Title
Word-medial syllabification and gestural coordination

Author
Garvin, Karee

Publication Date
2021-10-01

Word-medial syllabification and gestural coordination

by

Karee Garvin

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Linguistics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Keith Johnson, Co-chair
Professor Sharon Inkelas, Co-chair
Professor Darya Kavitskaya

Fall 2021

Word-medial syllabification and gestural coordination

Abstract

Word-medial syllabification and gestural coordination

by

Karee Garvin

Doctor of Philosophy in Linguistics

University of California, Berkeley

Professor Keith Johnson, Co-chair

Professor Sharon Inkelas, Co-chair

The focus of this study is to use Electromagnetic Articulography (EMA) data to investigate the difference between word-initial and word-medial coordination, how coordination is affected by segment quality and stress, and to model coordination as it relates to syllabification. There are four key results of this study. First, segments in word-medial position have looser constrictions than segments in word-initial position, demonstrating that articulatory pressure is a driving factor in word-medial lenition. Second, word-medial syllables differ in coordination from word-initial sequences; however, current methods of analysis do not provide adequate tools to analyze such variability. Third, jaw oscillation does provide a metric for analyzing variability, crucially, demonstrating that word-medial syllables are more variable than initial syllables and unstressed medial syllables are more variable than stressed medial syllables. In addition, clusters coordinated within the same jaw phase show an increased likelihood to show c-centering and rightward shift. Thus, there is a gradient likelihood for medial clusters to syllabify according the ONSET MAXIMIZATION PRINCIPLE, where initial stressed syllables are most likely and medial unstressed syllables are least likely, with the cluster syllabifying instead as a coda and onset. Fourth, coordination stability is influenced by by both coupling to jaw phase and intrinsic coordination between segments; both homorganic clusters and heterorganic clusters show coordinative stability, showing that jaw rather than gestural overlap between articulators determines stability; however, cluster composition and vowel quality interact and influence patterns of coordination such that a cluster is more likely to be syllabified as an onset when the cluster and vowel match in backness. Together, these findings demonstrate that word-medial coordination is dynamic and influenced by both top-down pressures, like stress, and bottom-up pressures, like segment quality. Furthermore, these findings demonstrate that jaw oscillation can be used as a diagnostic for motor planning and syllabification, providing a tool to improve existing models of planning and syllabification.

For my mom and dad

Contents

Acknowledgments

My research and my path in academia have been made possible through the support and guidance of so many wonderful humans, but first and foremost, I want to express my gratitude for my parents. Growing up, my parents filled my world with learning, creativity, and curiosity. On Saturdays, our family worked in and around our house. I always had to help, but I got to pick which chores I wanted to help with and I often chose to work out in the yard with my dad. He taught me how to drive stick so I could drive a little truck around the yard and pick up leaves. He taught me how to do carpentry and electrical work and about the importance of using the right tool for the job and that you can never have too many clamps. And whenever I felt like I couldn't do something or that I might do it wrong, he would say 'not my daughter, my daughter can do anything.' I think my dissertation is really an ode to my dad because it really is about problem solving and finding the right tool for the job.

After the work was done, we'd play games with our mom. She used games as a way to teach how to read, do math, strategize, and solve problems. My mom always knew how to make something joyful. She would hide treats behind the clutter in our rooms so that we'd find little surprises while we worked or put a TV on a rolling cart so we could watch movies while canning. My mom taught me to be determined and strong and to be kind and understanding. I don't think I am nearly as tough or as kind as my mom was, but I think my determination to keep going even when things are hard is because of my mom, so I really owe getting through three masters programs and a PhD to mom. My parents both passed away the summer before I started at Berkeley, and while I wish they had been here to see me cross the finish line, thinking of them and the lessons I learned from them has helped shape me into the person I have become both in and outside of academia.

I am also incredibly grateful for my dissertation committee who have been amazingly supportive throughout this whole process and who have helped shape my ideas, writing, and my research so much. Thank you to Dasha Kavitskaya whose advising and friendship is a much needed source of efficiency, honesty, confidence, and pure delight. And thank you to my wonderful co-chairs and advisers, Keith Johnson and Sharon Inkelas. During my first meeting with Sharon, we talked about sonority and throughout my time at Berkeley, Sharon has since asked me many times something along the lines of: so do you think this relates to what were talking about with sonority? I somehow get the feeling that Sharon knew where I was headed all along and has been patiently guiding me as I figured it out for myself. Sharon is also an amazing role model. I have often found myself wondering *what would Sharon do?* And when I am not sure, she has been there to provide expert advice.

Keith has been a huge support to me in my research and in my life. As an adviser, I can't tell you how many times Keith has walked me through the same questions again and again because I am still not sure I understand. And he does it with patience, kindness, and enthusiasm. He is also the kind of adviser I can ask advice from on a DIY project or even someone I can ask to borrow tools, which has really meant a lot to me. I particularly appreciate the support Keith provided me early in the program in grieving my parents. It was really difficult moving to a new place and starting in a new program so soon after my parents' deaths, but Keith welcomed me and supported me and it made a huge difference. Keith, Sharon, and Dasha are all people I can go to for advice about basically anything and people I can share and celebrate my life with. These three are the dream team.

My dissertation would not be possible without the endless support and advice of the brilliant staff at Berkeley. Ron Sprouse has been so patient in teaching me how to code and helping me through more questions than I can count. I really could not be writing this dissertation without Ron. We are also so so lucky to have had Belén Flores and Paula Floro in our department. These two have the expert know-how to solve any problem. Trips to the field, experiments in the lab, and conference organizing would have been a nightmare without Paula, but she always knows what needs to be done and how and she helps with cheer and warmth. And for every question from how to find housing to how to file my dissertation, Belén has had an answer. Belén has been someone I can talk to about anything and I am thankful for her help and kindness.

I have had so many fantastic mentors during my time at Berkeley. I am grateful for the guidance of Andrew Garrett, whose brilliant sense of style and writing have brightened my time at Berkeley and whose wit and wisdom I greatly admire. I am grateful to Susanne Gahl for her care and her thoughtful guidance and support. Thank you to Larry Hyman whose kindness and support got me into Berkeley and through it. I always know that I can ask him for help with anything and he will respond yes, promptly. Thank you to Gašper Beguš whose advising always makes me feel like a million bucks and to Terry Regier for being the kind of teacher who calls to check in on the weekend when he knows you are struggling through a difficult homework assignment. Thank you to Line Mikkelsen for teaching me so much about compassion in the classroom and to Lev Michael and Christine Beier who have helped me navigate so much of my fieldwork experience. Thank you to Peter Jenks, Susan Lin, and Amy Rose Deal, who have helped shape my writing and research and have all guided my time at Berkeley.

I have also worked with some really incredible undergraduate students who have been a huge help in my research. So much of the data collection and data processing for this project would not have been possible without their diligent efforts. Thank you to Phoebe Killick, Ray Mason, Anahita Farshi Haghro, Qing Yang Wang, Cassandra Serrano, Emma Brown, Emily Li, Elissa Chau, Thomas Lu, Cooper Bedin, Joanne Chuang, Jacky Chen, Ian Neer, Esther Ju Hyun Yeo, Emily Tsai, Amelia Fineburg, Nick Carrick, Ruby Dong, Cheyenne Tsai, and Tracy Lee.

Coming to Berkeley would not have been possible without some really incredible teachers and mentors at the University of Iowa and Idaho State University. Thank you to Bill Davies, Jill Beckman, Jerzy Rubach, Cathy Ringen, Liz Cartwright, and Lenore Grenoble whose guidance helped me to grow as a scholar and helped prepare me for this PhD. Thank you to Chris Loether who sparked my love for linguistics and whose confidence in me as a reseacher helped me get started down my academic path. And thank you to Bethany Schultz Hurst, who has been my teacher, adviser, and friend since she taught my very first college class during my senior year of high school. My decision to start college was largely inspired by her excellent teaching, and through each new milestone, I think of her with admiration and gratitude. I am so lucky to have had her as a mentor through so much of this journey.

Thank you to the Shoshone people and in particular, Drusilla Gould, who taught me so much about community and fieldwork. And thank you to the Nafaanra community who welcomed me with open arms. Their enthusiasm for and knowledge of their language has made my job as a fieldworker easy. A special thank you to my consultant Sampson Atta for welcoming me into his home and for patiently answering my endless questions. Thank you to Kofi Joachim Perry who has worked so hard in helping me to collect data and is a true friend. And thank you to Job Ababio, Anane James, and Henry Sales who have all shared so much about their languages with me.

I am grateful to have such a supportive family, like my siblings, Dan Garvin, Angie Jackson, Janice Whitman, Jeff Garvin, and JD Garvin who are calling even as I write this to wish me well, encourage me, and tell me that they are proud of me. I am also grateful to their spouses Melanie Garvin, Steven Jackson, Aaron Whitman, and April Garvin, and my in-laws, Brian Palmer, Joann Palmer, Hannah Pasquesi, Andy Pasquesi, Evan Palmer, and Haley Aubrey, who have supported me in everything I do. And I am thankful to all of my nieces and nephews who bring me endless pride and delight: William Garvin, Isabelle Garvin, Jackson Garvin, Ronan Garvin, Matthew Jackson, Ella Jackson, Grace Jackson, Tim Whitman, Sam Whitman, Mark Whitman, Jon Whitman, Luciana Whitman, Nate Garvin, Sarah Garvin, Lex Garvin, Bryce Garvin, and Eloise Pasquesi. (Yes, I have 17 nieces and nephews!)

I am particularly grateful to my friends and peers at Berkeley. Thank you to Hannah Sande and Marcus Ewert for all of the games, puzzles, movies, climbing adventures, and bird watching. What luck to find friends who share all of my hobbies! Thank you to the best possible mentor, Emily Clem who is just brilliant and gives the best advice. Thank you to Julia Nee for our weekly lunches, which were a much needed reprieve from grad-school stress. Thank you to Noah Hermalin and Eric Wilbanks for being such excellent cohort mates. Thank you to Myriam Lapierre for all the shared sour beers and last calls at Fieldwork and who has helped me to grow so much as a researcher and a person. Thank you to Ginny Dawson, Sarah Bakst, Meg Cychosz, Emily Remirez, Allegra Robertson, Katie Russell, Martha Schwarz, Zach Wellstood, Nico Baier, Zach O'Hagan, Andrew Cheng, Dmetri Hayes, Noga Zaslavski, Kenny Baclawski, Matthew Faytak, Madeline Bossi, Emily Drummond, Alice Shen, Maks Dabkowski, Anna Björklund, Emily Grabowski, Nik Rolle, Tessa Scott, Phuong Khuu, Edwin Ko, Raksit Lau, and Tyler Lemon. These colleagues have provided me with valuable feedback on my work and have filled my time at Berkeley with board games, coffee, happy hours, and good conversation.

And thank you to all my non Berkeley people both in and outside of academia. To my very first cohort mate and dear friend, Meg Tracy thank you for all the dance parties, rock climbing, adventures, happy hours, and coffee breaks and for your brilliant advice and much needed pep talks. You are magic. Thank you to the very best of office mates, Doug and Eli, whose kindness, support, and brilliance inspired me during my time Iowa. Thank you to my fantastic peers at Iowa, Carly Sommerlot and Maria Sumner, whose enduring friendship has meant so much to me. I especially want to thank my friends, Rachel Dierken and Julie Baker. Thank you to Rachel for long phone calls, handmade gifts, and for teaching me it's okay if I sometimes have emotions, I really needed that. And thank you to my oldest friend and my partner in crime, Julie Baker. Here is to all the sonic slushies, pumpkin pie milk shakes, late viewings of Cool Runnings, and to all the childhood shenanigans I'd better not put into print. Thank you for convincing me to join drumline and being a source of love, joy, and fearlessness in my life. You said it best: 'We were friends before braces and acne. This will be a friendship for the ages.'

And finally, thank you to my sweet little family. Thank you to the two cutest pups in the whole world, Jackson and Bacchus, for making sure I took some much needed study breaks. And last, but certainly not least, thank you to my husband and partner, Ryan Palmer. Your support means the world to me and I am so incredibly grateful for a brilliant partner who has spent hours talking about dynamics with me and helping me through all the ups and downs of grad school. Thank you for all the dinners, coffee, board games, and adventures. I am so happy to have you by my side.

Chapter 1

Introduction and background

1.1 Introduction

The syllable has been a powerful tool in phonological theory because it provides a simple mechanism for constraining the environment for a wide array of phonological patterns (e.g., Blevins, 1995; Hooper, 1972; Kahn, 1976; Prince and Smolensky, 2004; Smolensky and Prince, 1993). In addition, syllables have been found to have acoustic and articulatory correlates, where segments are produced distinctly depending on whether they occur in the onset or the coda (Krakow, 1999) and there is both articulatory and acoustic evidence that syllables are coordinated as a unit rather than discrete sequences of segments (e.g., Browman and Goldstein, 1988; Marin and Pouplier, 2010; Munhall et al., 1992; Katz, 2012).

Despite support for syllables in phonological theory and in speech production, theories of syllable structure and coordination typically assume that syllable margins behave uniformly regardless of whether they occur at a word boundary or word internally. For example, the Coupled Oscillator Model predicts that word-initial and word-medial syllables should have the same patterns in coordination as onset maximization is presumed to part of motor planning. However, there is widespread evidence that word-medial margins differ from word boundary margins, e.g., lenition patterns differently across margins (Gurevich, 2011; Katz, 2016; Kingston et al., 2008). As a result, differences between syllables at word boundary margins and word-medial margins are less well understood; likewise, theories to model these differences between margins are less widely agreed upon. For instance, asymmetries in segmental distribution and syllable complexity between word-boundary and word-medial syllable margins are common across languages, but theories differ as to whether these effects should be captured through syllable structure or through prosodic constraints (Blevins, 1995; Hayes, 1982, 2009; Kahn, 1976; Jensen, 2000).

While sonority was influential early in the development of syllable theory (Clements, 1990), theories that rely on sonority to predict where syllable breaks occur fail to account for a number of common cross-linguistic patterns and makes a number of spurious predictions. For example, sonority predicts that stop+nasal clusters should be preferred to nasal+stop clusters, but nasal+stop clusters are far more common. Theories relying on more general principles of perception (Fujimura et al., 1978; Steriade, 1997) and production (MacNeilage, 1998; Vallée et al., 2009) to model syllabification thus provide appealing frameworks for syllabification that capture many of the same patterns that sonority accounts for, while improving predictions of cross-linguistic patterns.

Although there are a number of open questions surrounding word-medial syllabification, studies of the articulatory correlates of syllable structure have focused on environments where syllabification is straightforward; however, because word-medial syllabification is often not straightforward, little is known about whether the same articulatory correlates hold for word-medial syllables or whether the coordination of word-medial syllables differ. In particular, stress has been shown to affect both the perception of syllables (Fujimura et al., 1978) and the gestural alignment of word-medial segments (Beckman and Edwards, 1994; Byrd et al., 2009); however, the extent to which stress affects word-medial syllabification and whether it is the only factor that affects word-medial syllabification is an open question.

This study will present data for syllables in word-initial and word-medial environments, analyzing both stressed initial syllables and stressed and unstressed medial syllables to investigate how word position and stress affect articulation and coordination. This study will demonstrate that there are significant differences between word-initial syllables and word-medial syllables both in segment articulation and syllable coordination. Crucially, this study will demonstrate that current methods of analyzing syllable coordination do not provide clear tools for analyzing variability in syllabification, and consequently, the results of these analyses are only partially interpretable. Thus, this study will provide an introductory framework and novel methodology for analyzing jaw oscillation to investigate differences in coordination and syllabification. The methods for analyzing jaw movement presented here provide significant results in analyzing differences in coordination across tasks and improve outcomes in analyzing syllable coordination. These results have important implications for theories of lenition, syllable structure, and syllabification.

1.1.1 Roadmap

The goal of this study is to analyze the effect of word position on syllabification and gestural coordination, using Electromagnetic Articulography (EMA) data to analyze differences in coordination in word-initial and word-medial contexts. The prediction of this study is that syllabification is sensitive to a number of factors such as stress and segment quality, and as such, syllabification is dynamic and variable. Existing methods for analyzing syllable coordination do not provide a clear framework for analyzing or accounting for variability, and thus, this study proposes using jaw gestures to analyze and interpret variability in syllabification.

This chapter will provide a background on syllables, presenting an overview of syllables in phonological theory (Section 1.2) and drawing on phonetic and typological evidence that provide support for syllables (Section 1.3). This chapter will focus in particular on word-medial syllabification and the pressures on syllabification. Section 1.4 will motivate the need for frameworks that can distinguish between syllable margins by discussing the differences between word-boundary syllable margins and word-medial syllable margins and how phonological processes, namely, lenition for the purposes of this section, pattern distinctly at each margin. Section 1.5 will discuss frameworks for dividing segments into syllables, including sonority, licensing by cue, and the Frame-Content Model, which provides an articulatory basis for syllables. Subsequently, this section will provide an overview of how stress and segment quality affect syllabification. Finally, this chapter will highlight the main goals and findings of this dissertation.

Chapter 2 will detail the methods of a two studies; study one analyzes /st/ tokens from MOCHA-TIMIT sentences; study two analyzes tokens controlling for word position and stress. Chapters 3 and 4 will present the results of the analyses on segment articulation and syllable coordination. Chapter 3 will discuss segmental articulation, focusing on the differences between gestures depending on their position in the word and whether they are stressed. Chapter 3 will highlight in particular the differences in articulation between word-initial and word-medial segments, which was found to have a stronger effect on segment articulation than stress, where word-medial segments are more reduced than word-initial segments. These results will be discussed in the context of lenition, providing support for articulatory pressures on lenition. Chapter 4 will present the results of analyses on syllable coordination, including compensatory shortening, c-centering, and shift to demonstrate that while these methods clearly illustrate the differences between word-initial and word-medial syllables, they do not provide a clear framework for analyzing variability in syllabification, making any interpretation of the results tentative.

Chapter 5, therefore, provides a novel methodology and foundation for using jaw oscillation as a framework for analyzing syllabification that can then be used in conjunction with existing syllable coordination frameworks. The results of Chapter 5 demonstrate that analyzing jaw coordination can identify patterns and improve the results of methods for analyzing syllable coordination. Finally, Chapter 6 discusses the implications and conclusions of this study for theories of syllabification and methods for analyzing syllable coordination to demonstrate how jaw coordination can be elegantly incorporated into existing frameworks of syllabification using constraints that capture jaw movement and coupling.

1.2 Introduction to syllables

Since Kahn (1976) and Selkirk (1981), the syllable has been canonically assumed to be a decomposable structure of onset, nucleus, and coda consisting of discretized vowel and consonant segments, where the onset and coda are filled by consonant or [-syll] segments skeletally represented by C, and the nucleus is filled by vowels and syllabic consonants, or [+syll] segments skeletally represented by V. Theories of syllable structure have been heavily based on a few key observations in syllable typology. First is the division of the nucleus and coda into the rhyme separate from the onset, as depicted in Figure 1.1. This structure is based on the observations that a) in poetic rhyme, the onset does not contribute to the rhyme scheme, but the vowel and coda do and b) both the vowel and coda are widely attested as contributing to syllable weight and subsequently stress attraction, whereas onsets do not typically contribute to syllable weight (Halle and Vergnaud, 1980; Hayes, 1989; Hyman, 1985).

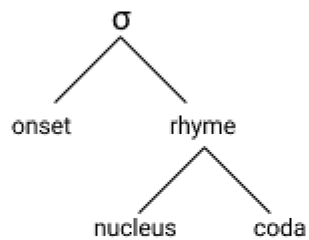


Figure 1.1: Syllable model

A second foundational observation regarding syllable structure is the privileged status of onsets. CV syllables are presumed to be the least marked syllable shape because CV syllables are the most common syllable across all languages and stand in an implicational relationship to other syllable shapes, i.e., if a language permits a more complex syllable shape, like a CVC, V, or CCV syllable, it implies the existence of CV syllables in the language; however, the reverse is not true (Gordon, 2016).

1.3 Support for syllables

1.3.1 The syllable in phonological theory

Building on the basic typological patterns of syllable structure, Hooper (1972) and Kahn (1976) propose an algorithm for syllabification and demonstrate how syllabification improves descriptive adequacy for a number of allophonic alternations. Kahn (1976, p. 55-56) formalizes the syllabification algorithm as follows:

(1) Syllabification Algorithm

- I Assign segments with a [+syll] feature to the syllable nuclei.
- II a Assign all prevocalic segments to the onset of the syllable, where all segments are permissible word-initial clusters (*bo.ston* but **a.fter*)
  - b Assign all remaining consonants to the coda. (*af.ter*)

Related to the privileged status of onsets is the Onset Maximization Principle (Kahn, 1976), which stipulates that if a segment can be syllabified in the onset, it will, as described in rule IIa. This stipulation is derived from the typological observation that onsetless syllables and codas are marked, and therefore, it is better for segments to be syllabified as onsets. However, another important component of rule IIa is that segments will only be syllabified as the onset if it is a legal onset in the language, preventing unnatural clusters from being syllabified. Segments that are not syllabified in the onset will then be syllabified in the coda.

More recent phonological models, e.g., Optimally Theory (OT)(Prince and Smolensky, 2004; Smolensky and Prince, 1993), include similar mechanisms for syllabification, but provide more flexibility to account for differences in syllabification across languages. In OT, syllable structure is operationalized through markedness constraints like Onset, *Coda, *ComplexOnset, *ComplexCoda. These constraints are formulated on markedness scales suggested by typological patterns in syllable shape, giving preference to onsets and dispreferring codas, complex onsets, and complex codas. Markedness constraints interact with faithfulness constraints like Dep, which penalizes epenthesis, and Max, which penalizes deletion. Arising from the interaction between markedness and faithfulness constraints are language specific repair strategies for illicit syllable structure. For instance, in Lardil, onsetless syllables are not permitted and are repaired through deletion, suggesting that Onset outranks Max, as shown in example (2) (Prince and Smolensky, 2004; Smolensky and Prince, 1993, pg. 108).

(2) /yukarpa+in/ ⟶ [yu.kar.pan] *[yu.kar.pa.in] 'husband'

However, in a language where onsetless syllables are tolerated, the constraints can be reranked such that faithfulness constraints outrank Onset, resulting in the surfacing of onsetless syllables. Thus, constraints like Onset and *Coda are structured around the same observations as the syllabification algorithm, but constraints like Dep and Max interact to give rise to wide range of grammars in line with the wide range of attested patterns cross-linguistically (Prince and Smolensky, 2004; Smolensky and Prince, 1993). Crucially, like the syllabification algorithm (Kahn, 1976), OT (Prince and Smolensky, 2004; Smolensky and Prince, 1993) shares the basic assumption of Onset Maximization in English.

The basic theoretical principles proposed by Kahn (1976) and OT frameworks (Prince and Smolensky, 2004; Smolensky and Prince, 1993) capture much of the distribution of syllable shape typology and likewise the frequency of individual syllable shapes within languages. For instance, markedness assumptions correctly capture the frequency of CV syllables and likewise that simpler syllable shapes like CVC and CCV are more common than highly complex syllable shapes. Likewise, the theories accurately predicts the relative frequency of syllables within a given language. Namely, the more complex the syllable, the less frequent it is both within a given language and across languages, i.e., languages that permit highly complex syllables are less frequent than those that don't and even in languages where complex syllables are permitted, they are less frequent than less complex syllables in the language (Gordon, 2016). Canonical theories of syllable structure also capture the attested patterns of complexity at a single margin. For instance, tolerance of greater complexity at one edge implies the existence of syllables of lesser complexity at the same margin (CCV implies CV) CVC > CCV (87%, 60%). CCV > CVCC (60%, 40%) (Gordon, 2016).

The power of syllables in phonological theory is the leverage syllables offer in accounting for cross-linguistic phonological patterns. For example, Hooper (1972) argues that nasal assimilation in Spanish is best accounted for through reference to syllable structure. The examples in (3) illustrate patterns of nasal assimilation in Spanish.

(3)
a. un beso ⟶ [umbeso], 'a kiss'
b. un charco ⟶ [untʃarko], 'a pool'
c. un gato ⟶ [uŋgato], 'a cat'

Nasal assimilation in Spanish occurs across word and syllable boundaries, except in the case of glides, where assimilation only occurs across word boundaries, as is shown in the contrasting examples in (4).

(4)
a. miel → [myel] *ɳyel, 'honey'
b. muevo → [mweβo] *ŋweβo, 'I move'
c. nieto → [nyeto] *ɳyeto, 'grandson'
d. nuevo → [nweβo] *ŋweβo, 'new'
e. un hielo → [uɳyelo], 'an ice'
f. un huevo → [uŋweβo], 'an egg'

When syllable structure is assumed, this pattern can be straightforwardly analyzed because nasals only assimilate when occurring before a boundary (word or syllable). However, in the case of glides, when the glide follows a word boundary, it functions as the syllable onset and thus triggers assimilation. When the glide occurs after a nasal onset, the onset does not trigger assimilation.

Likewise, Kahn (1976) demonstrates that in English stop aspiration can be restricted to the onset of stressed syllables; a generalization that is not as neatly captured without reference to syllables. The simplicity in accounting for patterns across languages through reference to syllable structure has made syllables a useful and lasting tool in phonological theory.

1.3.2 Phonetic support for syllables

Syllables are an intuitive unit to speakers, with studies in perception and neural encoding demonstrating that English speakers are attuned to the beats of a syllable (Morton et al., 1976; Oganian and Chang, 2018). In addition, studies of articulation and gestural coordination provide evidence for syllables. For instance, Krakow (1999) presents key data showing that segments are articulated differently depending on whether the segment occurs in the onset of the coda of a syllable. Differences include tighter constriction of gestures in the onset position, along with tighter coordination among articulators in onsets, e.g., between the velum and tongue in producing /n/, resulting in greater coarticulation between vowels and nasals when the nasal occurs as a coda.

A formal framework of the phonetic evidence of syllables is the Coupled Oscillator Model of gestural timing, which argues that the typological preference for CV syllables is based in motor planning. The syllable is generally agreed to be a unit in speech planning (Cholin, 2011). Models of motor planning and production (e.g., Hickok, 2014; Houde and Nagarajan, 2011) demonstrate the role of feedback in motor planning and speech production, illustrating that speakers rely on auditory and somatosensory feedback in speech production, particularly in early language acquisition where learners use auditory feedback to inform somatosensory targets and both play a role in motor planning. Hickok (2014) further argues that the two types of feedback inform different levels of motor planning, where the auditory feedback informs syllable targets and their production and the somatosensory target informs the segmental and subsegmental organization of gestures within the syllable. The Coupled Oscillator Model assumes that the onset and coda are planned and coordinated in-phase with one another, while the coda is anti-phase to the nucleus. In-phase planning of CV sequences is suggested by studies showing that the initiation of the motor plan for both C and V gestures are synchronous (Kozhevnikov and Chistovich, 1966; Löfqvist and Gracco, 1999; Nam et al., 2009). The synchronized planning of the CV onset results in a tighter and more stable timing relationship, giving rise to patterns of global timing in CV onsets and to the privileged status of onsets cross-linguistically.

Both articulatory and acoustic studies provide phonetic evidence for syllables, supporting the idea that syllables are a unit in motor planning. Acoustic studies demonstrate that there is compensatory shortening of the vowel in the presence of one or more onset consonants or one or more coda consonants (Katz, 2012; Marin and Pouplier, 2010; Munhall et al., 1999). Furthermore, at least in languages like English, there is increased shortening of the nucleus in the presence of additional consonants in the onset. This reorganization of timing is referred to as complex onset timing or global timing. Results of incremental shortening for complex codas have varied, with overall less support for incremental shortening of vowels in the presence of additional coda segments; Katz (2012) finds that there may be difference in compensatory shortening based on the quality of segments in the coda, where /l/ codas do cause incremental shortening. Katz (2012) argues that this effect may be due to the perception of segments in different syllable positions, advocating for a theory of syllable coordination that incorporates perception, unlike the Coupled Oscillator Model.

Studies on the gestural organization of syllables have found similar results to acoustic studies, illustrating that in English there is tighter coordination of segments in the onset than in the coda and additionally that there is a reorganization of gestures between the onset and the nucleus, an effect referred to as c-centering (Browman and Goldstein, 1988; Byrd, 1995). Articulatory studies also find slight differences in coordination based on segmental quality. For instance, in English, the timing of /pl/ clusters is less variable and more tightly coordinated with the vowel than /sp/ clusters. Goldstein et al. (2009) again attribute this effect to the /l/, but rather than perceptual differences, the difference is the articulation of /l/, which has two active tongue gestures causing it to be more closely coupled with the following vowel. A similar proposal for coda /l/ has been argued by Walker and Proctor (2019).

While much of the research in this area has been done on English, there have been similar studies conducted on a number of other languages. For instance, similar patterns have been found for German (Bombien et al., 2010) and Georgian (Goldstein et al., 2007). Moroccan Arabic, on the other hand, has been found to have local timing in onset clusters, resembling patterns in English codas (Shaw et al., 2009). Italian provides yet another organization of onsets, with different coordination structures for different kinds of complex onsets. In stop+liquid clusters, Italian patterns similarly to c-centering clusters in English; however, sibilant+stop clusters do not show the same strong c-center effect as stop+liquid clusters, and thus, Hermes et al. (2013) argue that the data support extrametricality for sibilant+stop clusters in Italian.

These studies demonstrate that the syllable is powerful both for phonological modeling and in accounting for articulatory and acoustic effects, where both segmental gestures and interaction across sequences of gestures are elegantly accounted for through syllable structure. These effects are less clearly accounted for without syllable structure; nevertheless, these studies concentrate on sequences where syllabification is straightforwardly presupposed, often analyzing initial and final sequences and sequences across word boundaries; however, as Section 1.4 discusses in detail, syllabification is often not straightforward, particularly in word-medial contexts, necessitating studies of coordination across a broader range of syllabification contexts.

1.4 Syllable margins

Given the absence of phonetic studies disentangling syllabification, phonologists have relied on frameworks of syllabification that infer syllable boundaries from phonotactic distributions; specifically, the hypothesis that if a sequence can occur at a word boundary, either onset or coda, it is presumed to be a licit word-medial onset or coda. This framework falls out of the assumption that all segments can be discretely and exhaustively syllabified into syllables. Under such an assumption, all onsets and codas are expected to behave uniformly, regardless of whether they are at a word boundary or are word internal. Likewise, if phonotactics and processes like lenition are sensitive only to syllable position, then it is predicted that word-initial and word-medial onsets and word-final and word-medial codas behave alike. However, as this section will demonstrate, there is abundant evidence that this is not the case. Cross-linguistic data shows that differences between word-boundary and word-medial positions are common, and furthermore, that phonological patterns differ across boundaries. This section will first discuss exceptional structures at word boundaries, and will then discuss patterns and processes common word-medially, exemplifying a few of the challenges to syllable structure and present the frameworks and mechanisms that have been proposed to account for these challenges.

1.4.1 Word boundary effects

Theories of syllabification and coordination like the Syllabification Algorithm and the Coupled Oscillator Model predict that onsets and codas pattern uniformly with other onsets and codas regardless of word shape; however, this assumption is problematic as differences between word-boundary syllable margins and word-medial syllable margins, referred to as edge-effects, are common cross-linguistically, as shown in Table 1.1. For instance, in Klamath, consonant clusters are permitted word-initially but not word-medially (Blevins, 1993).

	iso	Complex Nucleus	Obligatory Onset	Complex Onset	Coda	Complex Coda	Edge-Effect
Totonac	tlc	yes	yes	yes	yes	yes	F
Klamath	kla	yes	yes	no	yes	yes	I
English	eng	yes	no	yes	yes	yes	F
Nisqually	slh	no	yes	yes	yes	yes	F
Gilyak	niv	no	no	yes	yes	yes	F
Finnish	fin	yes	no	no	yes	yes	no
Tunica	tun	no	yes	no	yes	yes	no
Tamazight Berber	tzm	no	no	no	yes	yes	F
Sedang	sed	yes	yes	yes	yes	no	I
Cairene	arz	yes	yes	no	yes	no	F
Spanish	spa	yes	no	yes	yes	no	F
Dakota	dak	no	yes	yes	yes	no	F
Italian	ita	no	no	yes	yes	no	I, F
Mokilese	mkj	yes	no	no	yes	no	F
Thargari	dhr	no	yes	no	yes	no	no
Cuna	kvn	no	no	no	yes	no	no
Arabela	arl	yes	yes	yes	no	no	no
Siona	snn	yes	yes	no	no	no	no
Pirahã	myp	yes	no	yes	no	no	no
Piro	pib	no	yes	yes	yes	no	I
Mazateco	mau	no	no	yes	no	no	no
Fijian	fij	yes	no	no	no	no	no
Hua	yga	no	yes	no	no	no	no
Cayuvava	cyb	no	no	no	no	no	no

Table 1.1: Cross-linguistic survey of edge-effects. In the edge-effect column, 'I' stands for initial effects and 'F' stands for final effects (Blevins, 1995).

Languages that have the same restrictions both at word boundaries and word-medially are referred to as symmetrical, whereas languages with differing restrictions between medial and boundary positions are referred to as asymmetrical (Blevins, 1995). Examples of languages with symmetrical syllable margin effects include Manam, in which only nasals occur in codas both medially and finally; Spanish, in which both medially and finally any consonant and an optional /s/ can occur in coda position; and Selayarese, in which all codas both medially and finally are characterized by a lack of place specification.

There are two major kinds of word-positional asymmetries, asymmetries in segment distribution and asymmetries in the behavior of consonants in syllable margins. The degree of cluster complexity a language allows can differ between word-medial and word-boundary positions, where word-boundary clusters are more complex than word-medial clusters. For instance, in both Eastern Ojibwe and Tojolabal, a wider range of segments both in manner and laryngeal features are permitted in word-final position. As shown in example (5), fricatives and nasals in Eastern Ojibwe are attested as codas both word-finally and word-medially; however, stops are only attested word-finally (Côté, 2011, p.3).

(5) a. [baŋgisin] 'it falls'
b. [mo:ʃkine:] 'it is full'
c. [wi:ja:s] 'meat'

d. [nindib] ‘my head’
e. [ninik] ‘my arm’

Another type of edge-effect is phonological invisibility, in which segments at word boundaries appear to be invisible to phonological processes. For instance, in both French and English, a larger number of consonants can appear in the coda word-finally than word-medially, and furthermore, in English, co-occurrence of vowel and coda consonants differ word-medially and word-finally. Stem-finally, English permits long vowels to be closed by any consonant; however, in medial contexts, long vowels in closed syllables occur only in restricted contexts: prior to a fricative+stop (e.g., *pastry auspices*, *after* (Côté, 2011, pg.4)) or a sonorant homorganic with the following onset (*council*, *chamber*, *example* (Côté, 2011, pg.4)). Phonological invisibility also manifests in stress patterns, where final consonants seem to be invisible to stress assignment; examples include German, Menominee, English, and Icelandic, shown in example (6) (c.f. Côté, 2011).

(6) Icelandic
a. [’pu:] ‘estate’
b. [’sta:ra] ‘stare’
c. [’pru:n] ‘edge’
d. [’θa:k^h] ‘roof’
e. [’sɛnta] ‘send’
f. [’flaska] ‘bottle’
g. [’t^{h}jalt] ‘tent’
h. [’rixs] ‘rich (GEN SG MASC)’

The most common framework for modeling asymmetrical edge-effects is extrametricality (Hayes, 1982) in which final consonants are invisible for the purposes of syllabification and stress assignment and are later adjoined to a higher level of structure, typically the prosodic word, as shown in Figure 1.2.

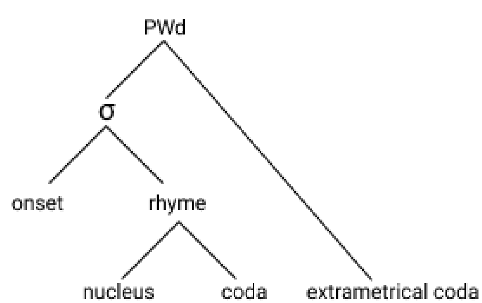


Figure 1.2: Prosodic word attachment

The ability for segments to attach to the prosodic word rather than the syllable is restricted to word boundaries, thus constraining the same skeletal slots from occurring word-medially and providing a structural basis for edge-effects. In addition to extrametricality as a phonological model, there is also phonetic support for extrametricality, e.g., coordination in Italian (Hermes et al., 2013). While extrametricality provides a straightforward model of edge-effects, neither the model nor the phonetic support for extrametricality provide an explanation for why edge-effects are common across languages. Nonetheless, patterns of extrametricality suggest that there may be differences between segments and syllables at word edges and word-medially, whether perceptually, articulatorily, or both.

1.4.2 Lenition

Like phonotactic differences between word margins, phonological processes like lenition differ depending on whether they occur at a word boundary or are word internal. Lenition is generally described as segmental weakening and often refers to processes where there is a reduction in the degree or length of constriction, e.g., a geminate becoming a singleton stop or a stop becoming a fricative (Foley, 1977); however, it is difficult to define precisely what weakening refers to as the term is used to refer to a wide range of processes. Broadly, it includes both patterns such as a stop becoming a fricative intervocalically, e.g., Tümpisa Shoshone /wisipin/ ⟶ [wiʃiɸi] 'thread' (Gurevich, 2011), as well as cases of final devoicing, Bulgarian *gradove* 'cities'   *grat* 'city'.

According to Gurevich (2011), while lenition is often treated as a unified class of effects at word edges and word-medially, constraining the nature of processes that fall within the frame of lenition is less clear because it includes both processes such as intervocalic voicing, e.g., VTV ⟶ VDV and final devoicing D# ⟶ T#. Kingston et al. (2008) and Katz (2016), however, divide word-boundary lenition and word-medial lenition into two types. The first type of lenition Katz (2016) terms *loss lenition*, where loss lenition typically involves the loss of features, length, or gestures, e.g., debuccalization, typically targets prosodically weak positions, such as word-final positions, and may result in contrast neutralization. The second type is termed *continuity lenition* and serves to minimize auditory disruption, and unlike loss lenition, occurs at prosodically strong positions and does not result in contrast neutralizations. In particular, continuity lenition often targets segments in intervocalic positions and results in consonants that are shorter or more intense than their counterparts in other syllable positions. By dividing lenition in this way and in particular by framing continuity lenition as a prosodic maintenance, Katz (2016) presents a clear typology of lenition processes, and furthermore, it is able to cast patterns of lenition and fortition in different prosodic positions in the same light.

Word-medial lenition is relatively predictable cross-linguistically because it often involves a predictable hierarchy of shift and often progresses synchronically along the same hierarchy, illustrated in (7).

(7) tt, t^h » t » d » θ, ð, ɾ » ʔ, h, j » ∅ (Foley, 1977)

Lenition typically involves a full class of segment shifting; for instance, all voiceless stops may become voiced intervocalicly. Additionally, lenition often begins in fast or casual speech as a type of reduction, but often progresses to become a systematic phonological alternation (Gurevich, 2011). Table 1.2 illustrates a lenition process in Spanish, where voiced stops surface as voiced stops word-initially or following a nasal, but surface as continuants between vowels, glides, and approximants (Katz, 2016, pg. 8).

#__		[+approx]_V		V_[+approx]		[+nas]_V	
goðo	'Goth'	bisiɣoðo	'Visigoth'	reɣla	'rule'	uŋgoðo	'a Goth'
beso	'kiss'	elβeso	'the kiss'	oβɾa	'work'	umbeso	'a kiss'
dia	'day'	ojðia	'nowadays'	siðɾa	'cider'	undia	'a day'

Table 1.2: Example of lenition in Spanish (Katz, 2016, pg. 8)

Both articulatory-based and perception-based explanations have been proposed to motivate word-medial lenition. From an articulatory stance, lenition effects have been attributed to a reduction in articulatory effort, which (Kirchner, 2001) proposed as a general pressure the grammar can encode. However, Kingston et al. (2008) argue that the difference in articulatory effort between, e.g., /tt/ and /t/, is too minuscule to motivate lenition and instead propose an alternative phonetic mechanism: prosodic maintenance. Kingston et al. (2008) argue that lenition occurs within a prosodic constituent, and serves to decrease disruption to the speech stream through sonority promotion, signaling to listeners where prosodic boundaries occur and where they do not. Katz (2016) likewise advocates that word-medial lenition is motivated by prosodic factors rather than syllable position, where lenition or the absence of lenition can communicate meaningful information to the listener about word and syllable boundaries. While prosodic maintenance clearly and accurately captures the differences between lenition at word boundaries and word internal lenition, it does not capture why lenition often starts as a process in fast speech which then progresses to be phonologized, whereas an articulatory account can capture this generalization. However, missing from the articulatory account is a difference between word-medial and word-boundary articulation, necessitating the analysis of the articulatory differences between these boundaries.

Hyman (2011) similarly demonstrates that if word-initial and word-medial boundaries are presumed to pattern uniformly, it is difficult to capture patterns of lenition. Hyman (2011) explains that the syllable cannot account for the the distribution of consonants in Gokana. In Gokana, any consonant may occur in the stem-initial consonant position; however in a sequence like $C_1VC_2(V)$, C_2 can only be occupied by one three archiphonemes /B, D, G/, which may surface as oral or nasal. In the oral condition, [b, l, g] are realized in final position and [v, r, g] are realized intervocalically. Hyman (2011) explains that syllable structure cannot explain this distribution as C_2 is sometimes in onset and sometimes in coda position, and furthermore, the distribution of C_2 differs from C_1 even when both are in onset position.

The studies presented in this section clearly demonstrate that syllables often do not behave identically at word boundaries and word-medially. In particular, lenition patterns provide a clear distinction in phonological patterns at word edges compared to word-medial positions. While loss lenition, which occurs at word boundaries, typically consists of the loss of a segment or feature, continuity lenition, which occurs word-medially typically involves a change in features. Ultimately, word-medial lenition is likely to be driven by by both articulatory and perceptual factors, as both capture components of cross-linguistic patterns. Regardless of whether the difference in word-medial lenition is driven by prosodic or articulatory factors, the differences in patterns between environments suggest that there is a difference in word-medial environments that should be captured in phonological models.

1.4.3 Ambisyllabicity

Another phenomenon that presents challenges for the idea that segments can be discretely and exhaustively syllabified and that those margins behave uniformly is ambisyllabicity. Phonological patterns, like lenition, as well as distributional patterns differ word-medially from word edges. For example, in English, the distribution of /ŋ/ is restricted to coda position, as shown in examples (8-a) and (8-b) (there is no known word in English beginning with /ŋ/). However, in word-medial contexts, a single word-medial consonant is expected to syllabify as an onset due to the constraint that all syllables have an onset, which presents a complication for the distribution of /ŋ/, as shown in (8-c), which would be syllabified as an onset according to the usual assumptions of syllabification (Hayes, 2009).

(8) a. *sing* [sɪŋ]
 b. *[ŋap]
 c. *gingham* [’gɪŋəm]

Much like extrametricality, ambisyllabicity provides structure to the basic syllable architecture that can account for patterns like the distribution of /ŋ/. In particular, ambisyllabicity assumes that segments can be simultaneously associated with multiple syllables, behaving both as a coda for one syllable and the onset for the next. Kahn (1976, p. 55-56) extends his proposed syllabification algorithm to include rules that generate ambisyllabicity in English:

(9) III At normal and fast speech rates, in a sequence of two syllables where the second syllable is unstressed, associate the coda of the first syllable with the onset of the second syllable where
 IV The the sequence of consonants does not for a universally prohibited cluster.
 V In connected speech in a sequence of two words, where the first word ends with a consonant and the second word ends with a vowel, associate the word final coda with the first syllable of the following word.

Rules III and IV are specific to fast speech in English and unlike nucleus and onset syllabification, ambisyllabicity is language-specific. Kahn (1976) demonstrates how these rules are able to accurately capture the distribution of voiceless coronal stops in English, where flapping is the result of ambisyllabicity. Rules III and IV create associations that link single segments to multiple syllables, segments associated with multiple syllables are referred to as ambisyllabic and ambisyllabic segments pattern differently from consonants that are only the onset or the coda of a syllable. For instance, /t, d/ in words like *butter* and *ladder* are produced as a /ɾ/. Kahn (1976) and Rubach (1996) propose that the onset of unstressed syllables is associated with the coda of the preceding syllable in fast speech, resulting in an ambisyllabic segment, which in the case of /t/ and /d/ in English, is flapped. Using the syllabification principals outlined in I-V, Kahn (1976) provides an account for a number of phonological alternations in English. To name a few: distribution of aspiration for voiceless stops and allophonic distribution of coronal stops, /ɹ/ loss in non-rhotic dialects, pre-/ɹ/ vowel quality, and simplification of /ŋ/.

A number of counter proposals to ambisyllabicity have been argued to account for the same phonological alternations, while improving predictions regarding the distribution of processes like flapping. For instance, Jensen (2000) argues that prosodic hierarchy is a better predictor of flapping than ambisyllabicity. Under such an account, a /t/ or /d/ remains the syllable onset and only the syllable onset, but the position of the syllable with regard to stress and foot determines the alternation. In particular, according to the prosodic hierarchy, words are divided into feet consisting of two syllables and this footing is the basis of stress assignment. Flapping occurs when /t/ or /d/ is between vowels and not foot initial. An advantages of the prosodic hierarchy is that it predicts a difference in the flapping of /t/ between examples like 'butter' and 'Mediterranean'. Additionally, Blevins (1995) and Jensen (2000) argue that ambisyllabicity undesirably complicates syllabification without evidence that there are any three way minimal pairs with segments being syllabified to the left to the right or across both[1]. Selkirk (1982) accounts for ambisyllabicity patterns through resyllabification, arguing that in cases where /t/ ⟶ /ɾ/, /t/, e.g., 'butter', begins as the onset of the second syllable, but the stress of the first syllable causes the segment to resyllabify as a coda of the stressed syllable. The process of resyllabification results in flapping.

While both resyllabification and prosodic accounts have their advantages, they are not without its own shortcomings. For example, neither resyllabification nor foot-based accounts accurately predict the flapping of /t/ in words like 'capitol'. Furthermore, while resyllabification and prosodic accounts argue against ambisyllabicity as the right modeling tool for such sequences, the underlying phenomena captured by accounts of ambisyllabicity is not under dispute nor is the need to account for such patterns beyond what core syllabification can provide.

Prosodic accounts approach the issue from the perspective of what motivates processes like flapping, arguing that the underlying mechanism is not syllable structure, but stress, making a prosodic approach preferable. However, there are likely a mix of underlying factors that motivate the differences between word-medial and word-boundary environments, as will be discussed in more detail in Section 1.5. Thus, despite the differences between word-medial and word-boundary environments, another way to approach the question of how best to account for these patterns is to ask whether there are shared properties between word-medial syllables and word-boundary

[1]However, Blevins (1995) also argues that syllabification is not underlying and that there are few if any syllabification minimal pairs; thus, the lack of minimal pairs is not unique to ambisyllabicity. Nevertheless, there are morphologically conditioned differences in VCV syllabification resulting in three levels of distinctiveness, e.g., 'light a match', 'lightning' and 'lie test.' In addition, where there are examples of contrastive syllabification, it often involves vowel syllabification. For instance, in Japanese and Turkish, there is an underlying distinction between /a.a/ and /a:/. Blevins (1995) likewise notes the distinction between Ida [ai.da] and Aida [a.i.da] might constitute an syllabification minimal pair in English; however, this distinction could also be analyzed as Ida [ai.da] and Aida [ai.i.da]. Together these facts can be taken to support that there may be some underlying architecture for syllables, but syllabification may not be underlyingly exhaustive.

syllables that give preference to a unified theory of syllables. Theories of coordination like the Coupled Oscillator Model make such predictions that syllable coordination should be the same word-initially and word-medially; however, these predictions have yet to be tested.

1.5 Segmenting the syllable

Theories distributing segments within the syllable are of central importance to theories of syllable structure, particularly where they inform the division of syllables at syllable margins. While vowels uncontroversially occupy the nucleus of the syllable, consonants occur in onsets, nucleus, and coda positions, and thus present an open question regarding syllabification. Consonants occurring at word boundaries can be straightforwardly assigned to the onset or coda, when occurring in word-initial or word-final position, respectively, or represented as extrametrical insomuch that extrametrical segments differ from other onset and coda segments. Word-medial consonants, however, may be syllabified as codas or onsets with a number of factors influencing syllabification and a number of theoretical approaches to syllabification. Relevant to the question of how segments are distributed within the syllable is syllabification itself, which is influenced by a complex set of factors such as segment quality, stress, perception, and articulation.

1.5.1 Segmental distribution

A number of theoretical proposals have been made in predicting cross-linguistic patterns of phonotactic distribution. While sonority was an early theory in accounting for syllabification patterns (Clements, 1990; Foley, 1970), later theories like licensing by cue (Steriade, 1997) and the Frame-Content theory (MacNeilage, 1998) demonstrate that more general principles of perception and production capture the same generalizations that sonority captures while avoiding problematic predictions that fall out of the principles of sonority. Despite theoretical developments in accounting for distributional patterns of syllabification, sonority remains a commonly relied on architecture in accounting for syllabification. In particular, phonologists have long noticed that the division of segment strings into units generally agreed upon as syllables are determined by two key generalizations that sonority and place of articulation can account for a large portion of phonotactic patterns; however, while theories of sonority have been widely developed, few theories incorporate observations about place of articulation into syllable architecture. This section will highlight the strengths and weakness of syllabification theories to exemplify cross-linguistic phonotactic patterns and how these distributional patterns are accounted for within the architecture of syllables established thus far. Furthermore, this section will identify how current theories account for syllabification to lay a foundation of how the findings of this dissertation, which incorporate observations about the role of place of articulation in segmental distribution, can be incorporated into existing frameworks of syllabification.

1.5.1.1 Sonority

According to the theory of sonority, the syllable is defined as a chunk of the word centered around a single relative sonority peak, where sonority rises from the onset to the vowel and falls from the vowel to the coda. This effect is referred to as the SONORITY SEQUENCING PRINCIPLE (SSP). According to the SSP, segments can be ranked on a scale of sonority, shown in (10) (Clements, 1990; Foley, 1970).

(10) Vowels » Glides » Liquids » Nasals » Fricatives » Oral stops

Although there are debates on the degree to which this scale is universal and a number of changes have proposed to this scale, the general pattern proposed for the sonority sequencing principle is able to account for a large swath of consonant distribution in onsets and codas. Cross-linguistically, there is a tendency for a rise in sonority in the onset and a fall in sonority in codas. For instance, for languages that permit complex onsets, stop+liquid clusters are widely attested, whereas the reverse is exceedingly rare and there are no such languages that permit a liquid+stop onset cluster, but ban stop+liquid clusters. The same pattern holds for the coda in reverse where a liquid+stop cluster is preferred in the coda. The robustness of this pattern is demonstrated by Greenberg's (1965/1978) survey of 104 languages. Greenberg's survey further demonstrates an implicational hierarchy for sonority: if a language has clusters that violate the sonority sequencing principle, it also has clusters that adhere to the principle (Gordon, 2016). Sonority sequencing makes many strong predictions about the clusters that are permitted in either margin.

Languages differ in the shape of the rise or fall in sonority in onsets and codas, respectively. For example, complex onsets in English can consist of both stops and fricatives in the first consonant position and either liquids or glides in the pre-vocalic consonant position; in contrast, Awa Pit likewise permits complex clusters, but pre-vocalic and post-vocalic segments are restricted to glides (Curnow, 1997). Cross-linguistic differences in sonority profile are accounted for through the MINIMUM SONORITY DISTANCE PRINCIPLE (Harris, 1983; Selkirk, 1981), which proposes that clusters combine based on their relative distance in the sonority scale and that languages can differ in the minimum required distance along the sonority scale.

While the overall sonority profile is mirrored between onsets and codas, there are differences in the sonority preferences of onsets versus codas. Cross-linguistically, highly sonorous segments, like /m/ or other nasals are typically preferred in coda position while low sonority segments like voiceless stops are typically preferred in onset position (Clements, 1990). Sonority is not the only distinction between the distribution of segments in onsets and codas. While there are typically few restrictions on what segments can occur in the onset, the distribution of segments in the coda is much more limited. For instance, in Nafaanra, a Senufo language spoken in Ghana, 29 out 30 total consonants may occur in the syllable onset, however, only /m/ may occur in coda position[2] (Garvin, 2017).

Sequences that do not adhere to the SONORITY SEQUENCING PRINCIPLE are typically syllabified

[2]In Nafaanra, /r/ is the only phone that does not occur word-initially as a singleton onset. /r/ only occurs word-initially in clusters and also occurs word-medially as a singleton onset (Garvin, 2017)

across a syllable boundary. Kahn (1976) stipulates as a part of the onset syllabification rule that the sequence must be a licit sequence word-initially to be syllabified as an onset, otherwise the initial portion of the sequence is left to be syllabified as a coda; this is where syllable contact comes in. According to the Syllable Contact Law (SCL) (Murray and Vennemann, 1983), in a sequence of two syllables, there is a preference for the first syllable margin to be of higher sonority than the segment of the following onset (Clements, 1990). These observations are similarly operationalized in OT, where constraints assess violations for sequences of segments that do not adhere to the SSP or the SCL. These constraints interact with language specific constraints on sonority slope and faithfulness constraints, resulting in repair strategies like epenthesis or deletion; languages differ in degree of preferred sonority slope and in repair strategies where sequences do not adhere to the SCL. For instance, Kazakh requires that sonority fall across the syllable boundary, but permits any sequence with falling sonority; however, in Sidamo, sonority must fall by a certain degree and Kirghiz requires an even steeper sonority slope (Gouskova, 2004).

While sonority provides an account for a number of cross-linguistic patterns in the distribution of segments within the syllable and syllabification, the theory also makes a number of spurious predictions and there are widespread exceptions to sonority sequencing, even in English. For example, coronals, and in particular the distribution of /s/ and /t/, is much freer than is predicted by sonority sequencing principles. For instance, English permits /s/+stop clusters in onsets and stop+/t/ and stop+/s/ clusters in codas, although it does not permit other fricative+stop or stop+stop clusters. Acoma provides another example, where /s/+stops are allowed but no other clusters are permitted (Gordon, 2016). Likewise, while sonority sequencing predicts that stop+nasal clusters should be common cross-linguistically, in fact, nasal+stop clusters are much more common across languages (Gordon, 2016; Wright, 2004).

Other counter examples to sonority sequencing are exceptions to the syllable contact law, where in both Icelandic and Faroese sonority may rise across the syllable boundary, although there are constraints on the degree of rise. Repair strategies for SCL violations also differ across languages. For instance, in Kazakh, coda+onset sequences that do not adhere to SCL result in onset desonorisation (Gouskova, 2004, pg. 202).

(11) a. /kol-lar/ ⟶ kol.dar ‘hands’
b. /murin-ma/ ⟶ mu.rin.ba ‘nose-INT
c. /koŋɯz-ma/ ⟶ ko.ŋɯz.ba ‘nose-INT
d. al.mma.lar ‘apples’
e. kol.ma ‘hands-INT’
f. ki.jar.ma ‘cucumber-INT’

In Icelandic and Faorese, where sonority rises excessively across the onset+coda, the coda is instead syllabified as a complex onset, resulting in vowel lengthening. The languages differ, however, in the degree of permitted rise, with less rise permitted in Faroese. A final example comes from Sidamo, in which complex onsets are not permitted, and thus, instead of resyllabification at the syllable boundaries, the segments are metathesized, e.g., /tn/ ⟶ [n.t], /lt/ ⟶ [l.t] (Gouskova, 2004). The significance of investigating syllable contact is not only in extending the role of sonority in syllabification, but the trade-off between onsets and codas in adjacent syllables and the significance of the rhyme in phonological patterns. Nevertheless, sonority provides a straightforward account for syllabification that is easily implemented into frameworks like OT. The simplicity and predictive power of the theory are perhaps why sonority persists in phonological theory despite its spurious predictions.

1.5.1.2 Licensing by cue

Licensing by cue approaches the question of segmental distribution and syllabification from the perspective of perceptual stability. According to Fujimura et al. (1978), the CV syllable structure is motivated by the acoustic clarity of the consonant vowel transition, which listeners are better able to perceive than VC transitions. In particular, Fujimura et al. (1978) conducted an experiment in which English and Japanese speakers were asked to identify an intervocalic consonant in which the VC transition cued one consonant, while the CV transition cued another consonant. The study found that for both languages, the CV transition was the more stable cue [3]. This stability in perception suggests that the preference for CV syllables may be driven by perception and that the stability in perception may influence syllabification.

Steriade (1997) proposes a framework within phonological theory that draws on these same principles, that the stability of perception cues in CV sequences drives the phonological processes in these environments. In particular, Steriade (1997) argues that contrasts in less perceptible positions, e.g., postvocalic consonants, are more likely to be neutralized than segments in easily perceptible sequences, e.g., prevocalic consonants. Steriade (1997) illustrates this claim with data from Klamath, in which contrastive aspiration is neutralized word-finally and before another obstruent (Blevins, 1993). Both contexts can be classified under coda position, and thus, a possible account for this laryngeal neutralization is that laryngeals are neutralized in coda position. Steriade (1997) argues that it is not innately the coda position that drives this neutralization, but instead that obstruent and word-final laryngeal segments are cue impoverished, resulting in neutralization. Thus, syllabification is emergent from other phonetic properties driven by pressure to minimize confusability, with particular emphasis on cue licensing.

The appeal of this outlook on syllable structure is that it captures many of the same patterns that sonority sequencing captures while avoiding some of the spurious predictions made by sonority sequencing. For instance, under a licensing by cue perspective, /s/+stop clusters are predicted to be common as /s/ is highly perceptible in this environment, and thus, need not be restricted to a pre-sonorant position. Likewise, nasal+stop clusters are predicted to be more common than stop+nasal clusters due to the perceptual salience of oral release bursts compared to nasal release bursts (Wright, 2004). Licensing by cue can also account for restriction on coda segments such that segments in codas are only licensed if they match in place of articulation to the following segment, e.g., Japanese, where there codas are only sufficiently cued for place of articulation when place is shared with the following onset, which has more robust place of

[3] The languages differed on the role of stress in the stability of this cue, which will be returned to in Section 1.5.2.2.

articulation cues (Gordon, 2016). However, while integrating perceptual salience into our understanding of syllabification provides an improvement over sonority in predicting typological patterns in phonotactics, it does not provide a solution to all of the question presented by sonority. For instance, perception cannot provide an explanation as to why coronals have a broader distribution than other stop segments, such as sequences of stop+stop clusters in English, e.g., 'kept' but *'ketp'. Likewise, in Finnish, only coronals may occur in coda position (Gordon, 2016). Furthermore, the mapping of global timing patterns, i.e., compensatory shortening and c-centering, or differences in the production of segments in onset compared to coda position, are not clearly deducible from a perception driven account of syllabification.

1.5.1.3 Articulation and the Frame-Content Theory

A final approach to understanding the distribution of segments within the syllable is through the articulatory pressures on phonotactics, which have the advantage of capturing both the basic principles of sonority and patterns influenced by place of articulation. Theories of speech production and motor-planning (e.g., Fowler and Saltzman, 1993) assume that intrinsic constraints like the muscular control and airflow also influence articulation. The predictions of the Frame-Content Model capture this, focusing on the evolution of syllables in speech production. MacNeilage (1998) argues that vowel and consonant productions correspond to opening and closing modulations in the jaw. MacNeilage (1998) explains that this modulation of the jaw constitutes the speech *frame* and that the segmental units then fill in the *content* of the syllable. Opening and closing frames are present in the babbling of early child language productions and likewise correspond to non-speech tasks as well, such as food ingestion [4]. Thus, MacNeilage (1998) hypothesizes that these modulations might have been well primed for extension to speech production.

Both typological evidence and articulatory evidence provide support for the Frame-Content Model. For instance, there is a correlation in the co-occurrence of vowels and consonants and of consonants at differing positions in the syllable. Vallée et al. (2009) find that coronal consonants are more likely to occur with front vowels, labial consonants to occur with central vowels, and velar consonants to occur with back vowels. They relate these patterns to the Frame-Content Theory, which argues that jaw movement plays an influential role in syllable structure. Vallée et al. (2009) additionally demonstrate a correlation between consonant co-occurrence in onset in coda and in C_1 and C_2 onsets in CV.CV sequences. They explain that the first C in either CVC or CV.CV sequence is more likely to be bilabial and the second consonants is more likely to be coronal. The authors state find that even in a sequence of two syllables, if consonants are sequenced as bilabial and coronal, then can be coordinated with a single jaw movement, whereas a coronal + bilabial sequence requires two jaw movements. Vallée et al. (2009) attribute this to the ability to prepare coronal articulation simultaneously with other gestures without the same degree of jaw coupling required for the gestures of bilabial and velar segments.

An articulatory approach can also account for the frequency of NC sequences compared to CN sequences. Vallée et al. (2009) explain that the high intraoral pressure of oral stops increases

[4]Hickok (2014) similarly argues for a production model that includes syllables based on early speech patterns

the volume of the vocal tract and pushes the velum upward, causing widening of the vocal tract and velar region. To produce a subsequent nasal, the velum must then be lowered to open to the velo-pharyngeal port despite the high intraoral pressure, which then results in a drop of the intraoral pressure. In the reverse sequence of nasal+stop, the increase of intraoral pressure actually facilitates closure of the velopharyngeal port. Thus, the gestures of nasal+stop sequences are facilitatory, whereas the gestures of stop+nasal sequence are inhibitory, providing a clear articulatory explanation for the pattern of nasal+stop, stop+nasal sequences cross-linguistically.

Additionally, while sonority sequencing accounts for much of the segmental distribution in syllables, there are, nevertheless, a number of languages for which sonority does not provide a sufficient solution to segmental distribution. For instance, Georgian permits stop+stop sequences and onset clusters of up to six segments (Butskhrikidze, 2002).

(12)
a. bgera ‘sound’ (Chitoran et al., 2002)
b. abga ‘saddle bag’ (Chitoran et al., 2002)
c. p^ht^hila ‘hair lock’ (Chitoran et al., 2002)
d. ap^ht^hari ‘hyena’ (Chitoran et al., 2002)
e. dgeba ‘s/he stands up’ (Chitoran et al., 2002)
f. adgeba ‘s/he will stand up’ (Chitoran et al., 2002)

(13)
a. gberavs ‘s/he is inflating you’ (Chitoran et al., 2002)
b. dagbera ‘to say the sounds’ (Chitoran et al., 2002)
c. t^hbeba ‘it is warming up’ (Chitoran et al., 2002)
d. gat^hba ‘it has become warm’ (Chitoran et al., 2002)
e. gdeba ‘to be thrown’ (Chitoran et al., 2002)
f. agdeba ‘to throw something in the air’ (Chitoran et al., 2002)

Sonority cannot account for these sequences, however, articulation may provide an account for Georgian phonotactics. Chitoran et al. (2002) compare the articulatory timing of sequences like those found in (12), where the segments are sequenced from front-to-back, with those in (13), where the segments are sequenced from back-to-front and found that those sequenced from front-to-back were more tightly coordinated than the back-to-front clusters. The authors attribute this effect to perceptual factors, where back-to-front segments need to be sequenced further apart to be perceptible, but an articulatory account complements the perceptual account, where front-to-back sequences can be coordinated in a single jaw opening gesture, but back-to-front sequences require multiple jaw gestures or at least less efficient jaw opening phase. Thus, assuming that articulatory pressures drive syllabification captures many of the same predictions that sonority captures and likewise captures a number of generalizations about place of articulation that other theories of syllabification do not predict.

1.5.2 Pressures contributing to dynamic and variable syllabification

Despite the evidence in favor of syllables as intuitive linguistic units and the utility of the syllable unit in accounting for cross-linguistic patterns, syllabification, i.e., dividing segments into onset and coda has remained a challenging issue. For instance, although studies have shown that speakers are attuned to syllable beats (Morton et al., 1976; Oganian and Chang, 2018), studies asking speakers to provide judgments on syllabification have found inconsistent results dependent largely on task rather than syllable structure (Côté and Kharlamov, 2011). This difficulty in obtaining syllabification judgments is likely due to the fact that there are a number of simultaneous pressures on syllabification and often multiple plausible syllabifications of a given sequence. For instance, a sequence of /str/ in a word like 'destruction', both *des.truction* and *de.struction* are plausible syllabifications. While the syllabification algorithm predicts that this sequence should be syllabified as *de.struction*, as /str/ is a licit onset sequence, licensing by cue and the Frame-Content Model predict both *de.struction* and *des.truction* to be possible.

Furthermore, there are additional factors that affect gestural coordination that may ultimately affect syllabification and related phonological processes. For example, in addition to word-medial flapping, flapping can also extend across word boundaries and even across phrase boundaries (Kilbourn-Ceron et al., 2016). This external flapping process is variable and Kilbourn-Ceron et al. (2016) demonstrate that a number of factors influence the likelihood of flapping, including phrase boundary strength and the frequency of the following word. Thus, there are likely a wide range of factors that influence articulation and coordination in production. Focusing on word-medial effects, this section will explore two additional pressures on coordination: segment quality (Section 1.5.2.1) and stress (Section 1.5.2.2) to raise the question of what competing pressures are likely to affect coordination and likewise syllabification.

1.5.2.1 Segment quality

Theories of motor planning and speech production (e.g., Fowler and Saltzman, 1993) assume that there are intrinsic pressures like muscular control and airflow that influence gestural coordination. For instance, segments that have gestural targets that are near each other will require less time for the articulators to move from point A to point B and therefore can be timed more closely to one another. The Frame-Content Model likewise assumes that intrinsic pressures of articulation affect syllable shapes cross-linguistically, focusing in particular on how the movement and articulatory coupling of the jaw shape sequences of segments. For example, as previously discussed, Vallée et al. (2009) find that there is a correlation between consonant and vowel quality, where front vowels occur more frequently with coronal consonants and back vowels occur more frequently with dorsal consonants. This may suggest that sequences matching in front or back feature may be more tightly coordinated or may be more likely to syllabify together than those that contrast in front and back feature. Similarly, Munhall et al. (1999) demonstrate that adjacent segments that share a given feature may even share the relevant gesture in fast speech conditions. For instance, in slow speech productions of 'Kiss Ted', the study found evidence for two glottal openings in the /s/ coda and /t/ onset; however, at fast speech rates, a single glottal opening was shared between the two segments. Thus, shared features seem to result in differing coordination patterns. Whether these differences result in differences in coordination or ultimately syllabification is an open question as the scope of studies on syllable coordination have not been extended to include these issues.

While models of gestural coordination within the tradition of Articulatory Phonology do not analyze coupling of the jaw, studies analyzing global and local timing have likewise found distinct patterns for different cluster compositions. For example, in English, while c-centering and rightward shift are found for both /s/+stop clusters and stop+liquid clusters, /s/+stop clusters are more variable than stop+liquid clusters (Browman and Goldstein, 1988; Goldstein et al., 2009). These results are shared by studies of compensatory shortening in English, where there are more overt patterns of compensatory shortening in stop+liquid clusters compared to /s/+stop clusters (Katz, 2012).

The Coupled Oscillator Model accounts for variation in coordination between cluster types for a given language through coupling strength. For instance, in English, differences in timing for complex onsets, e.g., between stop+liquid and sibilant+stop clusters, are attributed to coupling strength. In the /pl/ cluster, /l/ requires both a tongue tip and tongue body gesture, resulting in a tighter coordination between the /l/ and V compared to the /sp/ cluster (Goldstein et al., 2009). However, fine-grained distinctions in the timing of different cluster types remain an area for further development within the Coupled Oscillator Model. For instance, as is exemplified by Marin (2013), coupling multiple gestures is insufficient for capturing the full range of timing patterns even within English. Marin (2013) demonstrates that for the onset cluster /sm/, where /m/ requires two gestures, the timing more closely resembles other /sC/ clusters than stop+liquid clusters.

Although the Coupled Oscillator Model does find differences between segments depending on composition, it does not specifically predict that there would be a difference in coordination between a sequence like [sti] and [ski]. Instead, the model has focused in particular on sequences of consonants involving two articulators, e.g., /pl/ or /sp/, leaving it unclear whether homorganic clusters would pattern with heterorganic clusters and how different clusters might pattern with differing vowel quality in the nucleus. Furthermore, the Coupled Oscillator Model inherently assumes onset maximization in syllabification, but the scope of stimuli tested are too limited to conclude whether this prediction is borne out in sequences where syllabification may vary. Put in another way, studies and syllable gestural coordination have been restricted to word-initial onsets and to stressed syllables, leaving the question of variation in coordination and syllabification open for further investigation. Thus, the question of segment identity affects coordination is ripe for development in understanding both how segment quality affects syllabification and if and how syllabification varies across contexts.

1.5.2.2 Stress and syllabification

Despite claims in favor of onset maximization at least for English (Goldstein et al., 2009; Kahn, 1976; Nam et al., 2009), a number of studies analyzing the role of stress in coordination and syllabification show that stress has a significant impact on segmental articulation, syllable coordination, and even cue perception. For example, Byrd et al. (2009) demonstrates that stress attracts segments from neighboring segments showing that the gestures of intervocalic nasals are coordinated with the stressed syllable, e.g., ‘VN.V and V.‘NV, suggesting that stress may impact syllabification.

Furthermore, Fujimura et al. (1978) finds a difference in the perceptual stability of cues in VCV sequences between English and Japanese, where in English, when stress was on the initial syllable, speakers tended to use VC cues rather than CV cues, suggesting that stress plays a role in the speech perception of syllables in English, though for Japanese speakers, CV cues were more stable regardless of stress. Furthermore, while Fujimura et al. (1978) argues that this difference is due to a difference in stress and coda patterns between English and Japanese, the findings of Byrd et al. (2009), demonstrate that there are likewise articulatory differences between stress initial and stress final VCV sequences. Therefore, articulatory differences in the production of VCV sequences may also affect the cues that English speakers are attuned to.

In addition, segmental articulation also differs between stressed and unstressed syllables. Studies such as Beckman and Edwards (1994) de Jong et al. (1993) and de Jong (1995) find that there. is less coarticulation in prominent syllables and that both the position and velocity of gestures in prominent syllables is more extreme than non-prominent syllables. These findings are similar to the differences between segments in onset and coda position, as discussed Section 1.3.2, where Krakow (1999) finds that segments differ in articulation depending on whether they occur as the syllable onset or coda, where there is more coarticulation between vowels and codas, and where segments in onsets are produced with tighter constrictions than the same segments occurring in the coda[5]. Studies looking at the effect of stress and prominence only considered segments in word-initial position; however, given the overlapping articulatory differences, there is an open question of how segments in word-medial position and how segments in unstressed syllable onsets will pattern regarding these findings.

1.6 Chapter summary and overview of the study

This chapter demonstrates the utility of syllable structure in accounting for a wide range of phonological patterns, supported by wide-spread typological evidence. In addition, there is support for syllables as a unit in speech production and perception, where syllable beats are salient to listeners and there are differences in gestural organization dependent on syllable structure. In summary, syllables provide a straight-forward structure in accounting for both phonetic and phonological patterns. Nevertheless, there are a number of open questions regarding word-medial syllables as they pattern differently than word-boundary syllable margins and likewise phonological processes pattern differently word-medially than at word boundaries. There are a number of factors that drive the differences between syllable margins, including the effects of stress and segment quality on perception and articulation. Despite the evidence that perception and production are affected by stress and that these factors are likely to affect syllabification, there are relatively few studies exploring to what extent pressures on word-medial syllabifica-

[5]The majority of studies in Krakow (1999) are word initial or word final. One exception includes **?**, which analyzes CVCVC sequencing with alternating stress and finds that onset stop release bursts have greater force of air pressure than codas, regardless of stress; however, this study assumes that the medial C is an onset and effect of word position is not considered.

tion affect coordination, and thus, these questions should be investigated to aid in developing a more fine-grained framework of syllabification.

In the following chapters of this dissertation, I present the results of this study, which responds to the open questions raised thus far and will analyze the effect of word position and stress on segmental articulation and syllable coordination across three task conditions: initial stressed syllables, medial stressed syllables, and medial unstressed syllables. Chapter 2 will outline the methods used in collecting and processing the Electromagnetic Articulography (EMA) data and will describe the task conditions in detail. Chapter 3 will demonstrate that segmental articulation differs depending on word position and stress environment. These results are consistent with patterns of continuity lenition, which is common across languages. Furthermore, these results support a theory of lenition that includes articulatory pressures, as a perceptual account does not predict that there would be articulatory differences where there is no phonemic alternation. Thus, word-medial differences in articulation contribute to differences in phonological patterns and likely to differences in the distribution of segments.

Chapter 4 will demonstrate that while existing metrics of coordination can provide some insight into the differences between word-medial coordination and word-initial coordination, namely, that the two margins differ, the existing methodology is insufficient for investigating what drives the differences between word-medial and word-initial coordination. In particular, while there are observable differences between medial stressed and medial unstressed syllables, the existing framework does not provide the tools to determine what underlies the differences in the two conditions and whether tokens in the condition are consistently coordinated despite differing from the word-initial condition, or whether there is variability in how utterances in the condition are coordinated.

Thus, Chapter 5 proposes a novel analysis of jaw oscillation to investigate patterns of coordination, identifying significant landmarks in jaw movement, and laying a foundation to develop additional methodologies. The chapter illustrates how jaw oscillation differs between productions of the same word, demonstrating how coordination structures may differ from token to token. By analyzing jaw organization, tokens can thus be classified as to how they are coordinated and patterns within task conditions can be identified. Ultimately, the ability to classify tokens into coordination groups provides a tool to distinguish between variable coordination across tasks, where within-task patterns are relatively consistent, and variability within a task, where coordination is overall more variable from utterance to utterance. This analysis has serious implications for phonological models of syllables as this dissertation proposes that the Frame-Content Model is a component of the phonological grammar in much the same way that Steriade (1997) proposed that confusability should be modeled in the phonological grammar. Moreover, modeling the jaw movement within the grammar provides the infrastructure to model competing pressures like perception and production, giving rise to variable patterns of coordination and syllabification.

Chapter 6 returns to the discussion of syllable structure and syllabification. Syllables are a useful tool, not only because they are able to capture phonetic and phonological patterns, but because they are well positioned between the segment and the word, between subsegmental effects and suprasegmental effects, and between intrinsic pressures and prosodic pressures. The power of this interface can be leveraged in a model of syllable structure that assumes both that there are properties of word-medial syllables that are consistent with word-boundary syllables and that there are also crucial distinctions between these margins. From an articulatory perspective, for instance, jaw modulation proceeds through opening and closing cycles in both word-initial and word-medial syllables, where jaw opening largely corresponds with syllable onsets and jaw closing largely corresponds with syllable codas. In addition, at both word-initial and word-medial margins, intrinsic factors constrain the magnitude of jaw movement. However, while the timing of the jaw cycle is relatively consistent across word-initial tokens, timing is much more variable word-medially, which in combination with intrinsic factors, produces a difference in segmental articulation. This study demonstrates the presence of articulatory pressures on word-medial processes like lenition and likewise how articulatory differences can constrain syllabification. As such, Chapter 6 advocates for a model of these processes that incorporates articulation, although perceptual and prosodic constraints may interact with articulatory constraints. Furthermore, the chapter proposes simple constraints for jaw modulation and coupling that can be used in conjunction with existing syllabification and perception constraints, i.e., licensing by cue, to capture patterns and improve typological predictions of syllabification.

Chapter 2

Methods

Syllable structure affects both the articulation of segments within the syllable and the coordination of segments across the syllable. For example, as was established in Chapter 1, stops in onset position are produced with tighter constrictions than stops in codas (Krakow, 1999). In addition, segments within a syllable show evidence of being coordinated as a unit, with both acoustic and articulatory evidence of a reorganization of timing between singleton onsets and complex onsets. Nevertheless, studies on the role of syllables in articulation have focused predominantly on contexts were syllabification is straightforward. However, both differences in segmental distribution and phonological processes like lenition demonstrate that word boundary and word internal syllable margins do not pattern uniformly (Blevins, 1995; Côté, 2011; Hayes, 1982; Katz, 2016; Kingston et al., 2008; Kirchner, 2001). In particular, there is evidence that both intrinsic pressure, like segmental quality, and prosodic pressure, like stress, also affect articulation. For instance, Byrd et al. (2009) demonstrates that segment gestures are attracted to the stressed syllable in productions of VNV alternating stress. Likewise, studies like (Beckman and Edwards, 1994) and de Jong (1995) demonstrate that stress also affects segment articulation, where segments in stressed syllables are produced at faster speeds, are less reduced, and have less coarticulation than segments in unstressed syllables. Although stress and intrinsic differences between segments have been shown to affect articulation and coordination, the interaction of word position with these effects is untested.

To address this gap, this dissertation presents two studies analyzing segments in both initial and medial positions and in stressed and unstressed syllables. The first study analyzes productions of sentences from the MOCHA-TIMIT corpus and focuses in particular on /st/ sequences as /st/ sequences occur very frequently throughout the corpus and occur both in initial and medial position as well as in stressed and unstressed syllables, making the application of statistical analyses possible. The second study analyzes nonce words in a carrier phrase, where the target word controls for word position and and stress and expands the scope of sequences. The second study consisted of three task conditions to alternate stress and word position, where targets in the initial stressed condition consisted of CVd and CCVd tokens and targets in the medial conditions consisted of VCV and VCCV tokens with stress on the initial syllable or stress on the final syllable, referred to and medial unstressed and medial stressed, respectively. In addition, both vowel and consonant cluster varied in study two in order to analyze intrinsic pressures on coordination and syllabification. NDI Wave Electromagnetic Articulography (EMA) was used to collect data for both studies. In addition to differences in stimuli, there were minor differences in data collection and processing between the two studies, namely, sensor placement, head rotation and correction, and landmark calculations between the two data sets. However, the resulting data sets are highly similar and comparable. This section will detail the methodology for data collection, processing, and analysis for both studies, providing comparisons between methodologies where relevant.

This dissertation will present results for three levels of analysis. Chapter 3 will analyze the role of stress and word position on segmental articulation to determine whether consonants and vowels are produced differently depending on whether they are initial or medial and depending on whether they are in stressed or unstressed syllables. Segments in stressed syllables are predicted to have gestures at higher speeds and more extreme positions than segments in unstressed syllables, as is established by Beckman and Edwards (1994). However, there are no known studies analyzing specifically the effect of word position on articulation. Still, given the difference between word-medial and word-boundary lenition, consonants are expected to be more reduced word-medially compared to segments in word-initial syllable onsets, whereas vowels are not expected to differ significantly based on whether the syllable is initial or final. This chapter will describe the variables considered for analyzing differences in segmental articulation and how those variables were calculated.

Chapter 4 will analyze the role of word position and stress in syllable coordination. While models of gestural coordination like the Coupled Oscillator Model predict that syllable coordination should be the same regardless of whether the syllable is word-initial or word-medial, studies demonstrating the role of stress on syllabification suggest that stress will affect coordination. Furthermore, theories of syllabification like licensing by cue (Steriade, 1997) and the Frame-Content Theory (MacNeilage, 1998) leave open the possibility for varying syllabification in word-medial contexts. Therefore, it is predicted that word-medial syllables are more likely to vary in coordination than initial syllables, and in particular, word-medial unstressed syllables are predicted to be more variable than medial stressed syllables. Studies analyzing syllable coordination have used two main measures of coordinative stability: c-center and shift. This chapter will discuss a) these variables in detail, b) the articulatory landmarks used to calculate these measures c) how articulatory landmarks were calculated for this study; and d) how c-center and shift were calculated for this study.

Chapter 5 will analyze traces of jaw velocity to analyze the role of jaw coordination in the gestural coordination of syllables, focusing in particular on how jaw coordination can inform variability in syllabification. The Frame-Content Theory (MacNeilage, 1998) predicts that jaw modulation is foundational to syllable coordination, and thus, is predicted to differ depending on how the utterance is syllabified. While the methods described for processing the data and calculating variables like velocity are consistent across all the analyses discussed in this dissertation, this study is the first to use jaw movement to analyze variability in jaw coordination; therefore, Chapter 5 will present the methodology employed in detail along with the results and implications of the analysis.

2.1 Stimuli and participants

For both study one and two, stimuli were displayed on a monitor outside of the NDI Wave magnetic field at a comfortable distance from the participant. Stimuli were presented using an OpenSesame (Mathôt et al., 2012) script that progressed manually. Participants were instructed to read the stimuli at a natural and comfortable pace. Each word or sentence was previewed before the participant produced the sentence to increase fluency. In the case of speech errors, the stimulus was immediately repeated and the participant was instructed to produce the stimulus again.

2.1.1 Study one stimuli

Stimuli for study one consisted of sentences from the MOCHA-TIMIT corpus, which consists of 460 phonetically balanced sentences (Wrench, 1999). Data was collected from eight native English speakers (5f, 3m). A subset of the collected MOCHA-TIMIT dataset was analyzed in study one, focusing on /tV/, /sV/, and /stV/ sequences, as /st/ sequences occur in both initial and medial syllables and in unstressed syllables frequently throughout the corpus. Sequences of /str/ were excluded from the analysis along with word-final sequences of /st/; however, in addition to word-initial sequences of /st/, sequences of both /VstV/ and /CstV/, e.g., /nstV/, were included to analyze the syllabification of both sequences and the resulting timing and coordination of these sequences. The resulting dataset consisted of 188 total words containing /t/ sequences, either /#tV/ or /VtV/, 127 containing /s/ sequences, either /#sV/ or /VsV/, and a total of 64 words with /st/, including 41 words containing /st/ in the onset of a stressed vowel and 23 words with /st/ in the onset of an unstressed vowel. Additionally, of the 63 total /st/ onsets, 35 were initial and 28 were medial.

2.1.2 Study two stimuli

The second study targeted singleton and cluster sequences controlling for stress, word position, and vowel quality. The tasks were divided into two sequence types: CVd/CCVd sequences, e.g., *sed*, *sted* and VCV/VCCV sequences, e.g., *oso,osto*. The latter intervocalic sequences were also divided into two separate tasks where sequences were produced with initial stress in the first task block, referred to as medial unstressed, and with final stress in the second task block, e.g., *ò.sto* and *o.stò*, referred to as medial stressed. All tokens were produced in the carrier phrase 'say the [x] again.' All speakers produced all target stimuli. There were seventeen total participants of study two (5m, 12f).

The singleton and cluster sequences analyzed in study two included stop+liquid clusters, /pl/, /pr/, /tl/, /tr/, /kl/, /kr/ and sibilant+stop clusters /sp/, /st/, /sk/. All consonants that occurred in clusters also occurred as singletons along with the segments /ʃ, tʃ/. In the CVd/CCVd condition, all English vowels occurred in the medial position and each token was repeated three times throughout the course of the task for a total of 318 utterances. For both the medial stressed and unstressed conditions of the VCV/VCCV stimuli, the consonant sequences occurred between three vowel qualities: /o, i, a ~ə/, e.g., *eeplee*. For the /a ~ə/ alternation, schwa occurred in the unstressed position and /a/ occurred in the stressed position, e.g., medial unstressed, /akɹə/, medial stressed /əkɹa/. Spelling was kept consistent for all vowels across all tokens and participants were instructed on vowel pronunciation and stress patterns as necessary providing examples distinct from the target stimuli, *ono*, where /n/ was not a target segment. Singleton sequences were repeated three times per token and cluster sequences were repeated five times per token for a total of 206 utterances for each of the stress conditions. The stimuli conditions and repetitions are broken down in 2.1 and examples are provided in 2.2.

Stimuli Set	**Initial Stress**	**Medial Unstressed**	**Medial Stressed**
Word Shape	CVd, CCVd	VCV, VCCV	VCV, VCCV
Carrier Phrase	Say the ___ again	Say the ___ again	Say the ___ again
Vowels	/i, ɪ, ɛ, æ, ʌ, u, ʊ, oʊ, ɑ, əɹ/	/oʊ, oʊ; i,i; ɑ,ə/	/oʊ, oʊ; i,i; ə,ɑ/
Clusters	sp, st, sk, pl, pr, tr, tl, kl, kr	sp, st, sk, pl, pr, tr, tl, kl, kr	sp, st, sk, pl, pr, tr, tl, kl, kr
Cluster Repetitions	3	5	5
Singleton Repetitions	3	3	3
Total Utterances	318	206	206

Table 2.1: Breakdown of stimuli in study two

	Examples	
Initial Stressed CVd CCVd	kud	spad
	tood	preed
	pid	stod
	sed	tlad
	rud	skod
Medial Unstressed VCV, VCCV	opo	apra
	eetee	ospo
	aka	eestee
	oso	akra
	eelee	osko
Medial Stressed VCV, VCCV	eepee	oplo
	ata	eetlee
	oko	asta
	eesee	oklo
	ala	eeskee

Table 2.2: Example stimuli for the initial stressed CVd, CCVd; medial unstressed VCV, VCCV; and medial stressed VCV, VCCV conditions

2.1.3 Data collection and processing

For both study one and study two, acoustic data was recorded using an AKG-C417 lavaliere microphone attached to a lanyard worn around each participant's neck. Likewise, in both study one and study two articulatory data was collected using a NDI Wave electromagnetic articulography (EMA) (Perkell et al., 1992; Schönle et al., 1987) system sampling at 200 Hz. The NDI wave system is a point tracking system with accuracy within approximately 0.5 mm (Berry, 2011). In both studies, NDI Wave 5DoF sensors were attached to the upper and lower lips near the vermilion border, lower jaw to the gums below the lower incisors (JW), and three sensors were attached along the sagittal midline of the tongue at the tongue tip (TT), body (TB), and dorsum (TD). The front most sensor (TT) was attached less than one cm from the tip of the tongue, the back most sensor (TD) was attached as far back as was comfortable for the participant, typically 4cm to 6cm, and the third sensor (TB) was placed mid way between the TT and TD sensors.[1]

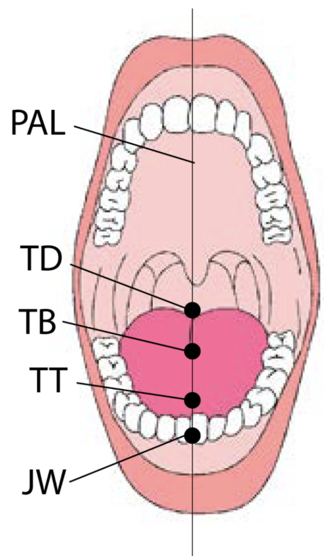


Figure 2.1: Tongue sensor placement

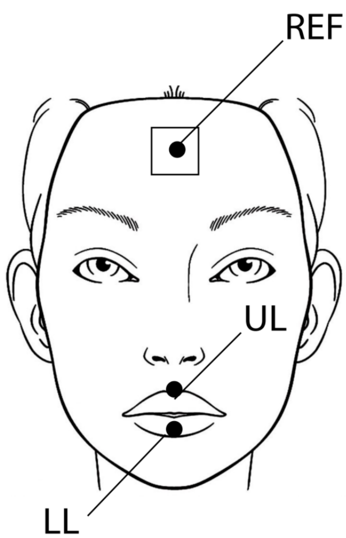
One major difference between study one and study two was the reference sensor placement, where study one used a 6DoF reference method and study two used a 5DoF method. These methods and the difference between these two methods are discussed in sections 2.1.3.1 and 2.1.3.2.

2.1.3.1 Study one reference sensor attachment and palate trace

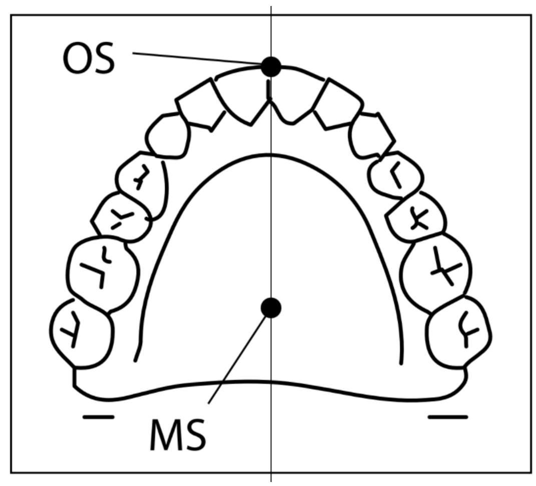
A 6DoF method was used for data collection in study one. This method entails three reference sensors consisting of two NDI Wave 5DoF sensors and one NDI Wave 6DoF sensor. The 6DoF sensor (REF) was attached to the bridge of the nose for the duration of the experiment. The two 5DoF sensors recorded the occlusal plane of each participant and were attached along the sagittal midline to a wax bite plate with one sensor aligned with the front incisors (OS) and the other

[1]Figure courtesy of Cassandra Serrano, University of California, Berkeley

aligned with the back molar (MS). Prior to the presentation of the stimuli, participants held the bite plate between their teeth[2]



(a) Reference and lip sensor placement



(b) Biteplate sensor placement

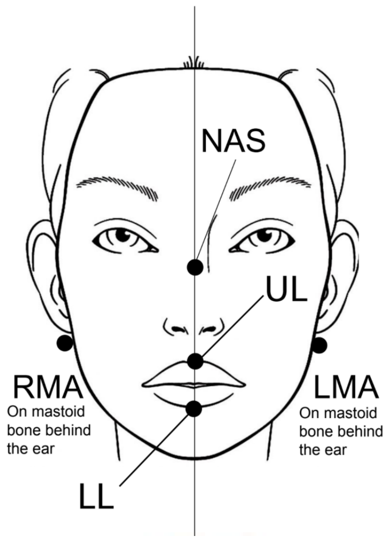
Figure 2.2: 6DoF reference sensor and lip sensor placement

These three sensors provide three stable points for rotation of the tongue, lip, and jaw sensors along the midsagittal plane. The 6DoF sensor method uses built-in software in the NDI wave system to correct for head movement. Additionally, the palate was traced using the NDI Wave Palate Probe to trace the midsagittal plane, the inner border of the upper teeth, and a back and forth trace from the back molar to the front of the palate to get a full 3D model of the palate. One of each per trace type was collected for each subject.

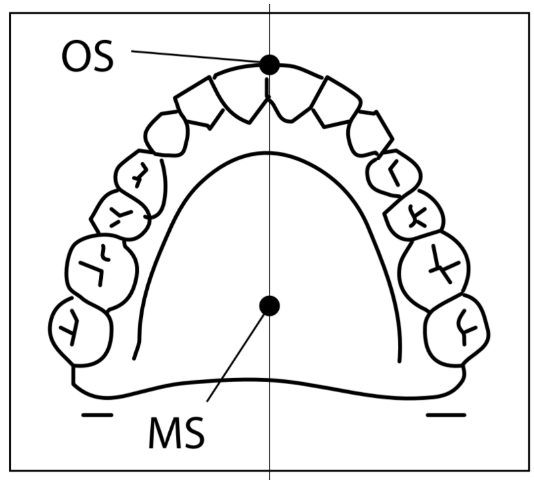
2.1.3.2 Study two reference sensor attachment and palate trace

In study two, a 5DoF method was used. For this method, reference sensors consisted of five NDI Wave 5DoF sensors (Ji et al., 2014; Wieling et al., 2016). One 5DoF sensor (REF) was attached to the bridge of the nose and two 5DoF sensors were attached behind the right (RMA) and left (LMA) ears, ensuring the orientation of the left sensor was offset from the right. These three reference sensors were attached for the duration of the experiment. Two additional 5DoF sensors recorded the occlusal plane of each participant and were attached along the sagittal midline to a wax bite plate with one sensor aligned with the front incisors (OS) and the other aligned with the back molar (MS). Prior to the presentation of the stimuli, participants held the bite plate between their teeth[3].

[2]Figures courtesy of Cassandra Serrano, University of California, Berkeley
[3]Figures courtesy of Cassandra Serrano, University of California, Berkeley



(a) Reference and lip sensor placement



(b) Biteplate sensor placement

Figure 2.3: 5DoF reference sensor and lip sensor placement

The 5DoF method provides five stable points for head rotation and correction and does not use the built-in NDI software for head movement correction. A comparison between rotation accuracy for each method is provided in Section 2.1.3.4. In addition to the reference sensor, the palate was traced using the NDI Wave Palate Probe to trace along the midsagittal plane, the inner border of the upper teeth, and a back and forth trace from the back molar to the front of the palate to get a full 3D model of the palate. For the trace of the midsagittal line, five traces were collected. For each of the 3D model of the inner border of the upper teeth and the 3D model of the palate, three traces were collected.

2.1.3.3 Study one head rotation and correction

Head movements in study one were corrected for in Python (Van Rossum and Drake, 2009) with the NDI Wave built in head correcting using a 6DoF reference sensor and the bite plate sensors. The NDI Wave 6DoF sensor tracks the x, y and z-axes; and the three rotation values roll, pitch and yaw. The bite plate and 6DoF sensors were then used to computationally rotate the head-corrected data along the occlusal plane so that the origin of the spacial coordinates corresponds to the front teeth. All articulatory points were smoothed using Garcia's robust smoothing algorithm (Garcia, 2010).

Acoustic data was segmented using the Penn Forced Aligner (Jiahong and Liberman, 2008). A team of research assistants[4] then hand corrected alignments in Praat (Boersma and Weenink, 2020).

[4]Thank you to Phoebe Killick, Ray Mason, Cassandra Serrano, Emma Brown, Elissa Chau, Emily Li, Thomas Lu, Anahita Farshi Haghro, Emma Brown, Elissa Chau , Emily Li, Thomas Lu, Jacky Chen, Joanne Chuang, and Cooper Bedin for assistance in data processing and segmentation.

2.1.3.4 Study two head rotation and correction

In the second study, all sensors were smoothed using Garcia's robust smoothing algorithm (Garcia, 2010). Then, head movement was corrected for in Python (Van Rossum and Drake, 2009) using a three-point 5DoF sensor method, where roll, pitch, and yaw are captured through position of the three reference sensors (nasion (REF), left (LMA), and right mastoid (RMA)) (Hoole and Zierdt, 2012; Shaw and Kawahara, 2018; Tilsen et al., 2015). The bite plate and reference sensors were then used to computationally rotate the head corrected data along the occlusal plane so that the origin of all spacial coordinates corresponds to the front teeth. Comparisons of the 6DoF rotation method and 5DoF method find improved accuracy in the 5DoF method over the 6DoF method, making the 5DoF method preferable (Johnson and Sprouse, 2019).

Acoustic data was segmented using the Montreal Forced Aligner (McAuliffe et al., 2017) and spot checked by hand for accuracy. Segmentation was found to be highly accurate for the relevant acoustic landmarks.

2.2 Measures of analysis for segmental gestures and syllable coordination

This section presents the methods used to calculate the variables used in analyzing differences in segmental gestures across the task conditions and coordinative stability across the task conditions. In order to calculate the variables to compare differences between both segmental gestures and syllable coordination, it is necessary to identify significant landmarks within the gesture. For instance, a given gesture includes the approach to the target, the constriction at the target, and the release from the target. The speed of movement for a given articulator during each of these portions will vary greatly. During the constriction the speed will be very low, at or near zero; however, during the approach and the release from the target, the articulator must move quickly to reach the upcoming target, and thus, the speed of the articulator during this time will be very high. To analyze speed meaningfully, then, the variable must be defined based on where gestures differ informatively from one another. Section 2.2.1 will first establish the articulatory landmarks necessary for calculating these variables and how the landmarks are calculated for this study. Sections 2.2.3 and 2.2.4 will describe how variables like speed were determined and calculated using these landmarks for the segmental gesture analysis and the syllable coordination analysis, respectively. In addition, to analyze the role of stress and word position, each token needed to be coded for a number of linguistic variables, Section 2.2.2 defines these variables and how tokens were coded.

For the comparison of segmental gestures, studies on stress, syllable structure, and lenition predict a number of variables that are likely to differ across the task conditions. For instance, stops in the syllable onset have tighter constrictions than stops in the syllable coda (Krakow, 1999); similarly, differences in word-boundary and word-medial positions predict that consonants are likely to have looser constrictions in unstressed syllables compared to stressed syllables. Furthermore, segments in stressed syllables have longer durations, less reduction, and higher velocities than stressed syllables (Beckman and Edwards, 1994). Therefore, this study analyzes the speed of the gesture for the active articulator, position of constriction for the active articulator, and the duration of constriction for the active articulator. Section 2.2.3 details how these variables were calculated for each study. Study one focuses on differences in the gestures of consonants, but study two analyses both consonants and vowels as vowels have been found to differ more significantly than consonants between stressed and unstressed syllables (Beckman and Edwards, 1994; Gay, 1978).

Studies on syllable coordination in English have demonstrated coordination between gestures within the syllable, where both the duration of vowels in the syllable and the timing between gestures is dependent on the composition of the syllable (Browman and Goldstein, 1988; Byrd, 1995; Katz, 2012; Munhall et al., 1992). Thus, in order to analyze the syllable coordination, three measures were calculated: vowel duration, c-center stability, and rightward shift. Section 2.2.4 defines these measures and how they were calculated in this study.

2.2.1 Articulatory landmarks

For both study one and study two, the following articulatory landmarks were assigned to each segment of each utterance: onset, achievement, steady state or plateau, and release.

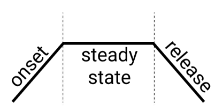


Figure 2.4: Articulatory landmarks

All articulatory landmarks across both studies were defined using the absolute value of the velocity profile, or in other words, speed, in the x and y-dimensions simultaneously for each of the articulatory sensors: TT, TB, TD, JW, UL, LL. The relevant sensor in landmark assignment depends on the active articulators for the relevant consonants. Table 2.3 illustrates which sensors were used in calculating articulatory landmarks. For /s/, speakers vary in production between the tongue tip and the tongue body. There was no significant difference found between articulators for participants in study one, and thus, the tongue tip was used in analyzing /s/ gestures. However, in the second study, speakers differed in the stability of /s/ production between the tongue tip and the tongue body. Thus, for eight speakers, the tongue tip was used in analyzing /s/ gestures, and for nine speakers, the tongue body was used.

The window of landmark assignment for each segment differed between study one and study two, largely due to the nature of the stimuli. Study one was able to consider a smaller window by relying more heavily on acoustic cues with a small buffer prior to the acoustic onset of about 0.5ms. Because the sequences in question were /st/ sequences, there was relatively little movement between the /s/ and /t/ segments and thus speed was a less reliable measure for parsing the two segments. Thus, articulatory landmarks provided the point of alveolar constriction while

Sensor	LL	JW	TT	TB	TD
Segment	p	h	t, d, s r, l	s tʃ, dʒ	k

Table 2.3: Sensor used to calculated articulatory landmarks for each consonant. For 8 speakers, /s/ was calculated using the TT and for 9 speakers /s/ was calculated using the tongue body.

acoustic landmarks segmented the respective segments. However, in study two, a wider window was used to calculate the landmarks relying instead on the speed profile, which was more reliable for segmentation given a more controlled dataset analyzing a wider range of segments and sequences. Clusters using the same active articulator were reliably disambiguated using the local speed minimum for each segment with reference to the acoustic landmarks where the local minimum occurring closest to the onset of the acoustic landmark provided a clear parse for the majority of segments regardless of whether the segments were homorganic or heteroganic. Additionally, a test of study one was analyzed using the study two windowing method to analyze differences between the two methodologies. The test of data from the first study showed that while the second study methodology was more likely to parse homorganic clusters than the first study, the difference in timing between the two methodologies was negligible.

In both study one and study two, the speed profile was used to assign articulatory landmarks. First, the x and y dimensions were analyzed simultaneously by calculating the sum of the two vectors. Next, the threshold for all sensors was calculated for each speaker using the full range of movement for each sensor. This provides a methodology that works across sequences, increasing the range of possible analyses to a wider range of vowel and consonant sequences, e.g., homorganic sequences.

Following threshold calculation, the steady state was labeled as the continuous sequence of data points where the speed fell below the given threshold, 20%. Finally, the first sample within the threshold was labeled as the onset, the local minimum was labeled as the point of achievement, and the final sample within the steady state was labeled as the release. This windowing and velocity profile methodology provided a clean parse in defining achievement for the majority of phones. Tokens that were not cleanly parsed were discarded.

2.2.2 Linguistic variables

To analyze the effect of word position, syllable position, environment, and stress on the gesture of each segment, the stimuli were coded with a number of linguistic features, including stress, word position, proceeding environment, and following environment. In the first study, stress was assigned based on stress transcription from the Carnegie Mellon University Pronouncing Dictionary (CMU) (CMU, 2014). The CMU uses ARPAbet transcription, which marks vowels with '0' for unstressed, '1' for stressed, and '2' for secondary stress. For example, a word like *potato* is transcribed in the CMU as 'P AH0 T EY1 T OW2', [pəˈteɪtoʊ], where 'AH0' is unstressed, 'EY1' is stressed, and 'OW2' has secondary stress. For this study, tokens with secondary stress were not analyzed. Of the total 64 words, there were 41 stressed syllable tokens , e.g., '***stem***', and 23 unstressed syllable tokens e.g., '*sy**stem***'. In the second study, stress was assigned based on the task, where all segments in the CVd/CCVd condition were stressed; in the medial unstressed condition, VCV/VCCV, the initial vowel was labeled as stressed and the remaining segments were labeled as unstressed; and in the medial stressed condition, the initial vowel was marked as unstressed and the remaining segments were labeled as stressed. In the medial stressed and unstressed condition, the token number was balanced with 206 tokens per task; for the CVd/CCVd condition, there were 318 tokens for a total of 524 stressed tokens and 206 unstressed tokens.

Word position was labeled in study one using a combination of segment and word acoustic landmarks, where if the singleton or cluster had the same start time, T1, as the word, then it was word-initial, otherwise it was labeled as word-medial. Of the total 64 tokens, 35 of them were initial and 28 were medial. In study two, the word position was again labeled using the conditions of the task where the coda /d/ was labeled final and the singleton or consonant clusters in the CVd/CCVd condition were labeled initial. In the medial stressed and unstressed conditions, the medial singleton or cluster sequence was always marked as medial.

Finally, in both studies, the proceeding and following segments were included for each phone. Where the previous or following phone was a vowel, the stress marking from the ARPAbet transcription was removed to disambiguate the effect of environment from the effect of stress, e.g., AH0, AH1, AH2 $\rightarrow$ AH. Together, these variables were modeled to analyze the effect of word position and stress on gestural composition.

2.2.3 Segmental gesture analysis

Four variables were identified to analyze differences in segmental gestures between initial and medial syllables and stressed and unstressed syllables: speed, x-position, y-position, and duration. All four variables were analyzed for both study one and study two, but the studies differed slightly in how the variables were calculated. Study one focused on /st/ sequences and did not analyze vowel gestures as vowels were too variable across tokens to conduct a reliable statistical analysis. However, both vowels and consonants were analyzed in study two, where tokens were balanced for vowel and consonant sequence composition.

In study one, the variables analyzed were: maximum speed of the sensor during the gesture, minimum speed of the sensor during the gesture, x-position averaged across the steady state, y-position averaged across the steady state, and the duration of the steady state. As study one only analyzed /st/ sequence and focused on consonants, the tongue tip was the only sensor analyzed for each of these variables, as was established in Section 2.2.1. The significance of these variables was determined using a linear mixed effects regression analysis, which is described along with the results of the analysis in Chapter 3.

Like study one, study two analyzed maximum speed of the sensor during the gesture, sensor position in the x-dimension at the point of achievement (where the x-dimension indicates degree of reduction), sensor position in the y-dimension at the point of achievement (where the y-dimension indicates degree of constriction), and duration. However, study two did not analyze minimum speed and differed in the calculation of the x and y positions. Minimum speed was not analyzed in study two as the steady state is expected to be at or near zero regardless or stress of word position, and thus, it was determined to be unnecessary Correspondingly, minimum speed was not found to be significant in study one. The calculations of the x and y dimensions were also changed in study two as study one did not find these variables to be significant. One hypothesis for why position was not found to be significant in study one is that by taking the average position of the steady state, the differences between tokens were minimized. Study two instead used x and y positions at the point of achievement for each segment. The relevant sensor in calculating these measures is established in Table 2.3 in Section 2.2.1. The significance of these variables was determined using a linear mixed effects regression analysis, which is described along with the results of the analysis in Chapter 3.

2.2.4 Syllable coordination analysis

To analyze syllable coordination, duration, c-center stability, and shift were calculated and analyzed in study two, but only c-center stability and shift were calculated and analyzed in the first study. The c-center is defined as the midpoint of the onset regardless of whether the onset is a singleton or a cluster. In English, the c-center has been found to be more stable than other landmarks in the onset, i.e., the left edge or right edge. The left edge is defined as the onset of the steady state for the first consonant in the onset, e.g., /s/ in ‘stop’ or /t/ in ‘top’. The right edge is defined as the release of the steady state of the prevocalic consonant, i.e., /t/ in both ‘stop’ and ‘top’. The c-center was determined using the mean of the means of the corresponding onset cluster steady states, e.g., /s/ and /t/ in study one. These calculations are made using the articulatory landmarks described in the previous section. In all cases, the anchor for the vowel was the minimum velocity of the relevant sensor for the post-vocalic segment (Browman and Goldstein, 1988; Byrd, 1995). The standard deviation and the relative standard deviation were used in both studies to calculate the coordinative stability for each of the measures, left edge, c-center, and right edge, and the sensor with the least variance is understood to be the most stable (Shaw et al., 2011).

The calculations of the left edge, c-center, and right edge for the velocity profile for study one are demonstrated in Figure 2.5. The figure shows the speed of the tongue tip, the threshold for determining the steady state, and the acoustic boundaries for each phone. In addition, the figure shows the left edge, c-center, and right edge as they correspond to the velocity profile.

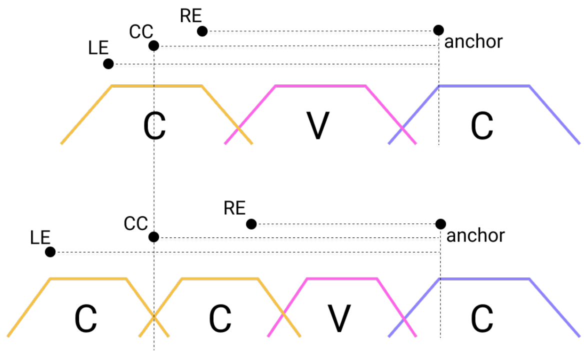


Figure 2.5: Left edge, c-center, and right edge measurements for measuring stability

In addition to the velocity profile, Figure 2.7 shows the trace of the palate and the trace of the TT and TB (x, y) movement for the same utterance shown in Figure 2.6, illustrating how the position of the tongue corresponds to the velocity of the tongue.

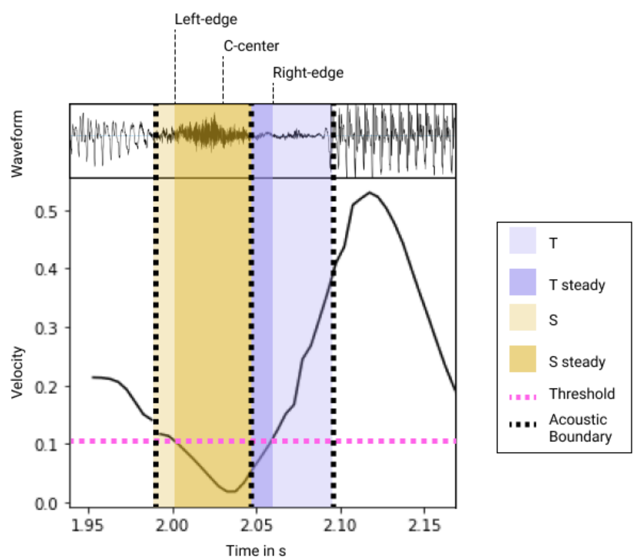


Figure 2.6: Wavevform and velocity profile of /V#stV/ sequence with target landmark for /s/ and release landmark for /t/. ‘Ralph controlled th<u>e sto</u>pwatch from the bleachers.’

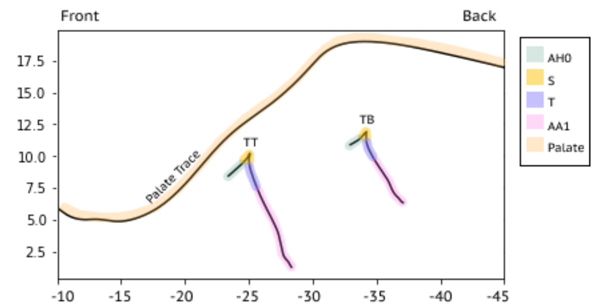
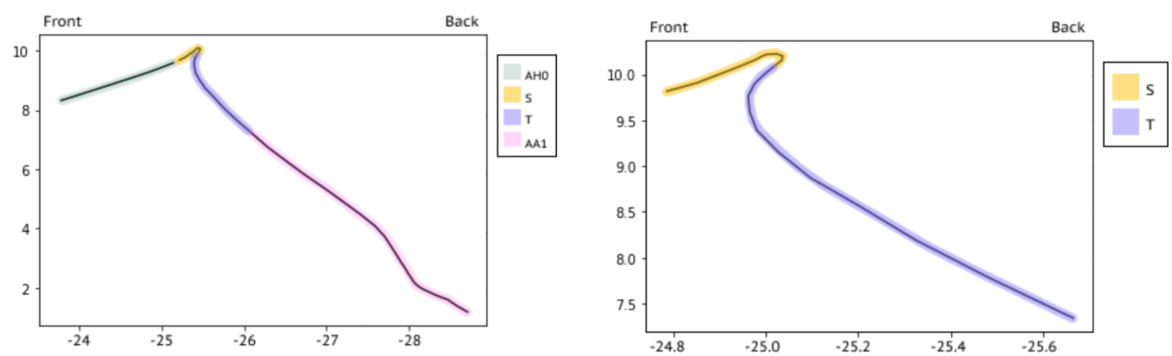


Figure 2.7: Trace of palate, tongue tip (TT) and tongue body (TB) for /V#stV/ sequence with target landmark for /s/ and release landmark for /t/. ‘Ralph controlled th**e sto**pwatch from the bleachers.’

Finally, Figure 3.12 shows the tongue tip movement in the production of the VstV and st sequence, respectively, in the sentence *Ralph controlled the stopwatch from the bleachers.* These figures demonstrate the trajectory of movement for the /st/ cluster along with the overall trajectory of movement for the utterance to demonstrate how the gesture itself corresponds with the speed of the gesture, which is used to calculate the articulatory landmarks.



(a) VSTV

(b) ST

Figure 2.8: (x, y) TT movement of (a) VSTV sequence and (b) zoomed in on the ST sequence for the sentence ‘Ralph controlled th**e sto**pwatch from the bleachers,’ the same token as in Figure 2.6

The measurements for shift used the same velocity profile landmarks and furthermore, the left boundary and vocalic anchor are the same in measuring shift as in measuring the stability heuristics; however, the right boundary differs between the c-center and stability heuristics. In measuring shift, the right boundary is the point of achievement of the prevocalic consonant.

In other words, the articulatory onset of /t/ for both *stop* and *top*. Figure 2.9 demonstrates the anchor points for the right and left boundaries in calculating shift. Leftward shift was calculated as the difference in lag time of the left edge for the complex onset from the singleton. Likewise, rightward shift was calculated as the difference in lag of the right edge for the complex onset from the singleton (Goldstein et al., 2009).

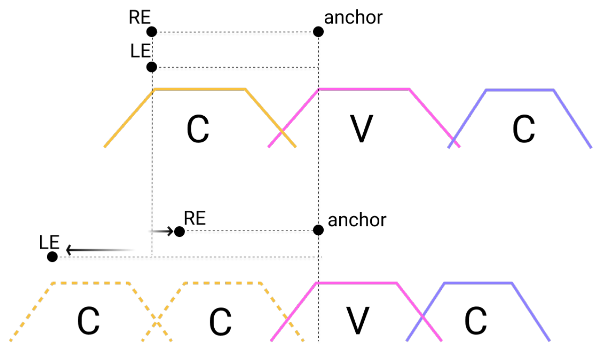


Figure 2.9: Left and right boundaries for measuring leftward and rightward shift

The final measure of syllabification is duration, where differences in the duration of the nuclei indicate differences in syllabification (Katz, 2012; Munhall et al., 1992). Specifically, in a $V^1C^1C^2V^2$ sequence, the V^1 becomes shorter if the C^1 syllabifies as a coda and the V^2 is shorter if the C^1 syllabifies as a complex onset. Thus, measuring differences in duration provides a diagnostic for patterns in syllabification. For each phone, the acoustic duration was calculated as the acoustic onset of the phone subtracted from the acoustic offset. The results of the initial stress condition include the duration of stressed vowels, comparing tokens with singleton onsets to tokens with complex onsets. The results of the medial conditions include the duration of both stressed and unstressed vowels, comparing tokens with singleton onsets to tokens with complex clusters. All three diagnostics for syllable coordination: c-center stability, shift, and duration, are analyzed and discussed in Section 4.

2.3 Chapter summary and conclusions

Both word position and stress are predicted to affect segmental articulation and gestural coordination. Two studies were conducted using EMA data to analyze articulation and coordination in initial stressed syllables and medial stressed and unstressed syllables. This chapter provided an overview of how the data was collected and processed for two studies, where study one used corpus data to analyze the effect of stress on the articulation of /st/ sequences and study two analyzed articulation and coordination across varying vowel and consonants sequences to analyze the role of both stress and word position on both consonants and vowels.

The variables analyzed in comparing segmental gestures for both studies include velocity, degree of constriction and reduction, and duration. Chapter 3 will detail the models used to analyze these variables and the results of the models, demonstrating that word position is significant in predicting differences in consonant articulation. While smaller differences were found between stress conditions for consonants in both study one and two, stress is also found to be a significant predictor of differences between consonant gestures. Furthermore, for study two, while word position is not found to distinguish between vowels, stress is significant in distinguishing between vowels. These results will be discussed in detail along with the implications of these results in Chapter 3.

Three methods are used to analyze gestural coordination: duration, c-center stability, and rightward shift. This section provided an overview of how these measures were calculated and Chapter 4 will present the results of these analyses in detail. Across all three measures, word position is found to differ significantly across conditions. However, the results for stress are ambiguous. These results and their implications will be discussed in detail in Chapter 4.

Chapter 3

The effect of word position and stress on segment articulation

Theories of the syllable often assume that syllables pattern identically at word boundaries and word internally. For instance, the canonical constraints for syllabification in OT models do not distinguish between word-medial and word-boundary margins (Smolensky and Prince, 1993). Likewise, the Coupled Oscillator Model predicts that onset maximization should apply to both word-medial and word-initial onsets due to the advantage afforded to onsets in motor planning (Nam et al., 2009). However, both typological patterns in segmental distribution (Blevins, 1993; Côté, 2011) and phonological processes like lenition (Katz, 2016; Kingston et al., 2008) provide clear evidence that these margins are not uniform.

Instead, word-medial syllabification is subject to additional pressures like segment quality and stress. For instance, Byrd et al. (2009) demonstrate that intervocalic nasal gestures are sensitive to stress, where the gestures for the nasal are attracted to the stressed syllable. Studies have likewise shown that segmental gestures differ in stressed and unstressed syllables, where segments in stressed syllables are produced with less coarticulation, at higher speeds, and in more extreme positions than segments in unstressed syllables (Beckman and Edwards, 1994). Nevertheless, studies have not analyzed the role of word position on segment articulation or the interaction between stress and word position.

Although cross-linguistic differences in word-boundary and word-medial lenition predict that these margins differ, studies like Katz (2016) and Kingston et al. (2008) argue that the driving factor in word-medial lenition is prosodic modulation, where the presence or absence of lenition signals meaningful information about the word boundary to listeners. In languages where word-medial lenition has been phonologized, this prediction is difficult to test as word-boundary and word-medial margins are necessarily different in articulation due to the phonemic alternation. However, in a language like English where only /t/ and /d/ are lenited word-medially, there are no predicted gestural differences between /p/ and /k/ which have no perceptible alternation due to lenition. Thus, English presents the optimal test case for analyzing whether articulatory pressures also drive word-medial lenition.

To investigate whether stress and word position affect articulation, this dissertation presents the results of two studies. The first study analyzes sequences of /st/ in stressed and unstressed syllables. The second study analyzes a wider range of segments, including both consonants and vowels, and analyses the effect of both word position and stress on articulation. The results of these studies are presented in Sections 3.1 and 3.2, respectively. For vowels, stress is predicted to influence articulation, consistent with Beckman and Edwards (1994), but word position is not. For consonants, both word position and stress are predicted to play a significant role in segment articulation, as is predicted by studies of stress such as Beckman and Edwards (1994), and lenition (e.g., Gurevich, 2011; Kingston et al., 2008; Katz, 2016).

3.1 Effects of stress on segment gestures: results of study one

Study one analyzed the effects of stress on the production of /st/ clusters, where stress is expected to affect the gestures of these segments. As established in Chapter 2, study one analyzed speed, x and y positions, and duration as dependent variables to demonstrate the role of stress on these aspects of articulation. A stepwise linear mixed effects regression (lmer) was used to analyze the effect of environment on gestural composition with a separate model for each of the articulatory variables. Specifically, the dependent variables analyzed were: maximum tongue tip speed, minimum tongue tip speed, tongue tip x-position averaged across the steady state, tongue tip y-position averaged across the steady state, and the duration of the steady state. Each of these variables were modeled for /s/ and /t/ separately for a total of ten models. Each model tested the significance of three predictors: preceding phone (X1), following phone (V1), and stress. Subject was included as a random effect.

Adjacent segment (X1 or V1) was found to be significant for all dependent variables, where the preceding phone (X1) was the best predictor of the /s/ gesture and the following phone (V1) was the best predictor of /t/ gestures. Stress was not found to be a significant predictor of minimum velocity or position for either phone, nor was stress predictive of /t/ duration or /s/ maximum speed. However, when controlling for the effect of adjacent segments, stress did significantly improve the model of the duration of /s/ and the maximum speed /t/, where the duration of /s/ is longer and the tongue tip speed of /t/ is faster in stressed syllables.

	npar	AIC	BIC	loglik	deviance	Chisq	Df	Pr(>Chisq)	
X1	15	-1946.6	-1883.6	988.30	-1976.6				
X1+stress	16	-1959.6	-1892.5	995.81	-1991.6	15.028	1	0.0001059	***

Table 3.1: ANOVA model comparison of duration of /s/ with the predictors X1 against X1+stress

	npar	AIC	BIC	loglik	deviance	Chisq	Df	Pr(>Chisq)	
V1	14	-160.52	-101.74	94.261	-188.52				
V1+stress	16	-167.38	-104.41	98.692	-197.38	8.8618	1	0.002912	**

Table 3.2: ANOVA model comparison of maximum speed of the tongue tip in /t/ with the predictors V1 against V1+stress

Summarizing the results of the first study, the most significant predictor in accounting for articulatory variance is the immediately adjacent segment, X1 for /s/ and V1 for /t/ in a sequence of X1stV1. Still, stress was also a significant predictor of variance for /s/ duration and /t/ maximum TT velocity. These findings suggest that stress is relevant in determining the gestural composition of a segment. However, these findings also leave several unanswered questions. For instance, while stress was found to be significant in these results, the first study did not analyze the role of word position due to the correlation between word position and stress. Furthermore, it is unclear whether the differences between /s/ and /t/, where stress was only predictive of duration for /s/ and maximum speed for /t/, are due to the differences between the manner of articulation or the environment of these segments. In particular, while the segment following /t/ was always a vowel, the segment preceding /s/ varied with /s/ occurring word-initially, following a consonant, e.g., /n/, and following a vowel. Thus, the degree of noise in the /s/ dataset may have washed out the effect of stress on velocity. Thus, further research is necessary to fully tease apart the effect of stress, word position, and adjacent segment.

3.2 Effects of word position and stress on segment gestures: results of study two

The stimuli in study two were designed to follow up on the remaining questions of study one, with particular emphasis on word position. In addition to word position, study two analyzed a broader range of segments, therefore allowing study two to test whether the findings of study one apply to other sequences of segments besides /st/. While consonants remain the focus of study two, analyzing both consonants and vowels offers the opportunity to test the prediction that while stress affects both consonants and vowels, word position effects are restricted to consonants. Furthermore, analyzing both consonants and vowels provides a clearer comparison to previous literature, as many studies only include or focus on differences in vowel gestures in stressed and unstressed syllables. Thus, this section outlines the models used in analyzing both consonants and vowels, and Sections 3.2.1 and 3.2.2 present the results for vowels and consonants respectively.

Like study one, a linear mixed effects model was used to analyze the role of task condition on each of the following articulatory variables: maximum speed, sensor position in the x-dimension at the point of achievement (degree of reduction), sensor position in the y-dimension at the point of achievement (degree of construction), and duration. As discussed in Chapter 2, although x and y positions were not significant in the first study, previous studies have found the degree of constriction (y-dimension) and the degree of reduction (x-dimension) to be significantly different between stressed and unstressed syllables, thus in the second study the manner of calculation was altered and they were included in study two. Together, a total of four dependent variables were analyzed: maximum speed, constriction (y-dimension), reduction (x-dimension), and duration.

For both vowels and consonants, the predictors used in the model included the preceding phone, the following phone, word position, and stress. In addition, each of the articulatory variables, with the exception of the relevant dependent variable, were included as fixed effects in the model. In other words, in the model analyzing maximum speed, constriction and reduction were included as predictors in the model as each of these factors is likely to play a role in the trajectory, and likewise, maximum speed. Finally, the random effects in the model were subject and phone. Like study one, preceding and following environment significantly improved model fit for all dependent variables. Additional details about what is included in the baseline model are provided on a variable-by-variable basis in the following sections. The variables word position and stress were then added in a stepwise fashion to test the effect of each of these variables on the relevant articulatory variable. An ANOVA was used to compare each of the models to the baseline model. Additional model specifications for each dependent variable are provided in the respective section. Section 3.2.1 provides the results of the analysis of vowels, discussing the results of the models analyzing duration, speed of gesture, x position of the gesture, and y position of the gesture. Section 3.2.2 likewise presents the results of the analysis for consonants.

3.2.1 Vowels

For vowels, studies like Beckman and Edwards (1994) predict that segments in stressed syllables are produced at higher magnitudes than segments in unstressed syllables, showing higher velocity, longer duration, and less reduced position. However, there is no clear basis for a difference between vowels depending on word position, and thus, word position is not predicted to affect the gestural composition of vowels. In analyzing the results of this study, these same variables were considered in a linear mixed effect model (lmer) for each of the following dependent variables: duration, maximum speed, x position, and y position. Subsequently, to control for the effect of syllable shape on these variables, the data was subset to the two VCV/VCCV task types, comparing both the stressed and unstressed syllables in both the medial stressed and unstressed conditions. The vowels included in this study were [o], [i], and [a]~[ə], where [a] occurred in stressed syllables and [ə] occurred in unstressed syllables.

3.2.1.1 Duration

For duration, the baseline model included phone label, preceding environment, following environment, maximum speed, and y-position. X-position was not included in the baseline model as it was not found to have a significant effect on duration. Word position and stress were added to the model in a stepwise fashion and the models were compared using an ANOVA model comparison.

	npar	AIC	BIC	logLik	deviance	Chisq	DF	Pr (>Chisq)	
baseline	24	-31710	-31535	15879	-31758				
+word position	25	-31890	-31708	15970	-31940	181.9	1	<2.2e-16	***
+stress	26	-32043	-31854	16048	-32095	155.51	1	<2.2e-16	***

Table 3.3: ANOVA model comparison for vowel duration showing baseline model for vowel duration compared to the model including word position and the word position model to the model including stress.

As shown in Table 3.3 both word position and stress were found to significantly improve model fit. For word position, stressed vowels in initial position were shorter than stressed vowels in final position. Unstressed vowels in initial position, however, were longer than unstressed vowels in final position. These results are shown in Figure 3.1 [1].

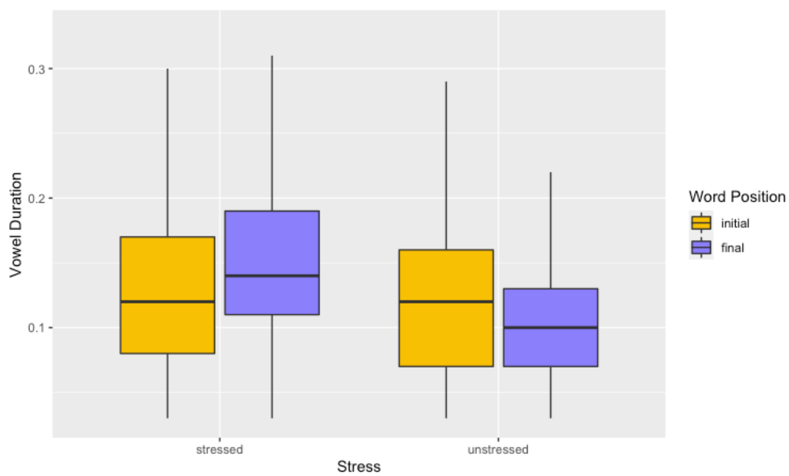


Figure 3.1: Vowel duration for stressed syllables in initial and final position and unstressed syllables in initial and final position

As previous studies have also found (e.g., Beckman and Edwards, 1994), the duration of vowels in unstressed syllables was significantly shorter than the duration of vowels in stressed syllables. This was true across vowel qualities and in both tasks, as shown in Figure 3.2.

[1]Syllabification is a major factor in the duration results and the role of syllabification in these results will be returned to in Chapter 4.

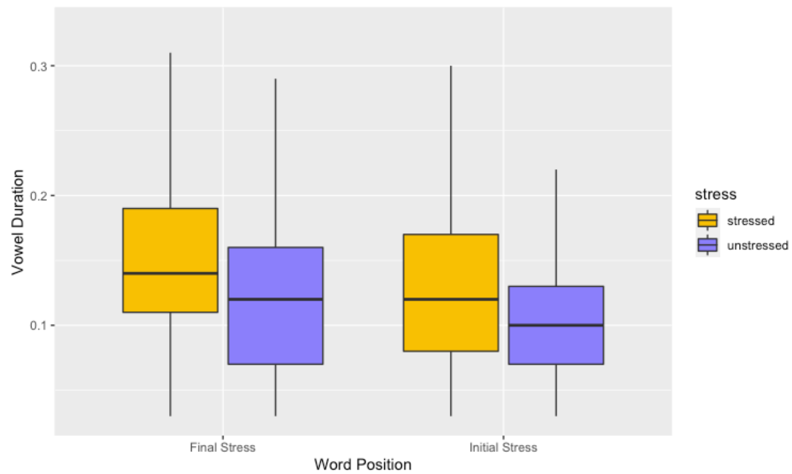


Figure 3.2: Vowel duration for stressed syllables and unstressed syllables in the medial stressed (final stress/VCV́) and unstressed tasks (initial stress/V́CV)

3.2.1.2 Maximum speed

Maximum speed is defined as the point of maximum speed of a given articulator during the trajectory of a given phone, where the speed of the gesture in stressed syllables is predicted to be higher than gestures in unstressed syllables. For maximum speed, the baseline model included phone label, preceding environment, following environment, duration, x-position, and y-position. Word position and stress were added to the model in a stepwise fashion and the models were compared using an ANOVA model comparison.

	npar	AIC	BIC	logLik	deviance	Chisq	DF	Pr (>Chisq)	
baseline	25	-8674.5	-8492.9	4362.2	-8724.5				
+word position	26	-8674.2	-8485.4	4363.1	-8726.2	1.6846	1	0.1943	
+stress	27	-8826.3	-8630.2	4440.2	-8880.3	154.13	1	<2.2e-16	***

Table 3.4: ANOVA model comparison for maximum speed showing the baseline model for speed compared to the model including word position and the word position model to the model including stress.

As shown in Table 3.4, while word position had no significant effect on maximum speed, stress did have a significant effect, with higher articulator speeds in stressed syllables than in stressed syllables across vowel qualities and tasks. Like duration, these results are consistent with previous studies, such as Beckman and Edwards (1994). The results are shown in Figure 3.3, which demonstrates that while there is very little difference between the two tasks, there is a clear difference in speed between stressed and unstressed syllables.

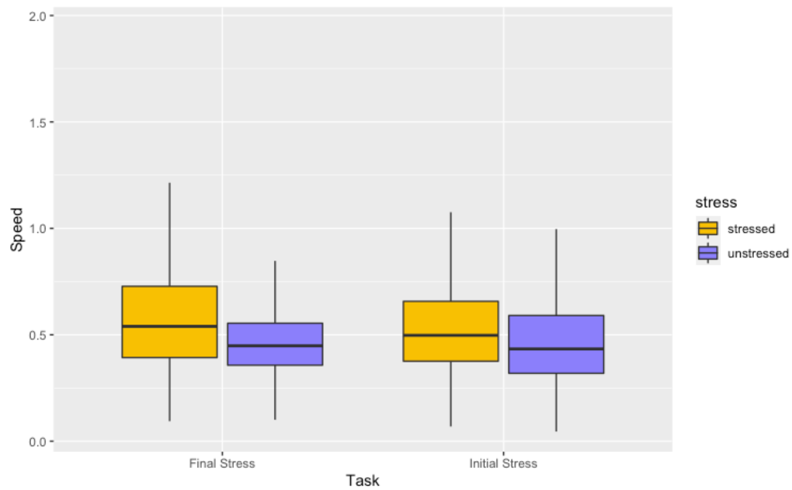


Figure 3.3: Maximum speed for stressed syllables and unstressed syllables in the medial stressed (final stress/VCV́) and unstressed tasks (initial stress/V́CV)

3.2.1.3 Reduction: x-position

X-position was calculated at the point of achievement or point of maximum constriction, where the x-position of the articulator in stressed syllables is expected to be more extreme (fronter for front vowels, backer for back vowels) in stressed syllables than unstressed syllables as unstressed vowels are reduced compared to stressed vowels. For x-position, the baseline model included phone label, preceding environment, following environment, maximum speed, and y-position. Duration was not a significant predictor of x-position, and therefore, was not included in the baseline model. Word position and stress were added to the model in a stepwise fashion and the models were compared using an ANOVA model comparison. However, for x-position, phones are not expected to have the same direction of effect; namely, while /i/ in unstressed syllables is expected to be backer than /i/ in unstressed syllables, /o/ and /a/ are expected to be fronter in unstressed syllables than stressed syllables. Thus, in addition to word position and stress, interactions between word position and phone and stress and phone were also added stepwise to the model.

	npar	AIC	BIC	logLik	deviance	Chisq	DF	Pr (>Chisq)	
baseline	24	-21930	-21756	10989	-21978				
+word position	25	-21954	-21772	11002	-22004	25.87	1	3.651e-07	***
+word:phone	27	-21963	-21767	11008	-22017	13.161	2	0.001387	**
+stress	28	-21962	-21758	11009	-22018	0.8753	1	0.3495	
+stress:phone	30	-21985	-21767	11022	-22045	27.169	2	1.26e-06	***

Table 3.5: ANOVA model comparison for x-position showing the baseline model for x-position compared to the model including word position, word position by phone, stress, and stress by phone.

As shown in Table 3.5, word position was significant regardless of interaction with phone, where vowels in final syllables were backer than those in initial syllables, as shown in Figure 3.4.

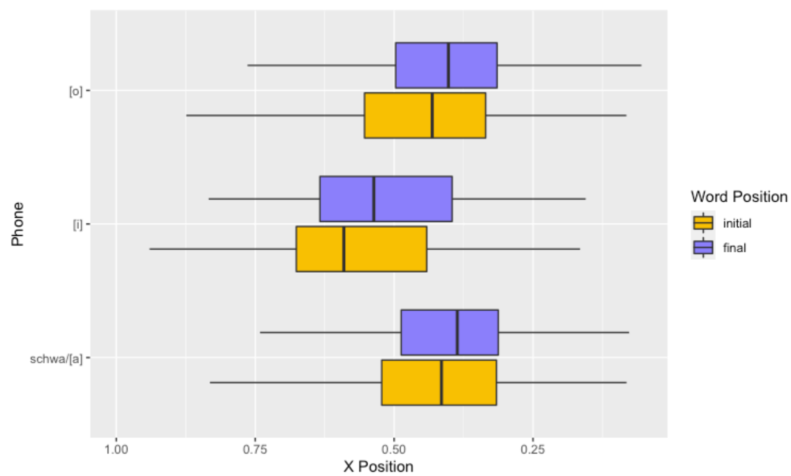


Figure 3.4: X position for [i], [o] and [ə]/[a] in initial and final syllables, where left is the front of the mouth and right is the back of the mouth

While initial prominence (Beckman, 1998) may be one explanation for this result, it is more likely that an effect of prominence would result in more extreme positions, similar to the effect of stress, as is shown in studies such as de Jong et al. (1993); de Jong (1995); thus, this result seems to instead be an effect of coarticulation.

Stress alone does not significantly improve the model; however, when including an interaction between stress and phone, stress is significant. This is because stress has a differing effect on each of the vowels, where vowels in stressed syllables are more extreme than those in unstressed syllables. Specifically, front vowels, /i/, in the case of this study, have a fronter gesture in stressed syllables than in unstressed syllables and back vowels have a backer gesture in stressed syllables than unstressed syllables. These results are shown in the Figure 3.5.

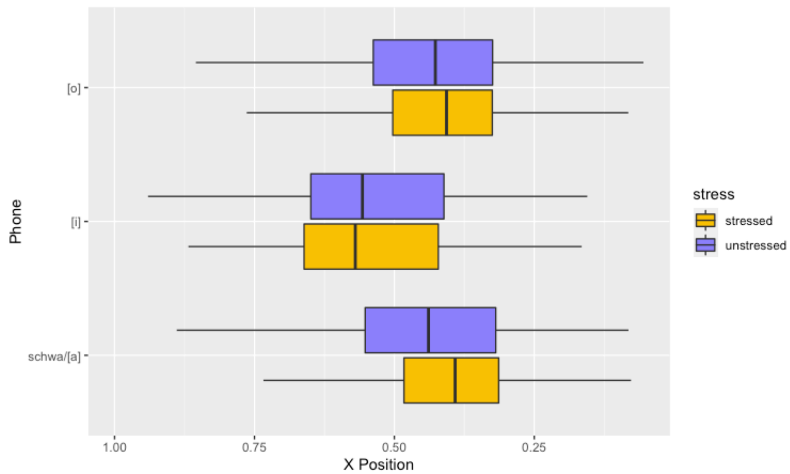


Figure 3.5: X position for [i], [o] and [ə]/[a] in stressed syllables and unstressed syllables, where left is the front of the mouth and right is the back of the mouth

3.2.1.4 Reduction: y-position

Y-position was calculated at the point of achievement or point of maximum constriction, where the y-position of the articulator in stressed syllables is expected to be more extreme (higher for high vowels, lower for low vowels) in stressed syllables than unstressed syllables as unstressed vowels are reduced compared to stressed vowels. For y-position, the baseline model included phone label, preceding environment, following environment, duration, maximum speed, and x-position. Word position and stress were added to the model in a stepwise fashion and the models were compared using an ANOVA model comparison. Like x-position, for y-position, phones are not expected to have the same direction of effect; namely, while /i/ in unstressed syllables is expected to be higher than /i/ in unstressed syllables, /a/ and /o/, which are both have the feature [+low], are expected to be lower in unstressed syllables than stressed syllables. Thus, in addition to word position and stress, an interaction between word position and phone and stress and phone was also added stepwise to the model.

	npar	AIC	BIC	logLik	deviance	Chisq	DF	Pr (>Chisq)	
baseline	25	-16780	-16598	8414.9	-16830				
+word position	26	-16862	-16673	8457.0	-16914	84.298	1	<2.2e-16	***
+word:phone	28	-17068	-16864	8561.8	-17124	209.62	2	<2.2e-16	***
+stress	29	-17079	-16868	8568.5	-17137	13.365	1	0.0002564	***
+stress:phone	31	-17142	-16916	8601.8	-17204	66.516	2	3.599e-15	***

Table 3.6: ANOVA model comparison for y-position showing the baseline model for y-position compared to the model including word position, word position by phone, stress, and stress by phone.

Table 3.6 demonstrates that both word position and stress as well as the interaction between word position and stress and phone were significant. As shown in Figure 3.6, final vowels were lower than initial vowels regardless of vowel quality. Like x-position, this effect is likely due to vowel to vowel coarticulation, where the carrier phrase, in which preceding and following vowel were consistent across tokens, affected production of vowels in the target.

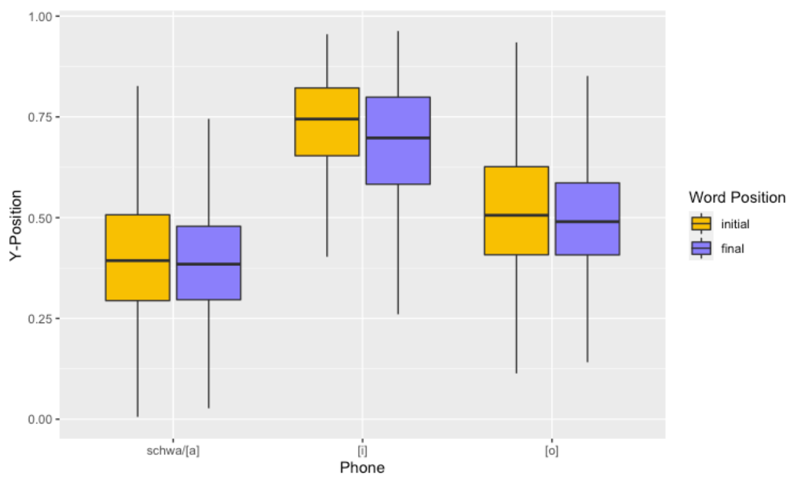


Figure 3.6: Y position for [i], [o] and [ə]/[a] in initial and final syllables

While both stress and stress with an interaction with phone were significant for the y-position model, the predictor with the interaction between stress and phone was a stronger predictor than stress alone. This is due to the differences in effect for the differing vowel qualities. Like x-position, stressed syllables are more extreme in position than unstressed vowels. In other words, the high vowel /i/ is higher in stressed syllables than unstressed syllables; and both [a] and [o], which both have the feature [+low], are lower in stressed syllables than in unstressed syllables.

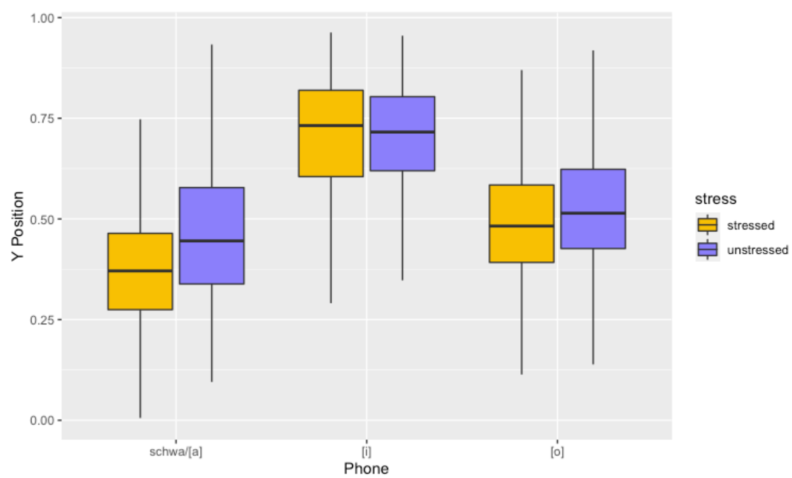


Figure 3.7: Y position for [i], [o] and [ə]/[a] in stressed and unstressed syllables

3.2.1.5 Interim summary

For vowels, the results of this study are consistent with previous literature, where vowel gestures in stressed syllables are of higher magnitude than unstressed syllables; in other words, speed is higher in stressed syllables than unstressed syllables, durations are longer in stressed syllables than unstressed syllables, and x and y position are more extreme in stressed syllables than unstressed syllables. These results are consistent with previous literature on the differences between stressed and unstressed syllables, e.g., Beckman and Edwards (1994); de Jong et al. (1993); de Jong (1995), and therefore, provide a verification of these methods before turning to consonant articulation.

While there were also significant effects for word position in predicting duration, x position, and y position, these effects are likely due to other conditions such as coarticulation, in the case of both position variables, and syllabification in the case of duration. Syllabification will be discussed in further detail in subsequent sections and chapters.

3.2.2 Consonants

Overall, the effect of stress on consonant articulation is less clear as results have been more varied across studies and participants (Beckman and Edwards, 1994). Furthermore, consonants have been found to be much less affected by stress than vowels (Gay, 1978). Nevertheless, where differences between consonants in stressed and unstressed syllables have been found, consonants in stressed syllables have been found to have greater magnitude than those in unstressed syllables, similar to the results for vowels (Beckman and Edwards, 1994). Likewise, the results of study one suggest that consonants in stressed syllables do have greater magnitudes than those in unstressed syllables, with longer /s/ durations in stressed syllables and higher speeds of /t/ in stressed syllables.

Typologically, both segmental distribution and commonly occurring phonological processes, like lenition, differ between word-boundary and word-medial environments, suggesting that word position may affect the gestural composition of a segment. While Katz (2016); Kingston et al. (2008) argue that differences in lenition are due to prosodic modulation, Kirchner (2001) argues that the differences between word-boundary, and word-medial lenition are due to a pressure for articulatory ease; however, few articulatory studies have been conducted to test whether there are articulatory differences between segments in word-initial vs word-medial position.

Thus, this section will analyze the effect of stress and word position on the following dependent variables: duration, speed, x position, and y position, where x position is interpreted as a proxy for reduction, and y position is interpreted as a proxy for degree of constriction. Generally, the predictors for these models were the same for consonants and vowels and thus included phone, preceding environment, and following environment as well as the remaining articulatory variables. Where any model varied from this set, it is discussed in the relevant section for each model. Subject was included as a random variable in all models.

3.2.2.1 Constriction results: y-position

The position of the sensor in the y-dimension at the point of achievement provides a measure of constriction degree. The baseline model analyzing constriction included previous phone, following phone, steady state duration, maximum speed, and sensor x-position. Word position and stress were added to the baseline model in a stepwise fashion. In addition, an interaction with phone was added for both word position and stress. An ANOVA model comparison of the model was used to compare the word position model to the baseline model and stress model to the word position model. The results of the model comparisons are provided in Table 3.7.

	npar	AIC	BIC	logLik	deviance	Chisq	DF	Pr (>Chisq)	
baseline	38	-25987	-25690	13032	-26063				
+word_position	39	-26013	-25708	13046	-26091	28.209	1	1.089e-07	***
+stress	40	-26012	-25700	13046	-26092	1.0635	1	0.3024	
+stress:phone	46	-26042	-25683	13067	-26134	41.868	6	1.952e-07	***

Table 3.7: ANOVA model comparison for sensor y-position when adding word position to the baseline model, when adding stress, and when adding an interaction between stress and phone

Because the shape of the data set is not a perfect two-by-two comparison, having both stressed and unstressed medial syllables, but only stressed initial syllables, an additional analysis over a subset of the data was analyzed to verify these results. For word position, this subset included only stressed syllables in initial and medial conditions. The dataset controlling for stress showed the same effects, with word position as a significant predictor of y position. As shown in Figure 3.8, the model accurately predicts a higher y-position of the sensor in initial syllables than medial syllables. Thus, segments in initial syllables are produced with tighter constriction than segments in medial syllables.

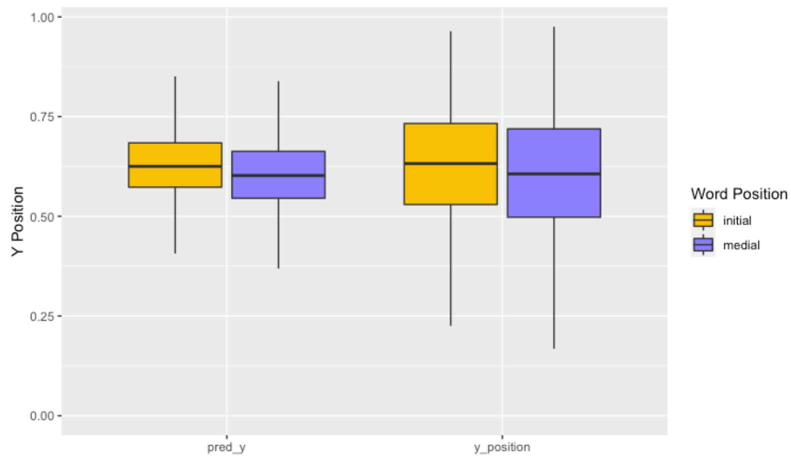


Figure 3.8: Predicted sensor y-position for initial and medial syllables in the model including word position alongside the measured values for sensor y-position

While stress was not significant on its own, when including an interaction between stress and phone, stress was significant, as shown in Table 3.7. Segments in unstressed syllables had a lower y-position, or looser constriction than segments in stressed syllables. This effect held across phone types, but to varying degrees. Moreover, when analyzing a subset of the data controlling for word position, which included only VCV/VCCV condition in both medial stressed and unstressed conditions, stress was significant without the interaction by phone, although the interaction did strengthen the effect. In both the full and subsetted dataset, the model accurately predicts that segments in stressed syllables have lower y-position values than stressed syllables, suggesting that consonants in stressed syllables have tighter constriction than consonants in unstressed syllables.

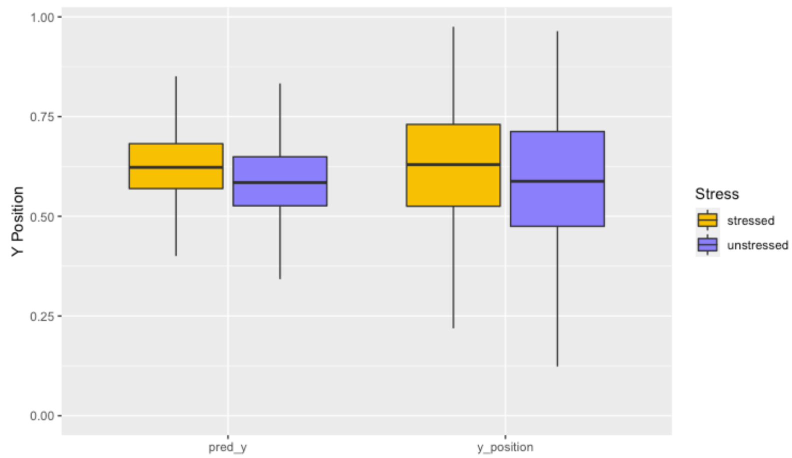


Figure 3.9: Predicted sensor y-position for stressed and unstressed syllables in the model including stress:phone along side the measured values for sensor y-position

3.2.2.2 Reduction results: x-position

Sensor position in the x-dimension provides a measure of reduction where, for example, coronal gestures may be backer and dorsal gestures may be fronter in reduced syllables compared to unreduced syllables. The model for reduction included previous phone, following phone, steady state duration, and maximum speed. Sensor y-position was not significant, and therefore, it was not included in the baseline model. Word position and stress were added to the linear regression mixed effects model in a stepwise fashion followed by an ANOVA model comparison. Furthermore, an interaction term with word position and phone and stress and phone were added to the model. Like sensor y-position, sensor x-position showed an effect of whether the segment was word-initial or word-medial. This effect was strengthened when adding the interaction with phone, but both were significant. Also like the y-dimension, stress was not significant on its own, but was significant when interaction with phone was added. The results of the model comparison are provided in Table 3.8.

	npar	AIC	BIC	logLik	deviance	Chisq	DF	Pr (>Chisq)	
baseline	38	-27541	-27044	13709	-27417				
+word_position	39	-27548	-27043	13713	-27426	8.8421	1	0.002944	**
+word:phone	44	-27509	-27165	13798	-27597	170.85	5	2.2e-16	***
+stress	45	-27507	-27155	13798	-27597	0.0154	1	0.9012	
+stress:phone	51	-27553	-27154	13827	-27655	57.819	6	1.247e-10	***

Table 3.8: ANOVA model comparison for sensor x-position when adding word position to the baseline model, when adding an interaction between word position and phone, when adding stress, and when adding an interaction between stress and phone.

As is indicated by the interaction between both word position and stress, the direction of the effect for x-position depends on the quality of the phone, where consonants like /t/ are fronter in both initial and in stressed syllables and consonsants like /k/ are backer. Additionally, the x-position for /p/[2] is backer in initial and stressed syllables. These results held in the controlled data sets as well, where word position was analyzed over a subset of the data containing only stressed syllables in initial and medial position and the stress was analyzed over a subset of the data containing stressed and unstressed syllables in only syllables in medial position. To illustrate these results, Figures 3.10 and 3.11 show the x-position for intervocalic /p/, /t/, and /k/.

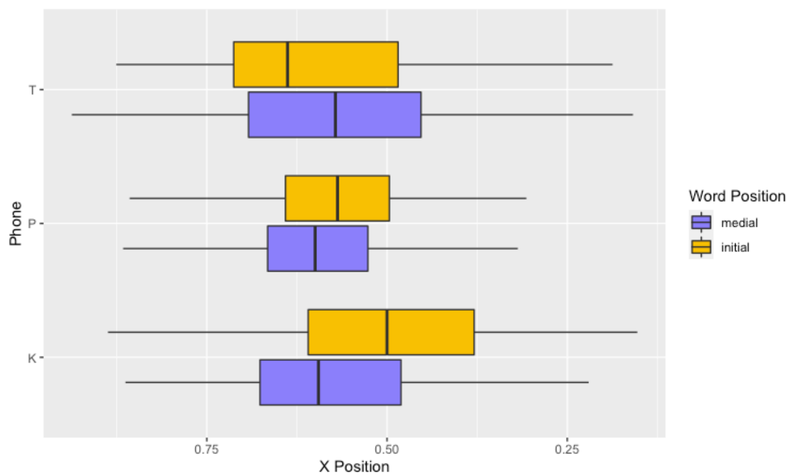


Figure 3.10: X-position for intervocalic /p/, /t/, and /k/ in initial and medial syllables, where left is the front of the mouth and right is the back of the mouth

[2]/p/ may be backer in unstressed syllables or fronter in stressed syllables do to jaw position; however this hypothesis is untested and is an area for future study.

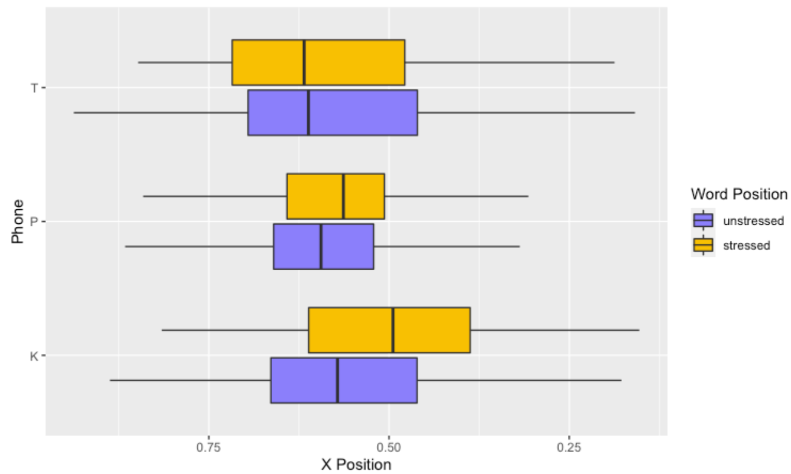


Figure 3.11: X-position for intervocalic /p/, /t/, and /k/ in stressed and unstressed syllables, where left is the front of the mouth and right is the back of the mouth

The results for x and y position differ from study one; however, as study one used a different method for calculating x and y position and as study one controlled for few variables, that there may be differences between the two studies is not surprising. The results are consistent, however, with the analysis of vowels, where segments in medial and unstressed syllables are more reduced than those in stressed syllables.

3.2.2.3 Duration results

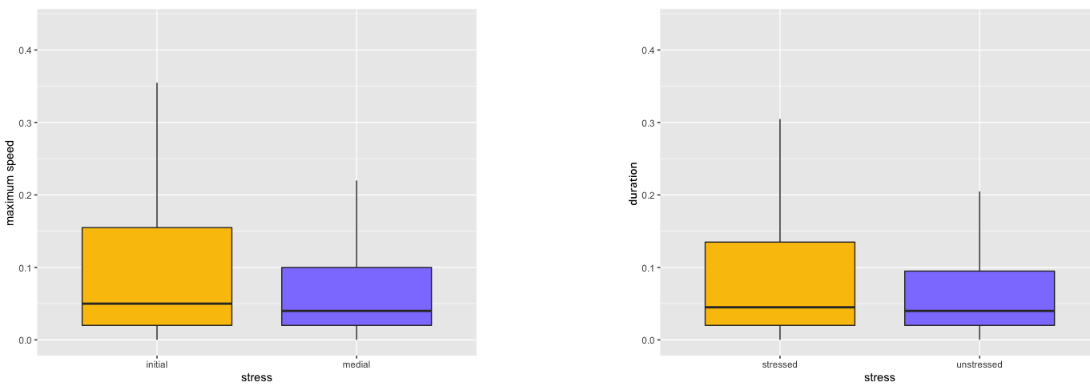
Duration was calculated as the time from the onset of steady state to the offset of the steady state. The baseline model for predicting steady state duration included previous phone, following phone, steady state, maximum speed, sensor y-position, and sensor x-position. The variables word position and stress were added to the models in a stepwise fashion and compared in an ANOVA model comparison. Neither word position nor stress significantly improved model fit, as shown by the results in Table 3.9; these results did not change when adding an interaction by phone or when subsetting the data for word position or stress.

	npar	AIC	BIC	logLik	deviance	Chisq	DF	Pr (>Chisq)
baseline	34	-49409	-49143	24738	-49477			
+word_position	35	-49408	-49134	24739	-49478	1.1666	1	<0.2801
+stress	36	-49409	-49135	24739	-49479	1.9997	1	0.1573

Table 3.9: ANOVA model fit for duration when adding word position to the baseline model and when adding stress to the baseline model

As shown in Figures 3.12a and 3.12b, there is a difference in variability in duration for both conditions where initial and medial syllables are more variable than stressed and unstressed syllables;

however, there is little difference in median duration when comparing initial and medial syllables or when comparing stressed and unstressed syllables. Based on the outputs of the model, any differences in duration are due to factors other than word position or stress, including the factors included in the baseline model, which include previous environment, following environment, y-position, x-position, and maximum speed.



(a) word position (b) stress

Figure 3.12: Measured duration for consonants in initial and medial and stressed and unstressed syllables in the model including word position and stress

These results again differ from both the first study and previous literature on the role of stress and segment duration. Regarding previous studies, this is likely in part due to the focus of this study on consonants rather than vowels. Additionally, segment identity, jaw oscillation, and syllable position also play a role in determining the differences between these gestures, and thus, further discussion of the interaction between these variables is provided in the following chapters.

3.2.2.4 Speed results

The baseline model predicting maximum speed included previous phone, following phone, steady state duration, sensor y-position, and sensor x-position. This baseline model was compared using ANOVA to a model for word position, which included all variables of the baseline plus word position. As shown in Table 3.10, adding word position to the model improved the fit of the model.

	npar	AIC	BIC	logLik	deviance	Chisq	DF	Pr (>Chisq)	
baseline	39	-17113	-16808	8595.4	-17191				
+word_position	40	-17170	-16857	8625.0	-17250	59.192	1	1.43e-14	***
+stress	41	-17179	-16859	8630.6	-17261	11.091	1	0.0008672	***

Table 3.10: ANOVA model comparison for predicting maximum speed when adding word position to the baseline model and when adding stress to the model including word position

As demonstrated in Figure 3.13, the model accurately predicts that segments in medial syllables are produced at higher speeds than segments in initial syllables. This effect held across phones.

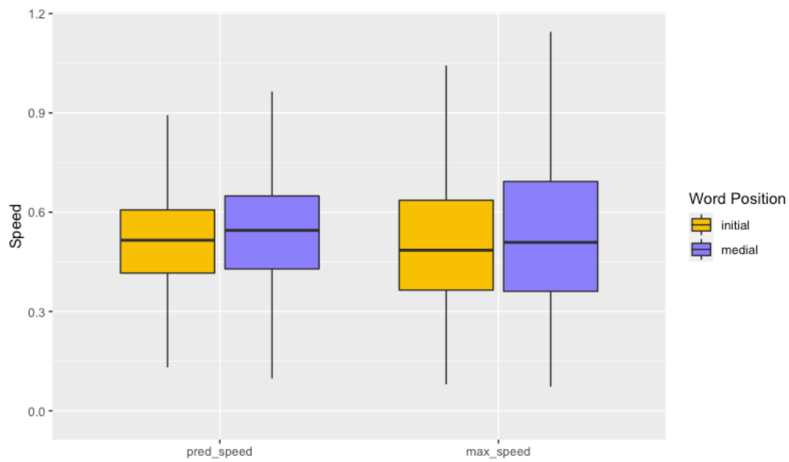


Figure 3.13: Predicted speeds for initial and medial syllables in the model including word position along side the measured values for maximum speed

Next, the model including word position was compared to a model including stress, which included the same variables as the word position model with the addition of stress as a predictor. Again, stress was a significant predictor of maximum speed where the model accurately predicts that unstressed syllables are produced at higher speeds than stressed syllables. The results of this model are of particular interest as they differ from both study one and previous studies, which found higher speeds in stressed syllables than unstressed syllables.

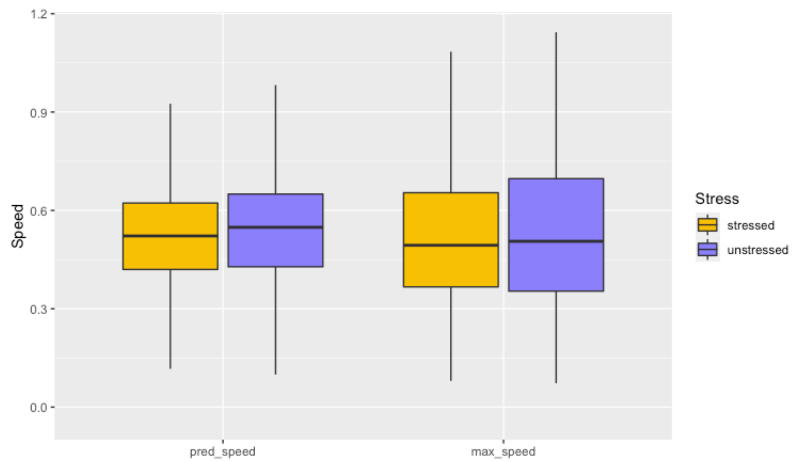


Figure 3.14: Predicted speeds for stressed and unstressed syllables in the model including word position and speed along side the measured values for maximum speed

3.3 Chapter summary and discussion

The results of this study show that word position has little effect on vowels outside of effects that are likely due to coarticulation. However, the results of this study do show that vowels differ significantly for speed, duration, x-position and y-position between stressed and unstressed syllables. In particular, vowels in stressed syllables occur at faster speeds, are of longer duration, and are less reduced than vowels in unstressed syllables. These results are consistent with previous literature such as Beckman and Edwards (1994).

However, for consonants both word position and stress were found to play a significant role in determining gestural composition. In particular, stress was significant in predicting constriction, reduction, and maximum speed, where speeds were faster for segments in unstressed syllables than those in stressed syllables; furthermore, segments in unstressed syllables were more reduced and had looser constrictions than segments in stressed syllables. Overall, these results are consistent with previous literature with the exception of speed, for which the findings of this study were counter to both study one and previous literature (Beckman and Edwards, 1994). However, given that vowels were consistent with previous findings, and in particular given that vowels are shorter in unstressed syllables than stressed syllables, the finding that speeds are higher for consonants in unstressed syllables may be due to the effect of vowels, where shorter vowel durations necessitate faster speeds for consonants in the same syllable.

Word position was also a significant predictor of speed, constriction, and reduction, where word-medial syllables have a higher velocity, are more reduced, and have a looser constriction than initial syllables. Furthermore, the findings are the same when analyzing a subset of the data comparing only stressed syllables in initial or medial positions. Thus, regardless of stress, there are significant articulatory differences between consonants that occur word-initially and those that occur word-medially.

These findings are generally consistent with differences in word external and word-medial lenition (e.g., Gurevich, 2011; Kirchner, 2001), as was presented at the beginning of this chapter and in Chapter 1; however, as word-medial lenition is restricted in English, where only /t/ and /d/ show phonemic alternation with /ɾ/, these results have crucial implications for the role of articulation in lenition. As discussed in Section 1, lenition differs word-initially and word-finally. Typical word-initial patterns include processes such as devoicing, like the pattern of final devoicing found in Bulgarian, e.g., *gradove* 'cities'   *grat* 'city'. Typical word-medial patterns include processes like spirantization, where a medial stop becomes a fricative. For instance, as shown in the following example, in Tümpisa Shoshoni /s/ becomes /ʃ/ and /p/ becomes /ɸ/ in word-medial position, e.g., /wisipin/ ⟶ [wiʃiɸi] 'thread' (Gurevich, 2011). Kirchner (2001) hypothesizes that lenition processes are driven by a grammatical pressure for economy of effort in articulation; however, both Kingston et al. (2008) and Katz (2016) argue that this difference between word-medial and word-initial stops is due to the desire for prosodic modulation. However, in an account of lenition where lenition is driven by perceptual factors, there is no expected difference in articulation in a language like English, where phonemic alternation is limited. Nevertheless, the results of this study show that there are articulatory differences between word-initial and word-medial segments even when there is little, if any, perceptible difference between the segments. In other words, in English, with the exception /t/ ⟶ /ɾ/ word-medially, most stops and fricatives do not alternate between initial and medial position; however, there is still an articulatory difference between these two positions. These results cannot be accounted for through prosodic modulation intended to help the listener, as predicted by a licensing-by-cue or segmentation-oriented model, necessitating instead an articulatory explanation for the difference between word-medial and boundary lenition.

Furthermore, these results are crucial in disambiguating between a perceptually driven account and an articulatory account. In languages where word-medial lenition has already been phonologized, there is no clear way to distinguish between an articulatory account and a perceptual account as the phonemic alternation necessitates an articulatory difference between environments. However, in a language like English, where word-medial lenition is not a phonological process, an articulatory difference between environments demonstrates that there are articulatory pressures that result in differences in constriction reduction, and speed. While these results demonstrate the need for articulatory constraints in modeling word-initial and word-medial differences, these results do not rule out perception constraints, nor do they disambiguate hyper-articulation in initial syllables as opposed to reduction in medial syllables. Indeed, perceptual pressure may drive lenition from the state of English, with relatively few phonologized alternations, and a language like Shoshoni or Spanish, with widespread alternations between boundary and word-medial positions. Instead, these results simply demonstrate that articulation is at least one of of the factors that results in differences between word-boundary and word-medial processes.

Chapter 4

The effects of word position and stress on syllable coordination

Principles of syllabification like the Onset Maximization Principle (Kahn, 1976; Selkirk, 1981) predict that syllables should behave uniformly regardless of whether they are word external or word-medial; however, as established in Chapter 1, there is abundant evidence that this is not the case. In addition, studies such as Katz (2012) and Browman and Goldstein (1988) provide both acoustic and articulatory support that syllables are coordinated as a unit and models of coordination like the Coupled Oscillator Model (Goldstein et al., 2009; Nam et al., 2009) predict that syllables should pattern uniformly regardless of position within the word; however, this hypothesis is untested as studies of syllable timing and coordination have focused on segments in environments where syllabification is straightforward. This chapter will analyze coordination using acoustic and articulatory data of sequences in word-initial and word-medial environments and in stressed and unstressed syllables to test whether word-medial syllables show the same global timing properties found in word-initial syllables and whether stress affects global timing patterns. In addition, this chapter will discuss the pressures corresponding syllabification schemata to guide the discussion of syllabification as it pertains to the results. The results of this study will demonstrate that word-medial syllables are distinct from word-initial syllables, consistent with the analysis that word-medial syllabification is variable. However, existing methods for analyzing coordination are not sufficient for determining what distinguishes word-initial and word-medial syllables, necessitating the development of new tools for analyzing variability in coordination.

4.1 Schematization of competing pressures on syllabification

Studies of stress (Byrd et al., 2009; Fujimura et al., 1978) and segment quality (Katz, 2016; Munhall et al., 1999) suggest that there are competing phonetic and phonological pressures that affect coordination and syllabification. This section will outline the predicted syllabification schemata of canonical models of syllabification and propose schemata for syllabification based on the effect of stress to guide the discussion of results provided in Sections 4.2 and 4.3.

Table 4.1 provides a position-based schema for syllabification, where consonants are syllabified either according to the Onset Maximization Principle; incrementally syllabified with segments being assigned to the onset then the coda in sequence; or to the theoretical possibility Coda Maximization Principle, in which consonants are preferentially syllabified as codas rather than onsets so long as they occur intervocalically. While coda maximization is a theoretical possibility, there is no theoretical or empirical support for coda maximization in English, and thus, the cell is grayed out as it is dispreferred. The examples in the table use ST as an example CC sequence where the CC forms a licit onset cluster, but can be substituted for any licit onset sequence.

Onset Maximization	**onset-coda**	**Coda Maximization**
V . STV	VS . TV	VST . V

Table 4.1: Position-based syllabification

Table 4.2 provides the syllabification schema according to a stress-based model of syllabification, where the examples are syllabified with stress having no effect, in which onset maximization is the default; with partial stress attraction, where stress may attract a single segment, but not the full consonant sequence; and with full stress attraction, where stress attracts all consonants into the coda, so long as they form a licit coda sequence. According to theoretical models of syllabification, either onset maximization or partial stress attraction are predicted to be more likely than the full stress attraction model because CV syllables are the least marked and because both jaw movement and perceptual cues should give preference to the CV syllabification in this context over a VCC syllabification. However, because sequences of multiple consonants have not been explored in studies analyzing the effect of stress on multiple segments, both are considered equally here.

	No Effect	**Partial Attraction**	**Full Attraction**
Medial Unstressed	V́ . STV	V́S . TV	V́ST . V
Medial Stressed	V . STV́	V . STV́	V . STV́

Table 4.2: Stress-based syllabification

Finally, Table 4.3 explores a combination of both position-based syllabification and stress-based syllabification[1], where the interaction of the two syllabification pressures may produce variability in syllabification, but the difference between medial stressed and unstressed tasks should produce at least a partial distinction in syllabification patterns.

[1]Adding ambisyllabicity to the schemata here would expand the predicted possible syllabifications. For example, using q-theoretic representations (Garvin et al., 2018; Shih and Inkelas, 2019), the initial stressed sequence may syllabify as VS(tt.t)V and the medial stressed sequence may syllabify as V(ss.s)TV. However, for the sake of simplicity, the question of ambisyllabicity will be returned to in the discussion.

Medial Unstressed	V́ST . V ∼V́S . TV
Medial Stressed	V . STV́ ∼VS . TV́

Table 4.3: Position + stress-based syllabification

These syllabification schemata make distinct predictions for gestural coordination. Comparing initial stressed syllables to medial stressed and unstressed syllables, the predictions are as follows: onset maximization predicts that syllable coordination should be the same regardless of stress and word position; however, onset-coda syllabification predicts that regardless of stress, word-medial syllables should pattern distinctly from word-initial syllables; stress-based syllabification predicts that while syllables with medial stress will pattern with the word-initial syllables, medial unstressed syllables will pattern distinctly from these two sets; finally, the schema with both position and stress based factors influencing syllabification predicts that all three conditions will pattern distinctly from each other.

In addition to the general patterns in syllabification, intrinsic pressures like segment quality, (e.g., MacNeilage, 1998; Vallée et al., 2009) or perceptual factors like strength of cue (e.g., Fujimura et al., 1978; Steriade, 1997) for a given sequence may also affect syllabification. Thus, analyzing segment quality aids in distinguishing the pressures on syllabification. The sequences analyzed in this study include both stop+liquid sequences, including /pl/, /pr/, /tl/[2], /tr/, /kl/, /kr/, and /s/+stop sequences including, /sp/, /st/, /sk/. A perceptual based model of syllabification predicts that classes of segments should pattern together based on what syllabification best cues place of articulation. For example, stops preceding stressed syllables should syllabify as onsets because the place of articulation for stops is best cued in the CV transition (Fujimura et al., 1978; Steriade, 1997). However, a Frame-Content based model of syllabification predicts that place of articulation will affect syllabification, where place of articulation affects jaw opening. Under this model an /sp/ cluster may be more likely to syllabify as VS.PV, whereas an /sk/ cluster may be more likely to syllabify as V.SKV, and an /st/ cluster may be most likely to syllabify V.STV.

The following sections will use three methods for analyzing syllable timing and coordination to test whether medial and unstressed syllables show evidence of global timing and to analyze patterns of syllabification across the data. Section 4.2 will present the methodology and results for compensatory shortening. Section 4.3 will present the methodology and results for analyzing the effect of c-centering and shift.

4.2 Compensatory shortening

There are two goals in analyzing compensatory shortening in this study. First, this study aims to determine whether medial and unstressed syllables show compensatory shortening of vowels between syllables with singleton onsets and syllables with cluster onsets. Comparing initial syllables to medial syllables and medial stressed syllables to medial unstressed syllables will

[2]While /tl/ is not a licit onset sequence in English, including /tl/ in the set helps to disentangle to what extent pressures like whether the sequence is a licit onset interact with other pressures like stress and intrinsic timing.

demonstrate whether stress and word position affect coordination. The second goal of this analysis is to identify which syllabification principles best account for the attested data in order to improve models of syllabification in English. This section will provide an overview of compensatory shortening. Section 4.2.1 will first analyze stressed syllables in initial and medial position in order to a) verify that the results of this study are consistent with previous studies, b) determine whether word position effects compensatory shortening. Next, Section 4.2.2 will analyze unstressed syllables to determine whether unstressed syllables pattern with stressed syllables. Finally, Section 4.2.3 will compare the results of the study to the syllabification schemata to analyze which syllabification principles best account for the results.

Studies using acoustic measures to demonstrate global timing have found that the number of segments within the syllable affects the duration of the vocalic nucleus. For instance, Munhall et al. (1999) demonstrates that vowels are shorter in the presence of an onset and/or a coda than in onsetless or open syllables, i.e., V » CV   VC » CVC. Furthermore, Munhall et al. (1999) and Katz (2012) demonstrate incremental compensatory shortening, where vowels become increasingly shorter in the presence of additional consonants in the onset, regardless of quality of the segments in the complex onset. While studies have consistently showed incremental compensatory shortening for onsets, evidence for incremental compensatory shortening in codas has varied. For instance, while Munhall et al. (1992) find evidence of incremental compensatory shortening in complex codas, Marin and Pouplier (2010) and Katz (2012) only find evidence of incremental compensatory shortening in complex coda clusters beginning with a liquid. Table 4.4 outlines the findings of previous studies to demonstrate when compensatory shortening is predicted for syllables in English.

	Onset	**Complex Onset**	**Coda**	**Complex Coda**
Compensatory Shortening	Yes	Yes	Yes	No*
Syllable comparison	V >C	CV >CCV	V >C	VC $\geq$ VCC

Table 4.4: The table shows the predictions for whether the addition of an onset or a coda segment will result in compensatory shortening. For complex codas, results have been mixed across studies, but Katz (2012) finds that compensatory shortening only occurs when the post vocalic coda segment of the cluster is an /l/, which does not occur in the dataset for this study (clusters are always stop+liquid, never liquid+stop). The table also provides a comparison of predicted vowel duration between syllables given the addition of the relevant onset or coda consonant.

The results of previous studies on compensatory shortening make predictions about whether compensatory shortening is expected depending how a given sequence is syllabified. For example, if the syllabification of the sequences VCV and VCCV is V.CV and VC.CV, the initial vowel is expected to be shorter in the cluster sequence compared to the singleton sequence because the coda in the cluster sequence causes compensatory shortening. However, if the syllabification is V.CV and V.CCV, there is no predicted difference between the vowels in the first syllable as both are a single onsetless codaless vowel. The following sections will outline whether compensatory shortening is predicted under each of the syllabification schemata for stressed and unstressed vowels, respectively.

4.2.1 Stressed syllable results

For stressed syllables, both the initial stressed task and the medial stressed task have the same expected compensatory shortening results regardless of the syllabification schema as both stress attraction and onset maximization predict the same syllabification. For the initial stressed task, tokens are monosyllabic and therefore there is only one possible syllabification: CV́C and CCV́C. For the medial stressed task, both basic syllabification schemata predict V.CV́ and V.CCV́ because the second syllable is the stressed syllable, and therefore, there is nothing to attract the intervocalic segments into the first syllable. Thus, under both syllabification schema, the vowel in the stressed syllable containing the cluster onset is expected to be shorter than the vowel in the syllable containing the singleton onset. However, for the medial stressed syllables, another less likely, but possible syllabification is that for the cluster, one segment will be syllabified as the coda for the first syllable and the other will be syllabified as the onset for the second syllable: VC.CV́. While this syllabification schema is attested in other languages, it is not supported for English and is, therefore, unlikely.

The medial unstressed condition, however, is more complex as there are several possible ways to syllabify the sequences. Table 4.5 outlines the predictions for the medial unstressed condition as well as the initial stressed and medial stressed conditions discussed above.

	Initial Stressed	**Medial Stressed**	**Medial Unstressed**
Onset Maximization	CV́C ∼ CCV́C	V.CV́ ∼V.CCV́	V́.CV ∼V́.CCV
	S >C	S >C	S = C
Partial Stress (a)	CV́C ∼ CCV́C	V.CV́ ∼V.CCV́	V́.CV ∼V́C.CV
	S >C	S >C	S >C
Partial Stress (b)	CV́C ∼ CCV́C	V.CV́ ∼V.CCV́	V́C.V ∼V́C.CV
	S >C	S >C	S = C
Full Stress	CV́C ∼ CCV́C	V.CV́ ∼V.CCV́	V́C.V ∼V́CC.V
	S >C	S >C	S $\geq$ C
Onset-Coda	CV́C ∼ CCV́C	V.CV́ ∼VC.CV́	V́.CV ∼V́C.CV
	S >C	S = C	S >C

Table 4.5: The table illustrates the predicted compensatory shortening results for each of the syllabification schemata and for each of the tasks for stressed syllables. 'S' stands for singleton and 'C' stands for cluster. Because the results for complex codas are mixed, where the relevant comparison has the difference between a singleton coda and a complex coda, the predicted compensatory shortening outcome is listed as $\geq$.

The main point to highlight in the table is that onset maximization, partial stress attraction (b), and full stress attraction all predict that there will be no compensatory shortening between the singleton and the cluster. For onset maximization, there will be no compensatory shortening because the initial stressed syllable is V (V́.CV and V́.CCV) for both sequences. For partial stress attraction (b), the initial stressed syllable is V́C (V́C.V and V́C.CV) for both sequences. And finally, for full stress attraction, the initial stressed syllable is V́C for the singleton and V́CC for the cluster, but because complex codas do not cause incremental compensatory shortening, no difference between the vowel durations is expected. However, under the schema partial stress attraction (a), compensatory shortening is predicted between the cluster and the singleton as there is predicted to be a difference between V́ (V́.CV) and V́C (V́C.CV).

As shown in Figure 4.1, both the initial stressed and medial stressed vowels show compensatory shortening where the sequence containing the singleton is shorter than the sequence containing the cluster. These results are consistent with previous studies such as Munhall et al. (1992) and Katz (2012) and demonstrate that medial stressed syllables pattern with the monosyllabic initial stressed condition, which is most similar to previous studies. In addition, although the difference between the singleton and cluster sequences in the medial unstressed condition are less than the initial and medial stressed conditions, the vowel in the singleton sequence is significantly shorter than the vowel in the cluster sequence ($p<0.024^*$). Thus, all three tasks show compensatory shortening.

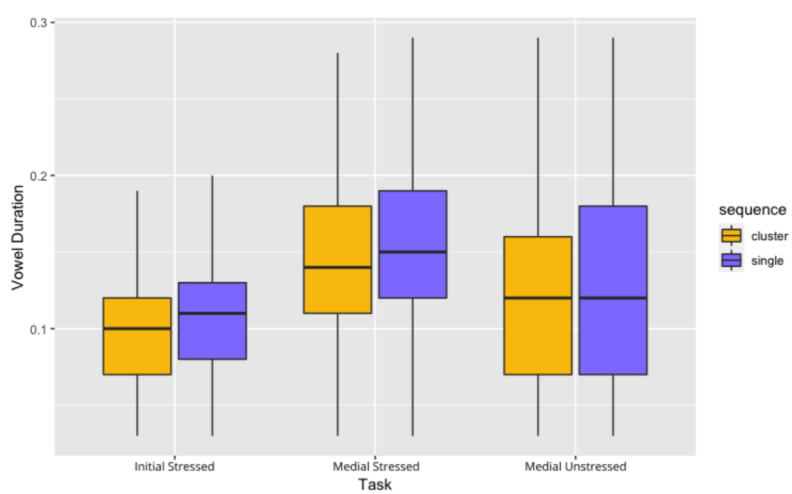


Figure 4.1: Comparison of duration for stressed vowels in tokens with singleton and cluster sequences across three task conditions: initial stressed and medial unstressed and medial stressed

As established above, neither the initial stressed results nor the medial stressed syllable results point to one syllabification schema over the other as both a stress-based model and onset maximization predict the same syllabification (although these results do rule out the unsupported onset-coda syllabification). For the medial unstressed condition, these results are most consistent with the partial stress (a) schema and the onset-coda schema[3]. Based on the results of the stressed

[3]These results could also be argued to be consistent with the full stress schema given the narrow results and the

syllables along with previous literature showing the effect of stress on syllabification, the most likely syllabification is the partial stress (a) schema.

While the compensatory shortening result for the medial unstressed syllables points to the partial stress (a) schema, another observation of the results for the medial unstressed syllables is the variability in this condition. The medial stressed condition is far more variable than the other two tasks, which may also suggest that syllabification in these contexts is variable. The discussion of variation will be returned to in Section 4.4.

4.2.2 Unstressed syllable results

Previous studies on compensatory shortening have not studied unstressed vowels, and therefore, whether unstressed syllables will show the same global timing patterns found in stressed syllables is unknown. Thus, this section will analyze the unstressed vowels in the medial stressed and medial unstressed conditions (there are no unstressed syllables in the initial stressed condition as the tokens are monosyllabic) to determine whether the results are consistent with previous literature and to further analyze which syllabification schema is most consistent with the results of this study.

For unstressed syllables, there is no expected difference between the singleton and cluster sequence in the medial stressed task condition for any of the basic syllabification schemata except the onset-coda syllabification schema as the unstressed syllable is expected to be only an onsetless, codaless vowel in either the onset maximization or stress-based syllabification schema. For the onset-coda condition, the singleton is expected to be shorter than the cluster sequence, but this outcome would be inconsistent with the results of the stressed syllables as the onset-coda schema would predict no compensatory shortening of the stressed vowels in the medial stressed condition, which is not the case. The predicted compensatory shortening outcomes are outlined for both conditions in Table 4.6.

For the medial unstressed condition, the predicted outcomes are again more complex as there are more possible syllabification outcomes. However, the most important predictions to note are first that for the partial stress attraction (a) schema, which is consistent with the stressed syllable results, there is no predicted difference between the singleton and cluster sequence. Furthermore, the only schema that predict compensatory shortening are the onset maximization schema and the partial stress attraction (b) schema. Overall, comparing the predictions for the stressed and unstressed syllables, for any given schema, if compensatory shortening is predicted for the stressed syllable, it is not predicted for the unstressed syllable and visa versa. Given that all three conditions showed compensatory shortening of stressed syllables, no compensatory shortening is predicted for the unstressed syllables.

---

ambiguity of compensatory shortening for complex clusters.

	Medial Stressed	Medial Unstressed
Onset Maximization	V.CV́ ∼V.CCV́	V́.CV ∼V́.CCV
	S = C	S >C
Partial Stress (a)	V.CV́ ∼V.CCV́	V́.CV ∼V́C.CV
	S = C	S = C
Partial Stress (b)	V.CV́ ∼V.CCV́	V́C.V ∼V́C.CV
	S = C	S >C
Full Stress	V.CV́ ∼V.CCV́	V́C.V ∼V́CC.V
	S = C	S = C
Onset-Coda	V.CV́ ∼VC.CV́	V́.CV ∼V́C.CV
	S >C	S = C

Table 4.6: The table illustrates the predicted compensatory shortening results for each of the syllabification schema and for each of the tasks for unstressed syllables. ‘S’ stands for singleton and ‘C’ stands for cluster.

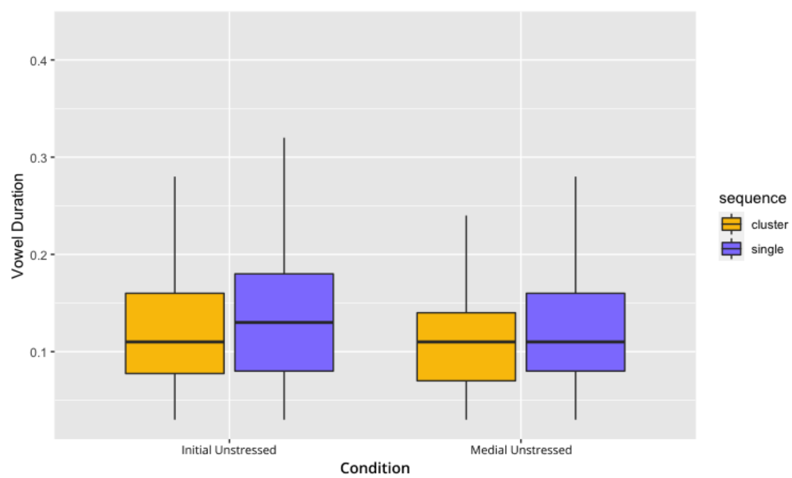


Figure 4.2: Comparison of duration for unstressed vowels in tokens with singleton and cluster sequences across the two task conditions: medial unstressed V́C(C)V and medial stressed VC(C)V́

Overall, the results shown in Figure 4.2 for unstressed syllables are consistent with previous literature, with both task conditions showing compensatory shortening. The results, however, are not consistent with the stressed syllables for either the medial stressed or the medial unstressed conditions. For both medial conditions, no compensatory shortening is predicted; however, vowels in the singleton sequence are longer than those in the cluster sequence. Likewise, vowels in the cluster sequence are significantly shorter than those in the singleton sequence in the medial unstressed tokens. Although the difference is again small, the results for the unstressed syllables are more significant than those in the stressed syllables (p<0.00269**).

While there are syllabification schemata that are consistent with the results of the stressed syllables or the results of the unstressed syllables, there is no syllabification schema that is consistent with the results of both the stressed and the unstressed syllables for either the medial stressed or the medial unstressed condition. Thus, the compensatory shortening results are most consistent with the variable syllabification schema represented in Table 4.3.

4.2.3 Interim summary and discussion

Overall, these results support the hypothesis that syllables in English are globally timed regardless of whether they are initial or medial and regardless of whether they are stressed or unstressed. However, a complicating factor of this finding is that none of the basic syllabification schemata can account for the attested results as there is evidence for compensatory shortening across all conditions; however, either a position-based or a stress-based model of syllabification predicts that if the stressed syllable shows compensatory shortening, the unstressed syllable should not, e.g., for V.CV V.CCV the second syllable will show compensatory shortening, but the first syllable will not. The most likely explanation for these results is that syllabification is variable, thus resulting in some degree of compensatory shortening in both stressed and unstressed syllables in both the medial stressed and medial unstressed vowels. Therefore, rather than demonstrating that syllables pattern uniformly regardless of stress or word position, the results indicate that word-medial syllabification is variable. Likewise, these results suggest that word-medial unstressed syllables are the most variable, as there was a smaller compensatory shortening effect and a wider range of vowel durations for both the stressed and unstressed vowels in this condition.

While variable syllabification is the most satisfactory analysis of these results, there is no direct method within the literature on compensatory shortening to confirm that syllabification is variable or to subset and reanalyze the data once it is grouped by syllabification pattern. Thus, further analysis is necessary to verify that these data present a unified analysis of compensatory shortening for unstressed syllables and for medial syllables once syllabification is determined. The next section will discuss the same questions of global timing and syllabification for unstressed and medial syllables, analyzing gestural coordination.

4.3 Gestural organization of syllables

Studies on articulatory timing and gestural coordination show global timing of segments in the syllable (Browman and Goldstein, 1988; Byrd, 1995). Studies on gestural timing have focused in particular on the coordination of onsets and complex onsets, as codas have been found to have little effect on the coordination of the syllable overall. For onsets, however, a number of studies (e.g., Browman and Goldstein, 1988; Byrd, 1995; Goldstein et al., 2009) have shown that onsets are coordinated with the following vowel. Global timing refers to a reorganization of timing between a sequence of segments with a singleton onset and a sequence of segments with a cluster onset, as in the compensatory shortening results; articulatory studies use two measures to analyze global timing: c-center and rightward shift. C-centering refers to an effect where the midpoint of the onset as a whole is timed to the following vowel, rather than each segment being timed only to immediately adjacent segments. Specifically, the midpoint of the onset has been found to be the most stable predictor of the following vowel showing less variability than the left or the right edge. This is true not only of singletons, but the average midpoint of all consonants in the onset is also the most stable predictor of the following vowel. These results suggest that the onset is timed as a unit to the following vowel. Rightward shift refers to a pattern similar to compensatory shortening, where in complex onsets the consonant gesture is shifted closer to the vowel resulting in a shorter lag time between the onset and following vowel in complex onsets compared to singletons; this effect would also cause compensatory shortening of the vowel.

Global timing patterns have been framed within the Coupled Oscillator Model, which proposes that onsets and vowels are timed synchronously, or within the same planning phase, i.e., in-phase, whereas vowels and codas are timed sequentially, e.g., anti-phase (Goldstein et al., 2009; Nam and Saltzman, 2003; Nam et al., 2009; Saltzman and Byrd, 2000). This model predicts that codas interact with vowels to a lesser degree than onsets[4]. Evidence for this pattern of planning come from studies such as Kozhevnikov and Chistovich (1966) and Löfqvist and Gracco (1999), which find that the initiation of the motor plan for both C and V gestures are synchronous. Proponents of the Coupled Oscillator Model argue that the synchronous timing of onsets and nuclei account for the stability and frequency of CV syllables cross-linguistically, as the stability of CV timing makes it more likely to occur than VC syllables. Furthermore, because of the privileged timing status of the CV sequence, the Coupled Oscillator Model inherently predicts that onset maximization syllabification should be preferable to other models of syllabification.

Studies analyzing the gestural coordination of syllables, however, have focused largely on onsets, syllables in stressed position and often initial stressed positions, and heterorganic sequences, leaving open questions about how inherent timing, word position, and stress impact global timing. This study seeks to explore such questions by analyzing a broader set of segments in both stressed and unstressed syllables and in initial and medial word positions. Section 4.3.1 will first present the results of study one. Subsequently, Section 4.3.1 will present the results of the c-center heuristics and the results of the leftward and rightward shift heuristics.

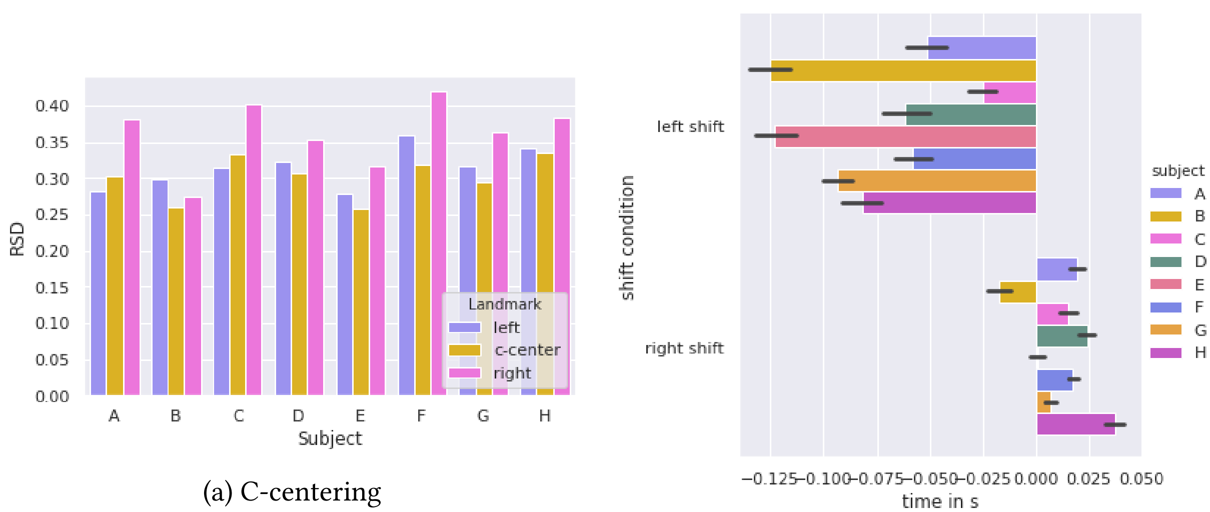
4.3.0.1 Results of study one

As has been established by previous chapters, study one was conducted to analyze the effect of stress and homorganic /st/ sequences on global timing. The study analyzed /st/ sequences in initial and medial stressed positions as well as medial unstressed syllables to determine whether

[4]Goldstein et al. (2009) propose that, in general codas, have little to no effect on the timing of the vowel. This provides an articulatory explanation for the asymmetry between complex onsets and complex codas discussed for compensatory shortening; however, it does not account for why some complex codas, like those beginning with a liquid, may cause more shortening than others, nor does it account for why single codas do cause compensatory shortening. Thus, Katz (2012) argues that the Coupled Oscillator Model cannot make correct predictions abut how syllables are timed. in particular, how codas are timed.

/st/[5] sequences showed the same indicators of global timing[6] as have been shown for heterorganic clusters, e.g., /sp/ and /pl/, in previous studies and to investigate whether unstressed syllables pattern with stressed syllables or show distinct timing patterns.

Overall, the results of the first study indicate that /st/ clusters in stressed syllables do pattern with previous results, showing both c-centering and rightward shift in stressed syllables. The results are presented in Figure 4.3. Figure 4.3a shows the results of the c-centering analysis for which the relative standard deviation (RSD) indicates the stability of the left edge, c-center, and right edge. Lower RSD values indicate less variability, and therefore, greater stability. For the stressed syllables, the c-center is typically the least variable with some minor variation across subjects. Figure 4.3b shows the results of the shift analysis, where the top portion of the graph shows shift of the left edge and all subjects show a leftward shift of the left edge. The bottom of the graph shows shift of the right edge, where the majority of subjects show a rightward shift of the right edge, with some minor variability between subjects.



(a) C-centering

(b) Shift

Figure 4.3: C-center and rightward and leftward shift results for stressed syllables in study one

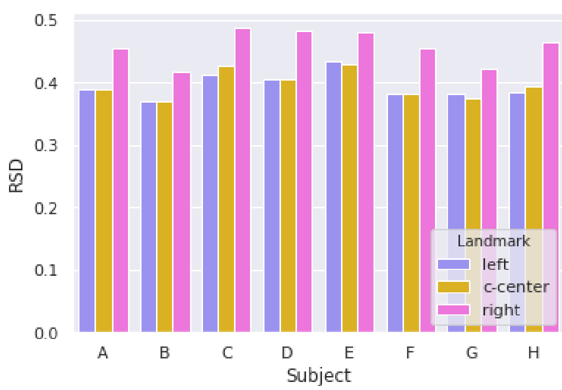
The results for the stressed syllables, therefore, show both global timing in stressed syllables and that for the stressed syllables, the cluster sequences are syllabified as complex onsets.

However, the patterns for unstressed syllables were less clear with overall weaker c-centering effect and no rightward shift, as shown in Figure 4.4. Figure 4.4a demonstrates that across subjects, the c-center is typically does not have the lowest RSD value; however the c-center does

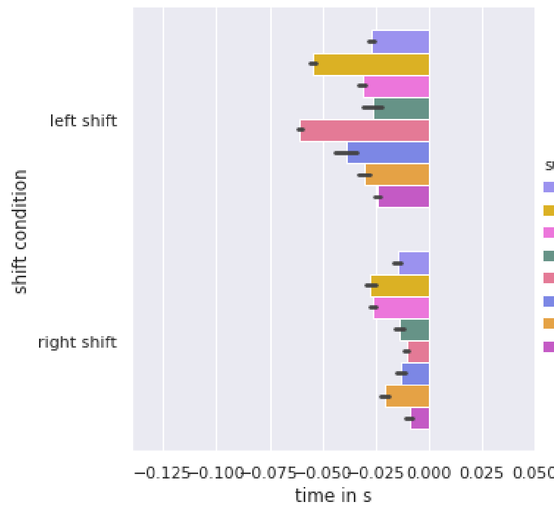
[5]Tokens with initial and medial stressed and unstressed singleton /s/ and /t/ sequences were analyzed along with the /st/ clusters in order to calculate the results for both c-centering and shift.

[6]Compensatory shortening was not analyzed in this study as the tokens consisted of real words from the MOCHA-TIMIT corpus, meaning that the token shapes were more varied, and therefore, the available comparisons were less straightforward.

trend with the left edge which does have the lowest RSD value. The RSD value is biased towards the left measure, and therefore, the fact that the c-center and left edge pattern together may still suggest some effect of c-centering; however, the pattern is notably different than the pattern in stressed syllables. Figure 4.4b shows the results of the shift analysis. While all subjects did show leftward shift of the left boundary, all subjects likewise showed leftward shift of the right boundary.



(a) C-centering



(b) Shift

Figure 4.4: C-center and rightward and leftward shift results for unstressed syllables in study one

These results do not provide strong support for global timing in unstressed syllables and may suggest that the complex consonant sequences in the unstressed tokens were not syllabified within the same syllable, instead syllabifying with one consonant as the coda of the previous stressed syllable and another as the onset of the unstressed syllable.

To investigate the role of syllabification in the difference between syllables, the study compared VCstV sequences and VstV sequences, e.g., ‘constantly’ and ‘hostages’, where syllabification of the /st/ as an onset is more likely in ‘constantly’ than in ‘hostages.’ The prediction is that if the /st/ sequence does occur in the onset, the right boundary should shift right, and thus, the VCstV sequences should show a rightward shift but the VstV sequences may not. However, there was no significant change in rightward shift between the two sequence types, leaving the explanation for why the unstressed tokens don’t show rightward shift unclear. Furthermore, the medial stressed syllables and initial stressed syllables did pattern together, suggesting that word position alone cannot account for the results of the unstressed syllables; however, there are too few initial unstressed clusters in English words and none in the MOCHA-TIMIT corpus, thus whether the results for unstressed syllables are due to word position, stress, or a combination is an open question. The second study follows up on these questions, analyzing sequences controlling for word position, stress[7], cluster composition, and vowel quality.

4.3.1 Results of study two

The results of the compensatory shortening analysis suggest that there is a difference between initial and medial syllables, where syllabification is variable in medial syllables, and between unstressed and stressed syllables, where medial unstressed syllables are syllabified more variably than medial stressed syllables. These results provide an inference for how the gestural studies are predicted to pattern, assuming that these acoustic and gestural measures correspond to the same underlying patterns of timing and coordination. Specifically, the initial stressed condition should show the strongest effect of c-centering, followed by the medial stressed condition, with the medial unstressed condition showing the least stability. Because the Coupled Oscillator Model predicts that onset maximization should be preferable to other syllabification schemata and because studies on gestural timing found little effect of the coda on timing (meaning that the first syllable of the VCV conditions cannot be analyzed and compared like in the compensatory shortening results), a single syllabification schema was considered for each of the tasks. The initial stressed condition was assumed to syllabify as CVC or CCVC, as the token is monosyllabic; the medial stressed condition was assumed to consist of either a stressed CV or a stressed CCV syllable, and the medial unstressed condition was assumed to consist of either an unstressed CV or an unstressed CCV syllable.

In this study, two measures were calculated to analyze the stability of the left edge, c-center, and right edge: the relative standard deviation (RSD) and the standard deviation (STD), where the relative standard deviation is the standard deviation divided by the mean. These measures provide a method for analyzing variability in the duration from each margin to the following vowel anchor. However, these measures are biased towards either the left edge for the RSD and the right edge for the STD. The RSD is biased toward the left edge as the left edge has the longest durations between the left edge and the vowel anchor, and therefore, there is a larger denominator in calculating the RSD, resulting in smaller left-edge values, or a bias for the left edge. The STD is biased toward the right edge for a similar reason, where the duration from the right edge to the following vowel anchor is the shortest duration, so it likewise has the smallest standard deviation. Thus, the RSD is most informative where either the c-center or the right edge is the lowest and the STD is most informative where the lowest STD is either the c-center or the left edge. To reduce the effect of bias in the stability measures, an additional level of analysis is included in the 'both' column of the results table, which combines outcomes of both the RSD and STD measures. The 'both' column is as follows: for each utterance, if the c-center has either the lowest RSD or the lowest STD, then the c-center is assumed to be the least variable because a low c-center outcome

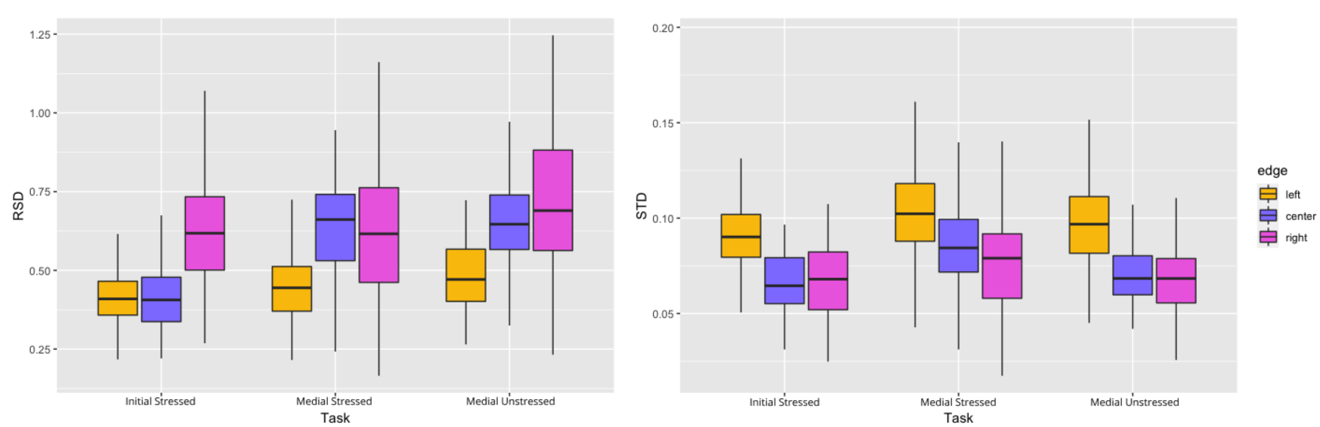
[7] Because English does not have unstressed monosyllabic words, such tokens were not included for comparison with the stressed monosyllabic tokens as speakers are unlikely to produce such sequences consistently. Another task in the study did include real word tokens that include initial unstressed clusters where they exist in the English lexicon; however these results are not included in this dissertation as they do not control for word length or number of tokens and speakers vary in whether the initial syllable containing the cluster is produced as reduced or secondary stress, thus, analysis of this task was left for a later time.

is more informative than a low RSD value for the left edge or a low STD value for the right edge. If the c-center was not the least variable according to either the RSD or the STD measures, the 'both' column indicates the margin with the lowest STD.

	Initial Stressed			**Medial Stressed**			**Medial Unstressed**			**Total**		
	RSD	STD	Both	RSD	STD	Both	RSD	STD	Both	RSD	STD	Both
Left Edge	49%	2%	2%	76%	14%	14%	86%	11%	11%	66%	5%	5%
C-Center	51%	56%	87%	10%	30%	35%	9%	38%	39%	28%	44%	60%
Right Edge	0%	42%	11%	14%	66%	61%	5%	51%	49%	6%	51%	35%

Table 4.7: Table showing percentage of each margin having the lowest RSD or STD and a combination of both measures, where if CC was the lowest for either the RSD or STD metric, then CC was analyzed as having the least variability; else, the margin with the lowest STD was analyzed as being the least variable.

As shown in Table 4.7, there is a clear distinction between the initial stressed condition and the two medial conditions. In the initial stressed condition, the c-center is most often the least variable measure across the RSD, STD, and combined metrics. However, for the word-medial conditions, the left edge is the least variable for the RSD and the right edge is the least variable in the STD and combined metrics. This outcome is indicative of the bias in the RSD and STD measures and can only be tentatively interpreted; however, these results illustrate two main findings. First, there is a clear effect of word position, where only the initial stressed condition clearly shows a c-centering effect. Second, rather than indicating the syllabification induced by the task environments, these results suggest that there is overall more variability for the medial conditions, where the increased variability among the utterances reduces stability, resulting in higher RSD and STD values for the c-center.



(a) RSD (b) STD

Figure 4.5: RSD and STD values for the left edge, c-center, and right edge

While the overall results indicate more variability in the VCV and VCCV conditions, there are additional trends in this data that can be meaningfully analyzed. First, looking at the combined RSD and STD column, the c-center in the medial unstressed condition is more likely to be the least variable compared to the medial stressed condition. This result is surprising given that the inference from the compensatory shortening results was that the medial unstressed condition would be more variable than the medial stressed condition. However, these results can be meaningfully interpreted by considering how vowel duration affects c-centering. While the medial unstressed condition was more variable, the vowels in the medial unstressed condition were shorter overall than the medial stressed condition. Furthermore, the syllable in question in the medial unstressed condition is the unstressed syllable, and as such, the vowel is expected to be shorter and more reduced than the stressed vowel in the medial stressed condition. According to Shaw et al. (2011), a shorter nucleus duration improves the stability of the c-center to anchor for simplex onsets, while compression of the nucleus is degraded for complex onsets; thus, the improvement in the c-center stability may not suggest that the c-center is more stable for medial unstressed than medial stressed , but instead that simplex onset timing is more common in the medial unstressed condition than the medial stressed condition and in combination with the shorter vowel duration, the c-center appears less variable in the medial unstressed condition than in the medial stressed condition.

Turning to the shift metric, there is a less distinguishable difference across tasks, where all three tasks show evidence of rightward shift, with the medial stress condition shifting the most and the medial unstressed condition shifting the least, as shown in Figure 4.6.

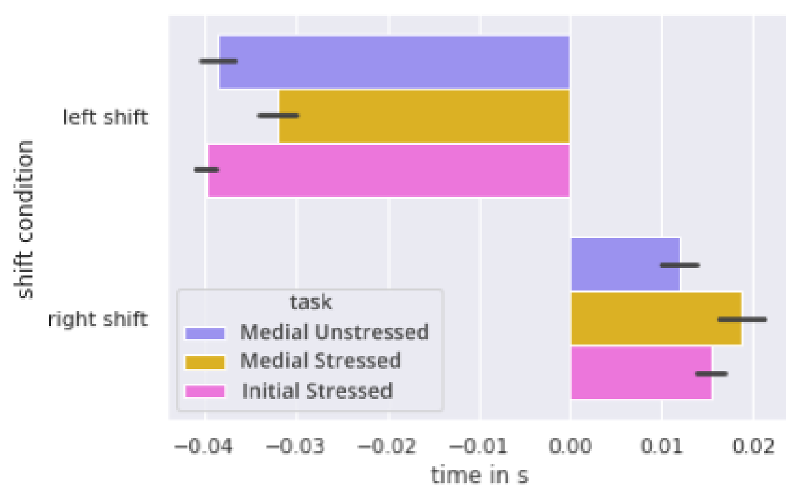


Figure 4.6: Rightward and leftward shift for each of the tasks

Similarly, across subjects, in the majority of cases, subjects showed rightward shift across all three tasks. While all subjects showed rightward shift for the medial stressed condition, three subjects did not show rightward shift for the initial stressed condition and three subjects did not show rightward shift for the medial unstressed condition; there was no overlap between subjects who did not show rightward shift in the initial stressed condition and subjects who did not show shift in the medial unstressed condition, as shown in Table 4.8.

	Initial Stressed	Medial-Stressed	Medial-Unstressed
A	N	Y	Y
B	N	Y	Y
C	Y	Y	Y
D	Y	Y	N
E	Y	Y	N
F	Y	Y	N
G	Y	Y	Y
H	Y	Y	Y
I	Y	Y	Y
J	Y	Y	Y
K	N	Y	Y
L	Y	Y	Y
M	Y	Y	Y
N	Y	Y	Y
O	Y	Y	Y
P	Y	Y	Y
Q	Y	Y	Y

Table 4.8: Presence of rightward shift for each task for each subject, where Y means there was rightward shift

Thus, much like study one, while the initial stressed condition is consistent with previous studies on stability and coordination, the results for c-center and shift measures in the medial conditions were mixed. However, unlike study one in which there was a weak c-centering effect, but no rightward shift; in study two, the c-center was typically not the most stable measure of the following vowel, but there was rightward shift across all three tasks. The difference between the first and second study could be due to a number of effects as there are several points of departure between the two studies. First, study one looked at real words in varying sentences, whereas study two looked at nonce words in a carrier phrase. As such, the difference between the tasks could be due to a difference between real words and nonce words or even between the tasks themselves. For instance, while the carrier phrase helped speakers to produce the target words in a more natural context while controlling for surrounding context and prosodic environment, the words in the MOCHA-TIMIT corpus were more naturalistic and more varied environments. Second, speakers had some instruction as to syllabification and stress in the second study, but there was no such instruction in study one. For instance, in study two, participants were given an example to highlight the stress pattern for the tasks, e.g., oNO or Ono. Despite the difference between the first and second study, both studies show a clear difference in coordination between the initial and medial conditions.

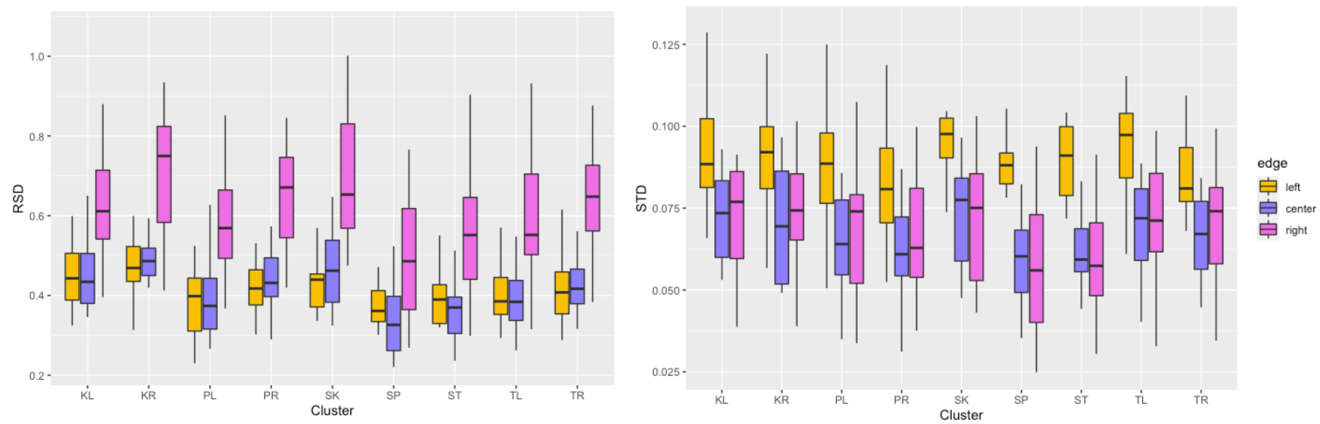
4.3.1.1 Coordination by cluster

Breaking down the results by cluster illustrates further patterns in the data and demonstrates how competing pressures like articulation and perception affect syllabification. Looking first at the initial stressed task condition, example (1) shows the the target clusters ranked by median c-center RSD and STD from highest to lowest, summarizing the results presented in Figure 4.7 and Tables 4.9 and 4.10; the c-center of the clusters with an asterisk have the lowest median RSD or STD values, illustrated in Figure 4.7.

(1) a. **RSD**: KR > SK > KL* > PR > TR > TL* > PL* > ST* > SP*
    b. **STD**: SK > KL* > TL > KR* > TR* > PL* > PR* > SP > ST

These results are consistent with previous literature, which has shown that /s/+stop clusters are more variable than stop+liquid clusters. Both stop+/l/ and stop+/r/ clusters were more likely compared to /s/+stop cluster to have lower c-center STD values than lower right edge STD values across all consonant types. Likewise clusters for which the c-center had the lowest median RST and STD values include the stop+/l/ clusters, with /l/ more likely to have lower RSD values than their /r/ counterparts across all three stop consonants. These results additionally provide support for perception pressures on syllabification, where these results pattern with the findings of Katz (2012) who found greater compensatory shortening for laterals compared to rhotics.

Next, the results suggest that homorganic clusters pattern with heterorganic clusters and likewise provide support for the effect of intrinsic pressures on coordination. Looking at the overall median ranking values illustrated in (1), clusters with /k/ had higher median STD and RSD values for both cluster types than either the clusters containing /p/ or /t/. The difference in coordination across place of articulation suggests that intrinsic pressure condition coordination. Furthermore, homorganic clusters, both stop+liquid and /s/+stop clusters patterned with the heterorganic PL, PR, and SP clusters, showing that global timing extends to homorganic clusters.



(a) RSD

(b) STD

Figure 4.7: RSD and STD values for the left edge, c-center, and right edge in the initial stressed task condition

Initial Stressed									
	SP			**ST**			**SK**		
	RSD	STD	Both	RSD	STD	Both	RSD	STD	Both
Left Edge	23%	0%	0%	23%	0%	0%	64%	0%	0%
C-Center	77%	29%	94%	77%	47%	88%	36%	64%	82%
Right Edge	0%	71%	6%	0%	53%	22%	0%	36%	18%

Table 4.9: Table showing percentage of each margin having the lowest RSD or STD and a combination of both measures for /sp/, /st/, and /sk/ in the initial stressed condition, where if CC was the lowest for either the RSD or STD metric, then CC was analyzed as having the least variability; else, the margin with the lowest STD was analyzed as being the least variable.

Initial Stressed																		
	PR			**PL**			**TR**			**TL**			**KR**			**KL**		
	RSD	STD	Both	RSD	STD	Both	RSD	STD	Both	RSD	STD	Both	RSD	STD	Both	RSD	STD	Both
Left Edge	64%	11%	11%	64%	0%	0%	64%	5%	5%	40%	0%	0%	58%	0%	0%	58%	0%	0%
C-Center	36%	53%	76%	36%	64%	82%	36%	70%	88%	60%	65%	100%	42%	63%	82%	42%	58%	83%
Right Edge	0%	36%	12%	0%	36%	18%	0%	24%	6%	0%	35%	0%	0%	37%	18%	0%	42%	17%

Table 4.10: Table showing percentage of each margin having the lowest RSD or STD and a combination of both measures for for /pr/, /pl/, /tr/, /tl/, /kr/, and /kl/ in the initial stressed condition, where if CC was the lowest for either the RSD or STD metric, then CC was analyzed as having the least variability; else, the margin with the lowest STD was analyzed as being the least variable.

The rightward shift results pattern similarly to the c-centering results, where across all clusters but the /kr/ cluster, there is clear rightward shift, as shown in Figure 4.8

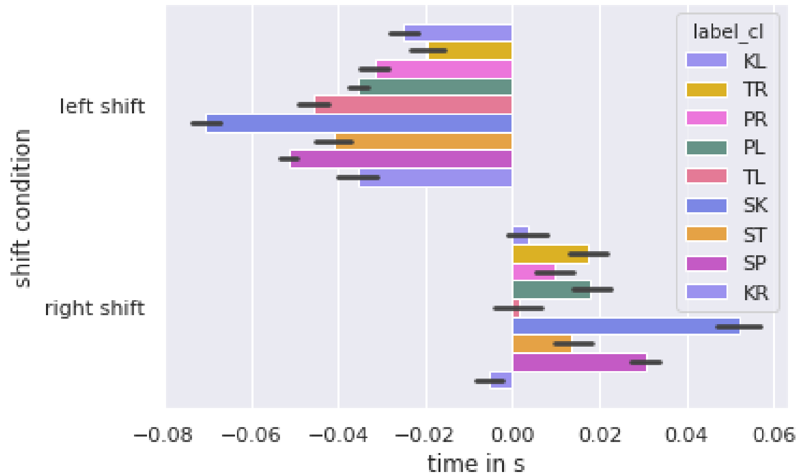
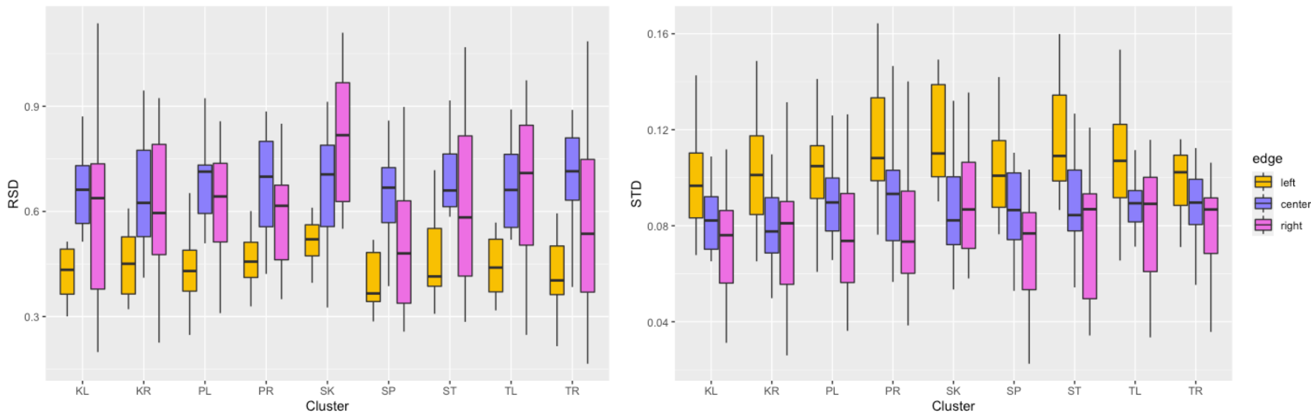


Figure 4.8: Rightward and leftward shift for the initial stressed task condition for each of the cluster types

The analysis of shift across clusters also provides additional insight into the differences across task conditions, where the effect of rightward shift was more consistent across cluster types in the initial stressed condition than the other two conditions, as shown in Figures 4.9 and 4.11. These results are consistent overall with previous literature, e.g., Goldstein et al. (2009). While the /kr/ cluster did not show rightward shift, stop+/r/ clusters have not been included in other studies analyzing rightward shift and previous studies have likewise found some variability across cluster types (Goldstein et al., 2009). Furthermore, the results for /kr/ were consistent across the RSD and shift conditions, where /kr/ cluster also had the highest RSD value.

While the medial task conditions yielded distinct patterns across the cluster types from the initial stressed condition, the overall differences between the initial stressed task and the other two tasks makes this outcome unsurprising. Figures 4.9 and Tables 4.11 and 4.12 provide the c-centering results.



(a) RSD (b) STD

Figure 4.9: RSD and STD values for the left edge, c-center, and right edge in the medial stressed task condition

Medial Stressed									
	SP			**ST**			**SK**		
	RSD	STD	Both	RSD	STD	Both	RSD	STD	Both
Left Edge	77%	0%	0%	72%	16%	16%	89%	0%	0%
C-Center	6%	11%	17%	6%	27%	27%	12%	66%	66%
Right Edge	17%	89%	84%	23%	67%	67%	0%	34%	34%

Table 4.11: Table showing percentage of each margin having the lowest RSD or STD and a combination of both measures for for /sp/, /st/, and /sk/ in the medial stressed condition condition, where if CC was the lowest for either the RSD or STD metric, then CC was analyzed as having the least variability; else, the margin with the lowest STD was analyzed as being the least variable.

Medial Stressed																		
	PR			**PL**			**TR**			**TL**			**KR**			**KL**		
	RSD	STD	Both	RSD	STD	Both	RSD	STD	Both	RSD	STD	Both	RSD	STD	Both	RSD	STD	Both
Left Edge	89%	82%	82%	84%	12%	12%	67%	7%	7%	76%	10%	10%	77%	0%	0%	77%	0%	0%
C-Center	0%	7%	7%	6%	6%	6%	0%	28%	28%	12%	33%	39%	7%	35%	35%	0%	17%	17%
Right Edge	12%	11%	11%	10%	82%	82%	33%	66%	66%	12%	56%	51%	16%	65%	65%	23%	83%	83%

Table 4.12: Table showing percentage of each margin having the lowest RSD or STD and a combination of both measures for for /pr/, /pl/, /tr/, /tl/, /kr/, and /kl/ in the medial stressed condition condition, where if CC was the lowest for either the RSD or STD metric, then CC was analyzed as having the least variability; else, the margin with the lowest STD was analyzed as being the least variable.

The most notable result demonstrated here is the variability in in the medial stressed condition compared to the initial stressed condition, demonstrated in particular by the tables demonstrating how frequently each edge was the least variable. Particularly, in the initial stressed condition the RSD and STD had relatively consistent results, where in the medial stressed condition there is a much clearer bias towards the left edge in the case of the RSD, and towards the right edge in the case of the STD, creating a bigger swing comparing between the two results. Furthermore, as shown in Figure 4.9, where the c-center was either the lowest or patterned with the lowest edge in the initial stressed task, the c-center is either the most variable or patterns with the most variable edge.

Despite the variability indicated by the c-center results for the medial stressed task, the shift results still indicate rightward shift for most clusters, as indicated by Figure, 4.10.

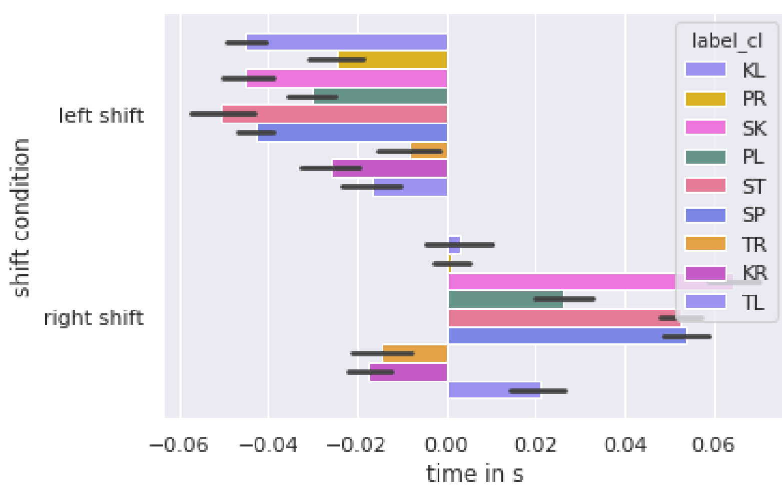
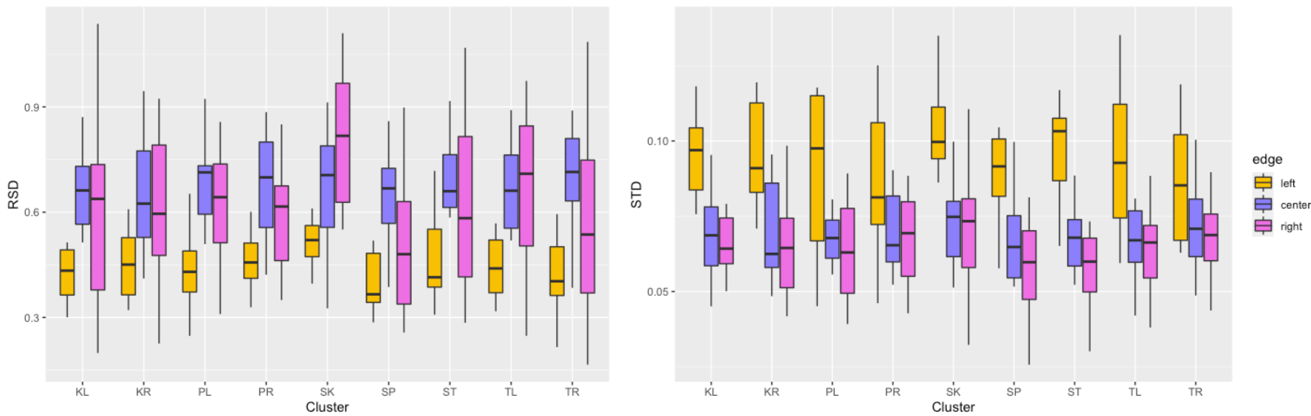


Figure 4.10: Rightward and leftward shift for the medial stressed task condition for each of the cluster types

In the medial stressed condition, fewer segments show rightward shift than in the initial stressed condition, and where there is leftward shift of the right boundary, there is a greater leftward shift in the medial stressed condition than in the initial stressed condition. Nonetheless, similarly to the c-centering patterns in the initial stressed condition, the stop+/r/ clusters are less likely to show rightward shift and show less rightward shift than any other cluster. Furthermore, the stop+/l/ clusters show less rightward shift than the /s/+stop clusters, which is likewise consistent with the initial stressed results in which the /s/+stop clusters also tended to have greater rightward shift than the other clusters, although there were mixed results between the stop+/l/ and stop+/r/ clusters.

Finally, as was also true of the medial stressed condition, there is clear evidence of variability in the medial unstressed condition, as shown in Tables 4.13 and 4.14. However, unlike the medial stressed condition, while the c-center does correlate with the most variable RSD, for the STD, the c-center correlates with the right edge. This result is consistent with the main results; however, analyzing the STD by cluster also elucidates that for the stop+/r/ clusters, the c-center tends to have the lowest STD. These results are illustrated in Figure 4.11.



(a) RSD (b) STD

Figure 4.11: RSD and STD values for the left edge, c-center, and right edge in the medial unstressed task condition

Medial Unstressed									
	SP			**ST**			**SK**		
	RSD	STD	Both	RSD	STD	Both	RSD	STD	Both
Left Edge	81%	6%	6%	81%	0%	0%	100%	0%	0%
C-Center	0%	27%	27%	7%	34%	34%	0%	50%	50%
Right Edge	20%	67%	67%	12%	66%	66%	0%	50%	50%

Table 4.13: Table showing percentage of each margin having the lowest RSD or STD and a combination of both measures for for /sp/, /st/, and /sk/ in the medial unstressed condition condition, where if CC was the lowest for either the RSD or STD metric, then CC was analyzed as having the least variability; else, the margin with the lowest STD was analyzed as being the least variable.

Medial Unstressed																		
	PR			**PL**			**TR**			**TL**			**KR**			**KL**		
	RSD	STD	Both	RSD	STD	Both	RSD	STD	Both	RSD	STD	Both	RSD	STD	Both	RSD	STD	Both
Left Edge	94%	24%	24%	85%	17%	17%	83%	68%	68%	83%	12%	12%	88%	6%	6%	84%	6%	6%
C-Center	7%	34%	34%	5%	25%	25%	18%	29%	34%	17%	40%	46%	12%	40%	40%	10%	51%	51%
Right Edge	0%	41%	41%	10%	58%	58%	0%	43%	37%	0%	48%	43%	0%	54%	54%	6%	43%	43%

Table 4.14: Table showing percentage of each margin having the lowest RSD or STD and a combination of both measures for for /pr/, /pl/, /tr/, /tl/, /kr/, and /kl/ in the medial unstressed condition condition, where if CC was the lowest for either the RSD or STD metric, then CC was analyzed as having the least variability; else, the margin with the lowest STD was analyzed as being the least variable.

This result is consistent with the results of the shift, in which the stop+/l/ rather than the stop+/r/ clusters were less likely to show rightward shift and had a lesser degree of rightward shift than other clusters, with the exception of the /kr/ cluster, which did not show rightward shift across any of the three tasks. These results are shown in Figure 4.12.

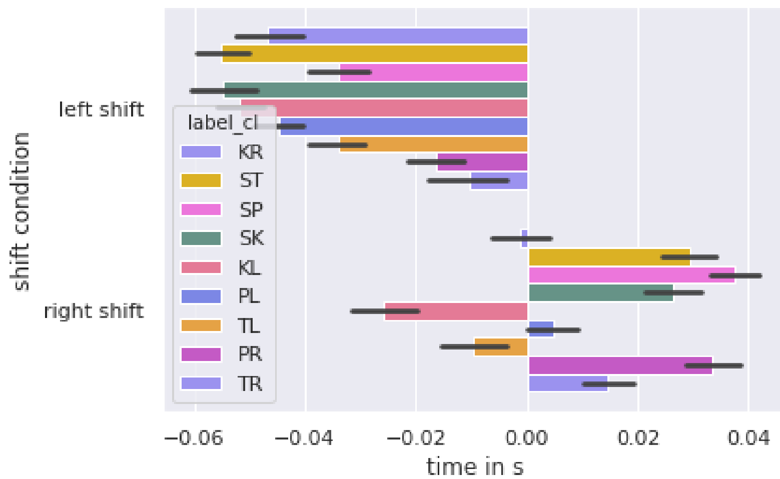


Figure 4.12: Rightward and leftward shift for the medial unstressed task condition for each of the cluster types

Thus, there is contrast between the initial and medial stressed conditions between the stop+liquid clusters, where in the medial stressed task, the stop+/l/ clusters had greater rightward shift and were more likely to show rightward shift, similar to the stop+/l/ clusters in the initial stressed task. Whereas in the medial unstressed task, the stop+/r/ clusters were more likely to show rightward shift and showed greater rightward shift than the stop+/l/ clusters. The similarity between the medial stressed and initial stressed tasks in the behavior of specific clusters provides tentative support that syllabification in medial stressed clusters is more comparable to initial stressed clusters than medial unstressed clusters are to initial stressed clusters.

Looking at the overall differences between segments, segments differ in stability depending on place of articulation. Table 4.15 illustrates differences in rightward shift by cluster for each subject, providing the number of subjects that showed rightward shift for a given cluster. The table indicates that /s/+stop clusters were the most likely to show rightward shift across participants, while clusters containing /k/ were the least likely to show rightward shift.

	SP	**ST**	**SK**	**PL**	**PR**	**TL**	**TR**	**KL**	**KR**
Initial Stressed	13	10	12	12	10	9	11	7	7
Medial Stressed	11	14	11	9	9	11	7	10	3
Medial Unstressed	14	13	13	10	13	6	11	7	9

Table 4.15: Number of subjects out of 17 where there was rightward shift for each cluster

These findings are supported by the results of a linear mixed effects model (lmer). The model predicted the overall RSD value with task, edge (left edge, c-center, and right edge), and cluster as fixed effects and subject as a random effect. The result of this analysis is that clusters containing /k/ were overall more variable than those containing /s/ or /t/, as is indicated in the results for all three tasks. Homorganic clusters /st/, /tr/, /tl/ along with /sp/ were overall the least variable with /sk/ being the most variable, a finding likewise supported across all three tasks. Finally, following vowel also had an effect on stability and rightward shift, where coronal clusters showed greater stability when followed by an /i/ than when followed by an /o/ or /a/ and dorsal clusters showed greater stability when followed by an /o/ or /a/.

4.3.2 Gestural coordination summary

There is a clear distinction in the results between the initial syllables and the medial syllables, where the initial syllables are consistent with previous studies on gestural coordination. In the initial syllables, the c-center tended to be the least variable across the RSD and STD measures, and likewise there was rightward shift of the right edge in initial syllables. The results of the medial syllables showed much more variability in the c-center, where for both RSD and STD, the c-center was either the most variable or patterned with the most variable margin. Therefore, coordinative stability between initial syllables and medial syllables is distinct.

Comparing stressed and unstressed syllables, the results are more ambiguous. While the rightward shift results do suggest syllabification of the cluster as the onset; the c-center results suggest that the c-center is not a stable predictor for the timing of the following vowel in medial syllables. Furthermore, the differences between medial stressed syllables and medial unstressed syllables was small, which suggests that word position may affect syllabification and coordination more than stress. Where there were differences between the two medial stress conditions, the effects did not pattern as predicted. While the medial unstressed syllables were predicted to be the most variable, the c-centering effect was stronger in the medial unstressed syllables than in the medial stressed syllables. However, nucleus duration has been shown to affect stability heuristics (Shaw et al., 2011) and thus, the surprising outcome may be due to the reliability of these measures rather than an underlying pattern of coordination in the data.

Analyzing differences in coordination between cluster sequences provides a number of additional key insights. First, there is a clear distinction in c-center and rightward shift across segments between the initial and medial tasks, consistent with the overall results. While the role of stress remains more ambiguous in these results, there are a few segmental patterns that may indicate similarities between stress conditions. Although the comparison of c-center stability across tasks suggests that initial stressed syllables are most similar to the medial unstressed syllables, comparison of coordination patterns for the varying clusters suggests that syllabification may be more comparable between the initial stressed syllables and the medial stressed syllables. This conclusion is particularly clear when considered alongside the compensatory shortening results, for which the initial stressed. and medial stressed syllables were more comparable.

In addition, there are more fine-grained distinctions between cluster types that provide insight into the pressures that affect syllabification. First, stop+/r/ and stop+/l/ clusters are distinct across the tasks. Furthermore, there is a distinction in place of articulation across the three tasks, where homorganic clusters are the most stable and and clusters containing /k/ are the least stable. Likewise, the clusters were more stable in the presence of a vowel that corresponds in backness, where coronals clusters were most stable in the presence of a front vowel and dorsal clusters were most stable in the presence of a back vowel. These results indicate that there are intrinsic pressures on syllabification and stability; there is a degree of caution in interpreting these results, however, given the indeterminacy of the overall results. Nevertheless, the results indicate that intrinsic differences in gestural composition affect the likelihood for a given sequence to syllabify within the same syllable or across syllables. Sequences of segments that are in close gestural proximity are more likely to syllabify together than gestures coordinated across further distances. While this causes some sequences to consistently syllabify together, for other sequences, proximity to both the preceding and following vowel may compete for coordination resulting in greater variability or ambisyllabicity, where the segment is coordinated with both the preceding and following vowel. Further investigation of these questions is needed in a framework that can investigate more fine-grained distinctions in syllabification.

Ultimately, the gestural results clearly demonstrate that initial sequences are coordinated distinctly from medial sequences. While the initial syllables show clear patterns of global timing, the results for medial syllables are ambiguous. Because of the ambiguity of the medial syllable results, the effect of stress on coordination is far less clear. Although the clearest explanation for the medial syllables is that syllabification in medial syllables is variable, the gestural results do not explicitly point to this explanation. Instead, c-center and rightward shift measures provide a largely binary conclusion, where results are either consistent with global timing or they are not, without further disambiguation of why the results differ from one another. Thus, while variability seems like the most logical conclusion, more fine-grained tools are needed to test this prediction.

4.4 Chapter summary and discussion

Overall, the results of this chapter provide clear insights on the differences in syllabification between word-initial and word-medial clusters. Across all three methods of analysis, compensatory shortening, c-centering, and shift, the results for the initial stressed task condition were most consistent with previous studies, showing compensatory shortening, c-centering, and rightward shift. These results were in clear contrast to the medial conditions. First, the compensatory shortening results were not consistent with any single set of syllabification predictions for the medial conditions. In particular, all syllabification schemata predict that if there is compensatory shortening in the stressed syllable, there should not be compensatory shortening in the unstressed syllable or visa versa. Instead, for the medial conditions, there is compensatory shortening in both the stressed and the unstressed syllable. This outcome suggests that syllabification is variable in both medial conditions, resulting in some degree of compensatory shortening in both syllables. Second, initial stressed syllables had a clear effect of c-centering; however, in the medial conditions, the c-center effect was weak. Finally, for rightward shift, the majority of clusters showed a rightward shift of the right boundary, where the results were mixed for the medial conditions, with some clusters showing rightward shift, but others showing leftward shift of the right boundary. Together, these results demonstrate that coordination in word-medial contexts differs for word-initial contexts. Furthermore, the analysis most consistent with these results is that word-medial sequences are variably syllabified, resulting in compensatory shortening in both syllables of the token and weakening the gestural coordinative stability.

Despite the clarity of the effect of word position, the effect of stress is ambiguous, particularly for the gestural analyses. While there were subtle differences in c-center stability and rightward shift between the medial stressed and medial unstressed conditions, it is difficult to determine whether the weak effects are indicative of coordination patterns in the sequences, or just the outcome of the overall weakened stability results. Nevertheless, inasmuch as the weak outcomes can be interpreted, syllabification patterns in the medial stressed syllables were most similar to the initial stressed syllables. This trend is most apparent when comparing differences between clusters, where clusters in the medial stressed condition showed similar patterns in rightward shift to the initial stressed condition, but clusters in the medial unstressed condition were distinct. The clearest differentiation between the medial conditions comes from the results of the compensatory shortening analysis. Both medial conditions were consistent with the analysis that medial syllabification is variabile; however, vowel durations in the medial unstressed condition were particularly variable, suggesting that there is additional variability in this condition compared to the medial stressed condition. Thus, the analysis most consistent with the effect of stress on syllabification is that word-medial syllables are variably syllabified and variability increases for medial unstressed syllables.

In addition to the overall syllabification patterns, the differences in coordination between clusters provide some support for both perceptual and intrinsic pressures on syllabification, supporting the licensing by cue and Frame-Content Model. For instance, the distinction between stop+/r/ and stop+/l/ evident in the data may be consistent with a difference in coordination due to a difference in perceptibility, consistent with the findings of Katz (2012) who finds greater compensatory shortening in the presence of rhotics than laterals and advocates for a model of compensatory shortening and coordination that incorporates duration and perception. Furthermore, while there may be articulatory motivations for the differences between /l/ and /r/ that additionally motivate the distinct patterns, an articulatory based explanation for these results might predict a greater difference between /tr/ and /tl/ clusters as opposed to the other corresponding stop+liquid clusters, which was less evident in the results, but homorganic clusters were found to have less variability than other clusters overall.

The results likewise provide support for intrinsic pressures on syllabification and the Frame-Content Model. For instance, clusters containing /k/ were found the be the most variable overall, with /kr/ clusters not showing rightward shift across any conditions. Likewise, sequences where consonants and vowels correspond in backness were more stable and more likely to show rightward shift and c-centering, consistent with the findings of Vallée et al. (2009). These findings cannot be captured by a model of syllabification that includes only perception as the basis for syllabification and indicates that there are both articulatory and perception pressures on syllabification and coordination.

Ultimately, these results highlight the difficulty in analyzing coordination where syllabification may be variable because none of the methods employed in this chapter have a clear mechanism for analyzing variability. Thus, variability must be inferred from a null or ambiguous result, rather than be directly signaled in the results. A consequence of this is that there is no clear way to test the variability hypothesis. Optimally, the data should be subset to where syllabification can be more accurately predicted and the tests should be re-run to diagnose whether the subsetted results pattern more consistently with initial syllables and previous literature; however, neither theoretical models of syllabification, empirical results, or analytical tools can accurately predict word-medial syllabification patterns with any degree of certainty. Furthermore, the existing tools do not provide a clear way to analyze what the underlying structure of these sequences is. While this chapter can draw the conclusion that medial syllabification is more variable, there are multiple hypotheses of what that variability might include. For instance, syllables may still be exhaustively syllabified but vary in where the segments are divided; medial segments may be ambisyllabic showing properties of coordination with both the preceding and following vowels; there may be no clear evidence of syllable structure whatsoever word-medially, with segments being timed locally rather than globally; or there may be some combination of these patterns word-medially. Existing tools simply do not provide the necessary mechanisms to disambiguate these possibilities, leaving a number of remaining questions. Thus, these results highlight the need for more granular tools for analyzing coordination. Chapter 5 proposes analyzing jaw movement to fill this gap.

Chapter 5

Analyzing jaw phase in interpreting patterns of coordination

The Coupled Oscillator Model proposes that syllable coordination is due to pressure in motor planning for overlap between segmental gestures. However, an alternative articulatory perspective on syllabification is the effect of jaw articulation influences the timing and coordination of segmental gestures. MacNeilage (1998) proposes the Frame-Content Model, according to which modulation of the jaw corresponds with syllable structure, referred to as the *frame*, which is filled with segmental gestures, referred to as the *content*. MacNeilage (1998) argues that the role of the jaw in early speech development as well as the significance of the jaw in non-speech tasks, like eating, has shaped the speech evolution of syllables, resulting in the strong typological trends in syllable structure observed cross-linguistically. Support for the Frame-Content Model is widespread. For instance, Green and Brock (2002) shows that jaw oscillation is indeed a critical component of early child language development and demonstrates that jaw coordination develops earlier than motor control for other articulators, like the lips. Likewise, Hickok (2014) argues that the syllable plays an active role in speech production, where auditory feedback corresponds with syllable-sized production targets and somatosensory feedback corresponds with segmental and subsegmental targets. Finally, typological patterns in segment co-occurrence support a jaw-based organization of syllables, as was shown in Chapter 1, where coronal consonants are more likely to occur with front vowels, labial consonants to occur with central vowels, and velar consonants to occur with back vowels. Likewise, Vallée et al. (2009) finds that the first C in either CVC or CV.CV sequences is more likely to be bilabial and the second consonant is more likely to be coronal. The authors explain that even in a sequence of two syllables, if consonants are sequenced as bilabial and coronal, then they can be coordinated with a single jaw movement, whereas a coronal + bilabial sequence requires two jaw movements. The findings of the present study are consistent with the typological patterns observed by Vallée et al. (2009), where coronal consonants were more stably timed when preceding /i/ and dorsal consonants were more stably timed when preceding /o/ and /a/.

Rather than coordination resulting from overlap between segmental articulators, a jaw based-model of coordination predicts that timing stability is the result of coupling between segmental content and jaw cycle. Coupling with jaw phase also likely corresponds to synchronous planning between segmental content and jaw content; however, jaw coupling adds a mechanical component to coordinative stability, where segments that are coordinated within the same jaw cycle are predicted to be more tightly coordinated than segments coordinated across jaw cycles. This prediction is consistent with the results of studies on syllable coordination. For instance, Goldstein et al. (2007) find that front-to-back clusters in Georgian are more stably timed than back-to-front clusters. The authors argue that this may be due to perceptual factors. However, another explanation is that front-to-back clusters can be articulated within the same jaw gesture, whereas back-to-front clusters require multiple jaw gestures. Furthermore, studies on coordination have predominantly focused on word-initial clusters or clusters in prominent positions, where word-initial clusters are particularly likely to be coordinated within the same jaw phase. Therefore, to test the predictions of this hypothesis, it is necessary to analyze sequences where syllabification and coordination are likely to vary, e.g., word-medial contexts. Thus, this study will test the predictions of the Frame-Content Model by analyzing jaw coordination across the three task conditions: initial stressed, medial stressed syllables, and medial unstressed syllables.

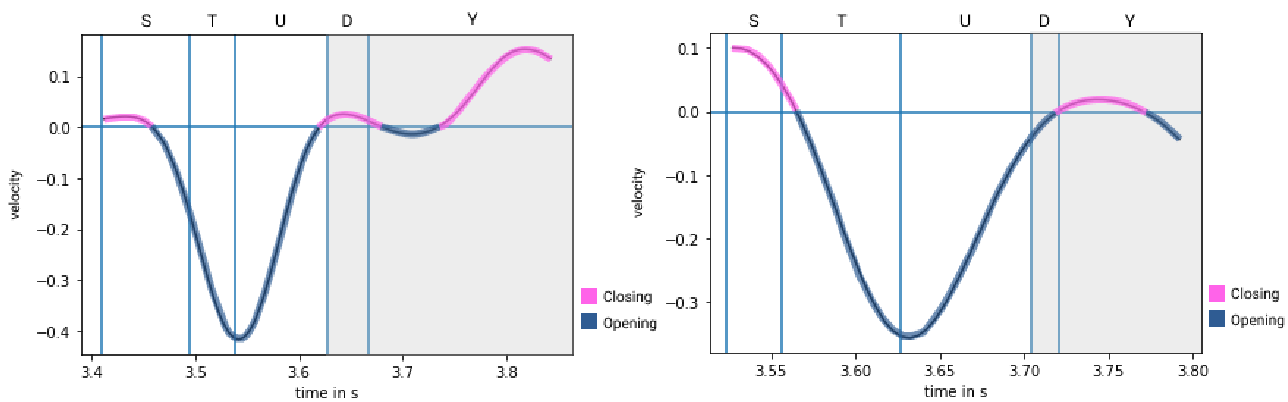
While the Frame-Content Model predicts a different underlying mechanism that motivates patterns of coordination, the predicted possible syllabifications under the Frame-Content Model resemble those discussed in Chapter 4. Namely, either onset maximization, stress, or a combination of these schemata are all possible syllabifications. However, whereas the Coupled Oscillator Model predicts that all syllables should adhere to the Onset Maximization Principle, the Frame-Content Model leaves open the possibility that syllabification may vary and may include coordination across syllables, i.e., ambisyllabicity. In particular, the Frame-Content model predicts that intrinsic pressures like place of articulation will influence syllabification. The results of Chapter 4 demonstrate that there is a difference in coordination depending on whether the sequence occurs word-initially or word-medially, suggesting that syllabification is variable in word-medial contexts; thus, this chapter will focus on how coordination varies word-medially and whether word-medial stressed and unstressed syllables can be clearly distinguished.

This study is the first to analyze jaw movement as a method for capturing word-medial syllabification patterns, and as such is still in the early stages of development; however, the landmarks identified in this study provide a foundation for analyzing jaw coordination that is able to capture differences between the syllabification paradigms analyzed in this study: initial stressed, word-medial stressed and word-medial unstressed. This chapter will present the methodology used to analyze jaw coordination, the results of the study, a discussion of the methodology and future directions for developing the methodology, and finally, a discussion of the results and implications of this study.

5.1 Methods

Like the methodology used for defining segmental landmarks, jaw landmarks were identified using both the position of the jaw over time and the velocity of jaw movement over time. However, there are two main distinctions in analyzing jaw movement. First, while the segmental gestures correspond with segment sized units with multiple steady states occurring throughout a given word, jaw oscillation consists of approximately one to two opening phases and one closing phase per word for the tokens in this study, which consist of (C)CVC and V(C)CV shapes. Jaw opening phases broadly correlate with vowels and jaw closing phases broadly correlate with consonants. Second, while the central task in identifying segmental landmarks is to identify the steady state of the target, the central task of identifying jaw coordination landmarks is to identify the alignment of the jaw trajectory with the segmental content.

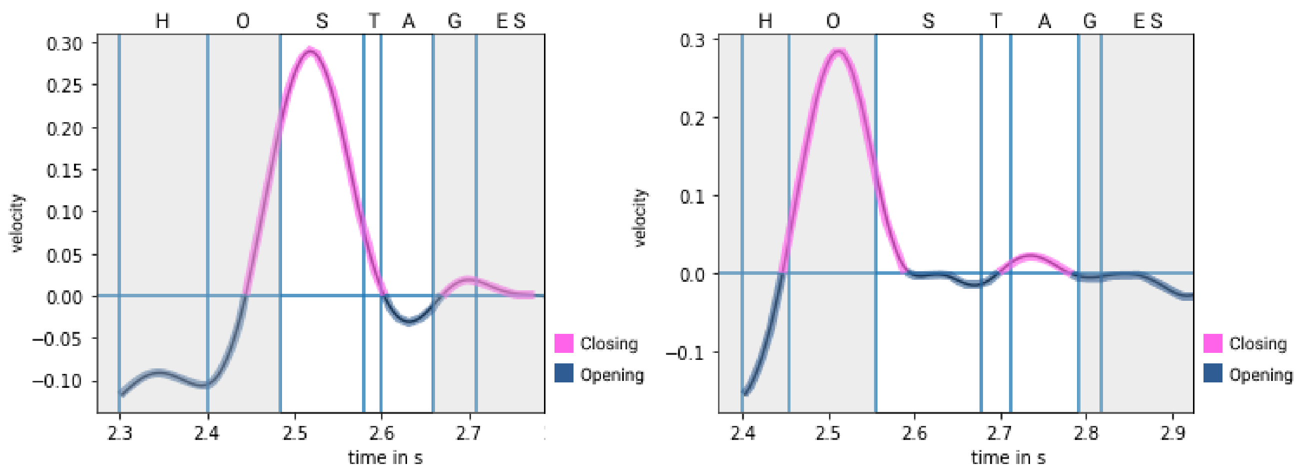
Study one provided the basis for the differences in syllabification between tokens, which illustrates key differences in jaw coordination that suggest differences in syllabification. Figure 5.1 provides two examples of jaw phase for the word 'study' along with the acoustic boundaries for each segment in the target utterance. Overall, the shape of the trajectory during the onset and stressed vowel are largely the same, with a low-speed upward trajectory during the /s/ and a high-speed downward trajectory that progresses across the stop+vowel sequence. There are subtle differences between the two sequences; for instance, the change in direction into the downward trajectory begins during the /s/ in Figure 5.1a, but begins during the /t/ in Figure 5.1b; however, there are clear similarities in the shape of the two trajectories.



(a) study (b) study

Figure 5.1: Jaw trajectory plotted with acoustic segmental boundaries for two utterances of 'study' in the sentences A. 'Biologists use radioactive isotopes to study microorganisms.' and B. 'That diagram makes sense only after much study.'

Figure 5.2 provides two contrasting examples for the word 'hostages'. Not only do these examples contrast from the productions of 'study', but these utterances bear striking differences from one another. Namely, in Figure 5.12a, the peak closing velocity occurs during the /s/, while in Figure 5.12b the peak closing velocity occurs in the /a/ prior to the consonant cluster. Furthermore, the change in direction from closure phase to opening phase occurs at the end of the /t/ in Figure 5.12a, but occurs during the beginning of the /s/ in Figure 5.12b.



(a) hos.tages (b) ho.stages

Figure 5.2: Jaw trajectory plotted with acoustic segmental boundaries for two utterances of 'hostages' in the sentence 'Special task forces rescue hostages from kidnappers.'

The difference in jaw coordination in these two productions of 'hostages' indicates differences in coordination between the two tokens, which likewise suggests a difference in syllabification. For the token in 5.12b, the /st/ cluster occurs within the same opening phase, and as such is syllabified as *ho.stages*, but the cluster in 5.12a occurs across the opening and closure phase and is syllabified as *hos.tages*. This is the strongest indication by far that, in fact, there are simply two different ways to syllabify 'hostages', which likely accounts for why analyses of word-medial coordination produced weak effects.

The overall patterns in peak closing velocity and initiation of the opening phase in jaw coordination persisted throughout the utterances containing /st/ that were analyzed in study one, with over 500 total utterances. Overall, syllabification of medial /st/ sequences varied in whether the peak closing velocity occurred prior to the start of the cluster and whether jaw opening initiated during the /s/ or the /t/. While medial stressed tokens were more likely to have the peak closure velocity before the start of the cluster, stress did not determine jaw closure, with other factors like the number of consonant segments in the sequence, i.e., VCCV or VCCCV, and surrounding vowel quality also playing a role in jaw alignment, e.g., 'prestige' was more likely to syllabify as *pres.tige*, while 'instead' was more likely to syllabify as *in.stead*.

Thus, based on the results of study one, two main landmarks were identified for exploring jaw phase coordination: whether the the maximum closing velocity occurred before the start of the C1 and whether the opening phase was initiated in the C1 of the cluster. Maximum closure velocity was identified using the point of maximum speed during the utterance. The initiation of jaw closing was calculated by determining whether there was a change in trajectory from closing to opening during the C1 of the cluster. In addition, whether a given token met all three conditions was also noted. To simplify the analysis for this initial study, each of these variables was binary for the quantitative portion of this study, meaning that, for example, a segment was labeled as containing the maximum peak velocity if the point of maximum closure velocity occurred at any point within the segment, regardless of when the maximum peak occurred within the segment; however, to gain a more holistic sense of the data, velocity traces were also analyzed individually and labeled not only for the presence of these landmarks, but the approximate timing of these landmarks. The holistic results and quantitative results are presented in the following section.

5.2 Results

Analyzing the landmarks identified in the methods section, peak closure velocity and the initiation of the downward trajectory, as well as the co-occurrence of those landmarks, a number of trends become apparent. Crucially, the alignment of the landmarks differed across the three tasks. First, the coordination of initial stressed tokens was highly consistent across utterances. Second, both word-medial tasks were more variable with regard to the two landmarks. Third, a key aspect of that variability was that in the medial unstressed condition, it was more likely that tokens would have the initiation of the jaw opening phase near the boundary of C1 and C2. In other words, the C1 and C2 did not occur within the same jaw phase; however, this boundary occurred earlier in both medial stressed tokens and initial stressed tokens, with more variability in the former.

5.2.1 Holistic results

5.2.1.1 initial stressed condition

Jaw traces for initial stressed tokens were highly consistent across clusters and participants. Overall, the traces can be described as sharing three properties: a) tokens consisted of a peak closure in the vowel preceding C1 or at the V-C1 boundary; b) tokens consisted of a change in direction signaling the initiation of the jaw opening phase in the first half of the C1; and c) peak opening velocity occurred during the C2. An example is shown in 5.3.

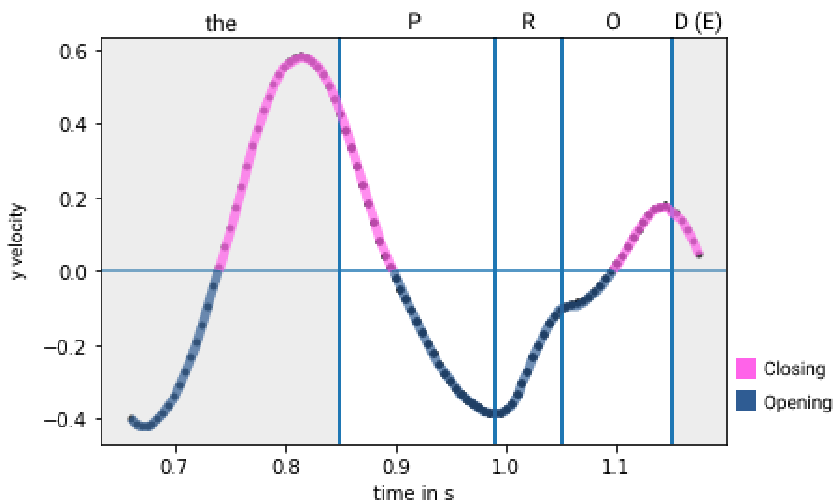


Figure 5.3: Jaw velocity trace plotted with acoustic segmental boundaries for the token ‘prode’, /proʊd/, reflecting a typical jaw pattern for CCVC tokens in the initial stressed condition

For singletons, both initiation of opening and peak opening typically occurred within the same segment. Figure 5.4 demonstrates a typical jaw trace in the initial stressed condition for a singleton onset.

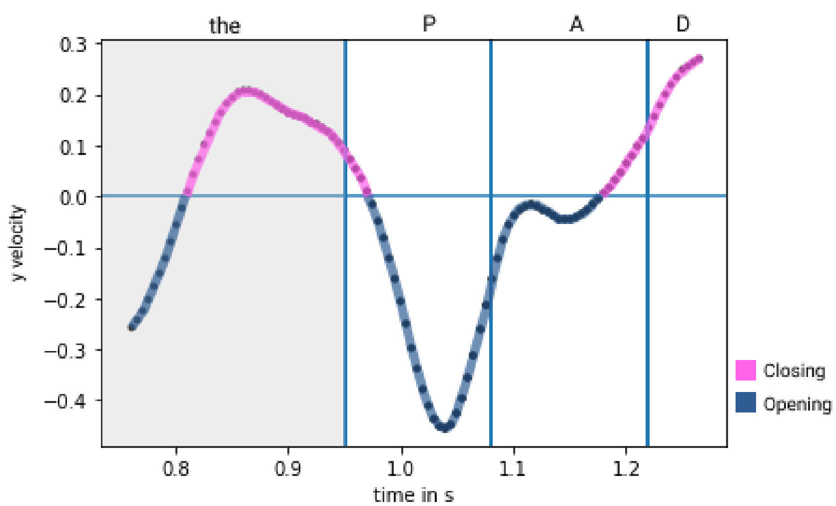
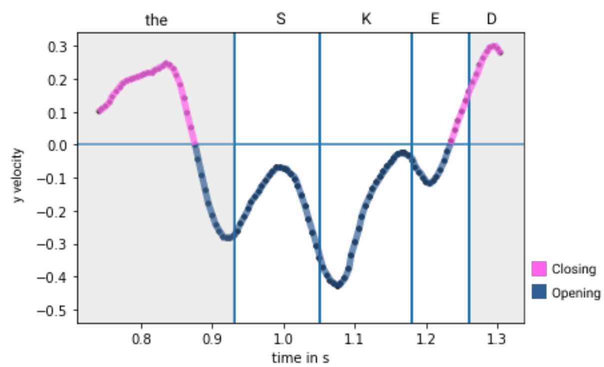


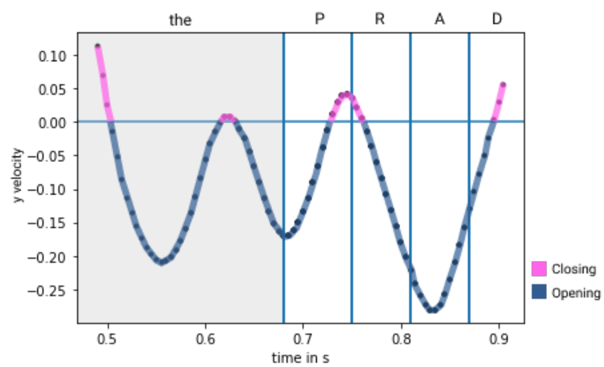
Figure 5.4: Jaw velocity trace plotted with acoustic segmental boundaries for the token ‘pad’, /pæd/, reflecting a typical jaw pattern for CVC tokens in the initial stressed condition

Overall, both clusters and singletons followed the patterns illustrated by these examples, with a similar pattern across both participants and cluster types.

Despite the consistency across utterances, there were a few notable points of departure from the main pattern of jaw cycle shape and alignment. First, two of the participants showed greater modulation of the jaw than other participants across all utterances. Figures 5.5 demonstrate traces for these subjects.



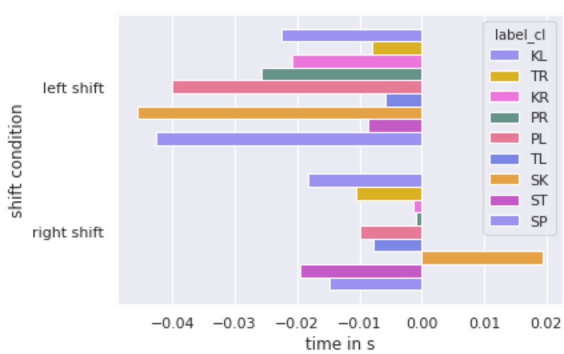
(a) Subject A, /skɛd/



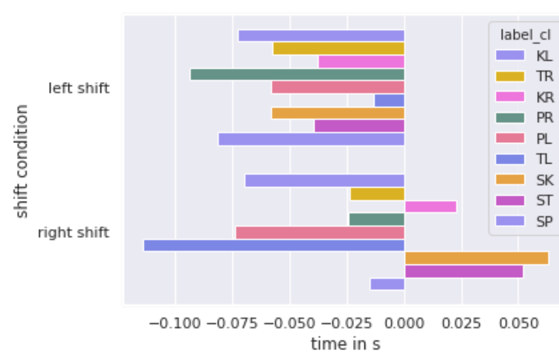
(b) Subject B

Figure 5.5: Jaw velocity trace plotted with acoustic segmental boundaries for the token ‘sked’, /skɛd/ produced by Subject A and ‘prad’, /præd/ produced by Subject B

While the trace for Subject A contains only a single downward trajectory, there is a decrease in speed that corresponds with each segment in the utterance. A similar pattern is observed for Subject B, but unlike Subject A, Subject B has a change in direction across the /pr/ cluster. The differences in jaw coordination for these two subjects may suggest less freedom of movement of the tongue from the jaw, resulting in greater coupling between the jaw and articulator movements. In addition, these subjects tended to show less rightward shift than average, as shown in Figure 5.6.



(a) Subject A



(b) Subject B

Figure 5.6: Rightward and leftward shift results for Subject A and B for the initial stressed task condition corresponding with the figures for Subject A and Subject B above, who exhibited a greater degree of jaw oscillation compared to other participants.

These results are consistent with the hypothesis that segments that occur across jaw cycles will show less coordinative stability. However, while increased jaw movement may drive decreased gestural stability across the syllable, more investigation is needed to determine whether these factors are indeed correlated.

Another notable departure from the common pattern of jaw oscillation is the 'stair-stepped' example in Figure 5.7. In these utterances, there is a plateau indicating a constant speed that tends to be near zero, suggesting a slow and constant speed of jaw movement.

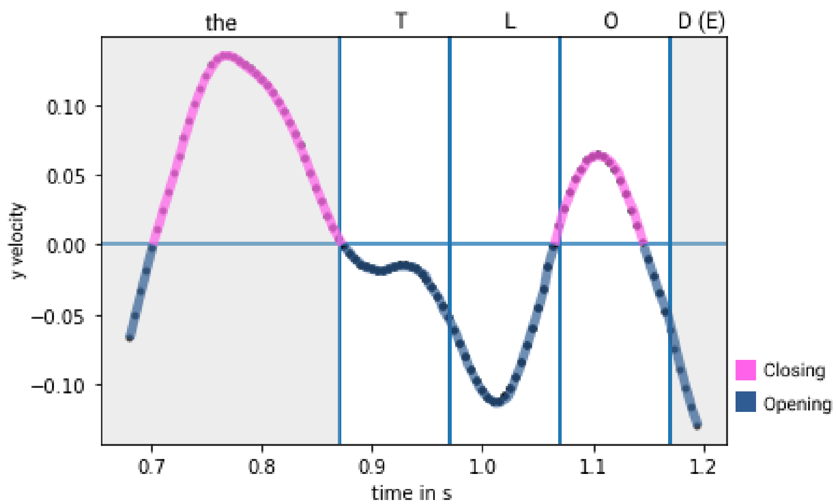


Figure 5.7: Jaw velocity trace plotted with acoustic segmental boundaries for the token 'tlode', /tloʊd/, reflecting a 'stair-stepped' jaw pattern that often occurs in tokens containing /l/ in the initial stressed condition

This pattern occurred in participants whose overall jaw movement fit within the typical framework of jaw movement described above; however, this pattern of movement was typical in tokens containing /l/. Again, while more investigation is needed into such patterns, this shape of movement may be consistent with the tighter coupling patterns described for /l/ by studies such as Goldstein et al. (2009). In addition to the patterns of variation described here, another common pattern was that tokens containing high vowels, e.g., /i/, had lower magnitudes of movement and in particular, the peak downward velocity was lower than for segments like /a/. This is unsurprising because the jaw opening for high vowels is expected to be less than low vowels.

These patterns of variation are important for future studies as they are likely to provide insights on ambisyllabicity and the influence of intrinsic pressures on jaw movement; however, the most notable finding in analyzing the initial stressed sequences is the consistency in the patterns of jaw oscillation alignment across the initial stressed condition. Crucially, across the majority of utterances in the initial stressed condition, both segments in the onset cluster occurred during the same jaw opening cycle. These results are consistent first with the hypothesis that clusters that occur within the same jaw cycle will show tighter coordination patterns, where the results of

Chapter 4 show stronger coordinative results in the initial stressed condition. In addition, these results support the conclusions of the Frame-Content Model, where jaw opening corresponds with the syllable onset and jaw closing corresponds with the coda segment.

5.2.1.2 Medial stressed and unstressed syllables

The results for the medial stressed and medial unstressed tokens were more variable, with tokens showing both more variability in the alignment of the landmarks with the segmental content and more overall movement of the jaw. In the medial stressed condition, about a quarter of tokens had overall greater jaw movement across subjects. In the medial unstressed condition, about a third of utterances had overall more jaw movement. Although this increased jaw movement may be relevant in determining what distinguishes word-medial coordination from initial coordination, this study focused on the alignment of the main landmarks identified in the first study, with future studies returning to the more variable patterns in jaw oscillation.

In the medial stressed condition, although there was greater variability than in the initial stressed condition, about half of utterances showed the same basic landmarks and patterns of alignment observed in the initial stressed condition. In other words, about half of utterances contained a peak closing velocity prior to the C1 of the onset cluster and an initiation of the jaw opening phase during the C1 of the cluster that was maintained during the C2 of the cluster. Figure 5.8 provides an example.

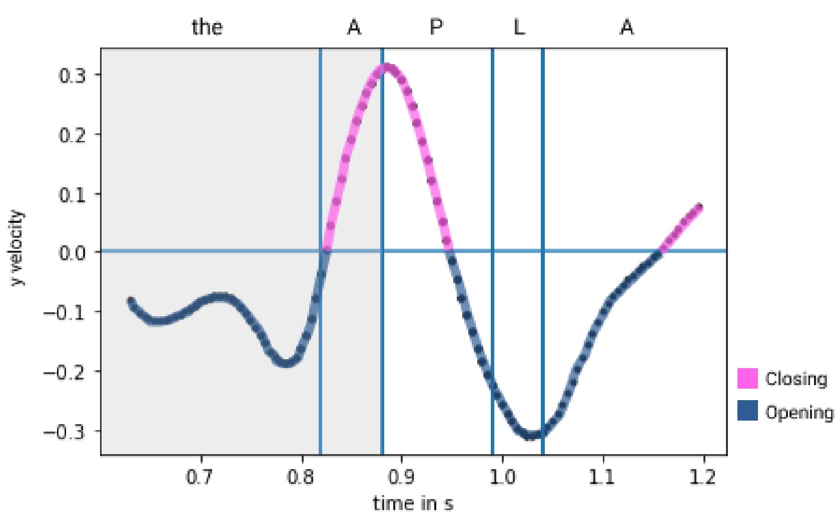


Figure 5.8: Jaw velocity trace plotted with acoustic segmental boundaries for the token ‘apla’, /əpla/, reflecting a typical jaw pattern in the medial stressed condition

These results suggest that these utterances were coordinated comparably to word-initial stressed clusters, with both segments of the cluster produced within the same jaw phase. Utterances varied subtly in the alignment of the jaw phase, where Figure 5.8 shows a later alignment of jaw movement with the segmental content. These slight differences may be consistent with an analysis like ambisyllabicity, where the C1 segment is coordinated with both the preceding and following syllables; however, further study will be necessary to determine the significance of the variability in alignment. For the purposes of this study, utterances where the peak closing velocity occurred prior to the start of the C1 and where the initiation of the opening phase occurred during the early portion of the C1 were analyzed as having the syllabification: *a.pla*.

Although about half of the utterances were consistent with the initial stressed results, in about a quarter of the utterances, the alignment of the peak closing velocity and initiation of the peak opening velocity occurred much later than was typical in the initial stressed condition. In these tokens, the max closing velocity occurred during the C1 of the onset and onset of the the opening gesture occurred at the boundary of the C1 and C2 consonants, as is shown in Figure 5.9.

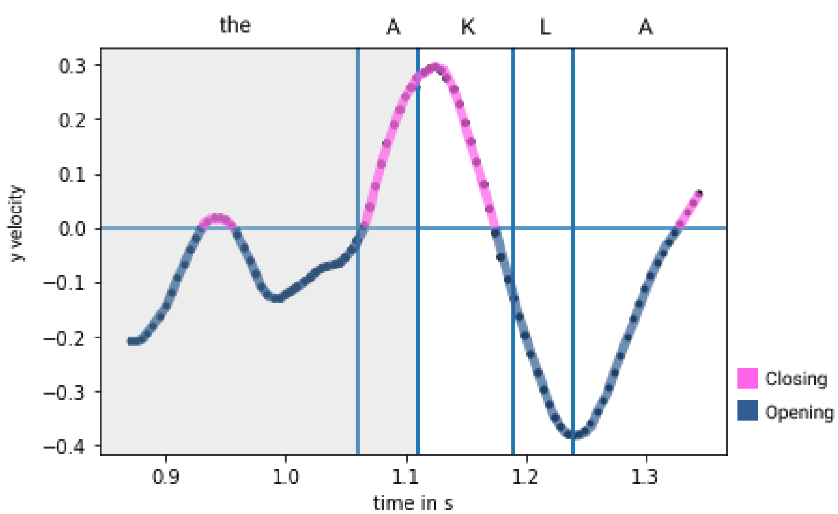


Figure 5.9: Jaw velocity trace plotted with acoustic segmental boundaries for the token ‘akla’, /əkla/, reflecting a late-aligned jaw pattern in the medial stressed condition

The alignment of the major landmarks in these tokens suggests that the consonants in the onset cluster were coordinated across a jaw closing and a jaw opening cycle, consistent with the syllabification pattern *ak.la*.

Finally, medial unstressed tokens were the most variable overall, with only about a third of tokens sharing the common trajectory found in the initial stressed condition. An example of a token where the peak velocity occurs before the C1 and the initiation of the opening happens in the early portion of the C1 is shown in Figure 5.10.

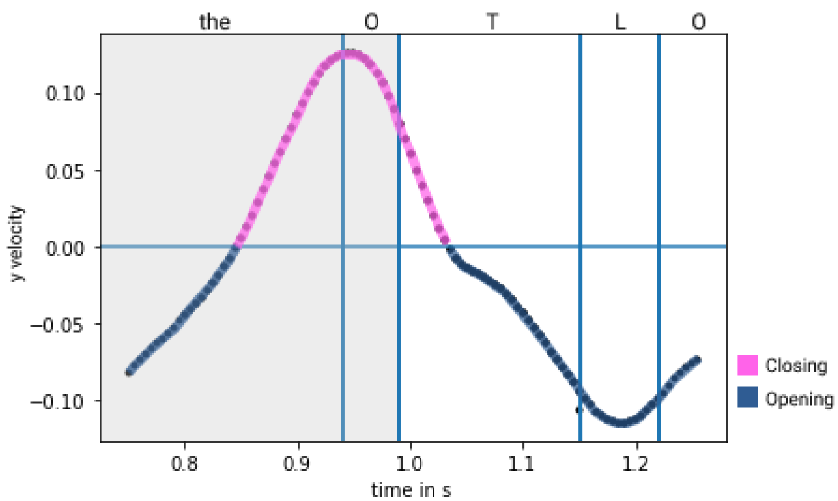


Figure 5.10: Jaw velocity trace plotted with acoustic segmental boundaries for the token 'otlo', /oʊtloʊ/, reflecting a typical jaw pattern in the medial unstressed condition

These tokens were analyzed as sharing the syllabification *o.tlo* and the pattern of coordinative stability in these utterances is predicted to be the same in these tokens as tokens with similar jaw alignment in the initial stressed task condition and the medial stressed task condition.

Instead, the remaining third of the medial unstressed tokens showed the later jaw alignment pattern, where the max peak closing velocity occurred during C1 and the initiation of downward movement occurred late in C1 or at the boundary of C1 and C2, as shown in Figure 5.11.

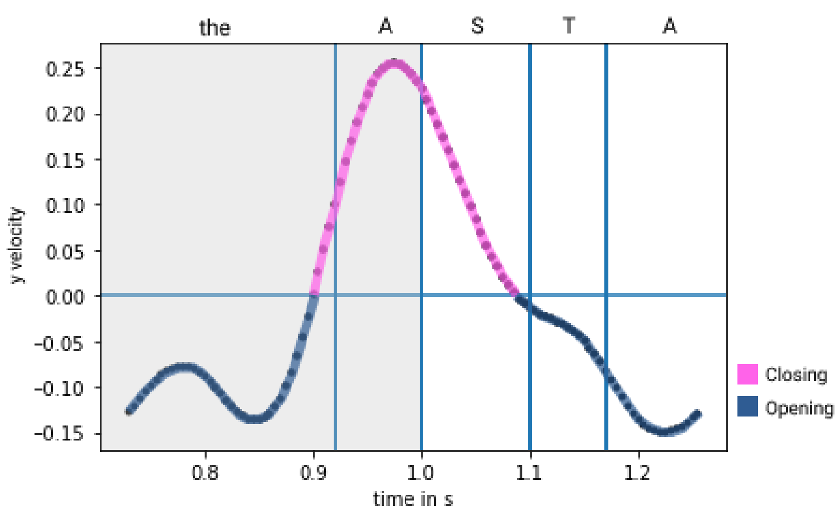


Figure 5.11: Jaw velocity trace plotted with acoustic segmental boundaries for the token 'asta', /astə/, reflecting a late-aligned jaw pattern in the medial unstressed condition

These utterances were analyzed as having the syllabification pattern *as.ta* and are expected to show less coordinative stability during the cluster, but showing a similar patterns in coordination to other onset-coda syllabification tokens. Furthermore, for medial stressed tokens, while the onset maximization coordinative pattern was about twice as likely as the later-aligned onset-coda pattern, in the medial unstressed tokens, the onset maximization pattern and the onset-coda pattern were about equally as likely.

Overall, the holistic results demonstrate distinct patterns across the three task conditions, with a gradient likelihood to show the onset maximization coordination pattern. While the medial unstressed condition was relatively consistent across tokens, both medial contexts showed variability in coordination, with the most variability in the medial unstressed condition. These results are consistent with the findings of Chapter 4; however, unlike Chapter 4, the jaw results indicate two main points of difference between the medial task conditions. First, unstressed syllables were overall more variable than the medial stressed syllables. Second, unstressed syllables were more likely to exhibit a jaw coordination pattern where both peak closing velocity and the initiation of the opening gesture happen later than in the stressed syllable conditions. Specifically, rather than corresponding with the preceding vowel segment and C1, the peak velocity and opening gesture correspond with the C1 and C2 segments. Thus, these results suggest that the C1 of the medial consonant sequence is more likely to syllabify as the coda of the first stressed syllable in the medial unstressed condition. The following section will implement a quantitative analysis of this methodology to demonstrate similar findings.

5.2.2 Quantitative results

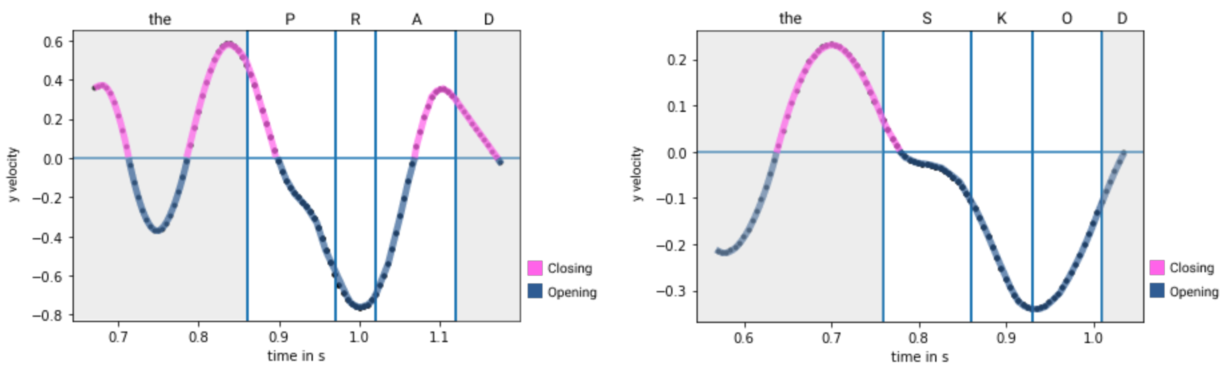
The quantitative study identified whether tokens were consistent with the observations discussed in the holistic results, focusing on peak closing velocity and the initiation of the downward trajectory. The quantitative results successfully demonstrate the gradient likelihood of each task to syllabify with both the C1 and C2 in the onset cluster. However, the binary method of assessing the alignment of landmarks proved too strict, resulting in overall modest results. The results are shown in Table 5.1, which provides the percentages of utterances that met one of the three conditions. Condition A required that the start of the downward trajectory be initiated during C1. Condition B required that the max peak in the closing trajectory occur prior to the start of the onset of C1. Finally, Condition C required that Conditions A and B both be met. While only a small proportion of overall tokens met both conditions, the difference between tasks was still significant.

	Initial Stressed	**Medial Stressed**	**Medial Unstressed**	**Significance**
Condition A	79%	81%	82%	
Condition B	52%	42%	35%	***
Condition C	27%	23%	17%	**

Table 5.1: Percentage of utterances that meet conditions for a given jaw landmark. Condition A: jaw opening begins in C1; Condition B: Maximum closing trajectory occurs prior to the start of C1; Condition C: Conditions A and B are met.

The majority of utterances met Condition A; while there were slight differences between tasks, the difference between tasks was not significant. This pattern appears much more common in the quantitative results than it appeared in the holistic results, predominantly because most of the tokens began the downward trajectory during C1, but crucially, tokens in the medial conditions were more likely for the downward trajectory to begin in the latter half of the segment, often at the boarder of the C1 and C2 boarder, and thus, the binary method was not sensitive enough to capture this distinction. While a smaller proportion of tokens met Condition B compared to Condition A, Condition B proved the most significant in identifying differences between tasks, where all three tasks differed significantly in the point of peak closure occurrence. Furthermore, Condition B successfully captures the gradient likelihood of the onset maximization syllabification observed in the holistic results. Condition C likewise shows this gradience, but with an even smaller proportion of tokens meeting both Conditions A and B, likely due to the binary method of determining whether a token met each condition. Nevertheless, both the alignment of the peak closure velocity and the initiation of the opening trajectory were significant in distinguishing between the three task conditions, suggesting that the alignment of jaw opening and closing landmarks are relevant in influencing coordination and syllabification.

Looking at the outcomes of each condition, the examples provided in Figure 5.12 demonstrate jaw velocity traces for utterances that met Condition C in the initial stressed condition. As shown in the figures, both utterances contain a peak closure velocity prior to the start of C1 and an initiation of the opening gesture in the early part of C1.

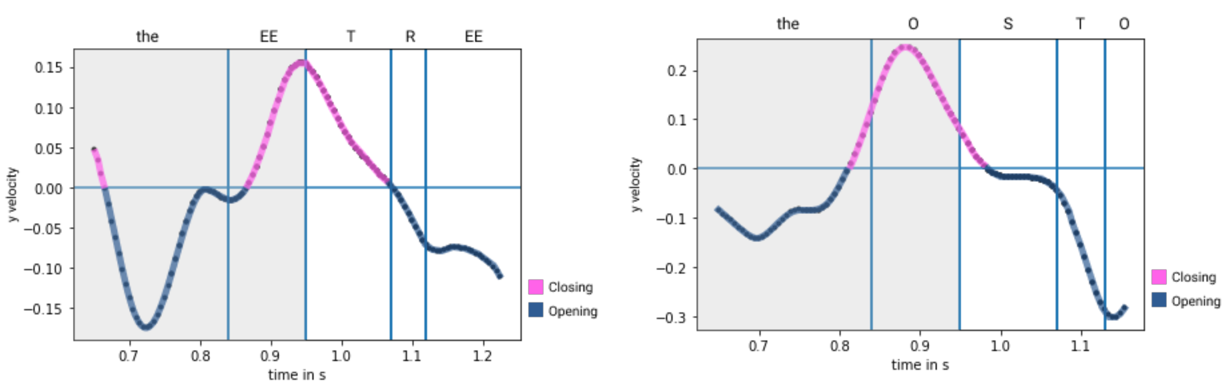


(a) /pɹæd/ (b) /skad/

Figure 5.12: Jaw trajectories plotted with acoustic segmental boundaries for two utterances of the targets 'prad', /pɹæd/, and 'skod', /skad/, produced in the carrier phrase 'Say the [x] again' in the initial stressed condition

These figures illustrate first, that the quantitative method employed in this study successfully identified velocity traces consistent with the results of the holistic analysis. Second, these results demonstrate that tokens in the initial stressed condition were relatively consistent in the alignment of the peak closure and initiation of opening landmarks.

Figure 5.13 provides examples of utterances that met Condition C in the medial stressed task condition. In particular, in both figures, the max peak occurs prior to the start of the C1 and the downward trajectory occurs prior to the start of C2.

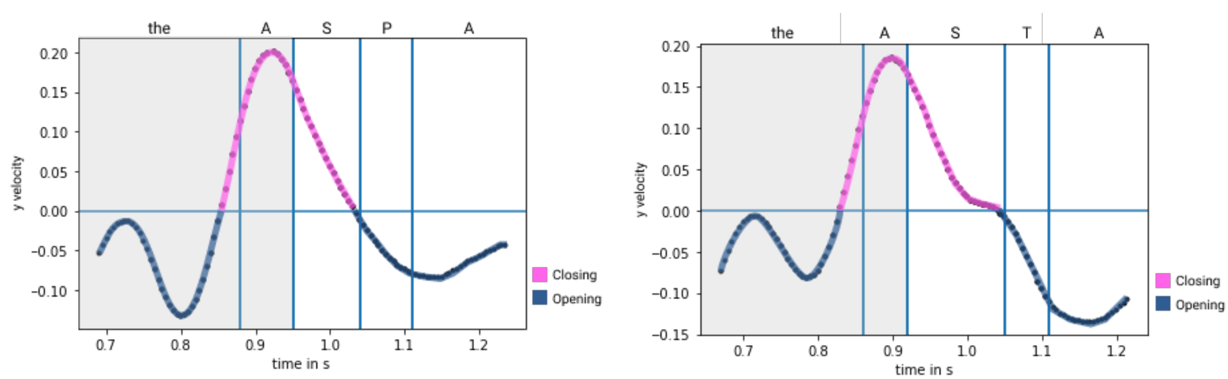


(a) /itɹi/ (b) /oʊstoʊ/

Figure 5.13: Jaw trajectories plotted with acoustic segmental boundaries for two utterances of the targets 'eetree', /itɹi/, and 'osto', /oʊstoʊ/, produced in the carrier phrase 'Say the [x] again' in the medial stressed condition

While these results are consistent in the alignment of max peak, the initiation of the opening phase occurs much later in Figure 5.13a than in Figure 5.13b. While further study is needed to analyze whether peak closure, initiation of the opening gesture, or both is most relevant in determining coordinative stability, these results may be better analyzed as having onset-coda syllabification, signaling points for improvement in future analyses, discussed in more detail below. However, for the purposes of this study, a combination of factors was used to determine coordination, and as such, both tokens were grouped with the onset maximization syllabification tokens.

Finally, Figure 5.14 provides examples of utterances in the medial unstressed condition that met Condition C. In both utterances, the peak velocity occurs prior to the start of C1, but the initiation of the opening movement occurs at the C1-C2 border.



(a) /aspa/

(b) /asta/

Figure 5.14: Jaw trajectories plotted with acoustic segmental boundaries for two utterances of the targets ‘aspa’, /aspa/, and ‘asta’, /asta/, produced in the carrier phrase ‘Say the [x] again’ in the medial unstressed condition

Insofar as each landmark can be interpreted individually, the alignment of the peak landmark indicates onset maximization, whereas the alignment of the opening trajectory indicates onset-coda syllabification. However, the significance of the difference in alignment between landmarks is yet to be determined. Based on the results of the present study, while the alignment of the initiation of the opening phase was significant, only the alignment of the peak opening velocity reflected gradience in syllabification between the tasks. Furthermore, the initiation of the opening phase was more variable even among the initial stressed tokens, where the example in Figure 5.1b, which shows an initial stressed token from study one, had an opening phase initiation at the border of the C1 and C2, indicating that this landmark may correspond with factors other than stress and word position. Because both landmarks are significant in distinguishing tasks and because the relationship between alignment landmarks is not yet clear, taking the combined outcome of the conditions was determined to be the best method for analyzing syllabification; however, future studies will follow up on this result to analyze the relevance of each landmark and the alignment of these landmarks respective to one another.

One prediction of this study is that identifying tokens that correspond in jaw coordination will result in an improvement of analyses of coordinative stability. Thus, in addition to analyzing differences in jaw coordination between tasks, the proportion of tokens that had a c-centering effect was calculated prior to subsetting for jaw coordination and after subsetting utterances to only those that met Condition C. Prior to subsetting, across all three tasks, the c-center was the least variable margin in 55% of tokens; however, after subsetting to only the tokens that met Condition C, the c-center was the least variable margin in 68% of tokens across all three tasks. Thus, identifying jaw coordination can aid in analyzing coordinative stability, and furthermore, this outcome suggests that the landmarks of jaw coordination identified in this study are positively correlated with coordination and likewise syllabification. Therefore, improvements in analyzing patterns of coordination are likely to continue to improve analyses of coordinative stability.

Finally, looking at the patterns of jaw coordination across cluster types, the results of this study suggest that intrinsic pressures on articulation affect jaw coordination. Table 5.2 provides the proportion of tokens that met Condition C.

	SP	**ST**	**SK**	**PR**	**PL**	**TR**	**TL**	**KR**	**KL**
Initial Stressed	29%	33%	25%	31%	26%	30%	31%	18%	23%
Medial Stressed	31%	28%	28%	21%	22%	30%	31%	17%	14%
Medial Unstressed	27%	26%	28%	18%	17%	19%	24%	10%	10%

Table 5.2: Percent of tokens per cluster that meet the requirements for Condition C

For both the initial stressed and the medial stress conditions, the /s/+stop clusters appear to be relatively equally likely to meet Condition C as the stop+liquid clusters. However, in the medial unstressed task, /s/+stop clusters were more likely to meet condition C than any of the stop+liquid clusters. This is consistent with the outcome of the stability analyses, where results for the initial stressed and medial stressed conditions were more similar than the medial unstressed condition. Additionally, like the stability results, clusters containing /k/, particularly the /k/+liquid clusters, were the least likely to meet Condition C. Finally, homorganic clusters were at least equally as likely and in some cases slightly more likely to meet Condition C than heterorganic clusters. The results both for clusters containing /k/ and the homorganic cluster suggest that there are intrinsic pressures that affect jaw coordination. Namely, segments with greater jaw coupling, like /k/, are less likely to be coordinated in the same jaw phase, particularly when the cluster is sequenced back-to-front, as is indicated by the results for the /k/+liquid clusters. On the other hand, coronal segments, like /s/ and /t/, have less jaw coupling, and thus, they are more likely to occur in the same jaw phase, as is indicated both by the homorganic results, and the improvement in results for the /s/+stop tokens in the medial unstressed task. Overall, the results suggest that coordinative stability is influenced by jaw coupling and jaw phase rather than overlap in articulation between segments or simultaneous planning alone, as is proposed by the Coupled Oscillator Model. Furthermore, both the stability results and the jaw results suggest that intrinsic pressures influence coordination, providing support for the Frame-Content Model.

5.2.3 Discussion of methodology

The landmarks identified in this study provide a foundation for determining patterns in jaw coordination and were overall able to capture differences in coordination across the three tasks. However, in analyzing the presence of these landmarks as binary properties, minor differences in timing or trajectory resulted in categorical outcomes that successfully served as an initial test and offer clear points for iteration, but which result in relatively conservative results. For instance, as described in the holistic results, the timing of the opening gesture within the C1 was crucial in distinguishing between tasks, but the binary metric was too coarse to capture the timing of this landmark in some cases. This section will highlight a number of key ways in which future analyses can improve upon the methods implemented in this study.

One common pattern that resulted in a token being categorized as non-conforming although the overall trajectory seemed to pattern with the conforming results, was the timing of the peak closure velocity. For instance, where the peak upward movement happened just after the start of the cluster C1, the target was marked as not meeting Condition B or subsequently Condition C. An example of such a token is provided in 5.15.

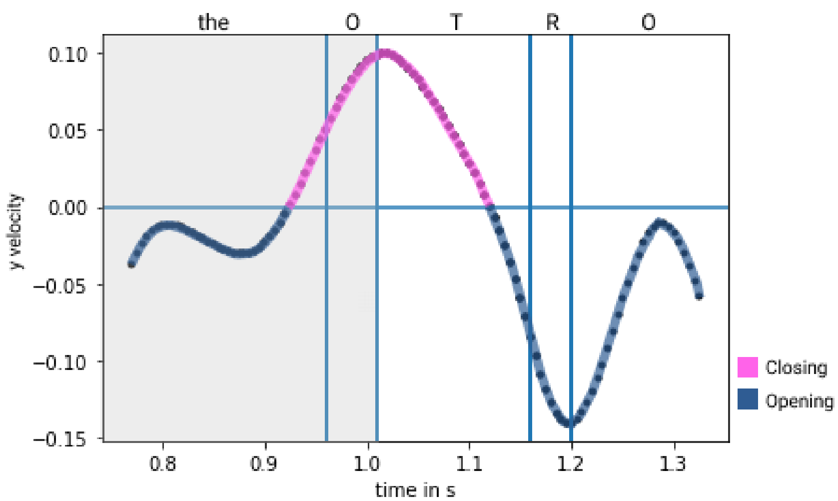


Figure 5.15: Jaw trajectories plotted with acoustic segmental boundaries for utterances of the target 'otro', /oʊtɹoʊ/, produced in the carrier phrase 'Say the [x] again' in the medial unstressed condition. The example demonstrates a slight difference in alignment of the max peak between targets conforming to Condition C and some targets that were not labeled as conforming, such as this example.

Another frequent pattern that caused a target to be marked as non-conforming to Condition C is that for the initial stressed tokens, the maximum peak displacement occurred in the coda of the target; thus, while another peak occurred before the start of the target cluster, the peak was not the maximum peak and was therefore labeled as non-conforming. An example of such a token is provided in 5.16.

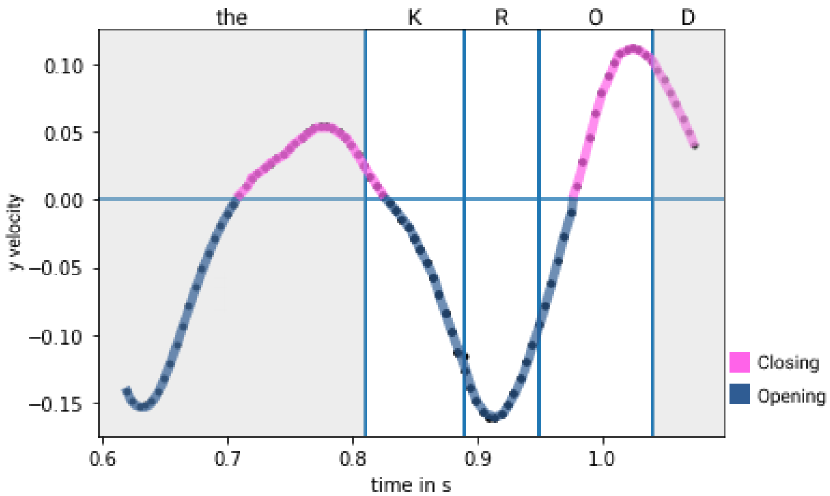


Figure 5.16: Jaw trajectories plotted with acoustic segmental boundaries for utterances of the target ‘krod’, /kɹad/, produced in the carrier phrase ‘Say the [x] again’ in the initial stressed condition. The example demonstrates a higher peak in the coda which resulted in the example being marked as not meeting Condition C.

In addition, another common pattern of coordination that was labeled as non-conforming were targets for which the initiation of the opening phase occurred just after the C2 rather than just before, as with many of the target examples labeled as conforming; examples where the downward trajectory did not begin in C1 did not meet Condition A. An example of such a token is provided in 5.17.

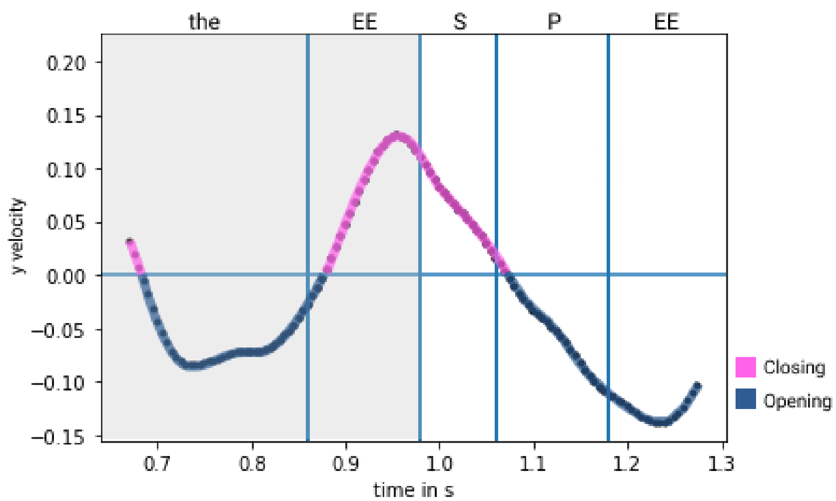


Figure 5.17: Jaw trajectories plotted with acoustic segmental boundaries for utterances of the target ‘eespee’, /ispi/, produced in the carrier phrase ‘Say the [x] again’ in the medial unstressed condition. The example demonstrates a slight difference in alignment of the change in direction between targets conforming to Condition A and some targets that were not labeled as conforming, such as this example.

Finally, there are additional landmarks that indicate coordination between the jaw and the segmental material, such as width and prominence of the peak and the point of peak downward movement that are not analyzed here that a more nuanced method of analysis would catch and might be relevant in analyzing jaw phase.

There are two apparent solutions to the issues faced in this study. First, incorporating more sensitive time-based measures that capture whether a jaw landmark is timed to the onset or the offset of the segmental articulators will aid in better assessing whether a segment occurs in-phase with the jaw or anti-phase to the jaw oscillation. Second, employing statistical methods analyzing full trajectories, e.g., Generalized Additive Model (GAM), rather than narrow landmarks, will likely yield the same benefits as employing more sensitive timing measures and will also likely uncover more nuanced patterns in coordination. Both methods are planned for the next steps of this research program.

The places where relevant data were not grouped with the conforming coordination landmarks explains why the proportion of data matching Condition C was only 27% even in the case of the initial stressed condition. However, despite the overall conservative nature of the measures employed in this study, the landmarks identified and analyzed here were successfully able to group highly similar candidates and to exclude alternating patterns of coordination, such as peak closure during C1 as opposed to prior to C1. For example, the utterance shown in Figure 5.18 patterns with Figure 5.12a and was excluded from the conforming tokens, as intended in identifying the relevant landmarks for analyzing jaw movement and the overall timing of these landmarks.

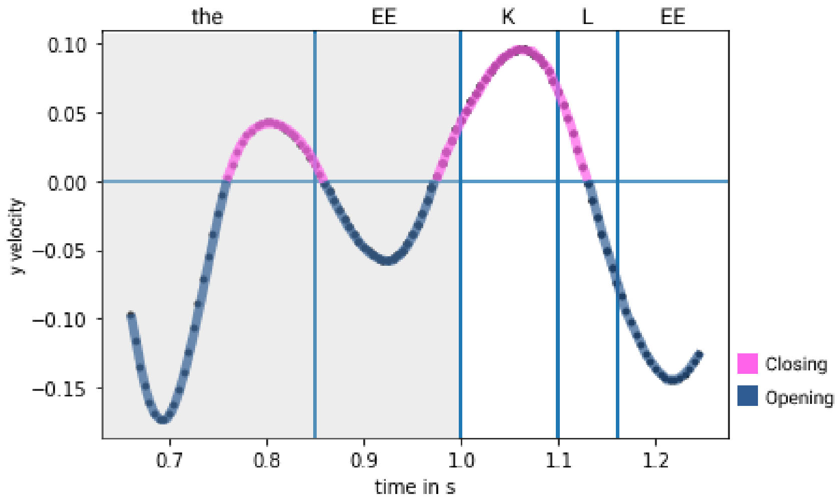


Figure 5.18: Jaw trajectories plotted with acoustic segmental boundaries for utterances of the target 'eeklee', /ikli/, produced in the carrier phrase 'Say the [x] again' in the medial unstressed condition. The example demonstrates a counter coordination example where the target was labeled as non-conforming and the coordination pattern does differ from those labeled as conforming.

The ability of these measures to show significant differences in jaw coordination despite the conservative nature of the landmarks highlights the importance of these landmarks in capturing alternating coordination patterns in the data and ultimately the success in the methods employed in this study.

5.3 Chapter summary and discussion

This study implements the Frame-Content model in a theory of syllable structure in which the cycle of jaw opening and closing correspond with onsets and codas, predicting that once primary jaw movement is included in the model, variability in secondary measures like c-centering will be explained; moreover, the jaw opening and closing provides a way to identify the durational window within which it is possible to answer questions regarding syllabification and coordination. Furthermore, this chapter provides a methodology for analyzing jaw coordination, where peak closure velocity and initiation of the opening gesture provide landmarks for analyzing jaw phase coordination. While the development of the methodology is still in its early stages, these results successfully demonstrate that jaw coupling can provide a clearer picture of coordination patterns and syllable structure.

The results of analyzing jaw phase were largely consistent with the results of Chapter 4, which concluded that medial syllables appeared to be more variable than initial syllables. The results of the jaw analysis likewise support the conclusion that medial syllables are more variable than stressed syllables. However, unlike the coordinative stability results, the results of the jaw analysis provide a clearer distinction between the medial stressed and unstressed conditions. Specifically, the results of this chapter illustrate that there is a pattern of gradient variability across the three task conditions, where initial stressed syllables are the least variable and medial unstressed syllables are the most variable. In particular, the alignment of the peak closing velocity was the least variable in the initial stressed syllable task and the most variable in the medial unstressed syllable task. The increased variability in the medial unstressed syllable condition compared to the medial stressed condition is consistent with studies such as Byrd et al. (2009), which demonstrates that stress attracts intervocalic nasal consonant gestures. In addition, this study provides novel evidence that medial syllables differ from initial syllables in that they are variably coordinated to the adjacent vowels, where both medial conditions show variability between the syllabification V.CCV and VC.CV, with V.CCV being more likely in the medial stressed condition.

In addition, the results of this study provide tentative support that intrinsic differences in the articulation of segments influences jaw coordination, where homorganic clusters and front-to-back clusters showed a greater likelihood to have corresponding jaw traces compared to heterorganic and especially back-to-front clusters, like /k/+liquid clusters. These results are consistent with the findings of Vallée et al. (2009). However, this chapter also demonstrates that there are a number of subtle differences in jaw coordination between varying sequences of segments that will require further study in order to better understand how intrinsic pressures affect coordinative stability. In particular, while this study identifies two contrasting patterns in jaw coordination, suggesting at least two different syllabifications, more fine-grained analyses of jaw movement provide the opportunity to ask more detailed questions about syllabification. For example, additional analyses of alignment and coordination might help demonstrate the presence of absence of ambisyllabicity and if the jaw can provide any insight into what makes word-medial lenition so common.

Although current proposals of the Coupled Oscillator Model do not assume that jaw coordination determines coordinative stability, the two theories can straightforwardly be reconciled by assuming that the in-phase coordination proposed by the Coupled Oscillator Model corresponds to in-phase jaw movement. The prediction of this combined model is that coordinative stability should improve when onset clusters occur within the same jaw phase. The results of this study support this prediction as there was an improvement in c-center effect when the data was subsetted to utterances where the onset occurred in the same jaw opening phase.

Crucially, the results of this study demonstrate that both stress and word position result in differences in coordination patterns and provide tentative support that intrinsic factors also drive differences in coordination. Thus, together the results illustrate a dynamic view of syllabification in word-medial environments, where coordination is variable and sensitive to both top-down and bottom-up pressures. These results have significant implications for phonological models of syllables, opening up exciting possibilities, like the ability to model variable phonological processes as related to variable syllabification. Thus, further development of the methodology outlined here and implementation of these findings in phonological modeling will improve our understanding and ability to predict cross-linguistic patterns.

Chapter 6

Implications and conclusion

Syllables are an important unit in phonological and phonetic theory as they provide an elegant architecture for accounting for a wide number of both phonetic and phonological patterns. However, segmenting word-medial consonants into discrete syllables and how word-medially syllables differ generally has proved less straightforward as there are a number of simultaneous pressures, e.g., stress and segment quality, affecting the coordination of word-medial syllables. Furthermore, the extent to which the coordination of word-medial syllables differs from coordination in word-initial syllables has not been not well understood as studies analyzing the effect of syllable structure on articulation have focused predominantly on tokens for which syllabification is straightforward.

This in-depth study has analyzed the effect of word position, stress, and segment quality on coordination, analyzing Electromagnetic Magnetic Articulography (EMA) data of utterances across three task conditions: initial stressed syllables, medial stressed syllables, and medial unstressed syllables with tokens varying in vowel and cluster composition. The most notable difference between tasks was the difference between word position, where both segmental gestures and coordination differ between word-initial and word-medial syllables. However, stress and segment quality were also found to affect articulation, particularly coordination. The results of this study have important implications for theories of syllable structure and syllabification, as will be discussed in this chapter.

6.1 The implications of differences between word-medial and word-initial segments and syllables

This study analyzed differences in segmental articulation between the three task conditions and differences in coordination across the three task conditions and across vowel qualities and cluster composition. Analyses of syllable coordination included existing methods like compensatory shortening, c-centering, and rightward shift as well as a novel method of analyzing coordination through jaw movement. All three analyses demonstrated clear differences between word-initial and word-medial and stressed and unstressed articulation and coordination. In addition, intrinsic pressure also influenced coordination.

6.1.1 Segmental articulation and lenition

Chapter 3 demonstrates that segments in word-medial positions are more reduced and have looser constrictions than segments in word-initial position. These results were true across stress conditions, indicating that while syllable position may factor into this result, i.e., segments in word-medial position occurred in coda position, reflecting the findings of Krakow (1999), the persistence and significance of the effect across segments and environments suggest word position more generally is a better explanation of the results than syllabification, although the explanations are not mutually exclusive.

The results of this study provide important insights into understanding lenition. First, articulation is likely to play some role in driving patterns of lenition, supporting Kirchner (2001). In particular, while prosodic modulation may also affect lenition, prosodic modulation does not predict there to be differences in articulation in a language like English were there is no phonological alternation between /k/ and /p/, but both showed greater reduction and looser constrictions word-medially. Not only is the difference in articulation between environments consistent with lenition, but the specific type of difference, namely, looser constrictions, is consistent with an alternation where stops become fricatives word-medially. Furthermore, these results are consistent with the observation of Gurevich (2011), which argues that there is often a clear diachronic progression in patterns of lenition, where lenition starts out as a pattern in fast or casual speech but progresses to become a phonological alternation overtime. This progression is best explained through an articulatory account of lenition, where the process begins with reduction in fast and casual speech that becomes routinized and phonologized overtime. Nevertheless, a prosodic account is not mutually exclusive with an articulatory account. What begins as articulatory reduction may have perceptual advantages that perpetuate the reduction until it becomes phonologized. Furthermore, exploring the interaction between prosodic and articulatory factors may be fruitful in explaining cross-linguistic differences in lenition patterns or within language differences between segment types. Overall, these results provide clear support for articulation as a contributing factor in lenition.

6.1.2 Syllable coordination and syllabification

Chapter 4 similarly showed an effect of word position, where word-medial syllables were less stably timed than word-initial syllables. The results of the compensatory shortening analysis suggest that unstressed syllables are more variable than stressed syllables; however, the results of the c-centering analysis suggest that both medial conditions were variably coordinated. Chapter 5 demonstrates jaw coordination differs between the three task conditions, in particular illustrating that word-medial syllables are more variably coordinated than word-initial syllables. In both medial contexts, the peak closing velocity and the onset of the opening phase was more likely to differ across utterances. In particular, the peak closing velocity and initiation of the opening gesture in the initial stressed condition occurred prior to C1 and during C1, respectively; however, while some tokens shared this pattern in the medial conditions, both medial conditions also had tokens with these landmarks aligned later in the utterance, where the peak closing velocity and initiation of the opening gesture occurred during the C1 and at the C1-C2 boundary, respectively. These two patterns of coordination are argued to correspond to syllabification, where the earlier alignment of the jaw landmarks corresponds with onset maximization syllabification and the later alignment of the jaw landmarks corresponds with a onset-coda syllabification.

While the methods employed in Chapter 4 did not provide a clear distinction between stressed and unstressed syllables, Chapter 5 demonstrates stress-based differences in coordination. In particular, Chapter 5 illustrates that patterns in variation and alignment were gradient across the three tasks. While the initial stressed syllable condition was least variable, medial unstressed tokens were both most variable and most likely to for utterances to have the later aligned onset-coda coordination pattern, providing support that consonant gestures are attracted to the stressed syllable, as illustrated by Byrd et al. (2009). These results demonstrate that both word position and stress influence coordination and syllabification, and furthermore, the results of Chapter 5 demonstrate the effectiveness of analyzing jaw cycles in interpreting patterns of syllabification; providing support for the Frame-Content Model (MacNeilage, 1998).

In addition to differences between tasks, analyses of differences coordination between varying clusters and vowel qualities provide tentative support that intrinsic pressures affect coordination and syllabification. Namely, Chapter 4 indicates that syllable onsets containing /k/ and /r/ were produced more variably than other clusters. While the difference in /r/ clusters may be explained through perceptual or articulatory factors, there is no reason why a /k/ should differ from other stop segments in a perceptual account. Thus, this result indicates greater jaw coupling for /k/ than for /t/ or /p/, resulting in less stability. In addition, this study found tentative support for intrinsic pressure in coordination between vowels and consonants, where consonants followed by a vowel matching in backness, e.g., /sti/ or /ska/ were more stably timed than sequences that did not match in backness; however, this outcome must be further investigated. Finally, the study also demonstrates that homorganic clusters were produced comparably to other clusters, indicating that syllable coordination is less the result of coupling among segmental gestures, but instead between segmental gestures and the jaw. In fact, a number of the homorganic clusters were among the most stable, even for /tl/, which is not a licit onset cluster in English. This outcome is consistent with a jaw-based framework of syllabification as homorganic clusters are more easily sequenced within a single jaw phase, at least to some extent, although an /lt/ onset would contrast with a /tl/ onset.

While the results of this study illustrate differences in coordination and syllabification based on stress, word position, and sequence composition, they also highlight similarities between sequences that provide support for syllables as a means of modeling patterns of coordination. For instance, while the alignment of these landmarks differed across the utterances, suggesting a difference in syllabification, the majority of tokens across all three tasks consisted of the same two main landmarks consistent with syllable onsets according to the framework of the Frame-Content Model. These shared landmarks thus suggest a difference in syllabification rather than a difference in the overall underlying structure. Therefore, this study concludes that syllable architecture provides both a necessary and sufficient architecture for capturing patterns of coordination and related phonological effects. Furthermore, syllables are well positioned as an interface between top-down prosodic pressure and bottom-up intrinsic pressure, making them an optimal tool for modeling dynamic interaction across these pressures.

Overall, the results of this study support a framework of coordination and syllabification that is dynamic and influenced by a number of factors. While this study provides support that both stress and intrinsic articulatory pressures affect coordination and syllabification, these pressures are likely not exhaustive. Other factors such as perception and planning are also likely to affect coordination, as is suggested by models of syllabification like licensing by cue Steriade (1997) and studies like Kilbourn-Ceron et al. (2016), which demonstrates that flapping is affected by frequency and planning. Ultimately, the results of this study support modeling syllabification with regard to a number of constraints including jaw alignment, gestural correspondence, and perceptual factors. Section 6.2 will propose constraints for modeling jaw cycles in syllabification in more detail.

6.1.3 Future directions in analyzing word-medial differences

The results of this study have implications for other phonological models of differences between word-initial and word-medial syllables that are not analyzed in detail in this study. Namely, analyzing jaw oscillation will likely provide a means to investigate phonetic support for extrametricality and ambisyllabicity. For instance, one reason why more complexity may be permitted at syllable margins is that multiple consonants can be articulated while the jaw is still relatively closed, whereas word-medial syllables are coordinated between opening and closing cycles, which may offer less time to articulate multiple consonants or less overall jaw closure, which may make articulating multiple segments less preferable. Potential tools for investigating this question are to compare the degree of closure at initial and medial margins and to analyze the stability between landmarks, e.g., the stability of the vowel-to-vowel timing, in order to determine how syllables are sequenced compared to how words are sequenced. Such studies are also likely to provide insight into ambisyllabicity. Furthermore, in analyzing whether there is phonetic support for ambisyllabicity, further analyses of the alignment of jaw cycle landmarks, like those employed in this study, may also provide a more fine-grained understanding of differences in word-medial timing that demonstrate whether segments are coordinated across multiple syllables. Future studies will thus seek to analyze more fine-grained differences in jaw velocity between utterances and to expand the variables of the analysis to further improve understanding of how word-initial and word-medial margins differ.

6.2 Operationalizing articulatory pressures into constraint based models of syllabification

Studies supporting the Frame-Content Model (MacNeilage, 1998; Vallée et al., 2009) indicate that jaw oscillation and coupling between the jaw and segmental articulators condition a large number of articulatory patterns in syllable typology. The results of this study are consistent with these claims and demonstrate that intrinsic articulatory pressures affect jaw coordination. Furthermore, this study demonstrates that prosody also affects jaw coordination and argues that a number of factors, likely including perception as well as articulation, interact to influence patterns of coordination. Thus, this section lays a foundation for capturing the role of jaw modulation in syllabification through two simple constraints that can be used in conjunction with perception-based and gestural-correspondence-based constraints. The first constraint provides a framework for the sequencing of segments according to jaw phase. Segments like stops and nasals[1] occur with the jaw at a closed or near closed position, vowels require the jaw to be relatively open, and fricatives and approximates fall between these two ends of the spectrum. This pattern can be captured in a constraint-based framework through the implementation of a basic constraint that adheres to the following Jaw Sequencing Hierarchy:

(1) **Closed** | nasals – stops – fricatives – approximates – glides – vowels | **Open**

As such, the Jaw Sequencing Constraint assesses a violation for sequences of segments that initiate a change in direction of jaw phase, where onsets are presumed to occur within a jaw opening phase and codas occur in a jaw closure phase.

The Jaw Sequencing Constraint constraint interacts with a constraint that provides a hierarchy for freedom of movement between a given articulator and the jaw, e.g., the tongue dorsum have relatively little freedom of movement from the jaw, whereas the tongue tip has relatively high freedom of movement from the jaw. Furthermore, classes of segments differ in the articulator gesture, so while alveolars have the highest degree of freedom from the jaw, retroflexes have far less. Thus, depending on the gestural composition of the segment, segments may have greater freedom to deviate from the Jaw Sequencing Hierarchy without triggering a change in jaw gesture direction. This hierarchy is implemented in the grammar through the Jaw Coupling constraint, which assesses a violation for sequences segments that occur above of the minimum distance of jaw range. This constraint allows for language specific differences both in the gestural composition of segments and in the distance permitted before a change in jaw direction is initiated.

A preliminary view into how these constraints might provide insights into language specific phonotactics comes through a comparison of complex segments and complex clusters in Nafaanra. In Nafaanra, intrinsic pressures determine which sequences are viewed by speakers as complex segments or complex clusters. Nafaanra permits only C and CC onsets and does not permit codas except for /m/ rarely. However, while the second segment in the cluster must be a sonorant, a range of segments occur in both the singleton and the first segment of the complex cluster, including complex segments. The combination of complex segments in the C1 position and sonorants in the C2 position give rise to complex onset sequences like /mbro/, 'fish'. In addition, Nafaanra permits complex clusters consisting of a stop and a nasal, such as /pnaa/, 'game' (Garvin, 2017). Complex segments such as pre-nasalized stops, e.g., /mb/, which follow the Jaw

[1]Nasals are ranked as more closed than stops because they are released, initiating jaw opening in a way that nasals do not.

Sequencing Hierarchy, can co-occur with an additional segment in the onset, e.g., /ndra/, *head*. However, stop+/n/ sequences, which violate jaw sequencing but at a permitted segment-jaw distance due to the alveolar in the C2 position, are timed further apart and do no allow another consonant in the onset and are instead interpreted as a cluster rather than a complex segment.

Another example of how these constraints might be implemented in accounting for cross-linguistic patterns comes from onset clusters in Georgian, where front-to-back clusters are more common and more complex than back-to-front clusters. While the front-to-back clusters do not incur any violations of either jaw constraint, back-to-front clusters are more likely to violate the Jaw Coupling constraint, limiting the ranging of complexity and frequency of these clusters (Butskhrikidze, 2002). Thus, these languages demonstrate an interaction between grammatical and motor constraints that interact to manifest language-specific distributional effects. Future work in this line of research will demonstrate how these constraints can be implemented to account for the English data presented in this study in addition to languages such as Georgian and Nafaanra.

While this jaw hierarchy draws on the principles of sonority constraints, there are crucial differences between the hierarchies, where the jaw sequencing principle avoids spurious predictions like the sequencing of stops and nasals, where sonority incorrectly predicts that stop+nasal should be more common, but nasal+stop sequences are more common cross-linguistically. Furthermore, sonority constraints fail to capture the broader distribution of coronals like /s/ and /t/, which are captured by the Jaw Coupling constraint. Ultimately, this research improves the predictions of syllable structure cross-linguistically. Furthermore, these findings provide a bridge between fine-grained phonetic data and quantized models of such patterns, making the tool kit more accessible in answering a broad range of questions regarding syllable structure without requiring articulatory analysis.

6.3 Conclusion

This study analyzes differences in segmental articulation and syllable coordination to demonstrate the significant articulatory differences between word-initial and word-medial syllables. In particular, word-medial segments are produced with looser constrictions than word-initial segments, providing an articulatory motivation to patterns of word-medial lenition. Furthermore, word-medial syllables are produced with greater variability in coordination, suggesting that word-medial syllabification is dynamic and sensitive to a number of pressures, including stress and segment quality, as is indicated here. Other pressures may also affect coordination. For instance, sequence and or word frequency and planning may determine how segments are aligned to the jaw gesture.

Ultimately, although this study finds that word position conditions the most significant results of this study, theories of syllable structure are best poised to capture these differences in articulation as they are a strategic interface for both intrinsic pressures like coordination between segments and prosodic factors like the pressure of stress. Thus, the results presented here can be captured through constraints on jaw sequencing and jaw coupling alongside existing syllable-based constraints to capture a range of cross-linguistic patterns.

This study serves as a foundation for analyzing jaw coordination and implementing the principles of the Frame-Content Model into models of phonology. Future studies will continue to develop the methodology for analyzing jaw movement by incorporating methods to analyze trajectories to analyze more fine-grained patterns in jaw coordination and timing-based relations between jaw movement landmarks to analyze stability in coordination. In addition, while the results of this study provide insights into English coordination, future studies must extend these findings to other languages to analyze the extent to which these findings are properties of English or of the jaw more generally. Likewise, modeling the articulatory results here in a phonological framework such as OT for English and for other languages will test the predictive power of the jaw in accounting for typological patterns in syllable structure. Ultimately, this study provides significant insights into the role of jaw coordination in syllabification, providing tools to better model and predict typological patterns in syllabification.

Bibliography

2014. The CMU Pronouncing Dictionary. URL `http://www.speech.cs.cmu.edu/cgi-bin/cmudict`.

Beckman, Jill. 1998. Positional Faithfulness. Doctoral Dissertation, University of Massachusettes, Amherst.

Beckman, Mary E, and Jan Edwards. 1994. Articulatory evidence for differentiating stress categories. In Phonological structure and phonetic form.

Berry, Jeffrey J. 2011. Accuracy of the NDI Wave Speech Research System. Journal of Speech, Language, and Hearing Research .

Blevins, Juliette. 1993. Klamath Laryngeal Phonology. International Journal of American Linguistics 59:237–279.

Blevins, Juliette. 1995. The syllable in phonological theory. In The handbook of phonological theory, ed. John A. Goldsmith, 206–244. Cambridge: Blackwell.

Boersma, Paul, and David Weenink. 2020. Praat: doing phonetics by computer.

Bombien, Lasse, Christine Mooshammer, Philip Hoole, and Barbara Kühnert. 2010. Prosodic and segmental effects on EPG contact patterns of word-initial German clusters. Journal of Phonetics 38:388–403.

Browman, Catherine P., and Louis M. Goldstein. 1988. Some notes on syllable structure in articulatory phonology. Phonetica .

Butskhrikidze, Marika. 2002. The consonant phonotactics of Georgian. Netherland Graduate School of Linguistics.

Byrd, Dani. 1995. C-Centers Revisited. Phonetica 52:285–306.

Byrd, Dani, Stephen Tobin, Erik Bresch, and Shrikanth Narayanan. 2009. Timing effects of syllable structure and stress on nasals: A real-time MRI examination. Journal of Phonetics 37:97–110.

Chitoran, Ioana, Louis Goldstein, and Dani Byrd. 2002. Gestural overlap and recoverability: Articulatory evidence from Georgian. In Laboratory phonology, volume 7, 419–448. De Gruyter Mouton.

Cholin, Joana. 2011. Chapter Nine. Do Syllables Exist? Psycholinguistic Evidence For The Retrieval Of Syllabic Units In Speech Production. In Handbook of the syllable, ed. Charles E. Cairns and Eric Raimy, chapter 9, 225–250. Brill.

Clements, G. N. 1990. The role of the sonority cycle in core syllabification. In Papers in laboratory phonology, 283–333. Cambridge University Press.

Côté, Marie-Hélène. 2011. Final consonants. In The blackwell companion to phonology, chapter 36.

Côté, Marie-Hélène, and Viktor Kharlamov. 2011. The Impact Of Experimental Tasks On Syllabification Judgments: A Case Study Of Russian. In Handbook of the syllable, ed. Charles E. Cairns and Eric Raimy, chapter 11, 274–293. Brill.

Curnow, Timothy Jowan. 1997. A grammar of Awa Pit (Cuaiquer): an indigenous language of south-western Colombia. Doctoral Dissertation, Australia National University.

Foley, James. 1970. Phonological Distinctive Features. Folia Linguistica 4:87–92.

Foley, James. 1977. Foundations of Phonological Theory. Cambridge University Press.

Fowler, Carol A., and Elliot Saltzman. 1993. Coordination and coarticulation in speech production. Language and Speech .

Fujimura, O., M. J. Macchi, and L. A. Streeter. 1978. Perception of stop consonants with conflicting transitional cues: A cross-linguistic study. Language and Speech 21:337–346.

Garcia, Damien. 2010. Robust smoothing of gridded data in one and higher dimensions with missing values. Computational Statistics and Data Analysis 54:1167–1178.

Garvin, Karee. 2017. Nafaanra Documentation Project. URL `http://dx.doi.org/doi:10.7297/X2V98672`.

Garvin, Karee, Myriam Lapierre, and Sharon Inkelas. 2018. A Q-theoretic approach to distinctive subsegmental timing. Proceedings of the Linguistic Society of America 3:9.

Gay, Thomas. 1978. Effect of speaking rate on vowel formant movements. The Journal of the Acoustical Society of America 63:223.

Goldstein, Louis M., Ioana Chitoran, and Elisabeth O. Selkirk. 2007. Syllable structure as coupled oscillator modes: Evidence from Georgian vs. Tashlhiyt Berber. In Proceedings of the XVI International Congress of Phonetic Sciences, ed. Trouvain J. and Barry W. J., 241–244. Saarbrücken.

Goldstein, Louis M., Elliot Saltzman, Ioana Chitoran, and Hosung Nam. 2009. Coupled oscillator model of speech timing and syllable structure. Frontiers in phonetics and speech science 239–250.

Gordon, Matthew K. 2016. Phonological Typology, volume 1. Oxford University Press.

Gouskova, Maria. 2004. Relational Hierarchies in Optimality Theory: The Case of Syllable Contact . Phonology 21:201–250.

Green, MC, and TC Brock. 2002. In the mind's eye: Transportation-imagery model of narrative persuasion. In Narrative impact: social and cognitive foundations, 315–341.

Gurevich, Naomi. 2011. Lenition. In The blackwell companion to phonology, 1–17. Oxford, UK: John Wiley & Sons, Ltd.

Halle, Morris, and Jean-Roger Vergnaud. 1980. Three dimensional phonology. Journal of Linguistic Research 83–105.

Harris, James W. 1983. Syllable structure and stress in Spanish : a nonlinear analysis, volume 8. MIT Press.

Hayes, Bruce. 1982. Extrametricality and English Stress. Linguistic Inquiry .

Hayes, Bruce. 1989. Compensatory lengthening in moraic phonology. Linguistic Inquiry 20:253–306.

Hayes, Bruce. 2009. Introductory phonology. John Wiley & Sons.

Hermes, Anne, Doris Mücke, and Martine Grice. 2013. Gestural coordination of Italian word-initial clusters: the case of 'impure s'. Phonology 30:1–25.

Hickok, Gregory. 2014. The architecture of speech production and the role of the phoneme in speech processing. Language, Cognition and Neuroscience 29:2–20.

Hoole, Phil, and Andreas Zierdt. 2012. Five-dimensional articulography. Speech Motor Control: New Developments in Basic and Applied Research .

Hooper, Joan B. 1972. The Syllable in Phonological Theory. Language 48:525.

Houde, John F., and Srikantan S. Nagarajan. 2011. Speech production as state feedback control. Frontiers in Human Neuroscience 5:82.

Hyman, Larry M. 1985. A theory of phonological weight. Dordrecht: Foris.

Hyman, Larry M. 2011. Does Gokana really have no syllables? Or: What's so great about being universal? Phonology 28:55–85.

Jensen, John T. 2000. Against ambisyllabicity. Phonology 17:187–235.

Ji, An, Michael T Johnson, and Jeffrey Berry. 2014. Articulatory space calibration in 3D electromagnetic articulography .

Jiahong, Yuan, and Mark Liberman. 2008. Speaker identification on the SCOTUS corpus. In *Proceedings of Acoustics '08.*.

Johnson, Keith, and Ronald L. Sprouse. 2019. Head correction of point tracking data. *UC Berkeley PhonLab Annual Report* 15.

de Jong, Kenneth J. 1995. The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation. *Journal of the Acoustical Society of America* 97:491–504.

de Jong, Kenneth J., Mary E Beckman, and Jan Edwards. 1993. The interplay between prosodic structure and coarticulation. *Language and speech* 36 (Pt 2-:197–212.

Kahn, Daniel. 1976. Syllable-based generalizations in English phonology.

Katz, Jonah. 2012. Compression effects in English. *Journal of Phonetics* 40:390–402.

Katz, Jonah. 2016. Lenition, perception and neutralisation.

Kilbourn-Ceron, Oriana, Michael Wagner, and Meghan Clayards. 2016. The effect of production planning locality on external sandhi: a study in /t/. In *Proceedings of the 52nd Meeting of the Chicago Linguistics Society*. Chicago.

Kingston, John, Laura Colantoni, and Steele Jeffery. 2008. Lenition. In *Selected proceedings of the 3rd Conference on Laboratory Approaches to Spanish Phonology*, 1–31. Somerville, MA: Cascadilla Press.

Kirchner, Robert Martin. 2001. *An effort based approach to consonant lenition*. Psychology Press.

Kozhevnikov, Valeriĭ Aleksandrovich, and Li͡udmila Andreevna Chistovich. 1966. *Speech: articulation and perception*, volume 30. Washington: U.S. Joint Publications Research Service, rev. [ed.] edition.

Krakow, Rena A. 1999. Physiological organization of syllables: A review. *Journal of Phonetics* .

Löfqvist, Anders, and Vincent L. Gracco. 1999. Interarticulator programming in VCV sequences: Lip and tongue movements. *The Journal of the Acoustical Society of America* 105:1864–1876.

MacNeilage, Peter F. 1998. The frame/content theory of evolution of speech production.

Marin, Stefania. 2013. The temporal organization of complex onsets and codas in Romanian: A gestural approach. *Journal of Phonetics* 41:211–227.

Marin, Stefania, and Marianne Pouplier. 2010. Temporal organization of complex onsets and codas in American English: Testing the predictions of a gestural coupling model. Technical report.

Mathôt, Sebastiaan, Daniel Schreij, and Jan Theeuwes. 2012. OpenSesame: An open-source, graphical experiment builder for the social sciences.

McAuliffe, Michael, Michaela Socolof, Sarah Mihuc, Michael Wagner, and Morgan Sonderegger. 2017. Montreal Forced Aligner.

Morton, John, Steve Marcus, and Clive Frankish. 1976. Perceptual centers (P-Centers). Psychological Review 83:405–408.

Munhall, Kevin, Carol Fowler, Sarah Hawkins, and Elliot Saltzman. 1999. "Compensatory shortening" in monosyllables of spoken English. Journal of Phonetics .

Munhall, Kevin, Carol A. Fowler, Sarah Hawkins, and Elliot Saltzman. 1992. "Compensatory shortening" in monosyllables of spoken English. Journal of Phonetics 20:225–239.

Murray, Robert W., and Theo Vennemann. 1983. Sound Change and Syllable Structure in Germanic Phonology. Language 59:528.

Nam, Hosung, Louis M. Goldstein, and Elliot Saltzman. 2009. Self-organization of Syllable Structure: A Coupled Oscillator Model. In Approaches to phonological complexity, chapter 16, 299–328.

Nam, Hosung, and Elliot Saltzman. 2003. A Competitive, Coupled Oscillator Model of Syllable Structure. In In Proceedings of the 15th international congress of phonetic sciences, 2253–2256.

Oganian, Yulia, and Edward Chang. 2018. A speech envelope landmark for syllable encoding in human superior temporal gyrus. bioRxiv .

Perkell, JS, MH Cohen, MA Svirsky, ML Matthies, I Garabieta, and MT Jackson. 1992. Electromagnetic midsagittal articulometer systems for transducing speech articulatory movements. The Journal of the Acoustical Society of America 92:3078–3096.

Prince, Alan, and Paul Smolensky. 2004. Optimality Theory. Oxford, UK: Blackwell Publishing Ltd.

Rubach, Jerzy. 1996. Shortening and Ambisyllabicity in English. Phonology 13:197–237.

Saltzman, Elliot, and Dani Byrd. 2000. Task-dynamics of gestural timing: Phase windows and multifrequency rhythms. Human Movement Science 19:499–526.

Schönle, PW, K Gräbe, P Wenig, J Höhne, J Schrader, and B Conrad. 1987. Electromagnetic articulography: use of alternating magnetic fields for tracking movements of multiple points inside and outside the vocal tract. Brain and language 31:26–35.

Selkirk, Elisabeth O. 1981. English Compounding and the Theory of Word Structure. In The scope of lexical rules, 229–278. De Gruyter Mouton.

Selkirk, Elisabeth O. 1982. The syllable. In The structure of phonological representations, ed. Harry van der Hulst and Norval Smith, volume 2, 337–383. Dordrecht: Foris Publications.

Shaw, Jason A., Adamantios I. Gafos, Philip Hoole, and Chakir Zeroual. 2009. Syllabification in Moroccan Arabic: Evidence from patterns of temporal stability in articulation.

Shaw, Jason A., Adamantios I. Gafos, Philip Hoole, and Chakir Zeroual. 2011. Dynamic invariance in the phonetic expression of syllable structure: A case study of Moroccan Arabic consonant clusters. Phonology 28:455–490.

Shaw, Jason A, and Shigeto Kawahara. 2018. The lingual articulation of devoiced /u/ in Tokyo Japanese .

Shih, Stephanie S., and Sharon Inkelas. 2019. Autosegmental aims in surface-optimizing phonology. Linguistic Inquiry 50:137–196.

Smolensky, Paul, and Alan Prince. 1993. Optimality Theory: Constraint interaction in generative grammar.

Steriade, Donca. 1997. Phonetics in Phonology: The Case of Laryngeal Neutralization. UCLA Working Papers in Phonology 3:25–146.

Tilsen, Sam, Debarghya Das, and Bruce McKee. 2015. Real-time articulatory biofeedback with electromagnetic articulography. Linguistics Vanguard 1:39–55.

Vallée, Nathalie, Solange Rossato, and Isabelle Rousset. 2009. Favoured syllabic patterns in the world's languages and sensorimotor constraints. In Approaches to phonological complexity.

Van Rossum, Guido, and Fred L. Drake. 2009. Python 3 Reference Manual.

Walker, Rachel, and Michael Proctor. 2019. The organisation and structure of rhotics in American English rhymes. Phonology 36:457–495.

Wieling, Martijn, Fabian Tomaschek, Denis Arnold, Mark Tiede, Franziska Bröker, Samuel Thiele, Simon N. Wood, and R. Harald Baayen. 2016. Investigating dialectal differences using articulography. Journal of Phonetics 59:122–143.

Wrench, Alan. 1999. The MOCHA-TIMIT articulatory database.

Wright, Richard. 2004. A review of perceptual cues and cue robustness. In Phonetically based phonology, ed. Bruce Hayes, Robert Kirchner, and Donca Steriade, 34–57. Cambridge University Press.