UNIVERSITY OF CALIFORNIA

Santa Barbara

Edgy prosody: an articulatory investigation of the role of lexical pitch accent in Tokyo Japanese boundary marking

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Linguistics

by

Karen Tsai

Committee in charge:

Professor Argyro Katsika, Chair

Professor Mary Beckman

Professor Eric W. Campbell

Professor Matthew Gordon

March 2023

The dissertation of Karen Tsai is approved.

Mary Beckman

Eric W. Campbell

Matthew Gordon

Argyro Katsika, Committee Chair

January 2023

Edgy prosody: an articulatory investigation of the role of lexical pitch accent in Tokyo Japanese boundary marking

by

Karen Tsai

For my parents

ACKNOWLEDGEMENTS

I have tremendous gratitude for everyone who has helped make this dissertation possible. I especially want to thank my entire committee — Mary Beckman, Eric Campbell, Matthew Gordon, and Argyro Katsika — for their guidance, support, and generosity with their time and energy. To Mary Beckman, for joining my committee despite being retired and busy with local election work. Thank you for the good energy you always sent my way, and for sharing your expertise and rich feedback even when you were sick with COVID. To Eric, who got me really excited about tone in Japanese during his tone seminar. To Matt, whose enthusiasm for my ideas, unwavering support, and kind feedback have always been an encouraging constant in the department.

To Argyro, who has been a monumental figure throughout my graduate career. The very first time we met, I knew immediately that I needed this amazing person to be my mentor. I'm so thankful to have had the opportunity to work with you as a research assistant, learning how to design and run EMA experiments, and as a teaching assistant, learning about articulatory phonetics and speech technology. I feel incredibly fortunate to have had such a dedicated, encouraging, and supportive advisor. I have learned so much from you, and I am glad and thankful to have grown as a scholar with you.

To all the Japanese speakers who participated in my experiments, as well as my research assistants Rie Waldon and Chisa Nagai. This study would not have been possible without them. To all of my friends who let me practice attaching EMA sensors on them. To all my teachers, mentors, friends, and host families from Japan who helped me achieve my dream of being fluent enough to conduct fieldwork in Japanese. I also acknowledge that this dissertation work was supported by NSF Grant #1551428.

To Yu Tanaka, whose Introduction to Phonetics course at UCLA sparked my love for phonetics, and Kie Zuraw, whose Phonology courses and mentorship inspired me to apply to graduate school so I could keep researching topics in Japanese phonetics and phonology.

To Caroline Crouch and Jiyoung Jang, for their friendship and support through all the EMA experiments, statistics struggles, and conferencing travel. To Sherry Chien, for all the work you have done (and will do) to continue my research. Our Phonetics Lab group will always have a special place in my heart.

To the many friends I made during graduate school who have made learning (and not learning) such a fun experience, especially: Chris Brendel, Alexia Fawcett, Chloe Willis, Jamaal Muwwakkil, Jiyoung Jang, Morgan Sleeper, and Julia Fine. To Caroline Crouch and Kevin Schafer, who took us out for drinks during my Open House and made me decide to choose UCSB for the wonderful people. To Phill Rogers and our many late-night chats in the lab about statistics, and Sandra Auderset, whose help with manipulating data in R and extremely useful tips on data management and analysis were absolutely game-changing.

I could not have completed this dissertation without the support of my peers who came together and committed to a Monday-through-Friday early morning writing group. We started as an East coast group of three, but we grew into an international group that included Alexia Fawcett, Chris Brendel, Sandra Auderset, Giorgia Troiani, Adrienne Tsikewa, Jamaal Muwwakkil, and Sherry Chien. I'm incredibly grateful to all of you for the accountability and companionship of our goal-setting, Karendoro timer-using group. I would never have finished this manuscript without the discipline I learned from everyone's commitment to the group and the momentum I gained from our knowledge-sharing and mutual encouragement.

Last but not least, to my family, who has always supported my love of languages. To my sister Josselyn, who helped me reset during the pandemic, introduced me to new joys in life, and reminds me that my successes don't come from sheer dumb luck. Your love and support lifted me out of one of the most difficult times of my life. To Avi, in remembrance of the sweetest dog who came with me on an adventure to New York City and was a huge source of comfort during the pandemic. And finally, to my kumquat-marmalade-making, coffee-roasting, kombucha-brewing parents: my mom, who has always nurtured and supported my love of learning, and my dad, the original Dr. Tsai and the crazy scientist who inspires me every day.

VITA OF KAREN TSAI

January 2023

EDUCATION

Bachelor of Arts in Linguistics, University of California, Los Angeles, December 2014 (departmental honors, cum laude)
Master of Arts in Linguistics, University of California, Santa Barbara, December 2018
Certificate in College and University Teaching, University of California, Santa Barbara, June 2021
Doctor of Philosophy in Linguistics, University of California, Santa Barbara, March 2023 (expected)

PROFESSIONAL EMPLOYMENT

2016-22: Teaching Assistant, Department of Linguistics, University of California, Santa Barbara
2022: Research Assistant, Department of History, University of California, Santa Barbara
2018-2021: Research Assistant, Department of Linguistics, University of California, Santa Barbara
2019-2021: Lead Teaching Assistant, Department of Linguistics, University of California, Santa Barbara
Summer 2020: Co-facilitator, Lead TA Institute, University of California, Santa Barbara
Summer 2020: TA Pedagogical Advisor, Instructional Development, University of California, Santa Barbara
2017: Teaching Assistant, Department of East Asian Languages and Cultural Studies, University of California, Santa Barbara
2017: Instructor, School Kids Investigating Language in Life and Society (SKILLS), University of California, Santa Barbara

PUBLICATIONS

Katsika, A. & Tsai, K. (2021). The supralaryngeal articulation of stress and accent in Greek. Journal of Phonetics, 88, 101085.

Tsai, K., & Katsika, A. (2020). Pitch accent and phrase boundaries: Kinematic evidence from Japanese. In Proc. 10th International Conference on Speech Prosody 2020 (pp. 215-219).

Tsai, K., & Katsika, A. (2019). The interaction between lexical prominence and pre-boundary lengthening in Japanese. The Journal of the Acoustical Society of America, 146(4), 3013.

Katsika, A., & Tsai, K. (2019). The scope of the kinematic effects of prominence in Greek. The Journal of the Acoustical Society of America, 146(4), 3084.

Tsai, K., Lester, N. A., & del Prado Martín, F. M. (2018). Bidialectal homophone effects in Kansai Japanese: An auditory lexical decision experiment. Acoustical Science and Technology, 39(2), 158-159.

AWARDS

Department of Linguistics Research Fellowship, University of California, Santa Barbara, 2021

GSA Excellence in Teaching Award, University of California, Santa Barbara, 2020-2021

ISSP Student Scholarship, International Seminar on Speech Production, 2020

Critical Language Scholarship Award, U.S. Department of State, 2018

Japanese-Language Program for Specialists Grant, Japan Foundation, 2017

ABSTRACT

Edgy prosody: an articulatory investigation of the role of lexical pitch accent in Tokyo Japanese boundary marking

by

Karen Tsai

This dissertation examines the interaction of word prosody with prosodic temporal and tonal events at the phrasal level in Tokyo Japanese and aims to provide a unified account of the coordination relations between them within the framework of Articulatory Phonology. Previous research on phrase-final lengthening and boundary tone coordination in stress languages such as Greek has found that there is some coordination between lexical and phrasal prosodic events, as well as between tonal and temporal phrasal events. Specifically, research on the interaction between word prosody and boundary events in Greek indicates that lexical stress presents similar timing patterns with phrase-final lengthening as with boundary tone initiation, which suggests that these two types of boundary-marking events (i.e., phrase-final lengthening and boundary tones) are timed with respect to each other. This dissertation investigates these relations in a language with lexical pitch accent and proposes an account of the interaction of word prosody with boundary marking within the framework of Articulatory Phonology.

In Articulatory Phonology, consonants and vowels are represented as sets of constriction gestures in the vocal tract. Tonal events are also represented as sets of tone gestures which unfold over time and can be coordinated with constriction gestures. Phenomena such as phrase-final lengthening are accounted for by π-gestures, which instantiate prosodic phrase boundaries and have clock-slowing effects on co-active speech gestures. We examine how these types of gestures are coordinated with each other in Japanese by analyzing kinematic and acoustic data of various Japanese words in controlled phrases. These data were collected using EMA (electromagnetic articulography) in experiments designed to test the interaction of lexical pitch accent position with phrase-final lengthening and boundary tone coordination. Three analyses were conducted on (i) the effect of pitch accent on the scope of phrase-final lengthening, (ii) the effect of pitch accent on boundary tone coordination, and (iii) the kinematic correlates of pitch accent per se. Lexical pitch accent position was found to have an effect on the scope of phrase-final lengthening such that the latter was initiated earlier in words with non-final pitch accent. These results imply that the tone gesture for the pitch accent is coordinated with the π-gesture. Similarly, pitch accent position was found to affect the timing of the boundary tone, which was initiated earlier in words with earlier pitch accent. Finally, no robust kinematic correlate of pitch accent was detected. On the basis of these finding, a final account is proposed where the lexical pitch accent gesture is coordinated anti-phase with the boundary tone, which in turn affects π-gesture coordination. This account considers tonal and temporal prosodic events together, which may provide a more complete view of intonation and of how word- and phrase-level events are connected.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Introduction

1.1. Introduction

This dissertation aims to better understand prosodic structure by examining prosodic phrasal boundary marking and its interaction with word prosody in Tokyo Japanese. Prosody is a component of grammar which encodes meaning via grouping and prominence. These functions include chunking the speech stream into larger cognitive units such as intonational phrases and marking units of relative importance such as prominent syllables within words or prominent words within phrases. This encoding has a fundamental role in speech production, perception, and acquisition, so looking at what happens at the edge of the phrase, at how phrase boundaries are marked, is of special interest to research which aims to better understand linguistic structure and communication. Prosodic boundaries are known to affect boundary-adjacent speech units in kinematic dimensions such as duration, displacement, and velocity, resulting in phrase-final constriction gestures that are longer, larger, and slower than their phrase-medial counterparts (Cho 2006). The temporal effects of prosodic boundaries are well-supported by acoustic and kinematic data (acoustics: Oller 1973; Berkovits 1993a, 1993b; Cambier-Langeveld 1997; White 2002; Turk and Shattuck-Hufnagel 2007; articulation: Edwards, Beckman and Fletcher 1991; Beckman and Edwards, 1992; Byrd and Saltzman 1998; Byrd 2000; Byrd, Lee, et al. 2005; Byrd, Krivokapić and Lee 2006; Krivokapić 2007; Katsika 2016; Kim, Jang and Cho 2017). However, the stretch of speech affected (i.e., the scope of the effects) and how these effects interact with word prosody are unclear. Previous work on Dutch, English, Greek, and Hebrew locates the largest and most systematic effect of phrase-final lengthening in the rime of the phrase-final syllable (Berkovits 1993a, 1993b, 1994; Cambier-Langeveld 1997; Nakatani, O'Connor and Aston 1981; Oller 1973; Wightman et al. 1992), a domain that seems to be important for the perception of boundaries as well (Wightman et al. 1992). There is also evidence that the scope of phrase-final lengthening interacts with the position of lexical stress (Turk and Shattuck-Hufnagel 2007; Katsika 2016). However, it is unclear whether the interaction of phrase-final lengthening with stress involves a continuous interval of speech such as the interval from the coda of the stressed syllable to the boundary (British English: White 2002), or two separate intervals, i.e., the rhyme of the final syllable and the rhyme of the stressed syllable, excluding intervening syllables (American English: Turk and Shattuck-Hufnagel 2007) (but see Byrd and Riggs 2008). An articulatory study that systematically disentangled the effect of lexical stress from that of phrasal pitch accent on boundary-related events in Greek found that phrase-final lengthening stretched over a continuous interval, which was affected by stress but not pitch accent (Katsika 2016); constriction gestures that were immediately adjacent to the boundary were affected when stress was final, but lengthening was initiated earlier when stress was earlier within the word. These results were accounted for using the π-gesture model (Byrd and Saltzman 2003) in which boundary effects on constriction gestures are modulated by clock-slowing π-gestures (Byrd and Riggs, 2008; Katsika 2016). Crucially, previous work on the scope of phrase-final lengthening is minimal and mainly focused on stress languages, and little is known about how spatio-temporal boundary events are coordinated with speech articulation in languages with different prosodic systems. This study of Tokyo Japanese contributes an analysis of a head/edge-prominence language with lexical pitch accents, which do not have prominence-lending functions (Venditti 2005; Venditti et al. 2008). In this way, Japanese differs from previously studied stress languages where lexically stressed syllables are potential carriers of phrasal pitch accents. Analyzing Japanese therefore enables us to separate word prosody from phrasal prominence while focusing on the effect of word prosody itself on the scope of phrase final lengthening.

In the tonal domain, boundary tones mark major phrase boundaries and are coordinated with speech gestures, usually those consisting of the phrase-final syllable (e.g., Silverman, Beckman, et al. 1992; Beckman and Pierrehumbert 1986). A study of Greek reported that the coordination of boundary tones is further fine-tuned by the position of the last stress in the phrase (Katsika, Krivokapić and Mooshammer 2014). Interestingly, these results were similar to the effects of lexical stress on phrase-final lengthening (Katsika 2016; see also Turk and Shattuck-Hufnagel 2007; Byrd and Riggs 2008). Both boundary tones and phrase-final lengthening were found to be initiated earlier in words with non-final stress as opposed to words with final stress. However, this evidence comes primarily from stress languages, and little is known about the role of word prosody in the coordination of boundary tones in languages with different prosodic typology. Of special interest are languages such as Japanese, which has a non-stress word prosodic system. Similar to stress languages, Japanese marks the head of prosodic units. However, this differs from stress languages, as head-marking in Japanese is achieved tonally with pitch accents; there are no strong correlations with duration or intensity. Furthermore, phrase-level prominence in Japanese is marked in the tonal domain, making it a head/edge-marking language (Jun 2014). Thus, in addition to the scope of phrase-final lengthening in Japanese, this dissertation examines the coordination of boundary tones, assessing whether the two effects present parallel patterns to each other in terms of their interaction with the pitch accent. Such parallel patterns would support the hypothesis that word prosody interfaces with phrase-level prosody, and that tonal and temporal phrasal events are interrelated more than currently assumed by theories of intonation.

The structure of this dissertation is as follows. First this chapter will provide an overview of prosodic structure and boundary-related kinematic effects, as well as the research questions for this dissertation. This will be followed by an overview of Articulatory Phonology, the theoretical framework for this study, in Chapter 2. Then, Chapter 3 will describe the common methodology across the experiments conducted for this study, followed by the specific methods for each experiment, their results, and interim discussions in Chapter 4 for phrase-final lengthening, Chapter 5 for boundary tone coordination, and Chapter **Error! Reference source not found.** for the kinematic correlates of lexical pitch accent. Finally, Chapter 7 will conclude with a general discussion and provide a theoretical account for our results.

1.2. Prosodic structure

Prosody is "the organizational structure of speech" (Beckman 1996) with which we structure information in the speech stream. It is the grammatical component which encodes meaning through grouping and prominence (see Fletcher 2010 for an overview). This grouping serves to chunk speech units into larger cognitive constituents, such as utterances into phrases and phrases into words. An example of prosodic grouping is provided in (1), where the two sentences are identical in terms of words but differ in prosodic boundary placement. Because grouping marks cognitive constituents, it is pivotal for speech production, perception, and acquisition.

(1) Grouping

a. I'm sorry I'm trying.

b. I'm sorry. I'm trying.

Prosody also uses prominence to mark units of relative importance, which also has an important role marking syllables within words or words within phrases as rhythmically or conceptually important. For example, in the examples provided in (2a), this marking occurs on the lexical level, where stress marks prominence on a syllable within a word; in the examples in (2b), this marking happens on the phrasal level, where prominence marks focus on a word within a phrase.

(2) Prominence

a. Lexical level: marks syllable within words
   i. **ob**ject (noun)
   ii. ob**ject** (verb)
b. Phrasal level: marks words within phrases
   i. The money was **donated**, not pledged.

Typologically, there are different prosodic systems. For example, in a stress language like English, prominence on the phrasal level can mark a word in a phrase as important with a pitch accent on the focused word. Within the word, the phrasal pitch accent is located on the stressed syllable, or the head of the word, as in the underlined syllable in (2b). This is called head-prominence marking, in which the head of a phrase is marked as prominent.

There are also different types of prosodic systems, such as edge-prominence languages and head/edge-prominence languages. Seoul Korean is an example of an edge-prominence language. Edge-prominence languages are "head-less", as there is nothing which specifies a word- or phrase-level head (Jun 2014). Prominence is marked at the word and phrasal levels with the edge of the prosodic unit, which is marked by a phrasal tone (Jun 2014). Meanwhile, Tokyo Japanese is an example of a head/edge-prominence language, as there are lexical pitch accents at the word level, but phrase-level prominence is primarily marked by means of the phrase edge (Jun 2014).

In this dissertation, we will use the X-JToBI prosodic annotation conventions, which is the extended version of J_ToBI (Venditti 2005), the Japanese version of the Tones and Break Indices system (ToBI) (Silverman, Beckman, et al. 1992). These models were designed as a standard for prosodic transcription and developed from the autosegmental-metrical theory (Goldsmith 1976; Pierrehumbert and Beckman 1988), and the X-JToBI model is intended for annotation of prosodic features in spontaneous speech (Maekawa, et al. 2002; Venditti 2005; Venditti, Maekawa and Beckman 2008). Within this model, there are two levels within the prosodic hierarchy: the Accentual Phrase (AP) and above that, the Intonational Phrase (IP) (Beckman and Pierrehumbert 1986; Maekawa, et al. 2002; Igarashi 2015). An AP is marked by a delimitative rise to a high (H) tone around the second mora (%L H-) and a gradual fall to a low (L) tone at the end of the phrase (L%). Pitch accent is marked by 'A', '%' refers to AP or IP boundary tones, and '-' to a phrasal tone. The following example in Figure 1. Pitch contour and X-JToBI annotation for test word in an unaccented AP: is an example from our experimental data which illustrates the tones in APs which have no lexical pitch accents. Figure 2 illustrates the tones in an AP which contains a pitch-accented word. The example utterance in Figure 1 is taken from the three-syllable word set and consists of an unaccented AP *honto ni* 'really' followed by another unaccented AP *namida* 'Namida (name)'. The example utterance in Figure 2 is also from the three-syllable word set and consists of an unaccented AP *honto ni* 'really' followed by an AP with word-initial pitch accent *na*mina* 'Namina (name)'. 'A' marks the location of the pitch accent, which is followed by a fall in

F0, and ‘pL’ indicates the beginning of the rise in F0 associated with the LH% boundary pitch movement, which signals question intonation.

Figure 1. Pitch contour and X-JToBI annotation for test word in an unaccented AP: %L H- L%.

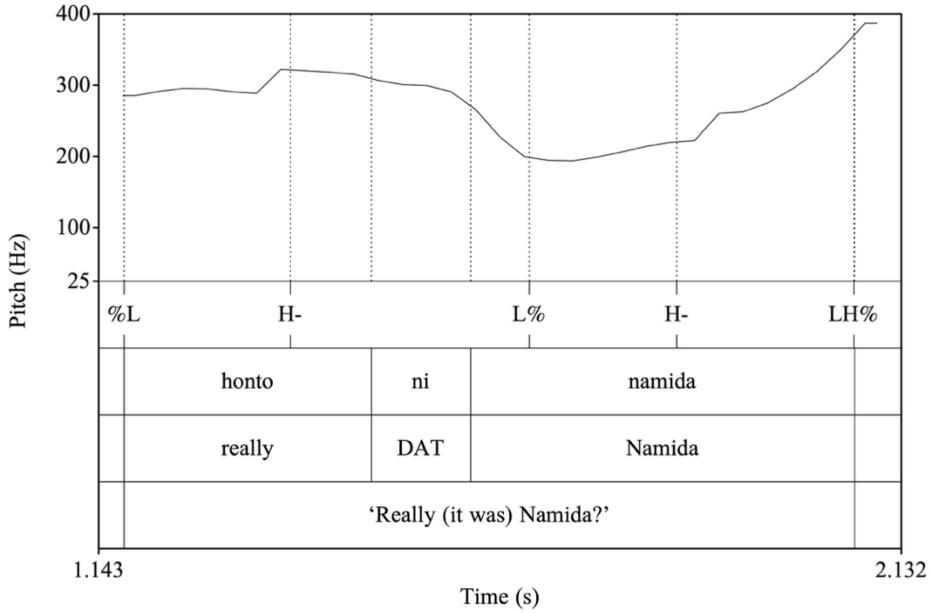
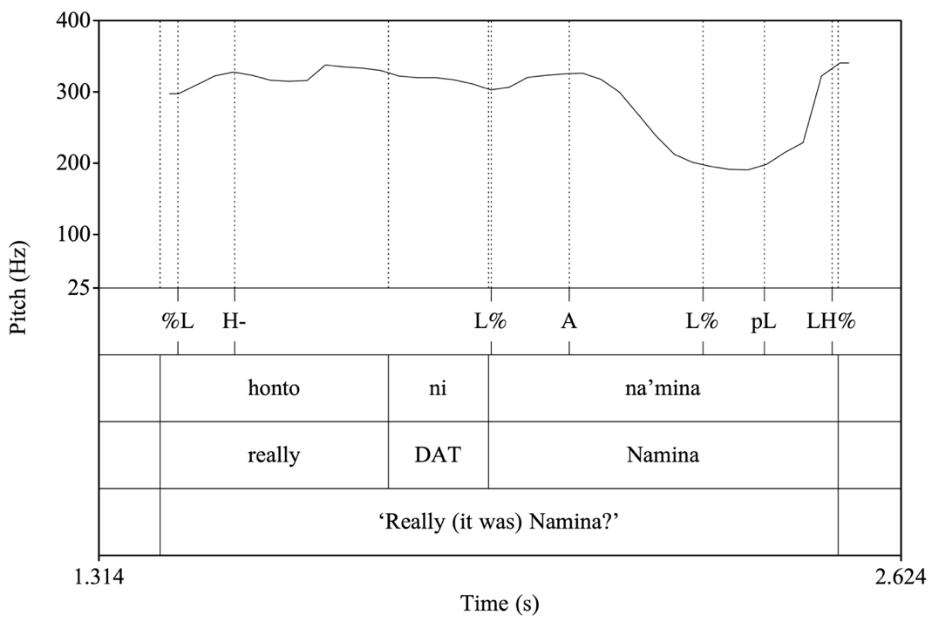
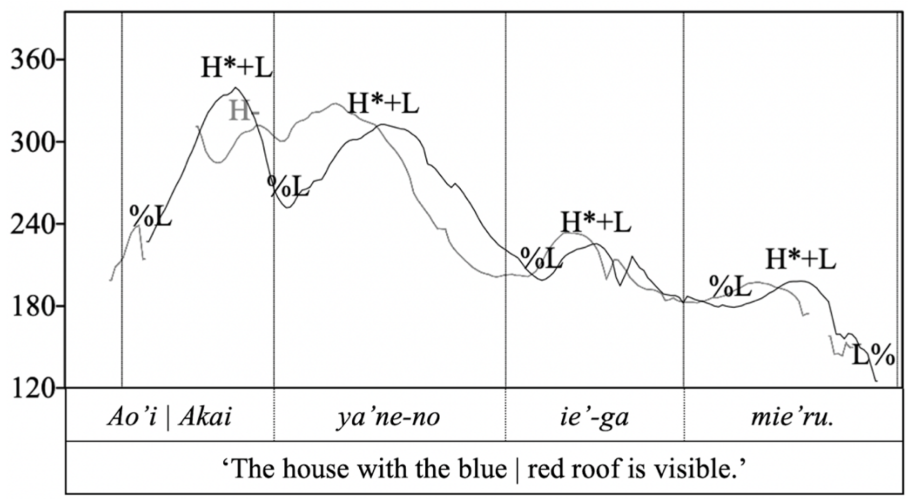


Figure 2. Pitch contour and X-JToBI annotation for test word in an accented AP: %L H- A L%.



The Intonational Phrase (IP), which we will refer to as the phrase level, is the prosodic domain immediately above the AP and is identified both tonally and by the degree of disjuncture (Venditti 2005). It is the domain in which pitch range is specified and is marked by pitch reset and downstep. Figure 3 from Venditti et al. (2008, p.11) illustrates downstep in an IP with multiple APs. The grey line shows the pitch for an IP with an unaccented word in the first AP; the black line shows the same for an accented word. Pitch accent is marked on the preceding syllable with the ’ symbol in the gloss. Note that this example uses J_ToBI annotation conventions, where pitch accent is marked as H*+L in the pitch track.

Figure 3. Overlaid F0 contours for utterances of two sentences showing that within one IP, the accent peaks of subsequent APs are lower than those in immediately preceding APs (from Venditti et al. 2008, p.11).



IPs are also marked tonally with boundary pitch movements, which follow AP tones and contribute to the pragmatic interpretation of the utterance, such as a question being indicated with a pitch movement from a low to high tone (LH%). Other boundary pitch movements in Japanese include L%, H%, HL%, LH%, and HLH% (Maekawa, Kikuchi and Igarashi, et al. 2002); the experiments in this dissertation only include LH% question intonation. This was by design, so that we could concurrently measure kinematic dimensions, such as duration and displacement of constriction gestures in test words, as well as the tonal dimension, by measuring the timing of tonal events.

At the word prosody level, lexical pitch accents are marked by a sharp fall in F0 on the mora following the accented one. Lexical pitch accent is privative and syntagmatic, i.e., it contrasts (i) presence vs. absence and (ii) location of accent. The pitch accent contrasts for a minimal set of two-syllable real words are illustrated in Table 1. Each word in the table is followed by enclitic nominative case particle to illustrate the full pitch contour (the tones for (b) and (c) would be the same (both LH) without the particle). Pitch accent is notated with an asterisk (*) on the preceding syllable. The images for F0 contours were generated using the Prosody Tutor Suzuki-kun tool from the Online Japanese Accent Dictionary (OJAD), in which pitch accent is notated with ˥ symbol and red font. Phonetic correlates of pitch accent are discussed in more detail in Section 1.6. Since the location of pitch accent is lexically specified and thus unpredictable, but its tone patterns are predictable, X-JToBI marks the location of the accent with an 'A', as can be seen in Figure 2 (Maekawa, Kikuchi and Igarashi, et al. 2002).

Table 1. Minimal set of unaccented vs. initial-accented vs. final-accented two-syllable real words.

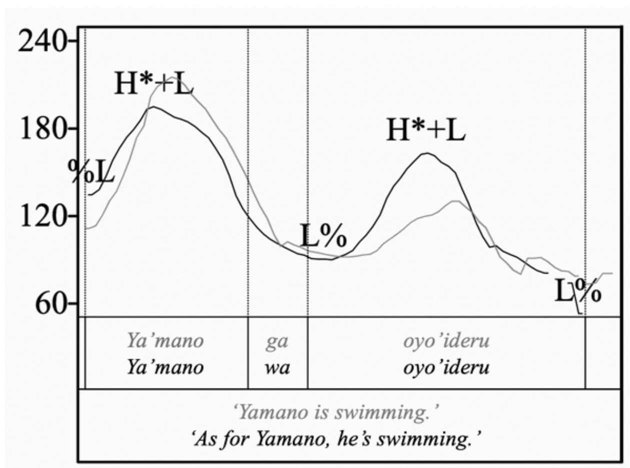
	Word	F0 Contour	Tone	Accent	Gloss
a.	**na***mi+ga	 なみが 奈美が	HLL	initial	'Nami (name) + NOM'
b.	na**mi***+ga	 なみが 波が	LHL	final	'wave+NOM'
c.	nami+ga	 なみが 並が	LHH	unaccented	'medium+NOM'

1.3. Focus marking in Japanese

Given the close relationship between information structure and prosodic structure in Japanese, it is also relevant to discuss how Japanese marks focal prominence. Units under focus are marked by means of local pitch range expansion, prosodic restructuring to mark the focal constituent, post-focal subordination, and prominence-lending boundary pitch movements (Venditti et al. 2008).

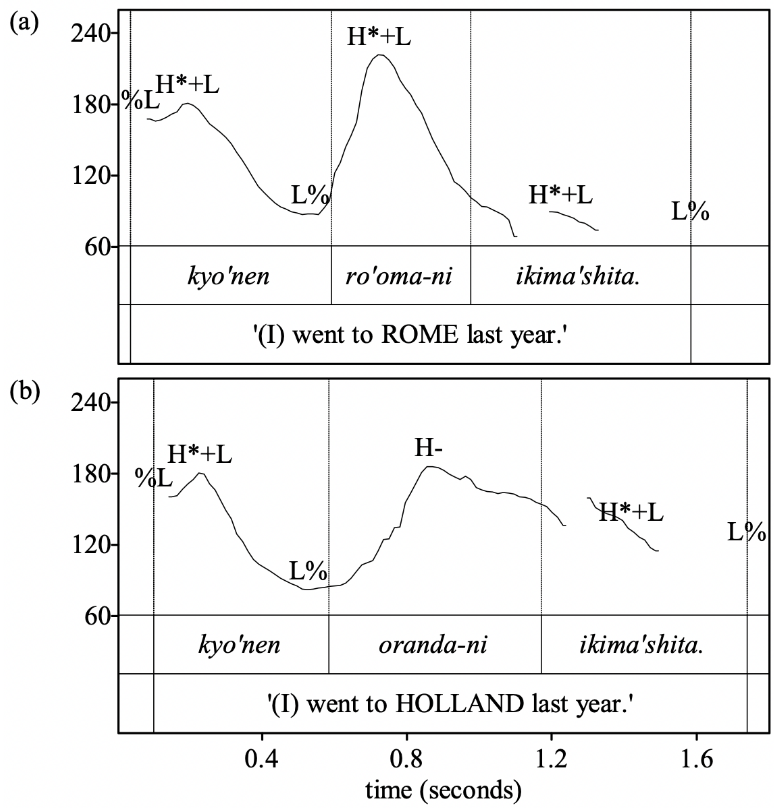
To illustrate pitch range variation in prominence marking, Figure 4 from Venditti et al. (2008, p.9) compares the F0 contours of two utterances that differ only in the nominative *-ga* or topic *-wa* particles. The accent peak in *Ya'mano-ga* (grey line) is higher than the accent peak in *Ya'mano-wa* (black line) because the former is in focus relative to the following predicate. On the other hand, *Ya'mano-wa* elicits the "thematic" interpretation of *wa*, which makes the following verb *oyo'ideru* 'swim' the focal constituent. Whereas the subject in *Yamano-ga* exhibits an expanded pitch range, the same is seen on the verb in *Ya'mano-wa*. The relatively greater difference between F0 peaks in *Ya'mano ga* compared to *Ya'mano wa* also shows the combined effects of downstep and post-focal prosodic subordination. In post-focal subordination, tone targets are realized in the F0 contour, but the pitch range of all words in the post-focal region is extremely reduced.

Figure 4. F0 contours illustrating different pitch ranges between *Ya'mano-ga oyo'ideru* 'Yamano is swimming' (grey line) and *Ya'mano-wa oyo'ideru* 'As for Yamano, he's swimming' (black line). From Venditti et al. (2008, p.9).



Post-focal prosodic subordination also often involves dephrasing, which Pierrehumbert and Beckman (1986) define as a total deletion of the [L% H-] AP-initial rise and prosodic weakening of any post-focal pitch accents. Kori (1997) describes the weakening as a reduced pitch range and a lack of the phrase-initial rise. This dephrasing in effect constitutes a restructuring under focus that inserts a new IP boundary at the beginning of the focused constituent and deletes any following AP-initial [L% H-] rises within the IP. The following example in Figure 5 illustrates post-focal prosodic subordination after narrow focus on an accented word (*ro'oma-ni* 'to Rome) versus an unaccented word (*oranda-ni* 'to Holland).

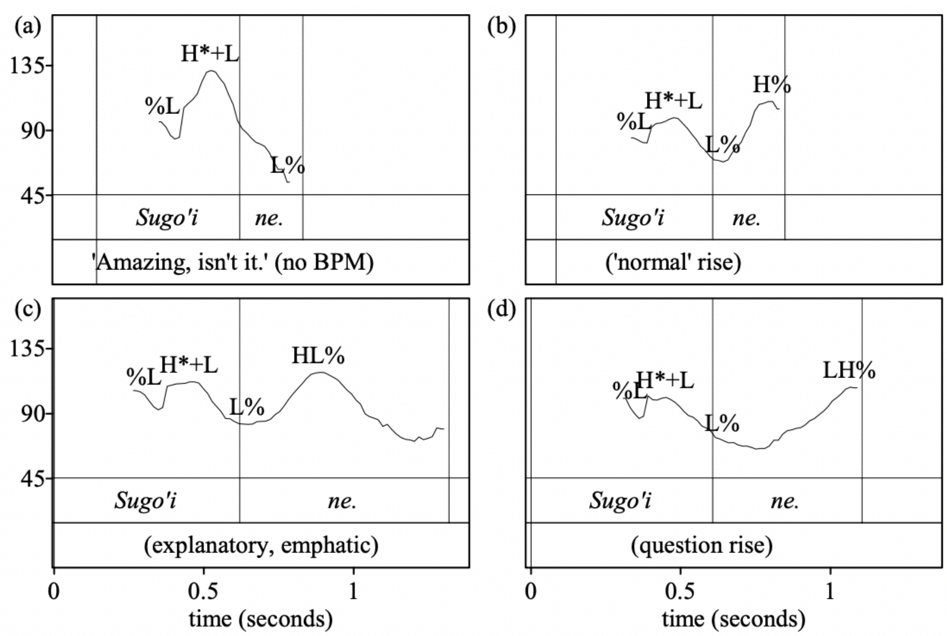
Figure 5. F0 contours of utterances illustrating post-focal prosodic subordination after an early narrow focus on (a) an accented word and (b) an unaccented word in the sentences *kyo'nen ro'oma/oranda ni ikima'shita* 'last.year Rome/Holland to go.PST' '(I) went to Rome/Holland last year' (from Venditti et al 2008, p.25).



Another main element of prominence-lending prosody in Japanese is boundary pitch movements, which are tones following the AP-level boundary [L%] that contribute to the pragmatic interpretation of the utterance (e.g., questioning, continuation, emphasis). There are four main types of boundary pitch movements in Japanese: H%, LH%, HL%, and HLH%. (22) illustrates the F0 contours for these types in the sentence *Sugo'i ne*. 'Amazing, isn't it.'

Boundary pitch movements were previously assumed to only occur sentence-finally (McCawley 1968, Haraguchi 1977, Poser 1984), but examples from the Corpus of Spontaneous Japanese (CSJ) have shown that they can also occur sentence-medially in spontaneous speech (Venditti et al. 2008).

Figure 6. F0 contours for the sentence Sugo'i ne. 'Amazing, isn't it.' Produced with four different boundary tones (from Venditti et. al. 2008, p.15).



According to Kori (1997), post-focal prosodic subordination is the only obligatory prosodic marker of prominence; boundary pitch movements and expansion of pitch range on the focal constituent are optional. Importantly, the literature on focus marking shows that prominence marking in Tokyo Japanese occurs primarily in the tonal dimension.

1.4. Boundary kinematic effects

Previous research has documented strengthening and lengthening effects of prosodic boundaries on the spatio-temporal profile of speech segments (cf. Fletcher 2010 for an overview). This dissertation focuses on boundary marking on gestures for segments at the ends of prosodic phrases, which are known to have larger, longer, and slower constriction gestures than their phrase-medial counterparts (Cho 2006). However, the scope of the effect, or which or how many segments are affected, is unclear. For instance, while many studies find boundary-related lengthening effects in the final syllable, some studies have also found lengthening effects in the penultimate syllable of an intonational phrase or throughout a final foot or word (Fletcher 2010, English: Lehiste 1977, Scott 1982, Turk and Shattuck-Hufnagel 2007; German: Kohler 1983, Kuzla et al., 2007; Hebrew: Berkovits 1994; Dutch: Cambier-Langeveld 2000; Estonian: Krull 1997). This dissertation aims to better understand the scope of lengthening effects, so that we may better understand speech production and the speech planning system, and their interaction with grammar. The scope of phrase-final lengthening can tell us about how the lexical level interacts with the phrasal level. In fact, there is evidence that the scope of boundary effects may interact with word prosody or prominence (Turk and Shattuck-Hufnagel 2007; Katsika 2016). However, most of the literature which explores the scope of boundary lengthening and how it interacts with word prosody has focused on stress languages where word prosody can carry phrasal prominence. Japanese is a language in which we can examine the same question while separating word prosody from phrasal prominence, as the head is marked in the tonal domain but does not carry phrasal prominence in the same manner as stress languages. The following sections give an overview of research on the kinematic effects of prosodic boundaries on speech gestures, the scope of these effects, and their interaction with word prosody.

1.4.1. Boundary marking: acoustic studies

The durational phenomenon of phrase-final lengthening, also known as pre-boundary lengthening, is a robust phenomenon which may be universal (but see Downing and Rialland 2016) and is supported by both acoustic and kinematic data derived from different languages and language varieties (e.g., acoustics: Oller 1973; Berkovits 1993a, 1993b, 1994; Cambier-Langeveld 1997; White 2002; Turk and Shattuck-Hufnagel 2007; articulation: Edwards, Beckman and Fletcher 1991; Beckman and Edwards 1992; Byrd and Saltzman 1998; Byrd 2000; Byrd, Lee, et al. 2005; Byrd, Krivokapić and Lee 2006; Krivokapić 2007; Katsika 2016; Kim, Jang and Cho 2017). The majority of previous research in this area described below has investigated stress languages.

In an acoustic study of American English, Oller (1973) analyzed the effect of position in an utterance (initial, medial, or final) on the duration of speech segments in nonsense words. Regardless of the number of syllables, syllable structure, or stress position in the word, boundary lengthening effects were found in word-final, phrase-final, and utterance-final positions. The strongest effect of final lengthening was on the word-final vowel. An additional effect of stress was also found, as stressed final syllables lengthened more than unstressed final syllables. Boundary lengthening was found to be cumulative (i.e., larger effects at constituents higher in the prosodic hierarchy), with greater lengthening in sentence-final conditions compared to word-final conditions.

Berkovits (1993a, 1993b) also provided evidence of lengthening in phrase-final syllables in Hebrew. These studies examined the distribution of utterance-final lengthening within the final syllable in words with final stress. Berkovits tested disyllabic words ending in fricatives or stops (1993a) and words with open or closed final syllables (1993b).

Regardless of syllable type, the greatest amount of lengthening occurred on segments closest to the boundary, and in some cases, lengthening extended to the penultimate syllable. Berkovits (1994) further tested utterance-final lengthening effects on initially-stressed and finally-stressed disyllabic words and found that this was also the case for unstressed syllables.

Cambier-Langeveld (1997) further explored the domain of phrase-final lengthening and offered evidence from another language. The study examined the role of different boundary types, stress, and syllable structure on the domain of phrase-final lengthening in Dutch and found that the amount of lengthening differed between two main levels of prosodic boundaries, namely prosodic word/phonological phrase vs. intonational phrase/utterance. Final lengthening effects were strongest closest to the boundary (agreeing with Berkovits 1993a, 1993b, 1994) and limited to the final syllable of the word. The domain of lengthening did not interact with boundary strength or stress position, but lengthening extended to the penultimate nucleus when the final syllable was super-light (i.e., when the final vowel was reduced to a schwa).

White (2002) considered different word lengths and the position of stress. This study examined phrase-final boundary lengthening in monosyllabic, disyllabic, and trisyllabic words with initial or final stress in British English. Utterance-final lengthening was found to affect stressed syllable codas in all positions (antepenultimate, penultimate, ultimate), as well as any following unstressed syllables. As with previous studies, the lengthening effect was found to be greater closer to the boundary. Word length also affected the amount of lengthening, as stressed syllables were lengthened less in a disyllable than a monosyllable, and even less in a trisyllable.

Turk and Shattuck-Hufnagel (2007) investigated the interaction between lexical prominence and the scope of phrase-final lengthening in American English by looking at various word lengths (1-4 syllables) and stress position (antepenultimate, penultimate, ultimate). They found that the phrase-final lengthening occurred reliably on the rime of the final syllable and the stressed syllable. Notably, they found that while lengthening was strongest in the final syllable, the effect did not operate in one continuous domain as lengthening skipped over the penultimate syllable in three-syllable words with antepenultimate stress.

In comparison, there is limited work on phrase-final lengthening in Japanese. An acoustic study by Shepherd (2008) examining phonemically short and long vowels in unaccented Japanese words found that immediately preboundary moras showed proportionally greater lengthening in short vowels than long vowels, suggesting that the temporal scope of lengthening in Japanese is one mora. However, more recent acoustic studies by Seo, Kim, et al. (2019a, 2019b) suggest that the scope of boundary lengthening in Japanese may begin earlier. These acoustic studies looked at two-syllable words with light (CV) and heavy syllables (CVN) and also introduced pitch accent as a factor to examine the interaction of word prosody with the scope of phrase-final lengthening in Japanese. While these studies did not investigate lengthening as a function of pitch accent position, their comparison of initial-accented and unaccented words found that accented words underwent less lengthening in the final rime compared to unaccented words. Seo, Kim, et al. (2019b) also found that the scope of lengthening was better accounted for by syllable structure rather than moraic structure, as lengthening on final CV syllables was comparable to that of the combined effect of lengthening in the rime of a final CVN syllable. Overall, lengthening was greatest closest to the boundary and extended up to the penultimate vowel in all accent conditions, regardless of mora count. It should be noted, however, that the stimuli used in these studies differ slightly from those in the current study, which have LH% boundary tones at the IP boundary. The relationship between the effects of phrase-final lengthening and boundary tone marking will be discussed further in Chapters 4, 5, and 7.

1.4.2. Boundary marking: articulatory studies

Articulatory studies have found similar patterns at phrase edges, showing that gestures become spatially larger, temporally longer, and less overlapped at prosodic boundaries (Cho and Keating 2001; Fougeron 2001; Fougeron and Keating 1997; Keating, et al. 2004; Byrd and Saltzman 2003; Byrd 2000; Byrd, Kaun, et al. 2000; Byrd and Saltzman 1998). The spatio-temporal effects of prosodic boundaries have also been found to increase with boundary strength, such as higher linguo-palatal contact with higher prosodic boundaries (Byrd 2000; Byrd and Saltzman 1998; Cho 2006; Tabain 2003; Tabain and Perrier 2005). This occurs together with lower gesture overlap at lower prosodic boundaries (Byrd 2000, Byrd and Choi 2010; Cho 2004).

In one articulatory study, Fougeron and Keating (1997) used electropalatography data to examine linguopalatal contact (contact between the tongue blade/front and the hard palate) at initial and final domain edges in American English. They found that most of the prosodic domains tested were delimited by strengthened articulations and that these effects were cumulative (increasing with boundary strength). Relevant to our discussion of final boundary marking, Fougeron and Keating (1997) found that the final edge of prosodic domains was marked by lengthening and decreasing contact for the vowel, a pattern which held regardless of lexical stress position.

Another articulatory study of boundary effects was Cho (2008), which used Electromagnetic Midsagittal Articulography (EMMA) data to look at V-to-V lingual movement in American English CV#CV contexts across various prosodic boundaries (word vs. IP). Cho found that both boundary and accent were marked by prosodic strengthening through spatial and temporal expansion. Vowel articulation was larger (greater displacement), slower (lower peak velocity), and longer (longer duration) at larger prosodic boundaries, and these patterns did not vary as a factor of post-boundary accent.

An articulatory study which used EMMA data to look at consonant gestures in boundary-adjacent constrictions was Byrd and Saltzman (1998). They analyzed labial /m/ gestures in /CVCV # CVCV/ contexts across five different boundary types (syllable, word, list, vocative, utterance). Lengthening of post-boundary closing movements and, to a lesser degree, phrase-final boundary opening movements were differentiated on three levels.

Another electromagnetic articulography study analyzing the scope of kinematic effects of prosodic boundaries in American English was Byrd, Krivokapić and Lee (2006) which looked at consonant constriction gestures in the word 'dodo.' All speakers showed a significant and large effect of phrase-final lengthening in the opening movement of the boundary-adjacent consonant (C1). Two speakers also showed longer time-to-peak-velocity for C1 opening movement. Other phrase-final spatial effects (displacement, maximum constriction) were inconsistent.

Byrd and Riggs (2008) investigated the scope of kinematic boundary effects using articulatory EMA data. Importantly, they looked at how prominence (stress and pitch accent) interacts with the scope of prosodic lengthening and examined consonant constrictions up to three syllables away from the boundary. In accordance with previous acoustic studies, phrase-final lengthening occurred in the consonant coda gestures immediately preceding the boundary. One speaker also showed a lengthening effect on the consonant onset of the final syllable, when the prominent syllable (stressed and pitch accented) was one syllable away from the boundary. Interactions between phrase-final lengthening and prominence appeared to vary by speaker.

Riggs and Byrd (2011) similarly examined the effect of prominence position (stress and pitch accent combined) on the scope of phrase-final lengthening. Acoustic and articulatory measurements were taken. Effects of phrase-final lengthening were most systematic immediately adjacent to the phrase boundary, but prominence could attract phrasal lengthening; one subject showed lengthening in a pitch accent syllable which was located three syllables away from the boundary. In cases where the remote prominent syllable was lengthened, gestures in-between the phrase edge and the prominent syllable also showed phrasal lengthening.

Katsika (2016) examined the scope of temporal boundary effects in an EMA study of Greek. Notably, this study experimentally tested the effects of lexical stress and pitch accent on boundary lengthening separately. Phrase-final lengthening was found to be initiated earlier in words with non-final stress, and these effects held regardless of accentuation.

1.5. Boundary tone coordination

While phrase-final boundary lengthening and boundary tones have mainly been considered independently, new research raises the possibility of parallelism between the two with regards to their interaction with word-level prosody. Katsika, Krivokapić, et al. (2014) conducted a study of boundary tone coordination in Greek, on the timing of tonal events with the constriction movements that make up the consonants and vowels of the speech signal, and found that boundary tone coordination and phrase-final lengthening effects were affected similarly by the position of stress in phrase-final words. To further investigate this relationship between tonal and temporal events, we look at boundary tone coordination in Japanese.

A relevant, widely used, term is tonal alignment, which refers to the timing of tonal events with acoustic events pertaining to segment production (e.g., the acoustic onset of a vowel). Previous research has found systematic patterns of tonal alignment, with phrasal tones, better approximated with F0 'turning points' (F0 minima and maxima), being coordinated in time with specific acoustic events in the segmental string (Ladd 2008). Within this literature, boundary tones, i.e., the tones that mark major phrase boundaries and which are the topic of focus here, are typically expected to occur within the phrase-final syllable (e.g., Silverman, Beckman, et al. 1992; Beckman and Pierrehumbert 1986).

A well-established finding in tonal alignment studies is tonal crowding, where tones are "too close together" and systematic adjustments occur. For instance, in cases of co-occurrence in the same syllable, tonal events may be altered in language-specific ways, such as compression or truncation (see Ladd 2008 for an overview). The alignment of boundary tones has received less attention; there is little data on the coordination of constriction movements with phrasal tones in general, and with boundary tones specifically. Previous work using Greek data reported that boundary tones were coordinated anti-phase (sequentially) with the vowel constriction gesture of the phrase-final syllable (Katsika 2016). Moreover, Katsika, Krivokapić, et al. (2014) was the first study to experimentally examine the effect of stress and pitch accent on timing of boundary tones. In an EMA study of Greek, their results for boundary tone coordination paralleled that of boundary lengthening; boundary tones were initiated earlier in words with non-final stress (see also evidence from tonal alignment in

Arvaniti and Ladd (2009) as well as Lickley, Schepman and Ladd (2005), Barnes, et al. (2006), Grice, Ladd and Arvaniti (2000)). Interestingly, boundary tone coordination was affected by lexical stress, but not phrasal pitch accent. These patterns held for both accented and de-accented phrase-final words, making a tonal crowding interpretation less plausible.

In this dissertation, we turn to Tokyo Japanese, a language which marks phrase heads and edges with tones, which could distinctively give rise to tonal crowding conditions at prosodic boundaries. We ask the following questions: (i) how are boundary tones coordinated with constriction gestures in Japanese, and (ii) how is this coordination affected by the position of lexical pitch accent while keeping phrasal prominence conditions constant? The specific hypotheses that emerge from these questions are listed in Section 2.4, after we present the theoretical framework of this work in Chapter 2.

1.6. Lexical pitch accent

The word prosodic system of Japanese is known as a "pitch accent" or "non-stress" system. We use pitch accent as a term to refer to the location of the pitch fall, and not based on the claim that pitch accent languages are a distinct typological category (see Hyman 2009 for more discussion). The following sections describe the phonetic correlates of lexical pitch accent in Japanese.

1.6.1. Primary cue

While in stress languages such as English word accent commonly has phonetic correlates such as duration or intensity, the primary correlate of Japanese accent is pitch, or fundamental frequency (F0) (Beckman 1986; Weitzman 1970). Pitch accent in Japanese is marked by a steep fall in F0, which is realized on the mora following the accented one. Pitch accent is culminative (every lexical word has at most one accent) but not obligatory (not every word is required to have at least one accent) (Hyman 2009). While the pitch accent systems of Japanese dialects are many and vary in their prosodic properties (see Haraguchi 1968, Kubozono 2011, 2012, Igarashi 2012), the current study focuses only on Tokyo Japanese.

In Tokyo Japanese, the number of possible pitch contrasts is n+1, where n is the number of syllables in the word. For example, in two-syllable words, there are three possible contrasts: initial-accented, final-accented, and unaccented words. Refer to Table 1 for illustrates of F0 contours for such a minimal set.

Since pitch accent is realized as a fall in F0 on the following mora, the tonal contour of the word is predictable based on the position of pitch accent in the word. Consequently, Japanese pitch accent has been analyzed both tonally and accentually (Hyman 2009). Figure 7 compares different approaches researchers have taken in analyzing Japanese pitch accent. Accentual models typically mark accent with a downstep arrow or A to indicate a pitch drop from H to L. Tonal models associate accented syllables to the H of an HL (or H*+L) sequence to produce an HL pitch drop. Notice that without a mora following the accented syllable, one would not be able to differentiate final-accented from unaccented words based on the pitch, as the tones associated with them are both LHH. The fact that final-accented and unaccented words are not distinguishable in isolation because they are both LHH pitch contour has led some researchers to argue that there is no difference (Kindaichi 1947; McCawley 1968). Other studies predict different surface representations for final-accented and unaccented words (Pierrehumbert and Beckman 1988). The approximate pitches of each syllable in Figure 7c reflect observations and findings by previous researchers that the phrasal H tone (associated with the second mora of the AP) is lower than the H of a pitch accent (Warner 1997; Matsui and Hwang 2018; Poser 1984; Pierrehumbert and Beckman 1984; Beckman and Pierrehumbert 1988; Venditti 2005). In short, there is some debate whether final-accented and unaccented words are different within the word itself.

Figure 7. Summary of accentual and tonal analyses of Tokyo Japanese pitch accent (from Hyman 2009, p. 218).

		'pillow' + NOM	'heart' + NOM	'head' + NOM	'fish' + NOM
a)	accentual	ma↓kura ga	koko↓ro ga	atama↓ ga	sakana ga
b)	tonal	makura ga │ H L	kokoro ga │ H L	atama ga │ H L	sakana ga
c)	approx. pitch	mákùrà gà	kòkóró gà	àtāmá gà	sàkānā gā

1.6.2. Secondary cues

While F0 is considered the most clear and consistent correlate of pitch accent in Japanese, studies on pitch accent in devoiced or whispered speech suggested the existence of secondary cues (Sugito, Higashikawa, et al. 1991; Sugiyama 2012; Sugiyama 2017). However, several studies have searched for correlations of duration or intensity with Japanese pitch accent and only found weak or inconsistent correlations (Beckman 1986, Cutler and Otake 1999; Kaiki, Takeda and Sagisaka 1992; Weitzman, 1970). Neustupný (1966) examined intensity and F0 of words produced in isolation and found that whereas in some words, accented moras showed a rise in F0 but not in intensity, in some other words, accented moras showed no rise in F0 but did in intensity. Neustupny concluded that accent in Japanese is realized by some inconsistent set of interacting features, which included F0 and intensity. Neustupný (1978) replicated their 1966 results with monomoraic and bimoraic minimal pairs of final- and unaccented words, and found a distinction between the two. However, findings form Weitzman (1970) and Beckman (1986) suggest that pitch accent and intensity are not so strongly correlated. Weitzman (1970) analyzed the intensity of six bimoraic words and controlled for vowel quality. They used six different measures for intensity and found that intensity was not consistently correlated with accent patterns, and was therefore only a secondary cue to F0. Beckman (1986) examined F0, duration, and amplitude of six minimal pairs of disyllabic words and also found that Japanese accent patterns did not have significant corelates in peak or average vowel intensity. Similarly, accented syllables in Japanese did not correlate with longer durations.

Articulatory studies have also investigated correlations between jaw displacement and pitch accent in Japanese. Jaw displacement is well correlated with stress in English, where stressed syllables show larger jaw opening (e.g., Erickson, Suemitsu, et al. 2012; Erickson, Kawahara, et al. 2014a; Erickson, Kawahara and Williams, et al. 2014b; Kelso, et al. 1985; Menezes and Erickson 2013; Vatikiotis-Bateson and Kelso 1993). Vatikiotis-Bateson and Kelso (1993) found that H-toned syllables in Japanese had smaller jaw displacement than L-toned syllables but did not specifically investigate the effect of pitch accent. Kawahara, et al. (2014a) conducted an EMA study of Japanese metrical structure and tested the effect of pitch accent on jaw displacement with a minimal set of stimuli that included unaccented, initial-accented, and final-accented two-syllable words. They found that there were no substantial differences for any speaker and concluded that there was no correlation between pitch accent and jaw displacement.

1.7. Research questions and predictions

This dissertation has two main goals: (i) to examine how phrase boundaries are marked in Japanese, especially with respect to lexical pitch accent, and, in order to further understand boundary marking, (ii) to investigate what, if any, are the kinematic correlates of pitch accent. To address these topics, our research questions are as follows:

i. What is the scope of phrase-final lengthening in Japanese? Is the scope affected by lexical pitch accent?

ii. What is the coordination of boundary tones in Japanese? Is coordination affected by lexical pitch accent?

iii. What are the kinematic correlates of pitch accent?

On the scope of phrase-final lengthening, one hypothesis is that the effects occur mainly in the final mora, as suggested by previous acoustic work on unaccented Japanese words (Shepherd 2008). An alternate hypothesis is that effects will begin in the penultimate syllable, following the results of previous acoustic studies which tested unaccented and initial-accented two-syllable words in Japanese (Seo, Kim, et al. 2019a, 2019b). The stimuli used in Shepherd (2008) and Seo, Kim, et al. (2019a, 2019b) suggest that moraic structure may have an effect on the scope of phrase-final lengthening effects, as Shepherd tested two-syllable words with a long vowel in the initial syllable, and short or long vowel in the second syllable (CVV.CV(V)). On the other hand, Seo, Kim, et al. (2019a, 2019b) tested the effect of moraic structure with heavy syllables with coda consonants (CVN.CV(N)) and found that the presence of a moraic N coda in the final syllable did not appear to have a significant effect on the scope of boundary-related lengthening, suggesting that the patterns in scope are better described by syllable structure rather than moraic structure. While the role of moraic structure in phrase-final lengthening in Japanese merits further research, it is not the focus of this dissertation, which has intentially selected test words consisting only of light syllables (CV.CV.(CV)) to control for possible effects of moraic structure.

As duration is not typically considered a correlate of pitch accent in Japanese (Weitzman 1970; Poser 1984; Beckman 1986; Pierrehumbert and Beckman 1988; Sugito 1998), it is unlikely that Japanese would exhibit a pattern similar to English, where the prominent syllable is lengthened as a function of boundary (e.g., Turk and Shattuck-Hufnagel 2007, but see Byrd and Riggs 2008). Previous studies also suggest that pitch accent may affect phrase-final lengthening in Japanese such that accented words undergo less lengthening in the final syllable compared to unaccented words (Seo, Kim, et al. 2019a, 2019b), although to our knowledge there are no previous studies examining these boundary effects as a function of lexical pitch accent position. However, it is important to note that these studies did not test stimuli in utterance-final position. Nevertheless, if pitch accent is orthogonal to stress, it is possible that in Japanese, we may find an interaction similar to that of Greek where the initiation of boundary lengthening varies as a function of pitch accent position, and we could expect boundary effects to be initiated earlier when pitch accent is earlier (Katsika 2016).

Regarding boundary tone coordination, we expect boundary tones to occur within the phrase-final syllable (e.g., Silverman, Beckman, et al. 1992; Beckman and Pierrehumbert 1986). We also expect an interaction between boundary tone timing and lexical pitch accent, with tones occurring later with respect to the phrase-final vowel in words with penultimate pitch accent than in words with antepenultimate pitch accent. This expectation arises both from previous effects of stress, another word prosodic marker, on boundary tone coordination (Katsika, Krivokapić, et al. 2014; see also Arvaniti and Ladd 2009), and from the tonal crowding hypothesis. No effect is hypothesized in words with ultimate pitch accent, since the latter is expected to delete phrase-finally (Matsui and Hwang 2018).

Lastly, if lexical pitch accent is found to interact with the timing of boundary lengthening and boundary tones, it follows that we should investigate whether there are any secondary cues to pitch accent in our kinematic data. While lexical stress has been found to interact with boundary lengthening, it is possible that this interaction arises because they both act in the same temporal dimension. However, most studies examining pitch accent and intensity or displacement have found weak or non-existent correlations, so we do not expect to find any robust correlation between duration or displacement with pitch accent. If there is such a correlation, we would expect syllables with pitch accent to be longer or larger than unaccented syllables (Beckman 1986; Cutler and Otake 1999; Kaiki et al., 1992; Weitzman 1970).

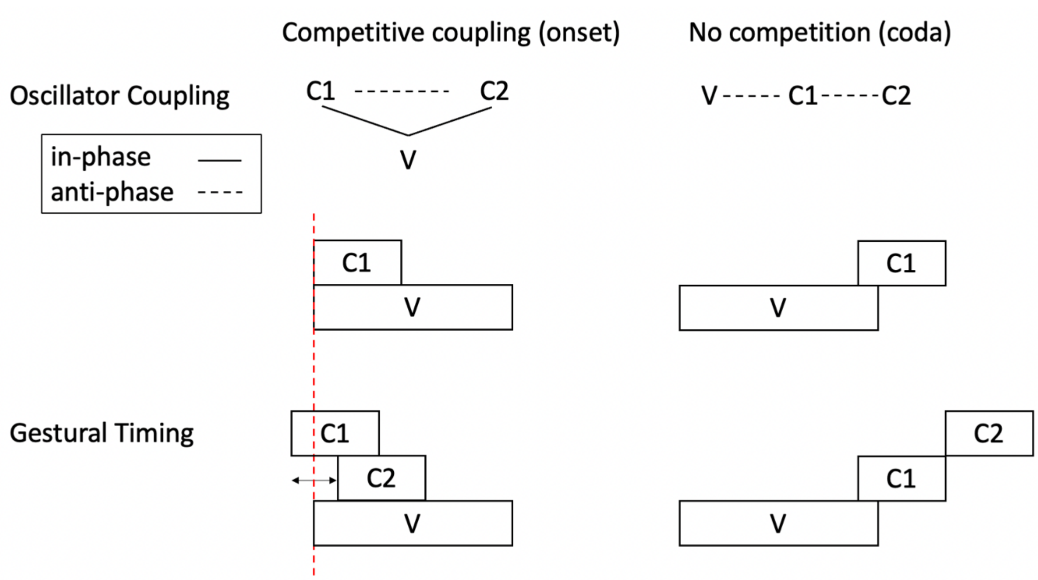
2. Articulatory Phonology

The theoretical framework for this dissertation is Articulatory Phonology (Browman and Goldstein 1986, 1989, 1990, 1991, 1995, 2000), a formal theory of phonology in which there are three types of gestures: (i) constriction gestures, (ii) tone gestures, and (iii) clock-slowing gestures. Like constriction gestures, which make up consonants and vowels, temporal and tonal events are also represented as phonological units known as gestures.

2.1. Constriction gestures

Constriction gestures form or release constrictions in the vocal tract. They take place across time and space and therefore have spatio-temporal properties. They are specified for abstract linguistic tasks and triggered by internal oscillators, which are coupled with each other. In Articulatory Phonology, consonants and vowels are made up of constriction gestures, where various articulators such as the lips, tongue, velum, and glottis form and release constrictions within the vocal tract. These gestural units are coordinated with each other to form consonants and vowels, which in turn are units of phonological contrast. In this way, consonants and vowels are considered constellations of consonant and vowel constriction gestures. These gestures are defined using the Task Dynamics model in which each gesture controls a specific articulator, which is specified for constriction degree and constriction location (Saltzman 1986; Browman and Goldstein 1989). These gestures are inherently spatiotemporal, unfolding over time and space. Their temporal relationship is modeled by dynamic coupling, where gestures are coordinated with each other and their initiation is controlled by planning oscillators or clocks which have two simple modes: in-phase (0°, where movements are activated synchronously) and anti-phase (180°, where movements are activated sequentially) (Saltzman and Byrd 2000; Nam and Saltzman 2003). For example, as Figure 8 illustrates, the C and V gestures in a CV syllable are coordinated in-phase, where the associated gestures for the consonant and vowel are triggered simultaneously. In contrast, the V and coda C gestures in a CVC syllable are triggered sequentially. In a CCV syllable with a complex onset, the planning oscillators for each consonant is in-phase with the vowel gesture but anti-phase with the other consonant gesture. This results in a competitive coupling relationship where the onset consonant gestures shift in relation to the vowel gesture. This is known as the c-center effect, where the timing of consonants in a complex onset are timed relative to the vowel gesture onset. In complex onsets, the anti-phase coordination between the two C gestures pushes the first C gesture earlier and the second C gesture later, resulting in both consonant onset gestures being centered around the V gesture.

Figure 8. In-phase and anti-phase oscillator coupling relationships. Adapted from Krivokapic 2020.



2.2. Tone gestures

As with constriction gestures, tonal events are represented in Articulatory Phonology as tone gestures. Tone gestures act on articulators in the larynx such as the thyroarytenoid, cricoarytenoid, and cricothyroid muscles (Hirose 2010), and are also associated with planning oscillators. Note that tone gestures for pitch targets in a speaker's lower pitch range also involve extra-laryngeal muscles (Honda, et al. 1999; Honda 2004). How tone gestures interact with global pitch range control is not understood (Erickson, Honda, et al. 1994; Fuchs, et al. 2015), and have yet to be fully incorporated into the theory of Articulatory Phonology. Tone coordination is determined by in-phase and anti-phase coupling mechanisms, and tone gestures are coordinated with C and V constriction gestures. For example, Gao (2008) applied this gestural approach to lexical tones. Gao proposed two tone gestures (an H and an L target) to model the four lexical tones of Mandarin, with tones 2 (rising) and 4 (falling) made up of a combination of two tonal targets. Similar to constriction gestures, tone coordination is determined by in-phase and anti-phase coupling mechanisms. In tone 2, the L and H gestures are coordinated in-phase to produce an overlap of L and H that results in a mid-level F0 followed by a rise to H. In tone 4, the H and L gestures are coordinated anti-phase to produce a sequential H to L falling tone contour. Figure 9 shows the proposed coordination of tone gestures with consonant and vowel constriction gestures for lexical tones in Mandarin, and Figure 10 provides an illustration of their relative timing in a gestural score.

Figure 9. Coupling graphs of Tones 1-4 from Gao 2008 (p. 95-96).

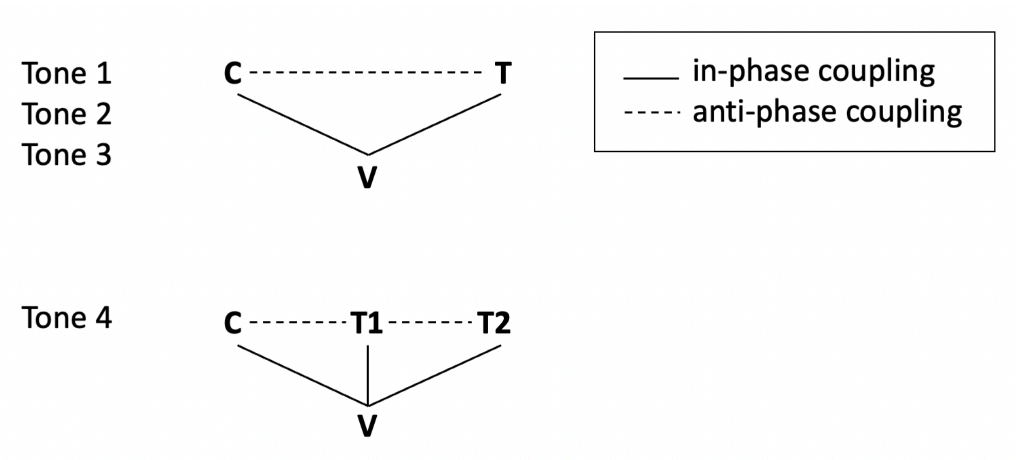
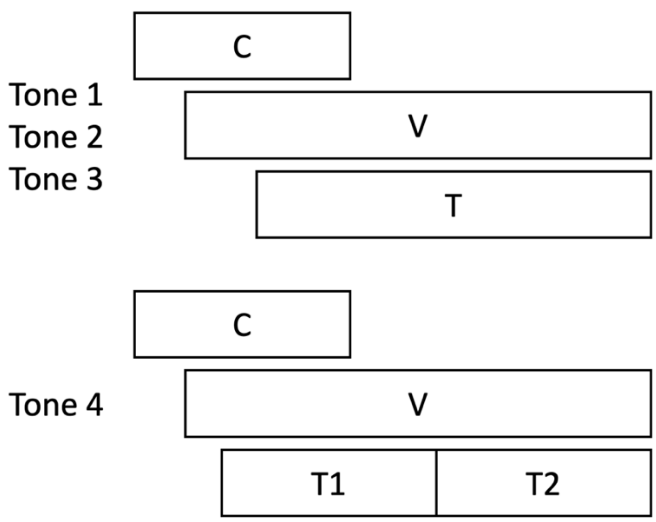
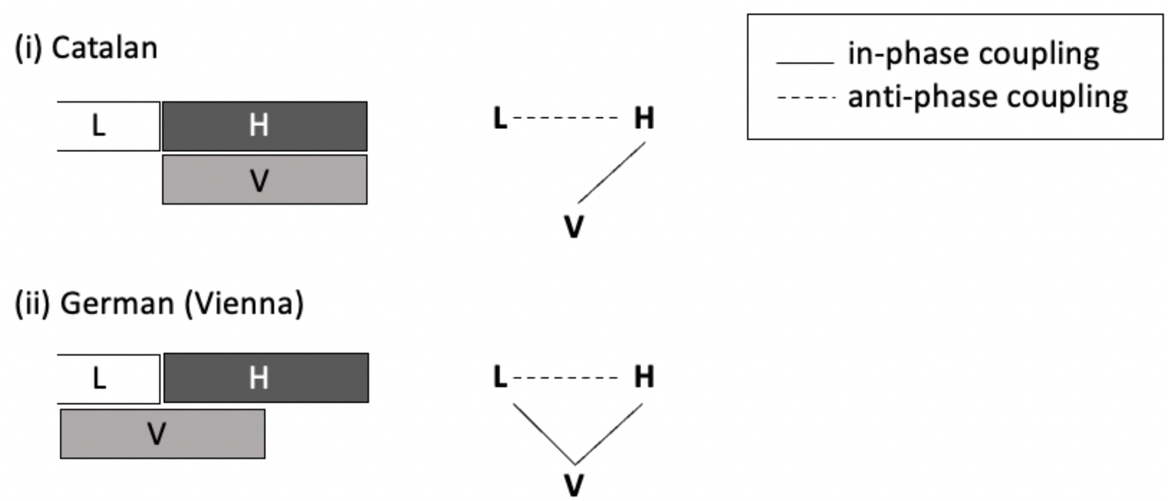


Figure 10. Gestural scores of CV syllables adapted from Gao (2008).



Tone gestures have also been proposed for post-lexical pitch accents. Mücke et al. (2012) examined coordination of rising (LH) nuclear pitch accents in Catalan and Viennese German with vowel and consonant gestures in a kinematic, acoustic, and modeling study. They found that in Catalan, the H gesture is in-phase coordinated with the vowel and anti-phase coordinated with the L gesture (see Figure 11). In Viennese German, the L and H are in-phase coordinated with the vowel and anti-phase coordinated with each other. This leads to competitive coupling and shifts the L gesture leftwards and the H gesture rightwards. This alignment may also differ based on the language variety, as the alignment of the rise in German varies across regional varieties.

Figure 11. Gestural score and coupling graphs for (i) Catalan and (ii) Viennese German. Coupling graphs show solid lines for in-phase and dotted lines for anti-phase target specifications. From Mücke, et al. (2012, p.20).



Notably, Mücke, et al. (2012) compares the results of post-lexical pitch accent tones in Viennese German and Catalan to the lexically-determined tones of Mandarin and observes that "when a Tone gesture is added to a syllable in Catalan it produces no effect on the C-V coordination, while in Mandarin a Tone gesture causes the C to precede the V, because of the competitive syllable structure" (p. 21-22). The competitive syllable structure mentioned here refers to the c-center effect, where a competitive coupling relationship results in a shift in gesture timing. So, the main difference between the two types of tonal coordination appears to be that lexical tone gestures in Mandarin are fully integrated into the coupling relations of the syllable, whereas post-lexical tone gestures in Catalan are coupled to V gestures without affecting coupling relations within the syllable.

Katsika, Krivokapić, et al. (2014) and Katsika (2016) considered the coordination of phrasal boundary tone gestures with constriction gestures and put forward a different proposal in which phrasal tones are indirectly connected to C and V constriction gestures by being activated through a third type of gesture, clock-slowing gestures (see Figure 15).

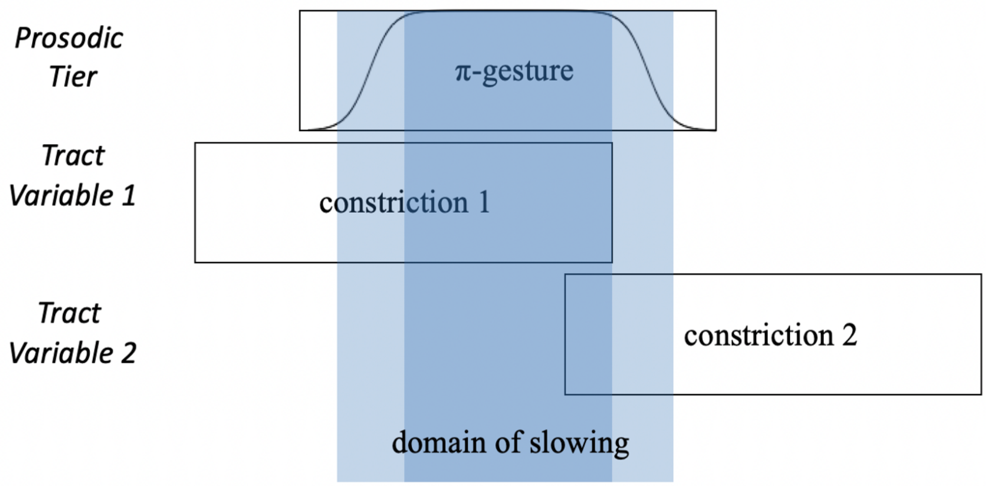
Tone gestures for pitch targets in a speaker's lower pitch range also involve extra-laryngeal muscles (Honda, et al. 1999; Honda 2004), and given that how tone gestures interact with global pitch range control is not understood (Erickson, Honda, et al. 1994; Fuchs, et al. 2015), these aspects of tone have yet to be fully incorporated into the theory of Articulatory Phonology.

2.3. Clock-slowing gestures

Like other types of gestures, clock-slowing gestures also extend over a temporal interval and overlap with constriction gestures. Unlike other types of gestures which control specific articulators, these gestures affect the temporal unfolding of constriction gestures by slowing the local "clock-rate" that controls them. Clock-slowing gestures are realized through their local slowing effect on concurrently active constriction gestures. They modulate the "clock-rate" that controls the temporal unfolding of speech gestures. One type of clock-slowing gesture is called the π-gesture, or prosodic gesture, which instantiates prosodic phrase boundaries; π-gestures are realized via their effect on the constriction dynamics, which locally slows all concurrently active constriction gestures (Byrd and Saltzman 2003). Another term for clock-slowing gestures is μ-gestures, for modulation gestures, because they can modulate the spatial or temporal activation of co-active gestures. Technically, this means all clockclosing gestures are modulation gestures. However, to make the distinction between prominence and boundaries, we will use the term π-gestures to refer to boundary-related gestures, and μ-gestures to refer to other prominence-related gestures. An example of a non-boundary related gesture would be a spatial or temporal μ-gesture associated with stressed syllables, which results in larger or longer constriction gestures.

The π-gesture captures the temporal effects of prosodic boundaries. The schematic figure presented in Figure 12 illustrates the activation of the π-gesture at a phrase boundary.

Figure 12. A schematic representation of a π-gesture (adapted from Byrd & Saltzman 2003).



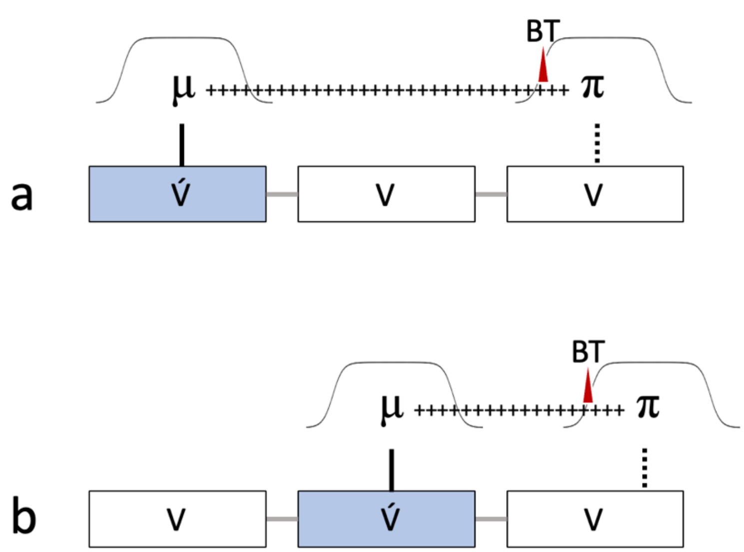
The π-gesture is also hypothesized to have a specific shape, shown in the Prosodic Tier, which shows its level of activation or the strength of its clock-slowing effects. The maximum height of the π-gesture corresponds with the boundary, where effect is the strongest. The end of constriction gesture 1 can be seen to overlap with the π-gesture's area of maximum activation, shown as a dark blue area in Figure 12. A smaller slowing effect further away from the boundary corresponds the light blue area of pi-gesture activation, which overlaps earlier with constriction 1. The overlap of the π-gesture with constriction 2 also shows that the π-gesture also has slowing effects on constriction gestures in the beginning of the following phrase.

Articulatory gestures at phrase edges have been found to undergo local lengthening in the temporal domain (increased duration) (Beckman, Edwards and Fletcher 1992; Byrd and Saltzman 1998; Byrd 2000) and sometimes kinematic slowing (decreased peak velocity) and an increase in spatial magnitude (larger displacement) as well (Cho 2006). These effects are accounted for in Articulatory Phonology by the presence of a π-gesture at the phrase boundary which results in less-overlapped constriction gestures which are slower, longer, and larger at the end of the phrase. This model understands boundaries as gestures, and π-gestures, like other articulatory gestures, have inherent durational properties and overlap with constriction gestures. As the π-gesture's maximum activation is approached at the phrase edge, the degree of slowing gradually increases; as the boundary recedes, the clock-slowing decreases. The clock-slowing effect is strongest closer to the boundary, and the level of activation depends on the strength of the boundary. Prosodic boundaries higher in the hierarchy are expected to have stronger clock-slowing effects. The shape of the π-gesture is also relevant, with maximum activation of the π-gesture occurring at the phrase boundary and a lesser clock-slowing effect at the edge of the π-gesture. This accounts for the progressive effect of phrase-final lengthening found in previous work, with gestures closest to the phrase edge undergoing stronger clock-slowing effects, and thus greater lengthening, than gestures further away from the phrase edge.

The coordination of π-gestures is related to the scope of prosodic effects, as only constriction gestures which overlap with the π-gesture undergo slowing. The amount of slowing, corresponding to the amount of overlap of dark or light blue in Figure 12, also indirectly reflects the π-gesture's coordination with constriction gestures. Therefore, understanding the scope of boundary lengthening effects can help us understand π-gesture coordination, which is not well understood. As mentioned in Chapter 1, in a study of the scope of boundary lengthening in Greek, Katsika (2016) found that lengthening began earlier in words with non-final stress. Based on these findings, a model within the Articulatory Phonology framework was put forth in which the π-gesture's coordination is affected by the stressed syllable, or the μ-gesture associated with the stressed syllable, which results in differences in the scope of lengthening. In the π-gesture framework, there are a couple of alternative proposals to account for the effects of stress on the scope of prosodic lengthening: (a) extension or (b) coordination shift of the π-gesture (Byrd & Riggs, 2008). One possibility is that the π-gesture extends towards the μ-gesture, the clock-slowing gesture that instantiates the lexical stress of the last word, resulting in greater overlap between the π -gesture and constriction gestures and thus a greater scope of effect. Another possibility is that the π-gesture might be attracted towards the μ-gesture, resulting in a shift in its coordination. Shifting the π-gesture closer to the μ-gesture would result in earlier overlap between constriction gestures and the π-gesture, also accounting for lengthening being initiated earlier in the word. A third proposal described in Katsika (2016) is the most recent account for the interaction between stress and boundary event and enriches the previous two accounts. Katsika (2016) proposed that the π-gesture is coordinated with both the final constriction gestures (the final vowel) and the final lexical μ-gesture; this account was further supported with post-boundary evidence and evidence from boundary tone coordination. Katsika, Krivokapić, et al. (2014) examined the role of lexical stress in boundary tone coordination in Greek and found that boundary tones were initiated earlier in words with non-final stress, findings which parallel those of from the

Greek study on phrase-final lengthening in Katsika (2016). The schematic in Figure 13 illustrates the proposed relationship between the μ-gesture, the phrase-final π-gesture, and the initiation of boundary tone. The timing of the π-gesture, and thus the timing of the boundary tone triggered by the π-gesture's activation level, with the constriction gestures in the final syllable is affected by its relationship with the μ-gesture associated with the lexically stressed syllable. In this way, the position of lexical stress in the phrase-final word may affect temporal and tonal boundary events.

Figure 13. A schematic representation of π-gestures and their coordination with final- and medial-stressed words. Figure adapted from Katsika (2016). Dotted lines indicate anti-phase coordination; solid black lines indicate in-phase coordination. Lines of crosses indicate uncertain types of coordination. Stress is indicated on the antepenultimate or penultimate syllable with by '´' and blue shading. Boundary tone initiation is indicated by the red triangle.



2.4. Research questions and predictions

This dissertation aims to expand on this work by examining the relationship between boundary events and word prosody in Japanese. While π-gestures can account for the phonetic findings, it is unclear how they are coordinated with other gestures. Understanding the scope of temporal and tonal boundary effects can help us understand this coordination. Findings from previous studies suggest that word prosody plays a role in the coordination of boundary events in stress languages; however, it is unclear whether it is word prosody or phrase-level prominence which drives this coordination. Japanese can help separate these issues, as the word prosodic system is not associated with phrase-level prominence in the same way as in stress languages.

Chapter 3 outlines the common methodologies for the following studies presented in Chapters 4-6. The general background and procedures regarding participants, experimental design, data collection, and analysis are described.

Chapter 4 addresses the question of whether or not word prosody affects the scope of boundary-related lengthening; in other words, does lexical pitch accent affect the coordination of π-gestures in Japanese? It is clear that lexical stress, which is associated with a μ-gesture, has an effect on π-gesture coordination in Greek. While typological variation of μ-gestures may also play a role in the interaction between word prosody and phrase-final lengthening, it remains to be seen whether lexical pitch accent, which may or may not be associated with a μ-gesture, also has an effect on π-gesture coordination. If, like stress, lexical pitch accent does affect π-gesture coordination, we expect to see an effect of pitch accent position on the scope of lengthening on constriction gestures where boundary lengthening begins earlier in the word when pitch accent is earlier in the word (Katsika 2016). Alternatively, we may find that like

Seo, Kim, et al. (2019a, 2019b), lengthening effects begin in the penultimate syllable, or that effects occur mainly in the final rime (Shepherd 2008).

Chapter 5 looks at boundary tone coordination, and, following Katsika, Krivokapić, et al. (2014) our first hypothesis is that boundary tone (BT) gestures in Japanese are anti-phase coordinated with the phrase-final vowel (V) gesture and have no coordination with the consonant (C) gesture. If this is the case, we expect to see the onset of BT gestures in synchronicity with the target of the V gesture. A competing hypothesis is that BT gestures are in-phase coordinated with the phrase-final V gesture, but do not interfere with the C-V coordination of this syllable, i.e., do not show the c-center effect, per Mücke, et al. (2012). The onset of BT gestures would occur synchronously with the onset of V gestures. A third hypothesis based on findings from Gao (2008) is that BT gestures are in-phase with V gestures but anti-phase with C gestures, creating a c-center effect. In other words, the c-center between the C and BT gestures would be in synchronicity with the onset of V gestures. Finally, based on findings on phrasal pitch accents (d'Imperio, et al. 2007), a fourth hypothesis is that BT gestures are initiated synchronously with the time of peak velocity of one of the phrase-final constriction gestures.

Chapter 6 takes a closer look at pitch accent per se. Finding a similar effect of pitch accent on π-gesture and BT coordination in the previous chapters would suggest that pitch accents are associated with μ-gestures, just as μ-gestures are associated with lexical stress, and that it is this μ-gesture that is coordinated with π-gestures and BT gestures. If this is the case, we expect to find evidence of a μ-gesture through the observation of modulation effects on co-active constriction gestures. In other words, we expect to see longer or larger constriction gestures in pitch accented syllables.

Finally, Chapter 7 concludes with a general discussion and provides an account for our results within the theoretical framework of Articulatory Phonology.

3. Methods

This section describes the shared methodology of the dissertation experiments such as data collection, labelling, and analysis. Experiment-specific methods will be described in their corresponding chapters.

3.1. Participants

Participants were 5 female native Tokyo Japanese speakers, all undergraduate students affiliated with University of California Santa Barbara (UCSB) (4) or Santa Barbara Community College (1) at the time of the study. The mean age was 21 years, with a range of 20-23 years old. No speech, hearing, or vision problems were reported, and all participants could read Japanese orthography. In order to control for regional varieties, information on participants' language background was also collected; we confirmed that all participants grew up speaking Japanese in the Kanto region (specifically, participants' hometowns were Tokyo, Saitama, Chiba, or Kanagawa prefectures). Participants were not aware as to the purpose of the experiment and were compensated for their time. The experiments were conducted at the UCSB Phonetics Lab.

3.2. Experimental design

Two experiments were designed with the goal of examining (a) spatio-temporal and (b) tonal boundary events and their interaction with lexical pitch accent position. The first experiment included real words as test stimuli, and the second included neologisms. Neologisms were made-up words which participants were trained to recognize as names of different dolls before the experiment began. The data from these experiments were also designed to allow for a systematic analysis of the kinematic correlates of pitch accent, so the same set of words were analyzed for spatio-temporal and tonal boundary effects.

In order to test the effect of phrase boundaries on speech gestures, test words for all experiments were embedded in frame sentences in a phrase-medial (PhM) control condition or at the end of an IP in a phrase-final (PhF) test condition, as illustrated in Figure 14.. Frame sentences consisted of a *honto ni X [verb.PST]* 'really X [verb.PST]' 'Really X verb-ed?' pattern, in which X is the test word. The verb was varied to produce sensical responses to context sentences, and the length (number of mora) and pitch accent of the verb were controlled. In this way, test utterances for phrase-medial and phrase-final conditions only differed in that phrase-medial test words were followed by a verb and then a phrase boundary *honto ni X [verb.PST]?* 'really X [verb.PST]?' 'Really, X verb-ed?', while phrase-final test words were immediately followed by a phrase boundary, with the verb following in a separate phrase *honto ni X? [verb.PST]* 'really X? [verb.PST]?' 'Really, X? Verb-ed?'. A couple of aspects of the stimuli which we keep in mind for interpretation are that the phrase-final frame puts an emphatic focus (incredulity question intonation) on both the target word and the following verb, while the phrase-medial frame could be putting a single emphatic focus on the following verb. So, it is possible that some of the effects that we document could be the effects of narrow focus. It is also the case that all phrase boundaries in the elicited utterances were associated with LH% boundary tones, which may also play a key role in the interaction between word prosody and boundary effects. All utterance elicitation utilized short dialogues, which were created to prompt frame sentences embedded with the target words to be read aloud as responses to a context sentence read silently. Frame sentences and dialogues were independently checked by two native Japanese speakers for grammaticality and plausibility.

All elicited test sentences were also checked by a native speaker for accurate and fluent production of the test sentences.

Figure 14. Sample context and frame sentences for test word na*mi 'Nami (name)'.

Context (read silently)

	Nami	*ni*	*katta!*
	Nami	DAT	won
	'I won against Nami!'		

Response (read aloud)

Phrase-medial:	[hoɴto: ni	**na*mi**	makasi̥ta?]$_{IP}$
	really	Nami	defeated?
	'Really (you) defeated Nami?'		
Phrase-final:	[hoɴto: ni	**na*mi** ?]$_{IP}$	[makasi̥ta?]$_{IP}$
	really	Nami	defeated?
	'Really Nami? (You) defeated (her)?'		

Test words included two- and three-syllable words (see Table 2). Accent patterns were verified using the Online Japanese Accent Dictionary (OJAD, n.d.). A set of three-syllable neologisms based on Katsika, Krivokapić, et al. (2014) and Katsika (2016) was also used (Table 2). Two-syllable real words and neologisms were minimal sets differing only in pitch accent position. However, no minimal set of three-syllable words within the constraints of stimuli measurable in an EMA experiment could be found, as there are no pairs of final-accented and unaccented words that differ only in pitch accent for three-syllable or longer words in Japanese. Constraints for the experimental design included requirements that words consist of only stop consonants (i.e., no liquids, glides or fricatives) and alternating vowel heights where possible for constriction gesture measurability. Stimuli were also restricted to voiced segments only, in order to measure F0 and boundary tone initiation.

Table 2. Test words for all experiments. Pitch accent is indicated on the preceding syllable with an asterisk (*).

Pitch Accent Position	*Real words*		*Neologisms*
Antepenultimate		**na*mina** 'Namina (name)'	**ma*mima**
Penultimate	**na*mi** 'Nami (name)'	**nana*me/nama*mi** 'slanted/living body'	**mami*ma**
Ultimate	**nami*** 'wave'	**nigami*** 'bitterness'	**mamima***
Unaccented	**nami** 'medium'	**namida** 'Namida (name)'	**mamima**

To compare the effects of phrase boundaries on segments in words of different lengths, the consonant (C) and vowel (V) gestures were coded according to their position from the right edge of the word, with the word-final syllable coded as C0 V0. This coding scheme is illustrated in Figure 15, where CV gestures in the penultimate syllable are labelled C1 V1, being one syllable away from the phrase edge, and so on.

Figure 15. Consonant and vowel coding scheme.

2-syllable words: e.g. *nami* 'Nami (name)'

n a m i

| | | |

C1 V1 C0 V0

3-syllable words: e.g. *namina* 'Namina (name)'

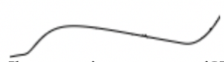
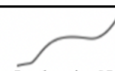
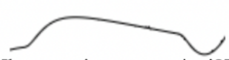
n a m i n a

| | | | | |

C2 V2 C1 V1 C0 V0

In order to allow for testing of boundary tone coordination, boundary tone type was controlled for in all experiments. The higher-level Intonation Phrase (IP) in Japanese is marked by boundary pitch movements at the right edge of the phrase, and all frame sentences for the experiments were produced with question intonation to elicit LH% boundary tones. The verbs in test utterances were varied to make sensible responses, while being controlled for length (number of mora) and pitch accent. Table 3 illustrates the basic tonal contour for all conditions tested. AP-level tones are notated as %L and L%. Pitch accent is notated with H*+L (instead of A) to show the tone targets associated with word accent. The example pitch contours shown here were generated automatically using the Prosody Tutor tool from the Online Japanese Accent Dictionary **(Prosody Tutor Suzuki-kun n.d.)**

Table 3. Tonal contour for test sentences.

	Phrase-medial	Phrase-final
Unaccented	 [honto: ni nami nakusi̥ta?]$_{IP}$ [%L H- L% H- L% H- L% LH%]$_{IP}$ really medium lost.PST 'Really (the) medium (one) was lost?'	  [honto: ni nami?]$_{IP}$ [nakusi̥ta?]$_{IP}$ [%L H- L% H- L% LH%]$_{IP}$ [%L H- L% LH%]$_{IP}$ really medium lost.PST 'Really? (The) medium (one) was lost?'
Accented	 [honto: ni na*mi makasi̥ta?]$_{IP}$ [%L H- L% A L% H- L% LH%]$_{IP}$ really Nami defeat.PST 'Really Nami defeated (them)?'	  [honto: ni na*mi?]$_{IP}$ [makasi̥ta?]$_{IP}$ [%L H- L% A L% LH%]$_{IP}$ [%L H- L% LH%]$_{IP}$ really Nami defeat.PST 'Really Nami? (She) defeated (them)?'

3.3. Procedures

Stimuli were presented on a computer screen placed about 60 inches away from the participant. Prior to recording, a short training session took place to introduce neologisms and their meanings and check that participants could produce correct pitch accents for all stimuli, as neologisms required some practice, and the three-syllable unaccented test word *namida* ‘Namida (name)’ has an initial-accented homophone *na*mida* ‘tears’ which needed to be clarified with participants. All sentences and stimuli were presented in Japanese orthography, which also served to disambiguate some homophones (e.g., 奈美 *na*mi* ‘Nami (name)’, 波 *nami** ‘wave’, 並 *nami* ‘medium’). Neologisms were presented in Hiragana and disambiguated with the ˥ symbol (ま˥みま *ma*mima*, まみ˥ま *mami*ma*, まみま˥ *mamima**, まみま *mamima*). This symbol was chosen to represent pitch accent after consultation with a native speaker. During the training session, participants were trained to recognize this notation, as well as practice the meanings of each neologism, which were presented as names of different dolls. They also received instructions on how to read sentences presented on the computer screen and were directed to read context sentences silently and target sentences out loud.

Data for other Japanese experiments were also collected within the same session for each speaker. This included a third experiment which used the same frame sentence structure but varied moraic structure for the test words, a fourth experiment for error elicitation with tongue twisters, and a fifth experiment with two passage readings (in English and Japanese). The data for Experiments 1 and 2 are analyzed in this dissertation. The order of data collection was as follows (one block includes one repetition of each stimulus, and additional blocks are run to collect multiple repetitions):

1. Two blocks of randomized one-syllable words to collect Japanese vowels

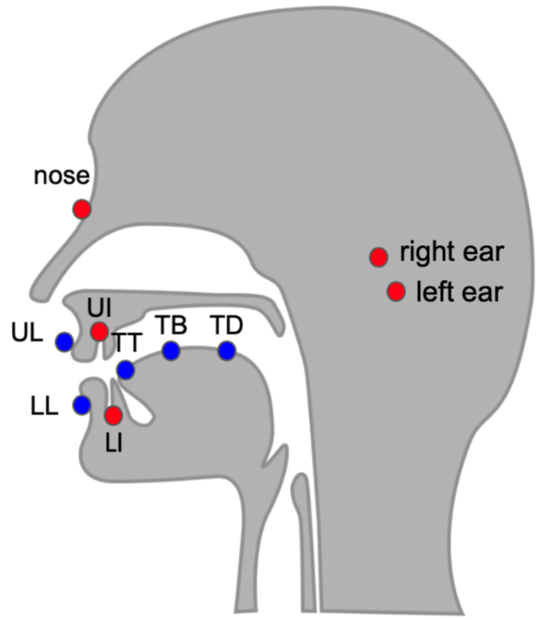
2. Four blocks of randomized utterances from Experiment 1 (real words), Experiment 2 (neologisms), and Experiment 3 (moraic structure)
3. One block of randomized utterances from Experiment 4 (error elicitation)
4. Five blocks of randomized utterances from Experiment 1-3
5. One block of randomized utterances from Experiment 4
6. Experiment 5: Japanese reading passage
7. Experiment 5: English reading passage

Each session took a maximum of three hours, which also included time spent obtaining consent, training, attaching sensors, multiple breaks as needed, recording data for all five experiments, and removing sensors.

3.4. Recording Apparatus

Data was collected via the AG501 three-dimensional electromagnetic transduction device (Carstens Medizinelektronik) at the UCSB Phonetics Lab. Sensors were attached to five areas of interest: tongue dorsum (TD), tongue body center (TB), tongue tip (TT), upper lip (UL), and lower lip (LL) (see Figure 16). An additional five sensors were attached as reference points: upper incisor (UI), lower incisor (LI), nose, and left and right ears. Audio recordings were collected simultaneously to the kinematic recordings with a Sennheiser shotgun microphone set at a sampling rate of 16 kHz.

Figure 16. EMA Sensor Placement. Red dots indicate reference sensors; blue dots indicate active articulators of interest.



3.5. Analysis

The following sections detail the labeling and analysis of data collected for this study.

3.5.1. Labelling

Consonant and vowel constriction gestures composing the test words were semi-automatically labeled using custom software in Matlab which detects kinematic landmarks on the basis of velocity criteria (Tiede, Haskins Laboratories). The kinematic landmarks used for labeling are illustrated in Figure 17. Labial consonants were labeled using lip aperture (LA), coronal consonants with tongue tip (TT) vertical displacement, and velar consonants and vowels with tongue dorsum (TD) vertical displacement. Landmark timepoints labeled for each constriction gesture included the onset, peak-velocity, target, and offset. Duration of constriction gestures were measured by the formation phase of the gesture (onset to release) and the release phase of the gesture (release to offset). Measurements for the antepenultimate and penultimate vowel gestures in three-syllable real words were not reliable enough for analysis due to coarticulation, so these were not able to be labeled for that set of data.

Figure 17. Kinematic landmarks in labeling of oral constriction gestures.

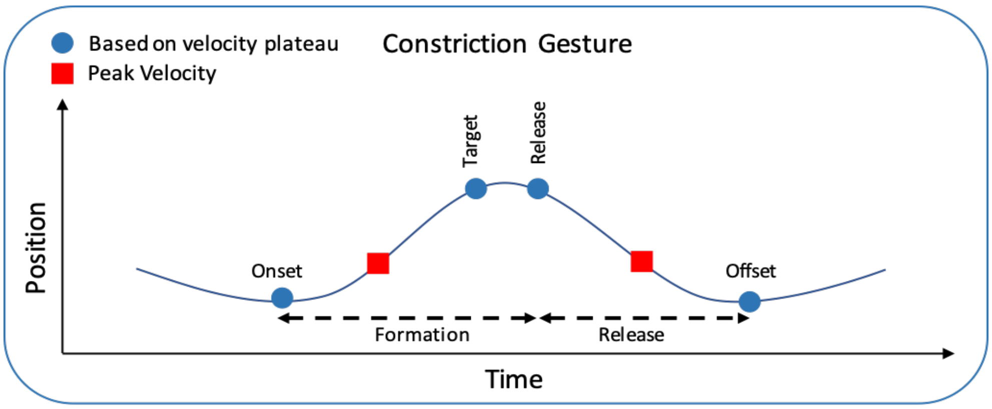
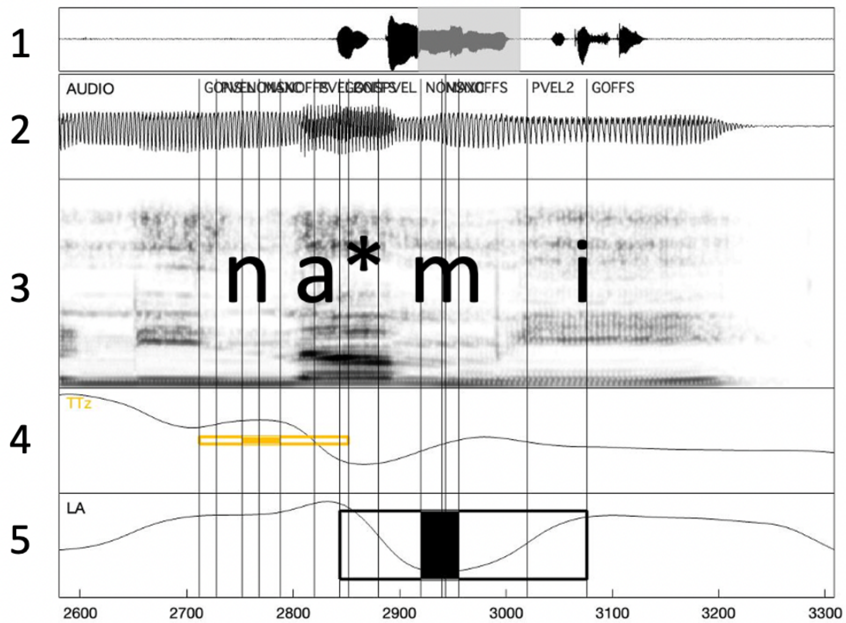


Figure 18 shows sample labels for the consonant constriction gestures in the test word *na*mi* ‘Nami (name)’. In (1) is the waveform for the entire utterance, with the target word in the shaded area. (2) shows the waveform for the shaded area, including the test word *na*mi*. (3) shows the spectrogram for the test word. (4) shows the measurement for the vertical trajectory of the tongue tip movement in which /n/ is labeled, and (5) shows the measurement for lip aperture (distance between upper and lower lips) labeled for the /m/. The start and end of the boxes in (4) and (5) align with the onset and offset landmarks in Figure 17; the solid-colored portion of the boxes align with the target and release landmarks in Figure 17. The height of the boxes corresponds to the size of the gesture.

Figure 18. Sample labels for /n/ and /m/ C gestures in test word *na*mi* ‘Nami (name)’.



3.5.2. Statistical analysis

The lmerTest package (Kuznetsova, Brockhoff and Christensen 2017) in R (R Development Core Team 2009) was used to run linear mixed effects models. Separate models were run for consonants and vowels. Emmeans (Lenth 2020) with Holm correction was used for pairwise comparisons ($\alpha = 0.05$). Specific information on the statistical models used will be given for each experiment separately in their respective chapters. In this study, we employ a deductive modeling approach which uses a “full” model containing all variables relevant to our research questions (Tizón-Couto and Lorenz 2021).

4. Boundary lengthening and pitch accent

This chapter investigates the scope of phrase-final lengthening as a function of lexical pitch accent position in Tokyo Japanese. The results will be analyzed within the framework of Articulatory Phonology, in which phrase-final lengthening effects are the result of clock-slowing effects of the π-gesture which instantiates the prosodic boundary. The scope of phrase-final effects can be interpreted as the degree and timing of overlap between consonant and vowel constriction gestures with the π-gesture, which imposes clock-slowing effects on co-active gestures. Thus, the scope of phrase-final lengthening effects, which can be measured by comparing the duration of phrase-medial and phrase-final constriction gestures, will allow us to examine the coordination of the π-gesture with constriction gestures in Japanese.

4.1. Questions and predictions

In order to understand how prosodic boundaries affect speech gestures and interact with word prosody in Japanese, we address the following questions:

(i) Which constriction gestures in the phrase-final word are affected by phrase-final lengthening?

(ii) Does lexical pitch accent affect the scope of phrase-final lengthening?

Based on previous research, we expect that phrase-final constriction gestures will be longer (Oller 1973; Berkovits 1993a, 1993b; Cambier-Langeveld 1997; White 2002; Turk and Shattuck-Hufnagel 2007; Edwards, Beckman and Fletcher 1991; Beckman and Edwards, 1992; Byrd and Saltzman 1998; Byrd 2000; Byrd, Lee, et al. 2005; Byrd, Krivokapić and Lee 2006; Krivokapić 2007; Katsika 2016; Kim, Jang and Cho 2017). On the scope of phrase-final kinematic effects, one hypothesis is that that effects occur mainly in the final rime, as suggested by previous acoustic work on Japanese (Shepherd 2008). An alternate hypothesis is that effects will begin in the penultimate syllable, following the results of previous acoustic studies which tested unaccented and initial-accented two-syllable words in Japanese (Seo, Kim, et al. 2019a, 2019b). The difference in scope between Shepherd (2008) and Seo, Kim, et al. (2019a, 2019b) may possibly be related to the difference in stimuli used, as Shepherd only tested two-syllable words with a long vowel in the initial syllable, and short or long vowel in the second syllable (CVV.CV(V)), and Seo, Kim, et al. (2019a, 2019b) only tested the effect of moraic structure with heavy syllables with coda consonants (CVN.CV(N)).)). This suggests that moraic structure may have an effect on the scope of phrase-final lengthening effects, a topic which is outside the scope of this dissertation. To control for possible affects of moraic structure, test words for the experiments in this disertation consist only of light syllables (CV.CV.(CV)). While Shepherd (2008) only tested CVV.CV(V) words, Seo, Kim, et al. (2019a, 2019b) also included CV.CV test words. Another notable difference is that these Shepherd (2008) used materials with statements, whereas the current study elicited question utterances with rising LH% boundary tones. Future research could possibly explore the interaction between boundary tone type and phrase-final lengthening effects using the Corpus of Spontaneous Japanese (Maekawa 2003; Maekawa, Kikuchi and Tsukahara 2004).

As duration is not typically considered a correlate of pitch accent in Japanese (Weitzman 1970, Poser 1984; Beckman 1986; Pierrehumbert and Beckman 1988; Sugito 1998), it is unlikely that Japanese would exhibit a pattern similar to English, where the prominent syllable is lengthened as a function of boundary (Cho 2006; Turk and Shattuck-Hufnagel 2007); (Byrd and Riggs 2008). To our knowledge, there are no previous studies examining boundary effects as a function of lexical pitch accent position in Japanese. However, if pitch accent has similar effects as stress, we could expect boundary effects to be initiated earlier when pitch accent is earlier. It is possible that we may find an interaction similar to that of Greek, in which the initiation of boundary lengthening varies as a function of pitch accent position (Katsika 2016). In fact, previous studies suggest that the presence of lexical pitch accent may affect the amount of lengthening in Japanese (Seo, Kim, et al. 2019a, 2019b). In the π-gesture framework, amount of lengthening may be related to the π-gesture's coordination, as a greater amount may result from greater overlap of the constriction gesture with the π-gesture's area of maximum activation. This was the account proposed in the case of Greek, where the earlier initiation of prosodic lengthening when stress was earlier in the word was accounted for by an effect of lexical stress on the coordination of the π-gesture, in that the π-gesture was shifted earlier. This will be further described and discussed in later chapters, as the Greek results were accounted for by coordination with a μ-gesture, which we investigate for Japanese in Chapter 6. For the purpose of this chapter, if the scope of prosodic lengthening in Japanese is affected by word prosody, it may be the case that the coordination of the π-gesture is affected, as it is in Greek.

4.2. Methods

4.2.1. Experimental design

As a reminder, there were three stimuli sets: two-syllable real words, three-syllable real words, and three-syllable neologisms. Stimuli were elicited in frame sentences in phrase-final (test) or phrase-medial (control) conditions to test the effect of phrase boundaries. To test the effect of pitch accent position, test words included all possible pitch accent conditions:

antepenultimate (possible only in the three-syllable word sets), penultimate, ultimate, and unaccented.

Table 4. Test words for all experiments. Pitch accent is indicated on the preceding syllable with an asterisk (*).

Pitch Accent Position	*Real words*		*Neologisms*
Antepenultimate		**na*mina** 'Namina (name)'	**ma*mima**
Penultimate	**na*mi** 'Nami (name)'	**nana*me/nama*mi** 'slanted/living body'	**mami*ma**
Ultimate	**nami*** 'wave'	**nigami*** 'bitterness'	**mamima***
Unaccented	**nami** 'medium'	**namida** 'Namida (name)'	**mamima**

4.2.2. Statistical Analysis

The experiments included 2 phrasal positions, 3 or 4 pitch accent positions (3 in two-syllable words and 4 in three-syllable words), 9 repetitions, and 5 speakers. 360 utterances were collected for neologisms, and 720 for the other two- and three-syllable words, yielding 1,080 utterances in total. Of these, a total of 1,035 utterances were analyzed (338 and 697 from each respective set), as some tokens were excluded due to speech errors or technical problems. We ran a linear mixed effects model on each stimuli set: (i) neologisms, (ii) two-syllable real words, and (iii) three-syllable real words. Each data set was run on a separate model due to different conditions resulting from word length (only three-syllable words have antepenultimate conditions) and different dependent variables (not all non-final V gestures were measurable).

Since we are interested in the effect of phrase boundaries, the role of pitch accent position on boundary effects, and the scope of these effects on consonant (C) and vowel (V) gestures, our fixed effects and their levels were BOUNDARY [phrase-medial (PhM), phrase-final (PhF)], PITCH ACCENT [antepenultimate, penultimate, ultimate, unaccented], and SYLLABLE POSITION [C0 V0 (word-final), C1 V1 (penultimate), C2 V2 (antepenultimate)]. Antepenultimate and C2/V2 conditions were only applicable to three-syllable words. C and V gestures were labeled for all stimuli sets, except non-final V gestures in three-syllable real words which were not reliable enough for measurement. Emmeans with Holm correction was used for pairwise comparisons ($\alpha = 0.05$).

To investigate boundary kinematics, we fit two models using the lmerTest package in R, each with formation duration (FormDUR) or release duration (RelDUR) as response variables, the same three factors as independent variables, and random intercepts by Speaker. The model included a three-way interaction between BOUNDARY, PITCH ACCENT, and SYLLABLE POSITION, as well as their two-way interactions. Separate models were run for data for C and V constrictions, and z-score normalization was used to normalize durations by phoneme across speakers.

4.3. Results

4.3.1. Neologisms: consonant constriction gestures

In neologisms, there was no significant three-way interaction between BOUNDARY, PITCH ACCENT, and SYLLABLE POSITION in formation duration or release duration of C gestures. Significant two-way interactions indicated that the formation duration ($p<.001$) and release duration ($p<.0001$) of final C constriction gesture (C0) underwent phrase-final lengthening. Figure 19 shows the results for normalized C formation duration. Figure 20 shows the raw durations of data for the formation phase of C gestures to provide context for model predictions which use normalized values. The reader is reminded that consonants (and vowels) in this study were coded based on their distance from the phrase edge, so that the word-final consonant is C0 (being at the boundary), the penultimate consonant is C1 (being one syllable away from the boundary), and the antepenultimate consonant is C2 (being two syllables away from the boundary). Planned pairwise comparisons showed that neologisms had phrase-final lengthening in the consonant formation gesture of the word-final syllable (C0) for antepenultimate ($ß_{Antepenult}$=0.42, SE=0.12, p<.001), penultimate ($ß_{Penult}$=0.30, SE=0.12, p<.01), and unaccented ($ß_{Unaccented}$=0.25, SE=0.12, p<.05) words.

Figure 19. Boundary lengthening for normalized durations of the formation phase of C constriction gestures in neologisms by pitch accent position.

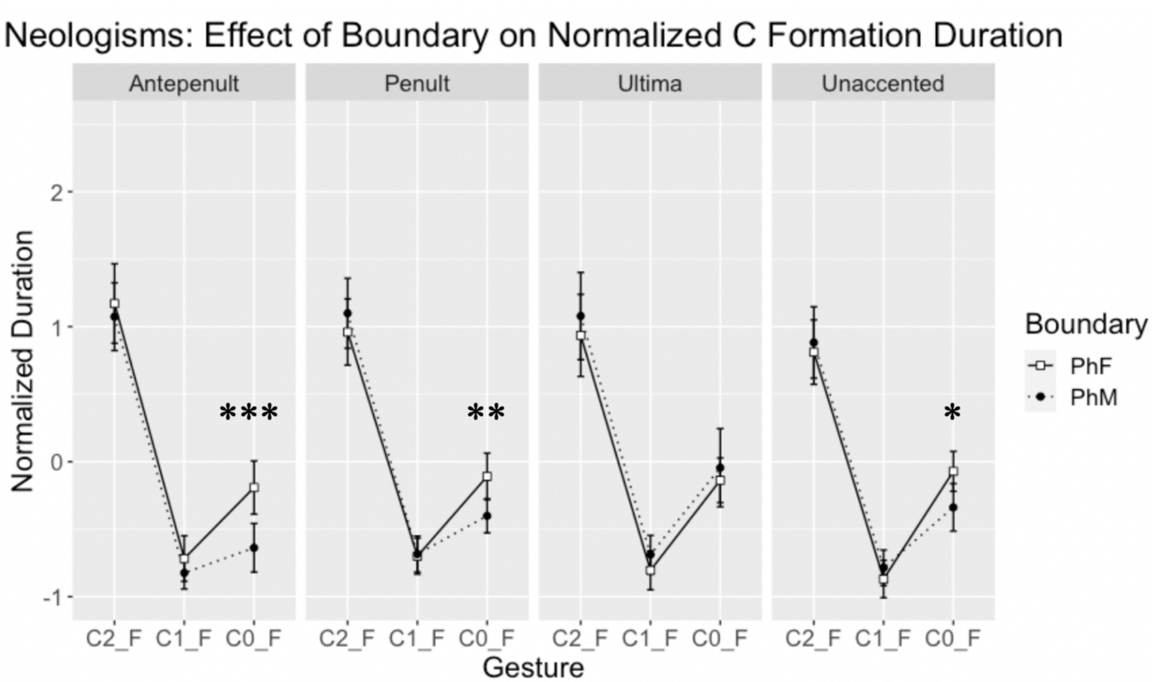
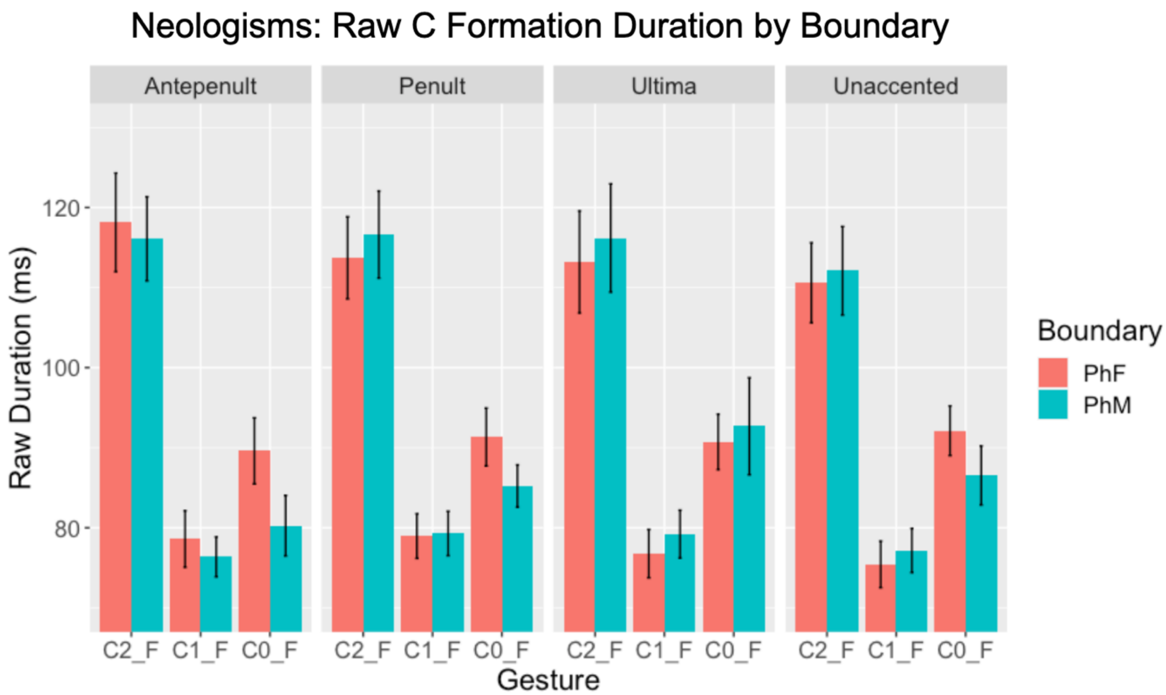


Figure 20. Raw durations for the formation phase of C constriction gestures in neologisms by boundary and pitch accent position.



The lack of C0 phrase-final lengthening in ultimate words can be attributed to C0 in ultimate words being longer than other accent conditions phrase-medially (ultimate: $ß_{Antepenult}$=-0.58, SE=0.12, p<0.0001; $ß_{Penult}$=-0.36, SE=0.12, p<0.01; $ß_{Unaccented}$=0.30, SE=0.12, p<0.05). In other words, the lack of a significant difference between phrase-medial and phrase-final conditions for ultimate-accented words may originate from a longer C0 duration phrase-medially, and not a shorter duration phrase-finally.

In the release duration of the final consonant (C0), there was boundary-related lengthening for all pitch accent conditions ($ß_{Antepenult}$=2.38, SE=0.11, p<.0001; $ß_{Penult}$=2.21, SE=0.11, p<.0001; $ß_{Ultima}$=1.86, SE=0.12, p<.0001; $ß_{Unaccented}$=2.10, SE=0.12, p<.0001). Figure 21 shows the results for the release durations of C gestures in neologisms Figure 22 shows the raw durations to provide context for model predictions which use normalized values.

Similar to formation gestures, phrase-medial C0 release gestures in ultimate-accented words were slightly longer than phrase-medial C0 release gestures for other pitch accent conditions (ultimate: $ß_{Antepenult}$=-0.42, SE=0.12, p<0.01; $ß_{Penult}$=-0.39, SE=0.11, p<0.01; $ß_{Unaccented}$=0.38, SE=0.12, p<0.01).

Figure 21. Boundary lengthening for normalized durations of the release phase of C constriction gestures in neologisms by pitch accent position.

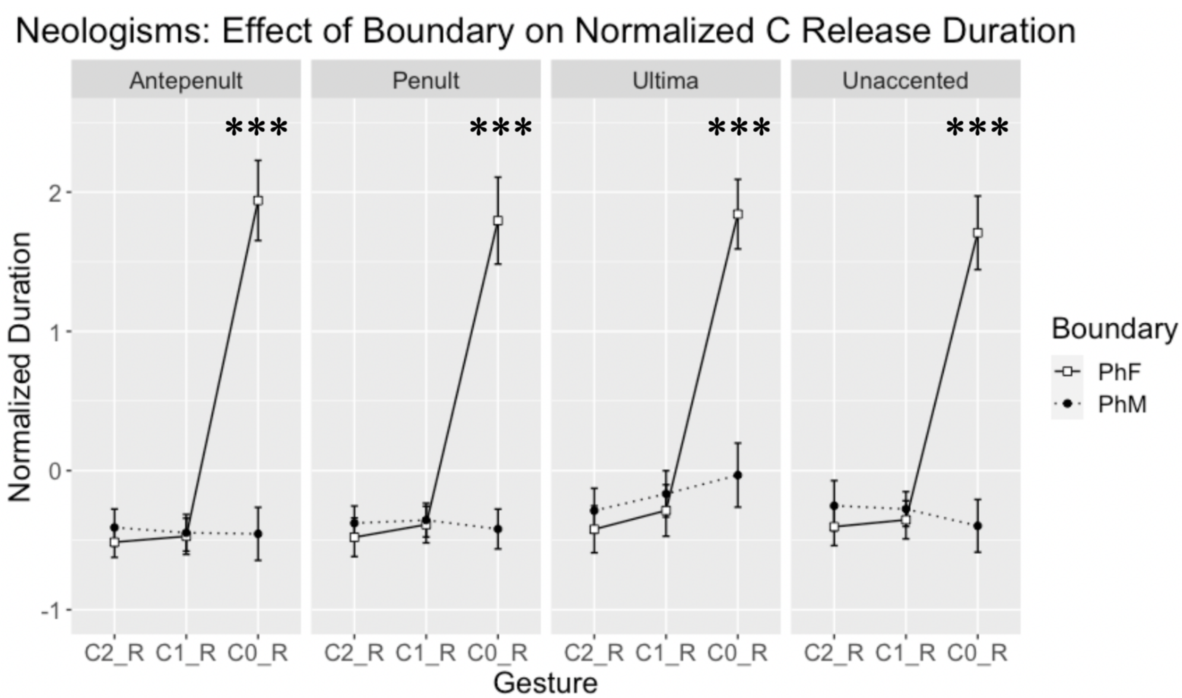
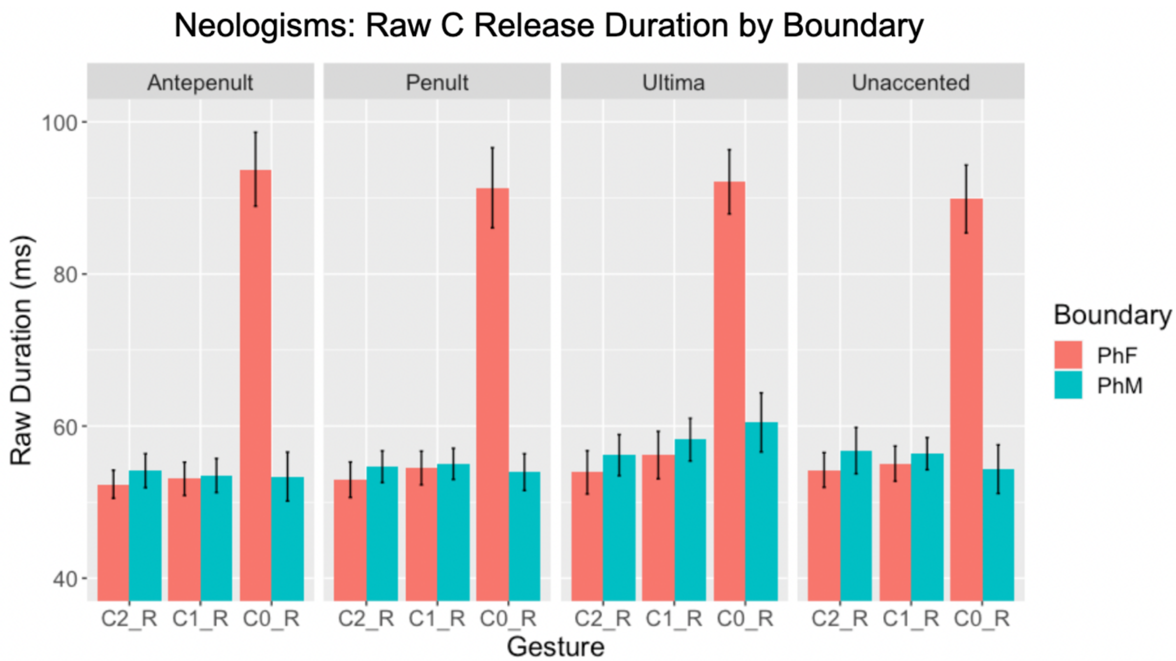


Figure 22. Raw durations for the release phase of C constriction gestures in neologisms by boundary and pitch accent position.



4.3.2. Neologisms: vowel constriction gestures

In vowel gestures for neologisms, there was a significant three-way interaction between BOUNDARY, PITCH ACCENT, and SYLLABLE POSITION (p<.05). Regardless of pitch accent position, there was significant boundary-related lengthening in the final V0 gesture ($ß_{Antepenult}$=1.92, SE=0.16, p<.0001; $ß_{Penult}$=1.71, SE=0.16, p<.0001; $ß_{Ultima}$=0.91, SE=0.17, p<.0001; $ß_{Unaccented}$=1.14, SE=0.16, p<.0001).

Figure 23. Boundary lengthening for normalized durations of the formation phase of V constriction gestures in neologisms by pitch accent position.

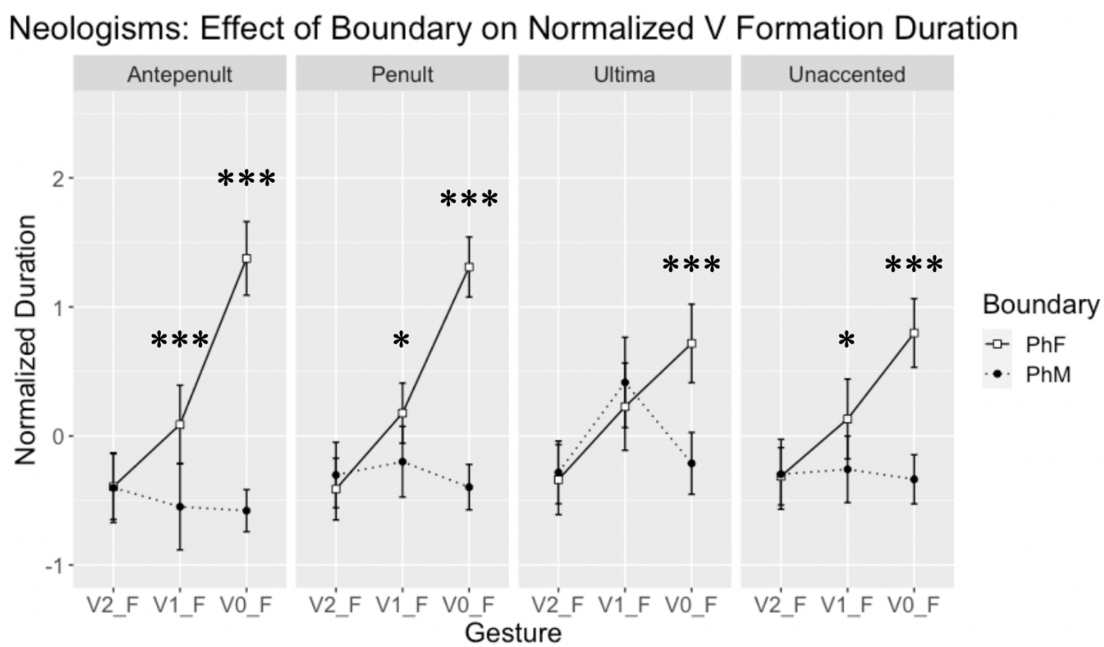
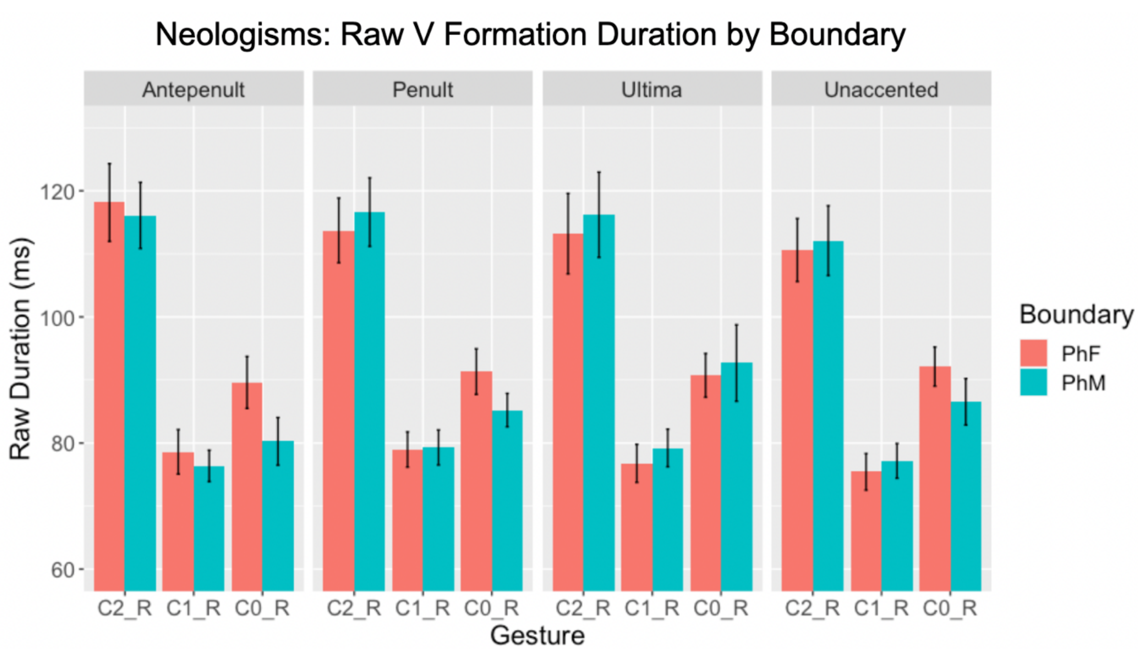


Figure 24. Raw durations for the formation phase of V constriction gestures in neologisms by boundary and pitch accent position.



As can be seen in Figure 23, lengthening extended to the penultimate vowel (V1) gesture in antepenultimate ($ß_{Antepenult}$=0.59, SE=0.17, $p<.001$), penultimate ($ß_{Penult}$=0.38, SE=0.16, $p<.05$), and unaccented ($ß_{Unaccented}$=0.39, SE=0.16, $p<.05$) words. The difference in scope for ultimate-accented words can be attributed to a longer phrase-medial duration in V1 ($ß_{Antepenult}$=-0.97, SE=0.16, $p<.0001$; $ß_{Penult}$=-0.64, SE=0.16, $p<.001$; $ß_{Unaccented}$=0.71, SE=0.17, $p<.001$). There is also a difference in the amount of lengthening in the final syllable (V0) which can be seen when comparing phrase-final V0 gestures for antepenultimate and penultimate words versus ultimate and unaccented words; V0 vowel formation gestures were longer phrase-finally for antepenultimate and penultimate words (antepenultimate: $ß_{Ultima}$=0.65, SE=0.16, $p<.001$, $ß_{Unaccented}$=0.57, SE=0.16, $p<.01$; penultimate: $ß_{Ultima}$=0.59, SE=0.17, $p<.01$, $ß_{Unaccented}$=0.52, SE=0.16, $p<.01$).

4.3.3. Neologisms summary

The following figures show the combined results for consonant and vowel constriction gestures in neologisms. Figure 26 shows the scope of boundary-related lengthening, which appears to vary by pitch accent position. In all pitch accent conditions except ultimate-accented words, phrase-final lengthening begins in the penultimate vowel gesture (V1). There is also a difference in the magnitude of effects, as shown in Figure 27, which shows the difference (delta) between phrase-final and phrase-medial durations for each constriction gesture by pitch accent condition. Words with antepenultimate pitch accent had a greater amount of lengthening in the constriction gestures of the penultimate vowel (V1) and the formation of the final consonant (C0). In the latter, the amount of phrase-final lengthening differed by pitch accent position, with antepenultimate words lengthening the most, followed by words with penultimate pitch accent, then unaccented words, and finally ultimate-accented words having no significant effect of phrasal boundary. Especially in the final vowel gesture (V0), antepenultimate- and penultimate-accented words saw a greater amount of lengthening compared to words with no pitch fall realized within the word.

Figure 25. Summary of results for the effect of phrase boundary on each constriction gesture in the test word by pitch accent position. Gestures are presented sequentially following orthographic conventions, but C and V gestures overlap in each syllable.

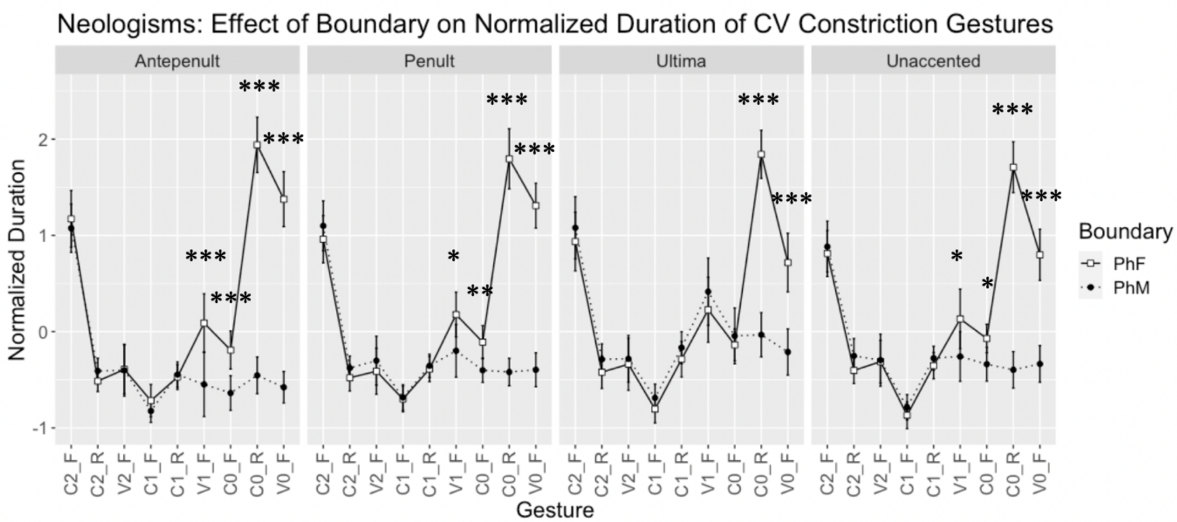


Figure 26. Summary of raw durations by phrase boundary and pitch accent for each constriction gesture in the neologism test word. Gestures are presented sequentially following orthographic conventions, but C and V gestures overlap in each syllable.

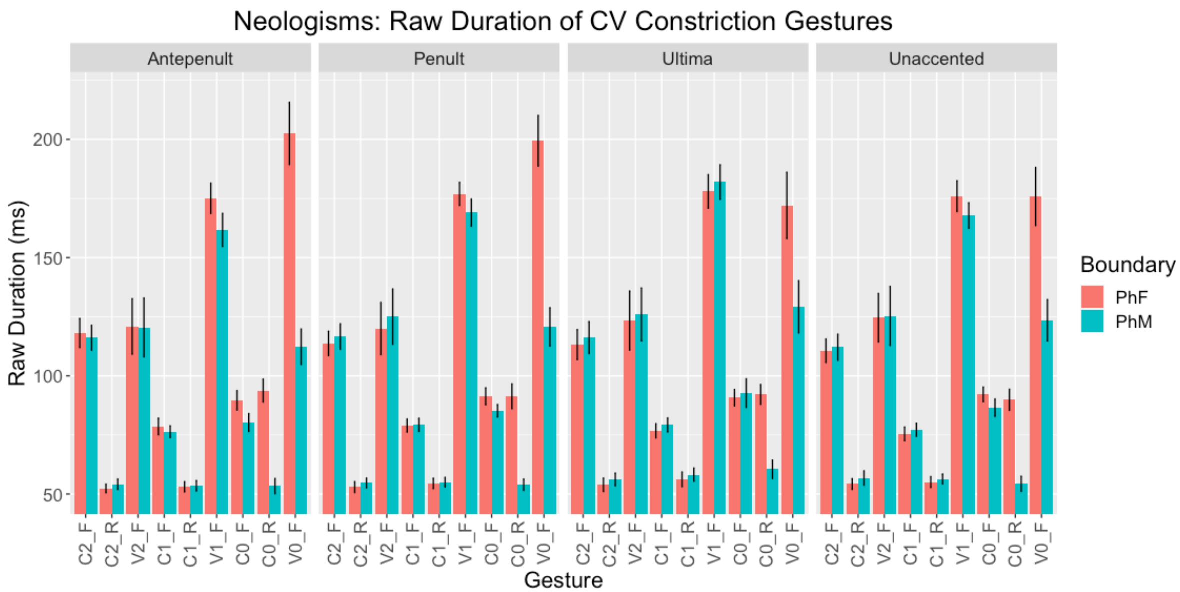
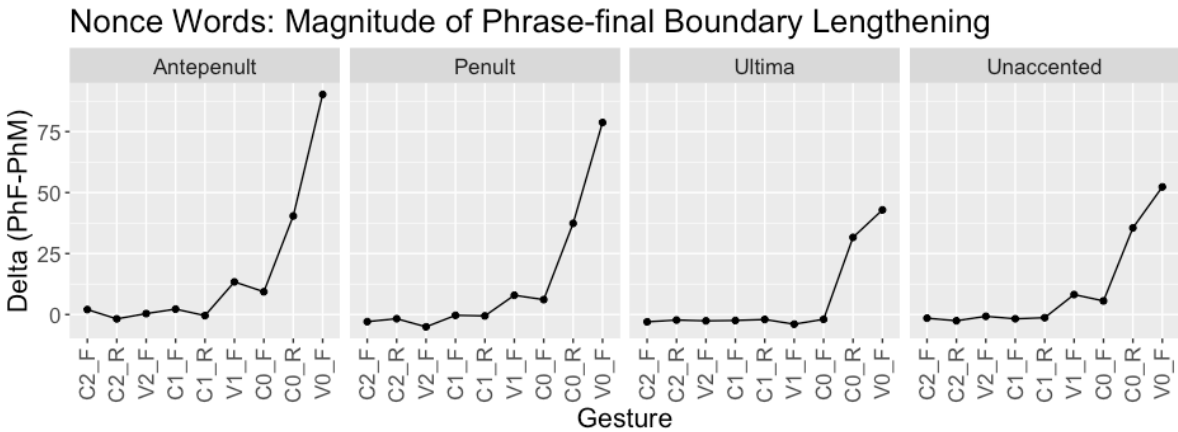


Figure 27. The absolute magnitude of increase in raw duration from phrase-medial to phrase-final boundary condition for each constriction gesture by pitch accent condition. Delta refers to the difference between boundary conditions (phrase-final minus phrase-medial). Consonant and vowel constriction gestures are coded so that gestures belonging to the word-final syllable are C/V0 and those belonging to the antepenultimate syllable (two syllables away from the boundary) are C/V2. Gestures

labeled with 'F' belong to the formation phase of that gesture; those labeled with 'R' belong to the release phase of that gesture.



4.3.4. Three-syllable real words: consonant constriction gestures

Three-syllable real words also did not have a significant three-way interaction in formation duration of C gestures. Figure 28 shows the results for normalized formation duration of consonant constriction gestures; Figure 29 provides raw durations for the same consonant constriction gestures for context.

Figure 28. Boundary lengthening for normalized durations of the formation phase of C constriction gestures in three-syllable real words by pitch accent position.

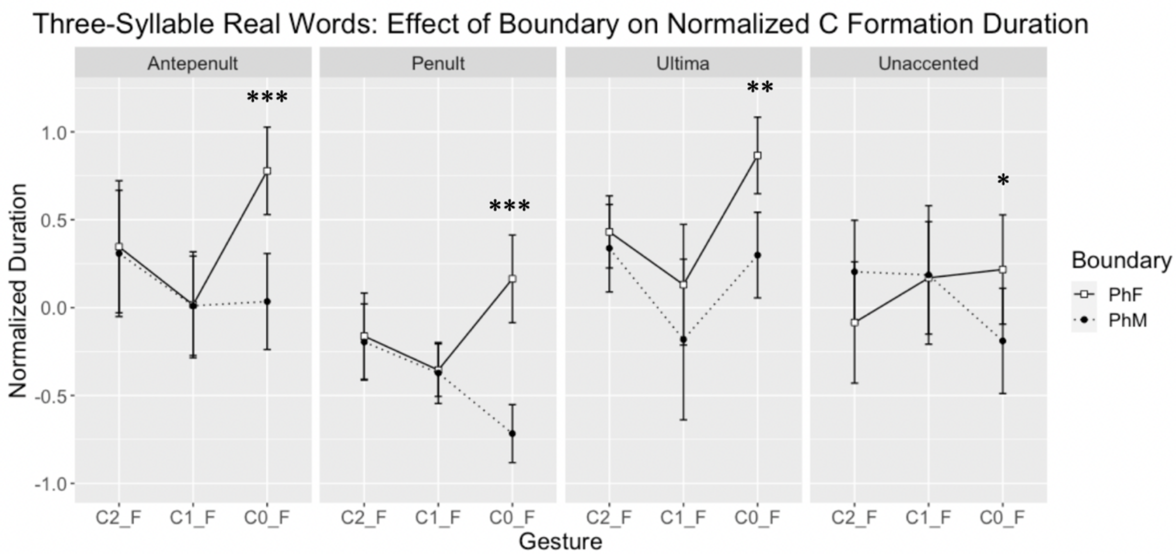
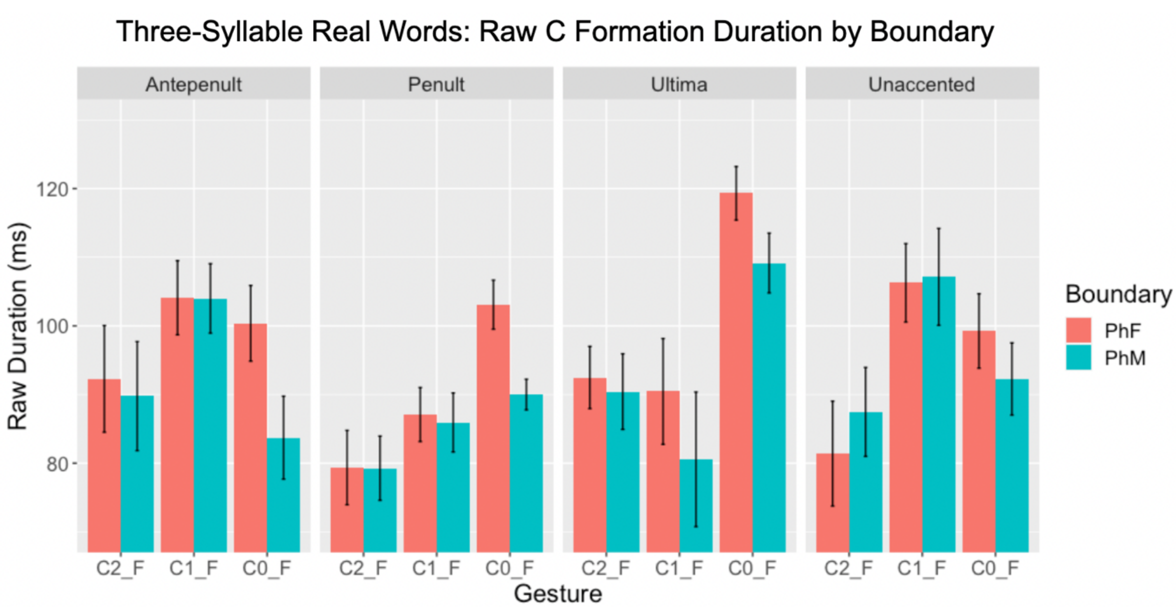


Figure 29. Raw durations for the formation phase of C constriction gestures in three-syllable real words by boundary and pitch accent position.



There was phrase-final lengthening in the formation of the final consonant constriction gesture (C0) for all pitch accent conditions ($ß_{Antepenult}$=0.71, SE=0.19, $p<.001$; $ß_{Penult}$=0.88, SE=0.13, $p<.0001$; $ß_{Ultima}$=0.54, SE=0.19, $p<.01$; $ß_{Unaccented}$=0.40, SE=0.19, $p<.05$). The larger difference between boundary conditions indicates that antepenultimate ($p<.001$) and penultimate words ($p<.0001$) had a greater amount of boundary lengthening compared to ultimate ($p<.01$) or unaccented words ($p<.05$).

Figure 30 shows the results for durations of the release phase of consonant constriction gestures in three-syllable real words; Figure 31 shows the raw durations for context. In release durations, there was a significant three-way interaction ($p<.01$). The C constriction gestures of word-final consonants (C0) were longer phrase-finally in antepenultimate, penultimate, and unaccented words ($ß_{Antepenult}$=1.44, SE=0.19, $p<.0001$; $ß_{Penult}$=1.25, SE=0.13, $p<.0001$; $ß_{Unaccented}$=0.81, SE=0.19, $p<.0001$). Similar to formation duration results for neologism consonants, in three-syllable real words, ultimate words did not show significant boundary lengthening in release durations because they were longer phrase-medially (ultimate: $ß_{Antepenult}$=-0.41, SE=0.20, $p<.1$; $ß_{Penult}$=-1.13, SE=0.17, $p<.0001$; $ß_{Unaccented}$=0.86, SE=0.20, $p<.001$). This may indicate an effect of pitch accent on boundary lengthening in the final syllable (C0) of ultimate-accented words.

Figure 30. Boundary lengthening for normalized durations of the release phase of C constriction gestures in three-syllable real words by pitch accent position.

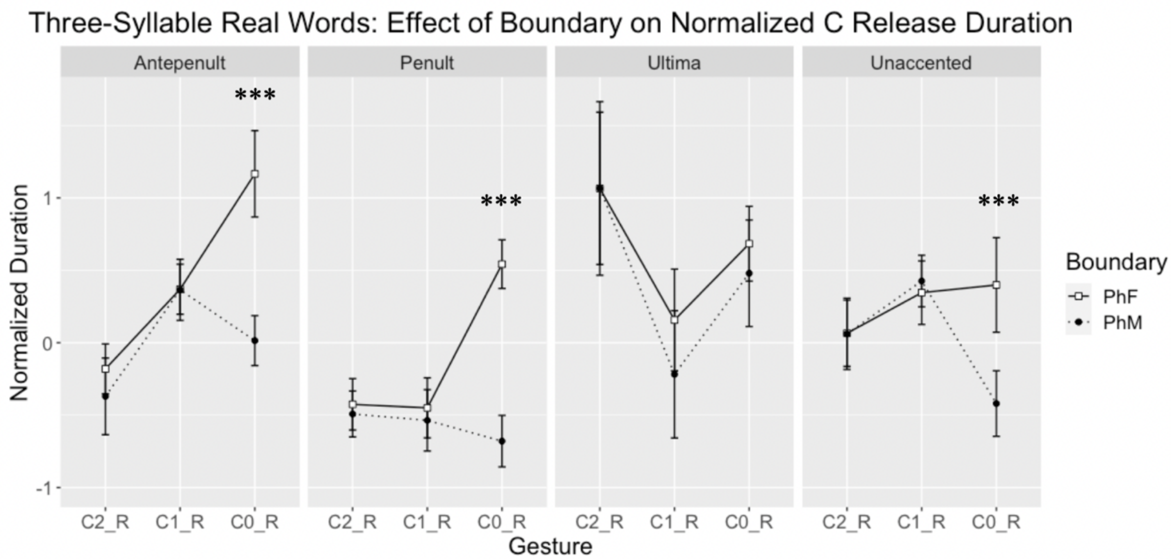
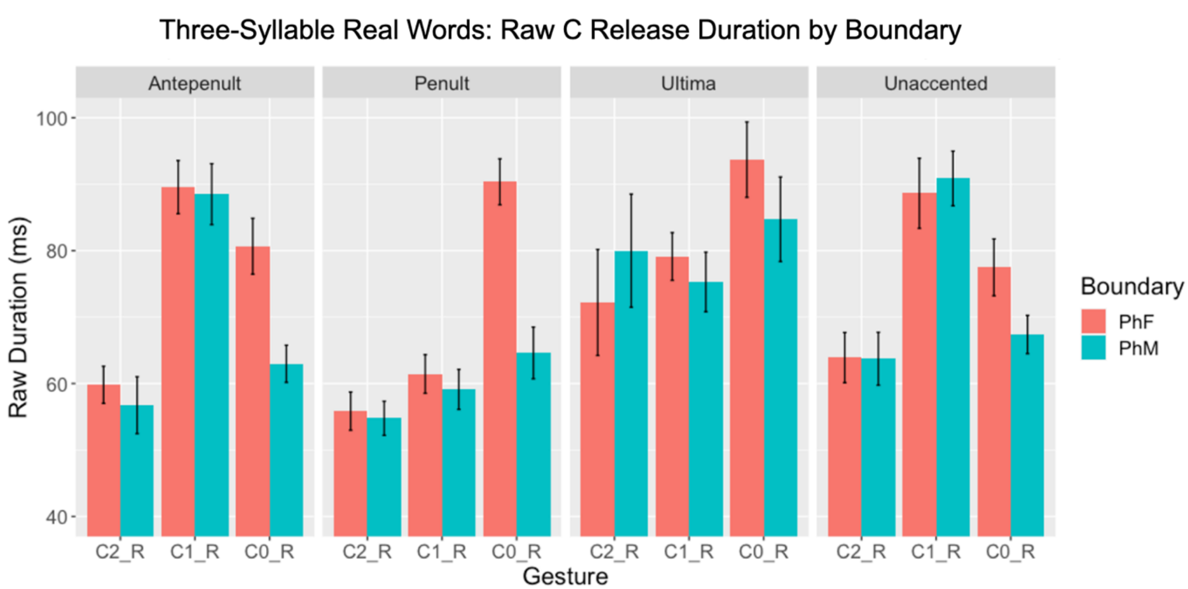


Figure 31. Raw durations for the release phase of C constriction gestures in three-syllable real words by boundary and pitch accent position.



4.3.5. Three-syllable real words: vowel constriction gestures

Only the final vowel was able to be measured in three-syllable real words, which meant this model did not include a factor of SYLLABLE POSITION. There was a significant two-way interaction between BOUNDARY and PITCH ACCENT (p<.001). Vowel gestures (V0) in all pitch accent conditions were significantly longer when the word was phrase-final ($ß_{Antepenult}$=2.34, SE=0.15, p<.0001; $ß_{Penult}$=1.93, SE=0.14, p<.0001; $ß_{Ultima}$=1.17, SE=0.13, p<.0001; $ß_{Unaccented}$=1.31, SE=0.16, p<.0001). There was also a difference in the amount of lengthening in V0; antepenultimate words were longer phrase-finally than all other words (antepenultimate: $ß_{Penult}$=0.53, SE=0.15, p<.001; $ß_{Ultima}$=1.54, SE=0.14, p<.0001; $ß_{Unaccented}$=0.68, SE=0.15, p<.0001).

Figure 32. Boundary lengthening for normalized durations of the formation phase of the final V0 constriction gesture in three-syllable real words by pitch accent position.

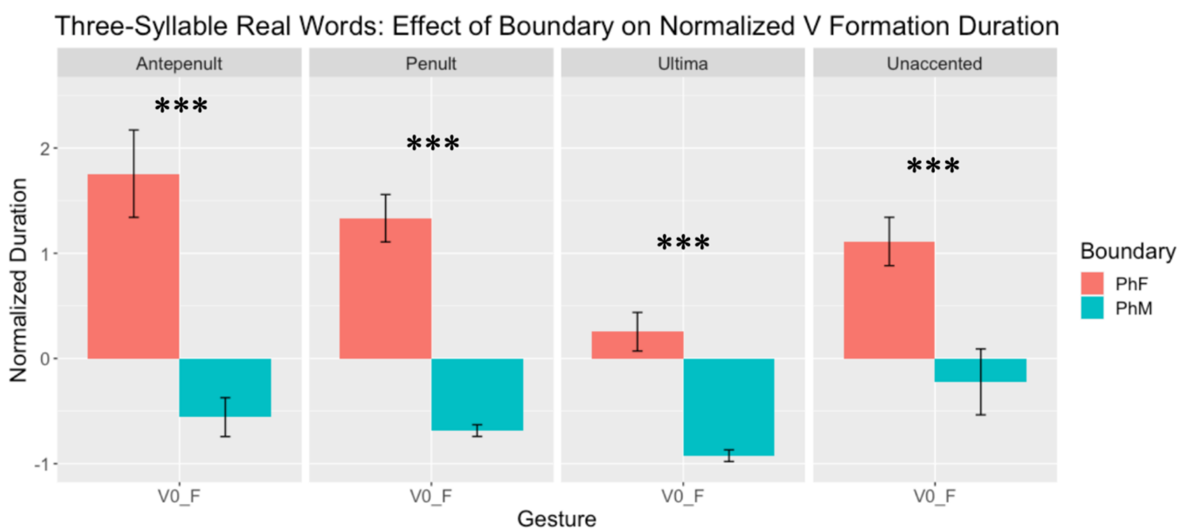
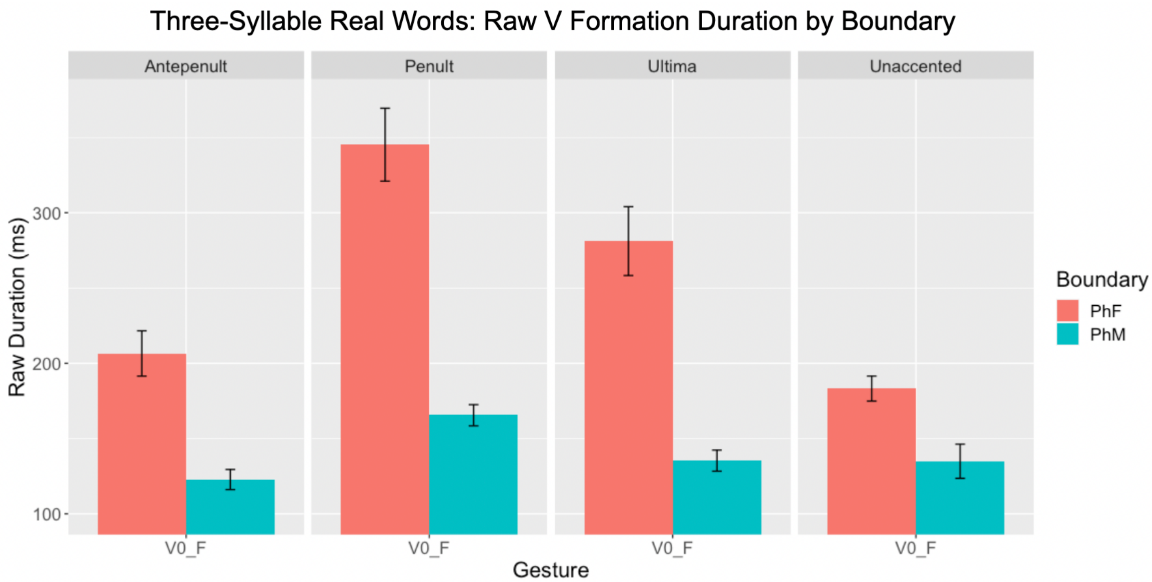


Figure 33. Raw durations for the formation phase of V constriction gestures in three-syllable real words by boundary and pitch accent position.



4.3.6. Three-syllable real words summary

Figure 34 and Figure 35 summarize the scope and amount of phrase-final lengthening in three-syllable real words, where phrase-final lengthening affected gestures in the final C0 and V0 gestures. In this set of data, penultimate and antepenultimate vowel gestures were not able to be measured, so we were unable to confirm whether or not phrase-final lengthening was initiated in the penultimate vowel formation gesture (V1). Unlike neologisms, phrase-final lengthening in three-syllable real words with ultimate pitch accent appears to affect at least the formation gesture of the final C0. In three-syllable real words, pitch accent also affected the amount of phrase-final lengthening in the formation gesture of the final consonant (C0), with antepenultimate and penultimate words having the greatest amount of lengthening, followed by ultimate words, and then unaccented words. In the release gesture of the final consonant (C0), antepenultimate words were longer phrase-finally than all other pitch accent conditions (antepenultimate: $ß_{Penult}$=0.62, SE=0.17, p<.001; $ß_{Ultima}$=0.51, SE=0.19, p<.05; $ß_{Unaccented}$=0.78, SE=0.19, p<.001). This was also the case for the final vowel gesture (V0) (antepenultimate: $ß_{Penult}$=0.53, SE=0.15, p<.001; $ß_{Ultima}$=1.54, SE=0.14, p<.0001; $ß_{Unaccented}$=0.68, SE=0.15, p<.0001).

Figure 34. Summary of results for the effect of phrase boundary on each constriction gesture in the test word by pitch accent position. Gestures are presented sequentially following orthographic conventions, but C and V gestures overlap in each syllable. Antepenultimate (V2) and penultimate (V1) vowel gestures are excluded from the figure as they were not measurable.

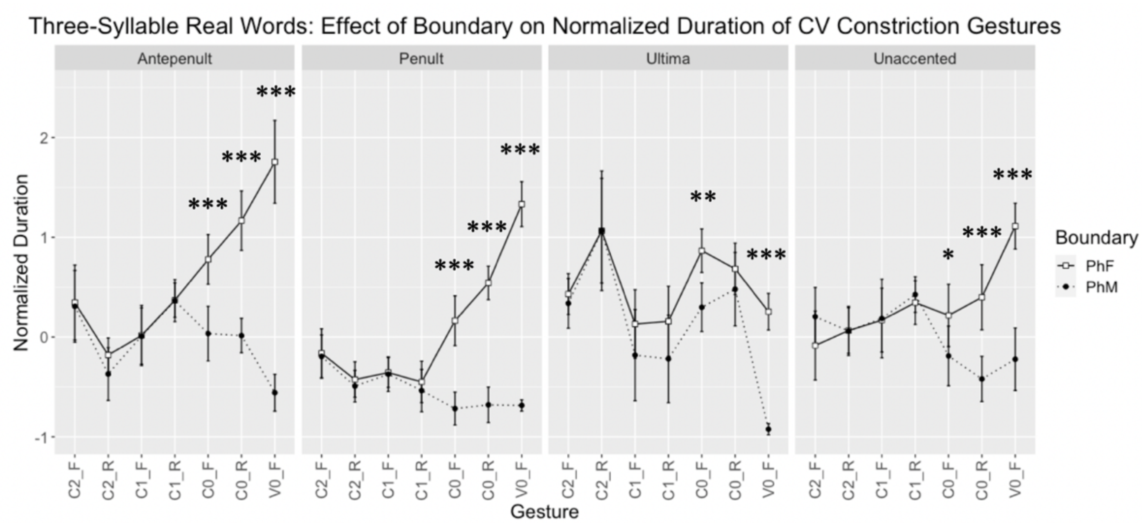
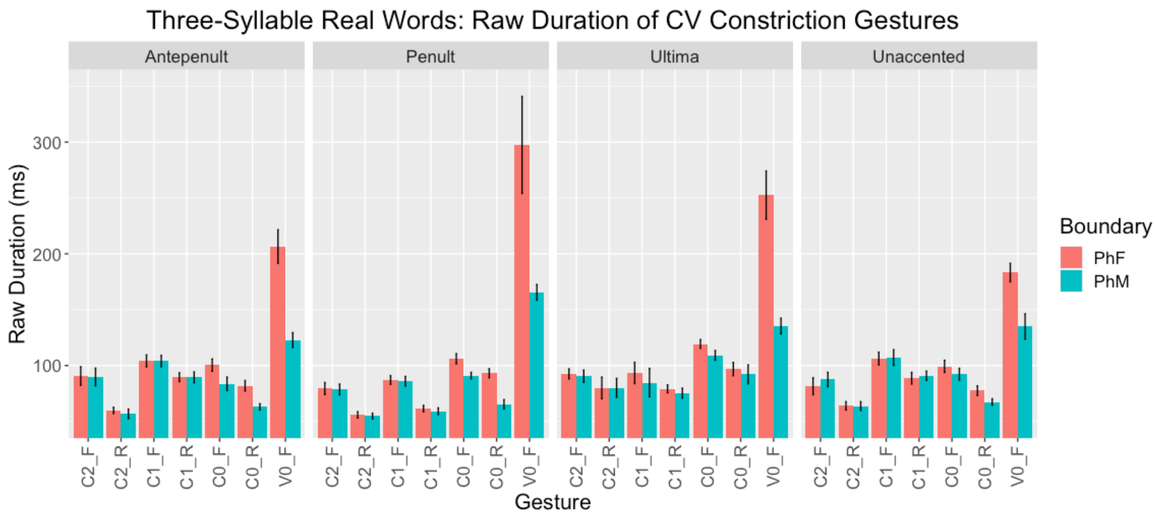


Figure 35. Summary of raw durations by phrase boundary and pitch accent for each constriction gesture in the three-syllable real word stimuli. Gestures are presented sequentially following orthographic conventions, but C and V gestures overlap in each

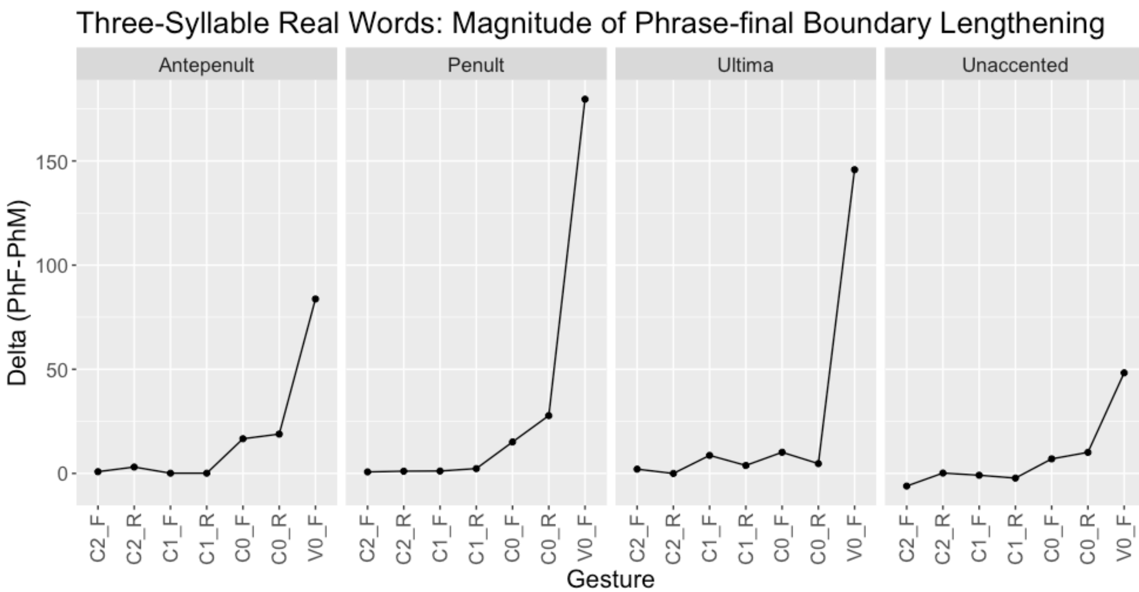
syllable. Antepenultimate (V2) and penultimate (V1) vowel gestures were not measurable and thus excluded from the figure.



There is also a difference in the magnitude of effects, as shown in Figure 36, which shows the difference (delta) between phrase-final and phrase-medial durations for each constriction gesture by pitch accent condition. In the case of three-syllable real words, the stimuli were not a minimal set, so the patterns in magnitude differences in raw duration is affected by inherent durational differences in vowel quality; the antepenultimate and unaccented test words ended in [a], whereas the final vowel was [e/i] for penultimate-accented words and [i] for ultimate-accented words.

Figure 36. The absolute magnitude of increase in raw duration from phrase-medial to phrase-final boundary condition for each constriction gesture by pitch accent condition in three-syllable real words. Delta refers to the difference between boundary conditions (phrase-final minus phrase-medial). Consonant and vowel constriction gestures are coded so that gestures belonging to the word-final syllable are C/V0 and consonant constriction gestures belonging to the antepenultimate syllable (two syllables away from the boundary) are C2. Gestures labeled with 'F' belong to the formation

phase of that gesture; those labeled with 'R' belong to the release phase of that gesture. Non-final vowel gestures were not measured and are not included in the figure.



4.3.7. Two-syllable real words: consonant constriction gestures

In the two-syllable real words, there was a significant three-way interaction between BOUNDARY, PITCH ACCENT, and SYLLABLE POSITION for formation duration (p<.01) and release duration (p<.001). Figure 37 and Figure 38 show the results for boundary lengthening by pitch accent condition for normalized and raw durations, respectively.

Figure 37. Boundary lengthening for normalized durations of the formation phase of C constriction gestures in two-syllable real words by pitch accent position.

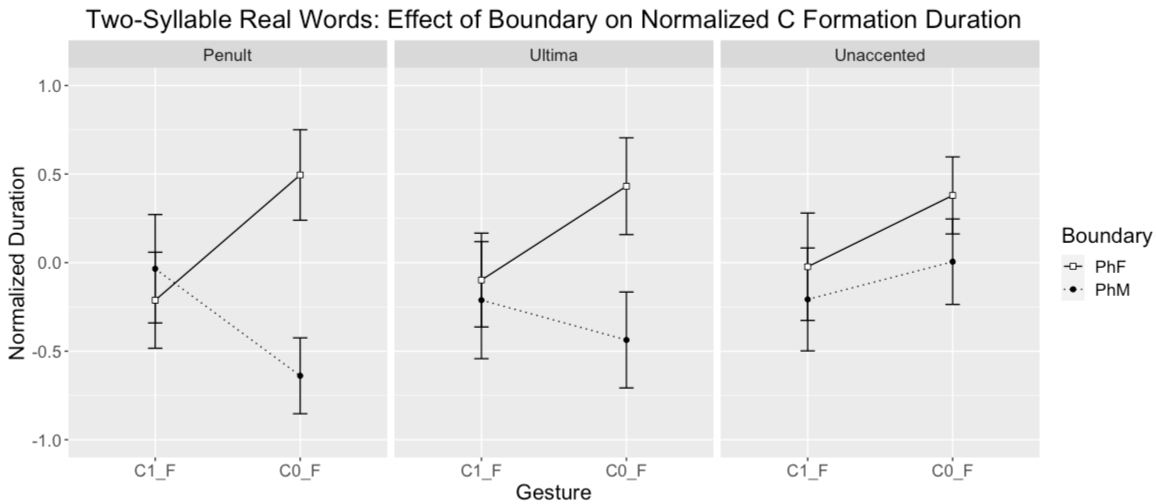
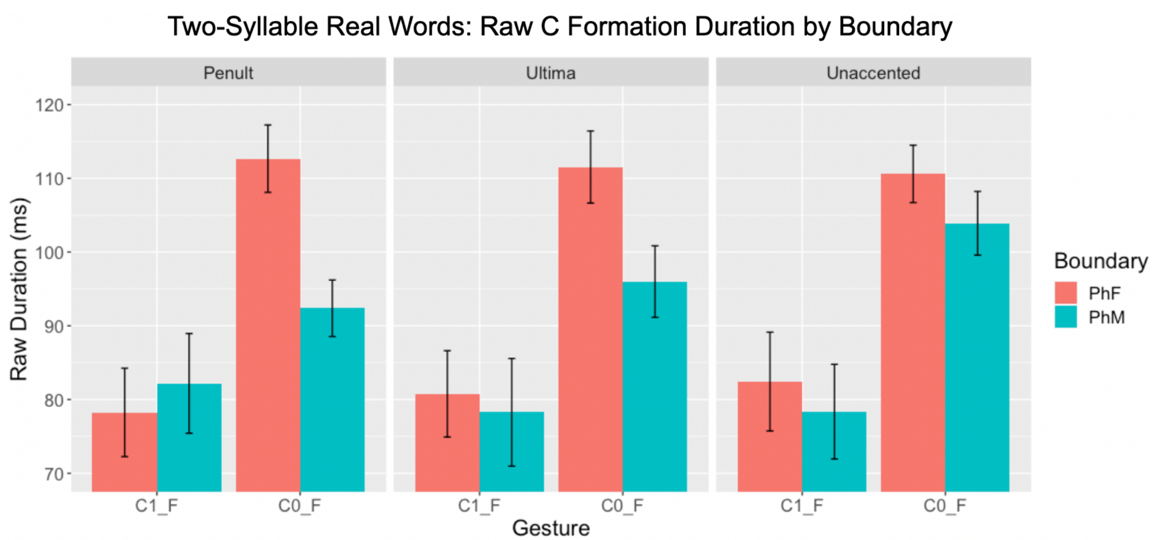


Figure 38. Raw durations for the formation phase of C constriction gestures in two-syllable real words by boundary and pitch accent position.



Phrase-final lengthening occurred in the formation gesture of the final consonant (C0) for all pitch accent conditions ($ß_{Penult}$=1.13, SE=0.17, p<.0001; $ß_{Ultima}$=0.87, SE=0.17, p<.0001; $ß_{Unaccented}$=0.38, SE=0.17, p<.05). In C0, penultimate and ultimate words appeared to have a greater durational difference between phrasal boundary conditions compared to unaccented words, but may be an effect of pitch accent, as in the phrase-medial condition, where there was no effect of phrase-final boundary, unaccented words were longer (unaccented: $ß_{Penult}$=-0.63, SE=0.17, *p*<.001; $ß_{Ultima}$=-0.45, SE=0.17, *p*<.05).

Figure 39 and Figure 40 show the results for the release phase of consonant constriction gestures by pitch accent position for normalized and raw durations, respectively.

Figure 39. Boundary lengthening for normalized durations of the release phase of C constriction gestures in two-syllable real words by pitch accent position.

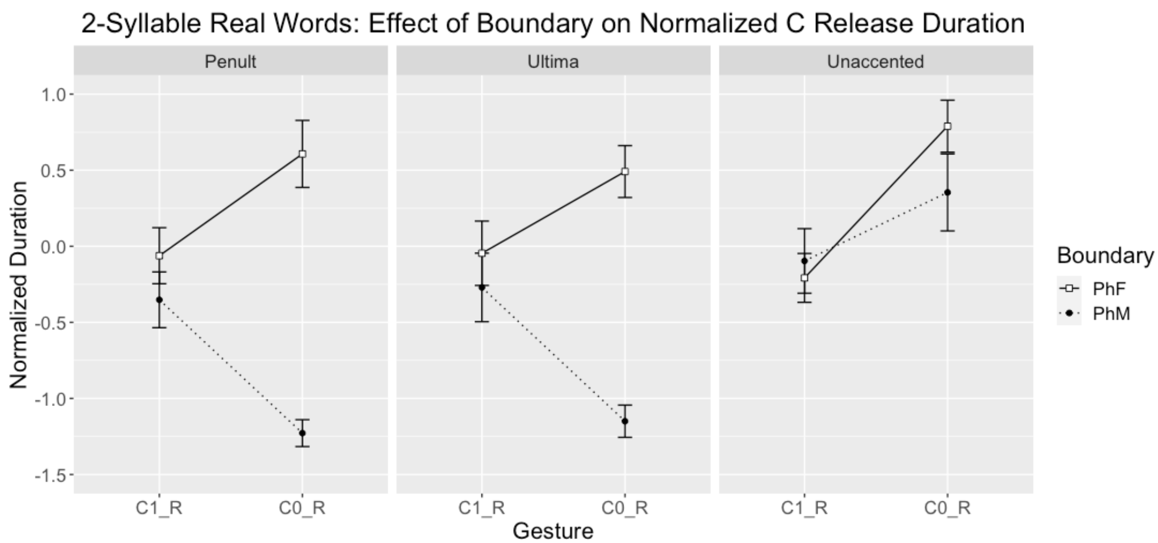
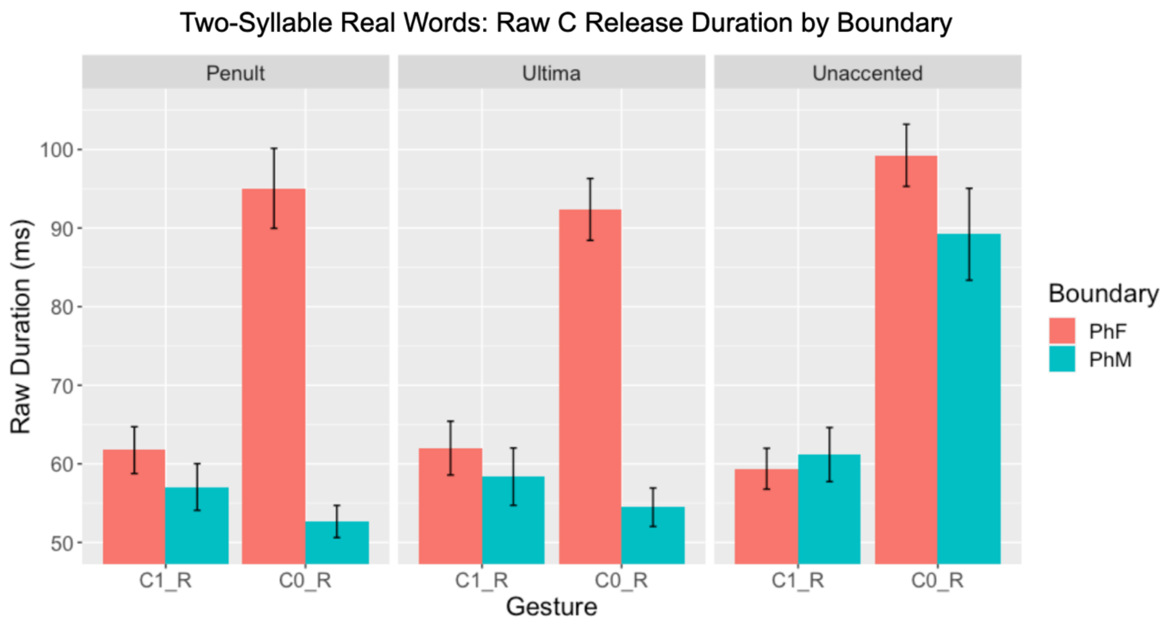


Figure 40. Raw durations for the release phase of C constriction gestures in two-syllable real words by boundary and pitch accent position.



All pitch accent conditions also saw phrase-final lengthening for release gesture of the final consonant (C0) ($ß_{Penult}$=1.18, SE=0.12, *p*<.0001; $ß_{Ultima}$=1.64, SE=0.12, *p*<.0001; $ß_{Unaccented}$=0.42, SE=0.11, *p*<.001). Again, there was a difference in the amount of lengthening, with penultimate- and ultimate-accented words undergoing more lengthening in C0. This was also due to unaccented words being longer phrase-medially in C0 (unaccented: $ß_{Penult}$=-1.57, SE=0.12, *p*<.0001; $ß_{Ultima}$=-1.52, SE=0.12, *p*<.0001).

4.3.8. Two-syllable real words: vowel constriction gestures

There was also a significant three-way interaction between BOUNDARY, PITCH ACCENT, and SYLLABLE POSITION in two-syllable word vowel gestures (*p*<.05). Figure 41 and Figure 42 show the results of boundary lengthening in the formation phase of vowel constriction gestures for normalized and raw durations, respectively.

Figure 41. Boundary lengthening for normalized durations of the formation phase of V constriction gesture in two-syllable real words by pitch accent position.

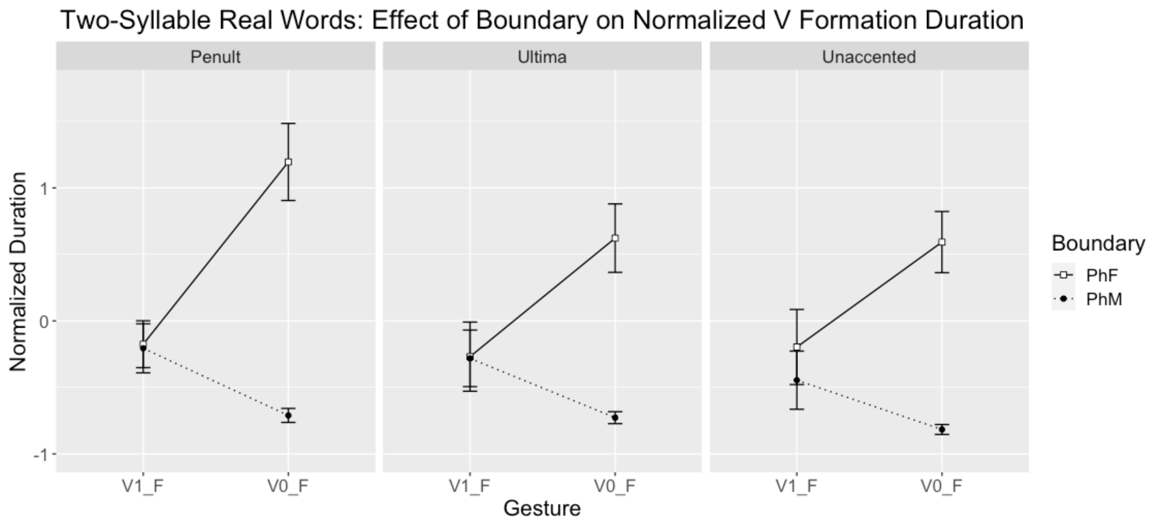
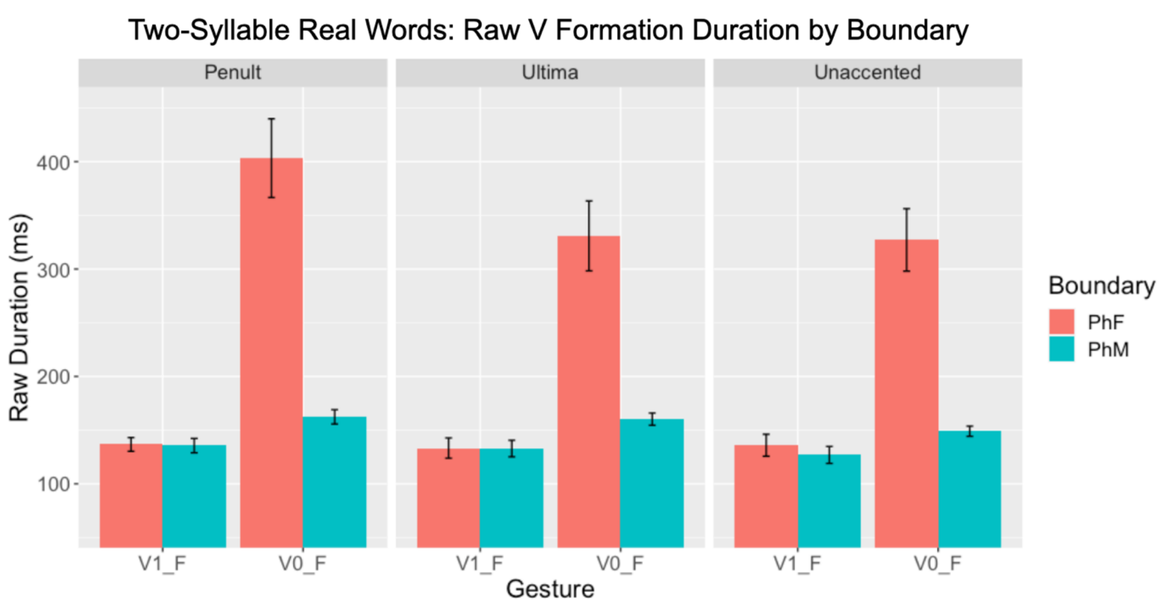


Figure 42. Raw durations for the formation phase of V constriction gestures in two-syllable real words by boundary and pitch accent position.



There was phrase-final lengthening in the final V0 gesture for all pitch accent conditions ($ß_{Penult}$=1.91, SE=0.13, *p*<.0001; $ß_{Ultima}$=1.35, SE=0.14, *p*<.0001; $ß_{Unaccented}$=1.40, SE=0.14, *p*<.0001). There was also a greater amount of lengthening in V0 for words with penultimate pitch accent (penultimate: $ß_{Ultima}$=0.57, SE=0.14, *p*<.001; $ß_{Unaccented}$=0.61, SE=0.14, *p*<.001).

4.3.9. Two-syllable real words summary

In Figure 43 and Figure 44, the results for two-syllable real words are summarized. In two-syllable real words, phrase-final lengthening was initiated in the formation phase of the final C constriction gesture. The amount of phrase-final lengthening also differed by pitch accent. In the formation and release gestures of the final consonant (C0), words with pitch accent (penultimate or ultimate) had a greater difference between boundary conditions than unaccented words. There was also an effect of pitch accent position in the final vowel gesture (V0), as words with penultimate pitch accent saw the greatest amount of phrase-final lengthening. The scope of boundary lengthening reached the formation of the final consonant gesture (C0).

Figure 43. Summary of results for the effect of phrase boundary on duration of C and V constriction gestures by pitch accent position.

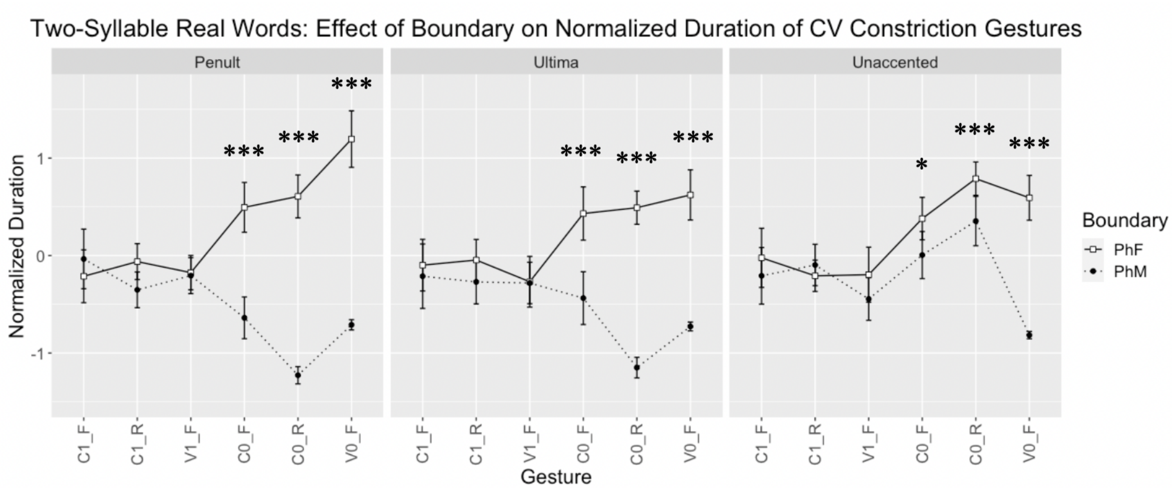
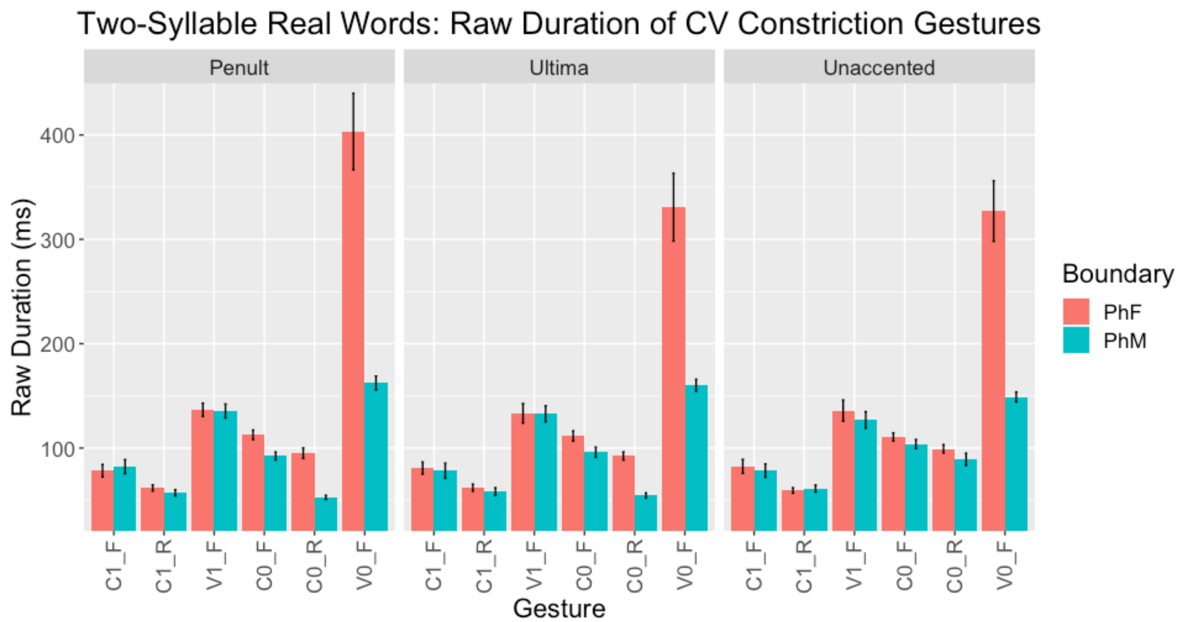


Figure 44. Raw durations of C and V constriction gestures by pitch accent position and boundary.



4.4. Discussion

In this study, we set out to better understand the interaction between word prosody and prosodic boundaries in Japanese by asking the following questions:

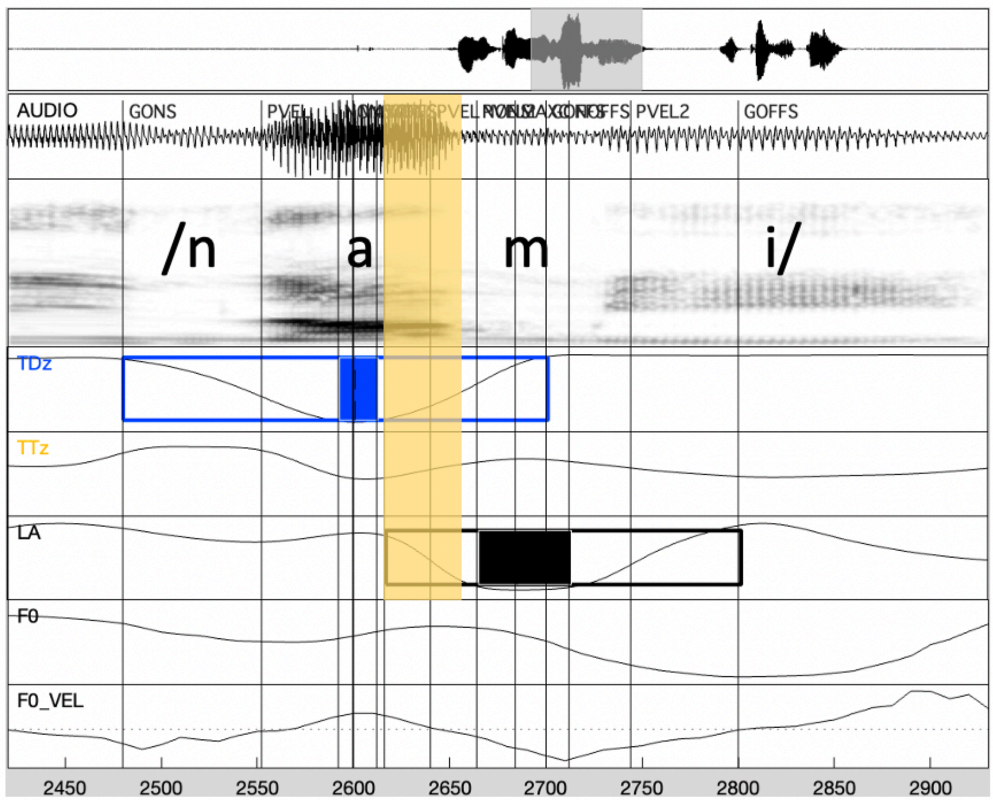
(i) Which constriction gestures in the phrase-final word are affected by phrase-final lengthening? Is it a continuous interval?

(ii) Does lexical pitch accent affect the scope of phrase-final lengthening?

We answered these by examining phrase boundary effects on duration in C and V constriction gestures in three sets of data: three-syllable neologisms, three-syllable real words (consonants and final vowel only), and two-syllable real words.

Based on previous work on Japanese, we had two hypotheses on the scope of phrase-final lengthening. The first was that lengthening effects occur mainly in the final rime (Shepherd 2008). The second hypothesis was that phrase-final lengthening effects begin in the vowel of the penultimate syllable (Seo, Kim, et al. 2019a, 2019b). The latter was the case in our neologism data, where boundary lengthening reached the V gesture of the penultimate syllable. In three-syllable real words, the penultimate V gesture was not measurable, but the C formation gesture was affected. In two-syllable real words, boundary lengthening was initiated also in the C formation gesture of the final syllable. However, constriction gestures in speech do not occur sequentially, and adjacent gestures overlap (see Figure 45). This means that here, the C formation gesture might coincide to some degree with the preceding vowel, so the finding that lengthening is initiated in C formation gestures of the final syllable does not necessarily contradict findings from acoustic studies where boundary lengthening reaches the penultimate vowel.

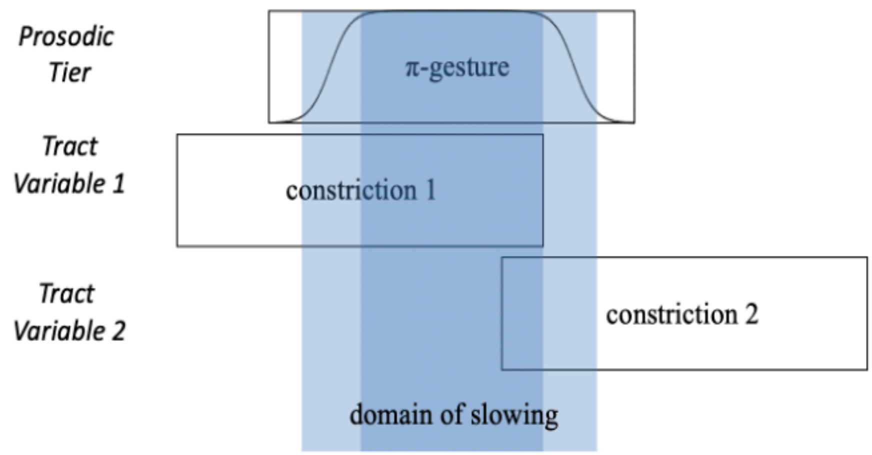
Figure 45. Overlap between /m/ gesture and the acoustic signal for /a/. The yellow highlighted area shows that the /m/ formation gesture occurs during the acoustic signal for the penultimate vowel /a/.



Overall, the final V0 constriction gestures, being immediately next to the phrase boundary, were more strongly affected than less proximal gestures, such as C formation gestures. This effect is also in concordance with the π-gesture account, in which the domain of prosodic lengthening is fixed at the boundary, and gestures closer to it experience larger clock-slowing effects.

On the effect of pitch accent on boundary-related lengthening, previous studies on the phenomenon in Japanese have also found that pitch accent affects the amount of lengthening (Seo, Kim, et al. 2019a, 2019b). While Seo, Kim, et al. (2019a, 2019b) found less lengthening in two-syllable words with initial pitch accent compared to unaccented words, the current study, which uses articulatory data and is able to compare antepenultimate vs penultimate pitch accent position, finds that words with pitch accent undergo a greater amount of lengthening. It is important to note that Seo, Kim, et al. (2019a, 2019b) used acoustic data and only examined two-syllable words with initial accent vs. unaccented words; they did not include stimuli to test the effect of different positions of pitch accent. Also, their test words did not come at the end of the utterance, which means their phrase-final test words did not have utterance-final boundary tones, an aspect which will be investigated and discussed further in the following chapters. Interestingly, in our data, pitch accent did interact with the scope of effects. Pitch accent position had an effect on the amount of lengthening, such as in the final V gesture in all three sets of data, where antepenultimate and penultimate words had greater phrase-final lengthening than ultimate or unaccented words. Within the π-gesture account, the degree of lengthening depends on a constriction gesture's overlap with the π-gesture, which has a steady increase in activation approaching the boundary and a maximum level of activation at the boundary itself. Figure 46 is a schematic representation of a π-gesture.

Figure 46. Schematic diagram of a π-gesture and two constriction gestures at the prosodic boundary. Figure adapted from Byrd and Saltzman (2003).



Blue shading indicates activation of the π-gesture, with darker blue indicating the maximum level of activation. Overlap of the blue shading with constriction gestures represents local slowing effects on those co-active gestures. Therefore, a greater amount of lengthening may be accounted for by greater overlap of the constriction gesture with the maximum activation area of the π-gesture. The finding that words with earlier pitch accent experience greater amounts of lengthening earlier in the word suggests earlier overlap with higher levels of π-gesture activation.

The π-gesture account has two proposals for such a difference: (a) coordination shift, in which the entire π-gesture is shifted towards the pitch accented syllable, or (b) extension, in which the π-gesture is extended towards the pitch accented syllable. However, we are primarily interested in the third account offered in Katsika (2014), which proposes a coordination between clock-slowing gestures, where the μ-gesture associated with the lexically stressed syllable is coordinated with the π-gesture at the phrase edge. So far, the interaction of pitch accent position with phrase-final lengthening in Japanese also suggests that something is affecting π-gesture coordination. However, it is unclear whether the cause is a μ-gesture, as in Greek, or something else. The account in Katsika (2014) also connected this issue with the coordination of boundary tones, which were also found to be initiated later in words with stress later in the word. Similarly, it follows that we investigate the role of lexical pitch accent position and boundary tone coordination to further understand the components involved in π-gesture coordination in Japanese.

5. Boundary tone coordination and pitch accent

This chapter investigates the coordination of boundary tones with constriction gestures in Tokyo Japanese and the role of word prosody in this relationship. Previous studies have found systematic patterns of tonal alignment, or the timing of tonal events with speech segments, and found that F0 turning points (F0 minima and maxima) are coordinated in time with specific acoustic events in the segmental string (Ladd 2008). Boundary tones which mark the end of major phrase boundaries are expected to occur within the phrase-final syllable (e.g., Silverman, Beckman, et al. 1992; Beckman and Pierrehumbert 1986). Results from recent research on Greek indicates that boundary tone coordination may be related to lexical stress, as boundary tones were found to initiate earlier when stress was earlier in the word (Katsika, Krivokapić, et al. 2014). However, unlike Tokyo Japanese, word prosody in languages like Greek or English can serve as anchors for phrasal pitch accents. The Greek study of boundary tone coordination in Greek systematically examined these issues separately and found that the result that boundary tones were initiated earlier when stress was earlier in the word held for accented and de-accented phrase-final words, indicating that this effect was not affected by phrasal prominence (Katsika, Krivokapić, et al. 2014).

The account proposed for Greek within the Articulatory Phonology framework involves μ-gestures, which are coordinated in-phase with the constriction gestures of stressed syllables. Katsika (2016), which analyzed the interaction between word prosody and boundary tone coordination along with the interaction between word prosody and phrase-final lengthening, proposed that the μ-gestures are also coordinated with the clock-slowing π-gesture at the prosodic boundary, which has anti-phase coordination with the phrase-final vowel. The boundary tone is triggered by the π-gesture, and its timing is therefore affected by effect of lexical stress on π-gesture coordination (Katsika 2016). In Chapter 4, we found that similar to Greek, the scope of phrase-final lengthening interacts with the position of word prosody. While we have not determined whether there may similarly be a μ-gesture involved in these patterns in Japanese, this chapter continues the investigation by examining the interaction of word prosody and boundary tone coordination in Japanese. Japanese is a language in which word- and phrase-level prosody are both realized in the tonal domain, so an interaction would not be surprising. We will look at the timing of the final H boundary tone initiation with constriction gestures, investigate boundary tone coordination as a function of lexical pitch accent position, and discuss the implications of these interactions between lexical and phrasal levels.

5.1. Questions and predictions

The current study aims to answer the following research questions: (i) how are boundary tones coordinated with constriction gestures in Japanese, and (ii) does boundary tone coordination interact with the position of lexical pitch accent?

Based on previous research, we expect boundary tones to occur within the phrase-final syllable (Silverman, Beckman, et al. 1992; Beckman and Pierrehumbert 1986). We also expect to see pitch accent marked by a high (H) tone on the accented syllable, followed by a peak and sharp fall in F0 to a low (L) tone on the following mora. As a reminder, pitch accent in Japanese is typically represented as 'H*+L', although the X-JToBI annotation system represents accent simply as 'A' because the sharp fall in F0 is predictable, but the location of pitch accent is lexically defined. Since pitch accent requires a following mora on which to realize the pitch fall, we also expect that phrase-final ultimate-accented words, where the pitch accented syllable is not followed by another syllable, may not show such a pitch fall. This follows the tonal crowding hypothesis, where in the case of multiple tones "too close together," with not enough syllables per tones, tones may be compressed or truncated (Ladd 2008). The pitch contour of test utterances containing phrase-final ultimate-accented words may therefore look similar to those of phrase-final unaccented words. Note that while some studies have investigated differences between the two such as in the height of F0 (Warner 1997; Matsui and Hwang 2018; Poser 1984; Pierrehumbert and Beckman 1984; Beckman and Pierrehumbert 1988; Venditti 2005), the distinction between ultimate-accented and unaccented words is beyond the scope of the current study and will not be discussed in depth.

There are a few possibilities for the gestural coordination of boundary tone gestures with the constriction gestures forming the V gesture of the phrase-final syllable, based on previous studies. Gao (2008) showed that word-level tones in Mandarin can affect C-V coordination, which could be seen in a c-centering effect on C gesture coordination, where C gestures were and lexical tones were in-phase with the vowel but anti-phase coordinated with each other. This results in respective left-ward and right-ward shifts in timing for the C gesture and tone gestures. Research on phrase-level pitch accents in German (Mücke, et al. 2012), suggests that one difference in word-level and phrase-level tonal coordination is that word-level tones are integrated in the coupling relations of the syllable, so that tone gestures may affect C-V coordination, while phrase-level tones are coupled with V gestures without affecting C-V coordination. Following these studies, we may expect that word-level pitch accents in Japanese are in-phase with the vowel, and may have effects on the coordination of consonant constriction gestures. Boundary tones are phrase-level; in this case, we expect that the LH% boundary tone used in the test utterances to be L and H tone gestures that are anti-phase with each other, being sequential to produce a rise.

Assuming that boundary tones are similar to other tones and also coordinated with constriction gestures, our first hypothesis is that the boundary tone gestures are in anti-phase coordination with the V gesture and occur sequentially with the V gesture, occuring near the onset of the V target (Hsieh 2011). A second hypothesis is that boundary tone gestures in Japanese have in-phase coordination with the phrase-final V gesture, beginning simultaneously with the start of the V gesture (Gao 2008; Mücke, et al. 2012). A third possibility is that boundary tone gestures are in-phase with the V gesture but anti-phase with the C gesture, which may result in a c-center effect where the onset of the C gesture and the onset of the boundary tone gesture shift in relation to the V gesture. Alternatively, boundary tones could be aligned with the time of peak velocity of the V gesture (d'Imperio, et al. 2007).

On the effect of lexical pitch accent on boundary tone coordination, previous studies on stress languages have reported that the onset of edge tones appears to vary with the position of lexical stress (Katsika, Krivokapić, et al. 2014). Our hypothesis for the effect of lexical pitch accent on boundary tone coordination is therefore that lexical pitch accent, which operates on the same tract variable as boundary tones (i.e. vocal fold vibrations), would also have some interaction with boundary tones. Acoustic studies describe pitch accent as an H*+L tone pattern which is expected to begin in the accented syllable and peak in the following mora; from an articulatory perspective, this means we can expect an H* tone gesture to begin in the accented syllable, followed by a L tone gesture which begins in the following mora. In the case of penultimate-accented words, this L tone would be expected within the word-final syllable, which is also where we expect the phrase-level boundary tones. Therefore, we predict that in penultimate-accented words, boundary tone initiation may be shifted later compared to words with antepenultimate pitch accent.

Given that word prosody in Japanese occurs in the tonal domain, an effect of pitch accent position on boundary tone coordination in Japanese may be explained through tonal crowding. In the phrase-final syllable, we may expect that a word with pitch accent in the preceding syllable (penultimate accent) may have more tones "crowded" at the end of the word compared to a word with pitch accent earlier in the word (e.g., antepenultimate accent), and thus expect boundary tones to be initiated later in penultimate-accented words because of this "crowding". However, another motivation for examining the effect of pitch accent position on boundary tone coordination is that in Greek, the effect of lexical stress on the timing of boundary tone initiation was proposed to be related to a similar effect of stress position on the scope of phrase-final lengthening. In Greek, both phrase-final lengthening and boundary tones began later in words with stress later in the word. The account proposed to connect these parallel findings in the temporal and tonal domains was that the μ-gesture associated with lexical stress was coordinated with the π-gesture associated with the phrase boundary. The π-gesture is assumed to trigger the boundary tone, so this coordination may explain the connection between lexical stress position and the scope of π-gesture-related slowing effects. This relates to our investigation on Japanese, as the results from Chapter 4 showed a similar pattern to Greek, as lexical pitch accent position also had an effect on the scope of phase-final lengthening. Thus, this prediction also arises from previous findings on the effects of stress, another word prosodic marker, on boundary tone coordination (Katsika, Krivokapić, et al. 2014; see also Arvaniti and Ladd 2009). If, like Greek, we find parallel patterns in the effect of pitch accent position on phrase-final lengthening and boundary tone coordination, it will be important to interpret the results in light of both the temporal and tonal interactions, as tonal crowding (which alone cannot account for the Greek findings) may not be the most complete explanation.

5.2. **Methods**

This chapter includes two analyses: (i) a tonal analysis using X-JToBI (Maekawa, Kikuchi and Igarashi, et al. 2002), and (ii) a kinematic analysis using linear mixed effects models to test boundary tone coordination as a function of pitch accent position.

The data for the analyses in this chapter are a subset of the phrase-final data presented in Chapter 3 (see Table 2 and Table 3). Data from four female native Tokyo Japanese speakers were included in this analysis; data from one speaker were omitted due to the speaker's naturally low, creaky voice, which did not allow for measurements of a turning point from L to H in the final syllable. Of the words in which we expected measurable H and L tones for the pitch accent (i.e., antepenultimate- and penultimate- accented words), about 12% of data were omitted due to creakiness.

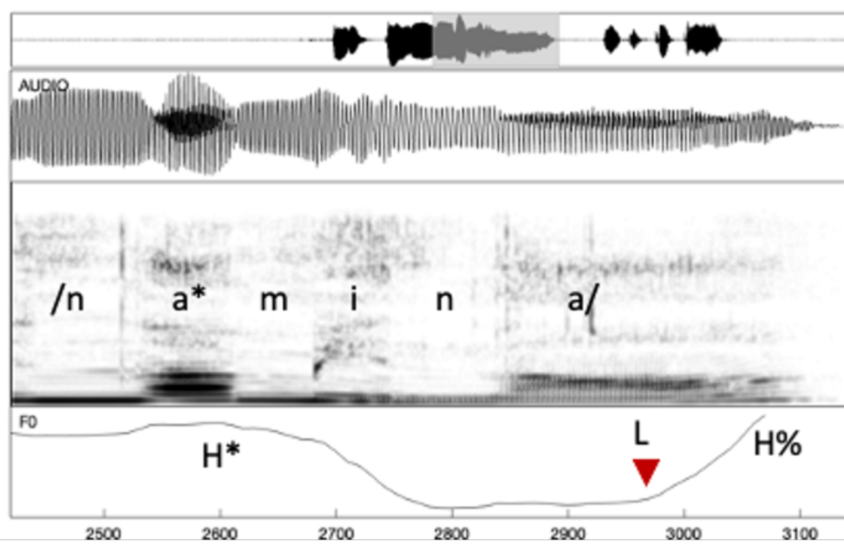
5.2.1. Labelling

Prior to annotation, all utterances were checked by a native Japanese speaker who confirmed each test utterance was produced fluently and perceived with the intended meaning. The tonal analysis was conducted in Praat (Boersma and Weenink 2022) using X-JToBI annotation conventions. X-JToBI was chosen over JToBI because its notation convention boundary tones aligns with the landmarks used for labels in our kinematic data. Namely, X-JToBI marks phrase-final complex boundary tones differently from JToBI. Whereas an LH%

complex boundary tone in JToBI is marked at the right edge of the IP, X-JToBI additionally marks the start of the L boundary tone with 'pL' to indicate where the rise from L to H begins. This is similar to the landmark we used to label the onset of the H boundary tone in the kinematic data, excepting that we also used available F0 velocity information to mark the turning point in the kinematic data.

The kinematic data were labeled using a custom program in Matlab (Tiede, Haskins Laboratories) to manually mark the F0 turning point from L to H% detected based on a 20% increase in the F0 velocity (see Figure 47). This turning point is used as the onset of the H tone in the LH% question boundary tone. Kinematic landmarks for the word-final consonant and vowel were also labeled (see Chapter 3 for a detailed description of these labeling methods). Then, the interval between H boundary tone turning point and the onset of the vowel in the final syllable was calculated.

Figure 47. Sample label of tonal landmarks for na*mina 'Namina (name).'



5.3. Results

5.3.1. X-JToBI analysis

The tonal analysis showed that words with antepenultimate and penultimate pitch accent had a clear LH% rise within the final syllable of the phrase-final word (representative examples provided in Figure 48 and Figure 49). There was also a clear fall from H to L to mark pitch accent in these words (marked by A for accent in the following figures).

Figure 48. F0 contour of antepenultimate-accented test word na*mina 'Namina (name)'.

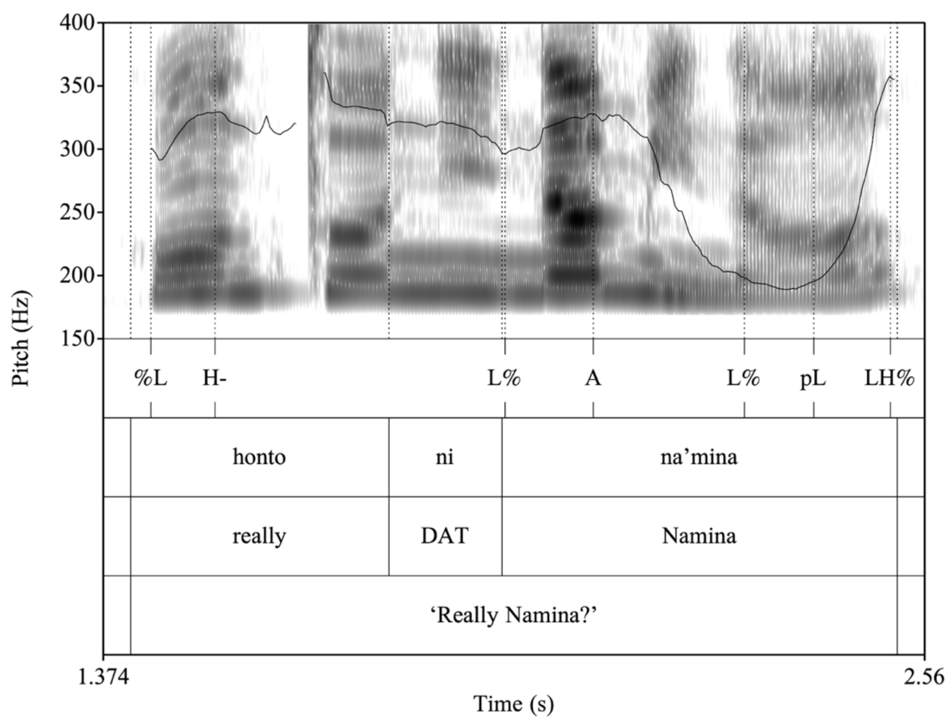
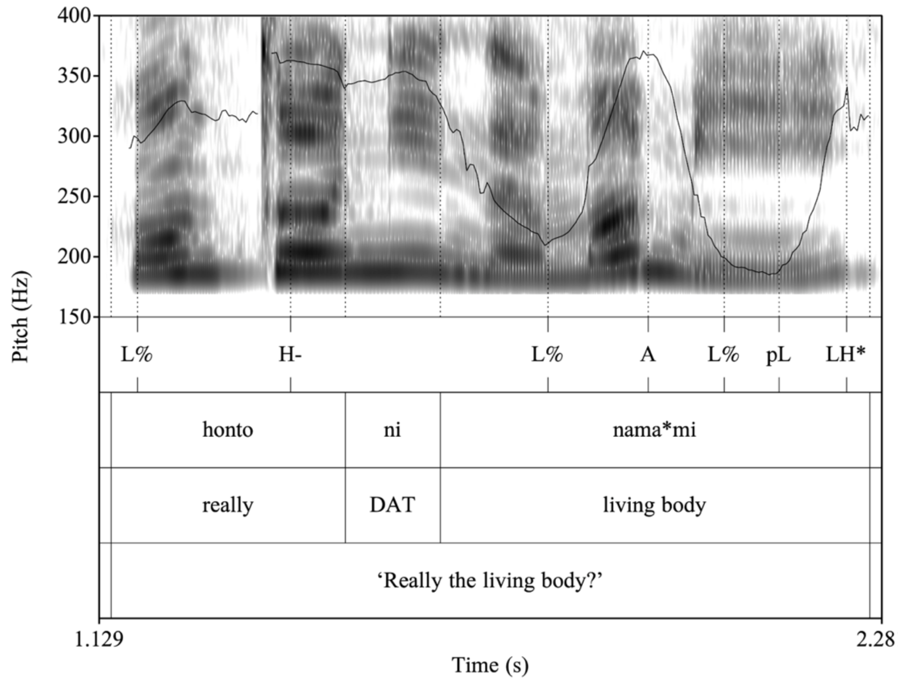


Figure 49. F0 contour of penultimate-accented test word *nama*mi* 'living body'.



Since pitch accent in Japanese is realized by a fall in F0 on the following mora, we expected that pitch accent located on the ultimate syllable of a phrase-final word has no room for realization, resulting in final-accented words which may be indistinguishable from unaccented words (Kindaichi 1947; McCawley 1968). The X-JToBI analysis of final-accented words in our data showed a distinct lack of any pitch fall in the word (compare **Error! Reference source not found.** and Figure 49 vs. Figure 50) suggesting that the pitch accent was not realized. Future work into F0 height could help clarify whether the tonal contour of ultimate-accented and unaccented conditions are indistinguishable, as previous studies have shown that phrasal H is lower than the H* of a pitch accent (Warner 1997; Matsui and Hwang 2018). These studies may also differ from the current study, which uses a complex LH% frame sentence, as opposed to a L% frame sentence. However F0 height was not measured for the current study, so this topic will not be discussed further here. In any case, ultimate-accented and unaccented words had similar pitch contours, showing a more gradual rise to a final H% such that there was no identifiable turning point based on a 20% increase in F0 velocity. Therefore, a turning point for the H boundary tone was not labeled for these pitch accent conditions. While a precise turning point could not be labeled, the figures below still show that there is a clear LH% rise which occurs within the phrase-final syllable.

The X-JToBI analysis for the tonal contour of two-syllable word revealed similar patterns. In Figure 52, there is a clear fall from high to low after the first syllable of *nami* 'Nami (name)' which indicates pitch accent on the initial syllable. However, the pitch contour for final-accented *nami* 'wave' in Figure 56 does not have the characteristic pitch fall indicative of accent and looks more similar to the unaccented *nami* 'medium' in Figure 57.

Figure 50. F0 contour of ultimate-accented test word *nigami 'bitterness'.**

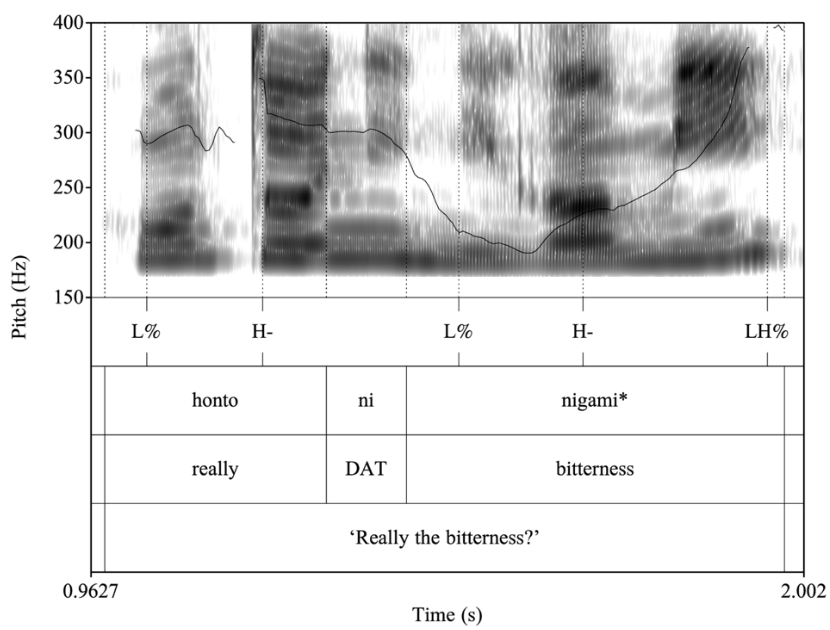


Figure 51. F0 contour of unaccented test word *namida* 'Namida (name)'.

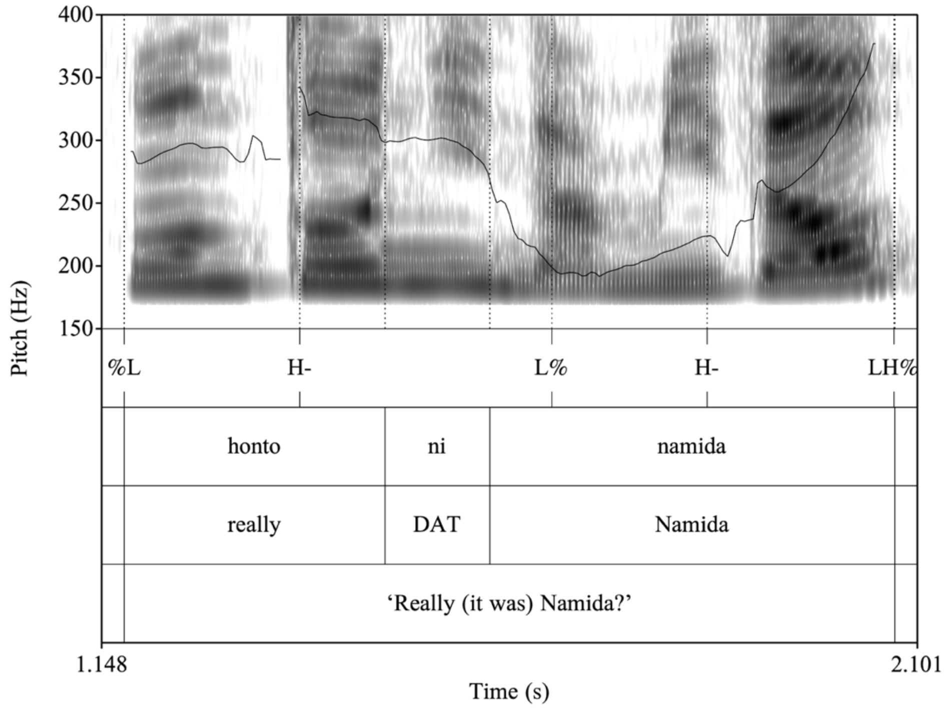


Figure 52. F0 contour of initial-accented test word *nami* 'Nami (name)'.

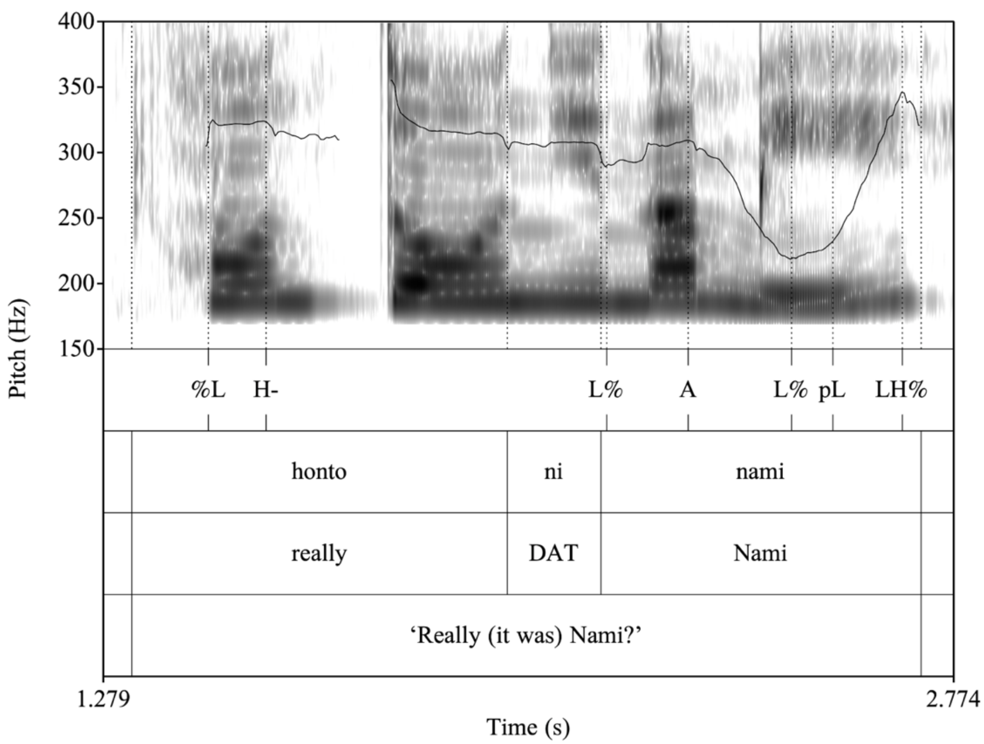


Figure 53. F0 contour of final-accented test word *nami* 'wave'.

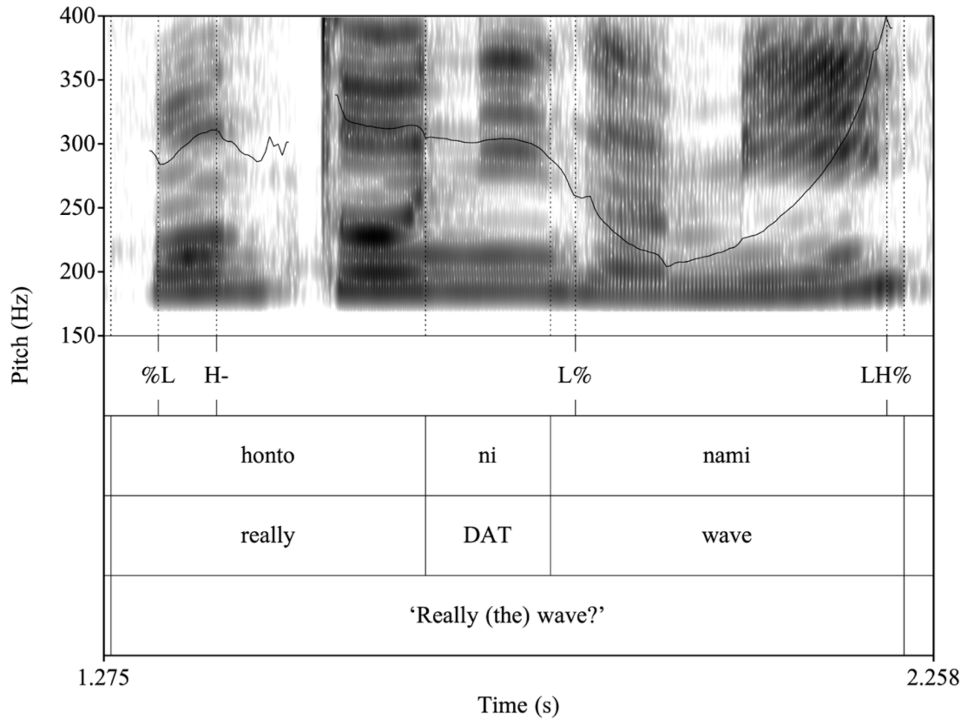
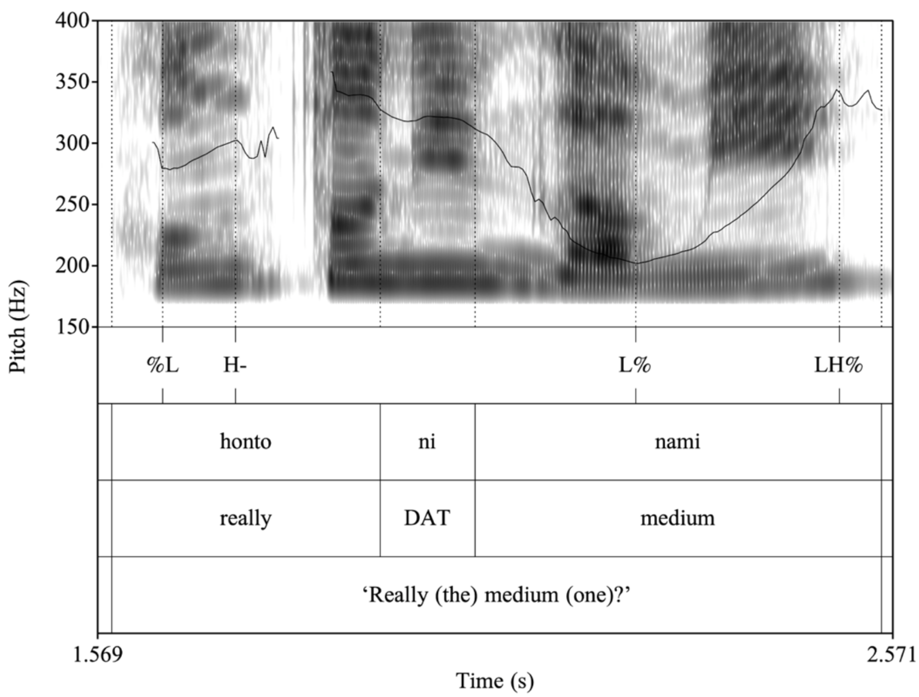


Figure 54. F0 contour of unaccented test word *nami* 'medium'.



5.3.2. Boundary tone coordination

We then conducted a series of analyses on boundary tone coordination and its interaction with pitch accent. Based on the results of the X-JToBI analyses, ultimate-accented and unaccented words did not have a measurable turning point and were therefore excluded from the following analyses. Test words were therefore a near-minimal pair of three-syllable real words and a minimal pair of neologisms with antepenultimate or penultimate pitch accent. A comparison of these two conditions allows us to examine the timing of the H boundary tone and the effect of pitch accent position. The results from Greek, which found that word prosody may affect the timing of boundary tones, along with conditions for tonal crowding, lead us to predict that in Japanese, penultimate-accented words will have a later boundary tone initiation compared to antepenultimate-accented words.

To examine the effect of pitch accent on boundary tone coordination, linear mixed effects models were run with the normalized interval between the onset of the vowel and the onset of the boundary tone turning point in the final syllable (i.e., V-onset to BT-onset) as the dependent variable, PITCH ACCENT [levels: antepenultimate, penultimate] as the fixed factor, and random intercepts by Speaker. The reader is reminded that V-onset refers to the kinematic onset of the V constriction gesture, and not the onset of the acoustic signal; this differs from older studies of tone timing such as Silverman and Pierrehumbert (1990) for English and Venditti, Maeda and Santen (1998). The interval was normalized by the length of the vowel to account for effects of vowel duration on the relative timing of boundary tone initiation, as vowel height may affect vowel duration (Han 1962; Port, Al-Ani and Maeda 1980; Beckman 1982; Homma 1981). A shorter interval would indicate earlier initiation of the H boundary tone relative to the vowel, and a longer interval a later initiation of the H boundary tone.

Separate models were run for real words and neologisms. The model selection process began with a maximal model that included random slopes and intercepts by the predictor per Speaker, but models with random slopes by predictor did not converge. The final models for real words and neologisms included random intercepts by Speaker.

Both models showed a significant effect of pitch accent position, with boundary tones being initiated earlier in antepenultimate words compared to penultimate words, as predicted. The effect of pitch accent position was larger in real words (β=1.42, SE=0.21, p<0.01) than neologisms (β=0.79, SE=0.26, p<0.01). The results shown in Figure 55 indicate the effect of lexical pitch accent position on the timing of H boundary tone initiation.

Figure 55. Normalized interval of BT-to-V onset per pitch accent in real words and neologisms. A longer interval indicates later initiation of the H boundary tone.

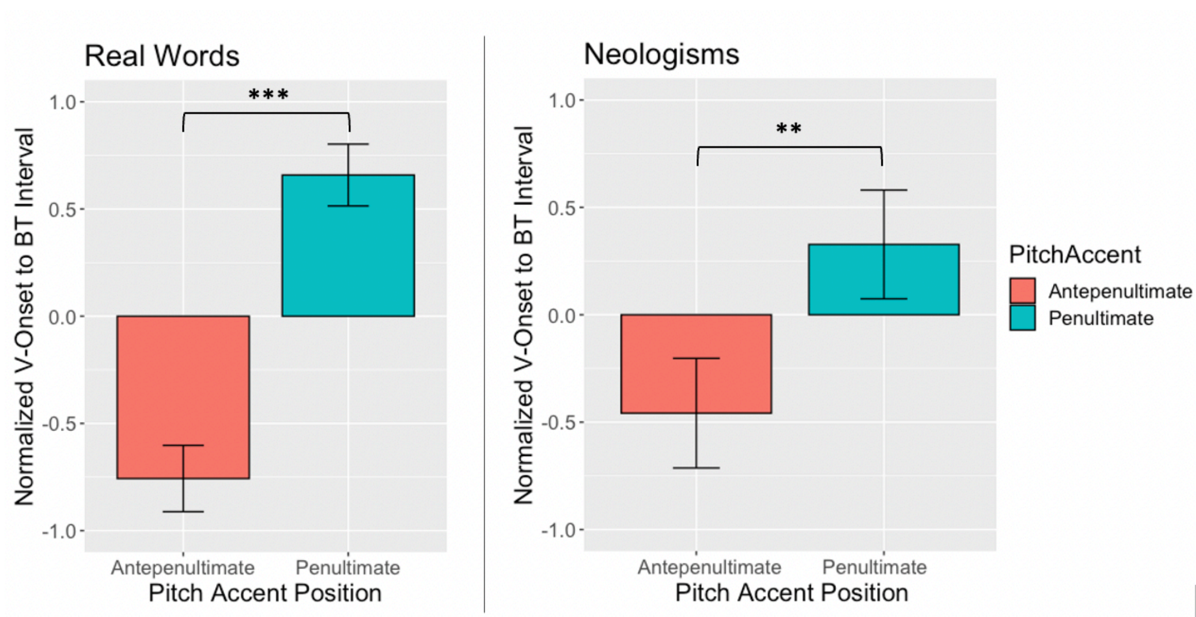


Figure 56 and Figure 57**Error! Reference source not found.** schematize these results as gestural scores to show the initiation of the H boundary tone relative to the onset of the phrase-final V gesture, which are set to 0 on the number line. The figures show consonant and vowel constriction gestures of the phrase-final syllable as well as the boundary tone H gesture. The start of the L tone is unclear and therefore not marked in the figures, only indicted with a grey dashed box.

Figure 56. Coordination of gestures in the phrase-final syllable by pitch accent position for three-syllable real words. The solid vertical line inside the solid rectangle indicates the time-point of peak velocity within the gesture. The dotted lines indicate the start and end of the gesture's target. Onset of vowel gestures are set to 0 on the number line, so that the timing of all gestures can be seen relative to the start of the V gesture.

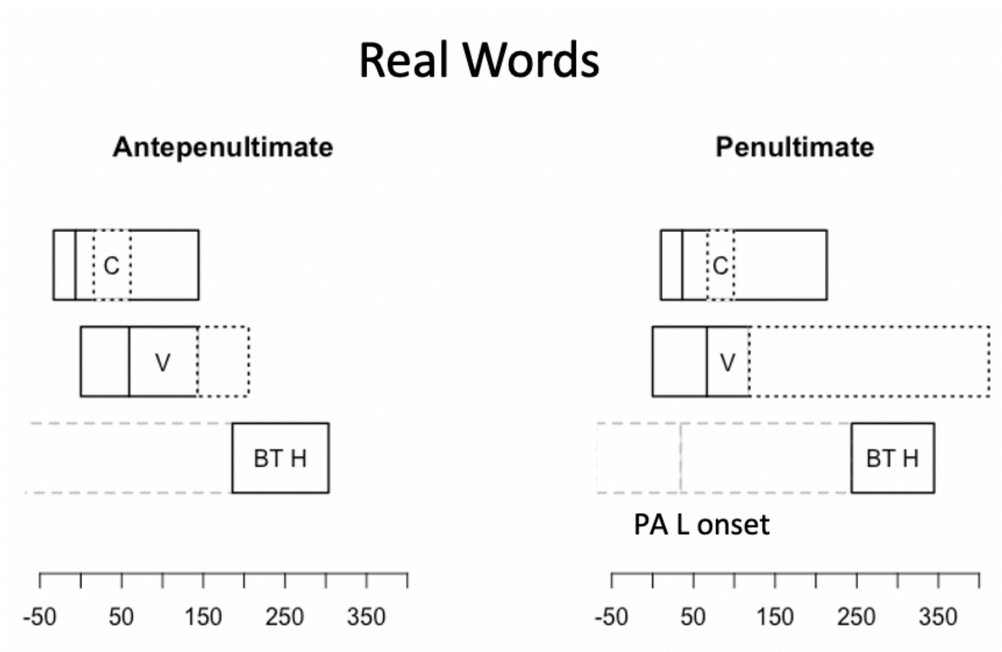
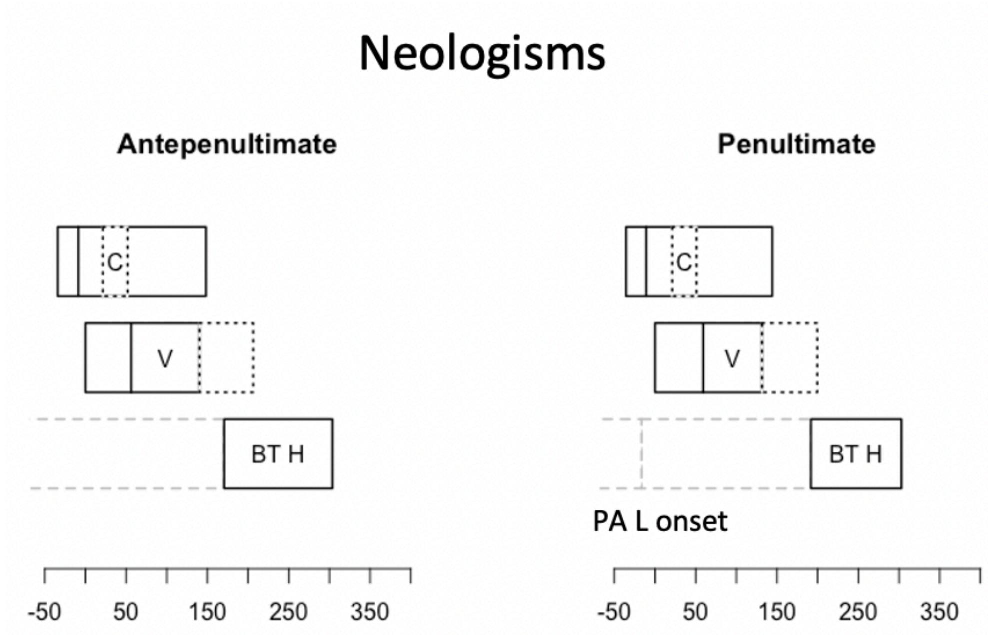


Figure 57. Coordination of gestures in the phrase-final syllable by pitch accent position for three-syllable neologisms.



Because the boundary tone in these utterances is a complex LH%, the H boundary tone is assumed to be anti-phase with the L boundary tone, as it occurs sequentially to produce a rise. Indeed, the timing of the H boundary tone occurs after the target of the final vowel constriction gesture is reached. The boundary tone offset is measured as the maximum F0, although it should be noted that this may be affected if there is creakiness or loss of phonation at the end of a phrase. In the penultimate-accented words, the onset of the pitch accent L tone gesture was also measurable; this is labeled in the figures with a vertical grey dashed line. The left edge of the pitch accent H gesture is left open, as the onset was not measured. Note that the gestural score for penultimate real words has a longer vowel target duration; this may be related to vowel quality, as the gestural score reflect means of raw duration and are not normalized to account for different phonemes. The penultimate real word condition is the only vowel which is /i/; all other vowels shown here are /a/.

There are two notable observations from these gestural scores. First, pitch accent position does appear to affect the timing of the boundary H tone, which occurs later in the vowel in penultimate-accented words. Second, the onset of the pitch accent L tone appears to be well-aligned with the peak velocity of the C gesture. Notably, regardless of whether the L tone begins slightly before or after the V gesture, the onset of the pitch accent L tone looks to occur in tandem with the C gesture's peak velocity timepoint. To further investigate this observation, the following analyses were conducted on words with penultimate pitch accent, where there timing relationships between the pitch accent L tone and H boundary tone in the final syllable could be measured. We examined the stability of the H boundary tone with the C gesture, alignment and stability of the pitch accent L tone with various kinematic landmarks, and the stability of the H boundary tone gesture with the pitch accent L tone gesture.

5.3.2.1. Stability of the H boundary tone gesture with kinematic landmarks

Alignment of the H boundary tone with kinematic landmarks was not tested here, since we expect certain predictions to hold. Namely, we expect that boundary tones will occur in the phrase-final syllable, and that the H tone should come after the other tones in the vowel, being anti-phase with its L component in a complex LH% boundary tone. To examine the stability of the H tone gesture's coordination with the V constriction gestures, the temporal intervals between the onset of the H tone and the following kinematic landmarks of the phrase-final syllable were calculated:

- Onset of V gesture (V-onset)
- Peak velocity of V gesture (V-pvel)
- Target of V gesture (V-target)

We ran linear mixed effects models for penultimate-accented real words and neologisms with standard deviation (in ms) as the dependent variable and INTERVAL ORIGIN (V-onset, V-pvel, V-target) as fixed factors, and the same model selection process as above. Models with random slopes did not converge, and the final model included random intercepts by Speaker. The models showed main effects of INTERVAL ORIGIN for real words ($F(2)=93.12$, $p<.001$) and neologisms ($F(2)=3.94$, $p<.05$). Figure 58 and Figure 59 show the standard deviations for each interval origin; a smaller standard deviation indicates a more stable coordination with the respective timepoint. Based on post-hoc pairwise comparisons using emmeans with Holm correction, the H tone to V target interval was the most stable, and the H tone to V onset interval was the least stable.

Figure 58. Stability of H boundary tone to kinematic landmarks within the phrase-final vowel in real words. Smaller standard deviation indicates more stable coordination with the respective timepoint.

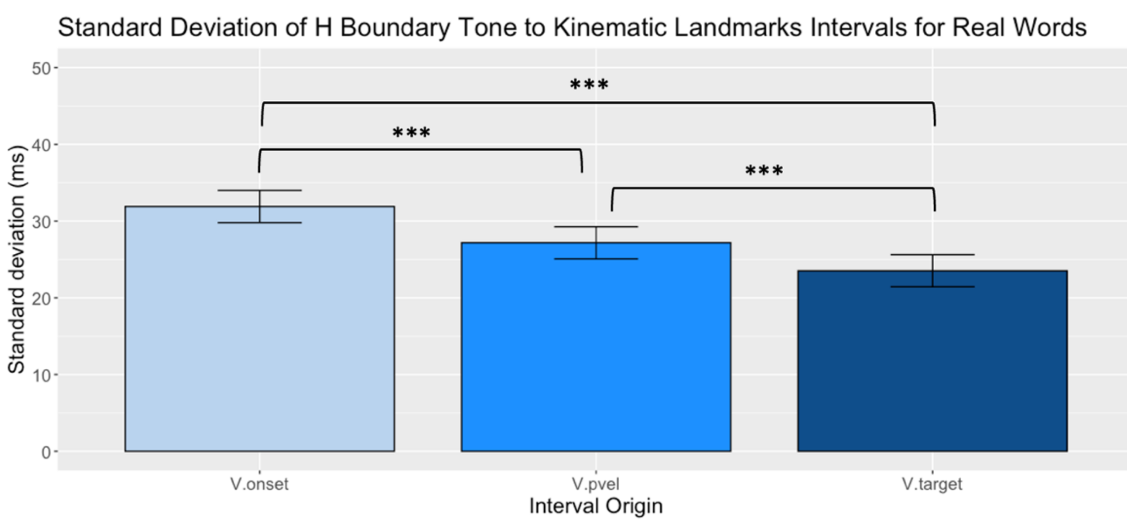
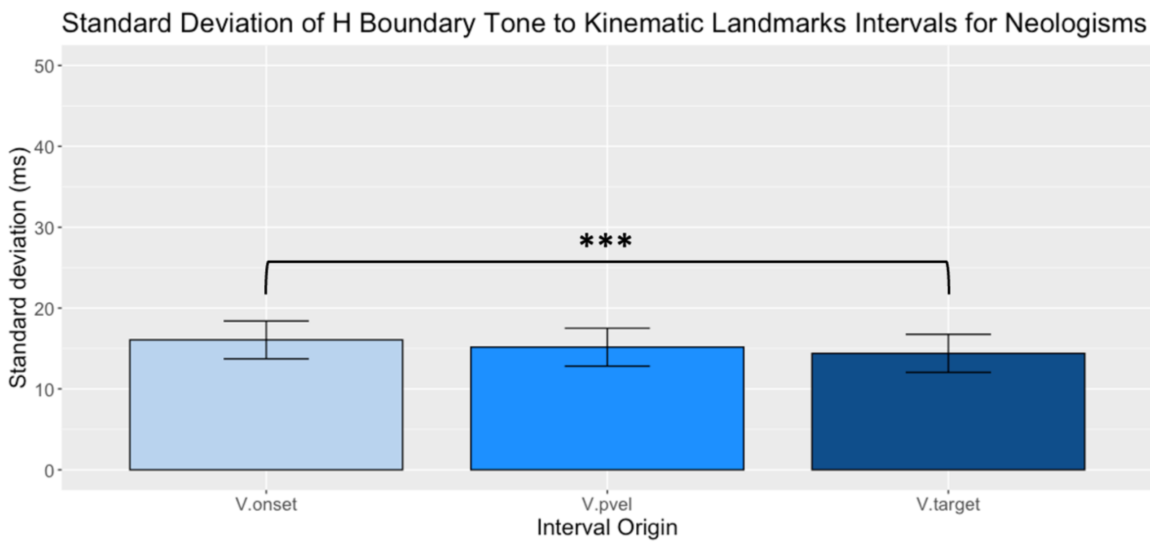


Figure 59. Stability of H boundary tone to kinematic landmarks within the phrase-final vowel in neologisms. Smaller standard deviation indicates more stable coordination with the respective timepoint.



5.3.2.2. Alignment of the pitch accent L tone gesture

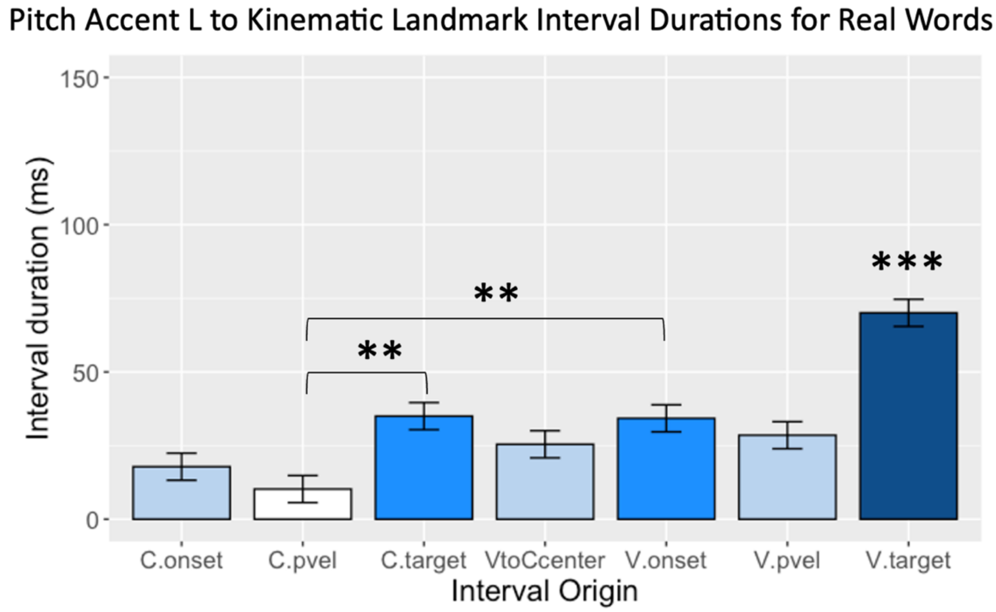
To better understand how pitch accent effects the coordination of the H boundary tone in the phrase-final syllable, we then examined the timing of the pitch accent L tone gesture with V and C constriction gestures by using the temporal intervals between the onset of the pitch accent L and the following kinematic landmarks of the phrase-final syllable:

- Onset of C gesture (C-onset)
- Peak velocity of C gesture (C-pvel)
- Target of C gesture (C-target)
- Interval between V onset and the c-center (VtoCcenter)
- Onset of V gesture (V-onset)
- Peak velocity of V gesture (V-pvel)
- Target of V gesture (V-target)

We also included the interval between the onset of the V gesture and the c-center. The c-center is calculated as the midpoint between the C gesture onset and the pitch accent L tone onset. If the interval between the c-center and the V onset is relatively small, that would indicate a c-centering effect.

In order to determine which articulatory landmark in the list above was most closely aligned with the pitch accent L onset, we ran linear mixed effects models for real words and neologisms with duration (in ms) as the dependent variable and INTERVAL ORIGIN (V-onset, V-pvel, V-target, C-onset, C-pvel, C-target) as the fixed factor. The model selection process began with a maximal model that included random slopes and intercepts by all predictors per Speaker. The models did not converge with random slopes, so the final models only included random intercepts by Speaker. The model for real words indicated main effects of INTERVAL ORIGIN ($F(6)=17.27$, $p<.001$). Figure 60 shows the results of the model for real words. Post-hoc pairwise comparisons using emmeans with Holm correction revealed that the interval between the onset of the pitch accent L tone and peak velocity of the C gesture (10.23ms) was shorter than the interval between the L tone and C target (ß=-24.75, SE=6.74, $p<0.01$), V onset (ß=24.02, SE=6.74, $p<.01$), or V target (ß=-59.83, SE=6.74, $p<.001$). This indicates that the pitch accent L tone is more closely aligned with the peak velocity timepoint of the C gesture, compared to the other timepoints. It should also be noted that the interval between the vowel onset and the c-center, which is calculated as the midpoint between C onset and L tone onset, is also relatively small. It is not significantly longer than the interval between the L tone onset and C peak velocity, meaning it is also relatively well aligned.

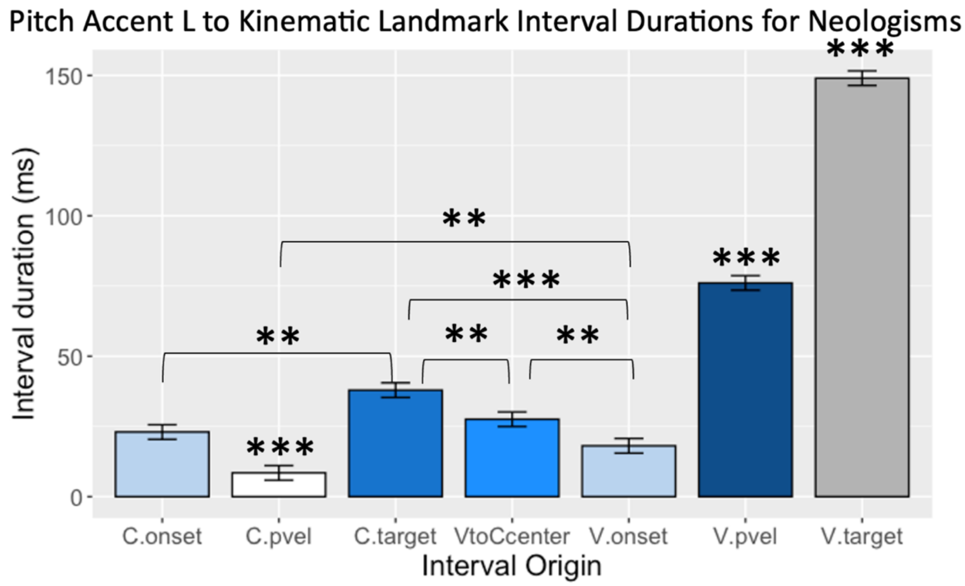
Figure 60. Alignment of the pitch accent L tone onset with various kinematic landmarks in real words with penultimate pitch accent. The interval between the L tone and peak velocity of the C gesture is smallest (except compared to C onset or V onset to c-center), indicating that the L tone is initiated closest to those timepoints.



For neologisms, the model also showed main effects of INTERVAL ORIGIN (F(6)= 534.57, *p*<.001). Figure 61 shows the results of the model for neologisms. Post-hoc pairwise comparisons using emmeans with Holm correction revealed that the interval between the onset of the pitch accent L tone and peak velocity of the C gesture (8.47ms) was shorter than all other landmarks (*p*<.01).

Figure 61. Alignment of the pitch accent L tone onset with various kinematic landmarks in neologisms with penultimate pitch accent. The interval between the L tone

and peak velocity of the C gesture is smallest, indicating that the L tone is most closely aligned with the peak velocity timepoint of the C gesture.

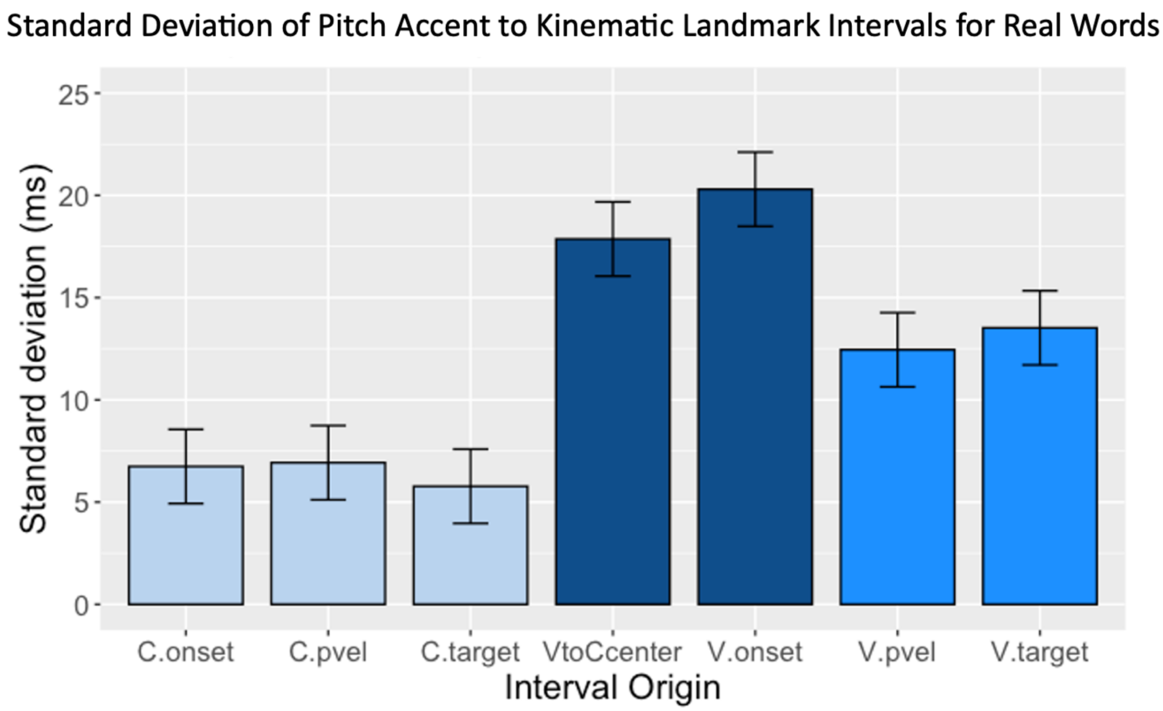


We followed this analysis with an analysis of their variance to evaluate the stability of the L tone's alignment in these intervals.

5.3.2.3. Stability of the pitch accent L tone gesture with kinematic landmarks

To analyze stability, we ran linear mixed effects models for real words and neologisms with standard deviation (in ms) as the dependent variable and INTERVAL ORIGIN (V-onset, V-pvel, V-target, C-onset, C-pvel, C-target, VtoCcenter) as fixed factors, and the same model selection process as above. Models with random slopes did not converge, and the final model included random intercepts by Speaker. The model for real words showed main effects of INTERVAL ORIGIN ($F(6)= 30.27$, $p<.001$). Figure 62 shows the standard deviations for each interval origin; a smaller standard deviation indicates a more stable coordination with the respective timepoint. Based on post-hoc pairwise comparisons using emmeans with Holm correction, L pitch accent tone to V onset and V onset to c-center intervals were least stable, followed by L pitch accent tone intervals with V peak velocity and V target. L tone intervals with C onset, C peak velocity, and C target were the most stable.

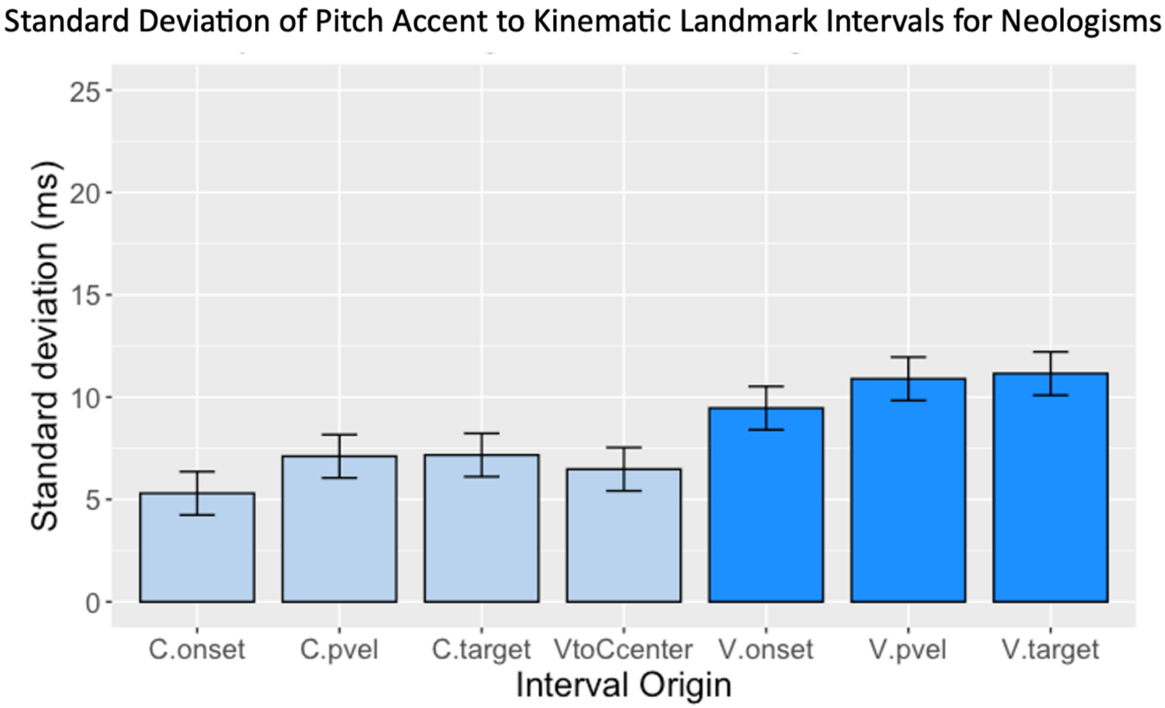
Figure 62. Stability of L tone with kinematic landmarks in real words.



The standard deviations for each interval origin for neologisms is provided in Figure 63. The model for neologisms indicated main effects of INTERVAL ORIGIN (F(6)= 16.89, *p*<.001), but post-hoc pairwise comparisons using emmeans with Holm correction revealed that interval origins using C gesture timepoints were more stable than those of V gestures. V onset to c-center interval was also stable. C gesture timepoints and V onset to c-center were not significantly different from each other, indicating that they were relatively equally stable.

Figure 63. Stability of L tone with kinematic landmarks in neologisms. Smaller standard deviation indicates more stable coordination with the respective timepoint. The

pitch accent L tone was more stable with C gesture timepoints than with V gesture timepoints.



5.3.2.4. Stability of the pitch accent L tone gesture with H boundary tone

Finally, we analyzed the stability of the relative timing between the pitch accent and the boundary tone in the final syllable of penultimate-accented words by examining the stability of the interval between the onset of the pitch accent L tone and the onset of the H boundary tone. Since we know that these tones occur sequentially, with other tones such as the L component of the LH boundary tone occurring in between, we expect anti-phase coordination. However, even if the two tones vary relative to kinematic landmarks (such as C or V onset), do they stay relatively stably timed with respect to each other? To examine this question, we looked at the coefficient of variation, which is a measure of dispersion which takes the standard deviation and the mean to show the relative dispersion of data points around the mean. The coefficient of variation is typically expressed as a percentage.

In real words, the coefficient of variation for the interval between the pitch accent L and the H boundary tone (L to H) was only 7.36%, a relatively low value compared to the CV of intervals with other landmarks which were all above 29%. In neologisms, the coefficient of variation for L to H was 9.64%, again the lowest out of all other landmarks, although V target was similarly low at 9.71%. These measures are a good indication the pitch accent L and the boundary tone H are quite stably timed relative to timing with other kinematic landmarks.

5.4. Discussion

In this study, we asked two questions: (i) how are boundary tones coordinated with constriction gestures in Japanese, and (ii) does boundary tone coordination interact with the position of lexical pitch accent? We began our investigation with a X-JToBI analysis which confirmed that, as expected, all test words had a LH% boundary tone rise in F0 in the phrase-final syllable. The analysis also confirmed that pitch accented words showed a fall in F0 from H to L in the syllable following the accented one. In final-accented words and unaccented words, where there is no pitch accent or no syllable following the accented one, there was no such fall in F0.

Gestural scores generated based on the timing of the H tone with C and V constriction gestures within the phrase-final syllable of antepenultimate- and penultimate-accented words showed that the BT H tone is initiated after the target of the V constriction gesture is reached. We then examined the alignment of the onset of the H boundary tone, which was measured based on the turning point from L to H in the final syllable. Intervals based on the relative timing of the vowel onset with various kinematic landmarks (e.g. onsets, peak velocity, and target onset/offset timepoints for C and V constriction gestures) were calculated. Linear mixed effects models for real and nonce words found that the onset of the H boundary tone occurs later when pitch accent is later in the word. In other words, regardless of the vowel's duration, relative to the onset of the vowel, the H boundary tone is initiated later in penultimate-accented words, compared to antepenultimate-accented words. So with regards to our research question, pitch accent position does indeed appear to affect boundary tone coordination.

To better understand the effect of pitch accent on the timing of the boundary tone H, we ran a linear mixed effects model on the stability of the H tone with kinematic landmarks in the phrase-final vowel in penultimate-accented words, where there is additionally a pitch accent tone gesture within the phrase-final syllable. We found that the BT H tone is most stably aligned with the target timepoint within the final V gesture.

We then similarly examined the coordination of the pitch accent L tone with kinematic landmarks. The linear mixed effects models indicated that pitch accent L was timed closely with the peak velocity timepoint of the C constriction gesture. In terms of stability, C gesture timepoints generally seemed more stable compared to those of V gestures. However, the L tone being timed with peak velocity is not predicted in the Articulatory Phonology framework, where boundary tone coordination is either in-phase (synchronous) or anti-phase (sequential), with the L tone aligned with the onset or target of the C or V gestures. Alignment with peak velocity of the C gesture suggests there may be a weak c-centering effect, where the tone gesture and the C gesture are shifted relative to the V onset.

To get a better sense of the coordination of tonal gestures with C and V constriction gestures, we also analyzed the stability of the pitch accent L tone with the boundary tone H gesture, to explore whether the tone gestures are stably timed with respect to each other. An analysis using the coefficient of variance found that the L and H tones are quite stably timed with each other. These two tones are assumed to be anti-phase with each other since they occur sequentially, with other L tones in between. Given that the timing between the L and H is relatively stable, it makes sense that the pitch accent "pushes" the H tone later in penultimate-accented words. This also fits predictions from tonal crowding account.

However, tonal crowding per se cannot account for the related temporal effects at the boundary. In the previous chapter, we found a similar effect of pitch accent position on the scope of phrase-final lengthening, which suggests some coordination between the pitch accent and the π-gesture. In Greek, boundary tone gestures were similarly found to be initiated earlier in words with non-final stress (Katsika, Krivokapić, et al. 2014). For Greek, an alternative explanation was proposed, in which the boundary tone is coordinated with (i) the μ-gesture associated with the lexical stress in the phrase-final word and (ii) the π-gesture associated with the phrase boundary (see Figure 13). Similar to Greek, the Japanese data so far indicate similar patterns of interaction between lexical pitch accent and tonal and temporal boundary-related events. However, while in Greek it is proposed that the μ-gesture affects the coordination of the π-gesture, which in turn triggers the boundary tone, there is no evidence this account can be similarly applied to Japanese. Chapter 6 therefore explores the kinematic correlates of pitch accent per se, in order to see if there is evidence of pitch accent correlation with longer duration or greater displacement in Japanese, and help us better understand the connection between word prosody and boundary marking.

6. Lexical pitch accent

We have seen in Chapters 4 and 5 that boundary marking – phrase-final lengthening and boundary tone coordination – interacts with lexical pitch accent in Japanese. In Greek, there was a similar interaction between lexical stress and boundary marking, and an account within the framework of Articulatory Phonology was proposed in which the π-gesture is coordinated with the μ-gesture associated with the stressed syllable of the phrase-final word. This coordination with the μ-gesture accounts for shifting of the π-gesture, which is evidenced by changes in the scope of phrase-final lengthening. The initiation of boundary tones is likewise shifted, being coordinated with the π-gesture. However, it remains to be seen whether an account involving the μ-gesture can be extended to and the similar findings so far regarding the effect of lexical pitch accent on the scope of boundary lengthening and boundary tone coordination, as it is also possible that the connection is due to some other relationship between word prosody and boundary events in Japanese, such as a connection in the tonal domain. This chapter investigates the durational and spatial correlates of lexical pitch accent in Japanese.

6.1. Questions and predictions

The main question this chapter seeks to answer is whether there is any kinematic evidence for durational or spatial correlates of lexical pitch accent in Tokyo Japanese. Since μ-gestures are expected to modulate the temporal or spatial unfolding of all co-active gestures, consonant and vowel gestures coordinated with a μ-gesture are expected to be affected in durational or spatial dimensions. In order to further investigate the relationship between our findings from Chapters 4 and 5, our questions for this chapter are the following:

(i) Are the consonant and vowel constriction gestures in pitch accented syllables longer than those in unaccented syllables?

(ii) Are the consonant and vowel constriction gestures in pitch accented syllables larger than those in unaccented syllables?

Duration is an unlikely correlate of accent in Japanese because vowel and consonant length is phonemic (Cutler and Otake 1999; Kaiki, Takeda and Sagisaka 1992; Mori 2001). Furthermore, while duration and intensity have been previously investigated as secondary correlates of pitch accent, studies have only found weak or inconsistent correlations. For instance, articulatory studies have found correlations between jaw displacement and stress in English, with stress syllables showing larger jaw opening (e.g., Erickson, Suemitsu, et al. 2012; Erickson, Kawahara, et al. 2014a; Erickson, Kawahara and Williams, et al. 2014b; Kelso, et al. 1985; Menezes and Erickson 2013; Vatikiotis-Bateson and Kelso 1993), while similar studies examining the effect of lexical pitch accent on jaw displacement in Japanese have shown no correlation between the two (Kawahara, Erickson, et al. 2014). Since duration and intensity are not typically considered strong correlates of lexical pitch accent in Japanese, we do not expect to find that consonant and vowel constriction gestures in pitch accented syllables are longer or larger than consonants and vowels constriction gestures in unaccented syllables.

6.2. Methods

In this chapter, we use the same kinematic data as the previous chapters to examine the relationship between pitch accent and (i) temporal length and (ii) displacement. The data were the same data described in Chapter 4 (two-syllable real words, three-syllable real words, and neologisms), subsetted for phrase-medial words only in order to test the effect of lexical pitch accent without any effects of boundary-related lengthening. As a reminder to the reader, antepenultimate and penultimate vowels in three-syllable real words were not reliable for kinematic labeling, so that data set only includes constriction gestures for consonants and the word-final vowels.

Duration was measured as the formation phase of consonant (C) and vowel (V) constriction gestures, and displacement as the absolute displacement of gestures (absolute values were used to account for the directions of high vs. low vowels). To test the effect of pitch accent on duration and displacement, we ran linear mixed effects models on C and V constriction gestures for each data set. Given that in Japanese, pitch accent is said to be realized as a sharp fall in F0 on the following mora, we decided to code our data to distinguish a three-way contrast between ACCENTED syllable in an accented word (AA), UNACCENTED syllable in an accented word (UA), and UNACCENTED syllable in an unaccented word (UU). In this way, we hope to be able to detect any effect of pitch accent on duration or displacement of consonant and vowel constriction gestures in the accented syllable as well as in the syllable following a pitch accented syllable, as that is where the main pitch correlate occurs.

Consonant and vowel constriction gestures were labeled in Matlab with the procedures described in Section 3.5.1. C and V duration were measured using formation duration of the constriction gesture. We used z-scores to normalize all durations by Phoneme to account for inherent differences by segment (Han 1961; Nishinuma 1979). Displacement was measured using absolute values and also normalized by phoneme (Erickson, Kawahara and Moore, et al. 2013; Kawahara, et al. 2014; Menezes and Erickson 2013; Williams, et al. 2013). Real words were normalized separately from neologisms to reduce noise; data sets were not able to be combined for models because of difficulties resolving different word lengths, number of factors dependent on word length, and missing measurements for three-syllable word vowels.

6.2.1. Statistical Analysis

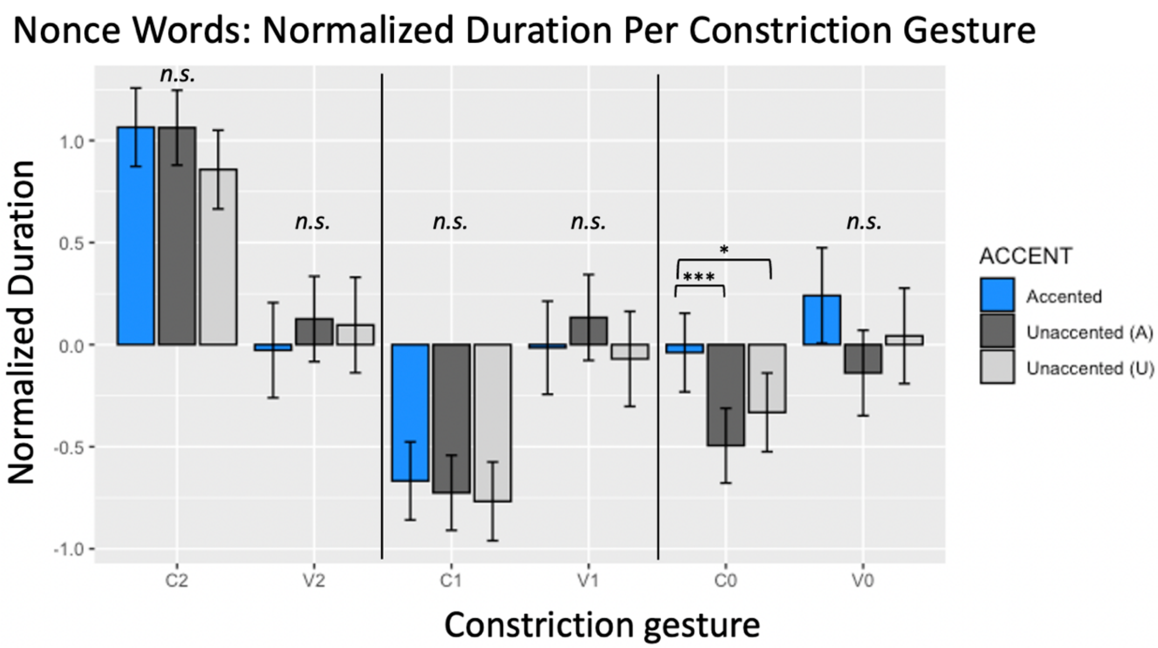
The linear mixed effects model included (i) duration or (ii) displacement as the dependent variable, ACCENT [levels: AA, UA, UU] and SYLLABLE POSITION for the model of C gestures [C2 (antepenultimate), C1 (penultimate), C0 (word-final)] as fixed factors, and random intercepts by Speaker. Since only the final vowel was measured in three-syllable real words, that model does not include a fixed factor of SYLLABLE POSITION.

6.3. Results: Duration

6.3.1. Neologisms

In neologisms, the data were a set of three-syllable words *mamima* with pitch accent being antepenultimate, penultimate, ultimate, or unaccented. In consonant constriction gestures, there was a significant interaction between ACCENT and SYLLABLE POSITION (p<.01). Figure 64 shows that the only significant pairwise comparisons from emmeans was in the final consonant (C0), where pitch accented syllables were longer than unaccented syllables in accented (i.e., antepenultimate or penultimate accent) or unaccented words (β_{UA}=0.46, SE=0.10, p<0.001; β_{UU}=0.29, SE=0.12, p<0.05). In vowel constriction gestures, there was no main effect of ACCENT or SYLLABLE POSITION.

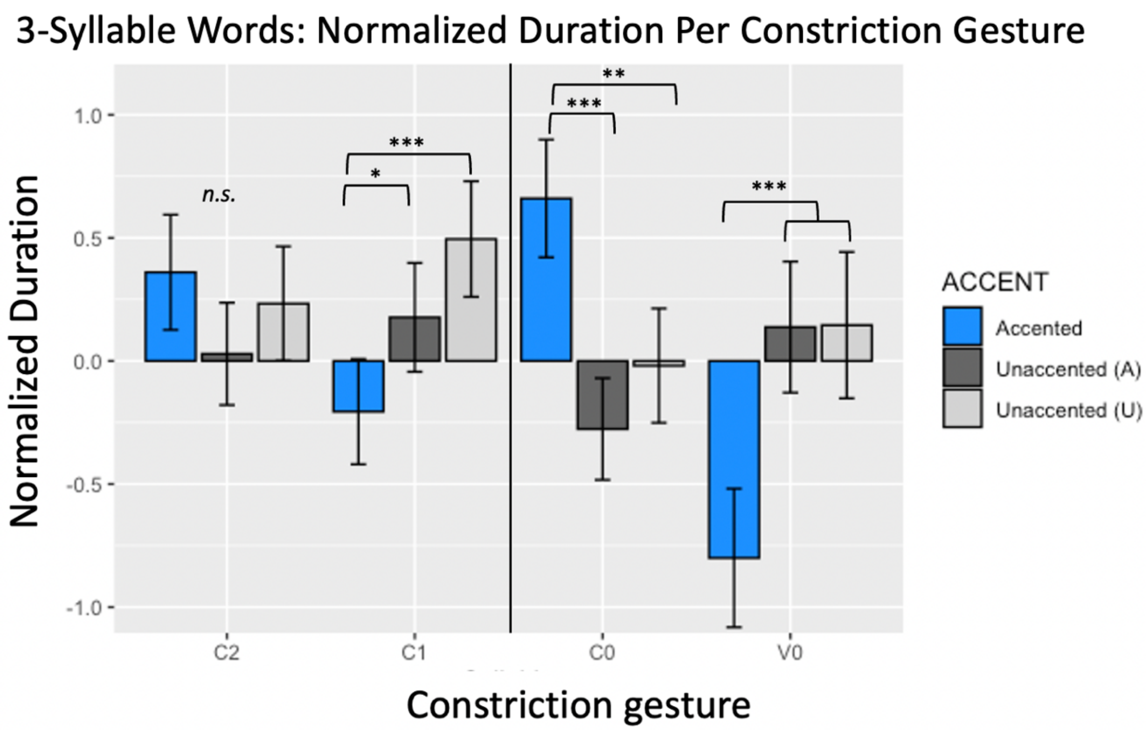
Figure 64. Model predictions for normalized duration per constriction gesture in neologisms. C2/V2 are in the antepenultimate syllable; C1/V1 in the penultimate syllable; and C0/V0 in the word-final syllable.



6.3.2. Three-syllable real words

In the three-syllable real word data, the model for consonant constriction gestures also showed an interaction between ACCENT and SYLLABLE POSITION (p<.0001). However, there was a complementary effect on duration in the final and penultimate syllables, which can be seen in Figure 65. In the penultimate consonant (C1), pitch accented syllables were shorter than unaccented syllables in accented or unaccented words (β_{UA}=-0.38, SE=0.15, p<0.05; β_{UU}=-0.70, SE=0.17, p<0.001). In the final consonant (C0), pitch accented syllables were longer (β_{UA}=0.94, SE=0.16, p<0.0001; β_{UU}=-0.68, SE=0.20, p<0.01). In the model for vowel constriction gestures, there was no factor of SYLLABLE POSITION because the data only included final vowels. There was a main effect of ACCENT (p<.0001) in the opposite direction as consonants, with V0 gestures being shorter in the final syllable when accented (β_{UA}=-0.94, SE=0.17, *p*<0.0001; β_{UU}=-0.95, SE=0.21, *p*<0.0001) (compare C0 and V0 patterns in Figure 65).

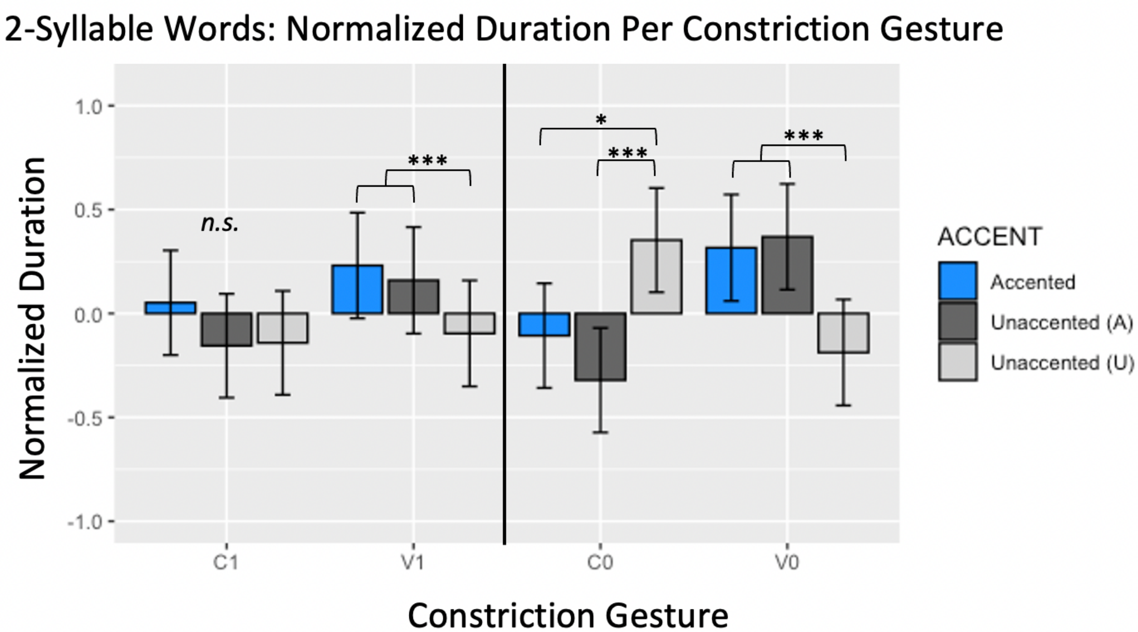
Figure 65. Model predictions for normalized duration per constriction gesture in three-syllable real words. C2 is the antepenultimate C gesture, C1 the penultimate, and C0/V0 are the word-final C and V gestures.



6.3.3. Two-syllable real words

The two-syllable word data were a minimal set consisting of the words *nami* versus *nami** versus *na*mi*. Again, there was a significant interaction between ACCENT and SYLLABLE POSITION (*p*<.01). However, in this case the final consonant constriction gestures (C0) of unaccented syllables in an unaccented word (UU) were longer than unaccented syllables in an accented word (UA) (β_{UA}=-0.67, SE=0.18, *p*<0.001) and longer than pitch accented syllables (AA) (β_{AA}=-0.46, SE=0.18, p<0.05), as can be seen in Figure 66. In vowel constriction gestures, there was no significant interaction between ACCENT and SYLLABLE POSITION. There was a main effect of ACCENT (p<.001), with duration of vowel gestures in unaccented words (UU) overall being shorter than accented words (β_{AA}=0.41, SE=0.11, p<0.001; β_{UA}=0.41, SE=0.11, p<0.001). There was no main effect of SYLLABLE POSITION.

Figure 66. Model predictions for normalized duration per constriction gesture in two-syllable real words.



6.3.4. Summary

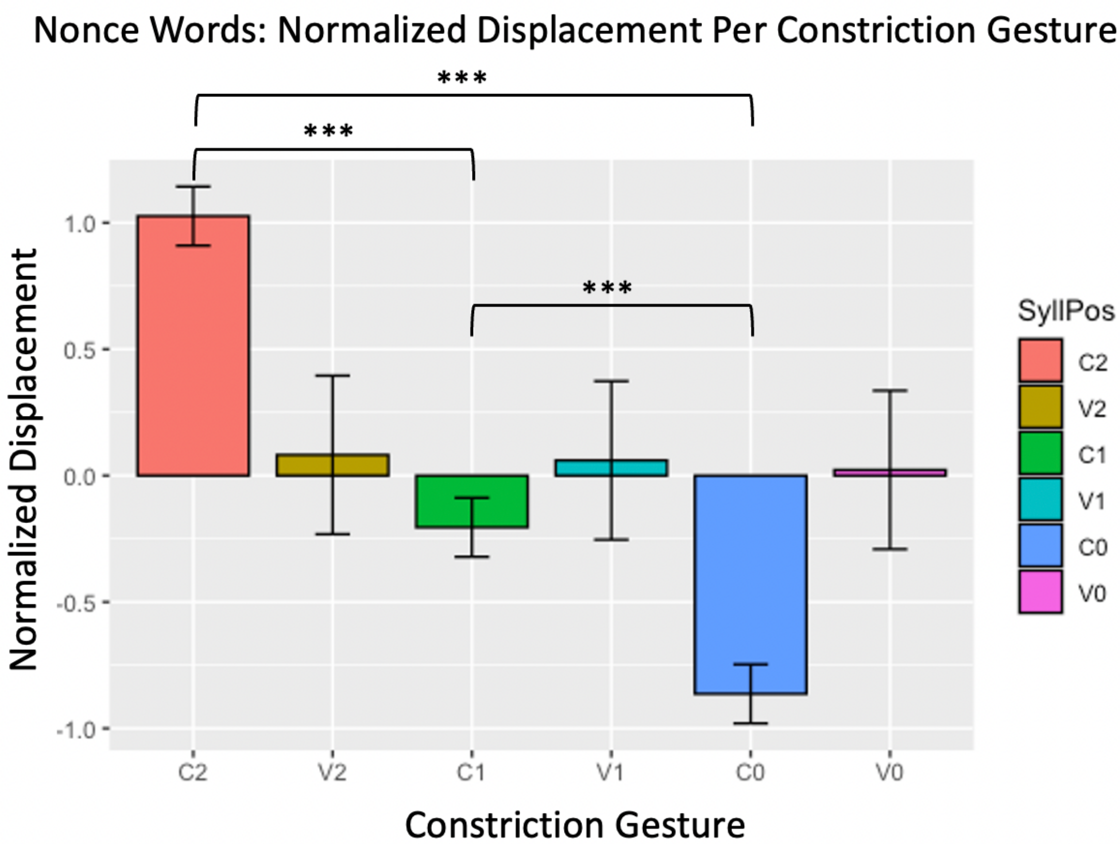
All in all, there were no consistent correlations found between pitch accent and duration. While the models for some data sets showed significant interactions between ACCENT and SYLLABLE POSITION, this interaction was only on the final consonant (C0) in two-syllable real words and neologisms, and on the penultimate consonant (C1) in three-syllable real words. Furthermore, the direction of the effect of pitch accent was not consistent. For example, although there was an effect of pitch accent on duration in vowels, in real words the constriction gestures were longer when accented in two-syllable words, whereas in three-syllable words the constriction gestures were shorter when accented. Neologisms showed no significant interaction at all in the models for vowel gestures. There was also inconsistent patterning between unaccented syllables in pitch-accented words vs. unaccented syllables in unaccented words. It is possible that rhythmic reasons may also play a role, as three-syllable real words and three-syllable neologisms both showed longer C0 formation duration when the final syllable was accented, while two-syllable real words did not show the same pattern.

6.4. Results: Displacement

6.4.1. Results: Neologisms

There was no significant interaction between ACCENT and SYLLABLE POSITION in consonant or vowel models, so a visualization of the model predictions for the interaction could not be generated. In consonant constriction gestures, the only effect on displacement was a main effect of SYLLABLE POSITION (p<.0001), with a three-way distinction in the size of gestures in antepenultimate, penultimate, and ultimate syllables (see Figure 67). Gestures in the antepenultimate syllable (C2) were the largest (β_{C1}=1.23, SE=0.06, p<.0001; β_{C0}=1.89, SE=0.06, p<.0001), and consonants in the final syllable (C0) were the smallest (β_{C1}=0.66, SE=0.06, p<.0001). This difference is illustrated in the figure below and may be explained by the fact that the antepenultimate syllable is AP-initial, an inherently strong position in this head/edge prominence language (Jun 2014). There were no significant interactions or main effects in the models for vowel gestures.

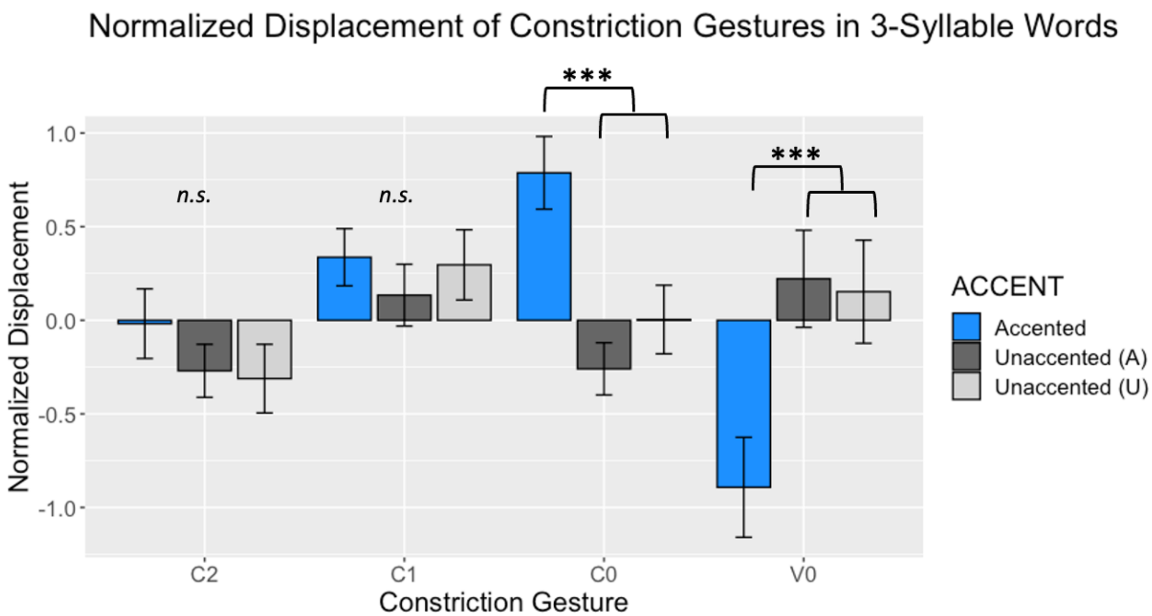
Figure 67. Normalized displacement per constriction gestures in neologisms.



6.4.2. Results: three-syllable real words

In three-syllable real words, there was a significant interaction between ACCENT and SYLLABLE POSITION for consonant constriction gestures (p<.01). Figure 68 shows the results of pairwise comparisons using emmeans where consonant constriction gestures of accented syllables in the final syllable (C0) were larger than those of unaccented syllables (β_{UA}=1.05, SE=0.18, p<.0001; β_{UU}=0.78, SE=0.22, p<.001). There was no significant difference between the two unaccented conditions. Since only the final vowel was measured in three-syllable words, the model for vowels did not include a factor of SYLLABLE POSITION. There was a main effect of ACCENT (p<.0001). Opposite to the results for consonant gestures, vowel gestures in the final syllable were smaller when accented (β_{UA}=-1.11, SE=0.12, p<.0001; β_{UU}=-1.04, SE=0.15, p<.001). The results for C0 and V0 gestures in the final syllable match with the patterns found for temporal effects.

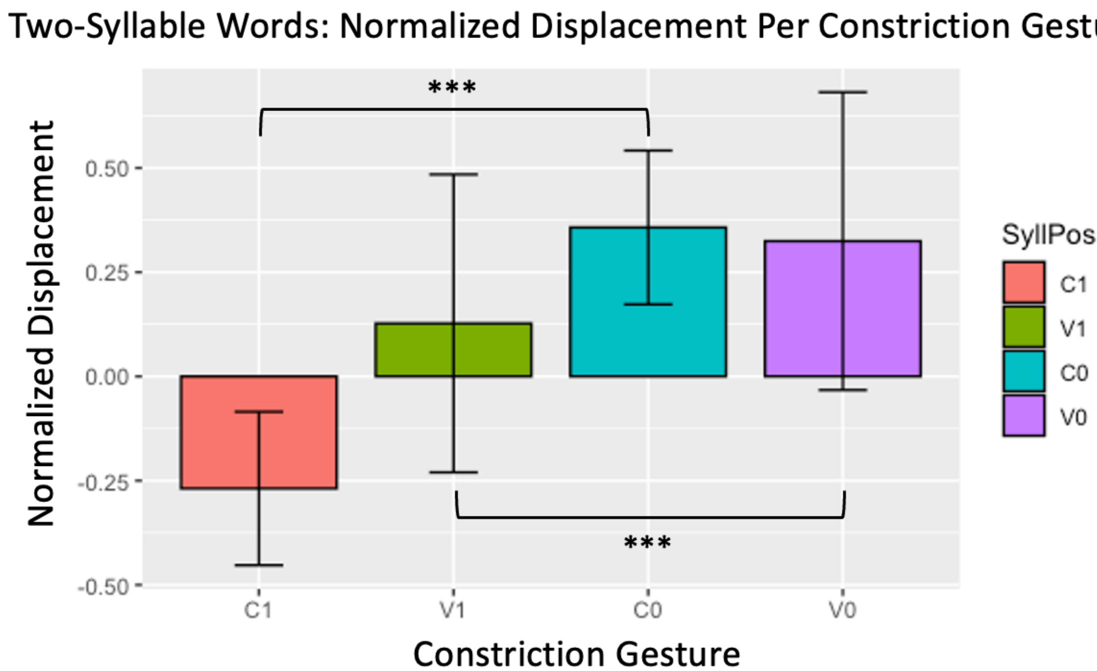
Figure 68. Normalized displacement per constriction gesture in three-syllable real words.



6.4.3. Results: two-syllable real words

In two-syllable real words, there was no significant interaction between ACCENT and SYLLABLE POSITION for consonant constriction gestures, so no figure to illustrate those comparisons was produced. Figure 69 shows the main effect of SYLLABLE POSITION (p<.0001), where word-final (C0) consonant gestures were larger than penultimate (C1) gestures (β_{C0}=-0.63, SE=0.08, p<0.0001). This was also the case for the vowel gestures, where there was no significant interaction, but a main effect of SYLLABLE POSITION (p<.05) where the word-final vowel (V0) gesture was larger than the penultimate (V1) gesture (β_{V0}=-0.12, SE=0.08, p<0.05). There was no main effect of ACCENT for consonants or vowels.

Figure 69. Normalized displacement per constriction gesture in two-syllable real words.



6.4.4. Summary

Except for consonants in three-syllable real words, none of the models for consonant or vowel constriction gestures showed a significant interaction between ACCENT and SYLLABLE POSITION. The model for three-syllable real words had an effect of ACCENT in the final C and V gestures, where the C0 gesture was larger when accented and V0 was smaller when accented. Neologisms and two-syllable real words had main effects of SYLLABLE POSITION with inconsistent directions. In neologisms, the antepenultimate C gesture was larger than the final C gesture, but in two-syllable real words, the penultimate C and V gestures were smaller than the final C and V gestures. In summary, there were no consistent correlations between pitch accent and displacement in any of the data. In other words, pitch accented syllables were not systematically larger than unaccented syllables.

6.5. Discussion

The kinematic data in this study confirms that there is no strong or consistent correlation between pitch accent and consonant or vowel duration in Tokyo Japanese, as expected based on previous literature. Across the three data sets, accented syllables were not consistently longer or larger than unaccented syllables. In languages like English, constriction gestures in a stressed syllable are correlated with longer durations or larger displacement, which may be modeled in Articulatory Phonology as a result of clock-slowing effects of a μ-gesture with concurrently active constriction gestures within the stressed syllable. However, we have found a lack of such a correlation between pitch accent and duration or displacement in consonant and vowel constriction gestures in Japanese. As μ-gesture would be expected to exert global slowing effects on all overlapping gestures, this suggests that μ-gesture account does not hold for pitch accented syllables in Tokyo Japanese.

One aspect of this study which merits further research is the role of rhythm and foot structure. Results across models were inconsistent, which may possibly be related to factors such as word length and stress. Word length is relevant because many phonological and morphological processes in Japanese can be generalized with the existence of bimoraic feet as a prosodic unit above the mora (e.g., compound clipping (Kubozono 1999), noun-noun compound accent rule (Haruo 1997), hypocoristic formation (Poser 1990)). Further studies to directly investigate the effect of moraic structure on kinematic and tonal effects in Japanese are warranted. The main finding from the current study is that the results from our analysis show that lexical pitch accent in Japanese does not correlate with longer duration or larger displacement of consonant and vowel constriction gestures.

The main conclusion from this analysis is that there is no robust correlation between lexical pitch accent and consonant and vowel duration or displacement. Within the Articulatory Phonology framework, we take this to mean that there is no evidence for the account that pitch accented syllables are associated with a temporal or spatial μ-gesture affecting consonant or vowel constrictions.

7. General Discussion/Conclusions

In this dissertation, we have taken a close look at how word-level prosody in Japanese interacts with phrase-level prosodic marking in the temporal and tonal domains. Using carefully designed experiments, we collected kinematic data to examine the effect of lexical pitch accent position on the scope of phrase-final lengthening and the coordination of phrasal boundary tones.

Recent work has found that in head-prominence languages, the position of lexical stress can affect the stretch of speech affected by phrase-final lengthening. In an articulatory study on Greek, Katsika (2016) found that lengthening began earlier in words with non-final stress. Within the framework of Articulatory Phonology, prosodic boundaries are instantiated by a prosodic gesture known as a π-gesture, which has slowing effects on any co-active gestures; the earlier the overlap with the π-gesture, the earlier the slowing effects begin. Furthermore, another study found that in Greek, the position of lexical stress in the phrase-final word also affects the initiation of the boundary tone (Katsika, Krivokapić, et al., 2014). Boundary tones were found to be initiated earlier when lexical stress was earlier in the word. An account for the similarity of these findings to those of the interaction of lexical stress with the scope of phrase-final lengthening was proposed in which the μ-gesture associated with lexical stress is coordinated with the π-gesture at the prosodic phrase boundary. The π-gesture triggers the boundary tone, so that lexical stress being later in the phrase-final word affects π-gesture coordination, which in turn affects boundary tone timing.

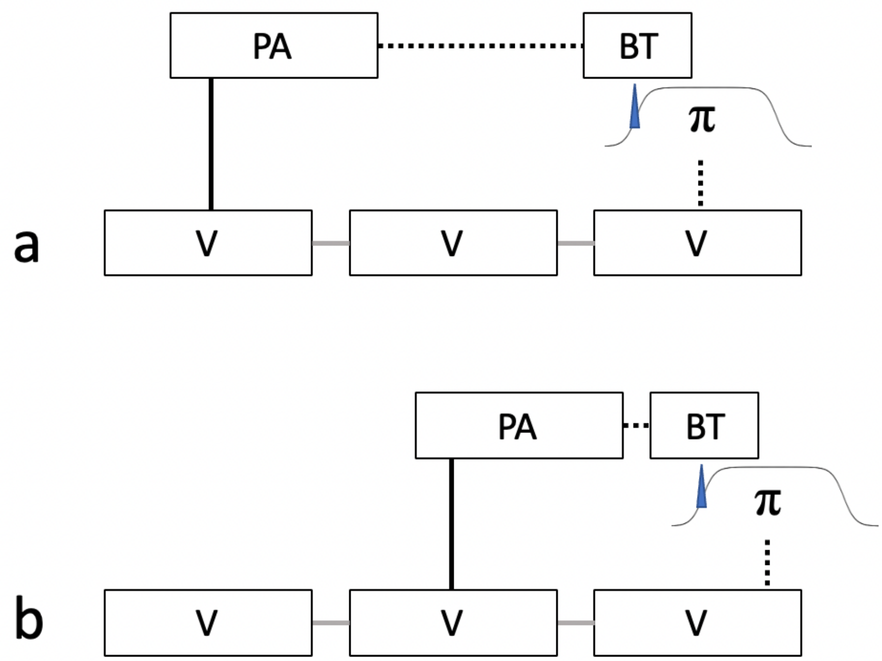
Turning to Japanese, we have similarly found effects of word prosody on the scope of phrase-final lengthening and the timing of boundary tone initiation. In particular, Chapter 4 found that there is more lengthening in words with pitch accent earlier in the word compared to words with pitch accent later in the word. Within the Articulatory Phonology framework, a greater amount of lengthening can be interpreted as greater overlap with the maximum activation of the π-gesture. Thus, pitch accent position in Japanese also affects the scope of slowing effects imposed by the π-gesture. In the tonal domain, we found in Chapter 5 that the final H of the complex LH% question boundary tone was also initiated later in penultimate-accented words compared to antepenultimate-accented words. Furthermore, an alignment analysis of the timing of tones in the phrase-final syllable of penultimate-accented words showed that the L tone of the lexical pitch accent was initiated near the peak velocity of the C constriction gesture, and there were indications of a weak c-centering effect where the timing of the C gesture and the pitch accent L tone may be shifted in relation to the V gesture onset. A c-centering effect is not unexpected, especially as lexical tones in Mandarin have also been proposed to influence C and V gesture coordination and have c-center effects (Gao 2008). In addition, an analysis of the interval between the pitch accent L and the boundary H tones found that the two tones were relatively stably timed with respect to each other. These findings together must also be interpreted in light of the results from Chapter 6, where no strong or consistent correlation between lexical pitch accent and longer duration or larger displacement of consonant and vowel gestures was found. We interpret these findings as no evidence supporting a spatial or temporal μ-gesture account for constriction gestures.

Altogether, a new account for the relationship between word-level prosody and temporal and tonal events at the phrase-level is proposed. Whereas the account for Greek involves the μ-gesture associated with the lexically stressed syllable, we have not found any correlation between duration or displacement with constriction gestures in pitch accented syllables.

However, Japanese word prosody is realized in the tonal domain, where an interaction with boundary tones is expected. Indeed, tonal crowding predicts the effects we see, where in the case of penultimate-accented words, more tones in the phrase-final syllable results in later initiation of the final tone gesture. Our analysis indicated that the pitch accent L tone and the H boundary tone may be timed most stably with respect to each other, in anti-phase coordination. Since the boundary tone is connected to the π-gesture at the phrase edge, it is possible that the coordination between the pitch accent L tone and the boundary H tone affects the π-gesture's coordination, which in turn is realized as an effect on the scope of phrase-final lengthening in the temporal domain. A schematic representation of the proposed account is provided in Figure 70.

Figure 70. Schematic representation of the coordination of phrase-final π-gestures, lexical and boundary tones, and V constriction gestures in three-syllable words with (a) antepenultimate and (b) penultimate pitch accent. Dotted black lines represent anti-phase coordination, while solid black lines represent in-phase coordination. The blue

triangle corresponds to the point where the boundary tone is triggered by the π-gesture's strength of activation. This model is based on the one proposed in Katsika (2016).



The main implications of this proposed account is that a more complete view of intonation might need to consider both tonal and temporal events. In Japanese, we have seen that similarly to Greek, word-level prosody and phrase-level prosody are interconnected, and the position of word accent has an effect on the timing and coordination of events at the phrase edge. By using articulatory data to closely examine the timing of temporal and tonal events with respect to consonant and vowel constriction gestures, we have found that the π-gesture's coordination can be affected by the coordination of lexical and phrasal tone gestures in the tonal domain. The interaction of boundary marking with word prosody suggests that the most complete account of intonation involves both temporal and tonal domains.

References

Arvaniti, Amalia, and D. Robert Ladd. 2009. "Greek wh-questions and the phonology of intonation." *Phonology* 26: 43–74.

Arvaniti, Amalia, D. Robert Ladd, and Ineke Mennen. 2006. "Phonetic effects of focus and “tonal crowding” in intonation: Evidence from Greek polar questions." *Speech Communication* 48 (6): 667-696.

Arvaniti, Amalia, D. Robert Ladd, and Ineke Mennen. 1998. "Stability of tonal alignment: the case of Greek prenuclear accents." *Journal of Phonetics* (Journal) 26 (1): 3-25.

Barnes, Jonathan, Stefanie Shattuck-Hufnagel, Alejna Brugos, and Nanette Veilleux. 2006. "The domain of realization of the L-phrase tone in American English." *Speech Prosody.* Dresden.

Bates, D, M Mächler, B Bolker, and S Walker. 2014. "Fitting Linear Mixed-Effects Models using lme4." *arXiv* 1406.5823.

Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. 2014. "Fitting Linear Mixed-Effects Models using lme4." *arXiv* 1406.5823.

Beckman, and Mary. 1982. "Segment duration and the 'mora' in Japanese." *Phonetica* 39 (2-3): 113-135.

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

Beckman, Mary E. 1986. *Stress and Non-Stress Accent.* Dordrecht: Foris Publications.

Beckman, Mary E. 1996. "The parsing of prosody." *Language and cognitive processes* 11 (1-2): 17-68.

Beckman, Mary E., and Jan Edwards. 1992. "Intonational categories and the articulatory control of duration." *Speech perception, production, and linguistic structure* 359-375.

Beckman, Mary E., and Janet B. Pierrehumbert. 1986. "Intonational structure in Japanese and English." *Phonology* 3: 255-309.

Beckman, Mary E., Jan Edwards, and Janet Fletcher. 1992. "Prosodic structure and tempo in a sonority model of articulatory dynamics." *Papers in Laboratory Phonology II: Gesture, Segment and Prosody* 68-86.

Berkovits, Rochele. 1994. "Durational effects in final lengthening, gapping, and contrastive stress." *Language and speech* 37 (3): 237-250.

Berkovits, Rochele. 1993a. "Progressive utterance-final lengthening in syllables with final fricatives." *Language and Speech* 36 (1): 89-98.

Berkovits, Rochele. 1993b. "Utterance-final lengthening and the duration of final-stop closures." *Journal of Phonetics* 21 (4): 479-489.

Boersma, Paul, and David Weenink. 2022. *Praat: doing phonetics by computer [Computer program]. Version 6.3.03.* http://www.praat.org/.

Browman, Catherine P., and Louis Goldstein. 1989. "Articulatory gestures as phonological units." *Phonology* 6 (2): 201-251.

Browman, Catherine P., and Louis Goldstein. 1992. "Articulatory phonology: An overview." *Phonetica* 49 (3-4): 155-180.

Browman, Catherine P., and Louis Goldstein. 2000. "Competing constraints on intergestural coordination and self-organization of phonological structures." *Les Cahiers de l'ICP. Bulletin de la communication parlée* 5: 25-34.

Browman, Catherine P., and Louis Goldstein. 1995. "Dynamics and articulatory phonology." *Mind as motion: Explorations in the dynamics of cognition* 175-194.

Browman, Catherine P., and Louis Goldstein. 1990. "Tiers in articulatory phonology, with some implications for casual speech." *Papers in laboratory phonology I: Between the grammar and physics of speech* 341-376.

Browman, Catherine P., and Louis Goldstein. 1986. "Towards an articulatory phonology." *Phonology* 3: 219-252.

Bruce, Gösta. 1977. *Swedish word accents in sentence perspective.* Lund University.

Byrd, Dani. 2000. "Articulatory vowel lengthening and coordination at phrasal junctures." *Phonetica* 57 (1): 3-16.

Byrd, Dani, Abigail Kaun, Shrikanth Narayanan, and Elliot Saltzman. 2000. "Phrasal signatures in articulation." *Papers in laboratory phonology V* 70-87.

Byrd, Dani, and Daylen Riggs. 2008. "Locality interactions with prominence in determining the scope of phrasal lengthening." *Journal of the International Phonetic Association* 38 (2): 187-202.

Byrd, Dani, and Elliot Saltzman. 1998. "Intragestural dynamics of multiple prosodic boundaries." *Journal of Phonetics* 26 (2): 173-199.

Byrd, Dani, and Elliot Saltzman. 2003. "The elastic phrase: Modeling the dynamics of boundary-adjacent lengthening." *Journal of Phonetics* (31): 149–180.

Byrd, Dani, and Susie Choi. 2010. "At the juncture of prosody, phonology, and phonetics -- The interaction of phrasal and syllable structure in shaping the timing of consonant gestures." *Laboratory phonology* 10: 31-59.

Byrd, Dani, Jelena Krivokapić, and Sungbok Lee. 2006. "How far, how long: On the temporal scope of prosodic boundary effects." *The Journal of the Acoustical Society of America* 120 (3): 1589-1599.

Byrd, Dani, Sungbok Lee, Daylen Riggs, and Jason Adams. 2005. "Interacting effects of syllable and phrase position on consonant articulation." *The Journal of the Acoustical Society of America* 118 (6): 3860-3873.

Cambier-Langeveld, Tina. 1997. "The domain of final lengthening in the production of Dutch." *Linguistics in the Netherlands* 14 (1): 13-24.

Cho, Taehong. 2006. "Manifestation of prosodic structure in articulatory variation: Evidence from lip kinematics in English." *Laboratory Phonology* 519-548.

Cho, Taehong. 2008. "Prosodic strengthening in transboundary V-to-V lingual movement in American English." *Phonetica* 65 (1-2): 45-61.

Cho, Taehong. 2004. "Prosodically conditioned strengthening and vowel-to-vowel coarticulation in English." *Journal of Phonetics* 32 (2): 141-176.

Cho, Taehong, and Patricia A. Keating. 2001. "Articulatory and acoustic studies on domain-initial strengthening in Korean." *Journal of phonetics* 29 (2): 155-190.

Cutler, Anne, and Takashi Otake. 1999. "Pitch accent in spoken-word recognition in Japanese." *The Journal of the Acoustical Society of America* 105 (3): 1877-1888.

d'Imperio, M, R Espesser, H Loevenbruck, C Menezes, N Nguyen, and P Welby. 2007. "Are tones aligned with articulatory events? Evidence from Italian and French." *Laboratory Phonology 9* 577-608.

d'Imperio, Mariapaola, Robert Espesser, Hélène Loevenbruck, Caroline Menezes, NNoël Nguyen, and Pauline Welby. 2007. "Are tones aligned with articulatory events? Evidence from Italian and French." *Laboratory Phonology 9* 577-608.

Downing, Laura J., and Annie Rialland. 2016. *Intonation in African tone languages.* Walter de Gruyter GmbH & Co KG.

Edwards, Jan, Mary E. Beckman, and Janet Fletcher. 1991. "The articulatory kinematics of final lengthening." *The Journal of the Acoustical Society of America* 89 (1): 369-382.

Erickson, Donna, Atsuo Suemitsu, Yoshiho Shibuya, and Mark Tiede. 2012. "Metrical structure and production of English rhythm." *Phonetica* 69: 180-190.

Erickson, Donna, Shigeto Kawahara, J. C. Williams, Jeff Moore, Atsuo Suemitsu, and Yoshiho Shibuya. 2014b. "Metrical structure and jaw displacement: An exploration." *Speech Prosody* 300-303.

Erickson, Donna, Shigeto Kawahara, Jeff Moore, Atsuo Suemitsu, and Yoshiho Shibuya. 2013. "Effect of vowel height position on jaw opening patterns in Japanese." *Talk presented at International Conference on Phonetics and Phonology (ICPP) 3.* Tokyo: NINJAL.

Erickson, Donna, Shigeto Kawahara, Yoshiho Shibuya, Atsuo Suemitsu, and Mark Tiede. 2014a. "Comparison of Jaw Displacement Patterns of Japanese and American Speakers of English: A Preliminary Report." *Journal of the Phonetic Society of Japan* 18 (2): 88-94.

Fletcher, Janet. 2010. "The prosody of speech: Timing and rhythm." *The handbook of phonetic sciences* 521-602.

Fougeron, Cécile. 2001. "Articulatory properties of initial segments in several prosodic constituents in French." *Journal of phonetics* 29 (2): 109-135.

Fougeron, Cécile, and Patricia A. Keating. 1997. "Articulatory strengthening at edges of prosodic domains." *The Journal of the Acoustical Society of America* 101 (6): 3728-3740.

Fujimura, Osamu. 2003. "Stress and tone revisited: Skeletal vs. melodic and lexical vs. phrasal." *Cross-linguistic studies of tonal phenomena: Historical development, phonetics of tone, and descriptive studies* (Tokyo: Tokyo University of Foreign Studies) 221-236.

Gao, M. 2008. "Tonal alignment in Mandarin Chinese: An articulatory phonology account." Yale University, CT: Unpublished doctoral dissertation.

Gao, Man. 2008. "Tonal alignment in Mandarin Chinese: An articulatory phonology account." Yale University, CT: Unpublished doctoral dissertation.

Goldsmith, John. 1976. *Autosegmental phonology.* Doctoral Dissertation, MIT Press.

Grabe, Esther. 1998. "Pitch accent realization in English and German." *Journal of Phonetics* 26 (2): 129-143.

Grice, Martine, D. Robert Ladd, and Amalia Arvaniti. 2000. "On the place of phrase accents in intonational phonology." *Phonology* 17: 143–185.

Gussenhoven, Carlos. 2004. *The phonology of tone and intonation.*

Han, Mieko S. 1962. "The feature of duration in Japanese." *Study Sounds* 10: 65-80.

Han, Mieko Shimizu. 1961. *Japanese Phonology: An Analysis Based on Sound Spctrograms.* Ph.D. dissertation, University of Texas, Austin.

Haruo, Kubozono. 1997. "Lexical markedness and variation: A nonderivational account of Japanese compound accent." *Proceedings of West Coast Conference on Formal Linguistics* (CSLI Publications) 15: 273-287.

Higurashi, Yoshiko. 1983. *The accent of extended word structures in Tokyo standard Japanese.*

Higurashi, Yoshiko. 1983. "The accent of extended word structures in Tokyo Standard Japanese."

Hirose, Hajime. 2010. "Investigating the physiology of laryngeal structures." *The Handbook of Phonetic Sciences* 130-152.

Homma, Yayoi. 1981. "Durational relationship between Japanese stops and vowels." *Journal of Phonetics* 9 (3): 273-281.

Hsieh, F Y. 2011. "A Gestural Account of Mandarin Tone 3 Variation." *ICPhS* 890-893.

Hsieh, Fang-Ying. 2011. "A Gestural Account of Mandarin Tone 3 Variation." *ICPhS* 890-893.

Hyman, Larry M. 2009. "How (not) to do phonological typology: the case of pitch-accent." *Language sciences* 31 (203): 213-238.

Jun, Sun-Ah. 2014. *Prosodic typology: By prominence type, word prosody, and macro-rhythm.* Oxford Universitiy Press.

Kaiki, Nobuyoshi, Kazuya Takeda, and Yoshinori Sagisaka. 1992. "Vowel Duration Control Using Linguistic Information." *電子情報通信学会論文誌 A* 75 (3): 467-473.

Katsika, A, J Krivokapić, C Mooshammer, M Tiede, and L Goldstein. 2014. "The coordination of boundary tones and its interaction with prominence." *Journal of Phonetics* 44: 62-82.

Katsika, Argyro. 2016. "The role of prominence in determining the scope of boundary-related lengthening in Greek." *Journal of phonetics* 55: 149-181.

Katsika, Argyro, and Karen Tsai. 2021. "The supralaryngeal articulation of stress and accent in Greek." *Journal of Phonetics* 88: 101085.

Katsika, Argyro, Jelena Krivokapić, Cristine Mooshammer, Mark Tiede, and Louis Goldstein. 2014. "The coordination of boundary tones and its interaction with prominence." *Journal of Phonetics* 44: 62-82.

Kawahara, Shigeto, Donna Erickson, Jeff Moore, Yoshiho Shibuya, and Atsuo Suemitsu. 2014. "Jaw Displacement and Metrical Structure in Japanese: The Effect of Pitch Accent, Foot Structure, and Phrasal Stress (<Feature Articles> Articulatory Phonetics: Focus on Japanese)." *Journal of the Phonetic Society of Japan* 18 (2): 77-87.

Kawahara, Shigeto, Donna Erickson, Jeff Moore, Yoshiho Shibuya, and Atsuo Suemitsu. 2014. "Jaw displacement and metrical structure in Japanese: The effect of pitch accent, foot structure, and phrasal stress." *Journal of the Phonetic Society of Japan* 18 (2): 77-87.

Keating, Patricia, Taehong Cho, Cécile Fougeron, and Chai-Shune Hsu. 2004. "Domain-initial articulatory strengthening in four languages." *Phonetic interpretation: Papers in laboratory phonology VI* 143-161.

Kelso, JA Scott, Eric Vatikiotis-Bateson, Elliot L. Saltzman, and Bruce Kay. 1985. "A qualitative dynamic analysis of reiterant speech production: Phase portraits, kinematics, and dynamic modeling." *The Journal of the Acoustical Society of America* 77 (1): 266-280.

Kim, Sahyang, Jiyoung Jang, and Taehong Cho. 2017. "Articulatory characteristics of preboundary lengthening in interaction with prominence on tri-syllabic words in American English." *The Journal of the Acoustical Society of America* 142 (4): EL362-EL368.

Kindaichi, Haruhiko. 1947. "Tookyoogo ni okeru "hana" to "hana" no kubetsu -- tookyoo akusento shinnidankan kyoochooron [The distinction between "hana 'flower'" and "hana 'nose'" in the Tokyo dialect -- the emphasis of the new two-level theory of the Tokyo accent]." *Kokugo.*

Kori, Shiro. 1997. "Nihongo no intoneeshon: Kata to kinoo [Intonation in Japanese: Patterns and functions]." *Akusento, intoneeshon, rizumu to poozu [Accent, intonation, rhythm and pause]* 169-202.

Krivokapić, Jelena. 2007. "The planning, production and perception of prosodic structure." Doctoral dissertation, University of Southern California.

Kubozono, Haruo. 2011. "Japanese Pitch Accent." *The Blackwell companion to phonology* 1-29.

Kubozono, Haruo. 1999. "Mora and syllable." In *The handbook of Japanese linguistics*, by Natsuko Tsujimura, 31-61.

Kubozono, Haruo. 2012. "Varieties of pitch accent systems in Japanese." *Lingua* 122 (13): 1395-1414.

Kuznetsova, Alexandra, Per B. Brockhoff, and Rune HB Christensen. 2017. "lmerTest package: tests in linear mixed effects models." *Journal of statistical software* 82: 1-26.

Ladd, D. Robert. 2008. *Intonational phonology.* Cambridge University Press.

Lenth, Russell. 2020. *Emmeans: Estimated marginal means, aka least-squares means.* https://CRAN.R-project.org/package=emmeans.

Lickley, Robin J., Astrid Schepman, and D. Robert Ladd. 2005. "Alignment of "phrase accent" lows in Dutch falling rising questions: Theoretical and methodological implications." *Language and Speech* 48 (2): 157-183.

Lickley, Robin J., Astrid Schepman, and D. Robert Ladd. 2005. "Alignment of “phrase accent” lows in Dutch falling rising questions: Theoretical and methodological implications." *Language and Speech* 48: 157–183.

Lickley, Robin J., Astrid Schepman, and D. Robert Ladd. 2005. "Alignment of “phrase accent” lows in Dutch falling rising questions: Theoretical and methodological implications." *Language and Speech* 48 (2): 157-183.

Mücke, D, H Nam, A Hermes, and L Goldstein. 2012. "Coupling of tone and constriction gestures in pitch accents." *Consonant clusters and structural complexity* 205-230.

Mücke, Doris, Hosung Nam, Anne Hermes, and Louis Goldstein. 2012. "Coupling of tone and constriction gestures in pitch accents." *Consonant clusters and structural complexity* 205-230.

Maekawa, Kikuo. 2003. "Corpus of Spontaneous Japanese: Its design and evaluation." *ISCA & IEEE Workshop on Spontaneous Speech Processing and Recognition.*

Maekawa, Kikuo, Hanae Koiso, Sadaoki Furui, and Hitoshi Isahara. 2000. "Spontaneous Speech Corpus of Japanese." *LREC* 6: 1-5.

Maekawa, Kikuo, Hideaki Kikuchi, and Wataru Tsukahara. 2004. "Corpus of Spontaneous Japanese: design, annotation and XML representation." *Reproduction* 16 (16): 5.

Maekawa, Kikuo, Hideaki Kikuchi, Yosuke Igarashi, and Jennifer J. Venditti. 2002. "X-JToBI: an extended j-toBI for spontaneous speech." *INTERSPEECH.*

Matsui, Mayuki, and Hyun Kyung Hwang. 2018. "Final vs. no accent in interrogative melodies of Tokyo Japanese: Implications for tonal clash resolution."

McCawley, James D. 1977. "Accent in Japanese." *Studies in stress and accent* 4: 261-302.

McCawley, James. 1968. *The Phonological Component of a Grammar of Japanese.* The Hague: Mouton.

Menezes, Caroline, and Donna Erickson. 2013. "Intrinsic variations in jaw deviations in English vowels." *Proceedings of Meetings on Acoustics.* Acoustical Society of America. 060253.

Mori, Yoko. 2001. "Effect of accentual fall on final-lengthening in Japanese." *Journal of Phonetic Society of Japan* 5-1.

Nakatani, Lloyd H., Kathleen D. O'Connor, and Carletta H. Aston. 1981. "Prosodic aspects of American English speech rhythm." *Phonetica* 38 (1-3): 84-105.

Neustupný, Jiri Vaclav. 1966. "Is the Japanese accent a pitch accent." *Onsei Gakkai Kaihoo* 121: 230-239.

—. 1978. *Post-structural approaches to language: Language theory in a Japanese context.* Tokyo: University of Tokyo Press.

Nishinuma, Yukihiro. 1979. *Un modèle d'analyse automatique de la prosodie: accent et intonation en japonais (Collection "Sons et parole" 1).* Centre National de la Reserche Scientifique.

OJAD, Project. n.d. *Online Japanese Accent Dictionary (OJAD).* The University of Tokyo. http://www.gavo.t.u-tokyo.ac.jp/ojad/.

Oller, D. Kimbrough. 1973. "The effect of position in utterance on speech segment duration in English." *The Journal of the Acoustical Society of America* 54 (5): 1235-1247.

Pierrehumbert, Janet, and Mary Beckman. 1988. "Japanese tone structure." *Linguistic inquiry monographs* 15: 1-282.

Port, Robert F., Salman Al-Ani, and Shosaku Maeda. 1980. "Temporal compensation and universal phonetics." *Phonetica* 37 (4): 235-252.

Poser, William John. 1990. "Evidence for foot structure in Japanese." *Language* 66: 78-105.

—. 1984. "The phonetics and phonology of tone and intonation in Japanese (Doctoral dissertation)." Massachusetts Institute of Technology.

n.d. *Prosody Tutor Suzuki-kun.* Grad. School of Eng., The University of Tokyo Hirose Laboratory, Grad. School of Info. Sci. and Tech., The University of Tokyo Minematsu Laboratory. Accessed 2022. https://www.gavo.t.u-tokyo.ac.jp/ojad/eng/phrasing/index.

Sakuma, Kanae. 1929. *Nihon onseigaku.* Kyōbunsha.

Saltzman, Elliot. 1986. "Task dynamic coordination of the speech articulators: A preliminary model." *Experimental brain research series* 15: 129-144.

Seo, Jungyun, Sahyang Kim, Haruo Kubozono, and Taehong Cho. 2019a. "Interaction between rhythmic structure and preboundary lengthening in Japanese." *Proceedings of the 19th International Congress of Phonetic Sciences.* Melbourne, Australia. 637-641.

Seo, Jungyun, Sahyang Kim, Haruo Kubozono, and Taehong Cho. 2019b. "Preboundary lengthening in Japanese: To what extent do lexical pitch accent and moraic structure matter?" *The Journal of the Acoustical Society of America* 146 (3): 1817-1823.

Shepherd, Michael A. 2008. "The scope and effects of preboundary prosodic lengthening in Japanese." *USC Working Papers in Linguistics* 4: 1-14.

Silverman, Kim E, M E Beckman, J F Pitrelli, M Ostendorf, C W Wightman, P Price, and J Hirschberg. 1992. "ToBI: A standard for labeling English prosody." *ICSLP* 2 (October): 867-870.

Silverman, Kim E., and Janet Pierrehumbert. 1990. "The timing of prenuclear high accents in English." *Papers in Laboratory Phonology I* (Cambridge University Press) 72-106.

Silverman, Kim E., Mary E. Beckman, John F. Pitrelli, Mari Ostendorf, Colin W. Wightman, Patti Price, and Julia Hirschberg. 1992. "ToBI: A standard for labeling English prosody." *ICSLP* 2 (October): 867-870.

Sugito, Miyoko. 1968. "Comparison between "high-final" and "level" tone in Toyko accent [A study of final-accented and unaccented disyllabic words in Tokyo using dynamic measurement]." *Onseigakkai Kaihoo* 129: 1-4.

Sugito, Miyoko. 1982. "Dootai sokutei ni yoru nihongo akusento no kaimei." *Nihongo akusento no kenkyuu* 49-75.

—. 1982. *Nihongo akusento no kenkyuu.* Tokyo: Sanseido.

Sugito, Miyoko. 1998. "Nihongo onsei no kenkyu 5: "hana" to "hana" [Studies on Japanese Phonetics 5]." *Osaka: Izumi Shion.*

Sugito, Miyoko, M Higashikawa, A Sakakura, and H Takahashi. 1991. "Perceptual, acoustical, and physiological study of Japanese word accent in whispered speech." *IEICE Technical Report 91 (Institute of Electronics, Information and Communication Engineers)* 1-8.

Sugiyama, Yukiko. 2012. "Perceiving pitch accent in the absence of F0." *Annual Meeting of the Berkeley Linguistics Societ* 38: 482-492.

Sugiyama, Yukiko. 2017. "Perception of Japanese pitch acent without F0." *Phonetica* 74 (2): 107-123.

—. 2012. *The production and perception of Japanese pitch accent.* Cambridge Scholars Publishing.

Sugiyama, Yukiko, CT Justine Hui, and Takayuki Arai. 2021. "The effect of fo fall, downstep, and secondary cues in perceiving Japanese lexical accent." *The Journal of the Acoustical Society of America* 150 (4): 2865-2878.

Tabain, Marija. 2003. "Effects of prosodic boundary on /aC/ sequences: articulatory results." *The Journal of the Acoustical Society of America* 113 (5): 2834-2849.

Tabain, Marija, and Pascal Perrier. 2005. "Articulation and acoustics of /i/ at prosodic boundaries in French." *Journal of Phonetics* 33: 77-100.

Team, R Development Core. 2009. *A language and environment for statistical computing.*

Tizón-Couto, David, and David Lorenz. 2021. "Variables are valuable: making a case for deductive modeling." *Linguistics* 59 (5): 1279-1309.

Toivonen, Ida, Lev Blumenfeld, Andrea Gormley, Leah Hoiting, John Logan, Nalini Ramlakhan, and Adam Stone. 2015. "Vowel height and duration." *Proceedings of the 32nd west coast conference on formal linguistics* (Cascadilla Proceedings Project) 32: 64-71.

Turk, Alice E., and Stefanie Shattuck-Hufnagel. 2007. "Multiple targets of phrase-final lengthening in American English words." *Journal of Phonetics* 35 (4): 445-472.

Vance, Timothy J. 1995. "Final accent vs. no accent: Utterance-final neutralization in Tokyo Japanese." *Journal of Phonetics* 23 (4): 487-499.

Vatikiotis-Bateson, Eric, and J.A. Scott Kelso. 1993. "Rhythm type and articulatory dynamics in English, French and Japanese." *Journal of Phonetics* 21 (3): 231-265.

Venditti, Jennifer J. 1997. "Japanese ToBI labelling guidelines." *Working Papers in Linguistics* (Ohio State University Department of Linguistics) 127-162.

Venditti, Jennifer J. 2005. "The J_ToBI model of Japanese intonation." *Prosodic typology: The phonology of intonation and phrasing* 172-200.

Venditti, Jennifer J., Kazuaki Maeda, and Jan PV Santen. 1998. "Modeling Japanese boundary pitch movements for speech synthesis." *The third ESCA?COCOSDA workshop (ETRW) on speech synthesis.*

Venditti, Jennifer J., Kikuo Maekawa, and Mary E. Beckman. 2008. "Prominence marking in the Japanese intonation system." *Handbook of Japanese linguistics* 456-512.

Warner, Natasha. 1997. "Recognition of accent patterns across dialects in Japanese." *Annual Meeting of the Berkeley Linguistics Society* 23 (1): 364-375.

Weitzman, Raymond S. 1970. "Word accent in Japanese." *Studies in the Phonology of Asian Languages IX.*

White, Laurence. 2002. "English speech timing: A domain and locus approach." Doctoral dissertatioin, University of Edinburgh.

Wightman, Colin W., Stefanie Shattuck-Hufnagel, Mari Ostendorf, and Patti J. Price. 1992. "Segmental durations in the vicinity of prosodic phrase boundaries." *The Journal of the Acoustical Society of America* 91 (3): 1707-1717.

Williams, J C, Donna Erickson, Yusuke Ozaki, Atsuo Suemitsu, Nobuaki Minematsu, and Osamu Fujimura. 2013. "Neutralizing differences in jaw displacement for English vowels." *Proceedings of International Congress of Acoustics*. 060268.

Appendix: Model outputs of analyses provided in the order presented in the dissertation

A. Model output for neologism C models. Normalized formation duration and release duration by boundary, pitch accent position, and syllable position.

	FormDUR_norm			**RelDUR_norm**		
Predictors	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*
(Intercept)	1.15	0.77 – 1.52	**<0.001**	-0.51	-0.78 – -0.24	**<0.001**
Boundary [PhM]	-0.07	-0.30 – 0.15	0.515	0.12	-0.10 – 0.34	0.280
PitchAccent [Penult]	-0.20	-0.43 – 0.03	0.085	0.04	-0.18 – 0.26	0.715
PitchAccent [Ultima]	-0.26	-0.49 – -0.03	**0.028**	0.09	-0.13 – 0.32	0.410
PitchAccent [Unaccented]	-0.38	-0.60 – -0.15	**0.001**	0.10	-0.12 – 0.32	0.363
SyllPos [C1]	-1.89	-2.11 – -1.66	**<0.001**	0.04	-0.18 – 0.26	0.707
SyllPos [C0]	-1.36	-1.59 – -1.14	**<0.001**	2.45	2.23 – 2.67	**<0.001**
Boundary [PhM] * PitchAccent [Penult]	0.20	-0.11 – 0.52	0.207	-0.02	-0.33 – 0.30	0.922
Boundary [PhM] * PitchAccent [Ultima]	0.25	-0.08 – 0.57	0.132	0.02	-0.29 – 0.34	0.882

Boundary [PhM] * PitchAccent [Unaccented]	0.17	-0.15 – 0.49	0.310	0.03	-0.28 – 0.35	0.829
Boundary [PhM] * SyllPos [C1]	-0.01	-0.33 – 0.31	0.957	-0.08	-0.39 – 0.23	0.615
Boundary [PhM] * SyllPos [C0]	-0.35	-0.67 – -0.03	**0.032**	-2.50	-2.81 – -2.19	**<0.001**
PitchAccent [Penult] *SyllPos [C1]	0.23	-0.09 – 0.55	0.161	0.05	-0.26 – 0.36	0.759
PitchAccent [Ultima] *SyllPos [C1]	0.15	-0.18 – 0.47	0.369	0.09	-0.22 – 0.41	0.566
PitchAccent [Unaccented] * SyllPos [C1]	0.21	-0.11 – 0.53	0.199	0.01	-0.30 – 0.32	0.963
PitchAccent [Penult] * SyllPos [C0]	0.29	-0.03 – 0.61	0.072	-0.18	-0.49 – 0.13	0.261
PitchAccent [Ultima] *SyllPos [C0]	0.29	-0.03 – 0.61	0.080	-0.19	-0.51 – 0.13	0.239
PitchAccent [Unaccented] * SyllPos [C0]	0.48	0.16 – 0.80	**0.003**	-0.34	-0.65 – -0.03	**0.031**
(Boundary [PhM] * PitchAccent [Penult]) *SyllPos [C1]	-0.11	-0.56 – 0.33	0.617	0.01	-0.43 – 0.45	0.962
(Boundary [PhM] * PitchAccent [Ultima]) *SyllPos [C1]	-0.02	-0.48 – 0.44	0.939	0.07	-0.38 – 0.51	0.774
(Boundary [PhM] * PitchAccent [Unaccented]) * SyllPos [C1]	0.02	-0.43 – 0.47	0.929	0.01	-0.44 – 0.45	0.978
(Boundary [PhM] * PitchAccent [Penult]) *SyllPos [C0]	-0.08	-0.53 – 0.37	0.719	0.18	-0.26 – 0.62	0.413

(Boundary [PhM] * PitchAccent [Ultima]) *SyllPos [C0]	0.30	-0.16 – 0.76	0.200	0.49	0.04 – 0.94	**0.031**
(Boundary [PhM] * PitchAccent [Unaccented]) * SyllPos [C0]	0.01	-0.44 – 0.46	0.964	0.24	-0.20 – 0.69	0.280
Random Effects						
σ^2	0.28			0.27		
τ_{00}	0.15 Speaker			0.06 Speaker		
ICC	0.34			0.19		
N	5 Speaker			5 Speaker		
Observations	1014			1014		
Marginal R^2 / Conditional R^2	0.567 / 0.716			0.667 / 0.730		

B. Summary of the results of linear mixed effects models with regard to boundary effects and the interaction between boundary and pitch accent (*p<.05, **p<.01, ***p<.001).

B x PA x S	C_F	C_R	V_F
Neologisms	*n.s.*	*n.s.*	$X^2(6) = 2.52$ *
Three-syllable real words	*n.s.*	$X^2(6) = 2.92$ **	$X^2(3) = 15.129$ ***
Two-syllable real words	$X^2(2) = 5.76$ **	$X^2(2) = 11.861$***	$X^2(2) = 3.8277$ *

C. Model output for neologism V models. Normalized formation duration by boundary, pitch accent position, and syllable position.

	FormDUR_norm		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	-0.40	-0.81 – 0.02	0.062
Boundary [PhM]	0.03	-0.29 – 0.35	0.838
PitchAccent [Penult]	0.01	-0.31 – 0.33	0.970
PitchAccent [Ultima]	0.06	-0.26 – 0.39	0.702
PitchAccent [Unaccented]	0.10	-0.22 – 0.41	0.555
SyllPos [V1]	0.48	0.16 – 0.80	**0.003**
SyllPos [V0]	1.79	1.48 – 2.11	**<0.001**
Boundary [PhM] * PitchAccent [Penult]	0.08	-0.36 – 0.53	0.716
Boundary [PhM] * PitchAccent [Ultima]	0.07	-0.38 – 0.53	0.751
Boundary [PhM] * PitchAccent [Unaccented]	-0.02	-0.47 – 0.44	0.946
Boundary [PhM] * SyllPos [V1]	-0.62	-1.07 – -0.17	**0.007**
Boundary [PhM] * SyllPos [V0]	-1.96	-2.41 – -1.51	**<0.001**
PitchAccent [Penult] * SyllPos [V1]	0.12	-0.34 – 0.57	0.614
PitchAccent [Ultima] * SyllPos [V1]	0.11	-0.35 – 0.56	0.650
PitchAccent [Unaccented] * SyllPos [V1]	-0.03	-0.48 – 0.42	0.902
PitchAccent [Penult] * SyllPos [V0]	-0.06	-0.51 – 0.39	0.794

PitchAccent [Ultima] * SyllPos [V0]	-0.71	-1.17 – -0.26	**0.002**
PitchAccent [Unaccented] * SyllPos [V0]	-0.67	-1.11 – -0.23	**0.003**
(Boundary [PhM] * PitchAccent [Penult]) * SyllPos [V1]	0.12	-0.51 – 0.75	0.700
(Boundary [PhM] * PitchAccent [Ultima]) * SyllPos [V1]	0.73	0.09 – 1.37	**0.026**
(Boundary [PhM] * PitchAccent [Unaccented]) * SyllPos [V1]	0.21	-0.42 – 0.85	0.508
(Boundary [PhM] * PitchAccent [Penult]) * SyllPos [V0]	0.13	-0.49 – 0.76	0.678
(Boundary [PhM] * PitchAccent [Ultima]) * SyllPos [V0]	0.94	0.30 – 1.58	**0.004**
(Boundary [PhM] * PitchAccent [Unaccented]) * SyllPos [V0]	0.80	0.16 – 1.43	**0.014**
Random Effects			
σ^2	0.54		
$\tau_{00\ \mathrm{Speaker}}$	0.16		
ICC	0.22		
N_{Speaker}	5		
Observations	1001		
Marginal R^2 / Conditional R^2	0.298 / 0.456		

D. Model output for three-syllable word C models. Normalized formation duration and release duration by boundary, pitch accent position, and syllable position.

	FormDUR_norm			RelDUR_norm		
Predictors	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*
(Intercept)	0.32	-0.11 – 0.75	0.144	-0.17	-0.50 – 0.16	0.303
Boundary [PhM]	-0.01	-0.38 – 0.37	0.976	-0.18	-0.55 – 0.19	0.336
PitchAccent [Penult]	-0.48	-0.80 – -0.16	**0.004**	-0.23	-0.55 – 0.09	0.162
PitchAccent [Ultima]	0.11	-0.26 – 0.48	0.555	1.25	0.88 – 1.62	**<0.001**
PitchAccent [Unaccented]	-0.41	-0.78 – -0.04	**0.031**	0.24	-0.13 – 0.61	0.204
SyllPos [C1]	-0.31	-0.68 – 0.07	0.106	0.54	0.16 – 0.91	**0.005**
SyllPos [C0]	0.43	0.06 – 0.80	**0.022**	1.35	0.97 – 1.72	**<0.001**
Boundary [PhM] * PitchAccent [Penult]	-0.01	-0.47 – 0.45	0.967	0.09	-0.37 – 0.55	0.697
Boundary [PhM] * PitchAccent [Ultima]	-0.03	-0.57 – 0.52	0.923	0.14	-0.41 – 0.69	0.614
Boundary [PhM] * PitchAccent [Unaccented]	0.28	-0.24 – 0.80	0.295	0.17	-0.35 – 0.70	0.512
Boundary [PhM] * SyllPos [C1]	0.03	-0.50 – 0.56	0.900	0.19	-0.35 – 0.72	0.493
Boundary [PhM] * SyllPos [C0]	-0.71	-1.23 – -0.18	**0.008**	-0.96	-1.49 – -0.44	**<0.001**
PitchAccent [Penult] * SyllPos [C1]	0.15	-0.31 – 0.61	0.521	-0.58	-1.04 – -0.12	**0.014**

PitchAccent [Ultima] * SyllPos [C1]	-0.16	-0.71 – 0.39	0.574	-1.47	-2.02 – -0.92	**<0.001**
PitchAccent [Unaccented] * SyllPos [C1]	0.56	0.03 – 1.08	**0.037**	-0.27	-0.79 – 0.26	0.319
PitchAccent [Penult] * SyllPos [C0]	-0.09	-0.54 – 0.37	0.706	-0.39	-0.84 – 0.07	0.096
PitchAccent [Ultima] * SyllPos [C0]	0.02	-0.50 – 0.55	0.930	-1.76	-2.28 – -1.23	**<0.001**
PitchAccent [Unaccented] * SyllPos [C0]	-0.15	-0.67 – 0.37	0.573	-1.02	-1.55 – -0.50	**<0.001**
(Boundary [PhM] * PitchAccent [Penult]) * SyllPos [C1]	-0.06	-0.71 – 0.59	0.853	-0.21	-0.86 – 0.44	0.531
(Boundary [PhM] * PitchAccent [Ultima]) * SyllPos [C1]	-0.26	-1.08 – 0.55	0.528	-0.59	-1.40 – 0.22	0.155
(Boundary [PhM] * PitchAccent [Unaccented]) * SyllPos [C1]	-0.27	-1.01 – 0.47	0.475	-0.11	-0.85 – 0.64	0.774
(Boundary [PhM] * PitchAccent [Penult]) * SyllPos [C0]	-0.16	-0.80 – 0.48	0.629	-0.20	-0.84 – 0.45	0.549
(Boundary [PhM] * PitchAccent [Ultima]) * SyllPos [C0]	0.20	-0.57 – 0.96	0.614	0.78	0.01 – 1.54	**0.046**
(Boundary [PhM] * PitchAccent [Unaccented]) * SyllPos [C0]	0.03	-0.70 – 0.77	0.928	0.16	-0.58 – 0.89	0.678
Random Effects						
σ^2	0.77			0.77		
τ_{00}	0.15 $_{Speaker}$			0.05 $_{Speaker}$		

ICC	0.16	0.06
N	5 Speaker	5 Speaker
Observations	1241	1241
Marginal R^2 / Conditional R^2	0.126 / 0.270	0.263 / 0.308

E. Model output for three-syllable word V model. Normalized formation duration by boundary and pitch accent.

	FormDUR_norm		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	1.78	1.51 – 2.05	**<0.001**
Boundary [PhM]	-2.33	-2.62 – -2.04	**<0.001**
PitchAccent [Penult]	-0.45	-0.74 – -0.17	**0.002**
PitchAccent [Ultima]	-1.53	-1.82 – -1.25	**<0.001**
PitchAccent [Unaccented]	-0.67	-0.98 – -0.37	**<0.001**
Boundary [PhM] * PitchAccent [Penult]	0.33	-0.06 – 0.72	0.100
Boundary [PhM] * PitchAccent [Ultima]	1.16	0.77 – 1.55	**<0.001**
Boundary [PhM] * PitchAccent [Unaccented]	1.02	0.59 – 1.44	**<0.001**
Random Effects			
σ^2	0.38		
τ_{00} Speaker	0.04		
ICC	0.09		
N Speaker	5		
Observations	303		
Marginal R^2 / Conditional R^2	0.686 / 0.713		

F. Model output for two-syllable word C models. Normalized formation duration and release duration by boundary, pitch accent position, and syllable position.

	FormDUR_norm			RelDUR_norm		
Predictors	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*
(Intercept)	-0.19	-0.64 – 0.25	0.391	-0.07	-0.36 – 0.22	0.639
Boundary [PhM]	0.20	-0.12 – 0.53	0.223	-0.24	-0.47 – -0.01	**0.044**
PitchAccent [Ultima]	0.11	-0.20 – 0.43	0.482	0.02	-0.21 – 0.24	0.888
PitchAccent [Unaccented]	0.19	-0.13 – 0.51	0.245	-0.12	-0.35 – 0.10	0.289
SyllPos [C0]	0.76	0.43 – 1.08	**<0.001**	0.65	0.42 – 0.88	**<0.001**
Boundary [PhM] * PitchAccent [Ultima]	-0.32	-0.78 – 0.14	0.167	0.01	-0.31 – 0.34	0.938
Boundary [PhM] * PitchAccent [Unaccented]	-0.38	-0.84 – 0.08	0.102	0.33	0.01 – 0.66	**0.044**
Boundary [PhM] * SyllPos [C0]	-1.34	-1.80 – -0.87	**<0.001**	-1.57	-1.90 – -1.25	**<0.001**
PitchAccent [Ultima] * SyllPos [C0]	-0.18	-0.64 – 0.28	0.447	-0.13	-0.45 – 0.19	0.421
PitchAccent [Unaccented] * SyllPos [C0]	-0.31	-0.77 – 0.15	0.193	0.30	-0.02 – 0.63	0.065
(Boundary [PhM] * PitchAccent [Ultima]) * SyllPos [C0]	0.59	-0.07 – 1.24	0.080	0.16	-0.30 – 0.62	0.505
(Boundary [PhM] * PitchAccent [Unaccented]) * SyllPos [C0]	1.13	0.48 – 1.79	**0.001**	1.06	0.60 – 1.52	**<0.001**

Random Effects		
σ^2	0.61	0.30
τ_{00}	0.19 Speaker	0.08 Speaker
ICC	0.24	0.21
N	5 Speaker	5 Speaker
Observations	523	523
Marginal R^2 / Conditional R^2	0.122 / 0.334	0.477 / 0.586

G. Model output for two-syllable word V model. Normalized formation duration by boundary, pitch accent position, and syllable position.

	FormDUR_norm		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	-0.16	-0.39 – 0.06	0.156
Boundary [PhM]	-0.04	-0.30 – 0.22	0.762
PitchAccent [Ultima]	-0.09	-0.35 – 0.18	0.522
PitchAccent [Unaccented]	-0.03	-0.31 – 0.24	0.817
SyllPos [V0]	1.37	1.11 – 1.63	**<0.001**
Boundary [PhM] * PitchAccent [Ultima]	0.02	-0.35 – 0.40	0.899
Boundary [PhM] * PitchAccent [Unaccented]	-0.20	-0.57 – 0.18	0.308
Boundary [PhM] * SyllPos [V0]	-1.87	-2.24 – -1.51	**<0.001**
PitchAccent [Ultima] * SyllPos [V0]	-0.48	-0.86 – -0.10	**0.013**
PitchAccent [Unaccented] * SyllPos [V0]	-0.58	-0.97 – -0.20	**0.003**

(Boundary [PhM] * PitchAccent [Ultima]) * SyllPos [V0]	0.54	0.01 – 1.07	**0.047**
(Boundary [PhM] * PitchAccent [Unaccented]) * SyllPos [V0]	0.72	0.18 – 1.25	**0.009**
Random Effects			
σ^2	0.37		
$\tau_{00\ Speaker}$	0.02		
ICC	0.06		
$N_{Speaker}$	5		
Observations	482		
Marginal R^2 / Conditional R^2	0.472 / 0.502		

H. Summary of the results of linear mixed effects models for H boundary tone initiation showing a main effect of pitch accent (*p<.05, **p<.01, ***p<.001).

	Real words	Neologisms
PITCH ACCENT	$X^2(1) = 43.093$ **	$X^2(1) = 6.5931$ *

I. Model output for linear mixed effects models for real words and neologisms. Interval between V gesture onset and onset of boundary H tone by pitch accent position.

		VONStoBT_norm			VONStoBT_norm	
Predictors	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*
(Intercept)	-0.76	-1.07 – -0.44	**<0.001**	-0.46	-0.97 – 0.06	0.080
PitchAccent [Penultimate]	1.42	0.99 – 1.84	**<0.001**	0.79	0.27 – 1.30	**0.004**
Random Effects						

σ^2	0.48	0.72
τ_{00}	0.00 Speaker	0.12 Speaker
ICC		0.14
N	4 Speaker	4 Speaker
Observations	43	44
Marginal R^2 / Conditional R^2	0.516 / 0.516	0.158 / 0.275

J. Model output for normalized formation duration of neologism consonant and vowel constriction gestures. Formation duration by accent status and syllable position.

	Consonants FormDUR_norm			Vowels FormDUR_norm		
Predictors	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*
(Intercept)	1.06	0.69 – 1.44	**<0.001**	0.10	-0.31 – 0.52	0.626
ACCENT [UA]	-0.00	-0.19 – 0.19	0.983	-0.02	-0.22 – 0.17	0.816
ACCENT [UU]	-0.21	-0.43 – 0.02	0.071	-0.04	-0.26 – 0.19	0.738
SyllPos [C1]	-1.73	-1.95 – -1.51	**<0.001**			
SyllPos [C0]	-1.10	-1.33 – -0.88	**<0.001**			
ACCENT [UA] * SyllPos [C1]	-0.06	-0.33 – 0.21	0.683			
ACCENT [UU] * SyllPos [C1]	0.11	-0.21 – 0.42	0.506			
ACCENT [UA] * SyllPos [C0]	-0.45	-0.73 – -0.18	**0.001**			
ACCENT [UU] * SyllPos [C0]	-0.09	-0.41 – 0.23	0.598			

SyllPos [V1]				-0.04	-0.23 – 0.16	0.700
SyllPos [V0]				-0.09	-0.28 – 0.11	0.382
Random Effects						
σ^2	0.28			0.82		
τ_{00}	0.15 Speaker			0.17 Speaker		
ICC	0.36			0.17		
N	5 Speaker			5 Speaker		
Observations	510			506		
Marginal R^2 / Conditional R^2	0.572 / 0.724			0.001 / 0.174		

K. Model output for normalized formation duration in consonant and vowel constriction gestures in three-syllable real words. Formation duration by accent status and syllable position.

	Consonants FormDUR_norm			**Vowels FormDUR_norm**		
Predictors	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*
(Intercept)	0.36	-0.10 – 0.82	0.124	-0.80	-1.36 – -0.24	**0.005**
ACCENT [UA]	-0.33	-0.64 – -0.02	**0.035**	0.94	0.61 – 1.26	**<0.001**
ACCENT [UU]	-0.13	-0.50 – 0.24	0.498	0.95	0.53 – 1.36	**<0.001**
SyllPos [C1]	-0.57	-0.89 – -0.24	**0.001**			
SyllPos [C0]	0.30	-0.09 – 0.69	0.127			
ACCENT [UA] * SyllPos [C1]	0.71	0.29 – 1.13	**0.001**			

ACCENT [UU] * SyllPos [C1]	0.83	0.34 – 1.32	**0.001**			
ACCENT [UA] * SyllPos [C0]	-0.61	-1.05 – -0.16	**0.008**			
ACCENT [UU] * SyllPos [C0]	-0.55	-1.08 – -0.02	**0.041**			
Random Effects						
σ^2	0.77			0.76		
τ_{00}	0.18 Speaker			0.31 Speaker		
ICC	0.19			0.29		
N	5 Speaker			5 Speaker		
Observations	613			155		
Marginal R^2 / Conditional R^2	0.077 / 0.254			0.143 / 0.390		

L. Model output for normalized formation duration of consonant and vowel constriction gestures in two-syllable real words. Formation duration by accent status and syllable position.

	Consonants FormDUR_norm			**Vowels FormDUR_norm**		
Predictors	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*
(Intercept)	0.05	-0.44 – 0.55	0.838	0.23	-0.27 – 0.73	0.365
ACCENT [UA]	-0.21	-0.55 – 0.13	0.233	-0.07	-0.38 – 0.24	0.652
ACCENT [UU]	-0.19	-0.53 – 0.15	0.265	-0.33	-0.63 – -0.02	**0.037**
SyllPos [C0]	-0.16	-0.50 – 0.19	0.367			
ACCENT [UA] * SyllPos [C0]	-0.01	-0.49 – 0.48	0.975			

ACCENT [UU] * SyllPos [C0]	0.65	0.17 – 1.14	**0.008**			
SyllPos [V0]				0.09	-0.23 – 0.40	0.589
ACCENT [UA] * SyllPos [V0]				0.12	-0.32 – 0.56	0.579
ACCENT [UU] * SyllPos [V0]				-0.18	-0.62 – 0.26	0.427
Random Effects						
σ^2	0.64			0.52		
τ_{00}	0.24 $_{\text{Speaker}}$			0.26 $_{\text{Speaker}}$		
ICC	0.27			0.34		
N	5 $_{\text{Speaker}}$			5 $_{\text{Speaker}}$		
Observations	257			250		
Marginal R^2 / Conditional R^2	0.049 / 0.308			0.052 / 0.373		

M. Summary of the results of linear mixed effects models for duration with regard to ACCENT and SYLLABLE POSITION (*p<.05, **p<.01, ***p<.001).

ACCENT*SYLLABLE POSITION	Consonants	Vowels
Neologisms	$X^2(4) = 3.86$ **	*n.s.*
Three-syllable real words	$X^2(4) = 11.316$ ***	--
Two-syllable real words	$X^2(2) = 4.8507$ **	*n.s.*

N. Model output for normalized displacement of consonant and vowel constriction gestures in neologisms. Normalized displacement by accent status and syllable position.

	Consonants DISP_norm			Vowels DISP_norm		
Predictors	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*
(Intercept)	1.08	0.84 – 1.33	**<0.001**	0.10	-0.52 – 0.73	0.743
ACCENT [UA]	-0.06	-0.18 – 0.06	0.309	-0.07	-0.22 – 0.08	0.357
ACCENT [UU]	-0.11	-0.25 – 0.03	0.129	0.05	-0.12 – 0.23	0.542
SyllPos [C1]	-1.23	-1.35 – -1.11	**<0.001**			
SyllPos [C0]	-1.89	-2.01 – -1.77	**<0.001**			
SyllPos [V1]				-0.02	-0.17 – 0.13	0.777
SyllPos [V0]				-0.06	-0.21 – 0.09	0.438
Random Effects						
σ^2	0.32			0.49		
τ_{00}	0.06 Speaker			0.48 Speaker		
ICC	0.15			0.49		
N	5 Speaker			5 Speaker		
Observations	510			506		
Marginal R^2 / Conditional R^2	0.617 / 0.676			0.003 / 0.494		

O. Model output for normalized displacement of consonant and vowel constriction gestures in three-syllable words. Normalized displacement by accent status and syllable position.

	Consonants DISP_norm			Vowels DISP_norm		
Predictors	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*
(Intercept)	-0.02	-0.38 – 0.35	0.922	-0.89	-1.42 – -0.36	**0.001**
ACCENT [UA]	-0.25	-0.60 – 0.09	0.152	1.11	0.89 – 1.34	**<0.001**
ACCENT [UU]	-0.29	-0.71 – 0.12	0.164	1.04	0.75 – 1.34	**<0.001**
SyllPos [C1]	0.35	-0.01 – 0.72	0.055			
SyllPos [C0]	0.81	0.37 – 1.24	**<0.001**			
ACCENT [UA] * SyllPos [C1]	0.05	-0.42 – 0.52	0.837			
ACCENT [UU] * SyllPos [C1]	0.25	-0.30 – 0.80	0.368			
ACCENT [UA] * SyllPos [C0]	-0.80	-1.29 – -0.30	**0.002**			
ACCENT [UU] * SyllPos [C0]	-0.49	-1.09 – 0.10	0.106			
Random Effects						
σ^2	0.97			0.37		
τ_{00}	0.06 Speaker			0.31 Speaker		
ICC	0.06			0.46		
N	5 Speaker			5 Speaker		
Observations	613			155		

Marginal R^2 / Conditional R^2	0.087 / 0.140	0.262 / 0.601

P. Model output for normalized displacement of consonant and vowel constriction gestures in two-syllable real words. Normalized displacement by accent status and syllable position.

	Consonants DISP_norm			**Vowels DISP_norm**		
Predictors	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*
(Intercept)	-0.26	-0.64 – 0.13	0.187	0.16	-0.55 – 0.87	0.661
ACCENT [UA]	-0.03	-0.23 – 0.17	0.765	-0.00	-0.20 – 0.20	0.991
ACCENT [UU]	-0.01	-0.21 – 0.20	0.959	-0.09	-0.29 – 0.11	0.359
SyllPos [C0]	0.63	0.46 – 0.79	**<0.001**			
SyllPos [V0]				0.20	0.03 – 0.36	**0.018**
Random Effects						
σ^2	0.45			0.43		
τ_{00}	0.15 $_{Speaker}$			0.62 $_{Speaker}$		
ICC	0.25			0.59		
N	5 $_{Speaker}$			5 $_{Speaker}$		
Observations	257			250		
Marginal R^2 / Conditional R^2	0.141 / 0.358			0.011 / 0.597		

Q. Summary of the results of linear mixed effects models for displacement with regard to ACCENT and SYLLABLE POSITION (*p<.05, **p<.01, ***p<.001).

ACCENT*SYLLABLE POSITION	Consonants	Vowels
Neologisms	*n.s.*	*n.s.*
Three-syllable real words	$X^2(4) = 3.95$ **	--
Two-syllable real words	*n.s.*	*n.s.*

R. Model output for tonal alignment of pitch accent L tone with kinematic landmarks in real words

		duration	
Predictors	*Estimates*	*CI*	*p*
(Intercept)	23.10	16.80 – 29.40	**<0.001**
GROUP [C.pvel]	-14.55	-22.00 – -7.09	**<0.001**
GROUP [C.target]	14.91	7.45 – 22.36	**<0.001**
GROUP [V.onset]	-4.91	-12.36 – 2.55	0.196
GROUP [V.pvel]	53.09	45.64 – 60.55	**<0.001**
GROUP [V.target]	126.00	118.55 – 133.45	**<0.001**
Expt [real]	-3.06	-11.80 – 5.68	0.491
GROUP [C.pvel] * Expt [real]	6.96	-5.27 – 19.19	0.263
GROUP [C.target] * Expt [real]	2.26	-9.97 – 14.49	0.716
GROUP [V.onset] * Expt [real]	21.34	9.11 – 33.57	**0.001**

GROUP [V.pvel] * Expt [real]	-42.40	-54.63 – -30.17	**<0.001**
GROUP [V.target] * Expt [real]	-73.75	-85.99 – -61.52	**<0.001**
Random Effects			
σ^2	157.17		
$\tau_{00\ Speaker}$	11.83		
ICC	0.07		
$N_{Speaker}$	4		
Observations	210		
Marginal R^2 / Conditional R^2	0.909 / 0.915		