production* since the act of motor planning causes the auditory system to be coupled to make predictions about the timing of auditory events. Further, the act of synchronization of one's own motor actions with a temporal prediction involves the production of an imagined beat (Patel & Iversen, 2014). The timescale of musical beats (~100 msec) is also a similar timescale for syllable production (Ding et al., 2017).

The notion of beats reflecting moments of integration between feedforward planning and feedback control for both speech and nonspeech is one shared by Kotz and Schwartze (2016). Kotz and Schwartze (2016) elaborated on how areas involved in language production interface with areas involved in timing processing, including in non-speech. The model of Kotz, Schwartze & colleagues (2016; 2011; Schwartze et al., 2012) emphasizes the need to view temporal structure as an important and valuable source of information for perception and planning of action, rather than a more-or-less trivial by-product of "processing in time." For this reason, they argue that any model is incomplete unless it provides an explanation for how the temporal structure of speech is generated.

Kotz and Schwartze (2016) further suggest multiple ways in which this cross-domain notion of a "beat" can be productively linked with related work. This includes linking the moment of occurrence of a syllable (i.e., the notion of a p-center; Morton et al., 1976) to the notion of a "theta-syllable" which associates syllabic timing with neural oscillations (Ghitza, 2013, 2014; Luo & Poeppel, 2007; Peelle & Davis, 2012; Pefkou et al., 2017), as well as an acoustic, spectrotemporal profile (Greenberg et al., 2003), and motor-articulatory movements as per the frame-content theory of speech, which dissociates syllable and segmental elements (MacNeilage, 1998). Relevant to the p-center phenomenon, rhyme onsets of syllables readily map to perceptual "beats" in speech which likewise afford temporal synchronization via tapping or adjustment of the

phase of distinct auditory streams (Morton et al., 1976), similar to music. For both speech and music, the acoustic correlate which most closely corresponds to a perceptual moment of occurrence of a beat is the amplitude envelope rise (Marcus, 1981). The importance of acoustic indices of beats associated with p-centers in speech is further supported by recent work by Oganian and Chang (2019), who showed that a well-defined zone in middle STG detects acoustic onset edges by responding to amplitude envelope rises, specifically the local maximum in envelope rate of change in both tone and speech stimuli. Local maxima in the speech amplitude envelope rate of change, correspond phonologically to rhyme onsets (Goldsmith, 2011). These arguments collectively support assertions that beats are temporal points of integration between top-down, feedforward auditory plans and ongoing, bottom-up sensorimotor feedback in speech as well as music, enabling planning adjustments in response to auditory signals of a variety of types. The relevance of a domain-general notion of "beats" to an account of findings from the core dissertation experiment, as well as to core deficits in stuttering will be developed gradually throughout this thesis.

1.9.2 Research with people who stutter

This section continues the development of the core argument in this dissertation, which links rhythm perception, beat movement affordances, auditory-motor integration, action preparation, and feedforward/feedback integration difficulty in stuttering as a speech disorder. Here, I review some additional findings relative to deficits and core features of stuttering, including findings that have used a wide variety of action-preparation and -execution tasks to study timing.

I further elaborate on patterns of neural activation and related findings that support a hypothesis of possible rhythm perception deficits in people who stutter.

It is widely held that stuttering involves difficulty with enacting a speech motor production plan, which is variably attributed to either a general motor deficit (Max et al., 2003; Zelaznik et al., 1997), a deficit in the coordination of the speech system (Hulstijn et al., 1992), and/or a generalized timing deficit (Boutsen et al., 2000). Impairment across *both* speech and non-speech motor control systems is widely held to be a critical component of the disorder of stuttering (Olander et al., 2010). Indeed, people who stutter may differ from those who do not primarily in their differential capacities to generate temporal structures of action (Kent, 1984). In general, these arguments are consistent with the view that timing deficits may be at the heart of speech motor difficulty in the disorder of stuttering. The goal of this section is thus to review findings related to timing abilities of people who stutter across a variety of speech and non-speech tasks, toward further developing the motivation and interpretation of the original experiment at the core of this dissertation.

A common approach in investigating putative timing deficits in people who stutter is to split tasks into those that require internal timing (i.e., self-generation of a timing structure) vs. external timing (i.e., synchronizing one's own motor responses to a provided external pacing cue). The internal timing network comprised of basal ganglia and SMA (Alm, 2004; Coull et al., 2013) is expected to be active during tasks performed in the *absence* of external timing cues. By contrast, an external timing network comprised of cerebellum and premotor cortex (Alm, 2004) is expected to be activated during tasks performed in the *presence* of external timing cues. It has been argued that stuttering entails dysfunction within the brain network supporting internal timing, such that

the external timing network may be used as a mechanism to compensate for a deficient internal timing system in people who stutter (Braun et al., 1997; Etchell et al., 2014).

Studies with adults who stutter have found evidence supporting dysfunctional processing within core structures of the internal timing network (i.e., basal ganglia and SMA). For instance, Chang and colleagues (2009) showed altered patterns of activity in the SMA relating to the perception and planning of speech for people who stutter compared to people who do not stutter. Similarly, Ingham and colleagues (2012) investigated neural activation in people who stutter during a speech task and a rest condition; they found differences in both the internal and external timing networks of people who stutter compared to people who do not stutter. Further, in Ingham and colleagues' (2012) study, people who stutter had significantly more activation in the basal ganglia during the rest condition, but significantly less activation in basal ganglia during the speech task, consistent with dysfunction in the internal timing network of people who stutter. This finding additionally showed that the basal ganglia may be over- or under-active compared with controls, depending on the task at hand. In other words, stuttering is better viewed not just a disorder of speech *per se*; rather, the dysfunction in stuttering also extends to other non-speech-specific domains of perception and action that require precise timing.

Other studies have specifically shed light on differential activation in a component of the external timing network – namely, the cerebellum – in people who stutter. For instance, Braun and colleagues (1997) reported heightened activity in the cerebellum in people who stutter compared to people who do not stutter during both fluent and stuttered speech. They proposed that this heightened activation in the cerebellum reflected a mechanism of compensation for a disordered internal timing network. Additionally, Ingham and colleagues (2012) found increased cerebellar activity in people who stutter compared with those who did not; this increased cerebellar activity

was negatively associated with stuttering during oral reading and monologues, reflecting a pattern of greater cerebellar overactivation in people who stutter compared with controls under conditions when stuttering occurred less. A pattern of increased cerebellar activation in people who stutter, compared with people who do not, has also been reported in other studies (e.g., De Nil et al., 2008; Watkins et al., 2008).

Tasks that assess individuals' abilities to maintain a steady internal rhythm for movements, such as tapping at one's own preferred pace (i.e., self-paced tapping), have been proposed to be useful for elucidating putative internal timing deficits in stuttering. Such tasks have revealed reduced abilities to maintain accurate timing in people who stutter compared with those who do not. For instance, Blackburn (1931) demonstrated a marked inferiority for people who stutter, compared to people who do not stutter, in abilities to execute rhythmical voluntary movements in speech-related tasks (i.e., move the tip of the tongue, or move the lower jaw), but not in abilities to execute voluntary rhythmical movements in non-speech-related tasks (i.e., tap steadily). Further, Cooper and Alan (1977) asked participants who did or did not stutter to speak passages steadily or else to tap steadily. They found that the group who stuttered was less accurate in controlling the timing of their rates of finger tapping, as well as their speech tasks, compared to the group that did not stutter. Cooper and Alan (1977) suggested this may be due to a less accurate "neural clock" in the group who stuttered compared to the group that did not. Further, Brown and colleagues (1990) investigated whether participants who stuttered differed from those who do not on several self-paced rhythmic tasks: finger tapping, jaw opening and closing, and jaw movement during the production of "ah". For all tasks, participants in both groups performed the tasks at their most comfortable rate, as well as at slightly faster and slower speeds. The results demonstrated that the variability decreased for people who stutter decreased for the self-paced speech task, as well as an

orofacial non-speech task and a finger tapping tasks. The authors interpreted their findings as indicating that people who stutter have less flexible timing systems which may be more susceptible to breakdown than people who do not stutter.

Tasks that investigate the use of externally-paced paradigms with adults who stutter have generally revealed a lack of group differences in mean accuracy or variability. Hulstijn and colleagues (1992) found decreased variability between participants who stutter compared to participants who do not stutter. The participants were asked to synchronize and continue tapping with the index finger of their dominant hand, their non-dominant hand, and with both hands; they further were asked to speak a voiced sound in a repetitive fashion or to tap synchronously with a syllable they spoke. No significant difference between groups was found during the speech and non-speech tasks; however, during the synchronization portion of the task, where speech and hand movements had to be synchronized with tones, participants who stuttered had a significantly larger standard deviations than participants who did not stutter during the most difficult condition. The authors suggested that the lack of observed group differences might have been due to the task being overall too simple, an assertion which is also echoed by other researchers that suggested task complexity is paramount to finding group differences (Boutsen et al., 2000; Max & Yudman, 2003; Zelaznik et al., 1994).

Tasks of synchronization with, and continuation of synchronized movements with a pacing stimulus (i.e., synchronize-continue tasks) are used to investigate how well a participant can tune into an external pacing stimulus; this involves is a very different process than internal rhythm generation. Boutsen and colleagues (2000) used a synchronize-continue task to investigate timing and intensity variability between adults who do and do not stutter during a task involving synchronization between a self-generated movement and an external pacing

signal. Participants were asked to produce syllables synchronized with an isochronous metronome, which were then used to calculate the durational variation between successive syllable onsets and the intensity variation of the beginning consonant and vowel in successive syllables. The results showed that although the intensity variation at the beginning consonant and vowel in successive syllables were similar across groups, the stuttering group displayed increased variability in the timing of successive syllables when compared to the non-stuttering group.

A synchronize-continue task was also used by Zelaznik and colleagues (1994), who compared external pacing by people who do and do not stutter by asking participants to produce a rhythmic flexion-extension movement ("tapping-like" motion side to side, rather than up and down) at the pace of the metronome and then to continue the imposed rhythm until the metronome was engaged again. A second similar task was to apply pressure to a force transducer using the right index finger. Trials were either paced or involved increasing/decreasing duration where the cycle duration increased or decreased, respectively. This work was based on Hulstijn and colleagues (1992) but expanding the number of intervals from 400 ms to 200-600 ms and adding the increase/decrease trials, as well as the force trials. Consistent with other work, Zelaznik and colleagues (1994) found a lack of significant group differences between people who do and do not stutter. Like Hulstijn et al. before them who had similarly found null results, Zelaznik et al. offered similar accounts positing that the task still might have been too simple to uncover what are possibly very subtle underlying group differences.

A compelling question is whether variability in timing observed for adults who stutter is also observed at younger ages(e.g., in children). Howell and colleagues (1997) investigated children who stutter in therapy (aged nine to ten) and children who do not stutter during an external

synchronization and continuation of non-speech lip movements of the upper lip, lower lip and jaw of the participant. Similar to Hulstijn et al. (1992) and Zelaznik et al. (1994), Howell and colleagues found no evidence of a difference between groups, suggesting no deficit in the timing mechanism of children who stutter compared to those that do not. However, the authors did find a consistent trend where the clock variance was higher relating to mouth movements for stuttering than non-stuttering participants, which might suggest the lack of effect was actually due to a small sample size.

Further, Wieland, McAuley, Dilley, and Chang (2015) conducted a study in which they found evidence that children who stutter perform less accurately on a rhythm discrimination task identifying if a test rhythm was the same or different than the comparison rhythm. Two studies conducted with children who stutter found mouth movements had greater timing variability (Howell et al., 1997), although clapping motions had more variable inter-clap-intervals (Olander et al., 2010) than typically-developing children. However, a follow-up study by Hilger and colleagues (2016) attempting to replicate these findings with a larger sample notably also did not find group differences. Both these studies suggest that a fundamental deficit may exist for children who stutter to internally generate consistent rhythmic motor behaviors, compared to those that do not stutter. Overall, the above findings support that stuttering involves multiple difficulties with enacting a speech motor production plan, where this may entail any of a number of possible deficits at the juncture of auditory-motor integration, including a general motor deficit, a deficit in the coordination of the speech system, and/or a generalized timing deficit.

1.9.3 Additional neuroanatomical and behavioral observations motivating an fMRI study of rhythm discrimination in people who stutter

In the previous sections I reviewed findings relevant to motivating the questions and experiment at the core of this dissertation. Having now reviewed above large bodies of work in brief which are relevant to speech and language, rhythm perception and production across domains, neuroscience, and deficits and fluency-enhancing conditions in stuttering, I consider some final topics that help close the gap in motivating a study of rhythm perception ability in people who stutter. In particular, I consider the issue of how rhythm perception relates to physical properties of acoustic signals, as well as issues related to studying neural structures for rhythm perception, including in people who stutter.

An important consideration in the study of rhythm is the relationship between stimulus properties and perception of rhythm and meter. Rhythms can be perceived as having beats due to a variety of different kinds of stimulus properties. In general, metrically prominent events in non-speech rhythms can sound like accented beats due to being louder, longer, and/or having distinctive pitch (Hannon et al., 2004). Similar acoustic characteristics increase the probability of a syllable's sounding prominent in speech (Kochanski et al., 2005; Ladd, 2008; Lehiste, 1970).

Making auditory events equal in frequency, duration, and inter-onset-interval renders these acoustic dimensions effectively irrelevant to the determination of rhythm and meter. A further stimulus property often thought to endow auditory events with the property of being rhythmic beats concerns the temporal similarity or isochronicity of intervals between onsets of the auditory events (e.g., syllables, tones, or percussive events). That is, it is frequently assumed that a sequence of non-speech onsets which demarcate isochronous or *regular* intervals confers "beat" status to

those events or that onsets which demarcate non-isochronous or *irregular* intervals will render those events less likely to be heard as beats (e.g., Patel, 2010).

In fact, the meter – the metrical arrangement of onset events into a perceived hierarchical structure – represents a top-down or high-level *abstraction* over the sequence of acoustic onsets. The metrical structure involves computing something like a "best fit" among competing candidate metrical hierarchies; assessment of this fit may be subject to context effects and/or top-down knowledge, such as knowledge of musical genre (Povel & Essens, 1985; Vuust et al., 2018). The onsets of auditory events need not correspond – and often do not correspond – to moments heard as beats, as in the case of rhythmic syncopation in music (Vuust et al., 2018). Nevertheless, acoustic regularity of onsets does make it more likely that some coherent metrical structure will be perceived, even though the sequence of intervals between onsets underdetermines the metrical structure that might be heard (Grahn & Brett, 2007). For instance, in absence of contextual, harmonic, or other genre-related cues to the metrical structure (e.g., instrumentation), a sequence of acoustically regular onsets might be heard readily as having a "straight" meter, in which abstract metrical prominences are coincident with the acoustic onsets, or else as having "groove," in which abstract metrical prominences are offset with acoustic onsets, where the latter are instead heard as syncopated "off-beats" (Vuust et al., 2018). Finally, it is noted that any discussion of regularity or irregularity among the intervals between acoustic onsets and its relation to meter perception must be tempered by knowledge that performed music often deviates substantially from strict isochrony of the underlying intervals for the intended metrical structure (Repp, 1998).

Having considered issues of how stimulus properties affect metrical perception, it is also noteworthy to consider neural structures or connections that may afford the ability to *act on* a perceived meter (e.g., through a motoric response). In this regard, additional motivation for

studying rhythm perception in people who stutter comes from convergent work appearing in the last five years relating to individual differences in white matter integrity in brains of people who do and do not stutter, and how these differences may affect motor action execution. This is because multiple studies of people who stutter have shown abnormal white matter tract characteristics. For instance, an important study of the resting brains of children who stutter (Chang et al., 2016; Chang & Zhu, 2013) showed aberrant network architecture in the white matter tracts connecting the default node network and attention, somatomotor, and frontoparietal networks of children who stutter relative to controls. Convergent work with people who do *not* stutter has shown that individual morphological differences in white matter tracts correlate well with non-speech rhythmic predictive abilities (an individual's degree of sensorimotor error during a tapping task) (Blecher et al., 2016; Vaquero et al., 2018). Disruptions to the transmission of timing information across abnormal white matter tracts likely manifests as a phase offset of planned coordinative timing events for syllables (Ross et al., 2018).

The research on white matter integrity differences across people who do and do not stutter is consistent with the emergent hypothesis that individual differences in *white matter tracts* influence the ability to accurately predict and/or act on sensory information. The finding that abnormal white matter tracts are a core anatomical correlate of stuttering (Chang et al., 2018) particularly highlights the need for explaining the functional implications of these differences (Warbrick et al., 2017). The literature cited above can be interpreted to suggest that a failure to faithfully translate timing predictions accurately into motor actions might arise from deficient integrity of white matter tracts (cf. Blecher et al., 2016). This would result in poorer-quality, statistically unreliable predictions about timing, or else larger-than-normal error signals upon comparison of feedforward and feedback information, leading to error propagation at different

levels of the speech production system. Regions in the dorsal pathway have rich connections to the basal ganglia and putamen, which are important in beat perception.

The above studies highlight the need for studies of rhythm to disentangle perception from associated motor actions. A challenge to be solved for studying metrical perception relates to the fact that, in order to study beats, overt behavioral responses may sometimes need to be made by participants; these responses almost always have a motoric (i.e., production) component (Grahn & Brett, 2007). For example, studies of beat perception might require participants to track an auditory stimulus and respond with a tap at (quasi-) periodic intervals at the moments in time associated with when they predict the beat or next periodic event to occur. From this perspective, it is relevant that planning a motor response to a perceived beat or to an otherwise quasi-periodic stimulus involves both perceiving and producing these rhythms (Grahn & Brett, 2007; Lewis & Miall, 2003).

To study neural indices of beat perception without the problematic element of a confounded overt motor response, researchers have investigated neural activation of participants while they merely *listen* to a rhythm with or without a beat. Prior studies of beat perception in people who stutter, though, have required an overt motor task, such as tapping (e.g., Cooper & Allen, 1977; Hulstijn et al., 1992; Zelaznik et al., 1997). Such studies have implicated the motor system of the cortex (i.e., SMA, premotor cortex, basal ganglia, and cerebellum) by virtue of differential activation within these systems when listening to rhythms with or without a beat (Chen et al., 2008; Grahn & Brett, 2007). A related fact is that in people who do not stutter, the same set of brain areas (i.e., premotor cortex, SMA, cerebellum, and basal ganglia) have been shown in multiple studies to be active during timing, rhythm perception, and rhythm production tasks (Coull, 2004; Grahn & Brett, 2007; Rao et al., 2001).

To further highlight findings regarding neural areas involved in rhythm perception when movement is controlled for, consider that for people who did not stutter, more activity existed in the SMA and the basal ganglia, when listening to beat-based and regular rhythms than when listening to non-beat-based and irregular rhythms (Geiser et al., 2012; Grahn & Brett, 2007). Like prior work, this research supports that the activation of the basal ganglia are crucial for normal beat perception, which have been shown to be more active in continuation of hearing a beat than of prediction or finding a beat (Cameron & Grahn, 2014; Grahn & Rowe, 2009, 2013; Li et al., 2019). The above findings suggest that a more complex temporal representation may be required for beat perception compared to basic perception or control of timing of individual intervals; thus, it is important to recognize additional functions of the identified brain regions. Note that the basal ganglia and SMA are also involved in attention to time (Coull, 2004) when predicting internally generated movements (Coull et al., 2013; Cunnington et al., 2002; Freeman et al., 1993) and in temporal sequencing (Shima & Tanji, 2000). Communication and coordination of these spatially distributed brain regions would be required to accomplish timing-related goals, and neural oscillations have been proposed as a way to achieve this task (Fries, 2005).

In addition to the basal ganglia, several additional brain regions have been implicated in rhythm perception in prior work. For instance, Grahn and Brett (2007) specifically showed increased activity within the basal ganglia, pre-SMA / SMA, and anterior STG during perception of rhythms that induce a sense of periodicity/beat compared to ones that do not. Further, Grahn and McAuley (2009) conducted a study of neural activation while participants performed a temporal judgement task while listening to non-speech auditory stimuli with ambiguous beat structure, where participants were expected to show individual differences in the strength with which they perceived a beat. Grahn and McAuley (2009) showed that participants with strong

inclinations to perceive a beat had greater activation in the SMA, left primary motor cortex, insula, and IFG than those with weak inclination to perceive a beat. By contrast, the participants with weak beat perception had greater activation in the auditory cortex and the right primary motor cortex.

I now turn to consideration of neural areas involved in rhythm perception or processing that have been highlighted in prior studies as relevant particularly to stuttering. As a preliminary, the production of fluent speech requires coordination between the frontal cortex, which controls movement planning and execution, and the auditory sensory regions in the temporoparietal cortex (Chang, 2011). Relevant to neural areas supporting rhythm perception and processing, important areas concern brain networks for *internal timing* (i.e., self-generated or endogenous timing); these include basal ganglia and SMA (Etchell et al., 2014). Consideration of brain areas for rhythm also include those responding to *external timing* cues (i.e., perceptual or motor timing with respect to externally-generated or exogenous sensory stimuli) to sequence speech movements include cerebellum and premotor cortex (Alm, 2004).

Concerning neural areas relevant for rhythm perception/processing in stuttering, additional pertinent evidence comes from studies comparing children who stutter to those who do not. For instance, Chang and Zhu (2013) found that children who stutter (aged 3-9) have reduced levels of connectivity between brain regions responsible for self-paced movements compared to children who do not stutter. In particular, reduced connectivity was observed among the putamen (an area within the basal ganglia) and the SMA, superior temporal gyrus (STG), and cerebellum in children who stutter, compared with those who do not. Further, the researchers observed reduced connectivity among the SMA, the putamen, STG, and cerebellum in children who stutter, compared with those who do not. Likewise, Beal and colleagues (2013) conducted a study using

voxel-based morphometry for neural structures of children who stutter and a control group of children who do not stutter. Beal et al. (2013) found less grey matter in the right Rolandic operculum and STG in the group of children who stutter compared to those who do not stutter. These results broadly support that differences exist in neural structures supporting timing and integration of auditory feedforward information with motor action plan in both adults and in children who stutter, compared with control groups.

Further, Chang, Chow, Wieland, and McAuley (2016) examined a putative relationship between individual rhythm discrimination abilities and functional connectivity in structures constituting the neural rhythm network. They built on the work of Wieland and colleagues (2015), which compared rhythm discrimination abilities in children who stutter compared to those who do not and identified a deficit in the former. Chang etal. (2016) compared the strength of functional connectivity in the putative rhythm network between groups, and then examined how this activity was associated with behavioral performance on the same-different rhythm discrimination task. They found that children who stutter had weaker rhythm network connectivity and demonstrated a lack of correspondence between the rhythm network connectivity and rhythm discrimination, in contrast to a clearly correlated relationship among these variables for children who do not stutter. They further found that in children who do not stutter, the putamen, SMA, PMC and STG appear to show greater involvement when rhythm processing involves the internal generation of a beat (i.e., simple vs. complex rhythms). These results provided a novel and important extension of understanding functional neural connectivity and rhythm behaviors that extends beyond related findings of subtle timing-related deficits in people who stutter (Kent, 1984; Mackay & Macdonald, 1984; Van Riper, 1982). These results add convergent support other research that has suggested a core deficit in the interaction among cortical and subcortical regions necessary for both rhythm

and speech processing (Alm, 2004; Chang et al., 2016; Chang & Zhu, 2013; Etchell et al., 2014; Lu et al., 2010; Wieland et al., 2015; Wu et al., 1997).

1.10 Summary and Research Questions

1.10.1 Synopsis of evidence and considerations motivating a study of rhythm perception in people who stutter

The above sections have reviewed large numbers of cross-disciplinary studies in service of highlighting motivations for investigating rhythm perception in people who stutter. In the present section, I briefly review some of the main points important for developing my argument. I then present my research questions, hypotheses, and the outline of an extension of emergent theoretical frameworks in anticipation of attempting to account for my results.

This Introduction outlined evidence that the brain areas implicated in stuttering behaviors and in timing behaviors involve shared areas of neural processing. For instance, the syllable-sequencing circuit for speech outlined by the DIVA model were the basal ganglia, left ventral premotor cortex, and thalamus, while the rhythm processing network (i.e., the network of brain areas involved in processing rhythms) corresponded to basal ganglia, premotor cortex, SMA, and insula (Coull, 2004; Grahn & Brett, 2007; Rao et al., 2001). The reviewed literature supports the idea that people who stutter have less accurate cognitive mechanisms for timing than people who do not stutter (Allen, 1973; Blackburn, 1931; Brown et al., 1990; Cooper & Allen, 1977). This hypothesized deficient "neural clock" for people who stutter further is consistent with findings of deficient temporal components of speech and non-speech tasks. Convergent evidence of

dysfunctional processing within core structures proposed for the internal timing network in people who stutter was also cited from multiple structural and functional studies (see for review Etchell et al., 2017). As one example, Chang and colleagues (2009) showed altered patterns of activity in the SMA relating to the perception and planning of speech for people who stutter compared to people who do not stutter. Additionally, Chang and colleagues (2016) correlated performance on an adapted version of this paradigm with functional connectivity in the rhythm perception loop in children who stutter compared to those that do not. They found that unlike children who do not stutter, children who stutter have a core deficit in rhythm processing associated with the ability to perceive temporally structured sound sequences, which is an underlying skill necessary for speech perception and production. Further, Ingham and colleagues' (2012) investigation of neural activation in people who stutter during a speech task and a rest condition showed that people who stutter had significantly more activation in the basal ganglia during the rest condition, but significantly less activation during the speech task; this finding that the basal ganglia are over- or under-active depending on the task at hand suggests that stuttering is not just a speech disorder, but extends into other domains that require precise timing.

These findings collectively motivated several aspects of the study at the core of this dissertation. First, I drew inspiration from findings that brain regions involved in speech production overlap with those utilized during rhythm processing, as detailed in multiple ways from the empirical studies reviewed above (e.g., Alario et al.; Fujii & Wan, 2014; Kotz & Schwartze, 2010). Second, this work built on research showing that people who stutter have demonstrated abnormalities in brain regions involved in rhythm processing as evidenced by structural and functional neuroimaging studies, compared to those that do not (see for review Etchell et al., 2017). Third, numerous studies and lines of argument support speech-music connections, particularly as

relevant to metrical hierarchy building and beat / rhyme onset / p-center perception (Dilley & McAuley, 2008; Ghitza, 2014; Grahn & Brett, 2007; Luo & Poeppel, 2007; Morton et al., 1976; Oganian & Chang, 2019; Patel, 2010; Patel & Iversen, 2014; Tierney et al., 2018b).

Grahn and colleagues' research with the simple-complex paradigm was selected as the basis for conducting the experiment, as this work has focused on the role of temporal intervals between onsets as the stimulus property that differentially affects meter and beat perception. Notably, Grahn and Brett's (2007) paradigm built on a specific, controlled property of temporal sequences, namely that the same set of intervals can be perceived as variably strong, weak, or even missing, depending on the specific sequencing and order of those intervals (Grahn & Brett, 2007; Povel & Essens, 1985; Vuust et al., 2018). As observed by Grahn and colleagues (Grahn & Brett, 2007; Grahn & McAuley, 2009), for a given set of intervals and fixed number of temporal events, different orderings of those intervals and auditory events could give rise to different degrees of salience of a metrical structure and associated strengths of implied beats. On the one hand, an ordering of intervals which produces a strong impression of a beat will be referred to as a *simple rhythm*; on the other hand, an ordering of intervals which fails to produce a strong impression of a beat will be referred to as a *complex rhythm*. I utilized a subset of the stimuli from Grahn and Brett (2007) in my experiment and adopted their operational distinction between simple and complex rhythms.

This led to the following research questions addressed in this dissertation:

1.10.2 Research Questions

Research Question 1: Do adults who stutter show a deficit in rhythm discrimination performance compared to adults who do not stutter in a task of rhythm perception?

Hypothesis 1: It was hypothesized that adults who stutter will have less accurate rhythm discrimination performance than adults who do not, particularly for complex rhythms (i.e., those not expected to readily induce the perception of a periodic beat). This hypothesis was based on findings that people who stutter have shown a deficit in rhythm production, and that children who stutter have a deficit in rhythm discrimination (Wieland et al., 2015). If adults who stutter are found to have less accurate rhythm discrimination performance than those that do not, the findings will support the idea of a core deficit in rhythm discrimination in people who stutter from onset to adulthood. Alternatively, if no deficit in rhythm discrimination is found in adults who stutter compared to those that do not, this would suggest that only children have a behavioral deficit in rhythmic performance (cf. Chang et al., 2016; Wieland et al., 2015). The latter outcome would lend support for the idea that adults who stutter may have developed compensatory mechanisms over time following years of stuttering which enable them to show levels of rhythm discrimination performance similar to people who do not stutter.

Research Question 2: Do adults who stutter show aberrant brain activation in the timing network during a rhythm discrimination task compared to adults who do not stutter?

Hypothesis 2: It was hypothesized that during the rhythm discrimination task adults who stutter will have: (a) less neural activation in the internal rhythm processing network (i.e., basal ganglia, SMA), and (b) possibly greater neural activation in compensatory regions (i.e., cerebellum, premotor cortex), than adults who do not stutter. This hypothesis is based on findings that people who stutter have shown aberrant neural activation compared to those that do not in brain regions involved in rhythm processing during structural and functional neuroimaging studies (Etchell et al., 2017). On the one hand, if *less* neural activation in the internal timing network is found during rhythm discrimination in adults who stutter compared with those who do not, it will most likely be due to adults who stutter not sufficiently accessing this neural network or using compensatory regions. On the other hand, if *more* neural activation in the internal timing network is found during rhythm discrimination in adults who stutter compared with those who do not, it may be due to adults who stutter being less adept at utilizing this internal timing network, such that they may have to utilize compensatory regions in the external timing network in order to perform well on the task.

1.10.3 Theoretical Framework

Should Hypothesis 1 and Hypothesis 2 be confirmed, I postulate that emerging predictive coding theoretical frameworks potentially can accommodate both of these findings (Brown et al., in press; Clark, 2013; Daube et al., 2019; Friston, 2010; Pickering & Garrod, 2013). To sketch here the outline of such an account, I begin by noting that within predictive coding accounts, top-down hierarchical representations comprising candidate metrical structures are proposed to compete to best match the transmitted bottom-up sensorimotor cues; such cues come from stimuli

that may be heard as comprised linguistic messages (e.g., noise-vocoded speech), or else they may derive from non-linguistic sources (e.g., musical tones) (Di Liberto et al., 2018; Donhauser & Baillet, 2020; Kaufeld, Bosker, et al., 2020; Meyer et al., 2020a, 2020b; Vuust et al., 2018).

Recently, Brown, Tanenhaus, and Dilley (in press) proposed a Syllable Inference account of reconciling prosodic / metrical induction and spoken word recognition in language. They hypothesized that a core process in predictive coding across both language and music involves positing top-down structures that provide candidate "good fits" in time to knowledge-based generative structures, including metrical structures. In language, the phrase-level metrical structures are derived from a generative repository of individual candidate lexical items, which are part of multiple hierarchical structures, including of embedded phonology (cf. syllables and phonemes), as well as ongoing phrasal-sentential structures implemented according to the linguistic rules of a speaker's language and dialect. Crucially, the Syllable Inference account posits that "good matches" between the imputed metrical structures in the relevant domain involve internal simulations of accurate temporal alignment between moments in time heard as "beats", whether in music or speech (in press). In both domains, "beats" tend to correspond to moments of amplitude increase which are highly salient in both domains and generate strong neural responses (Kato et al., 2003; Luo & Poeppel, 2007; Oganian & Chang, 2019; Peelle & Davis, 2012), but in language these "beats" usually correspond to the moment of a consonant-to-vowel (i.e., C-to-V) transition, i.e., the rhyme onset of a syllable, or its "p-center" (Morton et al., 1976; Oganian & Chang, 2019; Zoefel et al., 2018). In music, non-speech tones or percussive sequences, the top-down generative process requires computing the temporal alignment and goodness-of-fit of a candidate metrical hierarchical structure. This structure is comprised of events of differential prominence with respect to (sensory antecedents of) acoustic onsets in the signal, under conditions

of contextual knowledge (e.g., musical genre). In language, the top-down generative process requires cognitive formulation of semantically and phonologically plausible hierarchical linguistic structures. This structure is comprised of words that must be imputed to provide a good match to the specific acoustic properties, most especially with respect to the implied temporal alignment of their rhyme / beat / p-center onsets to the amplitude envelope rises in the speech signal.

Based on the considerable evidence of various deficits and difficulties in stuttering outlined above, I hypothesize that the neural networks propagating bottom-up sensorimotor feedback introduce temporal errors as a function of the structure-specific integration demands, such as task-specific neural computations of lateralized prosodic variation (Assaneo et al., 2019; Chien et al., 2020; van der Burght et al., 2019). In cases where temporal errors are introduced into bottom-up sensorimotor feedback, this is proposed to lead to disruption of the normal integration functions/processes by various competing top-down representations of hierarchical metrical structures with bottom-up sensorimotor stimuli. This account of stuttering essentially amounts to a specific elaboration of the widely-held theoretical view that the disorder involves difficulties with the integration of feedback information with feedforward motor or linguistic plans (e.g., Bohland et al., 2010; Civier et al., 2013), updated within a predictive coding framework.

Such a problem with integration of top-down representations and bottom-up sensorimotor feedback would be expected to have several negative potential consequences. First, it might be expected that errored propagation of bottom-up information about timing and/or sensorimotor dynamical states would lead to system-internal quantitative indices of such timing and/or states that were gradiently incorrect with respect to objective state of external sensorimotor indices. Further, the error would be expected to introduce inaccuracies into the quantitative neural evaluation of gradient goodness-of-fit by high-ranked candidate top-down representations that

“seek” to minimize the error with respect to the bottom-up cues in order to become the “winning” representation. One potential consequence of such a situation would be an expected inability to “resolve” the error by virtue of attempting a match by a supposedly good top-down representation; instead of the expected local error minimum by the “best” top-down representation (Brown et al., in press; Donhauser & Baillet, 2020), a residual error would be in its place, but one that was actually itself gradiently inaccurate. One consequence of this is that the perceptual consolidation that normally is assumed to be expected to occur subsequent to identification of a “best fit” top-down representation to minimizing the error (Clark, 2013) would instead fail to occur. A lack of consolidation by a top-down representation – i.e., an explanatory metrical structure for a complex rhythm – may lead to failure of a top-down representation to achieve perceptual salience, a failure to reach conscious awareness in perception, and/or a failure to leave an imprint on memory. Under conditions of such a failure of consolidation to a top-down metrical and overall linguistic representation to the extent that neurotypical individuals would. As a result, we would expect people who stutter to evidence a weaker or null “memory trace” for a recently heard complex rhythm compared with a simple rhythm, or compared to a control group overall. In the event that the predictions from this nascent predictive coding account find support from the experiment to be presented in this dissertation, I will reiterate and elaborate on these ideas in the Discussion.

CHAPTER 2: METHODS

2.1 Participants

Participant data was collected from 18 adults who stutter (12 M, 6 F; age: *M* = 29.22, *SD* = 11.44, range = 18.33-53.08 years) and 18 adults who do not stutter (12 M, 6 F; age: *M* = 25.32, *SD* = 6.94, range 18.67-44.00 years). Both groups were recruited ensure an approximately equal average and range of age, and age did not significantly differ between the groups (see Table 1). At the first session of this study, all participants had to pass a speech, language, and cognitive evaluation to ensure typical speech and language developmental history except for presence of stuttering in the stuttering group (Stuttering Severity Instrument-4; Riley, 2009; scores ranged 13 to 37, *M* = 24.1, *SD* = 7.65). The stuttering group could be split into sub-groups based on the SSI-IV results as: 3 very mild, 7 mild, 4 moderate, 4 severe. All participants were right-handed (Edinburgh Handedness Inventory), monolingual, native speakers of English, with normal hearing (audiometer attenuator set to 20 dB, frequencies presented at 1000 Hz, 2000 Hz, 4000 Hz, 8000 Hz, 500 Hz, 250 Hz in right and then left ear), without concomitant developmental disorders such as dyslexia, ADHD, learning delay, or other confirmed developmental or psychiatric conditions (based on self-report), and were not taking any medication affecting the central nervous system. Additionally, the two groups were not significantly different on measures of expressive vocabulary, receptive vocabulary, education, working memory, or years of musical training (see Table 1 and Appendices).

The participants were recruited from East Lansing, MI and up to 100 miles surrounding the area through the Speech Neurophysiology Lab at Michigan State University. Participants were

compensated $10 per hour for their time at Michigan State University, and a mileage reimbursement was given for participants who drove over 50 miles roundtrip (note: only participants who stuttered were recruited from longer distances). All research procedures were approved by the Michigan State University Institutional Review Board, and each participant signed an informed consent.

2.2 Speech, language, and cognitive evaluation

At the first session of this study, all participants were given a battery of standardized speech, language, and cognitive tests to ensure typical development and homogeneity between the groups. The tests included the Peabody Picture Vocabulary Test (PPVT-4), Expressive Vocabulary Test (EVT-2), Goldman-Fristoe Test of Articulation (GFTA-2), and Operation Span (working memory).

Stuttering severity was assessed by reviewing video-recorded samples of speech elicited through conversational, monologue, and a reading passage (Friuli sample from the Stuttering Severity Instrument, SSI-4) with a certified Speech-Language Pathologist (CCC-SLP) or second-year Master's student in the Communicative Sciences and Disorders Department. The SSI-4 was used by a CCC-SLP to assess frequency and duration of stuttering-like disfluencies occurring in the speech sample, as well as any physical concomitants associated with stuttering. These measures were incorporated into a composite stuttering severity rating.

Measure	Group		t-test	
	Adults who do not stutter	Adults who stutter	*t*(34)	*p-value*
Age	25.33 (6.87)	29.38 (11.36)	1.29	0.205
Handedness	83.33 (16.45)	78.33 (15.05)	0.95	0.348
Education	15.28 (1.9)	15.83 (2.53)	0.75	0.461
Years of musical training	4.31 (3.86)	4.19 (4.39)	0.08	0.936
Peabody Picture Vocabulary Test	108.67 (8.86)	109.61 (9.58)	0.31	0.761
Expressive Vocabulary Test	108.11 (9.75)	115.11 (11.16)	2.00	0.053
Goldman-Fristoe Test of Articulation	100.83 (1.25)	100.83 (1.5)	0.00	1.000
Operation Span Score (Working Memory)	43.17 (16.79)	34.5 (17.87)	1.50	0.143
Percent of stuttering-like disfluencies per sample	NA	6.76 (4.05)	NA	NA
Stuttering Severity Instrument (SSI) total	NA	24.11 (7.65)	NA	NA

Table 1: Average (SD) scores for the measures collected relating to background, disfluencies, speech, language and cognitive tests.

2.3 Apparatus

Behavioral. The experiment programs were presented to participants using E-Prime v2.0 Professional (Psychology Software Tools, Inc.). Sounds were presented over Sennheiser HD 280 headphones. Responses were made by pressing buttons on a computer mouse, marked buttons on a computer keyboard, or a magnetic resonance compatible hand paddle.

Imaging. The scans were acquired on a 3T GE HDx scanner with an eight-channel head coil. The functional scans consisted of 6 runs of 6 min 45 sec echo-planar imaging datasets, starting from the most inferior regions of the cerebellum, and were acquired with the following parameters: 44 contiguous 3-mm axial slices in an interleaved order, time of echo (TE) = 27.7 ms, time of repetition (TR) = 2500 ms, parallel acceleration factor = 2, flip angle = 80 degrees, field of view (FOV) = 22 cm, and matrix size = 64 × 64.

After the functional data acquisition, the structural scan consisted of 180 T1-weighted 1-mm3 isotropic volumetric inversion recovery fast spoiled gradient-recalled images (10 minute scan time), with cerebrospinal fluid (CSF) suppressed, and were obtained to cover the whole brain with the following parameters: TE = 3.8 ms, TR of acquisition = 8.6 ms, time of inversion (TI) = 831 ms, TR of inversion = 2332 ms, flip angle = 8°, FOV = 25.6 cm × 25.6 cm, matrix size = 256 × 256, slice thickness = 1 mm, and receiver bandwidth = ± 20.8 kHz.

2.4 Stimuli

The stimuli were 12 simple and 12 complex rhythms selected from a larger set of simple and complex rhythms (Grahn & Brett, 2009). Rhythms were 5, 6, or 7 intervals long and all intervals within a rhythm were integer multiples of a base time unit, notated in Table 2 by a '1'. Notated values of 2, 3, and 4 indicate that the temporal intervals were two times, three times, or four times the duration of the base unit, respectively. The base unit will hereafter be referred to as the base inter-onset-interval (IOI) as it indicates the time interval between successive tone onsets delineating the interval. The base IOI was presented at 220, 245, and 270 ms for each rhythm. The frequency of the tones marking the rhythms also varied randomly from trial to trial and took on one of six values: 294, 353, 411, 470, 528, or 587 Hz.

For simple rhythms, intervals were organized into a sequence designed to elicit perception of a periodic accent every four base IOIs, which was predicted to induce the strong perception of a periodic beat (Povel & Essens, 1985; see Figure 3). In contrast, intervals comprising the complex rhythms were organized into a sequence so the accents were not periodic, and thus not expected to induce the perception of a periodic beat. Each simple rhythm had a corresponding complex rhythm

that was matched in the number of intervals. The ‘different’ variant of a rhythm involved swapping the order of a pair of nearby intervals; the different rhythms were the same as those used by Grahn and Brett (2009).

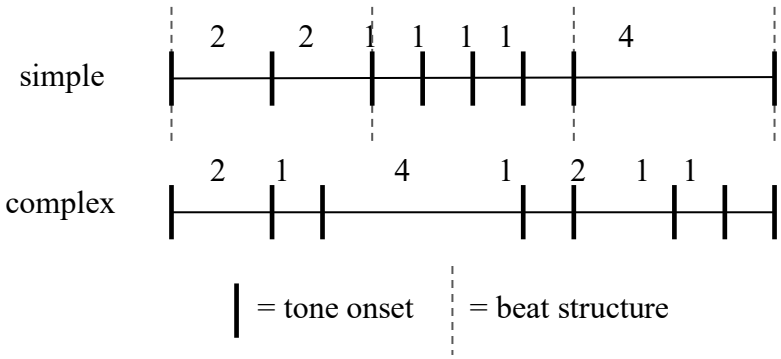

Figure 3: A schematic example of the types of rhythmic sequence stimuli used. The numbers represent the relative length of intervals in each sequence with 1 = 220–270 msec (value chosen at random on each trial), in steps of 10 msec.

	Simple		*Complex*	
	Standard	Different	Standard	Different
5 intervals	31413	31431	13242	31242
	31422	13422	23241	23214
	41331	43131	33141	31341
6 intervals	112314	112134	121233	122133
	211413	211431	214221	214212
	311322	313122	321411	324111
	422112	422211	421311	412311
7 intervals	1122114	1121124	1132131	1131231
	1123113	1123131	1132212	1132122
	2113113	2113131	2123211	1223211
	2211114	2112114	2141211	2142111
	4111131	4111113	3221112	3212112

Table 2: Table of simple and complex rhythm sequences used split by interval.

2.5 Procedure

Upon completion of tasks from session one (Section 2.2), participants were scheduled to participate in the fMRI paradigm of session two. Before entering the scanner, participants were given four practice trials outside of the scanner, consisting of same and different variants of one simple and one complex rhythm without any feedback (rhythms were different from the test trials). If a participant seemed confused by the task, the experimenter re-explained the instructions and re-ran the same practice session; this only occurred twice.

Inside the scanner, participants heard two successive presentations of a standard rhythm and was asked to judge whether a third comparison rhythm was the same or different from the standard (see Figure 4). The inter-onset-interval between presentations of each rhythm was 1300 ms, and the participant had 2100 ms to respond using an MR compatible hand paddle before the next trial started (index = “same”, middle = “different”). The trials were presented using an event-related design divided into six functional runs (6.75 min) of 24 trials where participants heard same and different variants of the 12 simple and 12 complex rhythms. The order of stimuli was determined by creating three sets of randomized orders of all 12 simple and 12 complex rhythms with half a same/different correct response and then the opposite correct response for other half of the pair. Each run consisted of 24 experimental trials with an additional 10 null trials lasting one or two TR. A total of 144 experimental trials were run across the session, which lasted approximately 40 minutes. The experimental stimuli methods of this study were based off previous research by Grahn and Brett (2007, 2009).

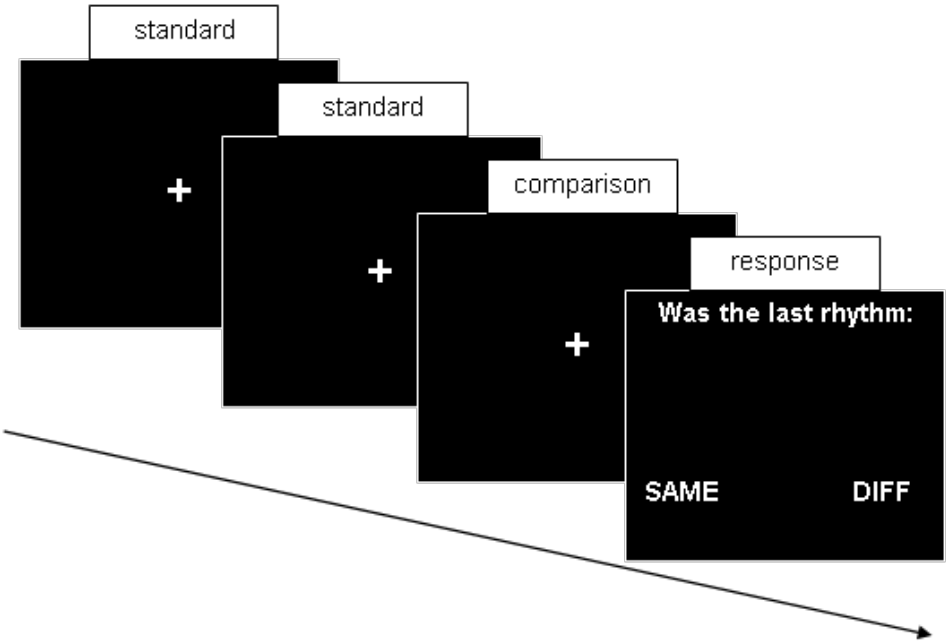

Figure 4: Stimuli used in the experiment. The standard rhythm was played twice, then the comparison rhythm was either the same or different, and then the participant answered "same" or "different".

2.6 Data Analysis

2.6.1 Behavioral

Performance on the rhythm discrimination task was first assessed using a signal detection analysis to distinguish between participants' ability to discriminate same and different rhythms from any general tendency to respond same or different (Macmillan & Creelman, 2005). Signal detection theory splits the data by whether the stimulus was present or absent, and then whether the participant determined the stimulus as present or absent. This method allows each trial to be sorted into one of four categories (see Table 3):

	Rhythm Different	Rhythm Same
Respond "Different"	Hit	False Alarm
Respond "Same"	Miss	Correct Rejection

Table 3: Table of possible trial categories in signal detection theory.

Then based on the proportion of each of these trial types, calculations can be made on a participant's sensitivity or response bias. This can be visually displayed as overlapping normal distribution curves (see Figure 5):

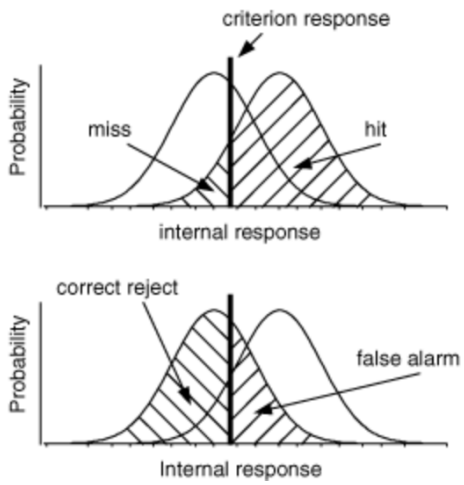

Figure 5: Internal response probability of occurrence curves for noise-alone and signal-plus-noise trials. Since the curves overlap, the internal response for a noise-alone trial may exceed the internal response for a signal-plus-noise trial. Vertical lines correspond to the criterion response.

Responding 'different' on trials when the comparison was different from the standard was treated as a 'hit' and responding 'different' on trials when the comparison was the same as the standard was treated as a 'false alarm'. Hit rates (HRs) and false alarm rates (FARs) were then used to calculate d' (a measure of sensitivity) and the response criterion, c (a measure of response bias) for simple and complex rhythms for each participant. Sensitivity, d', is determined by z(HR) – z(FAR), and the criterion, c, is determined by –0.5*[z(HR) + z(FAR)]. Values of $d' = 0$ correspond to chance performance, with larger values corresponding to better discrimination.

Values of $c = 0$ indicates no response bias, with negative values of c can be interpreted as a liberal response strategy (i.e., a tendency to respond 'different'), and positive values of c can be interpreted as a conservative response strategy (i.e., a tendency to respond 'same').

Separate 2 (Rhythm Type: simple, complex) x 2 (Group: AWS, Control) repeated measures ANOVAs were conducted on d' and c with Rhythm Type as a within-subject factor and Group as a between-subjects factor. ANCOVAs were also conducted to investigate possible sources of extraneous variability due to covariates of musical training and working memory.

Second, performance on the rhythm discrimination task was analyzed using linear mixed-effects modeling. Generalized linear models offer the option to model relationships between predictor variables and outcome measures in a flexible way based on normal (i.e., Gaussian) distributions. Mathematical extensions of the *generalized* linear model involve linking functions that enable dealing with data that involves probability density functions other than the normal Gaussian distribution (e.g., binomial distributions for data with binary outcomes, where two mutually exclusive outcomes of a trial exist such as accurate/inaccurate; Jaeger, 2008). Linear mixed-effects models also allow for simultaneously taking multiple random sources of variance (e.g., subject and item-level variance) into account within the same analyses, where this is a limitation of ANOVA, which can only account for one source of variance. A linear mixed-effects model with a logit linking function was employed here to simultaneously model subject and item random effects in a single analysis in order to understand how they influenced individual performance on a trial-by-trial basis. This form of modeling allows the encoding of assumptions about random effects regarding how sampling units (i.e., subjects and item) impact the variation/dependency in the data. These data were analyzed using logit mixed-effect models (Barr

et al., 2013; Jaeger, 2008) in the lme4 package (Bates et al., 2014) written for the R statistical package (version 2.11.1; the R foundation for statistical computing).

Published work has extended linear mixed effects models to signal detection theory (e.g., Wright & London, 2009), enabling separation of subjects' bias and sensitivity, while building on the power and advantages of these mathematical tools. A linear mixed-effects model was therefore constructed based on separate bias of individual subjects from subjects' sensitivity to fixed factors of theoretical interest. This involved including Trial Type as a fixed effect and having the predictor be 'responding different' instead of 'accuracy'. The model had three fixed factors of Group (control, stuttering), Rhythm Type (simple, complex), Trial Type (same, different), and all their interactions. Additionally, the model included random intercept terms for Item and Subject as well as random slopes of Subject indexed by Rhythm Type and Subject indexed by Trial Type. The independent variables included in the model were contrast coded to allow for easy interpretability by modifying the intercept to be the overall mean of the data, and the coefficients to be the distance of level from the mean. The coding was done as -1/1 respectively: Group (stuttering, control), Rhythm Type (simple, complex), and Trial Type (same, different).

Determining the appropriate fixed and random effects structure in linear mixed-effects models involves careful consideration by the researcher about the theoretical questions of interest and tradeoffs between Type I and Type II error (Barr et al., 2013; Bates et al., 2015; Matuschek et al., 2017). In a widely-cited paper, Barr and colleagues (Barr et al., 2013) argued through mathematical modeling that Type I error may be inflated by modeling subject- and/or item-level random effects in terms of random intercepts without also including relevant random slope terms across levels of any within-subjects or within-items fixed factors. Barr and colleague's (Barr et al., 2013) solution to this possibility of Type I error inflation was to argue for use of a maximal model

(i.e., a model that includes the maximal random effects structure that is justified by the experimental design). Subsequent published modeling research by Matuschek and colleagues (2017) instead showed that use of a maximal linear mixed effect model can be over-conservative with respect to Type I error and instead result in inflated Type II error and lost power. A compromise to balancing tradeoffs between Type I and II error is to select a model where the fixed effects structure enables testing theoretical predictions while the random effects structure is based on the pertinent dataset (Bates et al., 2015; Matuschek et al., 2017).

Building on current best practices, a linear mixed effect model was therefore constructed which involved fixed factors of Rhythm Type (simple, complex), Group (stuttering, control), Trial Type (same, different), and their interactions. This fixed effects structure showed the main effects and interactions for Rhythm Type and Group effects during the different trials (i.e. interaction with Trial Type), which allowed the hypotheses regarding performance to be answered. Additionally, the random effects structure justified by the experimental design (Bates et al., 2015; Matuschek et al., 2017) included random intercept terms for Item and Subject as well as random slopes of Subject indexed by Rhythm Type and Subject indexed by Trial Type. The random effects structure allowed the model to capture extraneous variance due to item-level differences in difficulty, subject-level factors (e.g., differing levels of sensitivity and bias), and differential performance by-subjects for each Rhythm Type (e.g., the possibility that some subjects performed more similarly in the two levels of Rhythm Type than other subjects) and Trial Type (e.g., the possibility that some subjects performed more similarly in the two levels of Trial Type than other subjects). Including this type of an interaction slope accounts for the theoretically relevant variation expected between groups and removes the significant three-way interaction, so by only looking at random slopes of Subject indexed by Rhythm Type and Subject indexed by Trial Type the model is able to fit the data set

more appropriately. The linear mixed-effects model design allowed for a structure similar to signal detection theory while accounting not only for subject sensitivity/bias, but also random subject/item level variance. This linear mixed-effects model was also tested against models with each intercept and slope term removed from the random effects structure. Utilizing a likelihood ratio test to compare each model, the model described above was deemed the best fitting model[1].

2.6.2 Imaging

Image preprocessing for the fMRI data was carried out using the Analysis of Functional Neuroimages (AFNI) software (Cox, 1996). The first 4 volumes were excluded to allow for fMRI signal stabilization. We modeled the correct and incorrect trials separately, and based on the design of this study and the lack of a significant difference between groups on overall incorrect trials[2] [$t(70) = 1.743$, $p = .086$], I will only discuss the correct trials in the analyses. Task-based functional images were corrected for slice acquisition timing differences, motion correction, and spatially blurred with a 4-mm FWHM kernel. The simple and complex conditions were modeled separately, and the BOLD percent signal change for each condition at each voxel was calculated using 3dDeconvolve in AFNI. Whole-brain analyses were carried in the ICBM452 standard template.

The group analyses for these data were analyzed using 3dANOVA3 (type 5), which allowed for a mixed-effects ANOVA analysis (fixed effects: Group, Rhythm Type; random effect: Subject) that accounted for subject-level variance. The contrasts between each condition utilized

[1] The maximal model was significantly better than: a model without Subject by Trial Type ($p < .001$), a model without Subject by Rhythm Type ($p < .001$), or a model without either slope term ($p < .001$).
[2] The proportion and standard deviation of correct responses for each group was adults who do not stutter simple: $M = 0.89$ ($SD = 0.11$), adults who do not stutter complex: $M = 0.77$ ($SD = 0.14$), adults who stutter simple: $M = 0.86$ ($SD = 0.12$), and adults who stutter complex: $M = 0.67$ ($SD = 0.16$).

pair-wise *t*-tests, which resulted in statistical maps for each contrast of interest. The identification of overlapping and distinct regions of activation for adults who do and do not stutter within the simple and complex rhythm conditions were based on individually thresholded statistical maps. 3dclustim was used to correct for multiple comparisons. A voxel-wise threshold at $p = 0.01$ and cluster threshold at 1350 voxels resulted in a corrected $p = .05$. Unless otherwise noted, all results are reported at this corrected $p = .05$.

CHAPTER 3: RESULTS

3.1 Behavioral

3.1.1 Analysis of Variance (ANOVA)/Analysis of Covariance (ANCOVA)

Figure 6 shows the mean response criterion, *c*, for simple and complex rhythms for the adults who do and do not stutter. An ANOVA on *c* revealed that participants did not vary in their response bias based on the rhythm type (i.e., tend to respond 'same' more to simple than complex rhythms) [$F(1,34) = .461$, $p = .502$, $\eta_p^2 = .013$; simple: $M = 0.11$, $SD = 0.30$; complex: $M = 0.15$, $SD = 0.39$]. Additionally, adults who stutter and adults who do not stutter did not statistically differ in their tendency to respond 'same' [$F(1,34) = .034$, $p = .856$, $\eta_p^2 = .001$; adults who stutter: $M = 0.12$, $SD = 0.36$; adults who do not stutter: $M = 0.14$, $SD = 0.34$], and no significant interaction of these main effects existed [$F(1,34) = 0.700$, $p = .409$, $\eta_p^2 = .020$].

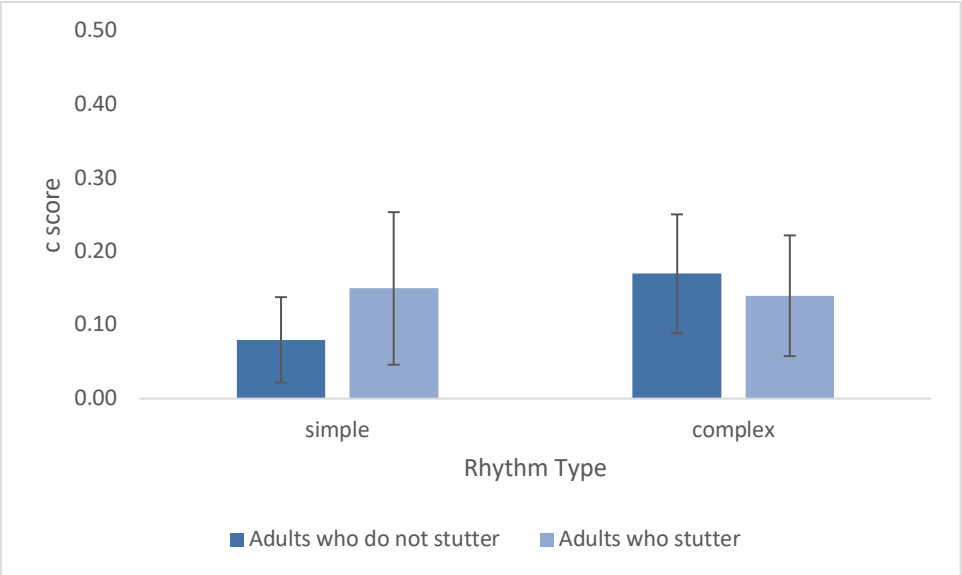

Figure 6: Mean *c* score for the adults who do and do not stutter for simple and complex rhythm types. Error bars show mean ± 1 standard error of the mean (SEM).

Figures 7-8 show mean *d′* scores for the simple and complex rhythms for the adults who do and do not stutter. An ANOVA on *d′* revealed better discrimination for simple than complex rhythms across both groups [$F(1,34) = 108.400$, $p < .001$, $\eta_p^2 = .761$; simple: $M = 2.70$, $SD = 1.19$; complex: $M = 1.41$, $SD = 1.12$]. Performance by adults who stutter and adults who do not stutter did not significantly differ [$F(1,34) = 1.876$, $p = .180$, $\eta_p^2 = .052$; adults who stutter: $M = 2.30$, $SD = 1.27$; adults who do not stutter: $M = 1.81$, $SD = 1.34$], and no significant interaction existed with $\alpha = 0.05$ [$F(1,34) = 3.046$, $p = .090$, $\eta_p^2 = .082$].

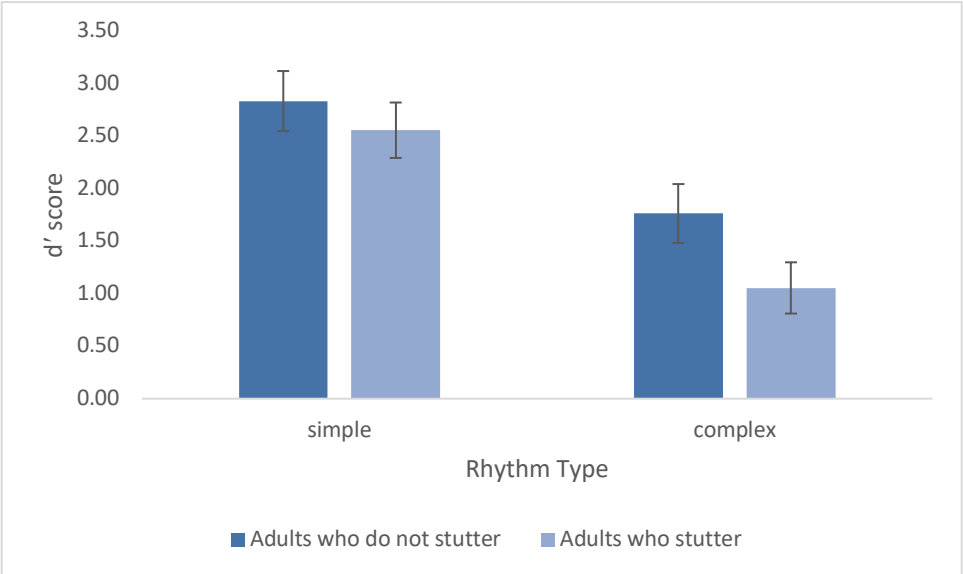

Figure 7: Mean *d′* score for the adults who do and do not stutter for simple and complex rhythm types. Error bars show mean ± 1 SEM.

The data were also examined in terms of possible covariates that might be driving the results for *d′* scores with a set of ANCOVAs. When using years of musical training as a covariate, the results remain the same, with a significant effect of rhythm type [$F(1,33) = 39.445$, $p < .001$, $\eta_p^2 = .544$], no group effect [$F(1,33) = 1.935$, $p = .173$, $\eta_p^2 = .055$], and no significant interaction [$F(1,33) = 3.150$, $p = .085$, $\eta_p^2 = .087$]. When using the measure of working memory (Operation

Span) as a covariate, the results are similar with a significant effect of rhythm type [$F(1,33) = 12.913$, $p = .001$, $\eta_p^2 = .281$], and no significant effect of group [$F(1,33) = 0.676$, $p = .417$, $\eta_p^2 = .020$] or interaction [$F(1,33) = 3.279$, $p = .079$, $\eta_p^2 = .090$].

An additional analysis was conducted by separating the d' scores for each of the stuttering severity sub-groups, which created smaller groups and decreased statistical power; this analysis is shown in Figure 5. An ANOVA showed the simple rhythm discrimination as being significantly better [$F(1,14) = 88.216$, $p < .001$, $\eta_p^2 = .863$], the main effect of severity group was not significantly different [$F(3,14) = 1.300$, $p = .313$, $\eta_p^2 =.218$], and no significant interaction was found between rhythm type or severity [$F(3,14) = 0.90$, $p = .965$, $\eta_p^2 =.019$]. An ANCOVA was further conducted using years of musical training as a covariate. Similar results were found to other analyses, with simple rhythm discrimination being better [$F(1,13) = 20.306$, $p = .001$, $\eta_p^2 = .610$], the main effect of severity group was not significantly different [$F(3,13) = 0.820$, $p = .506$, $\eta_p^2 = .159$], and no interaction was found [$F(3,13) = 2.288$, $p = .127$, $\eta_p^2 = .346$]. An ANCOVA was also run using working memory (Operation Span) as a covariate. In this analysis, simple rhythm discrimination was better [$F(1,13) = 10.232$, $p = .007$, $\eta_p^2 = .440$]; further, while no significant interaction was found [$F(3,13) = 0.222$, $p = .879$, $\eta_p^2 =.049$], there was a significant effect of severity group [$F(3,13) = 3.435$, $p = .049$, $\eta_p^2 =.442$]. This significant effect of severity group was driven by the very mild group having a greater mean than the other groups; however, pairwise comparisons did not result in significant differences between any pair of groups (very mild and mild: $p = .152$, very mild and moderate: $p = .197$, very mild and severe: $p = .054$).

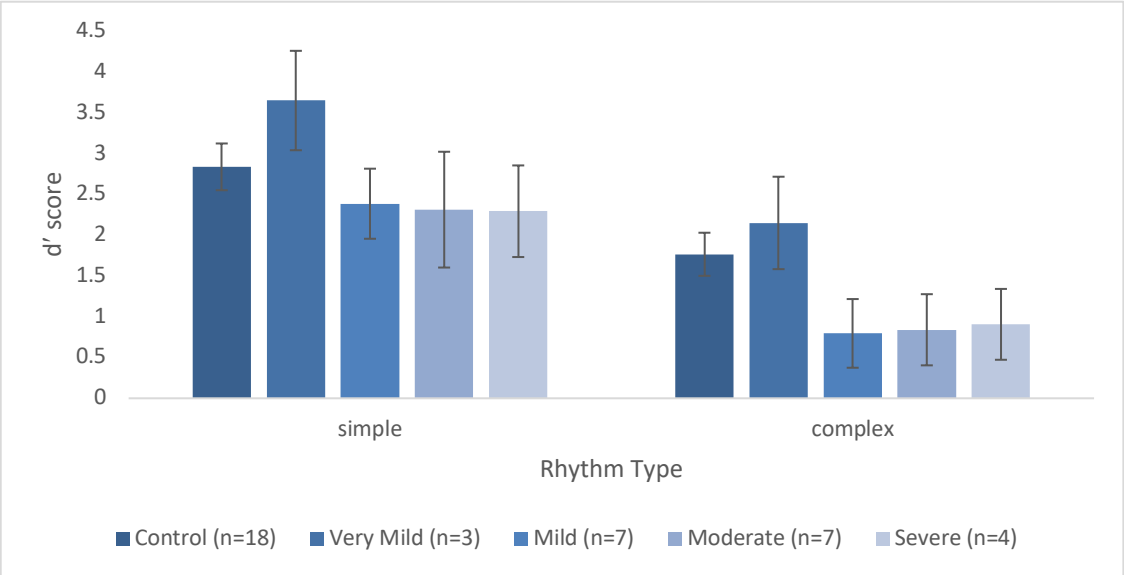

Figure 8: Mean *d′* scores for the adults who do and do not stutter for simple and complex rhythm types split by stuttering severity sub-groups. Error bars show mean ± 1 SEM.

3.1.2 Linear Mixed-Effects Modeling

Figure 9 shows the proportion of "different" responses for simple and complex rhythms for the adults who do and do not stutter. The linear mixed-effects model did not show significant main effects for Group (β = -0.04528, *SE* = 0.09665, *z* = -0.469, *p* = .6394) or Rhythm Type (β = -0.0114, *SE* = 0.010, *z* = -0.118, *p* = .058); however, Trial Type was significant (β = 1.91, *SE* = 0.185, *z* = 10.340, *p* < .001), demonstrating that a "different" response was given more for different trials. The interactions were not significant for Group by Rhythm Type (β = 0.068, *SE* = 0.057, *z* = 1.202, *p* = .230) or Group by Trial Type (β = -0.252, *SE* = 0.173, *z* = -1.459, *p* = 0.145), but the Rhythm Type by Trial Type interaction was significant (β = -0.500, *SE* = 0.069, *z* = -7.223, *p* < .001). Additionally, a significant three-way interaction was found for Group by Rhythm Type by Trial Type (β = -0.083, *SE* = 0.041, *z* = -2.005, *p* = .045). The amount of variance in the data that was accounted for by the random effects structure of the model was: Item intercept = 0.131 (*SD* =

0.362), Subject intercept = 0.258 (*SD* = 0.510), slope of Subject indexed by Rhythm Type = 0.049 (*SD* = 0.221), and slope of Subject indexed by Trial Type = 0.989 (*SD* = 0.994).

The analysis of simple effects for the Rhythm Type by Trial Type interaction showed that although both simple and complex rhythm types had more responses of "different" when trials were different, there was a greater change in numbers of "same" and "different" responses for simple ($\beta = 2.416$, $SE = 0.210$, $z = 11.553$, $p < .001$) than for complex rhythms ($\beta = 1.406$, $SE = 0.195$, $z = 7.216$, $p < .001$). This can be interpreted as simple rhythms discrimination being better than complex rhythms, as was expected with this paradigm. The analysis of the simple effects for the three-way interaction was done by splitting the data by Group and examining the simple effects of Rhythm Type by Trial Type. Both groups showed the same pattern where each rhythm type had more responses of "different" on trials involving different stimuli than trials involving the same stimuli, where this response was greater for simple (adults who stutter: $\beta = 2.239$, $SE = 0.252$, $z = 8.871$, $p < .001$; adults who do not stutter: $\beta = 2.526$, $SE = 0.296$, $z = 8.535$, $p < .001$) than for complex rhythms (adults who stutter: $\beta = 1.001$, $SE = 0.238$, $z = 4.205$, $p < .001$; adults who do not stutter: $\beta = 1.714$, $SE = 0.288$, $z = 5.960$, $p < .001$). By splitting the data by Rhythm Type, the Group by Trial Type interaction was significant ($\beta = -0.314$, $SE = 0.152$, $z = -2.060$, $p = 0.039$), which indicates that the participants who stutter were significantly less accurate at complex rhythms than participants that do not stutter.

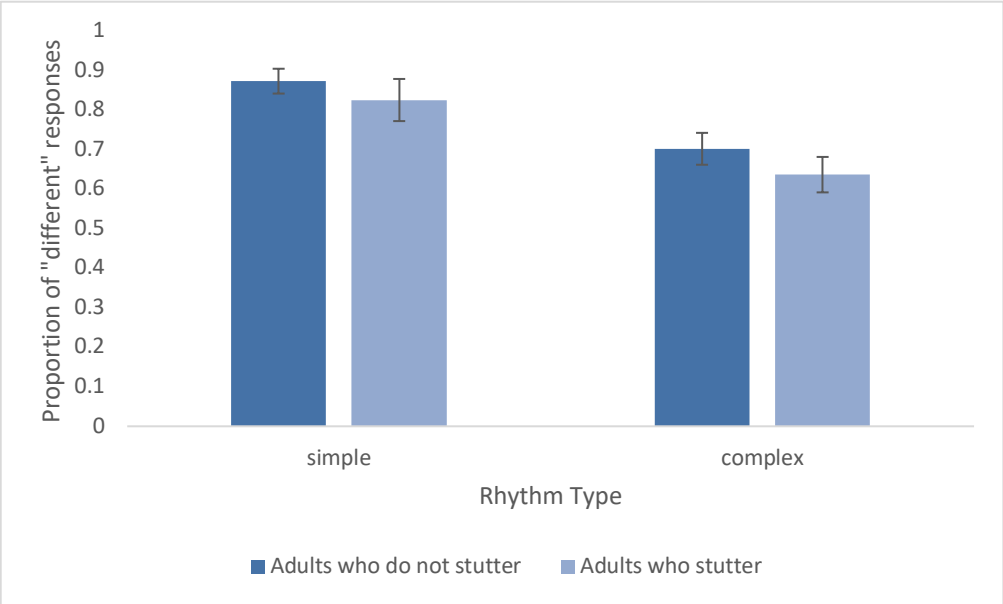

Figure 9: Proportion of "different" responses for adults who do and do not stutter for simple and complex rhythm types. Error bars show mean ± 1 SEM.

3.2 Imaging

3.2.1 Brain activity associated with simple and complex rhythm discrimination

The whole brain analysis for the Rhythm Types of simple and complex showed similar activation in many regions of the rhythm processing network (Figure 10, Tables 4-6). During the simple rhythms (compared to null trials), significant activity was observed in the bilateral superior temporal gyrus (STG), putamen, supplementary motor area (SMA), and premotor areas. During the complex rhythms (compared to null trials), significant activity was observed in the bilateral STG, SMA, and premotor areas, as well as left putamen. When the simple and complex rhythms were contrasted collapsing across both groups, the left STG, left inferior frontal gyrus (IFG), and bilateral putamen were more active for simple rhythms than for complex rhythms.

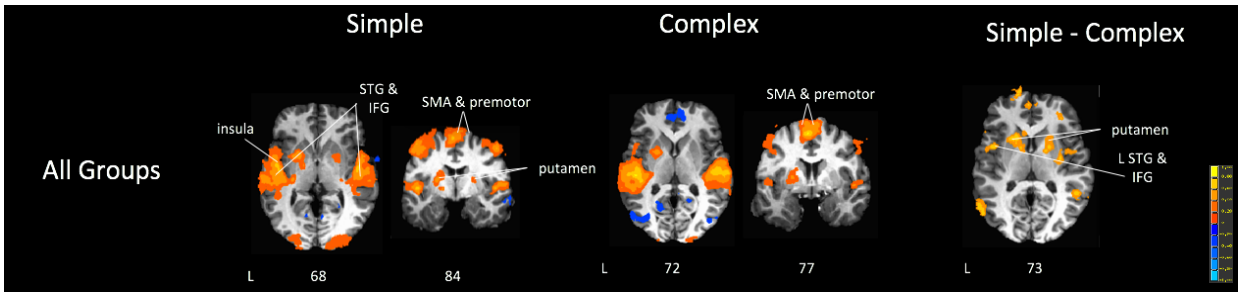

Figure 10: Contrasts for simple, complex, and simple – complex rhythms across both groups. Areas with significant activation are labeled.

Region	x	y	z	Size	Peak t
L STG, L insula, L IFG	-52	-22	10	44065	10.947
R cuneus	17	-87	25	37362	-6.245
R STG/BA42	63	-25	15	25502	9.085
L declive	-30	-84	-16	17211	5.841
L SMA	-4	4	57	16293	9.061
R lingual gyrus	27	-88	-6	12217	5.743
L precentral gyrus/BA6	-49	-3	57	10944	7.223
R precuneus	2	-37	45	8281	-4.913
R culmen	25	-52	-25	6418	5.146
R medial frontal gyrus/BA9	6	43	17	6131	-4.842
R fusiform gyrus	28	-52	-9	5392	-6.131
R precental gyrus/BA6	50	-4	51	4681	6.646
L parahippocampal gyrus/BA36	-28	-39	-8	4338	-4.939
R lentiform nucleus/putamen	24	5	8	2865	5.386
R STG/BA39	52	-60	18	2206	-4.829
R cerebellum	24	-58	-50	1687	5.099

Table 4: Regions activated during the rhythm discrimination task for simple rhythms. A voxel-wise threshold at $p = 0.01$ and cluster threshold at 1350 voxels resulted in a corrected $p = .05$.

Region	x	y	z	Size	Peak t
R cuneus/BA19	16	-87	26	53473	-6.217
L STG/BA41 (insula?)	-43	-27	12	30107	11.43
R STG	51	-17	7	20788	8.978
R medial frontal gyrus/BA9	8	44	17	19181	-5.647
R precuneus	1	-37	44	16196	-5.53
L fusiform gyrus/BA18	-28	-86	-15	13906	5.453
L medial frontal gyrus/BA6 (L SMA)	-5	3	59	13113	9.305
R inferior occipital gyrus	29	-89	-10	10854	5.513
L precentral gyrus	-50	-3	56	7373	6.96
R culmen (cerebellum)	17	-56	-9	5689	-5.929
L fusiform gyrus	-26	-51	-9	4396	-4.909
R precentral gyrus/BA6	50	-5	52	3477	6.105
R culmen	35	-57	-24	3115	4.944
L lentiform nucleus/L putamen	-22	6	8	2587	6.111
L superior frontal gyrus/BA8	-21	29	43	2192	-4.382
R cerebellar tonsil	25	-58	-49	1428	4.779

Table 5: Regions activated during the rhythm discrimination task for complex rhythms. A voxel-wise threshold at p = 0.01 and cluster threshold at 1350 voxels resulted in a corrected p = .05.

Region	x	y	z	Size	Peak t
Rhythm Type (Simple-Complex Contrast)					
R precuneus	12	-38	46	5872	-5.152
R parahippocampal gyrus	14	-17	-12	4736	-4.389
R inferior temporal gyrus	56	-60	-10	4227	-4.116
L middle temporal gyrus	-57	-68	12	4068	-5.802
L caudate & L putamen	-16	17	11	3544	-4.722
R inferior semi-lunar lobule	25	-76	-37	2590	-4.332
R middle temporal gyrus/BA19	38	-58	16	2231	-4.684
L postcentral gyrus/BA3	-36	-20	45	2149	-4.048
R medial frontal gyrus/BA10	6	54	12	2053	-4.805
L STG & L IFG	-44	7	8	1429	-4.24

Table 6: Regions activated during the rhythm discrimination task for the simple-complex rhythm contrasts. A voxel-wise threshold at p = 0.01 and cluster threshold at 1350 voxels resulted in a corrected p = .05.

3.2.2 Group-specific findings

The whole brain analysis was conducted separately for adults who stutter and adults who do not stutter. Both adults who do and who do not stutter showed activation in many regions of the rhythm processing network (Figure 11, Tables 7-8). For the adults who do not stutter, significant activity during the simple and complex rhythm conditions were similarly located in the bilateral insula, STG, SMA, and premotor area. When the simple and complex rhythms were contrasted within the adults who do not stutter, the left STG and left putamen were more active for simple rhythms than complex rhythms, though slightly below the statistical threshold (both at a cluster threshold of 900). For the adults who stutter, a much greater extent of activity was observed encompassing the rhythm processing network. Significant activity during the simple and complex rhythm conditions were similarly located in the bilateral insula, STG, SMA, premotor area, putamen, and IFG. When the simple and complex rhythms were contrasted within the adults who do stutter, the right insula and putamen were more active for simple rhythms than complex rhythms.

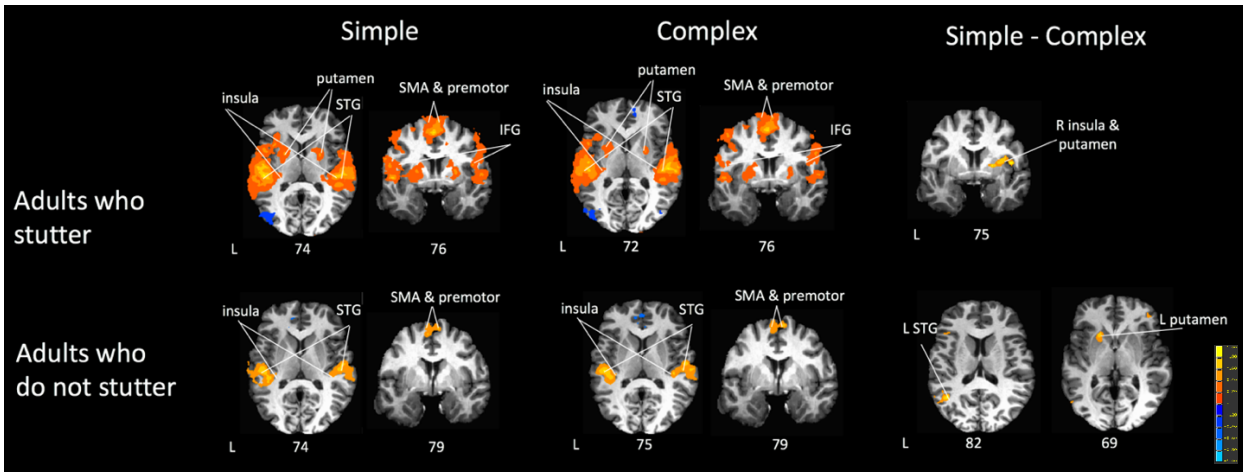

Figure 11: Contrasts for each Group (adults who stutter and adults who do not stutter) at each Rhythm Type (simple and complex). Areas with significant activation are labeled.

Region	x	y	z	Size	Peak t
Adults who stutter (Simple rhythm)					
L STG, L insula, L IFG, L putamen	-50	-23	9	46201	11.73
R STG, R insula, R IFG	64	-30	16	35434	13.575
L SMA, L BA6 (premotor)	-3	5	55	16472	11.619
R SMA, R BA6 (premotor)					
L cuneus/BA19	-4	-85	27	15828	-6.525
R culmen	31	-56	-24	13724	7.241
L tuber	-37	-58	-25	11504	5.481
L precentral gyrus	-47	-4	56	11477	9.074
R putamen	24	4	8	3206	6.487
R culmen	17	-56	-9	2338	-5.303
R inferior partietal lobule/BA40	53	-39	50	1889	4.861
L declive	-15	-55	-11	1677	-5.5
Adults who stutter (Complex rhythm)					
L STG, L insula, L IFG	-41	-29	12	37116	11.838
L putamen					
R STG, R insula	56	-14	4	30018	13.209
R cuneus	17	-87	25	17370	-6.364
L SMA, L BA6 (premotor)	-3	5	56	14098	10.946
R SMA, R BA6 (premotor)					
R culmen	32	-55	-24	12993	7.847
L precentral gyrus	-47	-3	55	10572	9.52
L inferior occipital gyrus	-29	-91	-5	7967	5.575
L middle temporal gyrus/BA39	-51	-76	11	4576	-5.446
R culmen	17	-56	-9	2477	-5.45
L precuneus	-19	-41	42	2220	-5.412
R inferior parietal lobule/BA40	52	-38	51	2020	5.031
R superior frontal gyrus (medial)	13	58	22	1881	-4.417
$^{+}$R putamen	24	0	9	994	5.613
Adults who stutter (Simple-Complex Contrast)					
R insula & R putamen	44	5	10	1359	5.523

Table 7: Regions activated during the rhythm discrimination task for the adults who stutter during simple and complex rhythms. A voxel-wise threshold at p = 0.01 and cluster threshold at 1350 voxels resulted in a corrected p = .05 ($^{+}$voxel threshold at 900).

Region	x	y	z	Size	Peak t
Adults who do not stutter (Simple rhythm)					
L STG, L insula	-37	-29	14	9322	6.868
R STG, R insula	56	-12	10	5522	4.994
R cingulate gyrus	9	35	27	5490	-5.213
R middle occipital gyrus	30	-78	16	2395	-5.002
R superior frontal gyrus	23	47	22	2274	-4.49
L precuneus	-1	-36	44	2227	-4.995
L declive	-31	-82	-16	1934	3.955
R middle frontal gyrus/BA8	22	25	39	1707	-4.922
R fusiform gyrus	24	-85	-12	1602	4.108
L SMA, L BA6 (premotor)	-6	3	59	1571	4.664
R SMA, R BA6 (premotor)					
Adults who do not stutter (Complex rhythm)					
R anterior cingulate/BA32	8	38	24	10288	-5.957
L STG, L insula	-37	-26	14	7549	6.7
R STG, R insula	56	-12	10	4858	5.118
L precuneus	-14	-45	49	4718	-5.785
L SMA, L BA6 (premotor)	-5	4	59	1648	4.986
R SMA, R BA6 (premotor)					
R middle frontal gyrus/BA8	22	25	38	1597	-4.823
L cuneus/BA19	-25	-86	37	1404	-4.356
Adults who do not stutter (Simple-Complex Contrast)					
R inferior semi-lunar lobule	37	-66	-45	4033	6.123
L postcentral gyrus/BA3	-39	-19	-51	2842	5.19
L angular gyrus	-30	-56	36	2096	4.578
$^{+}$L STG/BA22 (middle)	-42	-56	17	1327	6.191
$^{+}$L putamen	-20	19	4	954	4.784
$^{+}$L middle frontal gyrus, L IFG	-52	26	28	924	5.313

Table 8: Regions activated during the rhythm discrimination task for the adults who do not stutter during simple and complex rhythms. A voxel-wise threshold at $p = 0.01$ and cluster threshold at 1350 voxels resulted in a corrected $p = .05$ ($^{+}$voxel threshold at 900).

3.2.3 Group Contrasts

The group contrast collapsed across the rhythm conditions revealed significantly greater activity for adults who stutter relative to controls in the right superior temporal gyrus and right inferior frontal gyrus. The right insula was significant at a slightly lower cluster threshold (k=1000). This contrast suggests that greater activity in right hemisphere's STG, IFG, and insula in adults who stutter relative to controls was associated with comparable levels of accuracy in the two groups in rhythm discrimination (Figure 10, Table 9).

When the two groups (adults who stutter, adults who do not stutter) were compared within the simple rhythm condition, greater activity was found in the right STG, insula, and IFG for adults who stutter relative to adults who do not stutter. When the two groups were compared within the complex rhythm condition, adults who stutter showed significantly greater activity in the right STG and IFG relative to adults who do not stutter.

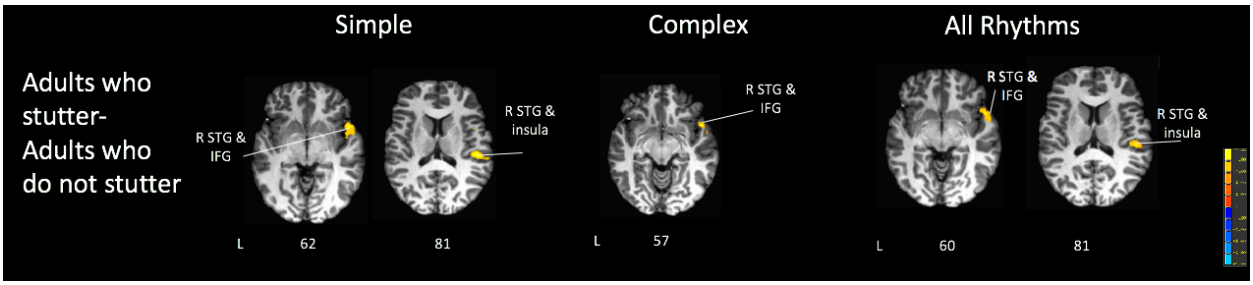

Figure 10: Contrasts for Rhythm Type simple (adults who do not stutter – adults who stutter) and complex (adults who do not stutter – adults who stutter) rhythms, and Group adults who do not stutter (simple – complex), adults who stutter (simple – complex), and the overall group contrast (adults who stutter – adults who do not stutter). Areas with significant activation are labeled.

Region	X	y	z	Size	Peak t
Simple Rhythms (Adults who stutter - Adults who do not stutter Contrast)					
R STG & R IFG	-53	-14	-3	2787	4.329
R STG & R insula	-47	25	16	1427	4.38
Complex Rhythms (Adults who stutter - Adults who do not stutter Contrast)					
R STG & R IFG	-51	-15	-8	2017	5.205
All Rhythms (Adults who stutter - Adults who do not stutter Contrast)					
R STG/BA38	-51	-15	-8	2232	4.896

Table 9: Regions activated during the rhythm discrimination task for Rhythm Type simple (adults who stutter – adults who do not stutter) and complex (adults who stutter – adults who do not stutter) rhythms, and Group adults who do not stutter (simple – complex), and the overall group contrast (adults who stutter – adults who do not stutter). A voxel-wise threshold at $p = 0.01$ and cluster threshold at 1350 voxels resulted in a corrected $p = .05$.

CHAPTER 4: DISCUSSION

4.1 Prior motivation and abbreviated summary

This dissertation was motivated by the hypothesis that people who stutter, compared to those who do not, have differences in neural circuits mediating timing prediction, and by extension, differences in timing accuracy related to the initiation and control of ongoing motor movements. This research built on several published papers (Grahn & Brett, 2007, 2009; Grahn & Rowe, 2009; Wieland et al., 2015) that used variants of a perceptual task developed by Grahn and Brett (2007) to investigate neural areas involved in beat perception during rhythm discrimination. The Grahn and Brett (2007) discrimination task involves judgments of whether a tone sequence is the same or different rhythm relative to a comparison sequence using sequences comprising both *simple* and *complex* rhythms; simple rhythms evoke a strong sense of periodic accent enabling one to "tap to the beat", while complex rhythms do not evoke a strong sense of periodic accent. This has been shown to lead the former to be more easily discriminated, an effect known as the "beat-based advantage" (Grahn & Brett, 2007, 2009). The present study particularly built on the work of Grahn and Brett (2009), who investigated neural activation during rhythm discrimination tasks for participants with Parkinson's Disease (PD), a group for whom similar brain regions are affected as in stuttering. In that study, Grahn and Brett (2009) found that PD patients showed poorer rhythm discrimination when compared with matched controls.

To test the hypotheses of poorer performance in rhythm discrimination and distinct neural activation for adults who stutter relative to those who do not, both groups performed the Grahn and Brett (2007) rhythm discrimination task while brain activity was acquired using fMRI. As

predicted, significantly reduced performance was shown in rhythm discrimination judgments for complex compared to simple rhythms for adults who stutter compared to those who do not. These findings support the hypothesis that dysfunction in the neural networks underlying processing rhythm/timing may be a causal component of stuttering as a neurodevelopmental fluency disorder. Combining behavioral and functional methodology represents a novel approach that affords new insight into neural activation involved in rhythm and timing judgments in people who stutter.

In the following, these findings are situated within theoretical frameworks as well as prior findings in the literature. First, I address the details of the behavioral and neural findings in this study. Next, I address how a predictive coding framework may permit an account of these results. Lastly, I consider converging evidence supporting the notion that timing deviations may be introduced that lead to error in the matching process between top-down models and bottom-up signals, and draw upon empirical findings from recent years to outline a new hypothesis relating to how neural structures engage to impute a top-down meter and how this relates to core deficits in stuttering.

4.2 Recap of findings of the research

The first research question focused on whether adults who stutter would show a rhythm discrimination deficit compared to adults who do not stutter. Based on previous research (Boutsen et al., 2000; Kent, 1984; Olander et al., 2010; Wieland et al., 2015), adults who stutter were hypothesized to have less accurate rhythm discrimination performance than adults who do not, particularly for complex over simple rhythms. In the task, participants heard two rhythms that varied in length from five to seven intervals (six to eight tones). Each simple rhythm was matched

to a complex rhythm in number of intervals, and in the temporal values of the inter-onset-intervals, which merely occurred in a different order across the paired simple and complex rhythm stimulus items.

Participants' performance on the rhythm discrimination task was analyzed using linear mixed-effects modeling, which allowed for modeling subject and item-level variance within the same analysis. The model included fixed-factors of Rhythm Type (simple, complex), Group (stuttering, control), Trial Type (same, different), and their interactions. As predicted, discrimination for simple rhythms across both groups was significantly better than complex rhythms, consistent with prior results (Grahn & Brett, 2009), with no significant difference observed between groups on discrimination for simple rhythms. Crucially, evidence was obtained for the proposed hypothesis in the form of the predicted significant three-way interaction. In particular, this result revealed that when the rhythms were more complex, harder, and/or afforded fewer cues for metrical prediction, participants who stutter were less accurate at discriminating these rhythms than people who do not stutter.

These findings revealed during the harder task raises questions about patterns of neural activation during these rhythm discrimination activities; this led to the second research question of whether hypothesized rhythm discrimination deficits for adults who stutter, compared to those that do not, would manifest as aberrant neural activation. Based on previous research (Alm, 2004; Chang et al., 2011; Chang & Zhu, 2013; Ingham et al., 2012), it was hypothesized that during the rhythm discrimination task adults who stutter would show less neural activation in the internal rhythm processing network, and possibly greater neural activation in compensatory regions, than adults who do not stutter.

To briefly recap the results as a preliminary to fuller interpretation within the context of the proposed extension of predictive coding framework, the findings showed that neural processing during rhythm discrimination was significantly different across the groups. Adults who stutter had greater overall activation in the rhythm processing network (i.e., bilateral superior temporal gyrus, STG; putamen; insula; supplementary motor area, SMA; and premotor area) during the rhythm discrimination task than adults who do not stutter, which held true for both simple and complex rhythms. Following standard functional fMRI data analysis procedures for this experimental design, the neural data were analyzed over correct trials only.

The findings revealed support for the main assertion of the second hypothesis; namely, as predicted, activation patterns were indeed aberrant in the rhythm network for adults to stutter compared to those that do not, in at least two respects. First, during processing of both simple and complex rhythms, adults who stutter showed activation that survived the contrast analysis in two regions of the brain which was not the case for the control participants, in the bilateral IFG and bilateral putamen. Second, adults who stutter showed differentially greater activation in right insula and right putamen for simple vs. complex trials, whereas controls showed differentially greater activation in left STG and left putamen for this comparison.

The unexpected pattern of results which went against the initial formulation of the second hypothesis concerned the amount of overall activation. That is, while I had predicted weaker activation in the rhythm perception network for adults who stutter compared with controls, the results instead revealed *greater* activation for adults who stutter compared to the control group diffusely and broadly distributed throughout the rhythm network. Having recapped these results, I now turn to developing an accounting of them within extensions of a predictive coding framework

for perception, action, cognition, and speech-language planning and processing, in the context of prior relevant research.

4.3 A predictive coding account of metrical structure-building

4.3.1 Introduction

I propose that the predictions of the two main experimental hypotheses, as discussed above, can be accommodated by developing a novel extension of predictive coding frameworks for speech, language and general auditory perception. While a full theoretical treatment is beyond the scope of this dissertation, in this section I develop a brief overview of the novel predictive coding account of my results. In the following section I consider how this proposal recasts issues of core deficiencies in stuttering. The proposal leads to a number of novel hypotheses and innovative new directions for future work regarding how metrical structure is computed in the brain.

This section is organized as follows. First, I review the main tenets of predictive coding frameworks, including how they have been applied to music and language with regard to dynamic, predictive reconstruction of acoustic patterns in the brain. Next, I will consider the role of knowledge and experience in how internal top-down representations of linguistic and musical structure are constructed over acoustic stimulus properties, including metrical hierarchical structure. I review a novel extension of predictive coding for linguistic structure-building, namely the Syllable Inference account of Brown and colleagues (in press), which posits top-down predictive processes for prosodic structure building as a component of overall top-down linguistic structure building. I further consider how notions of “beat” and “meter” in speech and music are

related to auditory-motor predictive affordances imputed through top-down metrical hierarchical structure. Next, I review issues related to integration of information across the brain, including the role of oscillatory coherence across white matter tracts in cortico-cortical and cortico-subcortical interactions. I then review core deficits and findings in stuttering literature before I finally reconsider findings from behavioral and neural evidence from the simple and complex conditions across groups.

4.3.2 Applications of predictive coding to language and speech processing

Predictive coding is a research framework that draws on insights of scientists over more than a century and posits that the brain is a "prediction engine" (Bever, 2018; Bever & Poeppel, 2010; Clark, 2013; Friston, 2010; Halle & Stevens, 1962; Helmholtz, 1860; Rao & Ballard, 1999). Predictive coding offers the core insight that the brain is fundamentally organized such that neural processes carry out perception, action, and cognition via internal generation and recapitulation of past experiences, attempting to match these to a series of intermediate hierarchical bottom-up representations (Friston, 2010; Rao & Ballard, 1999). Similar principles appear to apply in both auditory and visual domains (Koelsch et al., 2019; Olasagasti et al., 2015). Crucially, perception, action, and cognition involve the brain's generation of top-down candidate representations that compete to provide the "best match" to bottom-up sensory indices of external input, including from signals heard as linguistic messages (e.g., noise-vocoded or natural speech) and non-linguistic information (e.g., music) (Di Liberto et al., 2018; Donhauser & Baillet, 2020; Kaufeld, Bosker, et al., 2020; Meyer et al., 2020a, 2020b; Vuust et al., 2018). The core assertion of predictive coding that the brain instantiates multiple layers of hierarchically linked top-down and

bottom-up representations is supported by studies of neural activation (e.g., Clark, 2013; Keitel et al., 2018).

Initial extensions of predictive coding theory from vision research (Friston, 2010; Rao & Ballard, 1999) to language focused on accounting for higher-level linguistic prediction processes involving, e.g., syntax, semantics, and lexical selection (Dell & Chang, 2014; Levy, 2008; Pickering & Garrod, 2013). There have also been more recent extensions to general auditory processing (Koelsch et al., 2019) and music (Vuust et al., 2018). Particularly compelling has been more recent work applying predictive coding to speech processing in the brain (Daube et al., 2019; Di Liberto et al., 2018; Hovsepyan et al., 2020; Olasagasti et al., 2015; Yi et al., 2019). Since work on predictive speech processing holds relevance for arguments I develop around predictive metrical structure-building by the brain, I briefly review this research below.

It is well-known that neurons in the brain show temporal alignment with respect to speech amplitude envelope modulation due to alternations of consonants and vowels (Goldsmith, 2011; Peelle & Davis, 2012; Stevens, 2002). In particular, considerable work has shown that neural oscillations particularly in the theta band (4-8 Hz) "keep pace" with rate of speech signal envelope modulation up to about 30% of the original duration, which is the approximate limit of intelligibility (Ahissar et al., 2001; Dupoux & Green, 1997; Luo & Poeppel, 2007; Peelle & Davis, 2012). Originally, it was proposed that this close temporal alignment between the phase of neural oscillations and speech amplitude envelope onsets reflected early perceptual processes of segmentation and edge detection (Ahissar et al., 2001; Ding & Simon, 2014; Giraud & Poeppel, 2012; Luo & Poeppel, 2007). Ghitza and colleagues (2013, 2014; 2009) particularly developed proposals that neurons in the theta-band held a causal function of speech segmentation, such that a failure to understand very time-compressed speech reflected temporal rate-limits for theta

neuronal dynamics to align with amplitude onsets, thereby causally limiting human speech processing beyond.

Further work instead has supported that temporal alignment of neural oscillations, including theta, with amplitude onsets in speech, reflects active processes of structure- and meaning-building in language, rather than mere segmentation (Bourguignon et al., 2020; Ghitza, 2013, 2014; Gross et al., 2013; Keitel et al., 2018; Klimovich-Gray & Molinaro, 2020; Lizarazu et al., 2019; Luo & Poeppel, 2007; Molinaro & Lizarazu, 2018; Peelle & Davis, 2012; Pefkou et al., 2017). A notable experiment in this regard was carried out by Pefkou et al. (2017), who tested Ghitza and colleagues' (2013, 2014; 2009) claim that failures to understand very time-compressed speech reflected temporal rate-limits of theta neuronal dynamical entrainment to amplitude onsets. Pefkou et al. (2017) found evidence against this theta-entrainment dependency for understanding; instead, they observed that with increasingly fast speech, theta continued to show synchronous phase-alignment with speech amplitude onsets at rates of speech even faster than the limits of human speech understanding. Pefkou et al's (2017) experiment thus supported the notion that rate limitations on human speech understanding were instead due to time constants on retrieval of top-down candidate linguistic representations from cortex, which needed to be synchronized with bottom-up information in the theta band. Further studies have shown additional evidence that synchronous coordination between neural oscillatory dynamics in theta, as well as delta (1-4 Hz) and gamma (40-80 Hz) bands reflect active top-down structure building in the brain (Bourguignon et al., 2020; Donhauser & Baillet, 2020; Kaufeld, Bosker, et al., 2020; Kaufeld, Ravenschlag, et al., 2020; Meyer et al., 2020b), including the active building of phonemic representations (Daube et al., 2019; Di Liberto et al., 2018; Di Liberto et al., 2015).

4.3.3 Comparisons of structure and affordance in language and music

While (psycho)linguistic theories and empirical research have highlighted an important role for prosody in overall linguistic structure-building in the brain (Keitel et al., 2018; Ladd, 2008; Patel, 2010), important questions remain about the building of prosodic, metrical and phonological structures in language and in music. Many linguistic theorists have noted the similarity of organizational structures of metrical hierarchies across both and music (Ladd, 2008; Liberman, 1975; Patel, 2010). In both domains, one speaks of *relative strength or prominence* among auditory events in sequence, which are layered hierarchies of variable strength as well as of *grouping into phrasal units* (Halle & Idsardi, 1995; Hayes, 1995; Jones, 2018; Kotz et al., 2018; Ladd, 2008; Nespor & Vogel, 2007; Patel, 2010). Empirical work supports that neither speech nor performed music consists of equal intervals (Lehiste, 1977; Repp, 1998). Likewise, the stimuli used in the present experiment entailed considerable temporal irregularity from onset to onset across all items for both simple and complex stimuli. As a result, "fitting a meter" to the stimulus is not a trivial matter of determining which stimulus has regular inter-onset-intervals, since this was true neither for simple nor for complex rhythmic stimuli in the present experiment.

An important observation is that in both language and music, metrical hierarchical structure can only *partially* be "read off the signal", via acoustic cues such as increased amplitude, longer duration, facets of pitch, and or statistical quasi-regularity of intervals (Arvaniti, 2009; Hannon et al., 2004; Kotz et al., 2018; Lehiste, 1970; Thomassen, 1982; Tilsen & Arvaniti, 2013; Tilsen & Johnson, 2008). In language, lexical stress entails language-specific, probabilistic coincidences of vowel quality, amplitude, temporal, and fundamental frequency cues, suggesting stress must in part be *imputed* (Brown et al., 2015; Kondaurova & Francis, 2008; Lehiste, 1970). Likewise in

music, it is well-known that metrical structure must be imputed top-down from a string of notes or percussive beats, since the timing in a given string of notes is technically physically compatible with many alternative underlying meters, not all of which are perceived by listeners (Koelsch, 2011; Povel & Essens, 1985; Vuust et al., 2018). Grahn & Brett (2007) described this concept as follows: "The beat is somehow conveyed solely by the temporal properties of the rhythm itself. It is still unclear, however, exactly what temporal properties are critical for beat perception to spontaneously occur" (p. 893).

Successful imputation of structures from probabilistic or ambiguous acoustic cues implies a role for knowledge derived from experience and learning. Both language and music attest examples of statistical learning in childhood which adults cannot match. Experiments have shown that adults fail to accurately discriminate and/or learn to reproduce certain linguistic or musical distinctions, motifs, and/or representations (Hannon & Trainor, 2007; Hannon & Trehub, 2005; MacKain et al., 1981). For instance, Bulgarian, Macedonian, and Turkish folk rhythms require exposure during early childhood to predict and reproduce them, e.g., (e.g., Hannon et al., 2012; Hannon & Trainor, 2007; Hannon & Trehub, 2005). Similarly, in language, anticipating metrical structures derives from language-specific knowledge of lexical items for the language, as well as rules for lexical and/or phrasal prominence, which are typically required in early childhood (Chen, 2018; Saffran, 2020). Anticipating musical or linguistic structures, including those applicable to understanding timing and meter, thus depends on application of specialized knowledge of these forms to determine the extent to which sensory evidence from timing and other acoustic cues are consistent with alternative culture-specific contrasts or motifs. In short, generating candidate top-down temporal and metrical structures for language or music often necessitates imputations or

inferences, and in some cases requires specialized culture-specific perceptual knowledge may be acquired only during childhood.

Returning to the topic of metrical hierarchies, the notion of "meter" and beat" in both language and music bear further scrutiny. An earlier section discussed perceptual consequences of acoustic onsets – amplitude rises in both language and music, including how neurons in the brain respond to them (Ahissar et al., 2001; Luo & Poeppel, 2007; Oganian & Chang, 2019). In both language and music are associated with perceptual "moments of occurrence" of auditory events, e.g., syllables or notes (Cutting & Rosner, 1974; Ding et al., 2017; Gordon et al., 2015; Ladányi et al., 2020; Oganian & Chang, 2019). In language phonology, rhyme onsets correspond to perceptual moments of occurrence known as "p-centers" (Morton et al., 1976; Oganian & Chang, 2019; Zoefel et al., 2018). Consistent with this, vowels of syllables have been proposed to be primary organizing units, serving as anchors for establishing a syllable representation (Hoequist, 1983; Marcus, 1981) and expectations about their upcoming temporal occurrence in language often generates significant psycholinguistic planning, even while consonants that elapsed even earlier in time have zero or minimal effects on planning or psycholinguistic prediction (Galle et al., 2019; Schreiber & McMurray, 2019). Amplitude envelope decreases or *offsets*, by contrast, are considerably less important in perception of both language and music (Kato et al., 2003; Patel, 2010). In the stimuli of the present experiment, onsets were marked by steep amplitude envelope rises, consistent with clear impressions of beats. Sensitivity to amplitude rise-time in spoken language is associated with enhanced linguistic skills (Gordon et al., 2015; Leong & Goswami, 2014). Vowel onsets constitute perceptually salient landmarks with developmental significance for language acquisition cross-linguistically; infants show an early perceptual vowel-related

bias/advantage which begins before birth and persists for several months before switching to become a "consonant-bias" (Nazzi & Cutler, 2019).

It is thus clear that there are compelling analogues between the notions of "beat" in music and "stress" or "prominence" in language (Lehiste, 1970). Both represent an affordance for certain actions in production. Perception of the beat often causes spontaneous synchronized movement, such as toe tap- ping or head nodding; the presence of a beat also affects the ability to remember and perform a rhythm (Grahn & Brett, 2007). Likewise, abstract lexical stress affords certain kinds of acoustic variation during speech production, such as intoning with a H* or L* pitch accent (Ladd, 2008; Pierrehumbert, 1980) as well as coordinating actions such as clapping or tapping with speech signals (Small, 2015).

To summarize, numerous accounts support structural similarities in language and music. Importantly, for both language and music, inference processes which depend on knowledge and (early) experience influence the building of representations of metrical structure from the signal; the relative prominence patterns or metrical structure associated with discrete events like syllables or notes cannot simply be "read off" of local correlates like amplitude (cf. loudness) or duration (Grahn & Brett, 2007; Hannon et al., 2004; Kochanski et al., 2005; Lehiste, 1970; Povel & Essens, 1985; Vuust et al., 2018).

4.3.4 Predictive coding and building metrical, prosodic, and phonological structures

The prior section highlighted that determining metrical and phonological structure in language depends on a complicated inference process that takes into account linguistic knowledge over the lifetime, together with a complex consideration of acoustic structures, realizations, and

dependencies. This idea of inferences about prosodic structure is further supported by numerous experimental papers by Dilley and colleagues on the *distal rhythm effect* and the *distal rate effect*. In the distal rhythm effect, resynthesis techniques are used to alter the prosody on initial parts of experimental utterances, e.g., by changing the pattern of F0 and/or duration, to give rise to alternatively strong-weak-strong or weak-strong-weak metrical expectations across phonetically identical context syllable sequences. Alternative prosodic contexts then generate prosodic expectations that cause acoustically identical upcoming material to be heard as consisting of different structures - e.g., *timer derby* vs. *tie murder bee* – including different lexical composition, different segmentation, and different lexical and phrasal stress – with large effect sizes (Breen et al., 2014; Brown et al., 2011; Brown et al., 2015; Dilley et al., 2010; Dilley & McAuley, 2008; McQueen & Dilley, 2021; Morrill, Dilley, & McAuley, 2014; Morrill, Dilley, McAuley, et al., 2014; Morrill, McAuley, et al., 2015). The distal rhythm effect has been demonstrated using both wordlists and continuous sentences, and using a variety of paradigms (free responses, speeded reaction time tasks, ERP/EEG, and eye-tracking). Moreover, Dilley and colleagues' distal rate effect stems from the observation that in continuous casual speech, not all syllable rhymes are marked by an amplitude onset or any spectrotemporal discontinuity (Dilley & Pitt, 2010; Shockey, 2008). The distal rate effect involves experimental demonstrations that perceiving heavily coarticulated short words or syllables from naturally coarticulated speech requires generation of linguistic-prosodic expectations from temporal cues in context speech; that is, listeners hear, or fail to hear, a separate heavily-coarticulated syllable or word solely based on manipulations of resynthesized speech rate of contexts (Baese-Berk et al., 2019; Baese-Berk et al., 2014; Dilley & Pitt, 2010). Experiments have shown that the intelligibility of context temporal signals is crucial, such that prosodic expectations "ride on top of" some linguistic structure scaffolding (Morrill,

Baese-Berk, et al., 2015; Pitt et al., 2016). This was shown most clearly in experiments in which noise-vocoded or sine-wave speech rendered intelligible via the "pop-out effect" (Davis et al., 2005) or via training enabled perceptual recovery of coarticulated words, while matched unintelligible signals did not (Pitt et al., 2016). Interestingly, across both the distal rhythm effect and the distal rate effect, gradient acoustic details of prosodic contexts provide gradient quantitative support for alternative upcoming linguistic-prosodic structures (Baese-Berk et al., 2014; Dilley & McAuley, 2008; Heffner et al., 2013; Morrill, Dilley, & McAuley, 2014). In summary, numerous psycholinguistic studies support the assertion that generation of alternative candidate metrical hierarchical structures is part of top-down linguistic structure-building regarding upcoming syllable durations and metrical stress relations.

These distal rate and rhythm effects highlight fundamentally intertwined functions of syllables in lexical perception and production. On the one hand, syllables have a *grouping* function of organizing speech segments into syllabified units, whether in the lexicon or dynamically at output (Levelt et al., 1999). On the other hand, syllables stand in a relative prominence relationship with respect to one another. Further, syllables are the basis for organization of lexical items, supporting critical functions of syllables as units for lexical search and retrieval, along with construction and coordination of broader prosodic structures (Brown et al., in press; McQueen & Dilley, 2021).

Building on observations of these multiple organizing functions of syllables, Brown, Tanenhaus, and Dilley (in press) recently proposed the *Syllable Inference account*, which proposes that the syllable is a pivotal unit that creates layers in lexical and prosodic structure building. The Syllable Inference account proposes that as part of language processing, predictive generation of candidate metrical and/or temporal prosodic structures from ongoing linguistic representations.

Candidates must provide for internal simulations (on the basis of "prosodified" word candidates) that are good acoustic matches in time and frequency, between imputed candidate lexical structures and (sensory indices of the) spectrotemporal "landmarks" in the speech signal, especially the vowel/rhyme onsets (Kato et al., 2003; Luo & Poeppel, 2007; Morton et al., 1976; Oganian & Chang, 2019; Peelle & Davis, 2012; Zoefel et al., 2018). Specifically, the Syllable Inference account proposes that top-down metrical/linguistic structure generation entails all of the following: (1) a good overall acoustic (cf. frequency-time) match between high-ranked alternative lexical structures and (sensory indices of the) linguistic signal; (2) an "explanation" for prosodic characteristics of (imputed) syllable rhymes, e.g., the timing of elapsed vowel onsets, where the "best match" will have minimized the error between top-down, internal representations of prosodified word strings and bottom-up (sensory experiences of) acoustic-prosodic cues (e.g., timing of rhyme onsets); and (3) affordances of predictions about timing and metrical structures to be contained in the next moments of an unfolding utterance by the next word(s).

While Brown and colleagues (in press) focused on perception, the core ideas outlined in predictive coding approaches extend to production, because metrical structure building has consequences for both linguistic perception and production processes. In *perception*, candidate words with lexical or phrasal stress patterns that do not fit the prosodic expectations will be down-weighted or discounted (Brown et al., in press). In *production*, top-down representations of lexical items selected based on perceptual recapitulation of past words and current accumulated evidence (cf. Anders et al., 2015; Dell & Chang, 2014) will be subject to articulation in a manner that conforms to prosodic context-derived expectations. This may affect computations of the dynamic perceptuomotor changes that must be imposed on newly-selected lexical items that will instantiate comportment with contextually-derived prosodic expectations by affecting: syllable-level rate and

timing (cf. Dilley & Pitt, 2010); fundamental frequency (F0) characteristics of stressed syllables, e.g., whether the pitch on a prominence should be higher or lower than preceding syllables (cf. Dilley & McAuley, 2008); and the timing of stressed syllables, drawing on the apparently universal metrical constraint that lexical stresses must be sufficiently separated (Nespor & Vogel, 2007). Brown and colleagues (in press) further proposed that entrainment of populations of delta- and theta-band neurons to context rhythm and rate cues, respectively, comprise mechanisms for predictive prosodic expectation-generation. These sustained neural oscillations from context generate affordances and expectations about prosodic variables (e.g., timing and stress) in upcoming moments of speech that form the basis of quantitative goodness-of-fit calculations for alternative lexico-prosodic models in perceived or produced speech. In summary, the Syllable Inference account extends proposals within a predictive coding framework of speech and language to specify how prosodic structure-building represents an important component affecting the goodness-of-fit of an overall language model which is ultimately derived during perception and/or production.

4.3.5 Neural communication across spatially distributed and specialized brain regions

The above suggests that computation of prosodic structures corresponds to an essential step in determining an overall top-down language model. Given the spatially distributed nature of neural centers for computing prosody versus those for computing morphosyntax, lexical items, and meaning (Hickok & Poeppel, 2004) it is important to consider how an imputed top-down metrical/prosodic structure is intimately linked with linguistic structure built in other brain areas, such as syntax and semantics (Assaneo et al., 2019; Hickok & Poeppel, 2004). This entails

consideration of how spatially distributed structures in the brain are coordinated to give rise to some overall coherent perceptual structure. Relatedly, Fries (2005, 2015) has proposed the *Communication Through Coherence* (CTC) hypothesis, which suggests how spatially disparate regions of the brain communicate with one another through resonance achieved by neural oscillations. Hasson et al. (2018) likewise recently argued that neural oscillations are an essential framework for understanding how language is organized and computed in the brain.

Conceptualization of neurolinguistic structures in communicative sciences has undergone a revolution in recent decades which highlights that the traditional Broca-Wernicke-Geschwind model is now obsolete. Instead, modern conceptualizations take stock of the fact that building linguistic structures entails many cortical and subcortical brain regions, which are connected by a complex networks of white matter tracts. By extension, building a top-down linguistic representation and coordinating it with some prosodic representation entails transmitting information across cortico-cortical connections and/or cortical-subcortical connections. Cortico-subcortical connections involve interfacing with stored syllabaries of phonemes to execute well-practiced articulatory plans (Bohland et al., 2010). Cortico-cortical connections for linguistic structure-building include fronto-temporal parietal connections, such as dorsal and ventral streams, as well as left-right hemisphere connections.

White matter tracts encompass both inter-hemispheric cortico-cortical connections, as well as cortical-subcortical connections (Dick et al., 2014). White matter tracts such as the arcuate fasciculus, once thought to be a language-specific pathway, instead has domain-general functions for communication of auditory information (Blecher et al., 2016). Further, these white matter tracts long neglected due to technical and conceptual limitations, have increasingly been studied in their own right (Chang et al., 2018; Warbrick et al., 2017). Importantly, individual differences in white

matter tract morphology have been shown to be associated with individual differences in degree of accuracy during sensorimotor tasks (i.e., greater timing variability in tapping tasks) (Blecher et al., 2016), foreign language imitation ability (Vaquero et al., 2017), and short- and long-term melody and rhythm learning in non-musicians (Vaquero et al., 2018). A recent systematic review demonstrated substantial evidence of correlations between a white matter morphology index (fractional anisotropy, FA), and BOLD fMRI response, suggesting an important, poorly understood, and oft-neglected structure-function relationship (Warbrick et al., 2017), one which is potentially applicable to elucidating the multifactorial cause of stuttering.

Laterality of language in the brain may reflect specialization of the hemispheres to process and plan speech at different timescales. Notably, it has been suggested that the left hemisphere is specialized for frequency analysis at a fast time-scale (i.e., short temporal windows), whereas the right hemisphere is specialized for frequency analysis at a longer time-scale (e.g., Poeppel, 2003). In view of a predictive coding framework, it can be inferred that linguistic representations of prosodic structures for language, computed presumably in the right hemisphere, must be adduced and temporally coordinated with the unfolding morphosyntactic structures for language computed in the left hemisphere. It is noteworthy that evidence of a right dorsal stream for prosody has now been found (Sammler et al., 2018), in addition to the left dorsal stream described by Hickok and Poeppel (2004) for mapping sound to meaning and articulation. Further, coordination of prosodic structures thought to be primarily right-lateralized with left-lateralized syntactic and semantic structures (Sammler et al., 2018) has been revealed to recent research instead to show variable laterality for prosody – on the left for structural cues vs. on the right for emotion-related cues (Chien et al., 2020; van der Burght et al., 2019). These considerations suggest that the demands for integration and the need for linguistic communication between spatially disparate brain regions

to traverse *particular* white matter tracts will likely be dependent on the *particular* linguistic structures to be built.

4.4 Extending the predictive coding account of metrical structure-building to a hypothesized explanation for stuttering

4.4.1 Brief review of core deficits and hallmarks of stuttering

In this section I build on the Syllable Inference account of top-down metrical/prosodic structure building (Brown et al., in press) to develop a conceptual proposal within a predictive coding framework for why stuttering occurs. I first review some of the core difficulties associated with stuttering as a disorder. Then I link ideas of disordered white matter tracts to certain core deficits found in stuttering and why a predictive coding framework can explain these particular features.

Stuttering has been shown to entail a variety of difficulties, including subtle timing and prosodic deficits (Kent, 1984; Mackay & Macdonald, 1984; Van Riper, 1982; Wieland et al., 2015). A morphological hallmark of stuttering is aberrant network connectivity involving the default mode network and its connectivity with attention, somatomotor, and frontoparietal networks (Chang et al., 2018), as well as abnormal functional connectivity in rhythm networks (Chang et al., 2016; Chang & Zhu, 2013; Etchell et al., 2014), to atypical usage of sensorimotor feedback (Cai et al., 2014; Cai et al., 2012; Daliri et al., 2018; Kalinowski et al., 1993; Kalinowski et al., 1996). A critical insight afforded under this predictive account is to reframe proposals pointing to problems with integration of feedforward and feedback cues (e.g., Bohland et al., 2010;

Civier et al., 2013) as resulting from a failure to faithfully transmit timing or other dynamical information due to white matter abnormalities (Chang et al., 2018). A disruption to the transmission of timing information across abnormal white matter tracts likely involves disruptions to the phase of planned coordinative timing events for syllables (Ross et al., 2018), resulting in poorer-quality and statistically unreliable predictions about timing. This failure to transmit accurate timing information would result in a larger-than-normal error signal upon comparison of feedforward and feedback information and propagate to predictions at different levels of the speech production system.

Although stuttering is typically thought of as being a disorder of *production*, a variety of *perceptual* differences and deficits relative to neurotypical individuals have been identified – for instance, different reactions to altered auditory feedback between these two groups (e.g., Kalinowski et al., 1993; Kalinowski et al., 1996). These results support the profile of stuttering deficits as entailing facets of *both* production and perception. Predictive coding accounts of language hold that speech *production* entails recapitulation and prediction of past *perceptual* experiences (e.g., Dell & Chang, 2014; Pickering & Garrod, 2013). On this view, linguistic structure-building in people who stutter during speaking should involve dynamic generation of alternative candidate top-down linguistic representations (including metrical representations) selected according to accumulated evidence (Anders et al., 2015) and informed by how well these candidates accord with bottom-up sensory evidence (Di Liberto et al., 2018; Hovsepyan et al., 2020). The proposal of Brown and colleagues (in press) suggests that a crucial component of top-down linguistic structure-building is metrical/phonological structure-building consisting of hierarchical embedding of syllables in words, and phonemes within syllables (where syllables are specified as having distinct, relative degrees of lexical and/or phrase-level prominence). The word

boundaries should then generate minima in error signals due to matches of top-down representations with bottom-up sensory evidence (Donhauser & Baillet, 2020).

4.4.2 Predictive coding framework: Conceptual overview

I propose that core deficits in stuttering entail introduction of temporal or other sensorimotor errors (relative to actual external sources) into bottom-up sensorimotor feedback, leading to disruption of evaluation of alternative competing top-down representations (including hierarchical metrical structures) in terms of how well they account for bottom-up sensorimotor stimuli. This predictive coding-inspired account of deficits in stuttering is reminiscent of the widely-held theoretical view that the disorder involves difficulties with the integration of feedback information with feedforward motor or linguistic plans (e.g., Bohland et al., 2010; Civier et al., 2013). Problems with integration of top-down representations and bottom-up sensorimotor feedback are expected to have multiple negative consequences. First, errored propagation of bottom-up information about timing and/or sensorimotor dynamical states should lead to gradiently incorrect system-internal quantitative indices of timing and/or information about actual external sensorimotor states. Further, errors introduced via faulty signal propagation putatively via disordered white-matter tracts are expected to introduce inaccuracies into quantitative evaluations of gradient goodness-of-fit across high-ranked candidate top-down representations. For instance, word boundaries are typically locations of local error minima in neural signals (Brown et al., in press; Donhauser & Baillet, 2020). I propose that in individuals who stutter, difficulties putatively due to errored propagation of information across white matter tracts yields a profile of local and global error minima which does not match those of neurotypicals, where the extent of mismatch

may be correlated with the severity of stuttering. In people who stutter, the local minima in the evaluation of prediction error may be temporally misaligned with respect to top-down representations of typical controls, and temporal profiles of imputed error minima may result in moments of stuttering.

Extending these ideas to language, "beats" are proposed to be (spectro-)temporal positions with respect to an unfolding speech stream where different aspects of linguistic structure must be aligned in time. A top-down language model which accurately predicts the timing and rhythm of the "beats" of speech would be one involving a string of words with appropriate prosodic variations – the string of onsets, some of which are more salient in a hierarchical metrical sense. Enacting this language model as an articulatory/ motor plan would require generating temporally aligned motor actions that dynamically realize the top-down embedded language model – the words imbued with prosodic variations. If error in motor articulation and planning were to propagate through the system, this could lead to a mismatch between temporal predictions associated with top-down hierarchically embedded language models and bottom-up sensory cues/projections. This would mean that the top-down language model would receive less sensory support, causing a temporal "disconnect" between the language plan and the imputed sensorimotor plan. This is the hypothesized core cause of stuttering-like disfluencies within the current predictive coding account.

4.4.3 Stimulus properties and recent modeling: Some further preliminaries

As a preliminary to explaining the present experimental results within the predictive coding framework extensions sketched thus far, it is worth recalling some properties of the stimuli.

Notably, both simple rhythms and complex rhythms entailed a highly irregular sequence of onsets for each item (cf. Table 2); that is, it was not the case that simple rhythms consisted of a regular, acoustically isochronous sequences of onsets. For instance, for one "simple" item consisting of 5-intervals the standard was "31413" while the different was "31431"; at a base IOI of 220 ms, the same-different task for the item translated to a comparison of 660 ms – 220 ms – 880 ms – 220 ms – 660 ms with 660 ms – 220 ms – 880 ms – 660 ms – 220 ms, reflecting the irregularity of onset spacing. According to Grahn and Brett (2007), simple rhythms entail a good match between metrically strong events predicted by top-down metrical representations and moments of acoustic onsets in (sensory indices of) the signal, whereas complex rhythms entail a quantitatively worse match overall. This is because onsets predicted by candidate top-down metrical representations – which are based on default binary structures in language and in music (Halle & Idsardi, 1995; Hayes, 1995; Kotz et al., 2018) failed to be met with acoustic onsets in the speech signal; for complex rhythms, the acoustic onsets tend to co-occur with moments heard as "offbeats" relative to a top-down meter. The regularities in simple rhythms involved a regular implicit grouping into elements of four, while the complex rhythms did not. To make this explicit, with respect to "31413" (simple) the grouping was (31)(4)(13) whereas for "31242" (complex), there is no analogous sequential grouping into summations of quadrenary-element groups possible. This means that there is intrinsically less acoustic evidence to support a given meter in the case of complex rhythms.

As a further preliminary to explaining the neural activation findings from my experiment, I additionally appeal to recent developments in modeling neural circuits (Egger et al., 2020). Egger and colleagues (2020) built on insight gleaned from empirical studies suggesting that the neural basis of sensorimotor coordination may be understood using the language of dynamical systems

(Churchland et al., 2012; Remington, Egger, et al., 2018; Remington, Narain, et al., 2018). Recent work involved development of a basic circuit module (BCM; Wang et al., 2018) that "acts as a flexible open loop controller for producing desired time intervals"(Egger et al., 2020; p. 2). Building on the insight that simple extensions of the BCM could account for dynamic patterns and a range of other timing behaviors, Eggers et al. (2020) extended the BCM model with components that involved a motor planning module (MPM) capable of producing isochronous rhythms. Further, they extended the model to include a sensory anticipation module (SAM) permitting the resulting neural circuit model to accommodate internal noise, prior expectations, and sensorimotor delays.

Egger et al (2020) demonstrated that the resultant neural circuit model consisting of the BCM, MPM, and SAM is capable of capturing key features of human behavior in a number of classic timing tasks (Di Luca & Rhodes, 2016; McAuley & Jones, 2003; Shi & Burr, 2016). They relate the work to alternative classes of circuit models of timing, each with its own strengths and weaknesses. For instance, one class is based on the accumulation of clicks of a central clock (Killeen & Fetterman, 1988; Meck, 1996; Simen et al., 2011; Treisman, 1963). Noting that these models rely on an assumption on ramping activity observed in individual neurons, they note that a weakness of the models is to fail to explain how the recurrent circuit interactions lead to such ramping activity. The second class of models is based on recurrent neural circuits that produce rich dynamics ad capable of producing activity patterns similar to those observed in actual neurons (Buonomano, 2003; DePasquale et al., 2018; Karmarkar & Buonomano, 2007; Remington, Narain, et al., 2018; Wang et al., 2018); however, a weakness is the inability to flexibly integrate sensory and motor feedback. The third class of models uses coupled oscillatory units (Large et al., 2015; Large & Jones, 1999; Large & Kolen, 1994; Miall, 1989; Todd et al., 2002). Egger et al (2020)

note that the modes can produce nuanced timing behaviors and integrate sensory information across timescales, but reflect a weakness in that the activity profile of neurons in brain regions causally involved in timing is typically not oscillatory (Wang et al., 2018).

Egger et al (2020) argue that their neural circuit model provides an understanding of the link between these model classes by explaining the ramping activity in terms of recurrent dynamics, the ability flexibly incorporate sensory and motor feedback, and the generation of oscillations at output. Further, Egger et al (2020) propose that the understanding of sensorimotor control afforded by their model has critical implications for potential functionality in basal ganglia, such that inhibitory pathways "may be the substrate for implementing the mutual inhibitory interactions needed for the temporal control of movements" (p. 11). I hypothesize that the insights afforded under this new modeling work provide a crucial missing link to an understanding of predictive auditory-motor interactions as per Patel and Iversen (2014), along with facets of human timing behavior and feedforward-feedback integration (Bohland et al., 2010). Development of the proposed predictive coding account, particularly in terms of quantitative modeling, is left for future work.

4.4.4 A conceptual predictive coding account of behavioral results

Having sketched these ideas, I now turn to the explanation of my findings within the proposed predictive coding framework. It is proposed that shared resources are drawn upon for metrical structure-building for both language and non-linguistic stimuli (e.g., music or tone sequences) (Brown et al., in press; Dilley & Breen, in press; Lerdahl & Jackendoff, 1983). In both domains, the best-fit top-down metrical structures – an important component of language and

music – are imputed by matching candidate top-down representations against bottom-up cues, particularly the timing of onsets (Oganian & Chang, 2019). The claim is that disordered functional connectivity in the rhythm networks of people who stutter (Chang et al., 2016; Chang & Zhu, 2013) putatively due to abnormal white matter tracts (Chang et al., 2018) induce inaccuracies into internal representations of timing, introducing failures of candidate top-down representations to minimize prediction error to the same extent in people who stutter, compared with people who do not stutter.

The behavioral results match this predictive coding framework for both groups. The discrimination of complex rhythms was inherently poorer than for simple rhythms. This can be explained by an intrinsically lesser degree of statistical support in the sequences of onsets for the binary-based groupings important for language and music (Halle & Idsardi, 1995; Kotz et al., 2018; Patel, 2010). By contrast, a complex rhythm will yield differentially less signal-based support for a top-down representation, such that it will be less likely to achieve perceptual salience, to cross a threshold of activation and reach conscious awareness in perception, and/or a to leave a salient imprint on memory. For people who stutter, reduced accuracy of estimates of bottom-up sensorimotor information is posited to lead to temporal asynchronies between bottom-up signal-based support and the respective top-down representations, relative to those adduced by neurotypical individuals. Under similar logic as for the difference between simple and complex rhythms, it would be expected that there would be a weaker "memory trace" for a recently heard complex rhythm in the brains of people who stutter, compared to neurotypicals, or perhaps no memory consolidation at all. This is consistent with the pattern of behavioral results.

4.4.5 A conceptual predictive coding account of neural results

Functional MRI was used to examine the neural response to a same/different discrimination of the two different rhythm types. Prior studies have identified a rhythm perception and timing network including the putamen (area of basal ganglia), SMA, and premotor area regions (e.g., Grahn & Brett, 2007; Grahn & Rowe, 2009) The SMA and putamen form the 'main core timing network' (Merchant et al., 2013). The basal ganglia are more active during the performance or tracking of simple rhythms, i.e., those that are easier to internalize, compared to complex rhythms (Geiser et al., 2012; Grahn & Rowe, 2009, 2013).

The results for the group of adults who do not stutter showed significant activity in the bilateral insula, bilateral STG, bilateral SMA, and bilateral premotor area, for both simple and complex rhythms. This is consistent with Grahn and Rowe (2009), who showed similar activation patterns in neurotypical individuals with variable amounts of musical training. Further, in the present study, when the simple and complex rhythms were contrasted, the left STG and left putamen were more active for simple than for complex rhythms. Both sets of results are therefore consistent with prior evidence by Grahn and colleagues (2007; 2009) as well as the areas outlined for the ASAP hypothesis (Patel & Iversen, 2014).

The novel extension of a predictive coding framework proposed here recast the results of Grahn and colleagues (2007; 2009) and other work. Specifically, it appears that in cases where the generative process for rhythms can readily be inferred, as for the simple rhythms, that the left STG and left putamen are more active than when the generative process for the rhythm cannot be readily inferred, as for the complex rhythms. The generative model would be expected to have been harder to derive and/or the evidence for a model was more equivocal for complex than for simple – less

evidence and hence weaker neural connections instantiating the complex than the simple. Deriving the generative model of the metrical structure relied more on left hemisphere structures for onset-detection (left STG) and meter (left putamen). In other words, there was more focal activity on the left when the generative model of the meter was easier to construct and/or could be constructed on the basis of statistically more evidence in the signal ("simple") compared to the case of when the generative model of meter was harder to construct and/or its construction relied on less evidence statistically compatible with it in the signal ("complex").

Prior studies with people who stutter have shown activation in the rhythm perception and timing network, which includes the putamen (area of basal ganglia), SMA, and premotor area regions (Chang et al., 2016; Chang & Zhu, 2013; Civier et al., 2013; De Nil et al., 2003; Giraud et al., 2008; Kell et al., 2009; Kronfeld-Duenias et al., 2016; Toyomura et al., 2011). Additionally imaging research with children who stutter (Chang et al., 2016; Chang & Zhu, 2013)have found significantly decreased functional connectivity between the SMA and putamen (the 'main core timing network'; Merchant et al., 2013).

Several noteworthy fMRI findings were observed for adults who stutter. First, a greater extent of overall activity was observed encompassing the rhythm processing network in people who stutter compared with those that do not. Regarding the specifics, multiple areas of the rhythm network were observed to be activated in adults who stutter, reflecting many similarities in activation of the rhythm network as compared with adults who do not stutter. During the simple rhythms there was significant activity in the bilateral insula, bilateral STG, bilateral SMA, bilateral premotor area, and – *unlike* control participants - bilateral putamen, and bilateral IFG. During complex rhythms there was significant activity in the same regions as in the simple rhythms (i.e., bilateral insula, bilateral STG, bilateral SMA, bilateral premotor area), as well as bilateral putamen

and bilateral IFG. The comparison of the active areas during simple and complex showed different activation for people who stutter compared with those that do not; namely the bilateral putamen and bilateral IFG. This suggests that one or both of these regions could have been functioning in a compensatory capacity to increase rhythm discrimination performance in people who stutter, a topic considered further in later sections.

A further finding for adults who stutter was that when the simple and complex rhythms were contrasted, the right insula and right putamen were more active for simple than for complex rhythms. Interpreting these results within the proposed predictive coding framework in cases where the generative process for rhythms could more readily be inferred – that is, for the simple rhythms – the right insula and right putamen are focally active in generative structure-building, whereas for the case when the generative process for rhythm cannot be readily inferred – that is, for complex rhythms - there was more diffuse activation. The predictive coding framework further affords an interpretation of the comparison with control participants; namely, the differential patterns of activation in simple vs. complex rhythms in adults who stutter compared with controls, suggests that activation in the right insula and right putamen act in a compensatory fashion in people who stutter to carry out structure-building.

One of the most robust neural signatures of stuttering is an excessive recruitment of the right frontal cortical areas while speaking, but it is difficult to gain a clear understanding of the mechanisms that may be causal (i.e. associated with the risk of developing stuttering) or maladaptive (i.e., a result of life-long stuttering) (Neef et al., 2018). For this reason, the right frontal regions in people who stutter show signs of both compensatory and causal or maladaptive activity. Some researchers have found the over-activation in the right IFG to be adaptive by showing negative correlations with stuttering severity (e.g., Kell et al., 2009; Preibisch et al.,

2003), and that right hemisphere activation was stronger in fluent than stuttered speech (Braun et al., 1997; Fox et al., 1996). However, other researchers suggest that over-activation in the right hemisphere is maladaptive, such as finding positive correlations between overactivation of the right hemisphere and stuttering rate (Fox et al., 2000), and that stuttering therapy can reduce the activation of the right hemisphere (De Nil et al., 2003).

In a study by Neef and collagues (2016), they found that the right IFG activity was related to the inhibition of speech responses, which led to a theory that the cause of stuttering may be due to an overly active global response suppression mechanism mediated though a pathway of subthalamic nucleus-right IFG-basal ganglia. Neef and colleagues (2018) then investigated whether the right IFG was compensatory or maladaptive in a follow-up study combining fMRI and diffusion tensor imaging (DTI). Their results showed that stuttering severity is linked to the strength of white matter connections of hyperactive right frontal brain regions (i.e., right posterior IFG, SMA, and pre-SMA), which suggests a possible maladaptive role of these tracts in adults who stutter. Neef (2018) said:

> "Overall, our investigation suggests that right fronto-temporal networks play a compensatory role as a fluency enhancing mechanism. In contrast, the increased connection strength within subcortical-cortical pathways may be implied in an overly active global response suppression mechanism in stuttering. Altogether, this combined functional MRI–diffusion tensor imaging study disentangles different networks involved in the neuronal underpinnings of the speech motor deficit in persistent developmental stuttering" (p. 191).

Multiple studies have found anomalous activation of the right IFG for people who stutter during a variety of speech tasks (e.g., Brown et al., 2005; Fox et al., 1996; Sowman et al., 2012),

and when examining the effect of external auditory pacing on speech in people who stutter, greater activation in the right IFG was found during choral speaking and talking in time with a metronome (Toyomura et al., 2011). Additionally, Kell and colleagues (2009) associated the left IFG with processing temporal and sensorimotor feedback, and suggested the right IFG may perform a similar function. The right IFG has been suggested as part of a 'core timing network' that may become more active when a task is more demanding (Kung et al., 2013; Wiener et al., 2010).

4.4.6 Summary

In summary, the results from the present experiment can be accommodated within a predictive coding framework for deficits found in stuttering. Several significant findings were demonstrated in the results. First, both groups showed activation in core rhythm networks during the rhythm discrimination task for both types of rhythms – namely, bilateral insula, bilateral STG, bilateral SMA, and bilateral premotor area, a finding consistent with, and expected from, prior work identifying those areas as the core neural rhythm network (Grahn & Brett, 2007; Grahn & Rowe, 2009). In contrast to controls, however, adults who stutter additionally showed activation in the bilateral putamen and bilateral IFG - areas which did not show significant activation for controls, suggesting one or both of these areas may be compensatory. Further, distinctive patterns of relative activation were found during simple vs. complex rhythm trials; whereas control participants relied differentially more on left STG and left putamen when processing simple vs. complex stimuli, adults who stutter relied differentially more on right insula and right putamen. These findings suggest that when top-down metrical induction is easier – which is expected when there is statistically greater bottom-up signal support for a metrical structure, as in the case of

simple rhythms compared with complex ones – that neurotypical individuals carry out focal induction through coordination of left-lateralized structures (left STG and left putamen); in contrast, adults who stutter show compensatory reliance on right-lateralized structures (right insula and right putamen) to carry out a similar metrical induction task.

4.6 Study limitations

Many strengths existed in this study, but some limitations were also present. The study design assumed that the population of people who stutter and people who do not stutter were similar within the groups and across the groups in all factors other than the disorder of stuttering for people who stutter. Although this study has similar participant group sizes to other published studies (e.g., Grahn & Brett, 2007; Grahn & Brett, 2009), a post-hoc power analysis conducted using G*Power (Software 3.1; Faul et al., 2009) found the power of the study to be small and resulted in the necessary use of more highly powered statistical analyses to find significant results. Increasing the sample size in this study would have increased the power and would also have allowed larger sub-groups of stuttering severity and resulted in more insight and solid evidence-based claims into the possible connection between severity and rhythm perception deficits. Although including larger sample sizes per group would have preferred, the participants included in this study were the result of over two years of active recruitment within lower Michigan to find people who stutter that fit the exclusion criteria and combining monetary funding from multiple resources to afford the cost of the fMRI scanner time. This means the power of the findings in this study need to be carefully considered within the context of converging evidence for a rhythm perception deficit in people who stutter and that they deserve ample consideration.

Since the population of people who stutter is so heterogeneous, trying to limit the variation on more factors may have reduced additional noise within the data. Stuttering is more prominent in males than females. With a male-to-female ratio in adults of approximately 3 or 4 to 1 (Yairi & Ambrose, 2013), including only men in this study may have eliminated possible variation inherit in the fact that girls recover more than boys, which may have something to do with sex-differences in the brain. Additionally, including a smaller age range might deal with brain differences related to the number of years utilizing compensatory mechanisms to deal with stuttering disfluencies. This could also be related to controlling the type and amount of speech therapy that participants received which would also be related to the age of the participant. It could also be more beneficial to only include participants that with moderate to severe classifications to possibly enhance the neural differences between the two groups (though it should be noted that people who stutter are notorious difficult to classify based on one speech sample since the severity of stuttering can vary by day and situation). As with most studies that have a heterogeneous sample, including a larger n would have additionally minimized the effect of the individual variation noted here (not accounted for within the mixed-effects models).

Another consideration in the interpretation of my results is to investigate the possible role of working memory in the performance of the participants during this task. The analysis of Operation Span (OSPAN) to test for working memory showed no significant differences were found between the groups ($p = 0.143$), and no significant changes were shown in the results of the ANOVA when OSPAN scores were used as a covariate in an ANCOVA (see section 3.1.1). Although the OSPAN is standard for testing working memory, additional measures of working memory or attention could have been collected to support the determination that the groups did not differ on these measures. Additionally, due to the smaller sample size and large variation in

the scores, the OSPAN analysis to compare the two groups would be underpowered and differences simply may not have been detected. Based on existing theoretical frameworks for stuttering, it is unclear exactly how inconsistently-detected group differences in working memory, such as these OSPAN data, might relate to group differences in metrical processing abilities as implicated in the rhythm discrimination task. In Section 4.8 I return to this topic to propose future directions related to theory development involving integrating views of working memory, attention, and rhythmic processing as implicated in the phonological loop.

Lastly, although the two groups were not significantly different on years of musical training (collected via a survey), research including only participants with very minimal musical training would have been more optimal to reduce any additional noise in the data caused by musical training modifying structural or neural difference during this task. Cameron and Grahn (2014) found that while listening to rhythms musically trained and untrained participants showed similar connectivity between subcortical and cortical regions as well as patterns of activity in the dorsal pre-motor cortex, supplementary motor area, inferior parietal lobule, and cerebellum. However, they additionally found that when a beat is induced by the temporal structure of a rhythm, musicians had a greater increase than non-musicians in coupling between the auditory cortex and the SMA as well as greater activity in frontal regions and the cerebellum depending on the complexity of the rhythms.

4.7 Clinical implications

In terms of clinical implications, this study provides converging evidence of a neural and behavioral deficit in the rhythm processing network of people who stutter compared to controls

during a rhythm discrimination task. This study was designed to build on previous research of speech and non-speech temporal prediction errors that can be viewed within a larger predictive coding framework. This framework allows a bridge between research in the field of linguistics and that of music to allow a novel view of deficits in people who stutter.

Although effect sizes in the behavioral task between groups were small, striking group-related differences and substantial effect sizes were observed in the neural data. These differences in both behavioral and neural data in the present experiment were predicted from examination of convergent evidence across studies – including work with a related population for which the neural timing networks are disordered – namely, Parkinson's patients – together with decades of careful behavioral and neuroscientific evidence. Given arguments in support of the view that the evidence can be accommodated within extensions of emergent predictive coding theories, the current basic scientific study can be argued to involve potentially transformative findings. However, there are no immediate clinical benefits to the participants or translational venues, which contextualizes and circumscribes the clinical implications of this study.

Several other points suggest potential basic scientific advances emerging from the present study of a clinical population. First, this study found preliminary evidence that people with a higher stuttering severity rating also have worse rhythm discrimination abilities. These findings are a first step toward further evaluation of a possible correlation between poorer timing accuracy in transmitting bottom-up sensorimotor feedback to be matched against top-down predictions and reduced perceptual salience in memory for the top-down metrical beat-based organization. The low number of participants with moderate-severe levels of stuttering was low, such that the statistically significant univariate effect of stuttering severity was not matched with statistical significance across post-hoc pairwise group differences across the multiple severity sub-groups. It

seems plausible that a Type II error in the analyses might have obtained, a possibility which warrants further investigation, especially given the overall sketch provided in this Discussion of how an extension of an important emerging, empirically well-supported branch of neuroscience – that is, predictive coding – together with the intrinsically inter-disciplinary research approach taken here, presents the appearance of a theoretically consolidated across a wide range of findings related to stuttering.

One of the most compelling ideas to come from the theoretical consolidation of findings alluded to above arguably may be the potential for insight into one of the most puzzling aspects of clinical effects known for stuttering populations: the reliability of certain fluency-inducing conditions to reduce stuttering. The field of stuttering research has known for a long time that fluency-inducing conditions such as metronome therapy are very effective (Azrin et al., 1968; Greenberg, 1970), but there have been few plausible explanations for the effectiveness of these conditions. Predictive coding provides a framework for explaining these effects in light of other recent empirical advances.

I propose that the effects of fluency-enhancing conditions can be explained in outline form with reference to the following facts and appeals to a predictive coding framework: (1) Speech production involves speech perception "in reverse"; that is, preparations for speech production involve cognitive recapitulation of past perceptual experiences and perceived structures, including linguistic stimuli (Dell & Chang, 2014; Pickering & Garrod, 2013). (2) Both speech production and speech perception entail a process of consolidating many good-fitting top-down representations to identify a single "best fit" representation for bottom-up sensory stimuli. (3) A top-down representation of overall linguistic structure entails imputation of a separate best-fit *prosodic* structure that includes factors like syllabic phoneme structures and metrical stress (Brown

et al., in press; Dilley & McAuley, 2008; Dilley & Pitt, 2010; Hovsepyan et al., 2020) (4) Hallmarks of stuttering which include anomalous white matter tracts in stuttering (Chang et al., 2018) and anomalous functional connectivity in neural rhythm networks (Chang et al., 2016) plausibly entail introduction of variable degrees of error into representations of sensorimotor bottom-up signals with respect to external states (Blecher et al., 2016; Chang et al., 2018; Chang et al., 2016; Patel & Iversen, 2014). (5) Computing the production plan (or perceptual representation) for an utterance involves combining candidate lexico-syntactic structures derived from computations in left-lateralized dorsal and ventral streams with little-understood computations for (i) deriving prosody which are generally held to be right-lateralized (Bohland et al., 2010; Hickok & Poeppel, 2004; Sammler et al., 2018) and (ii) retrieving syllabified phoneme sequences that involve cortico-subcortical loops (Dick et al., 2014; Schwartze & Kotz, 2013). (6) Although prosody has often been claimed to be right-lateralized, recent empirical advances have shown that prosodic computations are in fact not consistently right-lateralized, but instead show variable lateralization as a function of the specific prosodic elements in question and the demands of speech constructive tasks (Assaneo et al., 2019; Chien et al., 2020; van der Burght et al., 2019). Such results imply variable manifestations of left-right hemisphere consolidation demands placed on speakers to produce a single cogent overall linguistic representation. This potentially helps explain not only certain correlations between stuttering and prosodic elements (Bergmann, 1986; Brown, 1938; Natke et al., 2002; Prins et al., 1991; Weiner, 1984; Wingate, 2012), but also their inconsistency in realization and variability in the degree to which they produce difficulties for fluency.

Together with points in (1)-(6), a predictive coding framework helps explain the effectiveness of fluency-enhancing conditions for reducing stuttering for the following reasons.

First, predictive coding posits that consolidation to a single overall top-down linguistic representation requires sorting through potentially many competing alternative good-fitting representations (2); this process obtains, whether the immediate communicative goal is production or perception (1). Moreover, determining a best-fitting top-down prosodic structure for the metrical and phrasal phonology is a distinct constructive step that independently and strongly affects the best-fitting *overall* linguistic representation (Dilley & McAuley, 2008; Dilley & Pitt, 2010) (3). This process of reconciling the best-fit prosodic structure with the best-fit lexico-syntactic structure involves both syllabification to retrieved phoneme sequences and internal (re)construction of metrical, phrase-level prosody, which in turn appears to involve coordination of information across potentially many white matter tracts that involve both cortico-cortical connections and cortico-subcortical connections (5). The demands of prosodic best-fit top-down structure building will place variable demands on the right hemisphere (6). Yet, a morphological hallmark of stuttering is anomalous white matter tracts, where these individual differences or anomalies likely involve variable degrees of error in internal representations of sensorimotor information relative to external events (4).

Predictive coding affords the insight that an external signal can independently provide top-down and bottom-up evidence of prosodic structure. Attending to external cues – whether an external stimulus is a regular pacing signal, such as a metronome (Azrin et al., 1968; Greenberg, 1970), or else choral speech (Kiefte & Armson, 2008) – would reduce neural resources devoted to computing a top-down best-fit prosodic structure, which instead is given by the external stimulus. Notably, conditions of equal syllable weight have been shown to change the demands for prosody construction place on the right hemisphere (Assaneo et al., 2019). Likewise, predictive coding suggests that external cues from another person –function as external evidence in support

of a top-down prosodic structure and overall linguistic representation, reducing the demands on the brains of people who stutter to faithfully represent information across faulty white-matter tracts. By providing independent evidence bottom-up in support of top-down representations, the brains of people who stutter may more readily consolidate to a motor action plan toward coordination of a predictive top-down structure and auditory-motor prediction in terms of consolidating a sense of implied moments of beats (i.e., p-centers).

4.8 Future directions

In the future it could be desirable to replicate this fMRI study with children who do and do not stutter to investigate task-based neural activation during a rhythm discrimination task for several reasons. First, although children who stutter have been shown to have a deficit in rhythm network connectivity when correlating rhythm discrimination performance and resting state fMRI data (Chang et al., 2016), utilizing a task-based design would give converging evidence and a way to directly compare the adults and children in these studies. Second, such a comparison would facilitate determining whether the heightened activation seen in adults who stutter compared to adults who do not stutter is similarly found in children who stutter, or whether a lack of greater activation may be the reason for the children's poorer behavioral performance during rhythm discrimination compared to children who do not stutter (Wieland et al., 2015).

Another future direction involves examining additional data that was collected during this study to examine possible converging evidence in support of hypotheses. For instance, task-based fMRI scans could be used to do additional region of interest analyses, seed analyses, or investigations of cerebellar activity which was not analyzed. The fMRI session also entailed

collecting a diffusion tensor imaging (DTI) scan to look at white matter differences, as well as a resting state scan to look at functional connectivity during rest. Additionally, although this experiment was not designed to be difficult enough to have enough incorrect trials to compare to correct trials (i.e., the simple rhythms were around 90% accurate), it may also be interesting to look at the existing data to see if the patterns of activation are the same as or different to the correct trials. Examining these data sets could give additional insight into structural and functional differences between groups and individuals considered important for rhythm processing. Additionally, finger tapping experiments and speech perception experiments were conducted with this population; these tasks could provide additional insight into the extent of correlation between speech and non-speech performance on behavioral tasks for which the neural data were already analyzed.

Additionally, more information could be gathered from looking at cerebellar activation of these two groups. The cerebellum and/or premotor areas have been shown to be responsible for more basic timing processes that are required to encode the time intervals (Grahn & Brett, 2007). Although the basal ganglia are important for beat perception and beat-based timing, the cerebellum has been shown as paramount for the perception of absolute time intervals (i.e., events are not relative to a beat) (Cameron & Grahn, 2014). However, both the cerebellum and the basal ganglia are active when listening to rhythms, which suggests that both types of timing are simultaneously engaged by rhythm processing. Additional analyses of the cerebellum would thus further elucidate neural group differences.

An important future direction relates to the potential for new insights that may arise from integrating theoretical literatures on rhythm processing, attention, working memory, and neural oscillations in the brain. Research has shown that people who stutter do not show differences in

working memory capacity compared to people that do not during simple digit span tasks which measure capacity (Oyoun et al., 2010; Pelczarski & Yaruss, 2016; Sasisekaran & Byrd, 2013; Smith et al., 2012; Spencer & Weber-Fox, 2014). However, when tasks becomes more challenging, differences in working memory and/or attention-related tasks can arise, such as in nonword repetition accuracy, which measures the quality of information held in the phonological loop (Anderson & Wagovich, 2010; Anderson et al., 2006; Byrd et al., 2017; Byrd et al., 2012; Hakim & Ratner, 2004; Pelczarski & Yaruss, 2016; Sasisekaran & Weisberg, 2014; Spencer & Weber-Fox, 2014; but see Dollaghan & Campbell, 1998; Sasisekaran & Byrd, 2013; Smith et al., 2012; 2010).

Recently, Bowers et al. (2018) proposed that anomalous sensorimotor timing has the potential to interrupt fluent speech in people who stutter as well as the properties of phonological working memory. Additionally, the authors suggest that neuroimaging evidence implicates the prefrontal cortex in stuttering, which can be related to neurobiological models that propose the prefrontal cortex and basal ganglia function to facilitate working memory during distracting conditions. The proposal of Bowers et al. (2018) regarding distracting conditions such as nonword repetition and dual-tasks in developmental stuttering holds promise for future work to integrate their proposals with a new model of metrical inference, as informed by these experimental results, within a predictive coding framework. Working memory and attention are implicated in frontal cortex activation, and Alexander and Brown (2018) recently devised the Hierarchical Error Representation (HER) model to provides a reconceptualization of this lateral prefrontal activity as important for anticipating prediction errors from multiple levels of description such as single neurons to all the way to behavior (i.e. a predictive coding framework).

This work was integrated with the novel modeling advances by Egger and colleagues (2020) who recently developed a circuit-level model which coordinated a motor planning module (controlling movement times) with a sensory anticipation module (which anticipates external temporal events). Theoretical integration may be informed by results showing that successful prediction can reduce working memory load and increase speed of perceptual organization of a sequence (Grahn & Rowe, 2009). This work suggests a link between difficulties of adults who stutter with deficits in temporal prediction within a predictive coding framework and new evidence that the bilateral putamen plays a role in meter perception (Li et al., 2019). An important link to be developed in future theoretical work involves the idea that attention is inherently rhythmic (Helfrich et al., 2019) and involves the alignment between the motor system and the timing of events in a task-relevant stream (Morillon & Schroeder, 2015).

Lastly, investigating the role of neural oscillations in the predictive coding framework of the results would provide a complementary picture of core deficits in stuttering relating to rhythm perception. Studies of speech production and perception indicate that multiple brain regions are closely coordinated in time in order for fluent speech to result. These brain regions are spatially distributed across the brain; it has been proposed that neural oscillations are a key mechanism permitting coordination among multiple, spatially-distributed brain regions (Fries, 2005). Neural oscillations represent cyclic changes of excitation and inhibition of neurons, and are usually described according to the speed of their cycle: delta band (5-8 Hz, cf. phrasal prosody, long distance prediction), alpha band (8-12 Hz), beta band (12-30 Hz, cf. conceptual planning), and gamma band (30-100 Hz, cf. phonemes) (Fries, 2005). Given that the periodic patterns of excitation and inhibition of different populations of neurons involve particular time constants, it is thought that neural oscillations may indeed constitute the neurophysiological substrate for

temporal prediction (Morillon & Schroeder, 2015). By extension, such (quasi-) periodic neural oscillations are thought to be responsible for entraining the brain to rhythmic or quasi-rhythmic auditory stimuli (Large & Jones, 1999). Neural oscillations are paramount for predicting rhythmic stimuli, so understanding that speech production engages many of the brain regions utilized in processing rhythm, it stands to reason that neural oscillations should be just as important for the production of fluent speech (Etchell et al., 2014).

4.9 Conclusions

Stuttering is a communicative disorder that involves disruptions to fluent speech, which are characterized by frequent repetition or prolongation of syllables or words, and/or by frequent hesitations or pauses. While stuttering can substantially negatively affect quality of life, the causes are not known. Prior research has identified a number of hallmarks and replicated deficits associated with clinical stuttering, including evidence of generalized timing deficits and reduced functional connectivity in the rhythm network, consisting of brain regions previously identified to be involved in perception of musical meter.

Building on assumptions of (1) shared neurocognitive resources exist for metrical structure-building for perception and production of auditory patterns across domains (e.g., music and language), and (2) functional similarity exists in processes involved in predictive action-preparation from metrical structure in auditory information, it was predicted that a core deficit in stuttering involves deficiencies in integrating candidate metrical structures with sensory evidence that would support them. To test this prediction, an experiment was designed using as stimuli

sequences of five to seven tones in a standard-comparison (i.e., same/different) discrimination task. All inter-onset-intervals in standard and comparison sequences involved an irregular sequence of inter-onset-intervals (IOIs). Critically, half the stimuli provided greater support for the induction of a beat/meter ("simple rhythms"), whereas the other half were matched in interval types but provided less signal-based statistical support for the induction of a beat/meter ("complex rhythms"). Participants were 36 adults meeting clinical criteria for stuttering and controls who did not stutter who were matched on age and sex. Both groups performed the rhythm discrimination task while undergoing fMRI. Rhythm discrimination performance was modeled both as a binomial variable (correct/incorrect) and using d' measures derived from signal detection theory as a function of independent variables of interest (Group: adults who stutter vs. adults who do not stutter; Rhythm Type: simple vs. complex; Trial Type: same/different).

For the behavioral results, statistical analyses using a generalized linear mixed model with random effects for participants and items revealed poorer performance at rhythm discrimination by adults who stutter than matched controls, as revealed by a significant three-way interaction among Group, Trial Type, and Rhythm Type on the likelihood of a correct response, relative to baseline accuracy level of matched controls (β = -0.314, SE = 0.152, z = -2.060, p = 0.039). However, no significant statistical evidence of a rhythm discrimination deficit was found when more traditional statistical measures were used (ANOVA, ANCOVA, p's > .05).

For the neural results, activation in the core rhythm network during the rhythm discrimination task was observed for both groups (bilateral insula, bilateral STG, bilateral SMA, and bilateral premotor area) for both simple and complex rhythms. However, adults who stutter additionally showed activation in the bilateral putamen and bilateral IFG, suggesting that one or both of these areas may perform a compensatory function during rhythm perception and predictive

action-preparation. Further, a within-group contrast comparison of activation across Rhythm Trial conditions, showed significantly greater activation in right insula and right putamen in adults who stutter, but in left STG and left putamen for matched controls.

These results can be interpreted with respect to predictive coding processes in the brain supporting perception, action, and cognition, as well as recent conceptual extensions to auditory processing of music, language, and speech, which propose that linguistic perception and production are "two sides of the same coin." Specifically, it was proposed that listeners attempt to build top-down metrical representations for structured auditory sequences (e.g., speech syllables, musical notes); during language processing, these top-down metrical representations are proposed to be merged with representations of other structures in language to give rise to a coherent overall linguistic representation. It was proposed that a core deficit in stuttering involves deficient processes for integration of top-down metrical/prosodic structure and/or bottom-up sensory indices of dynamic sensorimotor states toward construction of a coherent overall linguistic representation; evidence for this proposal came from findings that: (1) distal context rate and rhythm cues in speech influence metrical/prosodic structures heard across identical acoustic material, thereby influencing goodness-of-fit evaluations of alternative top-down candidate representations of lexico-syntax; (2) a hallmark of stuttering in children and adults is anomalous white matter connectivity and reduced functional organization of rhythm networks in the brain compared with controls.

This was the first study to investigate non-speech rhythm perception in adults who stutter. Results have implications for understanding the effectiveness of certain fluency-enhancing conditions, including provision of external pacing signals (e.g., metronome therapy), as well as delayed and/or pitch-shifted auditory feedback. Findings also suggest new hypotheses regarding

how dynamic connections among brain structures such as the basal ganglia and STG perform computations toward the imputation of timing and meter from acoustically highly variable auditory signals.

APPENDICES

APPENDIX A: Participants who do not stutter table

Count	Age (yrs)	Education (yrs)	Music Training (yrs)	Working Memory (OspanScore)	Handedness	PPVT Score	EVT Score	GFTA Score	% SLD	SSI Total	SSI Severity
1	27.83	14	0	44	100	107	118	101	N/A	N/A	N/A
2	23.42	19	8	31	100	111	106	96	N/A	N/A	N/A
3	34.58	16	7	55	100	123	115	101	N/A	N/A	N/A
4	44.00	16	2	15	80	113	103	101	N/A	N/A	N/A
5	23.33	17	4	37	50	107	97	101	N/A	N/A	N/A
6	22.00	16	0	49	100	91	104	101	N/A	N/A	N/A
7	20.25	14	0	35	80	98	97	101	N/A	N/A	N/A
8	19.33	12	0	49	80	111	96	102	N/A	N/A	N/A
9	21.00	14	2	68	90	109	102	101	N/A	N/A	N/A
10	18.67	13	9	61	70	126	120	102	N/A	N/A	N/A
11	22.33	14	5	42	60	100	106	101	N/A	N/A	N/A
12	22.25	15	7	43	60	100	106	101	N/A	N/A	N/A
13	20.25	13	9	52	70	109	108	101	N/A	N/A	N/A
14	21.83	16	10	39	100	116	116	101	N/A	N/A	N/A
15	20.25	14	9.5	24	100	114	130	101	N/A	N/A	N/A
16	31.5	17	0.5	75	100	101	110	101	N/A	N/A	N/A
17	34.67	18	6	10	80	104	118	101	N/A	N/A	N/A
18	28.42	17	0	48	80	116	94	101	N/A	N/A	N/A

Table 10: Survey, speech, language, and cognitive evaluation information for participants who do not stutter.

APPENDIX B: Participants who do stutter table

Count	Age (yrs)	Education (yrs)	Music Training (yrs)	Working Memory (OspanScore)	Handedness	PPVT Score	EVT Score	GFTA Score	% SLD	SSI Total	SSI Severity
1	25.92	14	8	25	90	106	102	101	4.92	20	Mild
2	26.50	16	4	6	90	110	112	101	5.19	26	Moderate
3	21.83	15	11	37	70	106	110	101	5.53	18	Mild
4	44.75	19	0	37	90	111	134	101	2.41	18	Mild
5	19.75	14	0	31	60	110	108	102	5.83	24	Mild
6	45.50	24	13	62	70	122	140	101	3.95	16	Very mild
7	22.92	16	11	48	90	117	106	101	7.61	25	Moderate
8	21.58	15	7	33	80	104	100	101	9.04	30	Moderate
9	18.33	12	5	43	50	101	110	102	9.51	26	Moderate
10	21.83	15	0	45	80	108	116	101	11.26	35	Severe
11	19.75	14	3	37	90	99	104	102	4.92	19	Mild
12	46.83	14	6	0	70	118	124	101	3.57	13	Very mild
13	21.92	16	0	45	60	111	129	101	4.50	20	Mild
14	53.08	17	0	28	100	111	117	101	5.53	37	Severe
15	22.83	16	7	23	90	132	123	95	3.39	14	Very mild
16	24.67	16	0	6	100	104	116	101	9.92	23	Mild
17	28.50	16	0	61	70	88	106	101	19.63	36	Severe
18	42.25	16	2	54	60	115	115	101	4.93	34	Severe

Table 11: Survey, speech, language, and cognitive evaluation information for participants who do stutter.

APPENDIX C: Advertisement to recruit participants

Title: Do You Stutter? MSU Experiment $10/hr

Do you stutter?

Participate in an fMRI listening study at MSU and earn $10/hr.

E-mail: Liz at msu.snl@gmail.com

You must be:
-A native speaker of English (monolingual)
-Right handed
-Normal hearing
-Not claustrophobic
-Not pregnant
-No history of developmental, psychiatric or other speech disorders (other than stuttering)
-Not be taking medication affecting the central nervous system (e.g., depression, anxiety, ADHD)

This study is investigating differences between stuttering and non-stuttering adults when listening to and tapping rhythms. This study will consist of 2-3 sessions that last between 1-2 hours each.

Compensation: $10/hr (plus mileage compensation if travel is over 50 miles roundtrip)

APPENDIX D: Google form questions to filter participants

- **Basic Information**

What is your name (last, first)?

What is your email address?

What is your phone number?

When is your birthday (day, month, year)?

What is your sex?
Male Female

- **Stuttering Information**

Do you stutter?
Yes No Now recovered Unsure

If you do stutter, what age did you start stuttering?

Do you have a family history of stuttering?
Yes No Unsure

If you have a family history of stuttering, who in your family stutters/stuttered? Make text

- **Filtering Information**

Are you a native speaker of English (monolingual)?
Yes No Unsure

Are you right handed?
Yes No Unsure/Ambidextrous

Do you have normal hearing?
Yes No Unsure

Are you claustrophobic?
Yes No Unsure

If female, are you pregnant?
Yes No Unsure

Do you have any non-removable metal in your body (tooth fillings and removable piercings are ok)?
Yes No Unsure

Do you have a history of developmental, psychiatric or other speech disorders (other than stuttering)?

Yes No Unsure

Are you taking medication affecting the central nervous system (e.g., depression, anxiety, ADHD)?

Yes No Unsure

- **Availability**

What is the FIRST date can you start the study?

What is the LAST date you can start the study?

When are you free on a typical Monday?

When are you free on a typical Tuesday?

When are you free on a typical Wednesday?

When are you free on a typical Thursday?

When are you free on a typical Friday?

When are you free on a typical Saturday?

When are you free on a typical Sunday?

Will you need a visitor's parking pass?

Yes No Unsure

Is there anything else you think the experimenters should know?

APPENDIX E: Hearing screening form

Read to the participant:

"You are going to hear a series of tones, low pitches and high pitches. I want you to raise your hand when you hear a tone, no matter how faint it is. Raise your right hand if the tone was presented in your right ear, and raise your left hand if the tone was presented in your left ear. If you are ready, we'll begin."

- The participant is seated in a chair with his/her back facing you, wearing the audiometer headphones (right ear-red, left ear blue).
- Set the audiometer attenuator dial to **20 dB HL**, and beginning with the **right ear**, present frequencies in the following order: 1000 Hz, 2000 Hz, 4000 Hz, 8000 Hz, 500 Hz, 250 Hz
- Repeat previous step testing the **left ear**.

[Note: If the participant does not respond at a certain frequency, present the tone again at 20 dB. If the participant again does not respond, raise the attenuator by 5 dB each presentation of the same Hz until the tone is heard. Then turn the attenuator back to 20 dB for the next Hz presentation level.]

Hearing Screening (20 dB HL)

Right ear:	Left ear:
1000 Hz _____	1000 Hz _____
2000 Hz _____	2000 Hz _____
4000 Hz _____	4000 Hz _____
8000 Hz _____	8000 Hz _____
500 Hz _____	500 Hz _____
250 Hz _____	250 Hz _____

*Poor performance by a participant *may* suggest hearing difficulties, but may *also* be attributed to additional noise in the testing room. Consult with your research advisor if this seems to be a regular problem with participants.*

Comments:

APPENDIX F: Handedness form

Edinburgh Handedness Inventory

Please indicate your preferences in the use of hands in the following activities by putting + in the appropriate column. Where the preference is so strong that you would never try to use the other hand unless absolutely forced to, put ++. If in any case you are really indifferent put + in both columns.

Some of the activities require both hands. In these cases the part of the task, or object, for which hand preference is wanted is indicated in brackets.

Please try to answer all the questions, and only leave a blank if you have no experience at all of the object or task.

		Left	Right
1.	Writing		
2.	Drawing		
3.	Throwing		
4.	Scissors		
5.	Toothbrush		
6.	Knife (without fork)		
7.	Spoon		
8.	Broom (upper hand)		
9.	Striking Match (match)		
10.	Opening box (lid)		
i	Which foot do you prefer to kick with?		
ii	Which eye do you use when using only one?		

L.Q._____ DECILE_____

APPENDIX G: Participant history form

MEDICAL HISTORY

Have you had any neurological problems (e.g., tremor, seizure)?
☐YES ☐NO

Have you had or are you being treated for any psychiatric illnesses (e.g., anxiety, depression)?
☐YES ☐NO

Do you have other speech-language-hearing related difficulties (e.g., articulation delay, stuttering, auditory processing delay, etc.)?
☐YES ☐NO

Have you been diagnosed with any developmental disorders (e.g., dyslexia, autism, Tourette's syndrome, learning disorder, ADHD)?
☐YES ☐NO

FAMILY HISTORY OF STUTTERING

Do you have any family members that stutter?

Maternal	**Paternal**	**Immediate**
☐Mother	☐Father	☐Brother
☐Grandmother	☐Grandmother	☐Half-Brother
☐Great Grandmother	☐Great Grandmother	☐Sister
☐Aunt	☐Aunt	☐Half-Sister
☐Great Aunt	☐Great Aunt	☐Son
☐Grandfather	☐Grandfather	☐Daughter
☐Great Grandfather	☐Great Grandfather	
☐Uncle	☐Uncle	
☐Great Uncle	☐Great Uncle	
☐Cousin (1, 2)	☐Cousin (1, 2)	
☐Niece	☐Niece	
☐Nephew	☐Nephew	

If family history of stuttering exists:

Family Member:_____ Onset _____ Recovered from stuttering? ☐YES ☐NO
If yes, at what age? _____

Family Member:_____ Onset _____ Recovered from stuttering? ☐YES ☐NO
If yes, at what age? _____

Family Member:_____ Onset _____ Recovered from stuttering? ☐YES ☐NO
If yes, at what age? _____

CHARACTERISTICS OF STUTTERING

What age did you begin to stutter?

- ☐ Before preschool (0-3)
- ☐ Preschool (3 - 5)
- ☐ Grade school (5 - 11)
- ☐ Middle school (11 - 15)

Comments: _____

What types of disfluencies do you notice in your speech?

☐ Part-word Repetitions	(m- my dog is Spot)
☐ Whole-word Repetitions	(my my dog is Spot)
☐ Prolongations	(mmmmy dog is Spot)
☐ Blocks	(…..my dog is Spot)
☐ Phrase Repetitions	(my dog my dog is Spot)
☐ Interjections	(my…uh…dog is Spot)
☐ Revisions	(my dog…my cat is Spot)

Do you have any body movements or unusual tension in your face and/or neck area associated with your stuttering?

_____
_____
_____
_____

Is the severity of your stuttering constant or does it fluctuate in different speaking situations (e.g., public speaking, with friends, etc.)?

_____
_____
_____

Are there any situations that make your stuttering worse/better?_____
_____
_____

What types of reactions do you have when you have difficulty speaking?

☐Anger	☐Frustration
☐Anxiety/Nervousness	☐Helpless
☐Difficulty Breathing	☐Irritation
☐Embarrassment	☐Shame
☐Fear	

☐Other _____
_____

Are there certain words you struggle to produce fluently? _____
_____
_____

Do you have any methods to minimize your stutter (e.g., substituting words, drawing out words, easy onset, pullout, etc.)? _____
_____

How often do you use these methods? _____
_____

FLUENCY THERAPY

Have you ever been seen by any professional to assist with your stuttering?
- ☐ Social Worker
- ☐ Speech-Language Pathologist
- ☐ Physician
- ☐ Psychologist
- ☐ Other Professional(s) _____

Have you ever been formally diagnosed with stuttering?
☐YES ☐NO

Have you participated in speech therapy to treat your stuttering? ☐YES ☐NO
Start of TX: _____
End of TX: _____
Type of TX: ☐Traditional ☐Program(s) _____
Freq of TX: _____
Setting: ☐School ☐Private ☐Other _____
Size: ☐Individual ☐Group ☐Other _____

Are you currently in TX? ☐YES ☐NO
Start of TX: _____
End of TX: _____
Type of TX: ☐Traditional ☐Program(s) _____
Freq of TX: _____
Setting: ☐School ☐Private ☐Other _____
Size: ☐Individual ☐Group ☐Other _____

APPENDIX H: Participant background form

Background

Age: _____

Sex: Male Female

Handedness: Right Left Ambidextrous

Ethnicity:

___ Hispanic or Latino ___ Not Hispanic or Latino

Race (select one or more):

___ Hispanic or Latino
___ American Indian or Alaska Native
___ Native Hawaiian or other Pacific Islander
___ Black or African American
___ Other
___ Asian
___ White

Is English your first language? Yes No

If no, at what age did you begin to be intensively exposed learn English? Please describe your type of exposure (e.g. home, school, etc.).

_____

_____

Have you studied any foreign languages?

If yes, please list all languages and years studied (1 semester = .5 years).

_____

_____

How many years of education have you received? _____

(For example, High school graduate = 12 years; 2 years of college =14 years)

What degrees have you earned (if any)?

_____

What special training have you received other than high school or college (e.g., Voc-ed classes, occupational training, etc.):

_____

What is your present occupation (title & employer)?

_____

What were your previous occupations (including military)?_____

Do you have any hearing impairments? Yes No

If yes, please describe_____

How many hours per week do you spend listening to music? _____

What types of music do you listen to?

_____

Do you have any formal musical training (either instrument or voice)? Yes No

If yes, how many years? _____

What instruments?

_____

What type of training did you receive?

___ School/Band ___ Friends/Family

___ Private Lessons ___ Self Taught

___ Religious ___ Other (please describe)

_____

Are you currently studying and/or performing music? Yes No

If yes, how many hours a week do you practice and/or perform? _____

Do you have absolute pitch? Yes No Don't Know

Do you have any formal dance training? Yes No

If yes, how many years? _____

What styles?

_____

What type of training did you receive?

___ School/Band ___ Friends/Family

___ Private Lessons ___ Self Taught

___ Religious ___ Other (please describe)

_____

Are you currently studying and/or performing dance? Yes No

If yes, how many hours a week do you practice and/or perform? _____

Do you play any rhythm or music-based video games (e.g. Guitar Hero, Rock Band)? Yes No

If yes, how many hours per week do you play? _____

Have you taken any prescribed or over-the-counter medications in the past 24 hours? Yes No

Please list any medications (prescribed or over-the-counter) taken before participating in the study today.

_____

APPENDIX I: Participant strategies form

Please answer the following two questions concerning your impressions of the study.

1. Did you use any particular strategies during the experiment that made it easier to perform the tasks asked of you? If so, please describe any strategies used.

2. What do you think the purpose of this study was?

Please circle the most appropriate response for the following questions.

3. How would you rate your understanding of what you were asked to do? (Circle one.)

1	2	3	4	5	6
I did not understand at all					I understood exactly what to do

4. How would you rate your effort during the study? (Circle one).

1	2	3	4	5	6
I did not try at all					I tried my best

5. How would you rate your attention during the **beginning** study? (Circle one).

1	2	3	4	5	6
I did not pay attention					I paid full attention

6. How would you rate your attention during the **middle** study? (Circle one).

1	2	3	4	5	6
I did not pay attention					I paid full attention

7. How would you rate your attention during the **end** study? (Circle one).

1	2	3	4	5	6
I did not pay attention					I paid full attention

8. How would you rate the difficulty of what you were asked to do? (Circle one).

1	2	3	4	5	6
Not difficult at all					Very difficult

9. How would you rate your comfort during the study (if not comfortable, please describe below)?

1	2	3	4	5	6

Not comfortable at all					Very comfortable

Why were you uncomfortable?

10. How would you rate your musical ability? (Circle one)

1	2	3	4	5	6
Very poor					Very good

11. How would you rate your rhythmic ability? (Circle one)

1	2	3	4	5	6
Very poor					Very good

12. How would you rate your singing ability? (Circle one)

1	2	3	4	5	6
Very poor					Very good

13. How would you rate your dancing ability? (Circle one)

1	2	3	4	5	6
Very poor					Very good

Please rate your disagreement/agreement with each of the following statements.

14. I am good at music.

1	2	3	4	5	6
Strongly disagree					Strongly agree

15. It is important for me to be good at music.

1	2	3	4	5	6
Strongly disagree					Strongly agree

16. My musical ability is important to my identity.

1	2	3	4	5	6
Strongly disagree					Strongly agree

Please answer the following three questions regarding your behavior prior to this session.

17. How much did you sleep last night? Is this typical?

18. How much caffeine did you consume within 6 hours of the scan? Is this typical?

19. How much nicotine did you consume within 6 hours of the scan? Is this typical?

20. Is there anything else you think we should know?

REFERENCES

REFERENCES

Ackermann, H., & Riecker, A. (2004). The contribution of the insula to motor aspects of speech production: a review and a hypothesis. *Brain and Language*, *89*(2), 320-328.

Ahissar, E., Nagarajan, S., Ahissar, M., Protopapas, A., Mahncke, H., & Merzenich, M. M. (2001). Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. *Proceedings of the National Academy of Sciences*, *98*(23), 13367-13372.

Alario, F.-X., Chainay, H., Lehericy, S., & Cohen, L. The role of the supplementary motor area (SMA) in word production. *Brain Research*, *1076*(1), 129-143.

Alexander, W. H., & Brown, J. W. (2018). Frontal cortex function as derived from hierarchical predictive coding. *Scientific Reports*, *8*(1), 1-11.

Allen, G. D. (1973). Segmental timing control in speech production. *Journal of Phonetics*, *1*, 219-237.

Alm, P. A. (2004). Stuttering and the basal ganglia circuits: a critical review of possible relations. *Journal of Communication Disorders*, *37*(4), 325-369.

Ambrose, N. G., & Yairi, E. (1999). Normative disfluency data for early childhood stuttering. *Journal of Speech, Language, and Hearing Research*, *42*(4), 895-909.

Anders, R., Riès, S., Van Maanen, L., & Alario, F.-X. (2015). Evidence accumulation as a model for lexical selection. *Cognitive Psychology*, *82*, 57-73.

Anderson, J. D., & Wagovich, S. A. (2010). Relationships among linguistic processing speed, phonological working memory, and attention in children who stutter. *Journal of Fluency Disorders*, *35*(3), 216-234.

Anderson, J. D., Wagovich, S. A., & Hall, N. E. (2006). Nonword repetition skills in young children who do and do not stutter. *Journal of Fluency Disorders*, *31*(3), 177-199.

Arbisi-Kelm, T. R. (2006). *An intonational analysis of disfluency patterns in stuttering* (Doctoral dissertation, University of California).

Arbisi-Kelm, T. R. (2010). Intonation structure and disfluency detection in stuttering. *Laboratory Phonology*, *10*(4), 405-432.

Archibald, L., & De Nil, L. F. (1999). The relationship between stuttering severity and kinesthetic acuity for jaw movements in adults who stutter. *Journal of Fluency Disorders*, *24*(1), 25-42.

Arenas, R. M. (2017). Conceptualizing and investigating the contextual variability of stuttering: the Speech and Monitoring Interaction (SAMI) framework. *Speech, Language and Hearing*, *20*, 15-28.

Armson, J., & Kiefte, M. (2008). The effect of SpeechEasy on stuttering frequency, speech rate, and speech naturalness. *Journal of Fluency Disorders*, *33*(2), 120-134.

Arnal, L. H. (2012). Predicting "when" using the motor system's beta-band oscillations. *Frontiers in Human Neuroscience*, *6*(225), 225.

Arvaniti, A. (2009). Rhythm, timing and the timing of rhythm. *Phonetica*, *66*(1-2), 46-63.

Assaneo, M. F., Orpella, J., Ripolles, P., Diego-Balaguer, D., & Poeppel, D. (2019). The lateralization of speech-brain coupling is differentially modulated by intrinsic auditory and top-down mechanisms. *Frontiers in Integrative Neuroscience*, *13*, 28.

Au-Yeung, J., Howell, P., & Pilgrim, L. (1998). Phonological words and stuttering on function words. *Journal of Speech, Language, and Hearing Research*, *41*(5), 1019-1030.

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.

Baese-Berk, M. M., Dilley, L. C., Henry, M. J., Vinke, L., & Banzina, E. (2019). Not just a function of function words: Distal speech rate influences perception of prosodically weak syllables. *Attention, Perception, & Psychophysics*, *81*(2), 571-589.

Baese-Berk, M. M., Heffner, C. C., Dilley, L. C., Pitt, M. A., Morrill, T. H., & McAuley, J. D. (2014). Long-term temporal tracking of speech rate affects spoken-word recognition. *Psychological Science*, *25*(8), 1546-1553.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255-278.

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). Parsimonious mixed models. *arXiv*, *1506.04967*.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *arXiv*, *1406.5823*.

Beal, D. S., Gracco, V. L., Brettschneider, J., Kroll, R. M., & Luc, 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.

Beal, D. S., Gracco, V. L., Lafaille, S. J., & Luc, F. (2007). Voxel-based morphometry of auditory and speech-related cortex in stutterers. *Neuroreport*, *18*(12), 1257-1260.

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*, 89.

Beckman, M. E., & Pierrehumbert, J. B. (1986). Intonational structure in Japanese and English. *Phonology*, *3*, 255-309.

Behroozmand, R., Johari, K., Bridwell, K., Hayden, C., Fahey, D., & den Ouden, D. B. (2020). Modulation of vocal pitch control through high-definition transcranial direct current stimulation of the left ventral motor cortex. *Experimental Brain Research*.

Behroozmand, R., Phillip, L., Johari, K., Bonilha, L., Rorden, C., Hickok, G., & Fridriksson, J. (2018). Sensorimotor impairment of speech auditory feedback processing in aphasia. *NeuroImage*, *165*, 102-111.

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.

Bengtsson, S. L., Ullen, F., Ehrsson, H. H., Hashimoto, T., Kito, T., Naito, E., Forssberg, H., & Sadato, N. (2009). Listening to rhythms activates motor and premotor cortices. *Cortex*, *45*(1), 62-71.

Bergmann, G. (1986). Studies in stuttering as a prosodic disturbance. *Journal of Speech, Language and Hearing Research*, *29*(3), 290-300.

Bernstein Ratner, N. (1997). Stuttering: A psycholinguistic perspective. In R. Curlee & G. Siegel (Eds.), *Nature and treatment of stuttering: New directions* (2nd ed., pp. 99-127). Allyn & Bacon.

Besozzi, T. E., & Adams, M. R. (1969). The influence of prosody on stuttering adaptation. *Journal of Speech, Language, and Hearing Research*, *12*(4), 818-824.

Bever, T. G. (2018). The unity of consciousness and the consciousness of unity. In R. d. A. L. Gleitman (Ed.), *On Concepts, modules, and language: Cognitive science at its core* (pp. 87-112). Oxford University Press.

Bever, T. G., & Poeppel, D. (2010). Analysis by synthesis: A (re-) emerging program of research for language and vision. *Biolinguistics*, *4*(2-3), 174-200.

Blackburn, B. (1931). Voluntary movements of the organs of speech in stutterers and non-stutturers. *Psychological Monographs*, *41*(4), 1-13.

Blecher, T., Tal, I., & Ben-Shachar, M. (2016). White matter microstructural properties correlate with sensorimotor synchronization abilities. *NeuroImage*, *138*, 1-12.

Bloodstein, O. (1944). Studies in the psychology of stuttering: XIX. The relationship between oral reading rate and severity of stuttering. *Journal of Speech Disorders*, *9*, 161-173.

Bloodstein, O., & Bernstein Ratner, N. (2008). *A handbook of stuttering* (6th ed.). Delmar.

Bock, J. K. (1982). Toward a cognitive psychology of syntax: Information processing contributions to sentence formulation. *Psychological Review*, *89*(1), 1-47.

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*.

Bourguignon, M., Molinaro, N., Lizarazu, M., Taulu, S., Jousmäki, V., Lallier, M., Carreiras, M., & De Tiège, X. (2020). Neocortical activity tracks the hierarchical linguistic structures of self-produced speech during reading aloud. *NeuroImage*, 116788.

Boutsen, F. R., Brutten, G. J., & Watts, C. R. (2000). Timing and intensity variability in the metronomic speech of stuttering and nonstuttering speakers. *Journal of Speech, Language and Hearing Research*, *43*(2), 513-520.

Bowers, A., Bowers, L. M., Hudock, D., & Ramsdell-Hudock, H. L. (2018). Phonological working memory in developmental stuttering: Potential insights from the neurobiology of language and cognition. *Journal of Fluency Disorders*, *58*, 94-117.

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.

Breen, M., Dilley, L. C., Kraemer, J., & Gibson, E. (2012). Inter-transcriber reliability for two systems of prosodic annotation: ToBI (Tones and Break Indices) and RaP (Rhythm and Pitch). *Corpus Linguistics and Linguistic Theory*, *8*(2), 277-312.

Breen, M., Dilley, L. C., McAuley, J. D., & Sanders, L. D. (2014). Auditory evoked potentials reveal early perceptual effects of distal prosody on speech segmentation. *Language, Cognition and Neuroscience*, *29*(9), 1132-1146.

Browman, C. P., & Goldstein, L. (1992). Articulatory phonology: An overview. *Phonetica*, *49*(3-4), 155-180.

Brown, C. J., Zimmermann, G. N., Linville, R. N., & Hegmann, J. P. (1990). Variations in self-paced behaviors in stutterers and nonstutterers. *Journal of Speech & Hearing Research*, *33*(317-323), 317-323.

Brown, M., Salverda, A. P., Dilley, L. C., & Tanenhaus, M. K. (2011). Expectations from preceding prosody influence segmentation in online sentence processing. *Psychonomic Bulletin & Review*, *18*(6), 1189-1196.

Brown, M., Salverda, A. P., Dilley, L. C., & Tanenhaus, M. K. (2015). Metrical expectations from preceding prosody influence spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *41*(2), 306-323.

Brown, M., Tanenhaus, M. K., & Dilley, L. C. (in press). Syllable inference as a mechanism for spoken language understanding. *Topics in Cognitive Science*.

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.

Brown, S. F. (1938). Stuttering with relation to word accent and word position. *The Journal of Abnormal and Social Psychology*, *33*(1), 112-120.

Brown, S. F. (1945). The loci of stutterings in the speech sequence. *Journal of Speech Disorders*, *10*(3), 181-192.

Budde, K. S., Barron, D. S., & Fox, P. T. (2014). Stuttering, induced fluency, and natural fluency: A hierarchical series of activation likelihood estimation meta-analyses. *Brain and Language*, *139*, 99-107.

Buonomano, D. V. (2003). Timing of neural responses in cortical organotypic slices. *Proceedings of the National Academy of Sciences*, *100*(8), 4897-4902.

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.

Byrd, C. T., Coalson, G. A., Yang, J., & Moriarty, K. (2017). The effect of phonetic complexity on the speed of single-word productions in adults who do and do not stutter. *Journal of Communication Disorders*, *69*, 94-105.

Byrd, C. T., Vallely, M., Anderson, J. D., & Sussman, H. (2012). Nonword repetition and phoneme elision in adults who do and do not stutter. *Journal of Fluency Disorders*, *37*(3), 188-201.

Cai, S. (2011). *Online control of articulation based on auditory feedback in normal speech and stuttering: Behavioral and modeling studies* (Doctoral dissertation, Massachusetts Institute of Technology).

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.

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.

Callan, D. E., Jones, J. A., Callan, A. M., & Akahane-Yamada, R. (2004). Phonetic perceptual identification by native-and second-language speakers differentially activates brain regions involved with acoustic phonetic processing and those involved with articulatory–auditory/orosensory internal models. *NeuroImage*, *22*(3), 1182-1194.

Cameron, D. J., & Grahn, J. A. (2014). Neuroscientific investigations of musical rhythm. *Acoustics Australia*, *42*(2).

Ćavar, M. E., & Lulich, S. M. (2020). Allophonic variation in the Polish vowel/i: Results of a 3D ultrasound study and their phonological implications. *Journal of Slavic Linguistics*, *28*(1), 1-21.

Chang, S.-E. (2011). *Using brain imaging to unravel the mysteries of stuttering*. Dana Foundation.

Chang, S.-E., Angstadt, M., Chow, H. M., Etchell, A. C., Garnett, E. O., Choo, A. L., Kessler, D., Welsh, R. C., & Sripada, C. (2018). Anomalous network architecture of the resting brain in children who stutter. *Journal of Fluency Disorders*, *55*, 46-67.

Chang, S.-E., Chow, H. M., Wieland, E. A., & McAuley, J. D. (2016). Relation between functional connectivity and rhythm discrimination in children who do and do not stutter. *NeuroImage: Clinical*, *12*, 442-450.

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., Garnett, E. O., Etchell, A., & Chow, H. M. (2019). Functional and neuroanatomical bases of developmental stuttering: current insights. *The Neuroscientist*, *25*(6), 566-582.

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.

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.

Chang, S.-E., & Zhu, D. (2013). Neural network connectivity differences in children who stutter. *Brain*, *136*(12), 3709-3726.

Chen, A. (2018). Get the focus right across languages: Acquisition of prosodic focus-marking in production. In *The Development of Prosody in First Language Acquisition* (pp. 295-314). John Benjamins.

Chen, J. L., Penhune, V. B., & Zatorre, R. J. (2008). Listening to musical rhythms recruits motor regions of the brain. *Cerebral Cortex*, *18*(12), 2844-2854.

Chen, J. L., Zatorre, R. J., & Penhune, V. B. (2006). Interactions between auditory and dorsal premotor cortex during synchronization to musical rhythms. *Neuroimage*, *32*(4), 1771-1781.

Chen, Q., & Mirman, D. (2012). Competition and cooperation among similar representations: Toward a unified account of facilitative and inhibitory effects of lexical neighbors. *Psychological Review*, *119*(2), 417-430.

Chien, P. J., Friederici, A. D., Hartwigsen, G., & Sammler, D. (2020). Intonation processing increases task-specific fronto-temporal connectivity in tonal language speakers. *Human Brain Mapping*, 1-14.

Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. Harper & Row.

Chon, H., Sawyer, J., & Ambrose, N. G. (2012). Differences of articulation rate and utterance length in fluent and disfluent utterances of preschool children who stutter. *Journal of Communication Disorders*, *45*, 455-467.

Churchland, M. M., Cunningham, J. P., Kaufman, M. T., Foster, J. D., Nuyujukian, P., Ryu, S. I., & Shenoy, K. V. (2012). Neural population dynamics during reaching. *Nature*, *487*(7405), 51-56.

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.

Civier, O., Tasko, S. M., & Guenther, F. H. (2010). Overreliance on auditory feedback may lead to sound/syllable repetitions: simulations of stuttering and fluency-inducing conditions with a neural model of speech production. *Journal of Fluency Disorders*, *35*(3), 246-279.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and BrainSsciences*, *36*(3), 181-204.

Clark, H. H. (1996). *Using Language*. Cambridge University Press.

Clopper, C. G., Turnbull, R., Cangemi, F., Clayards, M., Niebuhr, O., Schuppler, B., & Zellers, M. (2018). Exploring variation in phonetic reduction: Linguistic, social, and cognitive factors. *Rethinking Reduction*, 25-72.

Conture, E. G., & Kelly, E. M. (1991). Young stutterers' nonspeech behaviors during stuttering. *Journal of Speech, Language, and Hearing Research*, *34*(5), 1041-1056.

Cooper, M. H., & Allen, G. D. (1977). Timing control accuracy in normal speakers and stutterers. *Journal of Speech, Language, and Hearing Research*, *20*(1), 55-71.

Coull, J. T. (2004). fMRI studies of temporal attention: allocating attention within, or towards, time. *Cognitive Brain Research*, *21*(2), 216–226.

Coull, J. T., Davranche, K., Nazarian, B., & Vidal, F. (2013). Functional anatomy of timing differs for production versus prediction of time intervals. *Neuropsychologia*, *51*(2), 309-319.

Cox, R. W. (1996). AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research*, *29*(3), 162-173.

Crystal, T. H., & House, A. S. (1988). Segmental durations in connected-speech signals: Current results. *The Journal of the Acoustical Society of America*, *83*(4), 1553-1573.

Cummins, F., & Port, R. (1998). Rhythmic constraints on stress timing in English. *Journal of Phonetics*, *26*(2), 145-171.

Cunnington, R., Windischberger, C., Deecke, L., & Moser, E. (2002). The preparation and execution of self-initiated and externally-triggered movement: a study of event-related fMRI. *Neuroimage*, *15*(2), 373-385.

Cutting, J. E., & Rosner, B. S. (1974). Categories and boundaries in speech and music. *Perception & Psychophysics*, *16*(3), 564-570.

Cykowski, M. D., Fox, P. T., Ingham, R. J., Ingham, J. C., & Robin, D. A. (2010). A study of the reproducibility and etiology of diffusion anisotropy differences in developmental stuttering: a potential role for impaired myelination. *Neuroimage*, *52*(4), 1495-1504.

Cykowski, M. D., Kochunov, P. V., Ingham, R. J., Ingham, J. C., Mangin, J.-F., Rivière, D., Lancaster, J. L., & Fox, P. T. (2008). Perisylvian sulcal morphology and cerebral asymmetry patterns in adults who stutter. *Cerebral Cortex*, *18*(3), 571-583.

Daliri, A., & Max, L. (2015). Electrophysiological evidence for a general auditory prediction deficit in adults who stutter. *Brain and Language*, *150*, 37-44.

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.

Darwin, C. J. (1975). On the dynamic use of prosody in speech perception. In *Structure and process in speech perception* (pp. 178-194). Springer.

Daube, C., Ince, R. A., & Gross, J. (2019). Simple acoustic features can explain phoneme-based predictions of cortical responses to speech. *Current Biology*, *29*(12), 1924-1937. e1929.

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. *International Journal of Language & Communication Disorders*, *49*(1), 100-112.

Davis, M. H., Johnsrude, I. S., Hervais-Adelman, A., Taylor, K., & McGettigan, C. (2005). Lexical information drives perceptual learning of distorted speech: evidence from the comprehension of noise-vocoded sentences. *Journal of Experimental Psychology: General*, *134*(2), 222.

Dayalu, V. N., Kalinowski, J., Stuart, A., Holbert, D., & Rastatter, M. P. (2002). Stuttering frequency on content and function words in adults who stutter: A concept

revisited. *Journal of Speech, Language, and Hearing Research*, *45*, 871-878.

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.

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), 77-80.

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.

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

Dell, G. S., & Chang, F. (2014). The P-chain: Relating sentence production and its disorders to comprehension and acquisition. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1634), 20120394.

DePasquale, B., Cueva, C. J., Rajan, K., Escola, G. S., & Abbott, L. (2018). full-FORCE: A target-based method for training recurrent networks. *PloS One*, *13*(2), e0191527.

Di Liberto, G. M., Lalor, E. C., & Millman, R. E. (2018). Causal cortical dynamics of a predictive enhancement of speech intelligibility. *Neuroimage*, *166*, 247-258.

Di Liberto, G. M., O'Sullivan, J. A., & Lalor, E. C. (2015). Low-frequency cortical entrainment to speech reflects phoneme-level processing. *Current Biology*, *25*(19), 2457-2465.

Di Luca, M., & Rhodes, D. (2016). Optimal perceived timing: Integrating sensory information with dynamically updated expectations. *Scientific Reports*, *6*, 28563.

Dick, A. S., Bernal, B., & Tremblay, P. (2014). The language connectome: new pathways, new concepts. *The Neuroscientist*, *20*(5), 453-467.

Dick, A. S., & Tremblay, P. (2012). Beyond the arcuate fasciculus: consensus and controversy in the connectional anatomy of language. *Brain*, *135*(12), 3529-3550.

Dilley, L., Shattuck-Hufnagel, S., & Ostendorf, M. (1996). Glottalization of word-initial vowels as a function of prosodic structure. *Journal of Phonetics*, *24*(4), 423-444.

Dilley, L. C. (2005). *The phonetics and phonology of tonal system* Massachusetts Institute of Technology].

Dilley, L. C., & Breen, M. (in press). An e*nhanced* autosegmental-metrical theory (AM +) facilitates phonetically transparent prosodic annotation: A reply to Jun. In J. Barnes & S. Shattuck-Hufnagel (Eds.), *Prosodic theory and practice*. MIT Press.

Dilley, L. C., & Brown, M. (2005). *The RaP (Rhythm and Pitch) labeling system, Version 1.0*. Michigan State University.

Dilley, L. C., Ladd, D. R., & Schepman, A. (2005). Alignment of L and H in bitonal pitch accents: testing two hypotheses. *Journal of Phonetics*, *33*(1), 115-119.

Dilley, L. C., Mattys, S. L., & Vinke, L. (2010). Potent prosody: Comparing the effects of distal prosody, proximal prosody, and semantic context on word segmentation. *Journal of Memory and Language*, *63*(3), 274–294.

Dilley, L. C., & McAuley, J. D. (2008). Distal prosodic context affects word segmentation and lexical processing. *Journal of Memory and Language*, *59*(3), 294-311.

Dilley, L. C., & Pitt, M. A. (2010). Altering context speech rate can cause words to appear or disappear. *Psychological Science*, *21*(11), 1664-1670.

Dilley, L. C., Wieland, E. A., Gamache, J. L., McAuley, J. D., & Redford, M. A. (2013). Age-related changes to spectral voice characteristics affect judgments of prosodic, segmental, and talker attributes for child and adult speech. *Journal of Speech, Language, and Hearing Research*.

Ding, N., Patel, A. D., Chen, L., Butler, H., Luo, C., & Poeppel, D. (2017). Temporal modulations in speech and music. *Neuroscience & Biobehavioral Reviews*, *81*, 181-187.

Ding, N., & Simon, J. Z. (2014). Cortical entrainment to continuous speech: Functional roles and interpretations. *Frontiers in Human Neuroscience*, *8*, 311. https://doi.org/doi: 10.3389/fnhum.2014.00311

DiSimoni, F. G. (1974). Preliminary study of certain timing relationships in the speech of stutterers. *The Journal of the Acoustical Society of America*, *56*(2), 695-696.

Dollaghan, C., & Campbell, T. F. (1998). Nonword repetition and child language impairment. *Journal of Speech, Language, and Hearing Research*, *41*(5), 1136-1146.

Donhauser, P. W., & Baillet, S. (2020). Two distinct neural timescales for predictive speech processing. *Neuron*, *105*(2), 385-393. e389.

Dupoux, E., & Green, K. (1997). Perceptual adjustment to highly compressed speech: Effects of talker and rate changes. *Journal of Experimental Psychology: Human Perception and Performance*, *23*(3), 914.

Egger, S. W., Le, N. M., & Jazayeri, M. (2020). A neural circuit model for human sensorimotor timing. *Nature Communications*, *11*(1), 1-14.

Etchell, A. C., Civier, O., Ballard, K. J., & Sowman, P. F. (2017). A systematic literature review of neuroimaging research on developmental stuttering between 1995 and 2016. *Journal of Fluency Disorders*.

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*.

Falk, S., & Dalla Bella, S. (2016). It is better when expected: Aligning speech and motor rhythms enhances verbal processing. *Language, Cognition and Neuroscience*, *31*(5), 699-708.

Falk, S., Lanzilotti, C., & Schön, D. (*2017*). Tuning neural phase entrainment to speech. *Journal of Cognitive Neuroscience*, *29*(8), 1378-1389.

Falk, S., Müller, T., & Dalla Bella, S. (2015). Non-verbal sensorimotor timing deficits in children and adolescents who stutter. *Frontiers in Psychology*, *6*, 847.

Falk, S., Rathcke, T., & Dalla Bella, S. (2014). When speech sounds like music. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(4), 1491.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149-1160.

Foundas, A. L., Mock, J. R., Corey, D. M., Golob, E. J., & Conture, E. G. (2013). The SpeechEasy device in stuttering and nonstuttering adults: Fluency effects while speaking and reading. *Brain and Language*, *126*(2), 141-150.

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*, 158-162.

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

Freeman, J. S., Cody, F. W., & Schad, W. (1993). The influence of external timing cues upon the rhythm of voluntary movements in Parkinson's disease. *Journal of Neurology, Neurosurgery & Psychiatry*, *56*(10), 1078-1084.

Friederici, A. D. (2011). The brain basis of language processing: From structure to function. *PhysiologicalRreviews*, *91*(4), 1357-1392.

Friederici, A. D., & Gierhan, S. M. (2013). The language network. *Current opinion in neurobiology*, *23*(2), 250-254.

Friederici, A. D., Kotz, S. A., Scott, S. K., & Obleser, J. (2010). Disentangling syntax and intelligibility in auditory language comprehension. *Human Brain Mapping*, *31*(3), 448-457.

Fries, P. (2005). A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. *Trends in Cognitive Sciences*, *9*(10), 474-480.

Fries, P. (2015). Rhythms for cognition: communication through coherence. *Neuron*, *88*(1), 220-235.

Friston, K. (2008). Hierarchical models in the brain. *PLoS Comput Biol*, *4*(11), e1000211.

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*, *11*(2), 127-138.

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

Fujii, S., & Wan, C. Y. (2014). The role of rhythm in speech and language rehabilitation: the SEP hypothesis. *Frontiers in Human Neuroscience*, *8*, 777.

Galle, M. E., Klein-Packard, J., Schreiber, K., & McMurray, B. (2019). What are you waiting for? Real-time integration of cues for fricatives suggests encapsulated auditory memory. *Cognitive Science*, *43*(1), e12700.

Garellek, M. (2020). Acoustic discriminability of the complex phonation system in! Xóõ. *Phonetica*, *77*(2), 131-160.

Garrett, M. (1980). Levels of processing in sentence production. In *Language production Vol. 1: Speech and talk* (pp. 177-220). Academic Press.

Geiser, E., Notter, M., & Gabrieli, J. D. (2012). A corticostriatal neural system enhances auditory perception through temporal context processing. *The Journal of Neuroscience*, *32*(18), 6177-6182.

Geschwind, N. (1970). The organization of language and the brain. *Science*, *170*(3961), 940-944.

Ghitza, O. (2013). The theta-syllable: a unit of speech information defined by cortical function. *Frontiers in Psychology*, *4*, 138.

Ghitza, O. (2014). Behavioral evidence for the role of cortical θ oscillations in determining auditory channel capacity for speech. *Frontiers in Psychology*, *5*, 652.

Ghitza, O., & Greenberg, S. (2009). On the possible role of brain rhythms in speech perception: intelligibility of time-compressed speech with periodic and aperiodic insertions of silence. *Phonetica*, *66*(1-2), 113-126.

Giraud, A. L., Neumann, K., Bachoud-Levi, A. C., von Gudenberg, A. W., Euler, H. A., Lanfermann, H., & Preibisch, C. (2008). Severity of dysfluency correlates with basal ganglia activity in persistent developmental stuttering. *Brain and Language*, *104*(2), 190-199.

Giraud, A. L., & Poeppel, D. (2012). Cortical oscillations and speech processing: Emerging computational principles and operations. *Nature Neuroscience*, *15*(4), 511-517.

Glennon, E., Svirsky, M. A., & Froemke, R. C. (2020). Auditory cortical plasticity in cochlear implant users. *Current Opinion in Neurobiology*, *60*, 108-114.

Glover, H., Kalinowski, J., Rastatter, M., & Stuart, A. (1996). Effect of instruction to sing on stuttering frequency at normal and fast rates. *Perceptual and Motor Skills*, *83*, 511-522.

Godino-Llorente, J., Shattuck-Hufnagel, S., Choi, J., Moro-Velázquez, L., & Gómez-García, J. (2017). Towards the identification of Idiopathic Parkinson's Disease from the speech. New articulatory kinetic biomarkers. *PloS One*, *12*(12), e0189583.

Goldsmith, J. (2011). The syllable. In J. Goldsmith, J. Riggle, & A. C. L. Yu (Eds.), *The Handbook of Phonological Theory* (pp. 164). John Wiley & Sons, Ltd.

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.

Gordon, M., & Ladefoged, P. (2001). Phonation types: a cross-linguistic overview. *Journal of Phonetics*, *29*(4), 383-406.

Gordon, R. L., Shivers, C. M., Wieland, E. A., Kotz, S. A., Yoder, P. J., & Devin McAuley, J. (2015). Musical rhythm discrimination explains individual differences in grammar skills in children. *Developmental Science*, *18*(4), 635-644.

Grahn, J. A., & Brett, M. (2007). Rhythm and beat perception in motor areas of the brain. *Journal of Cognitive Neuroscience*, *19*(5), 893-906.

Grahn, J. A., & Brett, M. (2009). Impairment of beat-based rhythm discrimination in Parkinson's disease. *Cortex*, *45*(1), 54-61.

Grahn, J. A., & McAuley, J. D. (2009). Neural bases of individual differences in beat perception. *Neuroimage*, *47*(4), 1894-1903.

Grahn, J. A., & Rowe, J. B. (2009). Feeling the beat: premotor and striatal interactions in musicians and nonmusicians during beat perception. *The Journal of Neuroscience*, *29*(23), 7540-7548.

Grahn, J. A., & Rowe, J. B. (2013). Finding and feeling the musical beat: striatal dissociations between detection and prediction of regularity. *Cerebral Cortex*, *23*(4), 913-921.

Greenberg, J. B. (1970). The effect of a metronome on the speech of young stutterers. *Behavior Therapy*, *1*(2), 240-244.

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.

Grice, M., Ladd, D. R., & Arvaniti, A. (2000). On the place of phrase accents in intonational phonology. *Phonology*, 143-185.

Gross, J., Hoogenboom, N., Thut, G., Schyns, P., Panzeri, S., Belin, P., & Garrod, S. (2013). Speech rhythms and multiplexed oscillatory sensory coding in the human brain. *PLoS Biology*, *11*(12), e1001752.

Grossman, R. B., Bemis, R. H., Skwerer, D. P., & Tager-Flusberg, H. (2010). Lexical and affective prosody in children with high-functioning autism. *Journal of Speech, Language, and Hearing Research*, *53*(3), 778-793.

Guenther, F. H. (1994). A neural network model of speech acquisition and motor equivalent speech production. *Biological Cybernetics*, *72*(1), 43-53.

Guenther, F. H. (1995). Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production. *Psychological Review*, *102*(3), 594.

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

Guenther, F. H., Espy-Wilson, C. Y., Boyce, S. E., Matthies, M. L., Zandipour, M., & Perkell, J. S. (1999). Articulatory tradeoffs reduce acoustic variability during American English/r/production. *The Journal of the Acoustical Society of America*, *105*(5), 2854-2865.

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.

Guenther, F. H., & Hickok, G. (2015). Role of the auditory system in speech production. In *Handbook of clinical neurology* (Vol. 129, pp. 161-175). Elsevier.

Guenther, F. H., & Hickok, G. (2016). Neural Models of Motor Speech Control In G. Hickok & S. L. Small (Eds.), *Neurobiology of Language* (pp. 725-740). Academic Press

Guenther, F. H., & Perkell, J. S. (2004). A neural model of speech production and its application to studies of the role of auditory feedback in speech. *Speech motor control in normal and disordered speech*, 29-49.

Guenther, F. H., & Vladusich, T. (2012). A neural theory of speech acquisition and production. *Journal of Neurolinguistics*, *25*(5), 408-422.

Guhe, M. (2020). *Incremental conceptualization for language production*. Psychology Press.

Hahn, E. F. (1942). A study of the relationship between stuttering occurrence and grammatical factors in oral reading. *Journal of Speech Disorders*, *7*(4), 329-335.

Hakim, H. B., & Ratner, N. B. (2004). Nonword repetition abilities of children who stutter: An exploratory study. *Journal of Fluency Disorders*, *29*(3), 179-199.

Hall, K. D., Amir, O., & Yairi, E. (1999). A longitudinal investigation of speaking rate in preschool children who stutter. *Journal of Speech, Language and Hearing Research*, *42*(6), 1367–1377.

Halle, M. (1983). On distinctive features and their articulatory implementation. *Natural Language & Linguistic Theory*, 91-105.

Halle, M., & Idsardi, W. (1995). General properties of stress and metrical structure. In J. Goldsmith (Ed.), *Handbook of Phonological Theory* (pp. 403-443). Blackwell Publishers.

Halle, M., & Stevens, K. (1962). Speech recognition: A model and a program for research. *IRE Transactions on Information Theory*, *8*(2), 155-159.

Hannon, E. E., Snyder, J. S., Eerola, T., & Krumhansl, C. L. (2004). The role of melodic and temporal cues in perceiving musical meter. *Journal of Experimental Psychology: Human Perception and Performance*, *30*(5), 956–974.

Hannon, E. E., Soley, G., & Ullal, S. (2012). Familiarity overrides complexity in rhythm perception: A cross-cultural comparison of American and Turkish listeners. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(3), 543.

Hannon, E. E., & Trainor, L. J. (2007). Music acquisition: Effects of enculturation and formal training on development. *Trends in Cognitive Sciences*, *11*(11), 466-472.

Hannon, E. E., & Trehub, S. E. (2005). Tuning in to musical rhythms: Infants learn more readily than adults. *Proceedings of the National Academy of Sciences*, *102*(35), 12639-12643.

Hargrave, S., Kalinowski, J., Stuart, A., Armson, J., & Jones, K. (1994). Effect of frequency-altered feedback on stuttering frequency at normal and fast speech rates. *Journal of Speech, Language, and Hearing Research*, *37*(6), 1313-1319.

Harper, S., Goldstein, L., & Narayanan, S. (2020). Variability in individual constriction contributions to third formant values in American English/ɹ. *The Journal of the Acoustical Society of America*, *147*(6), 3905-3916.

Hasson, U., Egidi, G., Marelli, M., & Willems, R. M. (2018). Grounding the neurobiology of language in first principles: The necessity of non-language-centric explanations for language comprehension. *Cognition*, *180*, 135-157.

Hayes, B. (1995). *Metrical stress theory: Principles and case studies*. University of Chicago Press.

Healey, E. C., & Adams, M. R. (1981). Speech timing skills of normally fluent and stuttering children and adults. *Journal of Fluency Disorders*, *6*(3), 233-246.

Heffner, C. C., Dilley, L. C., McAuley, J. D., & Pitt, M. A. (2013). When cues combine: How distal and proximal acoustic cues are integrated in word segmentation. *Language and Cognitive Processes*, *28*(9), 1275-1302.

Helfrich, R. F., Breska, A., & Knight, R. T. (2019). Neural entrainment and network resonance in support of top-down guided attention. *Current Opinion in Psychology*, *29*, 82-89.

Hellbernd, N., & Sammler, D. (2016). Prosody conveys speaker's intentions: Acoustic cues for speech act perception. *Journal of Memory and Language*, *88*, 70-86.

Helmholtz, H. v. (1860). *Treatise on Physiological Optics*. Dover.

Hickok, G. (2009). Eight problems for the mirror neuron theory of action understanding in monkeys and humans. *Journal of Cognitive Neuroscience*, *21*(7), 1229-1243.

Hickok, G. (2012). Computational neuroanatomy of speech production. *Nature Reviews Neuroscience*, *13*(2), 135-145.

Hickok, G., Houde, J., & Rong, F. (2011). Sensorimotor integration in speech processing: computational basis and neural organization. *Neuron*, *69*(3), 407-422.

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

Hickok, G., & Poeppel, D. (2004). Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition*, *92*(1-2), 67-99.

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, *8*(5), 393-402.

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*, *59*(4), 674-685.

Hoequist, C. E. (1983). The perceptual center and rhythm categories. *Language and Speech*, *26*(4), 367-376.

Houde, J. F., & Jordan, M. I. (1998). Sensorimotor adaptation in speech production. *Science*, *279*(5354), 1213-1216.

Houde, J. F., & Jordan, M. I. (2002). Sensorimotor adaptation of speech I. *Journal of Speech, Language, and Hearing Research*, *45*, 295–310.

Hovsepyan, S., Olasagasti, I., & Giraud, A.-L. (2020). Combining predictive coding and neural oscillations enables online syllable recognition in natural speech. *Nature Communications*, *11*(1), 1-12.

Howell, P., & Au-Yeung, J. (2002). The EXPLAN theory of fluency control applied to the diagnosis of stuttering. *Amsterdam Studies in the Theory and History of Linguistic Science Series 4*, 75-94.

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. H. H. M. v. Lieshout (Eds.), *Speech production: Motor control, brain research and fluency disorders* (pp. 573-578). Elsevier Science.

Hubbard, C. P. (1998). Stuttering, stressed syllables, and word onsets. *Journal of Speech, Language, and Hearing Research*, *41*(4), 802-808.

Hulstijn, W., Summers, J., J., van Lieschout, P. H., & Peters, H. F. (1992). Timing in finger tapping and speech: A comparison between stutterers and fluent speakers. *Human Movement Science*, *11*(1), 113-124.

Humeniuk, E., & Tarkowski, Z. (2017). Overview of research over the efficiency of therapies of stuttering. *Polish Annals of Medicine*, *24*(1), 99-103.

Ingham, R. J., & Carroll, P. J. (1977). Listener judgment of differences in stutterers' nonstuttered speech during chorus-and nonchorus-reading conditions. *Journal of Speech, Language and Hearing Research*, *20*(2), 293-302.

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.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*(4), 434-446.

Jantzen, M. G., Large, E. W., & Magne, C. (2016). Overlap of neural systems for processing language and music. *Frontiers in Psychology*, *7*, 876.

Jin, X., & Costa, R. M. (2015). Shaping action sequences in basal ganglia circuits. *Current Opinion in Neurobiology*, *33*, 188-196.

Jones, J. A., & Striemer, D. (2007). Speech disruption during delayed auditory feedback with simultaneous visual feedback. *The Journal of the Acoustical Society of America*, *122*(4), 135-141.

Jones, M. R. (1976). Time, our lost dimension: Toward a new theory of perception, attention, and memory. *Psychological Review*, *83*(5), 323-355.

Jones, M. R. (2018). *Time Will Tell: A theory of dynamic attending*. Oxford University Press.

Jäncke, L., Hänggi, J., & Steinmetz, H. (2004). Morphological brain differences between adult stutterers and non-stutterers. *BMC Neurology*, *4*(1), 1-8.

Kaasin, K., & Bjerkan, B. (1982). Critical words and the locus of stuttering in speech. *Journal of Fluency Disorders*, *7*(4), 433-446.

Kadi-Hanifi, K., & Howell, P. (1992). Syntactic analysis of the spontaneous speech of normally fluent and stuttering children. *Journal of Fluency Disorders*, *17*(3), 151-170.

Kalinowski, J., Armson, J., Stuart, A., & Gracco, V. L. (1993). Effects of alterations in auditory feedback and speech rate on stuttering frequency. *Language and Speech*, *36*(1), 1-16.

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.

Kalveram, K. T., & Jäncke, L. (1989). Vowel duration and voice onset time for stressed and nonstressed syllables in stutterers under delayed auditory feedback condition. *Folia Phoniatrica*, *41*(1), 30-42.

Karabanov, A., Blom, Ö., Forsman, L., & Ullén, F. (2009). The dorsal auditory pathway is involved in performance of both visual and auditory rhythms. *Neuroimage*, *44*(2), 480-488.

Karmarkar, U. R., & Buonomano, D. V. (2007). Timing in the absence of clocks: encoding time in neural network states. *Neuron*, *53*(3), 427-438.

Karniol, R. (1995). Stuttering, language, and cognition: A review and a model of stuttering as suprasegmental sentence plan alignment (SPA). *Psychological Bulletin*, *1*, 104-124.

Kato, H., Tsuzaki, M., & Sagisaka, Y. (2003). Functional differences between vowel onsets and offsets in temporal perception of speech: Local-change detection and speaking-rate discrimination. *The Journal of the Acoustical Society of America*, *113*(6), 3379-3389.

Kaufeld, G., Bosker, H. R., Ten Oever, S., Alday, P. M., Meyer, A. S., & Martin, A. E. (2020). Linguistic structure and meaning organize neural oscillations into a content-specific hierarchy. *Journal of Neuroscience*.

Kaufeld, G., Ravenschlag, A., Meyer, A. S., Martin, A. E., & Bosker, H. R. (2020). Knowledge-based and signal-based cues are weighted flexibly during spoken language comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46*(3), 549.

Kazanina, N., Bowers, J. S., & Idsardi, W. (2018). Phonemes: Lexical access and beyond. *Psychonomic Bulletin & Review*, *25*(2), 560-585.

Keitel, A., Gross, J., & Kayser, C. (2018). Speech tracking in auditory and motor regions reflects distinct linguistic features. *bioRxiv*, 195941.

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.

Kelly, E., M., & Conture, E. G. (1992). Speaking rates, response time latencies, and interrupting behaviors of young stutterers, nonstutterers, and their mothers. *Journal of Speech and Hearing Research*, *35*, 1256-1267.

Kent, R. D. (1984). Stuttering as a temporal programming disorder. *Nature and treatment of stuttering: New directions*, 283-301.

Kiefte, M., & Armson, J. (2008). Dissecting choral speech: Properties of the accompanist critical to stuttering reduction. *Journal of Communication Disorders*, *41*(1), 33-48.

Killeen, P. R., & Fetterman, J. G. (1988). A behavioral theory of timing. *Psychological Review*, *95*(2), 274-295.

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), 548–559.

Klich, R. J., & May, G. M. (1982). Spectrographic study of vowels in stutterers' fluent speech. *Journal of Speech, Language, and Hearing Research*, *25*(3), 364-370.

Klimovich-Gray, A., & Molinaro, N. (2020). Synchronising internal and external information: A commentary on Meyer, Sun & Martin (2020). *Language, Cognition and Neuroscience*, 1-4.

Kochanski, G., Grabe, E., Coleman, J., & Rosner, B. (2005). Loudness predicts prominence: Fundamental frequency lends little. *The Journal of the Acoustical Society of America*, *118*(2), 1038-1054.

Koelsch, S. (2011). Toward a neural basis of music perception - a review and updated model. *Frontiers in Psychology*, *2*, 110.

Koelsch, S., Vuust, P., & Friston, K. (2019). Predictive processes and the peculiar case of music. *Trends in Cognitive Sciences*, *23*(1), 63-77.

Kondaurova, M. V., & Francis, A. L. (2008). The relationship between native allophonic experience with vowel duration and perception of the English tense/lax vowel contrast by Spanish and Russian listeners. *The Journal of the Acoustical Society of America*, *124*(6), 3959-3971.

Koopmans, M., Slis, I., & Rietveld, T. (1992). Stotteren als uiting van spraakplanning een vergelijking tussen voorgelezen en spontane spraak. *Stem-, Spraak-en Taalpathologie*, *1*(2), 87-101.

Kotz, S. A., Ravignani, A., & Fitch, W. (2018). The evolution of rhythm processing. *Trends in Cognitive Sciences*, *22*(10), 896-910.

Kotz, S. A., & Schmidt-Kassow, M. (2015). Basal ganglia contribution to rule expectancy and temporal predictability in speech. *Cortex*, *68*, 48-60.

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

Kotz, S. A., & Schwartze, M. (2016). Motor-timing and sequencing in speech production: A general-purpose framework. In *Neurobiology of Language* (pp. 717-724). Academic Press.

Kronfeld-Duenias, V., Amir, O., Ezrati-Vinacour, R., Civier, O., & Ben-Shachar, M. (2016). The frontal aslant tract underlies speech fluency in persistent developmental stuttering. *Brain Structure and Function*, *221*(1), 365-381.

Kung, S.-J., Zatorre, R. J., & Penhume, V. B. (2013). Interacting cortical and basal ganglia networks underlying finding and tapping to the musical beat. *Journal of Cognitive Neuroscience*, *25*(3), 401-420.

Ladd, D. R. (2008). *Intonational phonology*. Cambridge University Press.

Ladányi, E., Persici, V., Fiveash, A., Tillmann, B., & Gordon, R. L. (2020). Is atypical rhythm a risk factor for developmental speech and language disorders? *Wiley Interdisciplinary Reviews: Cognitive Science*, e1528.

Lane, H., Denny, M., Guenther, F. H., Matthies, M. L., Menard, L., Perkell, J. S., Stockmann, E., Tiede, M., Vick, J., & Zandipour, M. (2005). Effects of bite blocks and hearing status on vowel production. *The Journal of the Acoustical Society of America*, *118*(3), 1636-1646.

Large, E. W., Herrera, J. A., & Velasco, M. J. (2015). Neural networks for beat perception in musical rhythm. *Frontiers in Systems Neuroscience*, *9*, 159.

Large, E. W., & Jones, M. R. (1999). The dynamics of attending: how people track time-varying events. *Psychological Review*, *106*(1), 119-159.

Large, E. W., & Kolen, J. F. (1994). Resonance and the perception of musical meter. *Connection Science*, *6*(2-3), 177-208.

Lashley, K. S. (1951). *The problem of serial order in behavior* (Vol. 21). Bobbs-Merrill Oxford, United Kingdom.

Law, T., Packman, A., Onslow, M., To, C. K.-S., Tong, M. C.-F., & Lee, K. Y.-S. (2018). Rhythmic speech and stuttering reduction in a syllable-timed language. *Clinical Linguistics & Phonetics*, *32*(10), 932-949.

Lehiste, I. (1970). Suprasegmentals. MIT Press

Lehiste, I. (1977). Isochrony reconsidered. *Journal of Phonetics*, *5*, 253-263.

Leong, V., & Goswami, U. (2014). Impaired extraction of speech rhythm from temporal modulation patterns in speech in developmental dyslexia. *Frontiers in Human Neuroscience*, *8*, 96.

Lerdahl, F., & Jackendoff, R. S. (1983). *A generative theory of tonal music*. MIT Press.

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

Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*, 1-38.

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

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

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*(3), 1126-1177.

Lewis, P. A., & Miall, R. C. (2003). Distinct systems for automatic and cognitively controlled time measurement: evidence from neuroimaging. *Current Opinion in Neurobiology*, *13*(2), 250-255.

Lewis, P. A., Wing, A., Pope, P., Praamstra, P., & Miall, R. (2004). Brain activity correlates differentially with increasing temporal complexity of rhythms during initialisation, synchronisation, and continuation phases of paced finger tapping. *Neuropsychologia*, *42*(10), 1301-1312.

Li, Q., Liu, G., Wei, D., Liu, Y., Yuan, G., & Wang, G. (2019). Distinct neuronal entrainment to beat and meter: Revealed by simultaneous EEG-fMRI. *NeuroImage*, *194*, 128-135.

Liberman, M. (1995). The sound structure of Mawu words: a case study in the cognitive science of speech. *An Invitation to Cognitive Science*, *1*, 55-85.

Liberman, M. Y. (1975). *The intonational system of English* Massachusetts Institute of Technology].

Liu, H., Auger, J., & Larson, C. R. (2010). Voice fundamental frequency modulates vocal response to pitch perturbations during English speech. *The Journal of the Acoustical Society of America*, *127*(1), EL1-EL5.

Lizarazu, M., Lallier, M., & Molinaro, N. (2019). Phase− amplitude coupling between theta and gamma oscillations adapts to speech rate. *Annals of the New York Academy of Sciences*, *1453*(1), 140.

Logan, K. J., & Conture, E. C. (1995). Length, grammatical, complexity, and rate differences in stuttered and fluent conversational utterance of children who stutter. *Journal of Fluency Disorders*, *20*, 35-61.

Loucks, T. M. J., Chon, H., & Han, W. (2012). Audiovocal integration in adults who stutter. *International Journal of Language & Communication Disorders*, *47*(4), 451-456.

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.

Luo, H., & Poeppel, D. (2007). Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron*, *54*(6), 1001-1010.

MacKain, K. S., Best, C. T., & Strange, W. (1981). Categorical perception of English/r/and/l/by Japanese bilinguals. *Applied Psycholinguistics*, *2*(4), 369-390.

Mackay, D. G., & Macdonald, M. C. (1984). *Stuttering as a Sequencing and Timing Disorder*. Citeseer.

Macmillan, N. A., & Creelman, C. D. (2005). *Detection Theory: A user's guide* (2 ed.). Lawrence Erlbaum Associates.

MacNeilage, P. F. (1998). The frame/content theory of evolution of speech production. *Behavioral and Brain Sciences*, *21*(4), 499-511.

Maguire, G. A., Yu, B. P., Franklin, D. L., & Riley, G. D. (2004). Alleviating stuttering with pharmacological interventions. *Expert Opinion on Pharmacotherapy*, *5*(7), 1565-1571.

Maner, K. J., Smith, A., & Grayson, L. (2000). Influences of utterance length and complexity on speech motor performance in children and adults. *Journal of Speech, Language, and Hearing Research*, *43*(2), 560-573.

Marcus, S. M. (1981). Acoustic determinants of perceptual center (P-center) location. *Perception & Psychophysics*, *30*(3), 247-256.

Marslen-Wilson, W. D., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology*, *10*(1), 29-63.

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, *94*, 305-315.

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*, *46*, 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*, 105-122.

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.

Mayville, J. M., Jantzen, K. J., Fuchs, A., Steinberg, F. L., & Kelso, J. S. (2002). Cortical and subcortical networks underlying syncopated and synchronized coordination revealed using fMRI. *Human Brain Mapping*, *17*(4), 214-229.

McAuley, J. D., & Jones, M. R. (2003). Modeling effects of rhythmic context on perceived duration: A comparison of interval and entrainment approaches to short-interval timing. *Journal of Experimental Psychology: Human Perception & Performance*, *29*(6), 1102-1125.

McAuley, J. D., Jones, M. R., Holub, S., Johnston, H. M., & Miller, N. S. (2006). The time of our lives: life span development of timing and event tracking. *Journal of Experimental Psychology: General*, *135*(3), 348-367.

McQueen, J. M., & Dilley, L. (2021). Prosody and spoken-word recognition. In C. Gussenhoven & A. Chen (Eds.), *The Oxford Handbook of Language Prosody* (pp. 509-521). Oxford University Press.

Meck, W. H. (1996). Neuropharmacology of timing and time perception. *Cognitive Brain Research*, *3*(3-4), 227-242.

Melnick, K. S., & Conture, E. G. (2000). Relationship of length and grammatical complexity to the systematic and nonsystematic speech errors and stuttering of children who stutter. *Journal of Fluency Disorders*, *25*(1), 21-45.

Merchant, H., Harrington, D. L., & Meck, W. H. (2013). Neural basis of the perception and estimation of time. *Annual Review of Neuroscience*, *36*, 313-336.

Meyer, L., Sun, Y., & Martin, A. E. (2020a). "Entraining" to speech, generating language? *Language, Cognition and Neuroscience*, *35*(9), 1138-1148.

Meyer, L., Sun, Y., & Martin, A. E. (2020b). Synchronous, but not entrained: Exogenous and endogenous cortical rhythms of speech and language processing. *Language, Cognition and Neuroscience*, *35*(9), 1089-1099.

Miall, C. (1989). The storage of time intervals using oscillating neurons. *Neural Computation*, *1*(3), 359-371.

Miller, J. E., Carlson, L. A., & McAuley, J. D. (2013). When what you hear influences when you see: Listening to an auditory rhythm influences the temporal allocation of visual attention. *Psychological Science*, *24*(1), 11-18.

Mitterer, H., Kim, S., & Cho, T. (2019). The glottal stop between segmental and suprasegmental processing: The case of Maltese. *Journal of Memory and Language*, *108*, 104034.

Molinaro, N., & Lizarazu, M. (2018). Delta (but not theta)-band cortical entrainment involves speech-specific processing. *European Journal of Neuroscience*, *48*(7), 2642-2650.

Morillon, B., & Schroeder, C. E. (2015). Neuronal oscillations as a mechanistic substrate of auditory temporal prediction. *Annals of the New York Academy of Sciences*, *1337*(1), 26-31.

Morrill, T. H., Baese-Berk, M., Heffner, C., & Dilley, L. C. (2015). Interactions between distal speech rate, linguistic knowledge, and speech environment. *Psychonomic Bulletin & Review*, *22*(5), 1451-1457.

Morrill, T. H., Dilley, L. C., & McAuley, J. D. (2014). Prosodic patterning in distal speech context: Effects of list intonation and f0 downtrend on perception of proximal prosodic structure. *Journal of Phonetics*, *46*, 68-85.

Morrill, T. H., Dilley, L. C., McAuley, J. D., & Pitt, M. A. (2014). Distal rhythm influences whether or not listeners hear a word in continuous speech: Support for a perceptual grouping hypothesis. *Cognition*, *131*(1), 69-74.

Morrill, T. H., McAuley, J. D., Dilley, L. C., Zdziarska, P. A., Jones, K. B., & Sanders, L. D. (2015). Distal prosody affects learning of novel words in an artificial language. *Psychonomic Bulletin & Review*, *22*(3), 815-823.

Morton, J., Marcus, S., & Frankish, C. (1976). Perceptual centers (P-centers). *Psychological Review*, *83*(5), 405.

Natke, U., Grosser, J., & Kalveram, K. T. (2001). Fluency, fundamental frequency, and speech rate under frequency-shifted auditory feedback in stuttering and nonstuttering persons. *Journal of Fluency Disorders*, *26*(3), 227-241.

Natke, U., Grosser, J., Sandrieser, P., & Kalveram, K. T. (2002). The duration component of the stress effect in stuttering. *Journal of Fluency Disorders*, *27*(4), 305-318.

Nazzi, T., & Cutler, A. (2019). How consonants and vowels shape spoken-language recognition. *Annual Review of Linguistics*, *5*, 25-47.

Neef, N. E., Anwander, A., Bütfering, C., Schmidt-Samoa, C., Friederici, A. D., Paulus, W., & Sommer, M. (2018). Structural connectivity of right frontal hyperactive areas scales with stuttering severity. *Brain*, *141*(1), 191-204.

Neef, N. E., Bütfering, C., Anwander, A., Friederici, A. D., Paulus, W., & Sommer, M. (2016). Left posterior-dorsal area 44 couples with parietal areas to promote speech fluency, while right area 44 activity promotes the stopping of motor responses. *Neuroimage*, *142*, 628-644.

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.

Nespor, M., & Vogel, I. (2007). *Prosodic Phonology: With a new foreword* (Vol. 28). Walter de Gruyter.

Oganian, Y., & Chang, E. F. (2019). A speech envelope landmark for syllable encoding in human superior temporal gyrus. *Science Advances*, *5*(11), eaay6279.

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), 876–886.

Olasagasti, I., Bouton, S., & Giraud, A.-L. (2015). Prediction across sensory modalities: A neurocomputational model of the McGurk effect. *Cortex*, *68*, 61-75.

Oppenheim, G. M., Dell, G. S., & Schwartz, M. F. (2010). The dark side of incremental learning: A model of cumulative semantic interference during lexical access in speech production. *Cognition*, *114*(2), 227-252.

Oyoun, H. A., El Dessouky, H., Shohdi, S., & Fawzy, A. (2010). Assessment of working memory in normal children and children who stutter. *Journal of American Science*, *6*(11), 562-566.

O'Donnell, J. J., Armson, J., & Kiefte, M. (2008). The effectiveness of SpeechEasy during situations of daily living. *Journal of Fluency Disorders*, *33*(2), 99-119.

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.

Packman, A., Onslow, M., Richard, F., & Van Doorn, J. (1996). Syllabic stress and variability: A model of stuttering. *Clinical Linguistics & Phonetics*, *10*(3), 235-263.

Pasinski, A. C., McAuley, J. D., & Snyder, J. S. (2016). How modality specific is processing of auditory and visual rhythms? *Psychophysiology*, *53*(2), 198-208.

Patel, A. D. (2006). Musical rhythm, linguistic rhythm, and human evolution. *Music Perception*, *24*, 99-104.

Patel, A. D. (2010). *Music, Language, and the Brain*. Oxford University Press.

Patel, A. D., & Iversen, J. R. (2014). The evolutionary neuroscience of musical beat perception: the Action Simulation for Auditory Prediction (ASAP) hypothesis. *Frontiers in Systems Neuroscience*, *8*(57).

Patel, A. D., Iversen, J. R., Bregman, M. R., Schulz, I., & Schulz, C. (2008). Investigating the human-specificity of synchronization to music. Proceedings of the 10th International Conference on Music Perception and Cognition,

Peelle, J. E., & Davis, M. H. (2012). Neural oscillations carry speech rhythm through to comprehension. *Frontiers in Psychology*, *3*, 320.

Peelle, J. E., Gross, J., & Davis, M. H. (2013). Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. *Cerebral Cortex*, *23*(6), 1378-1387.

Pefkou, M., Arnal, L. H., Fontolan, L., & Giraud, A.-L. (2017). Theta-and beta-band neural activity reflect independent syllable tracking and comprehension of time-compressed speech. *Journal of Neuroscience*, *37*(33), 7930 –7938.

Pelczarski, K. M., & Yaruss, J. S. (2016). Phonological memory in young children who stutter. *Journal of Communication Disorders*, *62*, 54-66.

Perani, D., Saccuman, M. C., Scifo, P., Anwander, A., Spada, D., Baldoli, C., Poloniato, A., Lohmann, G., & Friederici, A. D. (2011). Neural language networks at birth. *Proceedings of the National Academy of Sciences*, *108*(38), 16056-16061.

Perkell, J., Matthies, M., Lane, H., Guenther, F., Wilhelms-Tricarico, R., Wozniak, J., & Guiod, P. (1997). Speech motor control: Acoustic goals, saturation effects, auditory feedback and internal models. *Speech Communication*, *22*(2-3), 227-250.

Perkell, J. S., Guenther, F. H., Lane, H., Matthies, M. L., Perrier, P., Vick, J., Wilhelms-Tricarico, R., & Zandipour, M. (2000). A theory of speech motor control and supporting data from speakers with normal hearing and with profound hearing loss. *Journal of Phonetics*, *28*(3), 233-272.

Perkins, W. H., Kent, R. D., & Curlee, R. F. (1991). A theory of neuropsycholinguistic function in stuttering. *journal of Speech, Language, and Hearing Research*, *34*(4), 734-752.

Perrachione, T. K., Stepp, C. E., Hillman, R. E., & Wong, P. C. (2009). The role of source and filter characteristics in human talker identification: Experiments with laryngeal and electrolarynx speech. *Journal of Speech, Language, and Hearing Research*, *48*, 766-779.

Perrachione, T. K., Stepp, C. E., Hillman, R. E., & Wong, P. C. (2014). Talker identification across source mechanisms: Experiments with laryngeal and electrolarynx speech. *Journal of Speech, Language, and Hearing Research*, *57*(5), 1651-1665.

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, *36*(4), 329-347.

Pierrehumbert, J., & Beckman, M. (1988). Japanese tone structure. *Linguistic Inquiry Monographs,* (15), 1-282.

Pierrehumbert, J. B. (1980). *The phonology and phonetics of English intonation* (Doctoral dissertation, Massachusetts Institute of Technology).

Pitt, M. A., Szostak, C., & Dilley, L. C. (2016). Rate dependent speech processing can be speech specific: Evidence from the perceptual disappearance of words under changes in context speech rate. *Attention, Perception, & Psychophysics*, *78*(1), 334-345.

Poeppel, D. (2003). The analysis of speech in different temporal integration windows: Cerebral lateralization as 'asymmetric sampling in time'. *Speech Communication*, *41*(1), 245-255.

Poeppel, D., Emmorey, K., Hickok, G., & Pylkkänen, L. (2012). Towards a new neurobiology of language. *Journal of Neuroscience*, *32*(41), 14125-14131.

Pollard, R., Ellis, J. B., Finan, D., & Ramig, P. R. (2009). Effects of the SpeechEasy on objective and perceived aspects of stuttering: a 6-month, phase I clinical trial in naturalistic environments. *Journal of Speech, Language, and Hearing Research*.

Postma, A., & Kolk, H. (1993). The Covert Repair Hypothesis: Prearticulatory repair processes in normal and stuttered disfluencies. *Journal of Speech and Hearing Research,*, *36*, 472-487.

Povel, D.-J., & Essens, P. (1985). Perception of temporal patterns. *Music Perception*, *2*(4), 411-440.

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.

Prins, D., Hubbard, C. P., & Krause, M. (1991). Syllabic stress and the occurrence of stuttering. *Journal of Speech, Language, and Hearing Research*, *34*(5), 1011-1016.

Proctor, M., Zhu, Y., Lammert, A., Toutios, A., Sands, B., & Narayanan, S. (2020). Studying clicks using real-time MRI. In *Click consonants* (pp. 210-240). Brill.

Pulvermüller, F., Huss, M., Kherif, F., del Prado Martin, F. M., Hauk, O., & Shtyrov, Y. (2006). Motor cortex maps articulatory features of speech sounds. *Proceedings of the National Academy of Sciences*, *103*(20), 7865-7870.

Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*(1), 79-87.

Rao, S. M., Mayer, A. R., & Harrington, D. L. (2001). The evolution of brain activation during temporal processing. *Nature Neuroscience*, *4*, 317 - 323.

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Computation*, *20*(4), 873-922.

Ratcliff, R., Smith, P. L., Brown, S. D., & McKoon, G. (2016). Diffusion decision model: Current issues and history. *Trends in Cognitive Sciences*, *20*(4), 260-281.

Rauschecker, J. P. (2011). An expanded role for the dorsal auditory pathway in sensorimotor control and integration. *Hearing Research*, *271*(1-2), 16-25.

Rauschecker, J. P., & Scott, S. K. (2009). Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nature Neuroscience*, *12*(6), 718-724.

Rauschecker, J. P., & Tian, B. (2000). Mechanisms and streams for processing of "what" and "where" in auditory cortex. *Proceedings of the National Academy of Sciences*, *97*(22), 11800-11806.

Ravignani, A., Dalla Bella, S., Falk, S., Kello, C. T., Noriega, F., & Kotz, S. A. (2019). Rhythm in speech and animal vocalizations: a cross-species perspective. *Annals of the New York Academy of Sciences*, *1453*(1), 79.

Redi, L., & Shattuck-Hufnagel, S. (2001). Variation in the realization of glottalization in normal speakers. *Journal of Phonetics*, *29*(4), 407-429.

Reinisch, E., Jesse, A., & Nygaard, L. C. (2013). Tone of voice guides word learning in informative referential contexts. *Quarterly Journal of Experimental Psychology*, *66*(6), 1227-1240.

Remez, R. E., Dubowski, K. R., Broder, R. S., Davids, M. L., Grossman, Y. S., Moskalenko, M., Pardo, J. S., & Hasbun, S. M. (2011). Auditory-phonetic projection and lexical structure in the recognition of sine-wave words. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(3), 968.

Remez, R. E., Rubin, P. E., Pisoni, D. B., & Carrell, T. D. (1981). Speech perception without traditional speech cues. *Science*, *212*(4497), 947-949.

Remington, E. D., Egger, S. W., Narain, D., Wang, J., & Jazayeri, M. (2018). A dynamical systems perspective on flexible motor timing. *Trends in Cognitive Sciences*, *22*(10), 938-952.

Remington, E. D., Narain, D., Hosseini, E. A., & Jazayeri, M. (2018). Flexible sensorimotor computations through rapid reconfiguration of cortical dynamics. *Neuron*, *98*(5), 1005-1019. e1005.

Repp, B. H. (1998). Obligatory "expectations" of expressive timing induced by perception of musical structure. *Psychological Research*, *61*(1), 33-43.

Ries, S., Piai, V., Perry, D., Griffin, S., Jordan, K., Henry, R., Knight, R., & Berger, M. (2019). Roles of ventral versus dorsal pathways in language production: An awake language mapping study. *Brain and Language*, *191*, 17-27.

Ritto, A. P., Juste, F. S., Stuart, A., Kalinowski, J., & de Andrade, C. R. F. (2016). Randomized clinical trial: The use of SpeechEasy® in stuttering treatment. *International Journal of Language & Communication Disorders*, *51*(6), 769-774.

Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition*, *42*(1-3), 107-142.

Rommers, J., Dell, G. S., & Benjamin, A. S. (2020). Word predictability blurs the lines between production and comprehension: Evidence from the production effect in memory. *Cognition*, *198*, 104206.

Ross, J. M., Iversen, J. R., & Balasubramaniam, R. (2018). The role of posterior parietal cortex in beat-based timing perception: A continuous theta burst stimulation study. *Journal of Cognitive Neuroscience*, *30*(5), 634-643.

Ryan, B. P. (1992). Articulation, language, rate, and fluency characteristics of stuttering and nonstuttering preschool children. *Journal of Speech and Hearing Research*, *35*(2), 333-342.

Ryan, B. P. (2000). Speaking rate, conversational speech acts, interruption, and linguistic complexity of 20 pre-school stuttering and non-stuttering children and their mothers. *Clinical Linguistics & Phonetics*, *14*(1), 25-51.

Saffran, J. R. (2020). Statistical language learning in infancy. *Child Development Perspectives*, *14*(1), 49-54.

Sammler, D. (2018). *The melodic mind: Neural bases of intonation in speech and music* (Doctoral dissertation, Max Planck Institute for Human Cognitive and Brain Sciences Leipzig).

Sammler, D. (2020). Splitting speech and music. *Science*, *367*(6481), 974-976.

Sammler, D., Cunitz, K., Gierhan, S. M., Anwander, A., Adermann, J., Meixensberger, J., & Friederici, A. D. (2018). White matter pathways for prosodic structure building: A case study. *Brain and Language*, *183*, 1-10.

Sasisekaran, J., & Byrd, C. T. (2013). A preliminary investigation of segmentation and rhyme abilities of children who stutter. *Journal of Fluency Disorders*, *38*(2), 222-234.

Sasisekaran, J., & Weisberg, S. (2014). Practice and retention of nonwords in adults who stutter. *Journal of Fluency Disorders*, *41*, 55-71.

Saur, D., Kreher, B. W., Schnell, S., Kümmerer, D., Kellmeyer, P., Vry, M.-S., Umarova, R., Musso, M., Glauche, V., Abel, S., Hauber, W., Rijntjes, M., Hennig, J., & Weiller, C. (2008). Ventral and dorsal pathways for language. *Proceedings of the National Academy of Sciences*, *105*(46), 18035-18040.

Schiavetti, N., Metz, D. E., & R., O. (2010). *Evaluating research in communicative disorders* (6 ed.).

Schirmer, A. (2004). Timing speech: a review of lesion and neuroimaging findings. *Cognitive Brain Research*, *21*(2), 269-287.

Schreiber, K. E., & McMurray, B. (2019). Listeners can anticipate future segments before they identify the current one. *Attention, Perception, & Psychophysics*, *81*(4), 1147-1166.

Schubotz, R. I., Friederici, Angela D., & Von Cramon, D. Y. (2000). Time perception and motor timing: A common cortical and subcortical basis revealed by fMRI. *NeuroImage*, *11*(1), 1–12.

Schwartze, M., Keller, P. E., Patel, A. D., & Kotz, S. A. (2011). The impact of basal ganglia lesions on sensorimotor synchronization, spontaneous motor tempo, and the detection of tempo changes. *Behavioural Brain Research*, *216*(2), 685-691.

Schwartze, M., & Kotz, S. A. (2013). A dual-pathway neural architecture for specific temporal prediction. *Neuroscience & Biobehavioral Reviews*, *37*(10), 2587-2596.

Schwartze, M., & Kotz, S. A. (2016). Contributions of cerebellar event-based temporal processing and preparatory function to speech perception. *Brain and Language*, *161*, 28-32.

Schwartze, M., Rothermich, K., & Kotz, S. A. (2012). Functional dissociation of pre-SMA and SMA-proper in temporal processing. *Neuroimage*, *60*(1), 290-298.

Senft, V., Stewart, T. C., Bekolay, T., Eliasmith, C., & Kröger, B. J. (2016). Reduction of dopamine in basal ganglia and its effects on syllable sequencing in speech: a computer simulation study. *Basal Ganglia*, *6*(1), 7-17.

Shattuck-Hufnagel, S. (1979). Speech errors as evidence for a serial oral mechanism in sentence production. In W. E. Cooper & E. C. T. Walker (Eds.), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett*. Lawrence Erlbaum.

Shattuck-Hufnagel, S. (1987). The role of word-onset consonants in speech production planning: New evidence from speech error patterns. In E. Keller & M. Gopnik (Eds.), *Motor and sensory processes of language* (pp. 17-51). Lawrence Erlbaum Associates Inc.

Shattuck-Hufnagel, S., & Klatt, D. H. (1979). The limited use of distinctive features and markedness in speech production: Evidence from speech error data. *Journal of Verbal Learning and Verbal Behavior,* (18), 1.

Shen, Y., & G, P. G. (1962). Isochronisms in English. *Studies in Linguistics, Occasional Papers (No. 9)*.

Shi, Z., & Burr, D. (2016). Predictive coding of multisensory timing. *Current Opinion in Behavioral Sciences*, *8*, 200-206.

Shima, K., & Tanji, J. (2000). Neuronal activity in the supplementary and presupplementary motor areas for temporal organization of multiple movements. *Journal of Neurophysiology*, *84*(4), 2148-2160.

Shockey, L. (2008). *Sound Patterns of Spoken English*. John Wiley & Sons.

Simen, P., Balci, F., deSouza, L., Cohen, J. D., & Holmes, P. (2011). A model of interval timing by neural integration. *Journal of Neuroscience*, *31*(25), 9238-9253.

Skeide, M. A., & Friederici, A. D. (2016). The ontogeny of the cortical language network. *Nature Reviews Neuroscience*, *17*(5), 323-332.

Small, L. H. (2015). *Fundamentals of Phonetics: A practical guide for students*. Pearson.

Smith, A., Goffman, L., Sasisekaran, J., & Weber-Fox, C. (2012). Language and motor abilities of preschool children who stutter: Evidence from behavioral and kinematic indices of nonword repetition performance. *Journal of Fluency Disorders*, *37*(4), 344-358.

Smith, A., Sadagopan, N., Walsh, B., & Weber-Fox, C. (2010). Increasing phonological complexity reveals heightened instability in inter-articulatory coordination in adults who stutter. *Journal of Fluency Disorders*, *35*(1), 1-18.

Smith, A., & Weber, C. (2017). How stuttering develops: The multifactorial dynamic pathways theory. *Journal of Speech, Language, and Hearing Research*, *60*(9), 2483-2505.

Smith, C. L., Browman, C. P., McGowan, R. S., & Kay, B. (1993). Extracting dynamic parameters from speech movement data. *The Journal of the Acoustical Society of America*, *93*(3), 1580-1588.

Soderberg, G. A. (1962). Phonetic influences upon stuttering. *Journal of Speech and Hearing Research*, *5*(4), 315-320.

Sommer, M., Koch, M. A., Paulus, W., Weiller, C., & Büchel, C. (2002). Disconnection of speech-relevant brain areas in persistent developmental stuttering. *The Lancet*, *360*(9330), 380-383.

Song, L., Peng, D., Jin, Z., Yao, L., Ning, N., Guo, X., & Zhang, T. (2007). Gray matter abnormalities in developmental stuttering determined with voxel-based morphometry. *Zhonghua yi xue za zhi*, *87*(41), 2884-2888.

Sowman, P. F., Crain, S., Harrison, E., & Johnson, B. W. (2012). Reduced activation of left orbitofrontal cortex precedes blocked vocalization: a magnetoencephalographic study. *Journal of Fluency Disorders*, *37*(4), 359-365.

Spencer, C., & Weber-Fox, C. (2014). Preschool speech articulation and nonword repetition abilities may help predict eventual recovery or persistence of stuttering. *Journal of Fluency Disorders*, *41*, 32-46.

Spencer, K. A., & Rogers, M. A. (2005). Speech motor programming in hypokinetic and ataxic dysarthria. *Brain and Language*, *94*(3), 347-366.

Stanovich, K. E. (2012). *How to think straight about psychology* (10 ed.). Pearson: HarperCollins Publishers.

Stevens, K. N. (2002). Toward a model for lexical access based on acoustic landmarks and distinctive features. *The Journal of the Acoustical Society of America*, *111*(4), 1872-1891.

Stuart, A., Kalinowski, J., Rastatter, M. P., Saltuklaroglu, T., & Dayalu, V. (2004). Investigations of the impact of altered auditory feedback in-the-ear devices on the speech of people who stutter: initial fitting and 4-month follow-up. *International Journal of Language & Communication Disorders*, *39*(1), 93-113.

Taylor, I. K. (1966). The properties of stuttered words. *Journal of Verbal Learning and Verbal Behavior*, *5*(2), 112-118.

Thomassen, J. M. (1982). Melodic accent: Experiments and a tentative model. *The Journal of the Acoustical Society of America*, *71*(6), 1596-1605.

Tichenor, S., & Yaruss, J. S. (2018). A phenomenological analysis of the experience of stuttering. *American Journal of Speech-Language Pathology*, *27*(3S), 1180-1194.

Tierney, A., Patel, A. D., & Breen, M. (2018a). Acoustic foundations of the speech-to-song illusion. *Journal of Experimental Psychology: General*, *147*(6), 888.

Tierney, A., Patel, A. D., & Breen, M. (2018b). Repetition enhances the musicality of speech and tone stimuli to similar degrees. *Music Perception: An Interdisciplinary Journal*, *35*(5), 573-578.

Tilsen, S. (2013). A dynamical model of hierarchical selection and coordination in speech planning. *PLoS One*, *8*(4), e62800.

Tilsen, S., & Arvaniti, A. (2013). Speech rhythm analysis with decomposition of the amplitude envelope: Characterizing rhythmic patterns within and across languages. *The Journal of the Acoustical Society of America*, *134*(1), 628-639.

Tilsen, S., & Johnson, K. (2008). Low-frequency Fourier analysis of speech rhythm. *The Journal of the Acoustical Society of America*, *124*(2), EL34-EL39.

Todd, M. N., Lee, C., & O'Boyle, D. (2002). A sensorimotor theory of temporal tracking and beat induction. *Psychological Research*, *66*(1), 26-39.

Tornick, G. B., & Bloodstein, O. (1976). Stuttering and sentence length. *Journal of Speech and Hearing Research*, *19*(4), 651-654.

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.

Toutios, A., Xu, M., Byrd, D., Goldstein, L., & Narayanan, S. (2020). How an aglossic speaker produces an alveolar-like percept without a functional tongue tip. *The Journal of the Acoustical Society of America*, *147*(6), EL460-EL464.

Toyomura, A., Fujii, T., & Kuriki, S. (2011). Effect of external auditory pacing on the neural activity of stuttering speakers. *Neuroimage*, *57*(4), 1507-1516.

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.

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.

Treisman, M. (1963). Temporal discrimination and the indifference interval: Implications for a model of the" internal clock". *Psychological Monographs: General and Applied*, *77*(13), 1.

Tsuboi, N., Francis, W. S., & Jameson, J. T. (2020). How word comprehension exposures facilitate later spoken production: implications for lexical processing and repetition priming. *Memory*, 1-20.

Ueno, T., Saito, S., Rogers, T. T., & Ralph, M. A. L. (2011). Lichtheim 2: Synthesizing aphasia and the neural basis of language in a neurocomputational model of the dual dorsal-ventral language pathways. *Neuron*, *72*(2), 385-396.

Van Borsel, J., Sunaert, R., & Engelen, S. (2005). Speech disruption under delayed auditory feedback in multilingual speakers. *Journal of Fluency Disorders*, *30*(3), 201-217.

van der Burght, C. L., Goucha, T., Friederici, A. D., Kreitewolf, J., & Hartwigsen, G. (2019). Intonation guides sentence processing in the left inferior frontal gyrus. *Cortex*, *117*, 122-134.

Van Riper, C. (1982). *The nature of stuttering*. Prentice Hall.

Vaquero, L., Ramos-Escobar, N., François, C., Penhune, V., & Rodríguez-Fornells, A. (2018). White-matter structural connectivity predicts short-term melody and White-matter structural connectivity predicts short-term melody and rhythm learning in non-musicians. *Neuroimage*, *181*(252– 2 252 –262), 252-262.

Vaquero, L., Rodríguez-Fornells, A., & Reiterer, S. M. (2017). The left, the better: White-matter brain integrity predicts foreign language imitation ability. *Cerebral Cortex*, *27*(8), 3906-3917.

Vasic, N., & Wijnen, F. (2005). Stuttering as a monitoring deficit. In R. J., Hartsuiker, R. Bastiaanse, A. Postma, & F. Wijnen (Eds.), *Phonological encoding and monitoring in normal and pathological speech*. Psychology Press.

Vigneau, M., Beaucousin, V., Herve, P.-Y., Duffau, H., Crivello, F., Houde, O., Mazoyer, B., & Tzourio-Mazoyer, N. (2006). Meta-analyzing left hemisphere language areas: phonology, semantics, and sentence processing. *Neuroimage*, *30*(4), 1414-1432.

Villacorta, V. M., Perkell, J. S., & Guenther, F. H. (2007). Sensorimotor adaptation to feedback perturbations of vowel acoustics and its relation to perception. *The Journal of the Acoustical Society of America*, *122*(4), 2306-2319.

Vuust, P., Dietz, M. J., Witek, M., & Kringelbach, M. L. (2018). Now you hear it: A predictive coding model for understanding rhythmic incongruity. *Annals of the New York Academy of Sciences*, *1423*(1), 19-29.

Wagner, M., & Watson, D. G. (2010). Experimental and theoretical advances in prosody: A review. *Language and Cognitive Processes*, *25*(7-9), 905-945.

Wang, J., Narain, D., Hosseini, E. A., & Jazayeri, M. (2018). Flexible timing by temporal scaling of cortical responses. *Nature Neuroscience*, *21*(1), 102-110.

Warbrick, T., Rosenberg, J., & Shah, N. J. (2017). The relationship between BOLD fMRI response and the underlying white matter as measured by fractional anisotropy (FA): A systematic review. *Neuroimage*, *153*, 369-381.

Warren, P. (2000). Prosody and language processing In L. R. Wheeldon (Ed.), *Aspects of Language Production* (pp. 71-114). Psychology Press.

Watkins, K. E., Smith, S. M., Davis, S., & Howell, P. (2008). Structural and functional abnormalities of the motor system in developmental stuttering. *Brain*, *131*(Pt 1), 50-59.

Watson, D. G., Arnold, J. E., & Tanenhaus, M. K. (2008). Tic Tac TOE: Effects of predictability and importance on acoustic prominence in language production. *Cognition*, *106*(3), 1548-1557.

Watson, D. G., Tanenhaus, M. K., & Gunlogson, C. A. (2008). Interpreting pitch accents in online comprehension: H* vs. L+ H. *Cognitive Science*, *32*(7), 1232-1244.

Weiner, A. E. (1984). Stuttering and syllable stress. *Journal of Fluency Disorders*, *9*(4), 301-305.

Wendahl, R. W., & Cole, J. (1961). Identification of stuttering during relatively fluent speech. *Journal of Speech, Language and Hearing Research*, *4*(3), 281-286.

Wheeldon, L., & Waksler, R. (2004). Phonological underspecification and mapping mechanisms in the speech recognition lexicon. *Brain and Language*, *90*(1-3), 401-412.

White, L. (2002). *English speech timing: A domain and locus approach* (Doctoral dissertation, University of Edinburgh).

Whitfield, J. A., Reif, A., & Goberman, A. M. (2018). Voicing contrast of stop consonant production in the speech of individuals with Parkinson disease ON and OFF dopaminergic medication. *Clinical Linguistics & Phonetics*, *32*(7), 587-594.

Wieland, E. A., Dilley, L. C., Burnham, E. B., & Chang, S.-E. (in preparation). Prosodic characteristics of fluent speech by children who do and do not stutter. *To be submitted to Journal of Fluency Disorders*.

Wieland, E. A., McAuley, J. D., Dilley, L. C., & Chang, S.-E. (2015). Evidence for a rhythm perception deficit in children who stutter. *Brain and Language*, *144*, 26-34.

Wieland, E. A., McAuley, J. D., Zhu, D., Dilley, L. C., & Chang, S.-E. (2014, November 6-10, 2014). Brain activity differences during rhythm discrimination in adults who stutter. Society for Neuroscience, San Diego, CA.

Wiener, M., Turkeltaub, P. E., & Coslett, H. B. (2010). Implicit timing activates the left inferior parietal cortex. *Neuropsychologia*, *48*(13), 3967-3971.

Wingate, M. E. (1966). Prosody in stuttering adaptation. *Journal of Speech, Language, and Hearing Research*, *9*(4), 550-556.

Wingate, M. E. (1984). Stutter events and linguistic stress. *Journal of Fluency Disorders*, *9*(4), 295-300.

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

Wingate, M. E. (2012). *The structure of stuttering: A psycholinguistic analysis*. Springer Science & Business Media.

World-Health-Organization. (2010). *International statistical classification of diseases and related health problems*. World Heath Organization. Retrieved 4/30/13 from http://apps.who.int/classifications/icd10/browse/2010/en#/F98.5

Wright, D. B., & London, K. (2009). Multilevel modelling: Beyond the basic applications. *British Journal of Mathematical and Statistical Psychology*, *62*(2), 439-456.

Wu, J. C., Maguire, G., Riley, G., Lee, A., Keator, D., Tang, C., Fallon, J., & Najafi, A. (1997). Increased dopamine activity associated with stuttering. *Neuroreport*, *8*(3), 767-770.

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.

Yairi, E., & Ambrose, N. G. (1999). Early childhood stuttering I: Persistency and recovery rates. *Journal of Speech, Language and Hearing Research*, *42*(5), 1097-1112.

Yairi, E., & Ambrose, N. G. (2013). Epidemiology of stuttering: 21st century advances. *Journal of Fluency Disorders*, *38*(2), 66-87.

Yang, Y., Jia, F., Siok, W. T., & Tan, L. H. (2016). Altered functional connectivity in persistent developmental stuttering. Scientific reports, 6. *Scientific Reports*, *6*, 1-8.

Yaruss, J. S. (1999). Utterance length, syntactic complexity, and childhood stuttering. *Journal of Speech, Language and Hearing Research*, *42*(2), 329–344.

Yaruss, J. S. (2010). Assessing quality of life in stuttering treatment outcomes research. *Journal of Fluency Disorders*, *35*(3), 190-202.

Yeatman, J. D., Dougherty, R. F., Rykhlevskaia, E., Sherbondy, A. J., Deutsch, G. K., Wandell, B. A., & Ben-Shachar, M. (2011). Anatomical properties of the arcuate fasciculus predict phonological and reading skills in children. *Journal of Cognitive Neuroscience*, *23*(11), 3304-3317.

Yi, H. G., Leonard, M. K., & Chang, E. F. (2019). The encoding of speech sounds in the superior temporal gyrus. *Neuron*, *102*(6), 1096-1110.

Zelaznik, H. N., Smith, A., & Franz, E. A. (1994). Motor performance of stutterers and nonstutterers on timing and force control tasks. *Journal of Motor Behavior*, *26*(4), 340-347.

Zelaznik, H. N., Smith, A., Franz, E. A., & Ho, M. (1997). Differences in bimanual coordination associated with stuttering. *Acta Psychologica*, *96*(3), 229-243.

Zimmermann, G. (1980). Stuttering: A disorder of movement. *Journal of Speech, Language and Hearing Research*, *23*(1), 122-136.

Zoefel, B., Archer-Boyd, A., & Davis, M. H. (2018). Phase entrainment of brain oscillations causally modulates neural responses to intelligible speech. *Current Biology*, *28*(3), 401-408. e405.