Variability, Stability, and Flexibility in the Speech Kinematics and Acoustics of Adults Who Do and Do Not Stutter

Eric S. Jackson
Graduate Center, City University of New York

How does access to this work benefit you? Let us know!
Follow this and additional works at: https://academicworks.cuny.edu/gc_etds
Part of the Cognitive Psychology Commons, and the Communication Commons

Recommended Citation
https://academicworks.cuny.edu/gc_etds/986

This Dissertation is brought to you by CUNY Academic Works. It has been accepted for inclusion in All Dissertations, Theses, and Capstone Projects by an authorized administrator of CUNY Academic Works. For more information, please contact deposit@gc.cuny.edu.
VARIABILITY, STABILITY, AND FLEXIBILITY IN THE SPEECH KINEMATICS AND ACOUSTICS OF ADULTS WHO DO AND DO NOT STUTTER

by

ERIC S. JACKSON

A dissertation submitted to the Graduate Faculty in Speech-Language-Hearing Sciences in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York

2015
This manuscript has been read and accepted for the Graduate Faculty in Speech-Language-Hearing Sciences to satisfy the dissertation requirement for the degree of Doctor of Philosophy.

Douglas H. Whalen, Ph.D.

____________________________

Date

Chair of Examining Committee

Klara Marton, Ph.D.

____________________________

Date

Executive Officer

Mark Tiede, Ph.D.

____________________________

Ben Watson, Ph.D.

____________________________

Supervisory Committee

Deryk Beal, Ph.D.

____________________________

Outside Reader

THE CITY UNIVERSITY OF NEW YORK
Abstract

VARIABILITY, STABILITY, AND FLEXIBILITY IN THE SPEECH KINEMATICS AND ACOUSTICS OF ADULTS WHO DO AND DO NOT STUTTER

by

ERIC S. JACKSON

Advisor: Professor Douglas H. Whalen

It is well known that people who do and do not stutter produce speech differently, at least some of the time, even when perceived as fluent. One way that investigators have assessed these differences is by measuring variability, or the inconsistency of repeated speech movements. Variability in speech has typically been quantified using linear analysis techniques (e.g., measures of central tendency), and results have indicated that people who stutter produce speech that is (sometimes) characterized by increased variability. However, variability is a complex phenomenon, one that cannot be assessed by linear methods alone. This dissertation employs linear and nonlinear analysis techniques to examine the nature of variability, stability, and flexibility in stuttering and non-stuttering speakers.

Two experiments are reported in this dissertation. The first is a pilot study in which 11 participants judged short utterances that were manipulated in gap (or pause) duration to be fluent or disfluent. This preliminary study facilitated the selection of “fluent” utterances for the primary experiment, which measured lip aperture kinematics...
and acoustics for 20 speakers who stutter and 21 speakers who do not stutter, under two manipulations: 1) audience and non-audience; 2) increasing linguistic complexity.

Results from the primary experiment corroborated results from prior studies that used linear techniques to show that 1) adults who stutter exhibit more effector variability than adults who do not stutter when target utterances are embedded in sentences of increased linguistic complexity, and 2) linear acoustic measures are as effective as linear kinematic measures for quantifying variability. Nonlinear analysis techniques demonstrated that adults who stutter exhibit more deterministic structure in lip aperture dynamics. Furthermore, cognitive-emotional stress (i.e., the presence of an audience) resulted in decreased surface variability, increased deterministic structure, decreased stationarity, and decreased signal complexity in speakers who stutter, but not in those who do not stutter. Thus, adults who stutter appear to exhibit less overall stability, which leads to a more rigid, less flexible approach to speech production, especially when cognitive-emotional stressors are placed on their speech motor systems.

These findings highlight the benefits of using nonlinear analysis techniques to examine variability in speech production. Specifically, the results demonstrated that speech movements that appear to be less variable on the surface, may in fact be overly deterministic and nonstationary—two attributes that indicate system instability in complex biological systems. Thus, a combination of linear and nonlinear approaches is warranted in future investigations of speech production.
Acknowledgements

Perhaps it is obvious, but this dissertation would not exist if not for the help and support of many people—too many, in fact, to be mentioned here. Still, I will acknowledge some of these individuals in the paragraphs that follow.

I have been fortunate to have Doug Whalen as my advisor throughout this doctoral program. Doug has shown me, through instruction and example, how to be a good scientist. He has been fully supportive of my work since the beginning of my doctoral program, and has provided me with numerous opportunities to collaborate with researchers and colleagues from various fields. This, along with his always-constructive feedback and criticism, especially regarding my written work, has contributed significantly to my development as a researcher of speech production. I am forever indebted to Doug for his help at this stage in my career.

My committee members, Mark Tiede and Ben Watson, played integral roles in my training. Mark has been exceedingly generous with his (much sought after) time. I would not have been able to conduct the data processing and analyses described in this manuscript without Mark’s guidance and expertise in developing and writing MATLAB procedures. Observing Mark’s meticulous approach to conducting research has made me a better researcher. I am also grateful for the support and guidance I received from Ben Watson. I benefitted from our philosophical, yet grounded, discussions about stuttering and speech production, which helped clarify my ideas, ultimately making them suitable for the scientific page. Ben’s consistent reminder that the dissertation is only an academic exercise helped me realize that this process has been the start, and not the culmination, of my work. Furthermore, my conversations with Deryk Beal about
stuttering and research have influenced how I think about being a clinician-scientist, and I am grateful to Deryk for his valuable feedback on this manuscript.

I thank my colleagues in the Speech Production lab, who have provided me with critical feedback and support throughout the dissertation process. I particularly thank Kate Dawson, who played a significant role as lab manager in the early days of the Speech Production lab, and who also provided moral support throughout the program. I am grateful to Kevin Roon for our many (high- and low-level) conversations about speech production, and specifically, my work. As a postdoctoral researcher, I aspire to be as helpful to doctoral students as Kevin was to me. My former student, Valerie Terranova, deserves recognition for devoting a substantial amount of time to helping conduct the experiments reported herein. Along with Valerie, many of my colleagues donated their time to serve as “coughers” during the audience condition of the primary experiment. Without these individuals, the study would have been far less interesting.

I am thankful to have developed personal and professional relationships through my work with the StutterTalk Podcast. My candid conversations with Bob Quesal, Scott Yaruss, Lee Caggiano, and Joe Klein, as well as many others, about stuttering, therapy, and professional issues related to clinical work, undoubtedly shaped (and continue to shape) my approach to conducting research. I also thank Patricia Zebrowski and John Spencer who two years ago, agreed to mentor me during postdoctoral work. Their feedback during the grant writing process enhanced my scientific writing and critical thinking skills, and I am very much looking forward to our work together.

My conversations with Pascal van Lieshout, as well as his research, have greatly informed my thinking regarding the nature of the variability of stuttering. Similarly, a
conversation with Anne Smith early in my doctoral program, as well as her group’s research, served as an impetus for developing a dissertation that focuses on the variability of stuttering.

I also appreciate the many vibrant conversations I had with my colleague and friend Jason Rosas about the nature of disordered speech production and experimental design. I thank the rest of my colleagues at Long Island University – Brooklyn, where I was employed for the majority of my program, for their understanding regarding the demands of doctoral work.

I am grateful to Mike Riley, Kevin Shockley, and their colleagues at the University of Cincinnati for developing and consistently offering the Advanced Training Institute on Nonlinear Methods. This week-long program provided me with a battery of analysis techniques and corresponding MATLAB procedures, but more importantly, refined my understanding of the analysis of complex biological systems.

I am forever grateful to my mother and father, Susan and Mark, for instilling in me the motivation and resilience necessary to complete a doctoral program. Their support and encouragement, as well as support from the rest of my family, allowed me to become the scientist that I am today.

Above all, I thank my beautiful wife, Julie, for her unwavering support, in every way, during the preparation and completion of this manuscript. Her patience and understanding of my being “busy” is what ultimately allowed for this dissertation to be finished. I am excited about what lies ahead for us.
This work was funded in part through NIH grant DC-002717 to Haskins Laboratories. Participant recruitment for the experiments was facilitated in part by the National Stuttering Association.
Dedication

This dissertation is dedicated to those individuals who know first-hand the devastating impact that stuttering can have on the life of a person who stutters. I hope that this work will help to improve our understanding of this all-too-often misunderstood disorder.
# Table of Contents

Chapter 1: Introduction ........................................................................................................1

Chapter 2: Literature Review/Background .........................................................................4

- Motor Control ..................................................................................................................4
- Divergent Perspectives .................................................................................................4
- Variability ......................................................................................................................10
- The Fluent Speech Paradigm ..........................................................................................15
- Stuttering and Variability ..............................................................................................19
- Nonlinear Analysis Techniques ....................................................................................22
- Contextual Influences on Stuttering ..............................................................................28
- Cognitive-Emotional Influences ....................................................................................31
- Linguistic Influences ......................................................................................................33

Chapter 3: Experiment 1 .....................................................................................................34

- Methodology ..................................................................................................................36
- Participants ....................................................................................................................36
- Stimuli ............................................................................................................................37
- Design ............................................................................................................................39
- Results ............................................................................................................................39
- Discussion .......................................................................................................................42

Chapter 4: Experiment 2 .....................................................................................................45

- Methodology ..................................................................................................................46
- Participants ....................................................................................................................46
- Stimuli ............................................................................................................................48
- Experimental Design ......................................................................................................50
- Data collection and analysis .........................................................................................54
- STI ..................................................................................................................................56
- RQA ...............................................................................................................................60
- Duration .........................................................................................................................64
- Amplitude Range ...........................................................................................................64
- Statistical Analysis ........................................................................................................64

Chapter 5: Results from Experiment 2 ...............................................................................70

- STI .................................................................................................................................70
- %DET ..............................................................................................................................75
- TREND ...........................................................................................................................78
- ENTROPY .......................................................................................................................81
- Duration .........................................................................................................................82
- Correlations with Duration ............................................................................................84
- Amplitude Range ...........................................................................................................85
- Summary .........................................................................................................................87

Chapter 6: Discussion ..........................................................................................................89
List of Figures

Figure 1. Time series of chaotic and white noise signals........................................13
Figure 2. Phase space reconstruction: raw trajectory and phase space plot.............25
Figure 3. (Auto-)recurrence plots with high and low recurrence..........................26
Figure 4. Percentage fluent vs. gap duration..........................................................40
Figure 5. Percentage fluent vs. percent gap time.....................................................41
Figure 6. Raw and registered trajectories...............................................................56
Figure 7. LA-STI and A-STI calculation presented visually....................................58
Figure 8. LA-NSTI{amp} and LA-NSTI{phase} calculation presented visually..........59
Figure 9. RQA calculation: raw trajectory, phase plot, recurrence plot.................63
Figure 10. Audience effect for shifters: LA-STI.....................................................73
Figure 11. Group differences by sentence: TREND...............................................79
List of Tables

Table 1. Stimuli from perception of fluency experiment........................................38
Table 2. Participant characteristics.........................................................................49
Table 3. Stimuli from primary experiment.................................................................49
Table 4. ANOVA comparisons of AT models with three fixed factors......................67
Table 5. ANOVA comparison to determine inclusion of trial as fixed factor............69
Table 6. ANOVA comparison for random slopes for trial by participant...............69
Table 7. LME model: LA-STI..................................................................................71
Table 8. LME model, only P2: LA-STI.................................................................71
Table 9. LME model, only shifter PWS: LA-STI.....................................................73
Table 10. LME model: %DET...............................................................................76
Table 11. LME model, only audience: %DET.............................................................77
Table 12. LME model, only PWS: %DET.................................................................77
Table 13. LME model: TREND.............................................................................78
Table 14. LME model, only PWS, including condition*sentence interaction: TREND...80
Table 15. LME model: ENTROPY.........................................................................81
Table 16. LME model: Duration.............................................................................83
Table 17. LME model, only PWNS: Duration..........................................................85
Table 18. Correlations between dependent variables and duration.......................85
Table 19. LME model: Amplitude Range.................................................................86
Table 20. LME model, only shifters: Amplitude Range...........................................87
Table 21. Summary of results.................................................................................88
List of Abbreviations

A-STI: acoustic spatiotemporal index
Base: base utterance ("Buy Bobby a puppy")
CNS: central nervous system
DIVA: directions into velocities of articulators
DST: dynamical systems theory
EOWPVT: expressive one-word picture vocabulary test
GODIVA: gradient order directions into velocities of articulators
LA-STI: lip aperture spatiotemporal index
LA-NSTIamp: nonlinear lip aperture spatiotemporal index, amplitude component
LA-NSTIphase: nonlinear lip aperture spatiotemporal index, phase component
LOI: line of identity
L1: longer-only sentence ("Four one three two five buy Bobby a puppy ten eight nine eleven")
MP: motor programs
OASES: overall assessment of the speaker’s experience of stuttering
PNS: peripheral nervous system
PWNS: people who do not stutter
PWS: people who stutter
P1: perspective embedment level 1 sentence ("He wants Karen to tell John to buy Bobby a puppy at my store")
P2: perspective embedment level 2 sentence ("You want Samantha to buy Bobby a puppy now if he wants one")
RP: recurrence plot
RQA: recurrence quantification analysis
SLP: speech-language pathologist
SSI-4: stuttering severity instrument, 4th edition
STI: spatiotemporal index
TD: task dynamics
Chapter 1: Introduction

Stuttering is a neurodevelopmental speech disorder that affects approximately five to eight percent of young children and one percent of older children and adults (Yairi & Ambrose, 1999, 2005, 2012). Numerous studies have shown that stuttering can have a significantly negative impact on mental health, social interactions and participation, and academic and vocational opportunities of those who stutter (Beilby, Byrnes, Meagher, & Yaruss, 2013; Bleek et al., 2012; Blumgart, Tran, & Craig, 2010; Bricker-Katz, Lincoln, & McCabe, 2009; Craig, Blumgart, & Tran, 2009; Craig, Hancock, Tran, & Craig, 2003; Daniels & Gabel, 2004; Franck, Jackson, Pimentel, & Greenwood, 2003; Klein & Hood, 2004; Yaruss & Quesal, 2004; Yaruss, 2010). Developing a more refined understanding of the underlying nature of stuttering, as well as its quantification, is critical to improving diagnostic protocols and clinical management for children and adults who stutter.

The most salient features of stuttering are atypical interruptions in speech, including part-word repetitions and audible or inaudible prolongations or cessations of sounds. However, stuttering moments do not always result in observable disfluency. There are at least two (related) reasons for this. First, speakers who stutter develop the ability to anticipate stuttering events, and subsequently delay or avoid speech production (Jackson, Yaruss, Quesal, Terranova, & Whalen, in press; Vanryckeghem, Brutten, Uddin, & Borsel, 2004) (for textbook reviews, see Bloodstein & Bernstein-Ratner, 2008; Guitar, 2013; Manning, 2009; Van Riper, 1971; Yairi & Seery, 2015). This can create the illusion that a stuttering event has not occurred, despite an underlying interruption/malfunction of some sort. People who stutter (PWS) develop avoidance
behavior as a result of the social penalty of stuttering (Bowers, Crawcour, Saltuklaroglu, & Kalinowski, 2010; Messenger, Onslow, Packman, & Menzies, 2004; Plexico, Manning, & Levitt, 2009; Sheehan, 1953; Yaruss & Quesal, 2004). Second, subtle differences between the speech of PWS and people who do not stutter (PWNS), whether observed through kinematic, myographic, acoustic, and/or neurological means, are not always perceptible to the human ear (or eye). This dissertation focuses on the latter—that is, the quantification of subtle differences in speech production between PWS and PWNS. Specifically, this dissertation examines the variability, stability, and flexibility associated with speech movements during fluent speech production. It 1) employs linear and nonlinear approaches to measuring kinematic and acoustic speech variability in adult PWS and PWNS, and 2) examines effects of contextual influences on these measures. This work serves the dual purpose of providing insight into the underlying, dynamical nature of speech production in stuttering and non-stuttering speakers, as well as enhancing techniques for the scientific measurement of typical and atypical speech production.

Chapter 2 reviews research in motor control and variability, and discusses variability in the context of stuttering. It then reviews studies that have examined the fluent speech of PWS, as well as controversies related to the so-called fluent speech paradigm. Chapter 2 also reviews linear and nonlinear approaches to quantifying kinematic and acoustic speech signals, as well as data on contextual influences on stuttering behaviors—specifically cognitive-emotional and linguistic variables. Chapter 3 presents a pilot study that examined perceptual thresholds of fluency-disfluency, which addresses some of the controversy related to the fluent speech paradigm discussed in
Chapter 2. Chapter 4 describes the primary experiment, which examined speech variability, stability, and flexibility in 20 adult PWS and 21 adult PWNS under two manipulations: 1) audience/no-audience; 2) differences in linguistic complexity. Chapter 5 synthesizes the first five chapters and provides a general discussion of findings. Chapter 6 concludes and offers potential directions for future work.
Chapter 2: Literature Review/Background

Motor Control

Motor control refers to the process(es) by which humans and animals move effectors (e.g., limbs, articulators) in space and time to achieve desired goals. Stuttering can be viewed as a disorder involving motor difficulty because, at least on the surface and intermittently, it manifests itself in atypical movements of the speech effectors (e.g., De Nil et al., 2008; De Nil & Brutten, 1991; Lieshout, Starkweather, Hulstijn, & Peters, 1995; Max & Gracco, 2005; Namasivayam & Van Lieshout, 2008; Namasivayam & van Lieshout, 2011a, 2011b; Smith & Kleinow, 2000; Smits-Bandstra, De Nil, & Saint-Cyr, 2006; van Lieshout, Ben-David, Lipski, & Namasivayam, 2014; Van Lieshout & Moussa, 2000). Stuttering should not be viewed as a disorder exclusively involving motor difficulty, though, as it may be the case that higher-order cognitive and linguistic factors play just as important a role as motor factors. The underlying causes of stuttering simply remain unknown. However, studying speech motor control in PWS and PWNS provides an established paradigm within which to examine the disorder.

Divergent Perspectives

There are primarily two divergent views in motor control research. The first (and more prevalent) involves motor programs (MP). Lashley (1951) spoke to the existence of central plans, which imply that the central nervous system (CNS) generates commands that inform the peripheral nervous system (PNS; e.g., muscle spindles, joints, articulators) how and when to move in order to achieve desired goals. Several observations support the existence of central plans: 1) movement realization in the
absence of consistent sensory feedback; 2) rapid movement sequences occurring too fast for sensory feedback to have occurred; and 3) speech errors (e.g., slips of the tongue; Lashley, 1951). In an homage paper to Lashley, Rosenbaum, Cohen, Jax, Weiss and vander Wel (2007) reviewed additional evidence: 1) the time to initiate a movement sequence can increase with length or complexity of the sequence; 2) the properties of movements occurring early in a sequence can anticipate later features (e.g., coarticulation); and 3) neural activity can indicate preparation of upcoming behavioral events.

Early views of motor control (e.g., Keele, 1968; Von Holst, 1954) were heavily focused on feedforward processes (i.e., commands from CNS to PNS). More recent MP accounts reflect the importance of feedback/afference. Desmurget and Grafton (2000) proposed that while a motor plan is assembled before movement onset, it is continuously updated and revised via feedback mechanisms. Central to these accounts are internal models, which represent sensorimotor mappings in both efferent and afferent directions. Forward models predict sensory consequences of a given action based on an efference copy of the motor plan; inverse models generate the necessary motor commands required to obtain a desired trajectory by integrating sensory information (Wolpert, Diedrichsen, & Flanagan, 2011; Wolpert, Ghahramani, & Flanagan, 2001). In another so-called hybrid model, aspects of feedforward and feedback commands are combined in an effort “to resolve the previously existing dichotomy between feedforward and feedback models of motor control” (Max, Guenther, Gracco, Ghosh, & Wallace, 2004, p. 109). The Directions into Velocities of Articulators (DIVA) model exemplifies how this hybrid perspective has been applied to
speech production and acquisition (Guenther, 1994, 1995; Guenther, Ghosh, & Tourville, 2006; Guenther & Perkell, 2004). According to DIVA, speech production begins with activation of a speech sound map, which consists of those phonemes that are most prevalent in the speaker’s environment. The speech sound map projects motor commands to the speech mechanism (i.e., lips, jaw, tongue, and larynx) via feedforward processes (Tourville & Guenther, 2011). Feedback processes involve an error-correction mechanism via forward mapping—that is, if the actual movement deviates from the expected consequence, the motor plan is adjusted. Thus, feedback mechanisms only “kick in” should the system require them. In this way, DIVA favors feedforward vs. feedback orientation. DIVA explains findings, according to Guenther, on contextual variability, motor equivalence, coarticulation, and speaking rate effects (Guenther et al., 2006). Gradient order (GO)DIVA is an extension to DIVA that introduces mechanisms for speech planning and the integration of multisyllabic production (Bohland, Bullock, & Guenther, 2010). A strength of the DIVA/GODIVA approach is that it is neurophysiologically grounded; that is, it associates specific brain regions with the various processes assumed necessary for speech production.

The second perspective on motor control is rooted in dynamical systems theory (DST). DST emerged in the early 1980s in response to MP and machine views of motor control and coordination. Early proponents of DST (e.g., Kelso, Holt, Kugler, & Turvey, 1980; Kugler, Kelso, & Turvey, 1980, 1982) argued that MP or machine accounts could not explain the control and coordination of movement because of the separate statuses allocated to controller (i.e., the brain) and that being controlled (i.e., the effectors). Proponents of DST asserted that this separation led to a “loan on intelligence” (Dennett,
since from the MP perspective the controller is the only party privy to the information related to the motor plan(s). Further, this loan on intelligence ultimately must be paid back; the control mechanisms need to be revealed at some point (Kugler et al., 1980; M. A. Riley & Turvey, 2002). MP accounts do not strive to explain the process(es) by which plans are generated; rather, they attempt to approximate the process(es) by relying on machine-like imitation (i.e., the central controller [i.e., CNS] as a computer that sends code/instructions to the PNS). DST posits that the dynamics, and not a central controller, are the control mechanism responsible for movement. Thus, movements of the effectors (e.g., limbs, articulators) are not prescribed via the CNS. Rather, motor control emerges as the product of lawful physical interactions between neurological, physiological, and environmental sub-systems. It is through explanation of how the components of the system change over time (using differential equations) that understanding of the system is garnered.

A challenge faced by all theorists of motor control is explaining how a system regulates its internal degrees of freedom, because human motor systems possess many more component parts than are needed to make goal-directed (and relevant) movements. To regulate degrees of freedom, DST proposes that muscles act in synergies, or coordinative structures, which share a common pool of afferent and/or efferent information (Kelso, Tuller, & Harris, 1983). Thus, the distinction between feedforward and feedback processes is not explicitly delineated—action and perception are (equally) intrinsic to the dynamical system. Evidence for coordinative structures has been found in locomotion (Herman, Wirta, Bampton, & Finley, 1976), speech (Tuller,
Another fundamental question in motor control relates to whether a movement trajectory needs to be specified before movement is initiated—that is, are planning and execution distinctly separate processes? (for discussion see Schöner, 1995; Turvey & Fonseca, 2009). Whereas MP retains a distinction between planning and execution, DST questions the feasibility of such a separation, and there is human and animal evidence that supports the notion that these processes are not separated. Prablanc and Martin (1992) reported that participants made continuous and gradual adjustments to hand positioning while reaching for an object that changed position. Importantly, these adjustments were made during saccades (i.e., times of vision suppression). Similarly, Hening, Favilla, and Ghez (1988) reported a gradient distribution of movement parameters during a reaching task in which reaction time was manipulated. Both of these results indicate that a distinct separation between planning and execution is unlikely, while also supporting the view that action and perception are inextricably linked.

DST has been applied to speech production via Task Dynamics (TD) (Kelso, Saltzman, & Tuller, 1986a, 1986b; Nam, Goldstein, & Saltzman, 2009; Saltzman, 1986; Saltzman & Byrd, 2000; Saltzman & Munhall, 1989). TD assumes that the same principles used in limb research can also be applied to the movement of speech articulators (though see Grimme, Fuchs, Perrier, & Schöner, 2011 for a conceptual review of limb vs. speech movement). TD was initially introduced to reconcile a fundamental issue in speech production: how categorical units (e.g., linguistic,
cognitive) relate to the continuous nature of speech production data. DST offers a unified and lawful account of articulatory patterning and the stability with which these patterns develop in the face of external and internal perturbation (Saltzman & Munhall, 1989). A challenge in speech production research has been developing a plausible account of the timing of speech movements/gestures. TD proposes that dynamic coupling between nonlinear planning oscillators (each associated with a speech gesture) facilitates the interaction between gestures during speech (Saltzman & Byrd, 2000). Evidence supporting this view comes from perturbation studies examining relative phasing of gestures (Saltzman, Löfqvist, Kay, Kinsella-Shaw, & Rubin, 1998) and studies related to speech production errors (Goldstein, Pouplier, Chen, Saltzman, & Byrd, 2007).

One criticism of DST is the absence (and sometimes outright denial) of representations. Representations are mental constructs intended to allow humans to talk about thoughts/ideas and feelings that are not otherwise explicit. Some proponents of DST claim that representations take too great a loan on intelligence, since those representations are descriptive and not explanatory. Lindblom and Macneilage (1986) accused proponents of DST of ignoring “mental” aspects of speech and language production. Their position, and that of others (e.g., Fujimura, 1986; Grillner, 1986; Kent, 1986), is that a specific focus on measuring articulators ignores important aspects of language (e.g., phonological segments). This is a valid concern, considering action and perception do not take place in the absence of cognition. However, this should not be taken to suggest that DST ignores the reality of cognition, intentionality, goal-making, and so on (for a reply to commentators, see Kelso et al., 1986b). While many
investigators of DST have focused primarily on the movements of effectors as their level of analysis (e.g., movement of the articulators for speech), the over-arching goal of the DST approach is analysis of the system in its entirety, which necessarily includes neural, behavioral, and environmental levels (for textbook introductions, see Kelso, 1995; Thelen & Smith, 1994).

Thus, there are fundamental differences between MP and DST approaches, especially related to the distinction between motor planning and execution, level of analysis (e.g., neural vs. behavioral), and the nature (and existence) of representations. While this dissertation does not provide a comprehensive review/comparison of MP and DST approaches to motor control, there is one critical theoretical distinction that warrants further discussion: the source and nature of variability.

Variability

In most areas of research, variability is broadly defined as the inconsistency of a signal over repeated measurements of that signal. The signal, theoretically, can represent any measurement that changes in spatial characteristics over time. Since it is often more straightforward to measure those signals that are (more) easily observed, motor control research has focused on measurement of the effectors. In speech research, measurement has focused on movement of the articulators via cineradiography (e.g., Zimmermann, 1980a, 1980b), infrared motion tracking (e.g., Jackson, Tiede, & Whalen, 2013; Kleinow & Smith, 2000, 2006; Smith & Kleinow, 2000), electromagnetic articulography (Cai et al., 2011; Murdoch, Cheng, & Goozée, 2012; Perkell & Zandipour, 2002; Schönle et al., 1987; Schötz, Frid, & Löfqvist, 2013; Tiede et al., 2012; Van Lieshout & Moussa, 2000), and ultrasound (Denby & Stone,
2004; Kelsey, Minifie, & Hixon, 1969; Shawker & Barbara, 1984; Zharkova, Hewlett, &
Hardcastle, 2011), as well as combinations of these methods. With these
measurements, variability of the effectors can be directly observed, and subsequently,
inferences about system stability and flexibility can be drawn.

Regarding nomenclature, the terms *variability and stability* are often used
antonymously in motor control research. This dissertation does not take this
perspective. Instead, variability (as stated above) refers to inconsistencies based on
signal measurement, whereas stability refers to the functioning of the speech motor
system, in its entirety, such that stable systems are “tolerant of small errors in control or
small changes in the environment” (or perturbations; Liebovitch, 1998, p. 234).

*Flexibility* is a related concept that will refer to a speaker’s ability to easily transition
between states, whether these states are abstract (e.g., underlying states) or tangible
(e.g., specific gestures). Biological systems require a balance between stability and
flexibility (Spencer & Schöner, 2003). In the current work, insights into stability and
flexibility will be based on measures of (overt) variability via kinematic and acoustic
measurements.

MP approaches (e.g., Smith, Goffman, Zelaznik, Ying, & McGillem, 1995)
assume that variability at the effector level implies system-wide instability. That is, more
variability over repeated productions of the same utterance indicates that the system is
less stable, more prone to interruption, and so on. From this perspective, variability is
typically quantified as the amount that a signal of interest deviates from the mean of the
set of signals that contains the signal of interest. Proponents of MP treat variability as
noise in an otherwise (fully) stable system. That is, the CNS generates a motor
command, which is subsequently exposed to neuromotor and environmental (and other) noise (Wolpert et al., 2011). This implies that the CNS generates a “clean” signal—one that in some way precisely reflects the actor’s goal. Proponents of DST question that human motor systems, given their complexity and diverse experiences, are capable of generating this kind of flawless transmission.

Inherent in DST are the physical realities of variability, stability, and flexibility, at biological, neural, behavioral, and environmental levels. To illustrate their alternative to the MP view of variability, Riley and Turvey (2002) use the simple formula,

\[ X(t) = M(t) + N(t), \]

in which \( X(t) \) represents the intended movement (i.e., the motor program), \( M(t) \) represents the deterministic component, and \( N(t) \) represents the random component (e.g., noise). In this equation, variability is equated with randomness. In DST, what is measured as variability (or \( N(t) \) in the above formula), actually consists of a deterministic component (including chaotic processes) and a truly random (or stochastic) component. That is, variability of an effector system from a MP perspective would necessarily be treated as random, neuromotor or neuromuscular noise imposed on the command. In DST, this so-called noise would consist of both deterministic and noisy signals—the deterministic component being critical to production. These concepts are illustrated by van Lieshout and Namasivayan (2010, p. 203) in Figure 1. The upper left panel of Figure 1 shows a one-dimensional chaotic signal; the lower left panel shows a pure white noise signal. Both signals are qualitatively “noisy,” but nonlinear analysis techniques reveal a different picture. Phase space reconstruction (described in more detail below) is achieved and visualized by plotting the time series on the x-axis.
and its time-delayed copy on the y-axis. The top right panel of Figure 1, representing reconstructed phase space based on the chaotic signal, shows a parabolic curve, governed by the logistic model,

\[ x_{n+1} = 3.95 x_n (1-x_n). \]

This equation clearly represents a deterministic signal. The bottom right panel, representing reconstructed phase space based on the white noise signal, is truly random, showing no clear pattern. This is one example of how a qualitatively noisy signal could contain highly deterministic patterns. Thus, a goal in movement research is calculating deterministic components in qualitatively noisy signals by means of nonlinear analysis techniques.

Figure 1. Time series of chaotic and white noise signals (top and bottom left panels, respectively), and their corresponding phase space reconstructions (top and bottom right bottom panels). See text for more details. Figure taken from van Lieshout and Namasivayam (2010).

To be sure, effector variability and system instability are related concepts. But it is the nature of this relationship that is critical to an increased understanding of speech
production, especially in pathological systems. In MP, it is assumed that more effector variability yields less system stability—that is, there is an inverse relationship between variability and stability. From this view, the system’s inability to converge on some external effector pattern (e.g., of arm motion, lower lip displacement) is reflected by measurements of variability. In DST, this inverse relationship does not exist because effector variability consists of a deterministic component as well as a truly random (noisy) component (Li, Haddad, & Hamill, 2005; Sternad & Dijkstra, 2004). The deterministic portion by definition is part of the plan (whether that be during motor planning or goal specification). Therefore, assessing the balance of deterministic-random components of the effector’s trajectory reveals insights that a researcher would not find should he/she assume a “clean” signal (i.e., should he take a MP view). Evidence that variability consists of deterministic patterns has been demonstrated in postural control (M. A. Riley et al., 1999), gait control (Li et al., 2005), and rhythmic ball manipulation (Sternad & Dijkstra, 2004).

It is reasonable to hypothesize that all human motor systems exhibit both variability and stability. It is also reasonable to hypothesize that pathological systems are (at least some of the time) less stable than non-pathological systems. However, answering this question requires that the complexity of the human motor system be appreciated—that 1) speech motor systems are comprised of variable, stable, and flexible processes, and 2) what looks on the surface to be random, or variable, may, in fact, be deterministic. Nonlinear techniques have already revealed that pathological systems, such as those associated with Parkinson’s disease (Schmit et al., 2006) and stroke (Ghomashchi, Esteki, Nasrabadi, Sprott, & Bahrpeyma, 2011), exhibit more
deterministic patterns in postural sway. This dissertation takes a nonlinear, dynamic approach to examine stuttering, a disorder known for its hallmark variable nature. Before discussing variability and stuttering, however, it is necessary to discuss the circumstances under which the variability of motor control has been and can be assessed in PWS—that is, during perceptibly fluent motor control (i.e., fluent speech).

The Fluent Speech Paradigm

There is a long scientific history of comparing the fluent speech of PWS and PWNS. Overall, this work points to subtle differences between speech production outcomes of PWS and PWNS, though this conclusion is not straightforward. The primary motivation for measuring the fluent speech of PWS (as opposed to measuring disfluent speech) is so that meaningful comparisons of speech data between PWS and PWNS can be made. For example, comparing the overtly fluent sentence, “Buy Bobby a puppy,” and disfluent sentence, “Buy B-B-B___Bobby a puppy” would reveal substantial variability in measures related to the acoustic and kinematic speech trajectories of each sentence (e.g., sentence duration, peak velocity of the lip opening gesture for “Bobby,” stop-gap time). It would also be evident perceptually that the two utterances are different, by both professional and lay observers. Thus, it would not be possible to compare disfluent and fluent speech on the same scale. In contrast, comparing the perceptually fluent speech of PWS and PWNS allows the researcher to potentially identify subtle differences between speech patterns of PWS and PWNS.

Work within the fluent speech paradigm has primarily focused on acoustics and kinematics. Acoustically, PWS demonstrate longer voice onset time (VOT) (Agnello, 1975; Healey & Ramig, 1986; Hillman & Gilbert, 1977; Metz, Conture, & Caruso, 1979);
slower laryngeal reaction time and/or phonation initiation (Adams & Hayden, 1976; Bakker & Brutten, 1989; Cross & Luper, 1979; Cross, Shadden, & Luper, 1979; Dembowski & Watson, 1991; Lees, 1988; Peters & Boves, 1987, 1988; Peters, Hulstijn, & Starkweather, 1989; Starkweather, Franklin, & Smigo, 1984; Stromsta, 1987; Till, Goldsmith, & Reich, 1981; Watson & Alfonso, 1987); longer segment durations (Borden, Kim, & Spiegler, 1987; Bosshardt, Sappok, Knipschild, & Hölscher, 1997; Colcord & Adams, 1979; Di Simoni, 1974; McMillan & Pindzola, 1986; Starkweather & Myers, 1979; Viswanath, 1989, 1991) and more variable segment durations (Jäncke, 1994; Janssen & Wieneke, 1987; Janssen, Wieneke, & Vaane, 1983; Wieneke & Janssen, 1987, 1991); slower speech rate (Bloodstein, 1944); more centralized (Klich & May, 1982) and faster transitions of (Robb & Blomgren, 1997) formant frequencies; reduced pitch variation (Healey, 1982) and increased shimmer (Newman, Harris, & Hilton, 1989); and differences in listener perceptions of fluency (Howell & Wingfield, 1990; Love & Jeffress, 1971; Wendahl & Cole, 1961). Kinematically, PWS exhibit longer latency of movement onset, as well as longer duration between movement onset and achievement of peak velocity (Zimmermann, 1980a, 1980b); timing irregularities between lip and jaw movement (Jäncke, Bauer, Kaiser, & Kalveram, 1997; van Lieshout, Hulstijn, & Peters, 1996; Ward, 1997; Zimmermann, 1980a); irregular sequencing of articulatory movements (Almé & McAllister, 1987; Caruso, Abbs, & Gracco, 1988; Guitar, Guitar, Neilson, O'Dwyer, & Andrews, 1988; Max & Gracco, 2005); and irregularities in non-speech movements (Max, Caruso, & Gracco, 2003). For a more detailed review related to acoustic and kinematic differences between the fluent speech (and non-speech movements) of PWS and PWNS, see Bloodstein and Bernstein-Ratner (2008).
While the differences during perceptually fluent speech between PWS and PWNS suggest deficits in the speech (and non-speech) systems of PWS, there are valid criticisms of this paradigm. First, several investigations did not report group differences, for example, for VOT (Borden et al., 1987; Jäncke, 1994; Watson & Alfonso, 1982), segment durations (Healey & Adams, 1981; Jäncke, 1994), centralization of formant frequencies (Prosek, Montgomery, Walden, & Hawkins, 1987), and listener perceptions of fluency (Young, 1964). Additionally, many of the above findings have not been replicated in children (Colcord & Gregory, 1987; De Nil & Brutten, 1991; Hall, Amir, & Yairi, 1999; Krikorian & Runyan, 1983; Zebrowski, Conture, & Cudahy, 1985). Second, there is no way to determine whether fluent speech is actually fluent speech—that is, whether it is contaminated by “imperceptible stutterings” (Armson & Kalinowski, 1994; Bloodstein & Bernstein-Ratner, 2008; Ingham, 1998). According to Armson and Kalinowski (1994), the fluent speech paradigm is based on the premise that speech difficulties associated with stuttering are “ever-present” during speech production. Third, it has been argued that examining the fluent speech of PWS cannot inform the nature of speech production difficulty in PWS because it is impossible to know whether acoustic and kinematic speech characteristics are the source of or compensation for stuttering (Armson & Kalinowski, 1994; Ingham, 1998). This is a fundamental issue in etiological research in stuttering because the disorder has been measured experimentally primarily by counting symptoms (i.e., overt stuttering behaviors; for discussion, see Jackson, Quesal, & Yaruss, 2012).

However, there is a different way to view this paradigm, as well as the concerns just raised. The assumption that differences in the fluent speech of PWS are “ever-
present” is puzzling, given that stuttering is known to be a variable disorder (i.e., overt stuttering moments are inconsistent in frequency). A more parsimonious interpretation is that the speech of PWS operates on a (nonlinear) continuum from observable disfluency to observable fluency (Adams & Runyan, 1981). From this perspective, speech that is characteristic of the acoustic and kinematic differences reviewed above represents “tenuous fluency”—fluent speech that only appears atypical when sophisticated tools/approaches are implemented. Therefore, it is not surprising that there are multiple studies that have not reported group differences, because abnormal speech characteristics are NOT ever-present, but variably manifest themselves at certain times and under certain conditions. Additionally, the criticism that fluent speech is contaminated by imperceptible stutterings is misdirected. The intriguing aspect of the fluent speech paradigm is that in fact there are subtle differences between the fluent speech of PWS and PWNS, differences that may intermittently manifest themselves, and that developing ways to quantify these differences will both improve diagnostic procedures as well as inform etiological research (e.g., by examining how these differences are influenced by certain variables). Thus, the fluent speech paradigm has much to offer stuttering research because “subtle stutterings,” or segments of tenuous fluency, are precisely what researchers need to better quantify.

One way that this tenuous fluency (or subtle stuttering) has been assessed is via the quantification of variability during speech production. The following section first discusses variability in the context of stuttering, and then reviews those studies that have assessed variability in PWS and PWNS.
Stuttering and Variability

Variability is a hallmark characteristic of stuttering (for textbook discussions, see Guitar, 2013; Manning, 2009; Van Riper, 1971; Yairi & Seery, 2015). In the broadest sense, variability in stuttering has referred to the qualitatively and quantitatively inconsistent patterns with which the overt features of stuttering (i.e., disfluencies) present themselves. Indeed, variability is arguably the most problematic feature of stuttering for PWS, clinicians, and researchers alike.

One approach to quantifying variability in PWS and PWNS has been to measure the consistency of speech effectors during fluent speech production in controlled tasks. Smith and colleagues (1995) developed the spatiotemporal index (STI), a linear, amplitude- and time-normalized index of speech variability originally based on lower lip movement trajectories. STI is a “global” measure in that it was designed to assess variability during connected speech (e.g., at the phrase level and higher). This is useful because it allows investigators to examine speech production during connected speech. This is particularly applicable to stuttering, which is known to occur more during complex, meaningful speech (Manning, 2009; Van Riper, 1971). The STI approach assumes that there exists an underlying template of trajectory motion on which repeated productions of simple utterances should converge. Smith and colleagues (Smith et al., 1995; Smith, Johnson, McGillem, & Goffman, 2000) argued that if speech is pre-programmed, then repetitions of a single utterance produced by a typically fluent speaker, in the absence of perturbation, should reveal similar movement patterns (and thus yield a low STI, or high stability). While STI for production of a simple utterance (i.e., “Buy Bobby a puppy”) generally overlaps in PWS and PWNS, PWS exhibit more
within group variability (Cai et al., 2011; Kleinow & Smith, 2000; Smith & Kleinow, 2000), and children who stutter demonstrate higher STI values than typically developing peers (Smith, Goffman, Sasisekaran, & Weber-Fox, 2012). STI has also been used to assess stability in typical development (Schötz et al., 2013; Smith & Zelaznik, 2004) and Parkinson’s disease (Anderson, Lowit, & Howell, 2008; Lowit, Anderson, Dobinson, & Howell, 2008).

Howell and colleagues (2009) extended STI to audio signals and reported that acoustic STI (i.e., A-STI\(^1\)) correlated with previously reported kinematic STI values. The rationale for using an acoustic-based STI was that it could provide speech-language pathologists and researchers with an attractive alternative to using the expensive and intrusive laboratory equipment required for kinematic data collection. In their exploratory study, Howell et al. (2009) included children, which could have significantly inflated STI values due to children having less developed speech motor systems (Smith & Zelaznik, 2004). Additionally, Howell et al. (2009) did not measure whether A-STI differentiated between groups (i.e., they only reported correlations between STI and A-STI.) Thus, a replication of A-STI results is warranted.

More sophisticated measures of sentence-level stability using nonlinear normalization techniques have also been proposed. Lucero et al. (1997), and subsequently Ward and Arnfield (2001), argued that a disadvantage to linearly normalized averaging is distortion caused by trial to trial timing differences of speech landmarks (e.g., lip closure). To account for these differences, Lucero and colleagues (1997) augmented STI with a time-warping function that minimized the difference

\(^1\) Howell et al. (2009) referred to their acoustic STI as “E-STI,” to reflect incorporation of the amplitude Envelope signal. This dissertation will refer to acoustic STI as A-STI, for clarity.
between each trajectory and the mean trajectory. Two advantages of this approach are:

1) separate amplitude and phase STI values are returned (in the current study, referred to as NSTIamp and NSTIphase, respectively); and 2) natural fluctuations in speech timing are (partially) accounted for (i.e., corrected). Examining a relatively large sample of 20 PWS and 20 PWNS, Cai et al. (2011, conference poster) found that NSTIamp was higher in PWS. However, because results from Cai et al. (2011) were reported using a baseline utterance in isolation as well as embedded in longer and more syntactically complex utterances, it is difficult to determine in which context(s) NSTIamp was higher in PWS (i.e., for Base or more syntactically complex utterances, or both).

Despite several advantages to using a composite approach such as STI, there is at least one significant disadvantage: because start and end points require alignment, the original trajectory shape and time-course are altered (Lucero et al., 1997; Ward & Arnfield, 2001). Thus, important information may be lost during this normalization procedure. This is true for both linear and nonlinear versions of STI (though nonlinear versions correct for some of this distortion). Defending their STI approach, Smith, Johnson, McGillem, and Goffman (2000) argued that for their purposes (i.e., measuring the effects of contextual variables on speech), linear normalization procedures were sufficient because their questions did not require the (somewhat arbitrary) identification of speech landmarks (e.g., peak velocity, lip closure; though, they do in fact specify start and end points themselves). Smith and colleagues (2000) also reported that a pattern detection algorithm they created was 96% successful at separating waveforms into slow-, typical-, and fast-rate bins, which confirmed their assertion that the trajectory converges onto an underlying mean trajectory. However, controlling for rate effects by
separating waveforms into such bins does not address variability within the bins. More importantly, it is unlikely that the movements required for speech are planned to be exactly the same from trial to trial. For example, sentences produced with different stress or prosody, while not being any less fluent, may yield different composite values (Maner, Smith, & Grayson, 2000). Thus, stability assessment will benefit from novel approaches that measure variability within sentence/trial, and that also preserve the timing dimension so critical to speech production. Recent developments in nonlinear time series analysis have made this possible.

**Nonlinear Analysis Techniques**

For all human behavior, there are observable and unobservable components. For speech, observable components consist of motion trajectories of articulators. Unobservable components of speech are more difficult to assess, and include cognitive functions (e.g., attention, executive function), linguistic ability (e.g., semantic and syntactic processing), emotions (e.g., anxiety), and so forth. A hurdle in behavioral research is identifying and subsequently understanding the unobservable features that give rise to (observable) behavior. Recurrence quantification analysis (RQA; Marwan, Carmen Romano, Thiel, & Kurths, 2007; Webber & Marwan, 2015; Webber & Zbilut, 2005) provides techniques by which the influences of unknown variables can be revealed when only one (observable) variable is known (e.g., lip aperture trajectory).

RQA is based on Takens’ (1981) theorem, which states that by embedding time-delayed copies of time series data, and finding patterns related to the distance(s) between these copies in phase space, it is possible to learn about the dynamics of higher dimensional variables in the system under study. This theoretical perspective is
particularly useful for stuttering, a disorder for which it is possible to measure symptoms but whose causes and contributing factors are still largely unknown. A preliminary step in RQA involves the construction of a distance, and subsequently recurrence, matrix. The distance matrix requires specification of the parameters DELAY and EMBED. DELAY refers to the number of samples used to create the embedded vectors, such that $1 + \text{DELAY}$ becomes the first point of the second dimension, $1 + 2 \times \text{DELAY}$ becomes the first point of the third dimension, $1 + 3 \times \text{DELAY}$ becomes the first point of the fourth dimension, and so on. A DELAY that minimizes the amount of mutual information for a given time series should be chosen (Fraser & Swinney, 1986). EMBED refers to the number of surrogate (embedding) dimensions to be analyzed in phase space. Selection of EMBED can be guided by the false nearest neighbor methodology (Abarbanel & Kennel, 1993), which determines whether adding surrogate dimensions provides new information about the system (where % false nearest neighbors approaches zero). Effectively, EMBED is increased by integer increments until the dynamics of the system stop changing. The following example, taken from Webber (2004), demonstrates distance and recurrence matrix construction. Given the time series,

$$TS = [3.7, 9.2, 2.1, -5.4, 0.0, -10.9, 9.2, 3.1, 1.7, 1.8, -0.3, -4.9, 2.7, 3.5, 7.5, -9.9, -9.9, -4.7, 1.3, 2.7, 7.6, 3.9, 7.3, 8.0, 0.3, -1.9, 5.1, 8.8, 8.2],$$

and implementing a DELAY of 8 and EMBED of 4 (and using only the first five data points), the following time-delayed vectors are constructed:

$$V1 = [+3.7, +1.7, -9.9, +0.3]$$
$$V2 = [+9.2, +1.8, -4.7, -1.9]$$
$$V3 = [+2.1, -0.3, +1.3, +5.1]$$
$$V4 = [-5.4, -4.9, +2.7, +8.8]$$
$$V5 = [+0.0, +2.7, +7.6, +8.2]$$
Euclidean distances between these time-delayed vectors can then be calculated. For example, the Euclidean distance between V4 and V5 is calculated as:

\[ \text{Euc. } V_5-V_4 = \sqrt{(-5.5-0)^2 + (-4.9-2.7)^2 + (2.7-7.6)^2 + (8.8-8.2)^2} = 10.55 \]

The distance matrix is constructed by finding the distances for each cell in the 5X5 matrix.

- \([1,4]=18.90; [2,4]=20.67; [3,4]=9.65; [4,4]=0.00\]
- \([1,3]=12.45; [2,3]=11.83; [3,3]=0.00\]
- \([1,2]=7.88; [2,2]=0.00\]
- \([1,1]=0.00\]

Note that only the upper triangle is shown above. This is because the lower triangle mirrors the upper triangle (i.e., yields the same values). Additionally, the center diagonal, or line of identity (LOI), is represented by “0” values, since these cells represent the comparison of a vector to itself.

The rescaling option is an additional parameter setting sometimes used to shrink the magnitude of the distance matrix, should it be unmanageably large. This can be achieved by dividing each element of the distance matrix by either the mean or maximum distance of the entire matrix (Webber & Zbilut, 2005). The values in the above example were not rescaled, though the data presented in the primary experiment are rescaled based on the overall distance mean, as described in Chapter 4.

The left panel in Figure 2 presents a time series for lip aperture during speech (taken from the data in Experiment 2); the right panel provides a graphic illustration of this time series in reconstructed phase space (corresponding to the distance matrix). Parameters are DELAY = 4, EMBED = 2. Note that Figure 2 does not directly
correspond to the example above, because that data would not have produced meaningful plots due to the small number of samples.

Figure 2. Raw (registered) trajectory for one trial with sample number on x-axis and amplitude on y-axis (left). Phase space plot with data point on x-axis and data point + DELAY on y-axis (right).

To construct the recurrence matrix, it is necessary to specify the radius parameter. Radius is selected such that it falls within a range for which there is a linear scaling relation (Webber & Zbilut, 2005). Essentially, radius determines which points in the distance matrix are to be registered as *recurrent*. Thus, points that fall within the radius are given a value of 1—that is, they are recurrent. All other points are given a value of 0. Revisiting the previous example (i.e., distance matrix), the recurrence matrix, using a radius of 8, is:

\[
\begin{align*}
[1,5] &= 0; [2,5] = 0; [3,5] = 1; [4,5] = 0; [5,5] = 1 \\
[1,4] &= 0; [2,4] = 0; [3,4] = 0; [4,4] = 1 \\
[1,3] &= 0; [2,3] = 0; [3,3] = 1 \\
[1,2] &= 1; [2,2] = 1 \\
[1,1] &= 1
\end{align*}
\]

The recurrence matrix can be visualized via a recurrence plot (RP). A RP represents the times in phase space that the states of the system recur (i.e., are neighborly), based on the selected initial conditions (i.e., the parameters; Eckmann, Kamphorst, & Ruelle, 1987). Figure 3 presents two RPs, a (relatively) low-recurrence RP (left) and high-recurrence RP (right) (from Experiment 2). Note that Figure 3 does
not correspond to the example above used for distance matrix calculation, since the example would not have produced meaningful plots due to the small number of samples (i.e., only five distance vectors).

Figure 3. (Auto-)Recurrence plots based on the time series in Figure 2 with low (left) and high (right) recurrence.

RPs provide a qualitative (and visual) assessment of the time series. RQA takes information from the RP (i.e., the recurrence matrix) and quantifies those patterns using a series of algorithms that have been constructed for this purpose. This dissertation focuses on four of these algorithms/indexes: percent recurrence (%REC), percent determinism (%DET), TREND, and ENTROPY. %REC quantifies the percentage of points, out of all possible points from the distance matrix, that are deemed recurrent. %REC is included in this discussion only to highlight its importance in establishing the parameter set (i.e., recurrence should be relatively low, between 3-6%). %DET quantifies the percentage of recurrent points that contributes to diagonal lines of at least LINE length, not including the line of identity (LOI) (Webber & Zbilut, 1994). %DET is a measure of the patterned structure of recurrence of the system under study (given the
specified parameters). %DET is a critical variable because it helps to differentiate chaotic (or semi-deterministic) from truly random or stochastic processes. TREND is a measure of stationarity of the time series, or how the repeatability of the time series evolves throughout a given trial. Linear methods assume a constant level, or mean, of variability throughout a time series—that is, the system is static during production of the signal under study. Nonlinear methods do not make this assumption. Rather, nonlinear systems have mean states that are theoretically moving (M. A. Riley et al., 1999). TREND measures this movement. ENTROPY measures the signal complexity of the deterministic structure of the system. Periodic signals (such as near-sine waves) should exhibit low ENTROPY, while more complex signals will exhibit higher values. The mathematical calculations related to these four variables will be described in more detail in the *Data Collection and Analysis* section in Chapter 4.

RQA has been applied in various fields, including but not limited to: seismology, climatology and biodiversity, photosynthetic activity, and biological, physiological, and cognitive systems (for review, see Webber & Marwan, 2015). Pertinent to this dissertation, RQA has been applied in speech research (Lancia, Fuchs, & Tiede, 2014; van Lieshout & Namasivayam, 2010), and has focused on pathological (Geman, 2011; Ghomashchi et al., 2011; Schmit et al., 2006; van Lieshout & Namasivayam, 2010) and non-pathological (Barbosa, Déchaine, Vatikiotis-Bateson, & Yehia, 2012; Lancia et al., 2014; Shockley, Baker, Richardson, & Fowler, 2007) systems. Practically, the indexes returned from RQA (e.g., %DET, TREND, ENTROPY) are easily calculable, and potentially provide broader characterization of the underlying dynamics of the system.
under study (including variability, stability, and flexibility) than widely used linear analysis techniques.

In summary, knowing one variable of a system (e.g., lip aperture trajectory for the speech production system) can provide information related to the neural dynamics of speech movement, assuming that higher dimensional variables (e.g., related to cognitive and language processes) exist during speech production. The challenge of not knowing the nature and number of underlying variables, therefore, can be plausibly sidestepped by using RQA. Thus, RQA can complement existing variability measures (i.e., STI) and provide new insights into the variability and stability of stuttering and non-stuttering systems. For stuttering, it will be particularly revealing to measure stability of systems that are exposed to different contextual stressors and influences.

**Contextual Influences on Stuttering**

It is well known that the observable features of stuttering manifest themselves sporadically. Thus, researchers in stuttering have been interested in determining in which contexts and under which conditions stuttering behaviors are more likely to emerge. Most contemporary views of stuttering implicate a host of contributing cognitive-emotional and linguistic factors (Adams, 1990; Conture et al., 2006; De Nil, 1999; Namasivayam & van LIEShout, 2011b; Smith & Kelly, 1997; Starkweather & Gottwald, 1990; van Lieshout, Hulstijn, & Peters, 2004; Walden et al., 2012). The term *cognitive-emotional* is used here to reflect the difficulty in separating these systems. For example, the presence of an audience is referred to as a *cognitive-emotional* condition because it potentially elicits both cognitive (e.g., increased awareness/attention) and emotional (e.g., anxiety, fear) responses. For purposes of this dissertation, which
measures influences on speech variability and stability, it was not necessary to examine cognitive and emotional processes separately.

Existing theories and frameworks of stuttering postulate, in one way or another, that PWS possess a speech production system vulnerable to breakdown when certain demands are placed on it. These demands (depending on the framework) have ranged from anxiety and other emotional factors, to increases in linguistic complexity or time pressure. The Dynamic Multifactorial model (DMM; Smith, 1999; Smith & Kelly, 1997), for example, describes stuttering as dynamic and nonlinear, meaning that small changes in one factor (whether biomechanical, neurological, or environmental) can lead to large qualitative changes in stuttering behavior. Critically, DMM postulates that stuttering can be “ongoing” in the absence of observable features. Thus, overtly stuttered disfluencies are not a prerequisite for labeling an utterance as containing stuttering. Evidence for the DMM has been found in studies that have measured speech motor variability (via STI) under varying conditions (Kleinow & Smith, 2000; Smith et al., 2012; Smith & Kleinow, 2000). The DMM predicts that overall system stability should be lower in PWS when their systems are subjected to increasing cognitive-linguistic or emotional demands. DMM subscribes to a MP view in this sense—increased variability of effectors yields reduced system stability, and vice versa.

Van Lieshout and colleagues apply a dynamical approach to stuttering theory. In their Speech Motor Skills (SMS) framework, the speech production systems of those who stutter are the “weak link” in the chain of sub-systems that are responsible for speech production (i.e., cognitive, linguistic, emotional, etc.) (Namasivayam & van Lieshout, 2011b; van Lieshout, 2004; van Lieshout et al., 2004). SMS postulates that
the speech motor systems of PWS are at the low end of a continuum, and that typical demands on cognitive-emotional and linguistic systems may cause their speech motor systems to break down. Thus, stuttering is not solely a disorder of speech motor control. Rather, stuttering is “a reflection of an innate limitation of the speech motor control system to prepare and perform complex motor actions in the presence of cognitive, linguistic, emotional, and speech motor influences” (Namasivayam & van Lieshout, 2011b, p. 478).

SMS is a particularly useful framework among multifactorial perspectives because it makes specific predictions related to kinematics (e.g., PWS exhibit reduced upper lip amplitude.) That is, while other multifactorial theories make general predictions related to the influence of cognitive, linguistic, and emotional factors (e.g., DMM postulates that increasing demands will result in decreased global stability), SMS makes specific predictions related to the effects of these variables on measures of speech motor control. For example, SMS postulates that typical sensorimotor systems require a balance of afferent and efferent amplitude gain to maintain system stability; this amplitude is reflected in kinematic measurements of articulator motion (e.g., of lips or tongue). In PWS, destabilization occurs when this balance is not met. Specifically, PWS compensate for underlying system difficulty by reducing amplitude gain, or increasing stiffness (for example, see Van Gemmert & Van Galen, 1997). Thus, SMS predicts that measures of amplitude range (e.g., vertical displacement of lips) will be smaller (i.e., more restricted) in PWS, especially when higher-order demands (e.g., anxiety, syntactic/phonological complexity) are placed on the system. Evidence for SMS has been found in studies of speech motor control that have manipulated sentence length.

30
Cognitive-Emotional Influences

Anxiety is often discussed by clinicians and researchers as playing a significant (though not causal) role in the development of stuttering (Alm, 2004, 2014; Beilby, 2013; Blumgart et al., 2010; Bowers, Saltuklaroglu, & Kalinowski, 2012; Craig et al., 2003; Iverach & Rapee, 2013; Messenger et al., 2004; Wischner, 1952). One possible source of anxiety in human speakers is the presence of an audience, which increases communicative pressure and the potential for negative evaluation or judgment (Arenas, 2012). For PWS, the presence of an audience has been linked to increases in stuttered speech production in several studies (Commodore, 1980; Porter, 1939; Steer & Johnson, 1936; Van Riper & Hull, 1955) (though Armson, Foote, Witt, Kalinowski, & Stuart, 1997 found no significant differences). However, there have been fewer investigations of the effects of cognitive-emotional factors directly on speech motor execution. A recent study by van Lieshout and colleagues (2014) incorporated both a classical and “emotional” Stroop task to examine speech motor control in PWS and PWNS. The classical Stroop task required participants to identify the colors of numbers/words visually presented for four types of words: neutral or non-linguistic (e.g.,
“0000”); reading (i.e., varying color names in white font); congruent (i.e., font name matched color); incongruent (i.e., font name did not match color). The emotional Stroop task required participants to identify the colors of words visually presented for four groups of words: neutral (e.g., “furniture”); general threat (e.g., “murder”); general communication threat (e.g., “audience”); and individual stutter threat (i.e., based on feared words for each individual). Results indicated that PWS exhibited smaller upper lip movement ranges across tasks, and greater inter-lip phase differences during the emotional Stroop task. It is possible that smaller movement ranges are reflective of a more restrictive speech pattern in PWS; that is, PWS may adopt a more rigid speaking strategy to compensate for underlying malfunction. Evans (2009) did not report significant differences in acoustic speech stability between PWS and PWNS in the presence of an audience. However, Evans (2009) focused on analysis at the phoneme level, which may not have placed enough demand on the speech system (i.e., the task was not complex enough). Pilot work from our lab (Jackson, Tiede, & Whalen, 2013) suggests that the presence of an audience yields lower STI values in PWS (i.e., decreased variability). This work will be discussed in more detail in Chapters 4-6.

Clearly, the influence of emotional factors on speech stability is not straightforward. One strength of a DST approach is that it embraces the notion that variability is a necessary component of all living systems—not simply random noise that changes an otherwise invariant control signal. Exploring the presumed impact of cognitive-emotional factors (e.g., the presence of an audience) on speech stability in PWS and PWNS can help to clarify this relationship.
Linguistic Influences

It also appears that the speech motor systems of PWS are more susceptible to breakdown (or interference) under conditions of increased syntactic complexity. Linear STI during connected speech (i.e., sentences) revealed increased variability for PWS when a simple utterance was embedded in more syntactically complex sentences (Kleinow & Smith, 2000). Interestingly, children show this same pattern of increased speech variability (according to linear STI) when phonemic complexity is increased (i.e., longer non-words; Smith et al., 2012). Van Lieshout et al. (1995) reported lower electromyographic activity associated with the lips in PWS compared to PWNS during longer sentence production, findings interpreted as evidence for a specific speech control strategy rather than a speech control deficit. It has also been shown that increased syntactic complexity leads to increased speech variability in PWNS (Ferreira, 1991; Maner et al., 2000).

In summary, one approach to examining system stability in PWS and PWNS is measuring sentence-level variability during various conditions (e.g., cognitive-emotional stress, increased linguistic complexity), with the goal of gaining insight into where and how these components fit into a larger “multifactorial” perspective of stuttering. Due to the nature of variability measurements (i.e., measuring repeated productions of the same utterance), these investigations exclusively examined the fluent speech of PWS and PWNS. Increased understanding of what constitutes fluent speech, then, is warranted.
Chapter 3: Experiment 1

One challenge for researchers interested in examining and measuring “fluent” speech production is identifying what constitutes fluent vs. disfluent speech. Finn and Ingham (1989) highlighted these challenges more than twenty-five years ago, emphasizing the need for agreed-upon definitions of fluency (and stuttering) and guidelines for identifying a fluent segment of speech. A major reason that it is difficult to categorize fluent vs. disfluent speech is because fluency/disfluency designations are perceptual ones, and different investigator backgrounds, experiences, biases, and training are likely to contribute to varying judgments of fluency. Although the presence of disfluency is a hallmark of stuttering, the experiment described in this chapter is not concerned with differentiating stuttered vs. non-stuttered disfluencies. Rather, this study aims to increase understanding of how short utterances that may be on the border of fluent-disfluent are judged to be one or the other, so that meaningful comparisons between the speech kinematics of PWS and PWNS can be made.

Fluency is typically defined as the uninterrupted and continuous flow of speech. Fluency therefore reflects the integration of cognitive, linguistic, and speech motor processes, since all of these processes necessarily precede overt speech production. Linguistic (and cognitive) fluency is evident when a speaker: 1) talks at length with minimal pauses; 2) applies semantic and syntactic knowledge appropriately; 3) exhibits pragmatic skill (i.e., the ability to “say the right thing”); and 4) is creative and imaginative in language use (Fillmore, 1979). Difficulty in any of these areas could feasibly lead to interruptions, subtle or obvious, in the acoustic speech signal. Speech fluency concerns
the physiology of speech production, and is comprised of parameters such as rate, continuity, effort (Starkweather, 1987). These parameters are more closely related to the behaviors associated with stuttering (i.e., part-word repetitions and audible/inaudible sound prolongations). While many investigators have attempted to categorize types of fluency, the pilot study presented here examines fluency from a more general perspective. That is, judgments made in the experiment described below do not differentiate between speech and linguistic (and cognitive) fluency/disfluency.

Investigations into the perception of fluency (and disfluency) have focused primarily on distinguishing between PWS and PWNS. Several studies have shown that listeners are able to distinguish between fluent speech segments produced by PWS and PWNS (e.g., Howell & Wingfield, 1990; Love & Jeffress, 1971; Wendahl & Cole, 1961). Since subtle differences in these speech signals must be responsible for the judgments, a question that emerges is what properties of the signal allow listeners to differentiate between groups. One indicator of disfluency that may be present in (seemingly) perceptually fluent speech is the occurrence of gaps (or pauses). Gaps are evident by the absence (or low level) of acoustic energy, and can be described as the absence of “phonological linking” between the words separated by pauses (Lickley, 1994). Pause occurrence and duration can be impacted by prosodic, syntactic, task, and speaker-specific factors (Krivokapić, 2007), but generally speaking, it is easy for listeners to detect pauses (J. G. Martin & Strange, 1968). Love and Jeffress (1971) and Prosek and Runyan (1982) reported that pauses represent a distinguishing factor between the fluent speech of PWS and PWNS, and Fayer and Krasinski (1995) found that for non-native speakers, listeners differentiated groups based on the pause percentage of the total
duration of the utterance. Importantly, the three aforementioned studies examined pauses that were at least 150 ms in duration. However, data from the primary experiment of this dissertation (described in detail in Chapter 4) suggest that perceptual ambiguity related to gaps/pauses in the speech signal occurs below 150 ms.

While several studies have examined the perception of fluency/disfluency (e.g., to differentiate speaker groups, to examine when listeners perceive disfluency), no study has attempted to quantify parameters of speech that lead listeners to perceive an utterance as fluent or disfluent. The study described in this chapter examines one particular parameter of disfluency, pause/gap time, by systematically manipulating gap time and determining how these manipulations contribute to the perceptual threshold of fluency/disfluency. These results will have important implications for the study of the fluent speech of PWS and PWNS.

Methodology

Participants

Eleven participants (six female, five male) served as listeners in this study. Listeners were comprised of ten graduate students in either Speech-Language-Hearing Sciences or Linguistics, and one postdoctoral associate (n = 11), at the Graduate Center of the City University of New York (GC-CUNY). Five participants were certified speech-language pathologists (SLP). Four participants were multilingual speakers, though English was the primary language spoken by all participants. One late learner of English was excluded from this study. Additionally, three undergraduates who originally participated in the study were excluded from the current analysis because it was unclear
if they had a clear conception of fluency. All participants reported normal hearing.

Participants were recruited in the Speech Production Laboratory at GC-CUNY.

Stimuli

Stimuli consisted of two fluent and two disfluent versions of the target utterance from the primary experiment described in Chapter 4 (i.e., “Buy Bobby a puppy”; also referred to hence forth as “Base”). The two fluent tokens were produced by a female PWS and male PWNS (TF1 and TF2, respectively); the disfluent utterances were produced by the same female PWS and another male PWNS (TD1 and TD2, respectively). Fluent and disfluent utterances were determined by the examiner, a licensed speech-language pathologist (SLP), and confirmed by another licensed SLP. Selection of appropriate tokens (from the 6,720 total tokens in the primary experiment) required that there be a gap/pause with little to no noise, at the same place in the utterance, for both the fluent and disfluent utterances. This gap occurred between “Buy” and “Bobby” in all tokens, and started at cessation of voicing for /aɪ/ in “Buy” and ended at stop closure for initial /b/ in “Bobby.” Durations for the original four tokens were: TF1 = 930.00 ms; TF2 = 1,053.69 ms; TD1 = 1,139.33 ms; TD2 = 1,276.53 ms. The gaps in TF1 and TF2 were both increased by 20 ms seven times, so that there were eight tokens per utterance. The gap in TD1 was reduced by 20 ms seven times (for a total of eight tokens). Since the gap in TD2 was especially long (i.e., 325.44 ms), it was first reduced 150 ms, followed by 200 ms, and then systematically six times by 20 ms (for a total of eight tokens). Gap times were chosen based on preliminary judgments by the examiner and another rater, who identified a fluency-disfluency threshold approximately between 6-9% of the total duration of the utterance. That is, judgments changed from
fluent to disfluent at approximately this threshold for both of these raters. The manipulations were made so that of the eight versions of each utterance (i.e., the original plus seven manipulations), three utterances fell below the threshold, three fell above the threshold, and two surrounded the threshold. Table 1 provides the parameters of each utterance and subsequent manipulations.

Table 1. Stimuli, including duration of Base, duration of gap between “Buy” and “Bobby,” and the percentage of the gap of the duration of Base. Original, un-altered productions in grey.

<table>
<thead>
<tr>
<th>Speaker</th>
<th>Sentence Duration</th>
<th>Gap Duration</th>
<th>% of sentence</th>
</tr>
</thead>
<tbody>
<tr>
<td>TF1</td>
<td>930</td>
<td>23.59</td>
<td>2.54%</td>
</tr>
<tr>
<td>TF1</td>
<td>950</td>
<td>43.59</td>
<td>4.59%</td>
</tr>
<tr>
<td>TF1</td>
<td>970</td>
<td>63.59</td>
<td>6.56%</td>
</tr>
<tr>
<td>TF1</td>
<td>990</td>
<td>83.59</td>
<td>8.44%</td>
</tr>
<tr>
<td>TF1</td>
<td>1010</td>
<td>103.59</td>
<td>10.26%</td>
</tr>
<tr>
<td>TF1</td>
<td>1030</td>
<td>123.59</td>
<td>12.00%</td>
</tr>
<tr>
<td>TF1</td>
<td>1050</td>
<td>143.59</td>
<td>13.68%</td>
</tr>
<tr>
<td>TF1</td>
<td>1070</td>
<td>163.59</td>
<td>15.29%</td>
</tr>
<tr>
<td>TF2</td>
<td>1053.69</td>
<td>38.2</td>
<td>3.63%</td>
</tr>
<tr>
<td>TF2</td>
<td>1073.69</td>
<td>58.2</td>
<td>5.42%</td>
</tr>
<tr>
<td>TF2</td>
<td>1093.69</td>
<td>78.2</td>
<td>7.15%</td>
</tr>
<tr>
<td>TF2</td>
<td>1113.69</td>
<td>98.2</td>
<td>8.82%</td>
</tr>
<tr>
<td>TF2</td>
<td>1133.69</td>
<td>118.2</td>
<td>10.43%</td>
</tr>
<tr>
<td>TF2</td>
<td>1153.69</td>
<td>138.2</td>
<td>11.98%</td>
</tr>
<tr>
<td>TF2</td>
<td>1173.69</td>
<td>158.2</td>
<td>13.48%</td>
</tr>
<tr>
<td>TF2</td>
<td>1193.69</td>
<td>178.2</td>
<td>14.93%</td>
</tr>
<tr>
<td>TD1</td>
<td>1139.33</td>
<td>185.33</td>
<td>16.27%</td>
</tr>
<tr>
<td>TD1</td>
<td>1119.33</td>
<td>165.33</td>
<td>14.77%</td>
</tr>
<tr>
<td>TD1</td>
<td>1099.33</td>
<td>145.33</td>
<td>13.22%</td>
</tr>
<tr>
<td>TD1</td>
<td>1079.33</td>
<td>125.33</td>
<td>11.61%</td>
</tr>
<tr>
<td>TD1</td>
<td>1059.33</td>
<td>105.33</td>
<td>9.94%</td>
</tr>
<tr>
<td>TD1</td>
<td>1039.33</td>
<td>85.33</td>
<td>8.21%</td>
</tr>
<tr>
<td>TD1</td>
<td>1019.33</td>
<td>65.33</td>
<td>6.41%</td>
</tr>
<tr>
<td>TD1</td>
<td>999.33</td>
<td>45.33</td>
<td>4.54%</td>
</tr>
<tr>
<td>TD2</td>
<td>1276.53</td>
<td>325.44</td>
<td>25.49%</td>
</tr>
<tr>
<td>TD2</td>
<td>1126.53</td>
<td>175.44</td>
<td>15.57%</td>
</tr>
<tr>
<td>TD2</td>
<td>1076.53</td>
<td>125.44</td>
<td>11.65%</td>
</tr>
<tr>
<td>TD2</td>
<td>1056.53</td>
<td>105.44</td>
<td>9.98%</td>
</tr>
<tr>
<td>TD2</td>
<td>1036.53</td>
<td>85.44</td>
<td>8.24%</td>
</tr>
<tr>
<td>TD2</td>
<td>1016.53</td>
<td>65.44</td>
<td>6.44%</td>
</tr>
<tr>
<td>TD2</td>
<td>996.53</td>
<td>45.44</td>
<td>4.56%</td>
</tr>
<tr>
<td>TD2</td>
<td>976.53</td>
<td>25.44</td>
<td>2.61%</td>
</tr>
</tbody>
</table>
Design

Participants were seated in front of a desktop computer and instructed to put on Sony MDR-7506 Studio Headphones. Participants were initially provided with the following instructions (on a black screen with white text):

*Many people exhibit disfluencies in their speech. Sometimes these disfluencies are obvious, other times they are very subtle (e.g., pauses or hesitations). Please identify whether the sentence you hear is disfluent (by pressing “1”), or fluent (by pressing “2”).*

Participants were first required to complete five practice trials, so that they could become familiar with the task; simple instructions to do so were provided on the screen (i.e., “You will first complete five practice trials. Press ‘1’ when you are ready to begin.”) The stimuli were then presented in randomized order. For each trial, the sound file was presented, followed by a black screen for 500 ms, followed by the response screen requesting that participants select “1” for disfluent and “2” for fluent. Eight tokens (i.e., original file and seven manipulations) were presented ten times for each of the four speakers: 8 X 10 X 4 = 320 trials per participant.

Results

There was a negative correlation between gap time and percentage designated as fluent across all speakers (r = -.62, p < .01). There was also considerable variation across listeners, as visually demonstrated by confidence interval bars in Figure 4, which plots the means of the percentages of items selected fluent against gap duration for each token across all four speakers/utterances (i.e., TF1, TF2, TD1, TD2). It appears, based on these graphs, that there is a continuous fluency-disfluency threshold, ranging from 50 ms to 110 ms (though as evident from the graphs, the continuum extends this 60 ms range). In any case, it is clear that this threshold lies below 125 ms, which is
lower than the lowest gap time reported in previous studies (i.e., 150 ms). Additionally, since the speakers exhibited different durations (e.g., see Table 1), and it is likely that the gap/pause threshold is at least partially dependent on speaker rate, percent gap time was also calculated (see Figure 5). As expected, there was a negative correlation between percent gap time (percent of total duration of utterance) and percentage designated as fluent across all speakers ($r = -.67, p < .01$). The threshold for percentage of gap duration also appears to be continuous, ranging from 6-10% (though again, this range does not represent all values).

**Figure 4.** Percentages of items selected as fluent against gap duration, for each token across all participants, for TF1 (top left), TF2 (top right), TD1 (bottom left), and TD2 (bottom right). Error bars represent 95% confidence intervals.
Visual inspection of Figures 4 and 5 (which are based on means) indicates that SLPs exhibited a lower threshold for disfluency—they were faster to mark an utterance as disfluent when gap duration decreased—than Non-SLPs (i.e., the blue lines are consistently lower than the red lines). However, there was significant variation across participants and groups, as indicated by the 95% confidence interval bars. To further probe these differences, linear mixed-effect models for each utterance with group and gap duration as fixed factors and participant and token as random factors, were computed (this statistical approach is described in detail in Chapter 4). Results did not indicate significant differences between SLPs and Non-SLPs.
Discussion

This small study examined how gap (or pause) duration impacts a listener’s perception of fluency. Gap duration was chosen as the variable of interest because it appeared to be the most salient factor in differentiating fluent/disfluent utterances for the primary experiment of this dissertation (described in Chapters 4-6).

Results indicated that most tokens yielded a (semi-)gradual decline in fluency judgments as gap duration increased (evident by viewing Figures 4 and 5). This supports prior claims (i.e., Adams & Runyan, 1981) that speech fluency operates on a continuum, though future work could more rigorously test these claims. Furthermore, listeners exhibited variability with respect to how they perceived fluency/disfluency. For example, generally speaking, P00 (SLP), P01 (SLP), P10 (non-SLP), and P11 (non-SLP) exhibited a low threshold for perceiving disfluency, suggesting an increased sensitivity to disfluency, while P02 (non-SLP), P09 (non-SLP), P12 (SLP), and P14 (SLP) exhibited a higher threshold. Interestingly, SLP (vs. non-SLP) status was not associated with lower thresholds. This is perhaps unsurprising in that all participants/listeners were speech scientists (albeit mostly graduate students), and likely represented a group that is generally more sensitive to identifying differences in speech signals. Furthermore, SLPs are more trained to identify stuttering-like disfluencies, not typical disfluencies. It is possible that identifying stuttering-like disfluencies, if that was the task, would have yielded group differences between SLPs and non-SLPs. It should be noted that the small sample (i.e., n = 11) may have contributed to null findings. Future studies should examine these differences in larger samples.
Above all, the current findings highlight the importance of establishing guidelines for making fluency/disfluency judgments in speech motor control studies that examine fluent speech, and especially those studies that examine disordered populations for which there may be subtle differences in speech signals. Prior studies examining the fluent speech of PWS and PWNS have relied exclusively on experimenter judgments (or those by an outside SLP) to determine fluent vs. disfluent speech. This process should be adequate for those studies interested in measuring frequency or qualitative characteristics of stuttering-like disfluencies. However, speech motor control studies interested in examining subtle differences that are imperceptible to the human ear or eye need to pay more attention to characterizing fluency (and disfluency). This is especially the case for studies attempting to measure variability over repeated trials, since subtle variations in the timing dimension of one trial can impact variability measures (e.g., Lucero, 2005). While the small study presented in this chapter does not provide a reliable and quantitative procedure to identify fluency status, it does suggest that examiners should 1) use consistent parameter selection (e.g., percentage or absolute gap time) within and perhaps across experiments, and 2) acknowledge the inherent difficulty in differentiating fluent vs. disfluent speech, and that other factors in addition to gap time (e.g., prosody) contribute to ratings. Thus, other parameters should be systematically studied. The primary experiment, described in the next chapter, used a 6% gap time threshold for excluding “disfluent” utterances. That is, if an utterance contained a gap or pause longer than 6% of the total duration of the utterance (i.e., “Buy Bobby a puppy”), that utterance was considered disfluent and excluded from analysis. 6% was selected because, based on visual inspection of Figure 5, it represented the
approximate threshold at which % Fluent dropped below 80 (and was deemed here to represent uncertainty in listener judgment). While somewhat preliminary, this criterion permitted consistency in judgment across all participants and conditions in the main experiment.
Chapter 4: Experiment 2

This chapter describes and presents results from the primary experiment, which examined speech variability, stability, and flexibility in PWS and PWNS using (linear and nonlinear) kinematic and acoustic approaches. Previous research has demonstrated, using linear techniques (i.e., STI), that PWS produce more variable speech movements than PWNS when linguistic stressors (i.e., grammatical complexity) are placed on the system. The current investigation will seek to replicate past findings, as well as employ nonlinear analysis techniques to enhance current understanding of speech dynamics in PWS and PWNS. Additionally, this experiment will test the effect of one cognitive-emotional stressor (i.e., presence of an audience) on lip aperture dynamics in PWS and PWNS. It is expected that the approaches used here will reveal novel information related to the quantification and underlying nature of speech dynamics in PWS and PWNS—demonstrating that while PWS exhibit more surface variability related to speech movements, these movements are also associated with complex deterministic structure, and reduced stability and flexibility.

This research protocol was approved by the Institutional Review Board of The Graduate Center of the City University of New York and the National Stuttering Association Research Committee.
Methodology

Participants

This study enrolled 24 PWS and 21 PWNS (i.e., controls), matched for age and sex. Three PWS were excluded because they did not produce at least 10 fluent trials for each sentence-condition set. Additionally, one PWS was excluded due to technical malfunction (i.e., data did not record for unknown reason(s)). Thus, the current analyses included 20 PWS (6 female) and 21 PWNS (7 female) (PWS: M =27.4, SD = 6.9; PWNS: M = 25.3, SD = 2.5). All speakers reported that English was their primary language, and all speakers reported learning English before six years old. Multilingual speakers were not excluded, as it was determined that the benefits of including them (e.g., larger sample, more heterogeneous group) outweighed any potential confounds (e.g., decreased language and/or speech ability due to less exposure to English). No participants exhibited a positive history of speech-language (other than stuttering for the PWS group; see next paragraph for diagnosis of stuttering), hearing, neurological, or psychological impairment. Speech-language abilities were assessed based initially on self-report, and then informally by the examiner (a licensed SLP). Additionally, the Expressive One-Word Picture Vocabulary Test – 3rd Edition (EOWPVT-3; N. A. Martin & Brownell, 2000) was informally administered as a screener of global language skills. The EOWPVT-3 is normed up to 18-11 years, and a raw score of 116 on the test for the highest age group (i.e., 18-8 to 18-11 years) converts to a standard score of 85, which is one standard deviation below the mean. Though scores could not be standardized due to the ages of participants in this study, raw scores served as an approximation of expressive language skills, and all participants received a raw score of at least 116. All
participants also passed a pure-tone hearing screening at 500, 1,000, 2,000 and 4,000 Hz at 20 dB HL. Identification of psychological or neurological impairment was based on self-report.

Both the Overall Assessment of the Speaker’s Experience of Stuttering (OASES; Yaruss & Quesal, 2008) and the Stuttering Severity Index – 4th Edition (SSI-4; G. D. Riley, 2009) were used to assist the examiner in determining whether to include participants as PWS. The OASES measures the subjective experience of stuttering; the SSI-4 measures severity of overt features of stuttering. Stuttering is not always observable (e.g., when a speaker chooses not to speak if he or she is about to stutter), and the OASES may reflect aspects of this information that can only be determined by the speaker him- or herself. Conversely, not all PWS exhibit covert reactions to stuttering, and the SSI-4 measures the more salient behaviors associated with stuttering (e.g., speech disfluencies). Thus, using both the OASES and SSI-4 protocols contributed to a more balanced assessment of stuttering than using either protocol independently. Additionally, the examiner obtained a detailed case history from each participant. Ultimately, the diagnostic protocols and the case history/interview assisted the examiner, an SLP with more than five years of experience and a particular focus working with individuals who stutter, in making a decision to include any participant as a PWS. This diagnostic approach was followed because the OASES and SSI-4 alone are not sufficient, especially because stuttering is known to be variable by situation/context, and because the tests were administered during one testing session (i.e., during one 60-120 minute interval). Furthermore, some PWS are characterized as “covert”—meaning that those speakers may be perceived as typically fluent speakers by even the
most skilled of observers. For example, in the current study, P10 received a score of 9 on the SSI-4. This score could be associated with a typically fluent speaker. However, P10 described vividly the experience of stuttering during the interview (e.g., “I’ll change words that I know I’m going to block on,” “I’ll avoid certain situations if I think I’m going to stutter”). Thus, the categorization of PWS vs. PWNS was best made by combining administration of established protocols, case history, and an interview conducted with an experienced SLP in the area of stuttering (i.e., the examiner). For a summary of participant characteristics, see Table 2.

A survey of responses to the anticipation of stuttering was also administered. This is described in detail in Jackson et al. (in press).

Stimuli

Stimuli were adapted from Kleinow and Smith (2000; see Table 3). The target phrase, “Buy Bobby a puppy,” produced in isolation was labeled Base. To address effects of utterance complexity, it was also embedded in one “longer-only” sentence (L1; i.e., “Four one three two five Buy Bobby a puppy ten eight nine eleven”), and two longer and more linguistically complex sentences (see below for explanation of linguistic complexity). The longer-only sentence was intended to lengthen the sentence with minimal linguistic complexity and a reduced probability of fluency enhancement due to rote counting (thus the numbers were shuffled). The two longer and more complex sentences followed perspective embedment guidelines (Whalen, Zunshine, & Holquist, 2012; Zunshine, 2006), which speak to mental states of actors, so that each state adds an additional level of perspective. For example, “she wanted to go the store” contains
Table 2. Participant characteristics, including sex, age, and OASES and SSI-4 scores.

<table>
<thead>
<tr>
<th>Participant</th>
<th>Group</th>
<th>Sex</th>
<th>Age</th>
<th>OASES</th>
<th>SSI-4</th>
</tr>
</thead>
<tbody>
<tr>
<td>P01</td>
<td>PWS</td>
<td>M</td>
<td>35</td>
<td>3.61</td>
<td>32</td>
</tr>
<tr>
<td>P02</td>
<td>PWS</td>
<td>M</td>
<td>22</td>
<td>1.61</td>
<td>30</td>
</tr>
<tr>
<td>P03</td>
<td>PWS</td>
<td>M</td>
<td>25</td>
<td>1.73</td>
<td>21</td>
</tr>
<tr>
<td>P04</td>
<td>PWS</td>
<td>M</td>
<td>31</td>
<td>3.18</td>
<td>24</td>
</tr>
<tr>
<td>P05</td>
<td>PWS</td>
<td>M</td>
<td>33</td>
<td>2.32</td>
<td>22</td>
</tr>
<tr>
<td>P06</td>
<td>PWS</td>
<td>M</td>
<td>18</td>
<td>2.35</td>
<td>33</td>
</tr>
<tr>
<td>P07</td>
<td>PWS</td>
<td>M</td>
<td>23</td>
<td>2.53</td>
<td>43</td>
</tr>
<tr>
<td>P08</td>
<td>PWS</td>
<td>M</td>
<td>27</td>
<td>2.59</td>
<td>29</td>
</tr>
<tr>
<td>P09</td>
<td>PWS</td>
<td>M</td>
<td>36</td>
<td>3.02</td>
<td>17</td>
</tr>
<tr>
<td>P10</td>
<td>PWS</td>
<td>M</td>
<td>27</td>
<td>1.46</td>
<td>9</td>
</tr>
<tr>
<td>P11</td>
<td>PWS</td>
<td>M</td>
<td>22</td>
<td>2.2</td>
<td>27</td>
</tr>
<tr>
<td>P12</td>
<td>PWS</td>
<td>M</td>
<td>28</td>
<td>2.03</td>
<td>11</td>
</tr>
<tr>
<td>P13</td>
<td>PWS</td>
<td>M</td>
<td>24</td>
<td>1.48</td>
<td>17</td>
</tr>
<tr>
<td>P14</td>
<td>PWS</td>
<td>M</td>
<td>26</td>
<td>2.85</td>
<td>19</td>
</tr>
<tr>
<td>P15</td>
<td>PWS</td>
<td>F</td>
<td>26</td>
<td>2.17</td>
<td>26</td>
</tr>
<tr>
<td>P16</td>
<td>PWS</td>
<td>F</td>
<td>27</td>
<td>1.4</td>
<td>19</td>
</tr>
<tr>
<td>P17</td>
<td>PWS</td>
<td>F</td>
<td>49</td>
<td>1.81</td>
<td>17</td>
</tr>
<tr>
<td>P18</td>
<td>PWS</td>
<td>F</td>
<td>21</td>
<td>1.77</td>
<td>27</td>
</tr>
<tr>
<td>P19</td>
<td>PWS</td>
<td>F</td>
<td>24</td>
<td>2.17</td>
<td>26</td>
</tr>
<tr>
<td>P20</td>
<td>PWS</td>
<td>F</td>
<td>24</td>
<td>1.98</td>
<td>8</td>
</tr>
<tr>
<td>P21</td>
<td>PWNS</td>
<td>M</td>
<td>28</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>P22</td>
<td>PWNS</td>
<td>M</td>
<td>26</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>P23</td>
<td>PWNS</td>
<td>M</td>
<td>25</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>P24</td>
<td>PWNS</td>
<td>M</td>
<td>26</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>P25</td>
<td>PWNS</td>
<td>M</td>
<td>30</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>P26</td>
<td>PWNS</td>
<td>M</td>
<td>19</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>P27</td>
<td>PWNS</td>
<td>M</td>
<td>30</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>P28</td>
<td>PWNS</td>
<td>M</td>
<td>23</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>P29</td>
<td>PWNS</td>
<td>M</td>
<td>25</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>P30</td>
<td>PWNS</td>
<td>M</td>
<td>22</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>P31</td>
<td>PWNS</td>
<td>M</td>
<td>27</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>P32</td>
<td>PWNS</td>
<td>M</td>
<td>26</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>P33</td>
<td>PWNS</td>
<td>M</td>
<td>25</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>P34</td>
<td>PWNS</td>
<td>M</td>
<td>26</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>P35</td>
<td>PWNS</td>
<td>F</td>
<td>25</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>P36</td>
<td>PWNS</td>
<td>F</td>
<td>24</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>P37</td>
<td>PWNS</td>
<td>F</td>
<td>24</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>P38</td>
<td>PWNS</td>
<td>F</td>
<td>26</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>P39</td>
<td>PWNS</td>
<td>F</td>
<td>23</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>P40</td>
<td>PWNS</td>
<td>F</td>
<td>25</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>P41</td>
<td>PWNS</td>
<td>F</td>
<td>26</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>

Table 3. Stimuli adapted from Kleinow and Smith (2000), following perspective embedment guidelines.

<table>
<thead>
<tr>
<th>Code</th>
<th>Sentence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Base</td>
<td>Buy Bobby a puppy</td>
</tr>
<tr>
<td>L1</td>
<td>Four one three two five buy Bobby a puppy ten eight nine eleven</td>
</tr>
<tr>
<td>P1</td>
<td>He wants Karen to tell John to buy Bobby a puppy at my store</td>
</tr>
<tr>
<td>P2</td>
<td>You want Samantha to buy Bobby a puppy now if he wants one</td>
</tr>
</tbody>
</table>
one level of embedment (i.e., she wanted), whereas, “He believed that she wanted to go to the store” contains an additional level (i.e., he believed and she wanted). The stimuli here included, “He wants Karen to tell John to buy Bobby a puppy at my store” (P1; level 1 perspective embedment), and, “You want Samantha to buy Bobby a puppy now if he wants one” (P2; level 2 perspective embedment). Levels of embedment can coincide with increases in grammatical/syntactic complexity, but do not have to. Using reading time as a proxy for complexity, Whalen et al. (2012) found that greater levels of embedment were indeed more complex. Although the current experiment was not broad enough in scope to test levels of embedment fully, this new source of complexity in language presented itself as a potentially valuable addition to the current discussion. The three embedded sentences (L1, P1, P2) each contain 17 syllables, so that length effects can be separated from syntactic/grammatical or perspective embedment effects.

**Experimental Design**

Each experimental session lasted approximately 90-120 minutes, including diagnostic testing and signing of consent forms. Diagnostic testing (i.e., case history, interview, SSI-4 and OASES administration) was administered by the investigator, a New York State licensed and American Speech-Language-Hearing Association (ASHA) certified SLP. After diagnostic testing, participants were seated in a chair approximately two meters from an Optotrak Certus 3020 (Northern Digital, Waterloo, Ontario). The Optotrak is a commercially available system used to track movement (in this case, upper and lower lip movement) in three dimensions. The system tracks movement by using three cameras to triangulate the location of infrared light emitting diodes (IREDS), which are affixed to the articulators of interest. The current study focused on lip aperture
(i.e., the Euclidean distance between the upper and lower lip IREDS), so head-correction procedures were not necessary. Two IREDS were placed at midline of the vermilion border of the upper and lower lips. An Audio-Technica MicroSet directional microphone with an AT8539 Power Module on a boom stand was placed ~20 cm in front of the participant’s mouth for audio recording, set at an angle at which it did not obstruct the space between the IREDS and the camera. It was possible for the investigator to continuously monitor mouth-to-microphone distance because he had a direct line of sight to the microphone. Participants were reminded between trials to “try” to keep their head relatively straight, if the investigator noticed a change in mouth-to-microphone distance.

Stimuli were presented on a 20” monitor (Dell - ST2320L Full HD LED Widescreen) using Presentation software (Neurobehavioral Systems). The monitor was placed approximately 12”-16” directly next to the Optotrak camera, which minimized potential interference emitted from the screen. Before data collection, participants were instructed to test the range of IRED detectability by moving their heads to the left and right; they were provided with verbal feedback when this movement caused the IREDS to go out of range. IREDS were monitored via First Principles (Northern Digital, Waterloo, Ontario), the proprietary software for Optotrak data collection. Participants were instructed to attempt to remain stationary during the experiment, though minimal movement was permitted as long as IRED view was not obstructed (see below regarding trials that were discarded due to IRED obstruction).

Participants were verbally instructed to use a typical speaking voice during the experiment. Then, after receiving a simple set of instructions via the monitor (i.e.,
“Please read the sentences as they appear on the screen”), the four sentences were presented twenty times each in pseudo-randomized order (for a total of 80 trials). The entire sequence was repeated in pseudo-randomized order for the “audience” condition (for a total of 160 trials). Audience and non-audience conditions were counter-balanced; half of the participants were exposed to the audience condition first, the other half, the non-audience condition first. During the audience condition, two unfamiliar observers, one adult male and one adult female, entered the testing room. The examiner provided observers with a "ballpark" time that they should enter the testing room. Since it was impossible to determine an exact time prior to the experiment, the examiner sent both observers a text message instructing them to enter the room at the appropriate time. The observers entered the room quietly and sat in two chairs directly behind the participants, so that participants were unable to see them. This step was taken to minimize potential biases related to gender, appearance, etc. Observers coughed three to five times throughout the block so that participants could realize that the observers were an adult male and an adult female. Only two participants failed to realize that there was a male and a female (both of these participants reported that there were two males). In addition to coughing, observers were instructed to scribble audibly on a pad approximately ten times during the session, with the goal of increasing cognitive-emotional stress on the participants.

During the experiment, sentences appeared on the monitor for five seconds. Each sentence was preceded by one second of silence and a blank screen. The investigator was able to delay trials if stuttering or other interruptions occurred; however, all participants either completed the sentences stuttered or non-stuttered within 6 s or
discontinued speech production when the blank screen appeared. After both conditions of kinematic and acoustic data collection, participants completed a short questionnaire assessing their ability to identify gender of the observers (after observers left the room), as well as subjective ratings of anxiety during the audience and non-audience conditions, using a Likert-scale questionnaire developed by the examiner (see Appendix).

Twenty productions of each utterance were collected to increase the probability that participants produced at least 10 fluent (use-able) utterances during each condition for all sentences. Ten trials have been used to calculate STI in past studies (e.g., Drome, Boyce, & Channell, 2014; Kleinow & Smith, 2000, 2006; MacPherson & Smith, 2013; Smith & Kleinow, 2000), and the first 10 trials were used here, unless otherwise indicated. Only fluent productions were used in the current analysis. Fluent/disfluent utterances were marked online by the examiner, as well as verified offline by the examiner and an additional licensed and certified SLP. Fluent utterances were those free from atypical and typical disfluencies, hesitations, pauses, interjections, re-wording, and aberrant prosody (as in Kleinow & Smith, 2000). Still, this represents an observational criterion. This served as motivation for the small study presented in Chapter 3. Other than obvious stuttering-like disfluencies in the current data set (i.e., part-word repetitions, blocks, prolongations), the most salient disfluencies were pauses (or hesitations). Thus, to further automate categorization of fluency/disflueny, utterances containing pauses that were longer than 6% of the total duration of the registered utterance were marked disfluent and excluded. These pauses were found between and within words. Though somewhat arbitrary, this criterion 1) provided for
consistency throughout the current analysis, and 2) set a precedent for other studies examining speech kinematics in PWS and PWNS. Regarding L1, P1, and P2, if the target utterance (i.e., “Buy Bobby a puppy”) was considered fluent, disfluency exhibited at other parts of the sentence did not preclude inclusion of the target utterance. Despite research that suggests excluding these utterances because of the potential influence of stuttering on surrounding kinematics (e.g., Pindzola, 1986; Prosek & Runyan, 1982; Shapiro, 1980), this study elected to examine all utterances that met the criteria set forth for fluency. The current examination is interested in revealing subtle differences in speech signals that appear on the surface to be typical. Examining utterances that are perceptually fluent but that are surrounded by clear instances of stuttering may yield information regarding the speech motor processes associated with a stuttering system that produces observably fluent speech. That is, by being near confirmed stuttering, the probability for atypical patterns in the observably fluent speech may increase.

Furthermore, excluding “fluent” utterances that are surrounded by stuttering begs the question, how close to (or far away from) stuttering does an utterance have to be to be considered contaminated. It was more straightforward to include ALL perceptually fluent target utterances in the current analysis. Of 2,980 PWS trials, 200 (or 6.7%) were disfluent and 56 (or 1.9%) yielded technical errors. Of 3,360 PWNS trials, 81 (or 2.4%) were disfluent and 42 (or 1.3%) yielded technical errors.

**Data collection and analysis**

Two types of data were collected (kinematic and acoustic). Kinematic signals were sampled at 250 Hz and subsequently low-pass filtered with a 3-order Butterworth filter at 10 Hz. IREDS for the first five PWS participants were sampled at 100 Hz, due to
experimenter error, and were subsequently up-sampled to 250 Hz (by up-sampling by 5, then down-sampling by 2). Acoustic signals were digitized at 16.5 kHz and hardware filtered at 7.5 kHz. The Optotrak Data Acquisition Unit (ODAU; Northern Digital, Waterloo, Ontario) synchronized all kinematic and audio signals. The analog (audio) data required conversion from voltage to waveform audio file format (.wav) prior to data processing. Custom functions in MATLAB (Mathworks, 2013), written by Mark Tiede (committee member) and in some cases written or altered by the investigator, were implemented for all data collection and STI analyses.

Lip aperture (LA) was calculated as the Euclidean distance over time between the upper and lower lip IREDS. To register start and end points in kinematic trajectories, audio files were first manually labeled to mark the target utterance (i.e., “Buy Bobby a puppy”), ensuring that the marking for the beginning of the utterance preceded “Buy,” and for the end of the utterance followed “puppy.” Since acoustic and kinematic files were synchronized, it was possible to transpose markings from the audio to kinematic files. To extract the registered target utterance from kinematic trajectories, a three-point central differencing method was used to first determine LA velocity at each sample. The beginning of the utterance was subsequently registered at peak velocity of the first opening movement (i.e., release of /b/ in “Buy”); the end of the utterance was registered at peak velocity of the last opening movement (i.e., the release of the second /p/ in “puppy”). Figure 6 illustrates both the raw (left) and registered (right) kinematic trajectories for one trial.
Figure 6. Raw (left) and registered (right) trajectories for one trial. Red lines represent peak velocity points following the first /b/ in “Bobby” and the /i/ in “puppy” (i.e., the registered start and end points).

**STI**

Separate analyses were performed for LA-STI for the first 10 records for each condition and sentence combination (i.e., eight measures per speaker) and for the last 10. The first 10 and last 10 records were used to assess familiarity or practice effects (only for LA-STI). Since not all sentence-condition sets consisted of 20 trials (due to disfluency, participant error, and/or technical failure), there was overlap in many of the first 10-last 10 sets. For example, if a participant produced 18 fluent trials, trials 1-10 were included in the first 10 trials, and trials 9-18 were included in the last 10 trials (i.e., trials 9 and 10 would overlap). Most speakers produced between 18-20 use-able utterances (overall mean = 18.8 use-able trials). Three PWS speakers exhibited a use-able utterance average of less than 17 (i.e., 14.9, 16.5, 16.1).

To calculate LA-STI, lip aperture signals were amplitude- and time-normalized following Smith et al. (1995). To normalize for amplitude, the mean was subtracted from each amplitude value of the trajectory and then divided by the standard deviation (SD). Time normalization was achieved through linear interpolation of the amplitude-normalized signal onto a consistent time-base of 1,000 points. The SDs were then
calculated for the 10 waveforms at 2% intervals for each condition. The sum of these 50 SDs for each sentence and condition combination resulted in LA-STI. A-STI was calculated similarly to LA-STI, with the trajectories being acoustic, not kinematic. This involved calculating the root-mean-square (RMS) amplitude of the acoustic signals based on 20 ms rectangular windows (cf. Howell et al., 2009). This signal then served as input to the LA-STI MATLAB function as described above. Figure 7 demonstrates LA-STI and A-STI methods applied to the same set of signals.

Calculations for the nonlinear STI amplitude (NSTIamp) and phase (NSTIphase) components generally followed the approach described above, with one additional step. That is, normalized alignment was determined by nonlinear optimization of a reference signal minimizing the difference between peak events across all waveforms (as in Lucero et al., 1997). After this algorithm was applied, differences in amplitude and phase for each contributing waveform from the reference were extracted, and NSTIamp and NSTIphase were calculated as above (i.e., SDs calculated at 2% intervals and summed). Figure 8 demonstrates NSTIamp and NSTIphase methods applied to the same set of signals as in Figure 7. For further details on nonlinear STI, see Lucero and colleagues (1997).
Figure 7. LA-STI and A-STI calculations presented visually. The top left panel shows 10 raw trajectories for one sentence-condition set (for one speaker); the top right panel shows the 10 corresponding RMS trajectories. The middle left and right panels show the 10 trajectories after amplitude and time normalization for LA-STI (rLA) and A-STI (eLA), respectively. The bottom panels show the SDs at 2% intervals for LA-STI and A-STI, respectively.
Figure 8. NSTIamp and NSTIphase calculations presented visually. The top left panel shows the same raw trajectories as presented in Figure 7. The top right panel shows the nonlinearly normalized trajectories, determined by an algorithm that minimized the distance between each trajectory and the mean trajectory. The middle panels show the nonlinearly normalized trajectories amplitude and phase components, respectively. The bottom panels show the SD at 2% intervals for the amplitude and phase components, respectively.
RQA

Recurrence quantification analysis (RQA) provides information related to the deterministic structure and (non-)stationarity of a system (here, a speech motor system) when only one time series of that system is measurable, and even when that time series, on the surface, appears to lack deterministic structure or be noisy. All MATLAB procedures used for the RQA calculations discussed in this section were obtained from the American Psychological Association 2014 Advanced Training Institute on Nonlinear Analysis Methods at the University of Cincinnati (Shockley, 2014a), and in some cases, altered by the investigator.

As described in Chapter 2, a preliminary step in determining RQA indexes (i.e., the dependent variables) involves parameter selection, including DELAY, EMBED, radius, and LINE length. A DELAY that minimizes the amount of mutual information for a given time series should be chosen (Fraser & Swinney, 1986). A DELAY of eight or nine seemed appropriate here, as these values represented the first local minima in the mutual information functions of 25 randomly selected trials from the data set. However, because the target signals are relatively short (i.e., ~200-250 samples), using this high a delay yielded errors in the RQA functions (e.g., zeros, 100%DET). As a result, lower DELAY values were probed, so that the RQA functions generated “good spreads” in %REC (i.e., ~3-6%) and %DET (i.e., ~80-99%). These values were obtained with DELAY set at 4.

EMBED refers to the number of surrogate dimensions to be analyzed in phase space. False nearest neighbors analysis on 25 randomly selected time series indicated that %false nearest neighbors “bottomed out” at approximately four dimensions.
However, setting this parameter to 4 yielded %REC values less than 1. This was likely due to the relatively short trajectory lengths that topologically exhibited sine-like properties. Thus, EMBED was set to 2, which generated good spreads in %REC (i.e., ~3-6%) and %DET (i.e., ~80-99%).

Radius determines which points in the distance matrix are to be registered as recurrent, and is the parameter responsible for transforming the distance matrix into a recurrence matrix—and ultimately the RP (Webber & Zbilut, 2005). Radius is selected such that it falls within a range for which there is a linear scaling relation, such that %REC values are kept relatively low (e.g., 1-5%) (Shockley, 2014b). A radius of 15% (of overall mean distance) was used for the current analysis. The final parameter, LINE, determines which points are included in the plot-based quantifications, specifically %DET (i.e., only points that are part of lines of at least LINE length are used to determine %DET). Typically, this parameter is set to 2. However, since the target trajectory (i.e., “Buy Bobby a puppy”) yielded sine-like times series’, LINE was set to 5 so that the RQA variables were not overly deterministic.

The RQA variables calculated in this study are based on distance (and subsequent recurrence) matrices. Given the one-dimensional time series (i.e., LA trajectory during production of “Buy Bobby a puppy”), and given the designated parameters (DELAY = 4, EMBED = 2, and radius = 15% of mean distance), the distance matrix was created by calculating the Euclidean distance between the time-delayed vectors. The values in each cell of the distance matrix were then rescaled by dividing them by the overall mean distance and multiplying them by 100. This rescaling procedure yielded appropriate %REC values (i.e., between 3 and 6%).
matrix was then derived by keeping all points within the set radius, and deleting all points that were not. These points were then fed into a series of algorithms, which are described next.

The following RQA variables were calculated: %REC, %DET, TREND, and ENTROPY. %REC quantifies the percentage of points out of all possible points from the distance matrix that are deemed recurrent. That is, it signifies which points fall within the established recurrence criteria (i.e., radius). Given window size $W$ (i.e., the number of samples),

$$%REC = \frac{\# \text{ of recurrent points}}{(W(W-1)/2)}.$$  

%REC was included only to establish the parameter set, and is not discussed in the Results or Discussion sections.

%DET quantifies the percentage of recurrent points that contribute to diagonal lines of at least $\text{LINE}$ length (here, 5 points), not including the LOI (Webber & Zbilut, 1994). It identifies which of the established recurrent points repeat in phase space given the radius parameter of 15%. %DET was calculated as,

$$%DET = \frac{\# \text{ points in diagonal lines}}{\# \text{ total recurrent points}}.$$  

TREND is a measure of stationarity of the time series, or how the repeatability of the time series evolves throughout a given trial. Mathematically, TREND is the slope of the least squares regression of percentage of recurrent points on long diagonals as a function of orthogonal displacement from the LOI (Webber & Zbilut, 2005). As a result, TREND is typically negative. TREND was calculated as,

$$TREND = 1000 \times \text{slope of \%local recurrence vs. displacement}.$$
ENTROPY is a measure of signal complexity. In RQA, it examines the length of the diagonals of the recurrence plot, and separates different lengths (in samples) into integer bins. Shannon's (1948) formula,

\[ \text{ENTROPY} = -\sum (P_{\text{bin}}) \log_2(P_{\text{bin}}), \]

was used to calculate the probabilities for each \( P_{\text{bin}} \) greater than LINE length of 5.

Figure 9 illustrates RQA for one production of the target utterance, “Buy Bobby a puppy.”

**Figure 9.** RQA of one trial using same registered start and end points as explained in the previous STI section/examples. The top left panel shows a registered (un-normalized) trajectory (i.e., time series); the top right panel shows this time series in reconstructed phase space, with \( x(t) \) on the x-axis and \( x(t+\text{DELAY}) \) on the y-axis. The bottom left panel shows the RP, whereas the bottom right panel specifies the parameters used and returns the RQA indexes.

- **DELAY = 4**  
- **EMBED = 2**  
- **radius = 15% mean distance**  
- **LINE (min) = 5 recurrent points**  
- **# recurrent points = 771**  
- **# lines = 53**  
- **%REC = 5.7684**  
- **%DET = 90.1427**  
- **TREND = -29.9876**  
- **ENTROPY = .6077**
Importantly, the present analyses were conducted with a second parameter set. Shockley (2014b) recommended changing one parameter (e.g., DELAY) while keeping all others constant to ensure that results are not due to artifact. Thus, a second analysis was conducted on the following parameter set: DELAY = 3; EMBED = 2; radius = 15%; rescaling = mean distance; LINE = 5.

Duration
Duration was calculated as the time in ms between the peak velocity point immediately following the release of /b/ in “Buy” and the peak velocity point immediately following the second /p/ in “puppy.”

Amplitude Range
LA amplitude range (ampRange) was calculated as the spatial difference between the first peak (i.e., during /aɪ/ in “Buy”) and first valley (i.e., during /b/ closure at start of “Bobby”) of the un-normalized, registered trajectories.

Statistical Analysis
The statistical analyses used the lme4 package (Bates, Maechler, Bolker, & Walker, 2014, p. 4) in the R statistical computing program (R Core Team, 2014) to construct linear mixed-effects models. Additionally, the LmerTest package (Kuznetsova, Brockhoff, & Christensen, 2014) was used to provide Satterthwaite p-value approximations for reader convenience. It is acknowledged that controversy exists regarding the estimation of degrees of freedom and p-values in analyzing linear mixed-models (LMM). However, the increase in use of LMM in speech and linguistics and
method articles/chapters on this subject (e.g., Baayen, 2008; Barr, Levy, Scheepers, & Tily, 2013), warranted the use of LMM for the current analysis.

LMM are regression methods that allow for the examination of (multiple) fixed and random variables concurrently. Fixed variables include those that are repeated by each participant in an experiment (e.g., condition, treatment); random effects refer to those variables that may be unique to each participant, trial, or condition (Baayen, 2008), or that simply reflect noise in the system. It is important to discuss some procedural issues before continuing to the results.

As with most statistical tests, it was important to remove “outliers.” The first such outliers were movements associated with disfluent speech; these were excluded so that meaningful comparisons could be made between repeated trials both within and across groups. Removal of disfluencies followed the guidelines outlined in Chapter 3. In total, 6.7% (or 200/2,980) of PWS utterances (not including those participants who were excluded altogether; see above) were marked as disfluent, whereas 2.4% (or 81/3,360) of PWNS utterances were marked as disfluent.

Additionally, local shape-preserving interpolation was used to correct for missing data points due to technical failure or IRED obstruction. Trials for which there were more than 25 consecutive data points missing (in the target utterance) were excluded from analysis; these included 1.6% (or 98/6,340) of total trials.

Models were fit using the restricted maximum likelihood technique. Two classes of models were built, relating to across-trial (AT) indexes (i.e., STI) and within-trial (WT) measures (i.e., RQA variables, amplitude range, duration). The model-building approach described by Baayen (2008) was followed. Regarding the AT models
(discussed first), STI measures (i.e., LA-STI, LA-NSTIamp, LA-NSTIphase, A-STI) served as the dependent variables. For illustration below, LA-STI will be used. However, any of the dependent variables can be (and were) inserted into the model.

Since previous research has indicated that PWS and PWNS exhibit different speech patterning during fluent speech production at least some of the time, a fixed factor of primary interest was group, which has two levels (i.e., PWS, PWNS). Since research also indicates that these patterns are influenced by cognitive-emotional and linguistic factors, both condition (i.e., audience/no-audience) and sentence (i.e., Base, L1, P1, P2) were included as fixed factors. Participant served as a random factor, to adjust for (generally expected) variation in intercept due to individual differences in production. These variables are included in the model that follows.

\[ \text{lm} \text{er}(\text{LA-STI} \sim \text{group} + \text{condition} + \text{sentence} + (1|\text{participant})) \]

However, to test the hypothesis that condition and sentence differentially affect STI in PWS and PWNS, it was necessary to probe interactions between group and condition and group and sentence. To do this, the log likelihoods of the models were compared using the \text{anova()} function. The \text{anova()} function compares all models to a baseline (in this case, \text{LA-STI} \sim \text{group} + \text{condition} + \text{sentence} + (1|\text{participant})). In Table 4, Akaike’s Information Criterion (AIC) increases until M3, at which point it decreases. This suggests that M3 provides the best fit. M3 also yields the lowest p-value, suggesting that including the two interactions in this model was justified. An additional measure of fit was obtained by calculating $R^2$ values (e.g., see Nakagawa & Schielzeth, 2013). This was achieved using the \text{r.squaredGLMM} function in the \text{MuMIn} package (Bartoń, 2015) in R, which returns marginal and conditional $R^2$ values. Marginal values estimate the
variance accounted for by the fixed effects of the model; conditional values represent variance accounted for by the fixed and random effects. Both marginal and condition values are highest for M3 ($R^2_{\text{marginal}} = .13$, $R^2_{\text{conditional}} = .42$), confirming that M3 was the most appropriate model given the variables and hypotheses of interest.

Table 4. ANOVA comparisons of AT models consisting of the three fixed factors (i.e., group, condition, and sentence), and varying interactions.

<table>
<thead>
<tr>
<th>Model</th>
<th>Df</th>
<th>AIC</th>
<th>BIC</th>
<th>logLik</th>
<th>deviance</th>
<th>Chisq</th>
<th>Chi Df</th>
<th>Pr(&gt;Chisq)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Base</td>
<td>8</td>
<td>1693.7</td>
<td>1723.8</td>
<td>-838.85</td>
<td>1677.7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M1</td>
<td>9</td>
<td>1695.6</td>
<td>1729.4</td>
<td>-838.78</td>
<td>1677.6</td>
<td>0.1399</td>
<td>1</td>
<td>0.70836</td>
</tr>
<tr>
<td>M2</td>
<td>11</td>
<td>1698.5</td>
<td>1739.8</td>
<td>-838.23</td>
<td>1676.5</td>
<td>1.0908</td>
<td>2</td>
<td>0.57961</td>
</tr>
<tr>
<td>M3</td>
<td>12</td>
<td>1694.2</td>
<td>1739.3</td>
<td>-835.10</td>
<td>1670.2</td>
<td>6.2638</td>
<td>1</td>
<td>0.01232 *</td>
</tr>
<tr>
<td>M4</td>
<td>18</td>
<td>1704.2</td>
<td>1771.9</td>
<td>-834.11</td>
<td>1668.2</td>
<td>1.9954</td>
<td>6</td>
<td>0.92012</td>
</tr>
</tbody>
</table>

One additional variable of interest was the subjective anxiety rating reported by each participant. There is a substantial literature base that suggests (and speculates about) influences of anxiety on stuttering, though no work has directly examined the impact of anxiety on speech motor variability in PWS and PWNS. As a result, a model including anxiety was probed, along with its interaction with group (i.e., c.(anxiety) and group*c.(anxiety)). “c” indicates that the anxiety ratings (values on a 7-point scale) were centered in order to avoid spurious correlations in the model. Centering was achieved by subtracting the overall mean from each data point without scaling (Baayen, 2008). $R^2$ values for the model only including anxiety were: $R^2_{\text{marginal}} = .15$; $R^2_{\text{conditional}} = .46$). $R^2$ values for the model including the interaction (group*anxiety) were: $R^2_{\text{marginal}} = .16$; $R^2_{\text{conditional}} = .48$. Thus, explanatory power of the model increased. However, the fixed factors condition and anxiety exhibited a significantly high degree of collinearity.
(i.e., > .75). This was not surprising in that for many speakers, it could be expected that the presence of an audience has a significant influence on the their anxiety levels. Thus, despite the marginal improvement in model fit, it was decided to not include anxiety as a fixed factor in the AT models because the research question involved the presence or non-presence of an audience. Therefore, the model template used for AT analyses was:

\[
\text{lmer}(\text{LA-STI} \sim \text{group} \ast \text{condition} + \text{group} \ast \text{sentence} + (1|\text{participant}))
\]

The WT analyses were more complicated in that trial effects needed to be accounted for. This is because there was an observation for each trial, unlike for STI, for which trials were averaged together (and which yielded one composite number for each condition/sentence set). For model illustration, the dependent variable here will be perDET_LA (i.e., %DET for lip aperture). The ‘baseline’ model was identical to the model used for AT:

\[
\text{lmer}(\text{perDET_LA} \sim \text{group} \ast \text{condition} + \text{group} \ast \text{sentence} + (1|\text{participant}))
\]

However, it is feasible that measures associated with speech motor control may be impacted by experimental familiarity (i.e., getting used to tasks) or fatigue, which would be reflected in performance over trials. Thus, to determine whether adding trial to the model provided a better fit for the data, a model including c.(trial) as a fixed factor was compared to a model without c.(trial) (See Table 5). A group*c.(trial) was not explored because there was no reason to suspect that groups performed differently according to trial. A decreasing AIC and low p-value, as well as higher R² values (i.e., R²marginal = .09, R²condition = .43) indicated that adding trial as a fixed factor improved the fit of the model. Furthermore, it is plausible that there are random trial effects by participant. A model including random slopes by participant for c.(trial) was compared to a model
without this factor (Table 6). Decreasing AIC and a low p-value justified the inclusion of the random slopes for trial by participant. This decision is further supported by a larger conditional R² value for the more complex model (R²\text{condition} = .44). The model used for the WT analyses was:

\[
\text{lmer(} \text{perDET} \sim \text{group*condition} + \text{group*sentence} + \text{c.(trial)} + (1+\text{c.(trial)}|\text{participant})\text{)}
\]

Table 5. ANOVA comparison to determine whether to include c.(trial) as a fixed factor in the AT models.

Baseline: \text{lmer(} \text{perDET_LA} \sim \text{group*condition} + \text{group*sentence} + (1|\text{participant})\text{)}

\begin{tabular}{lcccccc}
Df & AIC & BIC & logLik & deviance & Chisq & Chi Df & Pr(>Chisq) \\
Base & 14 & 24346 & 24438 & -12159 & 24318 & \\
M1 & 15 & 24306 & 24404 & -12138 & 24276 & 43.007 & 1 & 5.453e-11 ** *
\end{tabular}

Table 6. ANOVA comparison to determine whether to include random slopes for c.(trial) by participant in the AT models.

\text{lmer(} \text{perDET_LA} \sim \text{group*condition} + \text{group*sentence} + \text{c.(trial)} + (1|\text{participant})\text{)}

\begin{tabular}{lcccccc}
Df & AIC & BIC & logLik & deviance & Chisq & Chi Df & Pr(>Chisq) \\
M1 & 14 & 24346 & 24438 & -12159 & 24318 & \\
M2 & 15 & 24306 & 24404 & -12138 & 24276 & 43.007 & 1 & 5.453e-11 ** *
\end{tabular}
Chapter 5: Results from Experiment 2

Results are presented by model class: across-trial (AT) and within-trial (WT). AT models included all variations of STI. WT models permitted across-trial analysis, while allowing for within-trial measurements, and included three RQA variables (i.e., %DET, TREND, and ENTROPY), duration, and amplitude range.

STI

Results from the AT model with LA-STI as the dependent variable, are presented in Table 7. For all LME analyses, t values of greater than two are considered to represent significant findings (Baayen, 2008), though p-values are also estimated (based on t-values) and included for reader convenience. PWS exhibited higher LA-STI values overall (t = 1.89, p = .06), indicating increased effector variability for PWS. However, LA-STI based on the last 10 fluent productions for each participant, for each sentence-condition set, did not reveal a significant group difference (LME model not shown), indicating a potential practice or familiarity effect (PWS became more PWNS-like). The group*sentence P2 interaction that approached significance (t = 1.78, p < .08) indicates that PWS and PWNS may respond differently to P2. After subsetting the data to only examine P2, removing the sentence interaction (i.e., group*sentence) from the model, and correcting for multiple comparisons (i.e., α/2 = .025 new significance level), a new model (presented in Table 8) demonstrated that PWS exhibited significantly higher LA-STI than PWNS when the target was embedded in P2 compared to produced in isolation (t = 3.31, p < .002). This indicates that P2 was driving the overall LA-STI difference reported above.
Table 7. Results of a LME model with LA-STI as the dependent variable, including $R^2$ values.

Linear mixed model fit by REML
t-tests use Satterthwaite approximations to degrees of freedom ['merModLmerTest']
Formula: LA.STI_first10 ~ group * condition + group * sentence + (1 | participant)
  Data: BBAP_STI

REML criterion at convergence: 1683.7

Scaled residuals:

<table>
<thead>
<tr>
<th></th>
<th>Min</th>
<th>1Q</th>
<th>Median</th>
<th>3Q</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-2.3769</td>
<td>-0.6544</td>
<td>-0.1577</td>
<td>0.5506</td>
<td>3.0914</td>
</tr>
</tbody>
</table>

Random effects:

<table>
<thead>
<tr>
<th>Groups</th>
<th>Name</th>
<th>Variance</th>
<th>Std.Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>participant</td>
<td>(Intercept)</td>
<td>5.120</td>
<td>2.263</td>
</tr>
<tr>
<td>Residual</td>
<td></td>
<td>9.494</td>
<td>3.081</td>
</tr>
</tbody>
</table>

Number of obs: 321, groups: participant, 41

Fixed effects:

|                | Estimate  | Std. Error | df     | t value | Pr(>|t|) |
|----------------|-----------|------------|--------|---------|---------|
| (Intercept)    | 14.13149  | 0.72552    | 108.700| 19.478  | <2e-16  *** |
| groupPWS       | 1.96439   | 1.04156    | 109.660| 1.886   | 0.0619  .  |
| conditionNAud  | 0.11702   | 0.47545    | 271.530| 0.246   | 0.8058  |
| sentenceL1     | 0.08095   | 0.67239    | 271.530| 0.120   | 0.9043  |
| sentenceP1     | 1.37595   | 0.67239    | 271.530| 2.046   | 0.0417  * |
| sentenceP2     | -0.79810  | 0.67239    | 271.530| -1.187  | 0.2363  |
| groupPWS:conditionNAud | 0.32607 | 0.69220    | 271.530| 0.477   | 0.6380  |
| groupPWS:sentenceL1 | -0.26893 | 0.96902    | 271.530| -0.279  | 0.7808  |
| groupPWS:sentenceP1 | -0.80201 | 0.98157    | 271.530| -0.817  | 0.4146  |
| groupPWS:sentenceP2 | 1.72297  | 0.96902    | 271.530| 1.778   | 0.0765  .  |

$R^2$ marginal = .11; $R^2$ conditional = .42

Table 8. Results of LME model for sentence P2 with LA-STI as the dependent variable, Bonferroni corrected at $\alpha/2$, including $R^2$ values.

Linear mixed model fit by REML
t-tests use Satterthwaite approximations to degrees of freedom ['merModLmerTest']
Formula: LA.STI_first10 ~ group * condition + (1 | participant)
  Data: BBAP_STI_P2

REML criterion at convergence: 440.3

Scaled residuals:

<table>
<thead>
<tr>
<th></th>
<th>Min</th>
<th>1Q</th>
<th>Median</th>
<th>3Q</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-1.5447</td>
<td>-0.7079</td>
<td>-0.1367</td>
<td>0.4278</td>
<td>2.7618</td>
</tr>
</tbody>
</table>

Random effects:

<table>
<thead>
<tr>
<th>Groups</th>
<th>Name</th>
<th>Variance</th>
<th>Std.Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>participant</td>
<td>(Intercept)</td>
<td>4.439</td>
<td>2.107</td>
</tr>
<tr>
<td>Residual</td>
<td></td>
<td>11.420</td>
<td>3.379</td>
</tr>
</tbody>
</table>

Number of obs: 81, groups: participant, 41

Fixed effects:

|                | Estimate  | Std. Error | df     | t value | Pr(>|t|) |
|----------------|-----------|------------|--------|---------|---------|
| (Intercept)    | 12.8186   | 0.8690     | 71.470 | 14.751  | <2e-16  *** |
| groupPWS       | 4.1214    | 1.2442     | 71.470 | 3.312   | 0.00145 ** |
| conditionNAud  | 1.1467    | 1.0429     | 38.090 | 1.100   | 0.27844 |
| groupPWS:conditionNAud | -0.4424 | 1.5060     | 38.560| -0.294  | 0.77054 |

$R^2$ marginal = .20; $R^2$ conditional = .43
Despite the lack of a significant \textit{group*condition} interaction, the impact of condition warranted further investigation for two reasons. First, the models in Tables 7 and 8 do not directly probe within group differences between conditions. This question is of particular interest (i.e., Do PWS respond differently to the presence of an audience?) Second, we could expect that only those PWS and PWNS who reported a significant shift in anxiety level between the non-audience and audience conditions would alter their productions. These “shifters” were participants who reported anxiety to be at least two points higher during the audience compared to non-audience condition (using the Likert scale in Appendix). No participants reported a higher anxiety rating in the non-audience compared to audience condition (as expected). There were six shifter PWS (all male) and five shifter PWNS (three male, two female), all of whom reported a change of two, except for one PWNS who reported a change of three. As shown in Table 9, “shifter” PWS demonstrated a significantly lower LA-STI during the audience compared to non-audience condition \((t = 3.27, \ p < .003)\), indicating lower effector variability during the audience condition. This finding is notable because of the significance achieved with a small sample (i.e., \(n = 11\)). PWNS did not follow this pattern. Thus, shifter PWS exhibited higher LA-STI than shifter PWNS during the non-audience condition, but reduced these values to meet those of shifter PWNS during the audience condition (see Figure 10).
Table 9. Results of a LME model with LA-STI as the dependent variable, but only including “shifter” PWS. Bonferroni correction at $\alpha/4$ (to reflect this test and one for PWNS [not shown]).

Linear mixed model fit by REML
t-tests use Satterthwaite approximations to degrees of freedom (‘merModLmerTest’)
Formula: LA.STI_first10 ~ condition + sentence + (1 | participant)
Data: BBAP_STI_swing_PWS

REML criterion at convergence: 254.9

Scaled residuals:

<table>
<thead>
<tr>
<th>Min</th>
<th>1Q</th>
<th>Median</th>
<th>3Q</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1.34042</td>
<td>-0.70025</td>
<td>-0.08749</td>
<td>0.61347</td>
<td>1.92135</td>
</tr>
</tbody>
</table>

Random effects:

<table>
<thead>
<tr>
<th>Groups</th>
<th>Name</th>
<th>Variance</th>
<th>Std.Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>participant</td>
<td>(Intercept)</td>
<td>13.42</td>
<td>3.663</td>
</tr>
<tr>
<td></td>
<td>Residual</td>
<td>14.70</td>
<td>3.834</td>
</tr>
</tbody>
</table>

Number of obs: 47, groups: participant, 6

Fixed effects:

| Estimate   | Std. Error | df  | t value | Pr(>|t|) |
|------------|------------|-----|---------|---------|
| (Intercept)| 14.6278    | 10.5800 | 7.528  | 1.45e-05 *** |
| conditionNAud | 3.6660    | 37.0000 | 3.270  | 0.00233 **  |
| sentenceL1   | 0.7017     | 1.5654   | 36.9700 | 0.448  | 0.65660 |
| sentenceP1   | 1.5670     | 1.5654   | 36.9700 | 0.976  | 0.33535 |
| sentenceP2   | 1.5683     | 1.5654   | 36.9700 | 1.002  | 0.32291 |

$R^2$ marginal = .12; $R^2$ conditional = .54

Figure 10. Effect of audience condition on LA-STI for shifters from both groups.
The LME models for LA-NSTIamp and LA-NSTIphase used the same template as LA-STI (and, therefore, are not included as tables). Overall, PWS exhibited a higher LA-NSTIamp than PWNS (t = 2.49, p < .02), but no difference in LA-NSTIphase. However, there was a significant group*sentence P2 interaction for LA-NSTIphase (t = 2.06, p < .05). Discarding the sentence factor from the LME model to examine P2 only (and implementing a Bonferroni correction at α/2 = .025), yielded a significant difference between PWS and PWNS (t = 2.66; p < .01), indicating that PWS exhibited higher variability in phase STI than PWNS when the target was embedded in P2. These results suggest that the differences in variability for P2 may be attributable to temporal features, and not amplitude. Furthermore, the six shifter PWS exhibited lower LA-NSTIphase (t = 2.84, p < .01) in the audience compared to non-audience conditions (similar to linear STI findings), with Bonferroni correction applied (α/4 = .0125); LA-NSTIamp approached significance with Bonferroni correction applied (t = 2.12, p = .04). Both nonlinear STI findings are compelling given the low shifter sample size (i.e., six PWS). Again, shifter PWNS did not exhibit differences between conditions for nonlinear STI measures.

The LME model for A-STI also used the AT template, and therefore is not presented in a table. Similar to LA-STI, PWS exhibited higher overall ASTI than PWNS (t = 2.64, p < .01). This effect also disappeared for the A-STI calculation based on the last ten trials, suggesting again a familiarity or practice effect. For all speakers, L1 (t = 2.28, p < .03) and P1 (t = 3.91, p < .001) yielded A-STIs that were higher than Base, indicating more surface variability for the longer and more linguistically complex sentences. Similar to linear and nonlinear STI findings, shifter PWS exhibited higher A-STI in the non-audience compared to audience condition (t = 3.06, p < .005, significant
at $\alpha/3 = .02$), again indicating a decrease in effector variability when speaking in the presence of an audience. PWNS did not show a difference between conditions. Shifter PWS were also more variable in A-STIphase during the non-audience compared to audience condition ($t = 2.78, p < .01$, significant at $\alpha/5 = .01$). This pattern was not present for A-STIamp. Thus, the differences in effector variability for the shifter PWS appear to be more related to timing than amplitude.

**%DET**

Table 10 presents the results from the WT model with %DET as the dependent variable. There was a significant group difference ($t = 3.14, p < .003$), indicating that PWS, across all sentence and conditions, exhibited higher %DET than PWNS. Post-hoc tests (with Bonferroni correction at $\alpha/5 = .01$) revealed that PWS exhibited higher %DET for Base ($t = 2.71, p < .01$) and P2 ($t = 2.51, p = .01$). This differs from the findings for LA-STI, which exhibited differences for P2 only. An additional analysis was conducted with end-point registration at peak velocity after the third opening gesture (i.e., after “Bobby”). Interestingly, %DET values were consistent for this shorter segment of the utterance ($t = 5.74, p < .001$).

The significant *group*\(^{\ast}\)condition interaction ($t = -3.77, p < .001$) indicated that PWS and PWNS responded differently to the presence of an audience. To determine whether one condition or the other was driving the group difference, post-hoc tests were run for each condition. Table 11 presents the model output for the audience condition. As shown, PWS exhibit higher %DET than PWNS during the audience condition ($t = 2.55, p = .01$). There is no difference between PWS and PWNS during the non-audience condition.
Table 10. Results of LME model with %DET as the dependent variable, including c.(trial) as a fixed factor and random slopes for c.(trial) by participant.

Linear mixed model fit by REML
t-tests use Satterthwaite approximations to degrees of freedom ['merModLmerTest']
Formula: perDET.LA ~ group * condition + group * sentence + c.(trial) + (1 + c.(trial) | participant)
Data: BBAP_RQA2

REML criterion at convergence: 26771.5

Scaled residuals:
    Min      1Q  Median      3Q     Max
-6.7875 -0.5143  0.1001  0.6049  7.8025

Random effects:
  Groups     Name        Variance  Std.Dev.  Corr
  participant (Intercept) 4.1055241 2.02621
   c.(trial)  0.0006532 0.02556   0.67
  Residual      5.0734667 2.25244
Number of obs: 5924, groups:  participant, 41

Fixed effects:
                     Estimate  Std. Error df   t value Pr(>|t|)
(Intercept)          8.705e+01  4.075e-01  5.400e+01 213.594  < 2e-16 ***
groupPWS            1.589e+00  5.055e-01  4.500e+01   3.144  0.00294 **
conditionNaud       -4.807e-02  1.498e-01  2.221e+03  -0.321  0.74827
sentenceL1           1.724e-01  1.118e-01  5.833e+03   1.541  0.12330
sentenceP1           2.288e-01  1.122e-01  5.833e+03   2.039  0.04147 *
sentenceP2           -3.668e-01  1.116e-01  5.833e+03  -3.286  0.00102 **
c.(trial)            -7.176e-03  4.053e-03  4.000e+01  -1.770  0.08437 
  groupPWS:conditionNaud  -8.491e-01  2.255e-01  2.071e+03  -3.766  0.00017 ***
groupPWS:sentenceL1    -4.360e-01  1.648e-01  5.834e+03  -2.666  0.00816 **
groupPWS:sentenceP1    -5.343e-01  1.670e-01  5.834e+03  -3.199  0.00139 **
groupPWS:sentenceP2    -7.403e-02  1.649e-01  5.834e+03  -0.449  0.65340

R²marginal = .04; R²conditional = .54

To further assess within condition differences, an LME model was run for PWS and PWNS separately, with the group factor removed and a Bonferroni correction at α/4 = .0125 (see Table 12). PWS exhibited significantly higher %DET during the audience compared to non-audience condition (t = -5.07, p < .001). There were no differences between conditions for PWNS. Thus, PWS and PWNS appear to exhibit similar deterministic structure during the non-audience condition, but under cognitive-emotional stress (i.e., the audience condition), PWS become more deterministic.
Table 11. Results of LME model with %DET as dependent variable, only audience condition, and including c.(trial) as a fixed factor and random slopes for c.(trial by participant). Bonferroni correction at $\alpha/3 = .017$.

Linear mixed model fit by REML
tests use Satterthwaite approximations to degrees of freedom ['merModLmerTest']
Formula: perDET_LA ~ group * sentence + c.(trial) + (1 + c.(trial) | participant)
Data: BBAP_RQA2_Aud

REML criterion at convergence: 13085.9

Scaled residuals:
Min 1Q Median 3Q Max
-4.8763 -0.5446 0.1086 0.6489 3.1819

Random effects:
Groups Name Variance Std.Dev. Corr
participant (Intercept) 2.1067067 1.45145
(c.(trial)] 0.0001204 0.01097 -0.27
Residual 3.9653793 1.99133

Number of obs: 3057, groups: participant, 41

Fixed effects:
Estimate Std. Error df t value Pr(>|t|)
(Intercept) 8.707e+01 3.353e-01 259.694 <2e-16 ***

Table 12. Results of LME model with %DET as the dependent variable, only including PWS. Bonferroni correction at $\alpha/5$ (to reflect this test and one for PWNS [not shown]).

Linear mixed model fit by REML
tests use Satterthwaite approximations to degrees of freedom ['merModLmerTest']
Formula: perDET_LA ~ condition + sentence + c.(trial) + (1 + c.(trial) | participant)
Data: BBAP_RQA2_PWS

REML criterion at convergence: 12270.2

Scaled residuals:
Min 1Q Median 3Q Max
-6.6584 -0.5037 0.0864 0.5604 7.7614

Random effects:
Groups Name Variance Std.Dev. Corr
participant (Intercept) 6.225224 2.49504
(c.(trial)] 0.001307 0.03615 0.80
Residual 5.242283 2.28960

Number of obs: 2691, groups: participant, 20

Fixed effects:
Estimate Std. Error df t value Pr(>|t|)
(Intercept) 8.839e+01 5.706e-01 2.050e+01 154.892 < 2e-16 ***
conditionNaud -8.804e-01 1.735e-01 1.167e+03 -5.074 4.52e-07 ***
sentenceL1 -2.637e-01 1.230e-01 2.646e+03 -2.144 0.032108 *
sentenceP1 -3.068e-01 1.258e-01 2.647e+03 -2.439 0.014787 *
sentenceP2 -4.417e-01 1.233e-01 2.647e+03 -3.582 0.000347 ***
c.(trial] -1.167e-02 8.164e-03 1.900e+01 -1.429 0.169130

R2marginal = .04; R2conditional = .64
TREND

Since TREND is computed as the slope away from the LOI (i.e., to the bottom right of the plot), TREND values were negative. The more negative (i.e., the smaller) TREND values are, the greater the magnitude of TREND. Table 13 presents the results of the WT model with TREND as the dependent variable. PWS appeared to exhibit greater TREND magnitude than PWNS (t = -1.69, p < .10), indicating that PWS, overall, produced the target utterance with less stationarity, or a more volatile frame of reference.

Table 13. Results of LME model with TREND as the dependent variable.

| Estimate | Std. Error | df | t value | Pr(>|t|) |
|----------|------------|----|---------|----------|
| (Intercept) | -3.942e+01 | 2.000e+00 | 4.500e+01 | -19.708 < 2e-16 *** |
| groupPWS | -4.681e+00 | 2.768e+00 | 4.100e+01 | -1.691 0.09833 . |
| conditionNaud | -2.185e-01 | 6.233e-01 | 4.550e+02 | -0.350 0.72613 |
| sentenceL1 | 5.177e+00 | 5.140e-01 | 5.833e+03 | 10.072 < 2e-16 *** |
| sentenceP1 | 6.935e+00 | 5.156e-01 | 5.833e+03 | 13.450 < 2e-16 *** |
| sentenceP2 | 7.241e+00 | 5.131e-01 | 5.832e+03 | 14.111 < 2e-16 *** |
| c.(trial) | -1.439e-02 | 8.853e-03 | 3.800e+01 | -1.625 0.11227 |
| groupPWS:conditionNaud | 1.840e+00 | 9.345e-01 | 4.300e+02 | 1.969 0.04962 * |
| groupPWS:sentenceL1 | 2.368e+00 | 7.572e-01 | 5.835e+03 | 3.127 0.00177 ** |
| groupPWS:sentenceP1 | 4.615e+00 | 7.676e-01 | 5.835e+03 | 6.012 1.94e-09 *** |
| groupPWS:sentenceP2 | 4.489e+00 | 7.577e-01 | 5.835e+03 | 5.925 3.30e-09 *** |

There were also significant differences for sentences L1 (t = 10.07, p < .001), P1 (t = 13.45, P < .001), and P2 (t = 14.11, p < .001), indicating that for all speakers,
increased length and/or linguistic complexity contributed to lower TREND magnitude, or increased stationarity. Additionally, significant group*sentence interactions were found for all sentences, indicating that PWS and PWNS responded differently to linguistic complexity. PWS exhibited greater TREND when the target utterance was produced in isolation, but this pattern reversed when the target was embedded in linguistically complex sentences (see Figure 11 for graphical representation). However, post-hoc tests did not reveal significant differences between PWS and PWNS for any of the utterances.

Figure 11. Trend group differences broken down by sentence. Only Base approached significance (t = -1.71, p < .10).

The significant group*condition interaction for TREND warranted further investigation into within-group condition changes. Initial post-hoc analyses using the same approach as above (i.e., removing the group factor) indicated that the difference
between the audience and non-audience conditions for PWS (after Bonferroni adjustment of $\alpha/3 = .017$) approached significance ($t = 2.12, p = .03$). However, it seemed plausible that a combined effect of linguistic and cognitive-emotional complexity may be present, especially for P2, because it appeared to drive the STI results reported above. Thus, additional models that included a condition*sentence interaction, were implemented. Table 14 presents results from this LME model for PWS, with an adjusted significance level of $\alpha/5 = .01$. PWS exhibited higher TRENDS in the audience compared to non-audience condition ($t = 3.15, p < .005$). Furthermore, this difference appears to be driven by P2 ($t = 3.90, p < .001$). PWNS did not exhibit this pattern.

### Table 14. Results of LME model for TRENDS for PWS including a condition*sentence interaction (table for PWNS not shown).

Linear mixed model fit by REML  
Formula: trend_LA ~ condition * sentence + c.(trial) + (1 + c.(trial) | participant)  
Data: BBAP_RQA2_PWS  
REML criterion at convergence: 20352.9  
Scaled residuals:  
   Min  1Q Median  3Q Max  
-4.1603 -0.5827  0.0230  0.5928  3.6331  
Random effects:  
   Groups   Name        Variance Std.Dev. Corr  
   participant (Intercept) 1.176e+02 10.84581  
                       c.(trial)  3.958e-03  0.06291 0.27  
   Residual              1.075e+02 10.36628  
Number of obs: 2691, groups: participant, 20  
Fixed effects:  
   Estimate Std. Error   df t value Pr(>|t|)  
   (Intercept) -4.364e+01  2.501e+00  2.100e+01 -17.446 5.55e-14 ***  
   conditionNaud  3.112e+00  9.873e-01  8.816e+02   3.152 0.00168 **  
   sentenceL1    7.620e+00  7.578e-01  10.056 < 2e-16 ***  
   sentenceP1    1.246e+01  7.774e-01  16.030 < 2e-16 ***  
   sentenceP2    1.381e+01  7.705e-01  17.922 < 2e-16 ***  
   c.(trial)    3.476e-03  1.498e-02  3.152   0.81920  
   conditionNaud:sentenceL1 -4.686e-02  1.171e+00  2.644e+03 -0.042 0.96656  
   conditionNaud:sentenceP1  1.921e+00  1.142e+00  2.644e+03 -1.683 0.09254 .  
   conditionNaud:sentenceP2  1.362e+00  1.118e+00  2.644e+03 -3.903 9.76e-05 ***  
R^2marginal = .10; R^2conditional = .58
ENTROPY

Sentences P1 and P2 yielded significant differences in ENTROPY for all speakers (t = 5.73, p < .001, t = 4.51, p < .001, respectively; see Table 15). That is, embedding the target utterance in more linguistically complex (but not longer-only) structures increased complexity (i.e., Shannon’s entropy) of the signal (i.e., utterance). Additionally, a group*condition interaction that approached significance (t = 1.54, p < .06) warranted examination of within group condition comparisons. PWS exhibited higher ENTROPY in the non-audience compared to audience condition (t = 2.84, p < .005, with α/3 ~ .02 correction), suggesting that complexity decreased for PWS when speaking in the presence of an audience, which is perhaps a compensatory strategy.

This pattern was not present in PWNS.

Table 15. Results of LME model with ENTROPY as the dependent variable.
Linear mixed model fit by REML
t-tests use Satterthwaite approximations to degrees of freedom [‘merModLmerTest’]
Formula: rel_entropy_LA ~ group * condition + group * sentence + c.(trial) + (1 + c.(trial) | participant)
Data: BBAP_RQA2

REML criterion at convergence: -22265.2

Scaled residuals:
    Min 1Q Median 3Q Max
-3.8172 -0.6731 -0.0426 0.6083 6.0338

Random effects:
  Groups     Name        Variance  Std.Dev.  Corr
participant (Intercept) 4.219e-04 0.0205410
                c.(trial) 3.076e-08 0.0001754 -0.08
  Residual              1.286e-03 0.0358605

Number of obs: 5924, groups: participant, 41

Fixed effects:
                                Estimate Std. Error   df t value Pr(>|t|)
(Intercept)                     5.348e-01 4.777e-03 111.962 < 2e-16 ***
groupPWS                       1.048e-02 6.843e-03 1.531 0.1322
conditionNaud                  8.130e-05 2.176e-03 0.037 0.9702
sentenceL1                     5.318e-04 1.780e-03 0.299 0.7652
sentenceP1                     1.024e-02 1.786e-03 1.531 0.1322
sentenceP2                     8.010e-03 1.777e-03 0.670 0.5030
                c.(trial) 3.076e-08 0.0001754 -0.08

R² marginal = .04; R² conditional = .30
Duration

Table 16 presents results from the WT model with duration as the dependent variable. Note that the number of observations for this model is greater than the previous WT models. This is because the RQA analyses did not return values for 129 trials. As expected, PWS exhibited longer durations than PWNS across all conditions and sentences ($t = 3.26, p < .005$). There were significant differences for L1 ($t = 2.69, p < .01$), P1 ($t = -6.87, P < .001$), and P2 ($t = -14.84, p < .001$); for both PWS and PWNS, L1 yielded longer durations, whereas P1 and P2 yielded shorter durations. Additionally, a significant trial effect was present ($t = -4.01, p < .001$), indicating a decrease in duration for all speakers (across conditions) as trials progressed. All speakers also exhibited increased duration during the audience compared to non-audience condition ($t = -2.00, p < .05$), indicating that all speakers reduced their rate to some degree when in the presence of an audience.
Table 16. Results of a LME model with duration as the dependent variable.

Linear mixed model fit by REML
t-tests use Satterthwaite approximations to degrees of freedom ['merModLmerTest']
Formula: duration ~ group * condition + group * sentence + c.(trial) + (1 +
c.(trial) | participant)
Data: BBAP_RQA2

REML criterion at convergence: 66945.7

Scaled residuals:
          Min      1Q  Median      3Q     Max
-6.7811 -0.5712 -0.0773  0.4601  9.3445

Random effects:
Groups     Name        Variance  Std.Dev. Corr
participant (Intercept) 1.190e+04  109.0917
       c.(trial)  2.528e-01   0.5028 0.33
Residual                3.530e+03   59.4120
Number of obs: 6053, groups:  participant, 41

Fixed effects:
               Estimate Std. Error df  t value Pr(>|t|)
(Intercept)  831.50724   23.37506  42  35.572  < 2e-16 ***
groupPWS    106.04366   32.57866 40  3.255  0.002313 **
conditionNaud -7.83643    3.91435 1950 -2.002  0.045426 *
sentenceL1    7.92059    2.94577 5962  2.689  0.007191 **
sentenceP1   -20.32171    2.95906 5962 -6.868 7.19e-12 ***
sentenceP2   -43.69407    2.94471 5962 -14.838 < 2e-16 ***
c.(trial)   -0.32265    0.08050  39 -4.008  0.000271 ***
groupPWS:conditionNaud  12.45563    5.78332 1846  2.154  0.031391 *
groupPWS:sentenceL1   -12.49754    4.28659 5962 -2.915  0.003564 **
groupPWS:sentenceP1   -8.78076    4.35600 5963 -2.016  0.043867 *
groupPWS:sentenceP2   -0.06173    4.30461 5962  0.014  0.988559

R2marginal = .18; R2conditional = .82

Significant group*sentence interactions warranted running the WT model using a subset of this data by removing the fixed factor of sentence, and implementing a Bonferroni correction for multiple comparisons (α/5 = .01). Compared to PWNS, PWS exhibited longer target utterance duration for Base (t = 3.16, p < .005), P1 (t = 2.94, p < .006), and P2 (t = 3.46, p < .005). The insignificant finding for L1 may be due to a rhythm effect, such that speaking in rhythm potentially normalizes durations across participants. Furthermore, the significant group*condition interaction in Table 16 warranted further examination into within-group differences between conditions.

Interestingly, PWS did not differ between conditions, but PWNS exhibited a difference that approached significance (see Table 17). That is, PWNS appeared to exhibit shorter
durations during the non-audience compared to audience condition (t = -2.35, p < .02, corrected at α/7 = .007). This finding suggests that PWS do not slow down when speaking in the presence of an audience.

Correlations with Duration

Since there have been questions regarding the influence of duration on kinematic speech measures, specifically those assessing variability, Pearson correlations were calculated between the dependent variables calculated in this study and utterance duration. Table 18 summarizes these results. There was a positive correlation between duration and LA-STI (r = .50, p < .01), suggesting that despite normalization procedures (or maybe because of them), the duration of the target production might have influenced LA-STI. This pattern is also exhibited by A-STI (r = .47, p < .01). The correlations were smaller for LA-NSTIphase (r = .41) and LA-NSTIamp (r = .21), suggesting that the nonlinear time-warping function may account for some of the distortion caused by the (exclusively) linear normalization procedures. There were no correlations for %DET (r = .18) and TREND (r = .01), suggesting that these measures are not affected by durational changes.
Table 17. Results of a LME model with duration as the dependent variable, for PWNS only. Bonferroni corrected at α/7 = .007

Linear mixed model fit by REML
t-tests use Satterthwaite approximations to degrees of freedom ['merModLmerTest']
Formula: duration ~ condition + sentence + c.(trial) + (1 + c.(trial) | participant)
Data: BBAP_RQA2_PWNS
REML criterion at convergence: 34476.6

Scaled residuals:
  Min  1Q  Median  3Q  Max
-2.8620 -0.6280 -0.0910  0.4847 11.4379

Random effects:
  Groups   Name       Variance  Std.Dev. Corr
  participant  (Intercept) 6.591e+03 81.1828
        c.(trial)    9.536e-02  0.3088  0.60
  Residual             2.360e+03 48.5788
Number of obs: 3237, groups: participant, 21

Fixed effects:  Estimate Std. Error   df   t value  Pr(>|t|)
(Intercept)     833.2722   17.85960 21.00000 46.657 < 2e-16 ***
conditionNaud  -6.98438  2.97441 315.00000 -2.348 0.019485 *
sentenceL1     7.91317  2.40860 3191.00000  3.285 0.001029 **
sentenceP1    -20.31971  2.41948 3191.00000 -8.398 < 2e-16 ***
sentenceP2    -43.70309  2.40775 3191.00000 -18.151 < 2e-16 ***
c.(trial)  -0.29560  0.06992 19.00000  -4.228 0.000453 ***
R^2marginal = .06; R^2conditional = .76

Table 18. Correlations with duration. AT measures (i.e., STI measures) were based on mean duration for each sentence-condition set, while WT measures were based on utterance durations associated with each trial.

<table>
<thead>
<tr>
<th></th>
<th>LA-STI</th>
<th>LA-NSTIamp</th>
<th>LA-NSTIphase</th>
<th>A-STI</th>
<th>%DET</th>
<th>TREND</th>
<th>ENTROPY</th>
</tr>
</thead>
<tbody>
<tr>
<td>LA-STI</td>
<td>.50**</td>
<td>.21</td>
<td>.41**</td>
<td>.47**</td>
<td>.18</td>
<td>.01</td>
<td>.17</td>
</tr>
</tbody>
</table>

Amplitude Range

Table 19 presents results from the WT model with amplitude range as the dependent variable. Random intercepts by participant were included in the model to account for differing jaw sizes. Furthermore, PWS and PWNS, as groups, do not significantly differ in jaw size (Daliri, Prokopenko, & Max, 2013), so comparisons across groups were appropriate. All speakers exhibited significant differences between Base and L1 (t = -7.45, p < .001), Base and P1 (t = -19.23, p < .001), and Base and P2 (t = -15.76, p < .001), indicating that embedding the target utterance in longer and/or more
complex structures reduced amplitude range of the first gesture in the target utterance.

Significant group*sentence interactions were also found for P1 (t = 2.99, p < .005) and P2 (t = 3.17, p < .005), indicating that PWS exhibit greater amplitude range than PWNS for the initial gesture when linguistic complexity is increased. Thus, PWS do not reduce their amplitude range as much as PWNS under these conditions.

Table 19. Results of a LME model for all speakers with amplitude range as the dependent variable.
Linear mixed model fit by REML
t-tests use Satterthwaite approximations to degrees of freedom [‘merModLmerTest’]
Formula: ampRange ~ group * condition + group * sentence + c.(trial) + (1 +
c.(trial) | participant)
Data: BBAP_ampRange

REML criterion at convergence: 22105.4

Scaled residuals:
  Min  1Q  Median   3Q  Max -4.2946 -0.6243 -0.0077  0.5978 5.3359

Random effects:
  Groups     Name        Variance  Std.Dev. Corr
  participant (Intercept) 8.5300185 2.92062
  c.(trial)   0.0001727 0.01314  0.02
  Residual                2.4047893 1.55074

Number of obs: 5843, groups:  participant, 40

Fixed effects:

(Intercept) 1.520e+01  6.417e-01  3.900e+01  23.685  < 2e-16 ***
groupPWS    2.388e-01  9.313e-01  3.900e+01  0.256  0.79897
conditionNaud 1.970e-01  1.030e-01  2.153e+00  1.913  0.05590 .
sentenceL1 -5.739e-01  7.699e-02  5.754e+00  -7.454  1.04e-13 ***
sentenceP1 -1.486e+00  7.724e-02  5.754e+00  -19.234  < 2e-16 ***
sentenceP2 -1.211e+00  7.686e-02  5.754e+00  -15.758  < 2e-16 ***
c.(trial)   3.545e-03  2.132e-03  3.700e+00  1.663  0.10475

R²marginal = .03; R²conditional = .79

Since the effect of condition approached significance (t = 1.91, p < .06), it was prudent to conduct these analyses on shifters exclusively (see Table 21). A significant group*condition interaction (t = -2.71, p < .01, with α/2 = .025) indicated that the shifter PWS exhibited lower amplitude ranges than the shifter PWNS on the first gesture of the target utterance during the non-audience, but not audience, condition. Interestingly,
shifter PWS exhibited a significant reduction in amplitude range compared to PWNS during L1 production ($t = -3.79, p < .001$). This effect only approached significance when tested for all participants. Thus, only the shifter PWS seemed to be affected by the potential rhythm effect of L1.

Table 20. Results of a LME model for shifters with amplitude range as the dependent variable.

| Formula: ampRange ~ group * condition + group * sentence + c.(trial) + (1 + c.(trial) | participant) |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Data: BBAP_ampRange Swing                                                                                                                                                                      |
| REML criterion at convergence: 5432.5                                                                                                                                                          |
| Scaled residuals:                                                                                                                                                                              |
| Min 1Q Median 3Q Max                                                                                                                                                                          |
|-4.1298 -0.5876 -0.0049 0.5816 3.7708                                                                                                                                                           |
| Random effects:                                                                                                                                                                                |
| Groups Name Variance Std.Dev. Corr                                                                                                                                                    |
| participant (Intercept) 6.5695248 2.56311                                                                                                                                            |
| c.(trial) 0.0001985 0.01409 -0.43                                                                                                                                                             |
| Residual 2.3646106 1.53773                                                                                                                                                                   |
| Number of obs: 1442, groups: participant, 10                                                                                                                                                   |
| Fixed effects:                                                                                                                                                                                |
| Estimate Std. Error df t value Pr(>|t|)                                                                                                                                                        |
| (Intercept) 1.299e+01 1.104e+00 9.100e+00 11.765 8.20e-07 ***                                                                                                                            |
| groupPWS 9.845e-01 1.485e+00 8.200e+00 0.663 0.52552                                                                                                                                         |
| conditionNaud 5.133e-01 2.092e-01 5.104e+02 2.454 0.01447 *                                                                                                                             |
| sentenceL1 5.446e-02 1.570e-01 1.414e+03 0.347 0.72870                                                                                                                                         |
| sentenceP1 -1.165e+00 1.568e-01 1.414e+03 -7.428 1.89e-13 ***                                                                                                                             |
| sentenceP2 -8.086e-01 1.572e-01 1.414e+03 -5.145 3.05e-07 ***                                                                                                                             |
| c.(trial) 6.347e-04 4.554e-03 8.900e+00 0.139 0.89226                                                                                                                                 |
| groupPWS:conditionNaud -8.285e-01 3.058e-01 4.354e+02 -2.709 0.00701 **                                                                                                                   |
| groupPWS:sentenceL1 -8.598e-01 2.271e-01 1.414e+03 -3.786 0.00016 ***                                                                                                                     |
| groupPWS:sentenceP1 4.732e-01 2.301e-01 1.415e+03 2.057 0.03991 *                                                                                                                       |
| groupPWS:sentenceP2 4.134e-01 2.295e-01 1.415e+03 1.802 0.07184 .                                                                                                                                 |
| R²marginal = .03; R²conditional = .76                                                                                                                                                       |

Summary

Table 21 presents a summary of the key findings related to differences found between PWS and PWNS. These findings are separated by model class (i.e., AT, WT), and include linear (kinematic and acoustic) and nonlinear STIs, RQA variables, duration, and amplitude range.
<table>
<thead>
<tr>
<th>Across-trial (AT)</th>
<th>Within-trial (WT)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>LA-STI</strong></td>
<td><strong>%DET</strong></td>
</tr>
<tr>
<td>PWS &gt; PWNS (only for first 10 trials); P2 drives this difference.</td>
<td>PWS &gt; PWNS for all sentences; audience condition drives this difference.</td>
</tr>
<tr>
<td>Shifter PWS audience &lt; Shifter PWS non-audience; not for PWNS.</td>
<td><strong>TREND</strong></td>
</tr>
<tr>
<td><strong>A-STI</strong></td>
<td></td>
</tr>
<tr>
<td>PWS &gt; PWNS (only for first 10 trials); for P2 and Base.</td>
<td><strong>ENTROPY</strong></td>
</tr>
<tr>
<td>Shifter PWS audience &lt; Shifter PWS non-audience; not for PWNS.</td>
<td>PWS non-audience &gt; PWS audience; not for PWNS.</td>
</tr>
<tr>
<td><strong>LA-NSTIamp</strong></td>
<td><strong>Duration</strong></td>
</tr>
<tr>
<td>PWS &gt; PWNS; across sentences.</td>
<td>PWS &gt; PWNS for all utterances (except L1)</td>
</tr>
<tr>
<td>Shifter PWS audience &lt; shifter PWS non-audience (approached significance with correction); not for PWNS.</td>
<td>No condition effects.</td>
</tr>
<tr>
<td><strong>LA-NSTIphase</strong></td>
<td><strong>Amplitude Range</strong></td>
</tr>
<tr>
<td>PWS &gt; PWNS, only for P2</td>
<td>Decreased when target was embedded, for all speakers.</td>
</tr>
<tr>
<td>Shifter PWS audience &lt; shifter PWS non-audience; not for PWNS.</td>
<td>PWS &gt; PWNS for P1 and P2.</td>
</tr>
<tr>
<td></td>
<td>Shifter PWS &lt; shifter PWNS during non-audience, but not audience condition.</td>
</tr>
</tbody>
</table>
Chapter 6: Discussion

The purpose of this study was to examine sentence-level speech motor variability, stability, and flexibility in PWS and PWNS using linear and nonlinear techniques. Additionally, this study examined the impact of the presence of an audience and linguistic complexity on these measures. Importantly, a distinction was made earlier in this manuscript that variability refers to the inconsistency of recorded speech movements over repeated trials, whereas stability and flexibility represent more abstract concepts related to system-wide speech motor functioning and the ability to transition between states (or adapt to perturbation), respectively. Thus, interpretations about stability and flexibility are (mostly) based on the patterning and/or inconsistencies of speech effector movement.

Prior studies have examined sentence-level stability using STI, a normalized index of the consistency of speech movements over repeated productions of an utterance (e.g., “Buy Bobby a puppy”). Current results corroborate prior studies that reported higher STI for PWS compared to PWNS when the target utterance was embedded in longer and more grammatically complex sentences, but not when the utterance was produced in isolation or embedded in longer-only sentences. Specifically, P2 (“You want Samantha to buy Bobby a puppy now if he wants one”) yielded higher LA-STI for PWS compared to PWNS, indicating that when the target utterance was embedded in this particular sentence, surface speech movements were more variable for PWS. Kleinow and Smith (2000) also indicated that their more basic version of P2 (on which the current P2 was based) yielded higher STI for PWS. Their results were based exclusively on lower lip movement, but the current work showed that STI values
for lower lip and lip aperture are comparable. The current study demonstrated that higher LA-STI values for PWS for sentence P2 were responsible for overall LA-STI differences between groups, as there were no group differences between Base and either L1 or P1. Given these findings, it is speculated that increased variability of movement over repeated trials for P2 may have been due to the upcoming conditional conjunction, “if.” That is, the anticipation of the boundary created by the conjunction could have impacted speech production differentially in PWS. This could be the result of the conjunction creating a context in which the speaker initiates a second utterance/clause, having to effectively “start over.” Indeed, it is well known that stuttering emerges primarily at the beginnings of utterances (for review, see Bloodstein & Bernstein-Ratner, 2008). However, this explanation requires additional testing (e.g., replacing “if” with “because,” controlling grammatical structure). Furthermore, the nonlinear STI technique revealed an important distinction regarding the nature of this P2 difference—namely, that it is due to timing (or phasing) irregularities by PWS, not amplitude irregularities. This was discovered by using the Lucero et al. (1997) technique which permitted calculation of separate amplitude and phase components of STI. Compared to PWNS, PWS exhibited higher LA-NSTIphase, but not LA-NSTIamp values.

Another novel finding related to LA-STI is that this difference is only present for the first 10 trials. Calculating STI based on the last 10 trials yielded no significant differences between PWS and PWNS. This indicates that as the experiment progressed, the speech movements of PWS became more consistent, perhaps due to the adaptation effect—the well-known phenomenon that overt stuttering reduces over
repeated readings of the same material. While overt stuttering was not examined in this study, it is likely that the underlying processes that contribute to stuttering are also present to (at least) some degree during perceptually fluent speech. Therefore, it is plausible that the adaptation effect is present in PWS during perceptually fluent speech. This result contrasts with Smith and Kleinow (2000), who administered a split-half reliability procedure for STI based on the first and last seven trials in their study, which indicated no difference. It is possible that the experimental task in that study precluded the emergence of an adaptation effect, since participants simply repeated the target utterance in blocks of five or ten productions. Repetition is a natural fluency enhancement, and may have facilitated the early productions in their experiment. Stimuli in the current study were presented in randomized order, individually, and visually on a screen. This delivery method is not associated with any known fluency enhancing conditions.

Current results also confirm results from Howell, Anderson, Bartrip, and Bailey (2009), demonstrating a strong correlation between LA-STI and A-STI. It is perhaps unsurprising then that A-STI both differentiated between PWS and PWNS, and also demonstrated the practice/familiarity effect. This replication in a larger study (i.e., the current study) suggests that acoustic techniques can be used instead of kinematic techniques if the investigator’s primary interest concerns linear, normalized measures (e.g., STI). A-STI may have important research implications, in that acoustic recording is more cost-effective than kinematic recording, and more importantly, can be used in conjunction with neuroimaging techniques, which sometimes may prohibit the
introduction of the electromagnetic or optical devices required for kinematic data recording.

**Duration**

Since STI is based on linear normalization, it is necessary to discuss the influence of utterance duration. Smith and colleagues (Kleinow & Smith, 2000; Smith & Goffman, 1998; Smith et al., 1995; Smith & Kleinow, 2000) proposed that the normalization procedures required to calculate STI effectively remove artifact due to variations in utterance duration. Indeed, time normalization requires some type of interpolation by which the samples of the signals of interest are transposed onto a time base of a fixed number of points, so that the signals being compared have common start and end points. As Smith and colleagues (2000) assert, STI is meant to be a *global* measure of variability, one not necessarily concerned with preserving the internal structure of the signals (e.g., speech landmarks). This is problematic because linear “stretching” will by definition change the shape of the signals, with subsequent differences potentially reflected in amplitude-specific measures. Indeed, current results for nonlinear STI indicated that PWS exhibited higher (overall) STI for the amplitude component, but not the phase component.

Rate has been shown to influence STI in adults (Dromey et al., 2014; Smith et al., 1995; Smith & Kleinow, 2000) and children (Smith & Goffman, 1998). That is, there seems to be a positive correlation between rate and STI. However, the literature on this subject provides conflicting results. For example, Smith and Kleinow (2000) reported low correlations between STI and duration within each of their three rate groups (i.e., typical, fast, and slow). Durational impact, however, may only be noticeable across rate
groups. Furthermore, Kleinow and Smith (2000) reasoned that if rate determined STI, then the fastest rate should yield the lowest STI, which in their study it did not. However, in their study, participants were asked to alter their own rate to fit into each of the typical, fast, and slow categories. As Dromey et al. (2014) pointed out, requesting that participants change their speech strategy introduces a confound since their speech production process is being altered. For example, some speakers naturally speak quickly, while others do not. Asking a fast talker to slow down may impact system stability just as asking a slower speaker to speed up would (both may increase system instability). In the current study (as in Dromey et al., 2014), participants were not given instructions regarding speech rate, and thus were free to use their preferred rate. Current results indicated moderately strong, positive correlations with duration for both LA-STI and A-STI. The nonlinear versions of STI (i.e., LA-NSTIlamp, LA-NSTIphase) yielded lower correlations with duration. Thus, it is evident that 1) STI may be significantly influenced by rate, and 2) the nonlinear time-warping function, which attempts to minimize the duration-related error in STI calculation (Lucero et al., 1997), appears to do so.

It is also well known that in controlled studies, PWS typically exhibit a slower speaking rate compared to PWNS. Thus, the finding that PWS exhibited longer target utterance durations than PWNS for all sentences was expected. Additionally, it was expected that the more grammatically complex sentences would yield shorter durations for the target phrase for all speakers (based on Dromey et al., 2014 results). Interestingly, the target utterance exhibited greater duration when embedded in the longer-only sentence compared to Base, which as stated above, could have been the
result of a speaker normalizing effect due to rhythm. That is, speakers may obligatorily produce speech that entrains to the so-called rhythm, which may be slower than natural production without this type of constraint. Given the moderately strong correlation between duration and STI, it follows that STI should be higher for those utterances with longer durations. This is mostly the case—PWS overall exhibited longer durations and greater LA-STI and A-STI values. However, there are also findings that do not support the hypothesis that duration directly influences STI. For example, for all speakers, P1 yielded higher than Base LA-STI and A-STI values. It may be expected then that P1 should also exhibit longer durations than Base, but this was not the case. For the first 10 trials, there were no durational differences between P1 and Base.

While there is strong evidence that rate influences linear STI measures, from the current and prior work, it cannot be concluded that fluctuations in linear STI are solely due to durational fluctuations. It was difficult to parse the effect of duration in this study because it was treated as an additional dependent variable—one that changed based on the same factors as STI and RQA (e.g., group, sentence, condition). Thus, it was not included as a fixed or random factor in any of the statistical models. However, it is clearly the case that time-normalization procedures alter the original trajectory signals, and in the process, may distort any analyses concerning them. Furthermore, linear and nonlinear STI analysis is a technique that measures consistency of movements between trials. Thus, there is an assumption that these movements converge on an “underlying template” of trajectory motion (Smith et al., 1995). In other words, it is assumed that repeated productions of an utterance at a speaker’s preferred rate should be the same, and that any divergence (or error) is reflective of noise in the system.
From a dynamical view, however, variability is a critical component of a healthy system. Therefore, it was desirable to examine other, complementary approaches to measuring stability that 1) do not assume that repeated utterances should converge on the same underlying template, 2) can measure both across trial and within trial variability, 3) are mostly immune to durational influences, and 4) respect the balance of variability, stability, and flexibility in biological systems. RQA provided one way to approach these challenges.

Lip aperture dynamics

The current study employed RQA techniques to gain a deeper understanding of the nonlinear structure in lip aperture dynamics in PWS and PWNS. Recall that Takens’ (1981) theorem indicates that information about the dynamics of a system can be garnered when only one variable—in this case, lip aperture—is known. To provide a more complete explanation of the current data set, three RQA variables were examined: %DET, TREND, ENTROPY.

%DET is a measure of the repeatability, or regularity, of time series data. The finding that PWS were more deterministic in lip aperture than PWNS may on the surface appear counter-intuitive. That is, it may be expected that more variable effector movements (higher STI) would be associated with lower %DET. However, this logic rests on a narrow view of variability as unwanted or random noise in the system. From a DST perspective, speech motor systems require a balance of variability, stability, and flexibility. Thus, a system that is too deterministic may represent a system that lacks stability and/or flexibility. Prior work has revealed that pathological systems are characterized by increased regularity or stereotypical (i.e., less flexible) behavioral
patterns (Goldberger, 1997), and both individuals with Parkinson’s disease (Schmit et al., 2006) and stroke patients (Ghomashchi et al., 2011) have been shown to exhibit higher %DET than control participants. More directly related to the current work, McClean, Levandowski, and Cord (1994) found that PWS were less variable than PWNS on various timing measures (e.g., onset of first vowel glottal cycle, maximum point of jaw displacement). Along these lines, Kalveram (1993) proposed a neural network model of sensorimotor learning in which excessively strong couplings between underlying neuronal populations (or dynamical variables) responsible for speech led to reduced motor variability, increased system instability, and subsequently, stuttering. In the only study to measure %DET in PWS and PWNS, van Lieshout and Namasivayam (2010) reported a main effect of rate showing a reduction in %DET (as related to inter-gestural coordination between bilabial closure and tongue body constriction) as rate increased. Comparatively, the current analysis did not find a correlation between %DET and duration, suggesting that %DET is not influenced by speech rate. However, as noted above, deliberately changing speech rate (which participants in their study did) may significantly alter the speech production process for a speaker, thus making it difficult to compare differences between so-called rate conditions. Van Leishout and Namasivayam (2010) did not report a %DET group difference. The small sample size in their study, the deliberate rate change condition, or the nature of the measurement all may have precluded any meaningful difference related to %DET. Current findings provide novel evidence that PWS exhibit more deterministic lip aperture dynamics than PWNS.
TREND is a measure of stationarity of the time series, or how the repeatability of the time series evolves throughout a given trial. TREND provides an indication of the degree to which the mean state of the system, or set-point or frame of reference, is fluctuating (Dijkstra, 1998; M. A. Riley et al., 1999). Thus, TREND may provide information related to stability, in that greater TREND magnitude (along with increased deterministic structure) may indicate reduced system stability. The finding that PWS exhibit increased TREND magnitude, but only for the Base phrase, suggests that the frame of reference (or set-point) for PWS for the Base utterance production is less stationary (i.e., fluctuates more), a finding in line with Ghomashchi et al. (2011) which showed that stroke patients exhibited greater TREND magnitude in postural control compared to controls. Interestingly, TREND magnitudes decreased for both the longer and more grammatically complex sentences for all speakers, but to a greater degree for PWS. These findings are in line with findings that during more complex postural tasks, TREND magnitude decreased (M. A. Riley et al., 1999), and also findings that increasing dual task cognitive difficulty reduced TREND magnitude related to postural stability (Mazaheri, Salavati, Negahban, Sanjari, & Parnianpour, 2010; M. A. Riley, Baker, Schmit, & Weaver, 2005). One explanation for this TREND reduction may be that diverting attention to the other task effectively constrains the postural control system (Jeka, 1995; M. A. Riley et al., 2005). An explanation for the effects of increased utterance length and linguistic complexity on the target might be that bounding the target utterance between other words (effectively diverting attention) imposes constraints on the speech production system. This assumes that the Base utterance represents a “starting point” for participants—which is reasonable because participants
learned very quickly in the current experiment that, “Buy Bobby a puppy” was common to all sentences. There is also anecdotal evidence that for PWS, speech is facilitated when the target utterance (or whatever utterance is being emphasized) is surrounded by other words/utterances, perhaps because emphasis (i.e., attention) is taken away from the target. For example, it is well known clinically that PWS often rely on “starter” or “filler” utterances (e.g., “umh... “you know...”, “well...”) to facilitate speech production. Coupled with the fact that stuttering typically emerges in utterance-initial position, it may be the case that embedding the target in larger, relatively simple, sentences has a facilitative effect on production. This facilitation may be responsible for the increase in stationarity for the target utterance when it is embedded, and why TREND values for PWS become more like those of PWNS when it is embedded. This interpretation acknowledges that for purposes of the current experiment, L1, P1, and P2 are longer and more linguistically complex sentences, but also that, relatively speaking, all of the sentences in this experiment are “simple” productions for adults without language or cognitive disorders.

*Impact of presence of an audience*

Several studies have demonstrated that the presence of an audience increases overt stuttering for PWS. However, this is the first study to examine the influence of the presence of an audience on speech kinematics for both PWS and PWNS. PWS exhibited increased lip aperture determinism (%DET) during the audience compared to non-audience condition. This finding is similar to that of Riley, Baker, Schmit, and Weaver (2005), who found that postural sway became less variable (i.e., more deterministic) when cognitive load increased. PWS appear to adopt a more restrictive or
rigid approach to speech production when speaking in the presence of an audience. It may be the case that in order to preserve forward-moving speech, PWS elect to adopt this more restrictive strategy.

An additional finding that supports the view of a more restrictive motor control strategy is based on ENTROPY, which provides a representation of complexity of the signal. That PWS exhibit lower complexity during the audience compared to non-audience condition, suggests a compensatory mechanism for PWS. That is, during increased communicative pressure (i.e., speaking in the presence of an audience), PWS elect to adopt a more conservative approach to speech motor control. PWNS do not exhibit this reduced complexity during the audience condition. It may be speculated that PWNS demonstrate a “luxury” of not having to adapt under stress. Conversely, PWS may feel as though they have to (or may actually have to) adapt under these conditions, specifically by decreasing the complexity of the output.

It was somewhat surprising that PWS did not exhibit a difference in LA-STI between conditions (even though they did for %DET). This triggered the investigator to examine only those participants who exhibited a significant “shift” in anxiety between the non-audience and audience conditions, indicated by a difference of at least two points on a Likert scale of subjective anxiety rating (see Appendix). Fortuitously, there were six PWS shifters and five PWNS shifters. Thus, since this study consisted of a relatively large sample for a speech production and stuttering study, it was able to capitalize on the small subset (or subgroup) of speakers who were impacted by the presence of an audience during a speaking task. For this subset of speakers, PWS exhibited a decrease in LA-STI and the same increase in %DET as described above during the
audience compared to non-audience condition. These findings indicate that those PWS who are more prone to experience anxiety during a speaking task will also be more susceptible to altering their speech production, by demonstrating less effector variability and more deterministic structure in movement. Furthermore, shifter PWS exhibited a higher TREND magnitude during the audience compared to non-audience condition, suggesting less stationarity, or more speech system volatility, and more underlying instability. This finding complements the increase in %DET in that a system that is less stable will attempt to compensate by becoming overly deterministic. For example, PWS may rely on more stereotyped behaviors because of an underlying speech motor deficit.

These findings complement the proposal by van Lieshout and colleagues (Namasivayam & van Lieshout, 2011b; van Lieshout et al., 2004) that PWS are less flexible in adapting to cognitive-emotional influences (e.g., while speaking in the presence of an audience). Importantly, findings of reduced variability, increased determinism, and lower system stability reflected speech that is perceptually fluent (i.e., free from overt disfluency). This highlights that despite overt or observable fluency, a speaker may experience concurrent underlying speech motor difficulty, which ultimately should be reflected when determining the magnitude of an individual speaker’s problem (i.e., severity).

To be clear, these results do not imply that anxiety causes stuttering or speech production difficulty in PWS. Rather, it suggests that those PWS who experience anxiety during speaking tasks are more likely to alter their approach to speaking than those PWS who do not experience anxiety. These findings do support the claim, however, that anxiety can play a significant role in how a PWS learns to manage his or
her speech clinically. Specifically, the approach that some PWS take to coping with stuttering (e.g., tensing, "pushing") is likely maladaptive. Indeed, most approaches to stuttering (e.g., stuttering modification, “normal talking”, fluency shaping) in one way or another propose that PWS produce speech with less tension (e.g., pull-outs, light articulatory contacts/approximations). Given that PWS who also exhibit higher anxiety levels are those who tend to change their speaking approach in the presence of an audience, to a greater degree, it follows that speakers who do exhibit communicative anxiety would benefit from desensitization procedures in therapy.

**Theoretical implications**

Van Lieshout and colleagues (Namasivayam & Van Lieshout, 2008; van Lieshout, 2004; van Lieshout et al., 2004; van Lieshout & Namasivayam, 2010), as well as others (e.g., Lancia et al., 2014; Saltzman, 1991; Smith & Kelly, 1997), have discussed the benefits of applying a dynamical perspective to speech production and stuttering. The current findings provide further support for this perspective. Specifically, this dissertation provides evidence that speech motor stability is not simply the inverse of speech motor variability—a well-known principle in nonlinear dynamics (for discussions, see Bernstein, 1967; M. A. Riley & Turvey, 2002; Stergiou & Decker, 2011; Turvey & Kugler, 1984; van Lieshout & Namasivayam, 2010).

For example, stuttering has long been viewed as multifactorial, in that its onset and progression is thought to be driven by a combination of motor, linguistic, cognitive, and emotional factors (e.g., Conture et al., 2006; De Nil, 1999; Namasivayam & van Lieshout, 2011b; Smith & Kelly, 1997; van Lieshout et al., 2004; Walden et al., 2012). Thus, one line of inquiry into stuttering has been measuring the impact of (any of) these
factors on speech motor control. Previous investigations that examined the impact of linguistic and cognitive-emotional factors on speech variability (e.g., Kleinow & Smith, 2000, 2006) have asserted that increased effector variability signifies system instability. Based on these findings, it might be expected that for some PWS, the presence of an audience (a communicative stressor for many PWS) would lead to increased speech variability. However, the current findings reveal that speaking in the presence of an audience leads to a reduction in effector variability (i.e., STI), and increases in both deterministic structure (%DET) and nonstationarity (TREND), for PWS. A MP (or linear) perspective would interpret increased variability as decreased system stability. From a dynamical perspective, it is the reduction in effector variability, coupled with the increased determinism and greater TREND magnitude, that signifies a system that is rigid, inflexible, and ultimately, unstable. Thus, the dynamical view provides a more comprehensive explanation regarding speech stability. This dissertation also demonstrates how a combination of linear and nonlinear techniques can be used to develop a deeper understanding of the variability, stability, and flexibility associated with stuttering and speech motor systems.

Clinical Implications

The indication that PWS are more restrictive in their movements during speech production suggests that PWS adopt a more deterministic (or rigid) speech pattern to maintain fluent speech in spite of potential underlying speech difficulty. This strategy is likely associated with increased muscle tension during speech production. It is well known that PWS exhibit significant tension as a result of stuttering—articulator tension, facial/neck tension, or tension in the chest or other parts of the body. And most
treatment approaches involve some aspect of reducing tension during speech production. For example, a “pull-out” is a strategy in which speakers identify articulatory tension during a stuttering event and subsequently (attempt to) reduce that tension to continue with speech production. Similarly, “light articulatory contacts” and “easy onsets” are speaking strategies in which speakers initiate phonation with reduced tension, either between the articulators (e.g., for consonant-initial) or in the larynx (e.g., for vowel-initial). However, there is little quantitative evidence to support the implementation of these strategies. Speculatively, it is possible that using a metric such as %DET would facilitate the therapeutic approach of reducing tension, since it provides a straightforward index (percentage), which theoretically represents the degree of inflexibility or restriction. There may be an ideal, individual specific, range of %DET values. The simple utterance used in this study (i.e., “Buy Bobby a puppy”) could be used to introduce speakers to the idea of a “sweet range” of production. Of course, the usefulness of such an approach needs to be tested.

Considerations

There are primarily three issues that should be considered in the context of this work. First, STI was criticized for the assumption that effector movements associated with a repeated utterance should converge on the same trajectory “template.” Indeed, any linear approach to measuring variability will by definition carry this same assumption. A significant strength of nonlinear methods is that they do not carry this same assumption, and are more concerned with how the trajectory evolves over time. However, nonlinear approaches assume cyclicity in time series data. The assumption that articulator trajectories during speech exhibit cyclicity is at the very least, arguable—
and at the most, rejected. This assumption was made in the current study because lip aperture trajectory for the target utterance (i.e., “Buy Bobby a puppy”) exhibits a clear “sine-like” pattern. Of course, selection of the target to be studied when implementing RQA requires careful consideration. It cannot be over-stated, though, that RQA is well-suited for the study of speech variability because it allows for the measurement of variability within trials, as opposed to STI, which provides one indexed value that is supposed to reflect (average) variability of a set of trajectories.

Second, data were collected in an “un-naturalistic” environment. That is, participants were required to read relatively simple sentences from a monitor in the confines of a laboratory. Since stuttering is a disorder that primarily manifests in meaningful communicative exchanges, there is concern related to how generalizable the current results are. However, the approach to examining speech motor output in PWS and PWNS is well accepted. Furthermore, the purpose of this study was to quantify and investigate subtle differences between the fluent speech of PWS and PWNS. While this study may lack ecological validity, the controlled nature of the approach taken revealed subtle speech differences that may not have been evident if data was collected in more naturalistic communicative contexts. That said, a goal of future research is to use the nonlinear approaches investigated in this study in more ecologically valid environments.

Third, RQA requires the a priori selection of parameters. Results from RQA are dependent upon these parameters, as output of the system is sensitive to the system’s initial conditions. While there are established guidelines for parameter selection, there is also significant uncertainty in this domain. The current work followed these established
guidelines, and also completed the entire analysis on different sets of closely related parameters (to ensure the results were not due to artifact in the data). Results from these additional analyses corroborated the results reported on in this dissertation.
Chapter 7: Conclusion

The application of nonlinear approaches to measuring sentence-level speech provided insight into the nature of variability, stability, and flexibility in stuttering and non-stuttering systems. Specifically, RQA complemented existing linear approaches (i.e., STI) by providing information about the underlying dynamics associated with observable speech movements (in this case, lip aperture). This approach is particularly useful for stuttering since quantifying both observable and unobservable behaviors associated with stuttering has challenged researchers and clinicians since the measurement of stuttering has become a scientific endeavor. This study demonstrated the feasibility of employing nonlinear approaches to better characterize the subtle differences that exist in the fluent speech of PWS and PWNS. RQA has provided novel evidence that PWS exhibit more deterministic and less stationary speech patterns during production of relatively simple utterances. These characteristics indicate that the speech motor systems of PWS may be less flexible than those of PWNS.

The notion that stuttering is “multifactorial”—that many factors contribute to its onset and progression—is now well accepted. However, understanding the nature of how these factors or their interactions influence stuttering remains a significant challenge in stuttering research. The current results add to the existing literature by demonstrating that one source of cognitive-emotional stress (i.e., the presence of an audience) reduces effector variability while increasing deterministic structure and nonstationarity in speech motor output for PWS (and not PWNS). Thus, evidence is provided that the speech motor systems of PWS are de-stabilized and susceptible to this kind of cognitive-emotional stressor. This finding highlights the importance of
acknowledging the dynamic nature of a complex system (i.e., the speech motor system). That is, a de-stabilized system often exhibits effector patterns that are more deterministic. Future work should employ a combination of linear and nonlinear techniques to examine the impact of other stressors on variability and system stability and flexibility. Additionally, it will be revealing to examine correlations between kinematics and neurophysiological methods using nonlinear approaches.
Appendix: Condition questionnaire

Audience/Non-Audience Debriefing Questionnaire

*These questions will be read to each participant.*

“Did anybody come into the room during testing?” Y / N

“If so, how many people were there (besides the examiner)? _______ What do you think the gender(s) of the observer(s) was(were)?” _______

Did you experience anxiety when people weren’t in the room (besides examiner)?

1 2 3 4 5 6 7

no anxiety moderate anxiety extreme anxiety

Did you experience anxiety when people entered the room?

1 2 3 4 5 6 7

no anxiety moderate anxiety extreme anxiety

“Did you feel differently when the observers were in the room vs. out of the room? If so, how?”
References


movements of multiple points inside and outside the vocal tract. *Brain and Language, 31*(1), 26–35.


