PROSODIC INFLUENCES ON THE ACOUSTICS OF VOWEL SEQUENCES

by

Miao Zhang

August 31, 2022

A dissertation submitted to the
Faculty of the Graduate School of
the University at Buffalo, The State University of New York
in partial fulfilment of the requirements for the
degree of

Doctor of Philosophy

Department of Linguistics

Acknowledgments

I owe great thanks to Dr. Christian Dicanio, who has given me much academic advice and helped, encouraged, and believed in me from my first days at the University at Buffalo till now. Without his help and encouragement, I would have never made the transition from a morphosyntactician to a phonetician. A number of people have given me insightful suggestions and help with my work. I would like to thank Dr. Matthew Faytak, Dr. Jean-Pierre Koenig, and Dr. Weirong Chen for reading and commenting on my dissertation and Dr. Richard Hatcher for his enthusiastic discussion with me on phonetics and phonology throughout my writing. Thanks also go to all my speakers for patiently getting through the many sentences in my experiments. I am also very grateful to Dr. Karin Michelson and Dr. Jeff Good for reading my qualifying paper. I greatly appreciate the help from all the UB PhonLab members: Pegi Bakula, Joshua Benn, Braden Brown, Genevieve Franck, and Jared Sharp, who were always helpful with helping my experiments and discussing research ideas. Thanks also to my best non-phonetician friend and colleague in the department: Dr. Yanwei Jin. It has always been a great joy talking about linguistic stuff with him. I would also like to thank all my colleagues in the Chinese language program in the department of linguistics. I would especially like to thank Yongbo Tian for making teaching Chinese a joyful task while I pursue my academic goals at UB. I am thankful to all students, faculty, and staff at the linguistics department for making it such a stimulating environment and a wonderful place to be. Special thanks go to my best student, an enthusiastic cook, Nathan Lynch, for making me feel at home in America while I cannot return to China due to the

pandemic.

Before coming to UB, I spent two and a half years at the University of Tsukuba, and I would like to thank the Department of International Area Studies for their warm welcome. I am incredibly grateful to Dr. Shingo Imai for his massive help at the beginning of my stay in Japan and for showing me the fascinating world of linguistics. I would also like to thank Prof. Chieko Kano for her warm-hearted encouragement when I hit the lowest point of my life in Japan. While at the University of Tsukuba, I also met Dizhogn Fang, Lizhen Huang, Ruizheng Yu, and Meng Yuan. They made my study experience in Japan a great time.

Further away, I would like to thank my professor at the Department of Foreign Languages at Renmin University of China, Dr. Yiqun Wang, for introducing me to the study of language. I am grateful to Xiaoshi Hu for all the years we spent together from high school to college, for sharing his enthusiasm for linguistics, and for believing in me from the start. I am lucky to have had my friends from Renmin University of China, Xin Chen, Fei Fu, Cong Li, and Wuda Pan, for all the great times we had in Beijing.

I am fortunate to have friends I made from Zhihu.com, Totolalatum, Heizhishenglei, Tansuanqinglei, Guillem, Erjintingyu, Chengzhir, Linga, Dr. Yunfan Lai, and from Douban.com, Dr. Jian Zhu. Thank you for your insights into linguistics and phonetics that have stimulated me and for your company while I was writing my dissertation, even though I do not know all your names. I want to thank Totolalatum for discussing teaching Chinese and studying phonetics.

My final thanks go to my family, my wife Wenjing Zhou, and my parents, Chuansui Zhang and Jurong Xiao. I could not have made such far without your support over the years. I am really thankful to them for always being there for me and going through every step with me. Without them, this dissertation would not have been possible and would not mean nearly as much.

.

Table of Contents

List of Tables

List of Figures

Abstract

This dissertation examines prosodic influence on the acoustic properties of tautosyllabic vowel sequences (TVS) using acoustic data. The analyses focused on how the duration of the TVS, the excursion of the formants, and the movement of the TVS are influenced by prosodic boundaries that follow the speech sound. The prosodic structure is understood as an abstract hierarchical structure of prosodic phrasing in this research. At the boundaries of prosodic constituents, prosodic phrase boundaries introduce systematic phonetic variation in the temporal and spatial properties of segments. Four different TVS (/ai, au, ae, ou/) in three languages with different prosodic characteristics (Chinese as a tone language, English as a stress language, and Japanese as a language with mora as its basic prosodic unit) were investigated. This research serves as the first research to cross-linguistically study how TVS are influenced by prosodic structure.

The results show that first, the pre-boundary lengthening is confirmed in all three languages but implemented differently cross-linguistically. Chinese TVS were less affected by prosody than those in English and Japanese. Monophthongs are less lengthened than TVS pre-boundarily. However, the difference between the lengthening of monophthongs and TVS is yet unclear.

Second, TVS are hyperarticulated by prosodic boundaries with more extreme acoustic properties indicating enhancement of the distinctive features of the vocalic targets. Japanese TVS are different than those in Chinese and English in that the onset vocalic targets are also influenced by prosodic boundaries while the onsets of those in Chinese and

English are not. This suggests that in Japanese TVS, the first vocalic target is more salient than in Chinese and English.

Third, the strategy of modulation on the TVS trajectory in the F1/F2 vowel space mainly involved stiffness reduction and target rescaling. This is somehow different than the result reported in the literature of articulatory study on pre-boundary strengthening since stiffness reduction has been found to be the major strategy in pre-boundary prosodic modulation. This result suggests a discrepancy between the acoustic and the articulatory domain.

In general, this dissertation demonstrated that TVS is influenced by prosodic structures, although the effect differs for languages and specific TVS. The effect is slightly different than those on monophthongs, and those found in the articulatory study.

Chapter 1

Introduction

The dissertation focuses on prosodic structure and its phonetic realization via investigating the spectral patterns of tautosyllabic vowel sequences in Chinese, English, and Japanese.

Speech production is highly variable and adaptive to meet contextual demands. Speakers tune their production according to the communicative and situational demands to optimize the interplay between speaker-oriented and listen-oriented factors. From the speaker-oriented perspective, one of the most important questions in phonetic theory is how speech sounds are produced in contexts. Understanding how speech production is structured in contexts is vital in understanding how humans produce language to serve different goals in verbal communication. This study investigates speaker-oriented factors in the production of tautosyllabic vowel sequences in prosodic contexts. This study specifically studies how prosodic boundaries introduce systematic variation in the production of vowel sequences. Acoustic studies have shown that the duration of segments lengthens at boundaries (Klatt, 1975; Oller, 1973; Wightman et al., 1992, etc.). In the articulatory domain, it was also found that the production of gestures slows down at boundaries (Byrd, 2000; Byrd & Saltzman, 1998; Edwards et al., 1991). I will analyze the prosodic boundary influence on the duration, the formant excursions, and the trajectory movement of vowel sequences.

1.1 Prosody

Prosody is one source of variance in segmental and suprasegmental production (Fry, 1965; B. Lindblom, 1983; B. Lindblom, 1968; van Santen & Shih, 2000, etc.). To scientifically study the influence of linguistic prosody on speech production, one has first to define what linguistic prosody is. Prosody has been approached in previous phonetic studies from different angles. Lehiste's pioneer study Lehiste (1970) defined the difference between several different variables in analyzing speech signals: "suprasegmental" and "segmental". Segmental features are those inherent to phonological phonemes and allophones, such as spectral patterns. Suprasegmentals refer to those features that are overlaid with phonologically inherent variables, such as pitch, duration, and intensity. The suprasegmental variables are involved in the signaling of paralinguistic affect and emotion as well as linguistic functions such as information structure and pragmatic functions. For example, English declarative sentences tend to be produced with a falling contour at the end of the utterance, as in "John wants to eat the ↘cake", while yes-no questions tend to be produced with a rising pitch contour, as in "Does ↗John want to eat the cake". In this regard, prosody was referred to as these low-level phonetic suprasegmental variables.

Later theoretical research on prosody drifted away from the aforementioned phonetic definition of prosody. Instead, it is assumed that the detailed suprasegmental and other phonetic variables are encoded in prosodic structure, which is created as a blueprint for motor execution of the utterance so that the abstract phonological (or segmental) representations that constitute the planned utterance are fleshed out with fine-grained phonetic content as specified by the prosodic structure. The mainstream theoretical phonetic/phonological studies have defined linguistic prosody in more abstract ways as the phrasal organization and accentual prominence in speech (Ladd, 2008; Pierrehumbert, 1980). Phrasal organization refers to that linguistic unit must be prosodically parsed into prosodic constituents such as syllables, words, feet, and larger phrases hierarchically in a way that is not in isomor-

phic relation with the structure parsed by lexico-phonological rules. For example, the coda consonant at the end of a lexical item is usually resyllabified to the onset of the following word if the following word does not contain a consonantal onset in many languages, as in "keep ahead [kiː.pə.hed]"('.' indicates syllable boundary), showing that the boundaries of lexical items may not strictly align with prosodic boundaries. This is called prosodic parsing. Prosodic parsing is considered obligatory by many researchers (Nespor & Vogel, 1986; E. O. Selkirk, 1986). Prosodic parsing is done in terms of prosodic constituents ranging from morae to utterances, with syllables, feet, phonological/prosodic words, accentual phrases, intermediate phrases, and intonational phrases in between. This is illustrated in Figure 1.1 with three representative theoretical models.

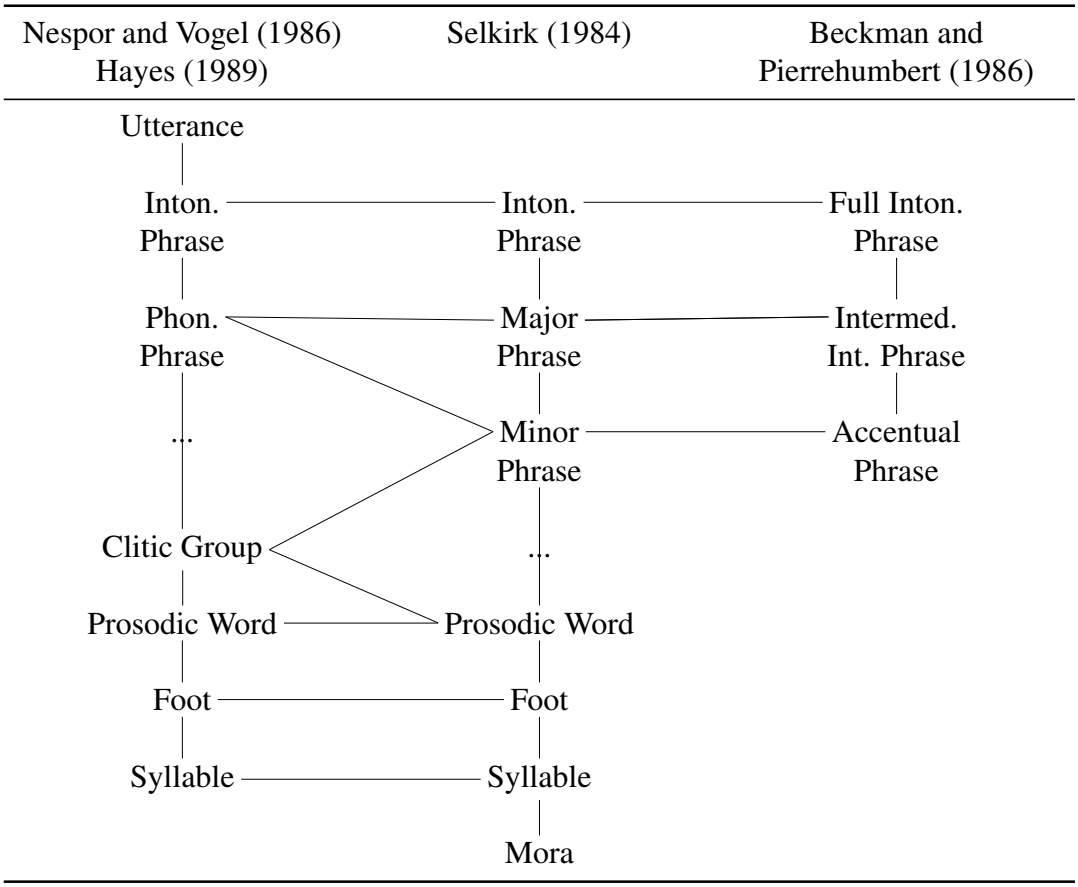

Figure 1.1: Prosodic theories proposed in the literature (figure adapted from Shattuck-Hufnagel & Turk, 1996).

Although theories differ in classifying intermediate prosodic levels, the consensus among researchers is that at the lowest level lies mora/syllable. Each level within the structure represents the groupings of units from a lower level and is, in turn, the materials to be grouped in a higher level up to the utterance or intonational phrase. A higher prosodic constituent dominates a lower one in the hierarchy, and most of the constituents are non-recursive except for intonational phrase (Ladd, 1986). The crucial assumption of hierarchical prosodic structure is that the constituents in the hierarchy have their pattern of phonetic encoding (e.g., durational patterns, formant excursion of vowels, hyper-hypoarticulation, tonal associations) to allow speakers to construct utterances and to enable the listeners to reconstruct the intended structure of an utterance. This structure thus allows a relatively unified and comparable way to describe the prosodic phenomena in languages around the world (S.-A. Jun, 2005; S.-a. Jun, 2014).

Within this framework, prosody is considered to have two major functions: DELIMITATIVE and CULMINATIVE function (Beckman, 1986; Cho, 2016), corresponding to the locations of the *edge* and the *head* of prosodic constituents. The delimitative function is to group meaningful linguistic units into prosodic phrases, the subunits more structurally cohesive to each other than to units from other prosodic chunks. The culminative function of prosody is to mark the head of each prosodic phrase structurally. They are also called the phrasing and prominence respectively in other studies (see D. L. Bolinger, 1958; Cutler et al., 1997; Krivokapić, 2007; Krivokapić, 2012, etc.). In addition to the two major functions above, prosody is widely used in marking the pragmatic prominence of linguistic expressions. Prominence reflects what information in the utterance is given or new and whether the information is highlighted to contrast with other information in the context (D. Bolinger, 1972). Prominence relates to information structure (Lambrecht, 1994). The current study will only focus on prosody's delimitative or phrasing function. Figure 1.2 below illustrates how prosody parses the structure of an utterance with two intonational phrases in English. In the figure, we see that prosodic structure provides the phrasing in an utterance

and conveys where the post-lexical tones should associate. The post-lexical tones do not associate with particular segments in the utterance but with the nodes at different levels in the structure. The prosodic constituents or prosodic *domains* serve as domains of tonal applications and phonological rules E. Selkirk (1995). For example, the flapping/tapping of coronal stops in American English can occur across higher boundaries (rider: [ɹaɪɾɚ] vs. ride a bike: [ɹaɪɹ#ə#baɪk]). Instead of being confined within the domain of prosodic words, it is sensitive to the position of the stress and pitch-accent (de Jong, 1998).

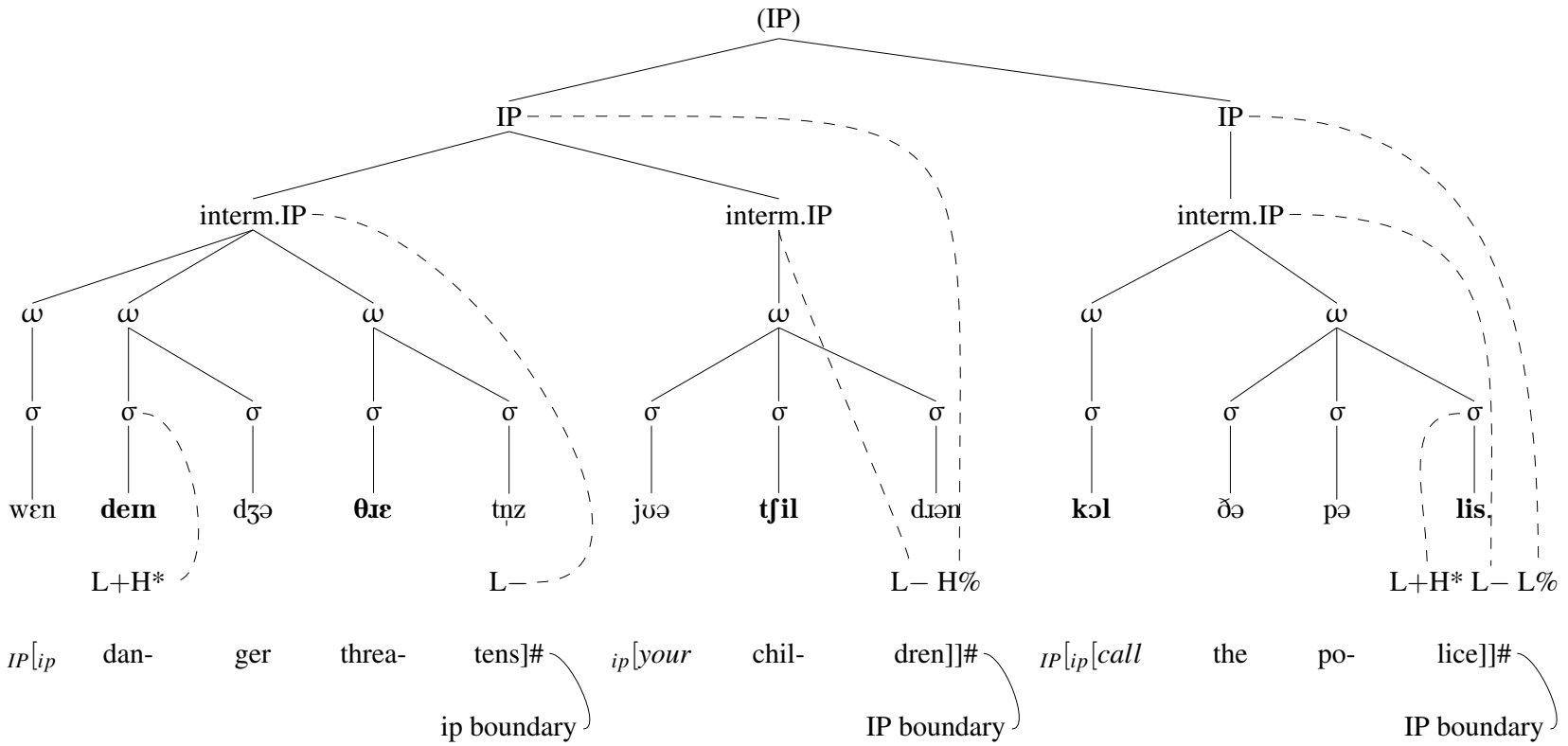

Figure 1.2: A prosodic structure of *When danger threatens your children, call the police* (Adapted from (Cho, 2016, p. 122)). It depicts a hierarchically organized phonological structure of the utterance in terms of phrasing and post-lexical association. Note that '−' in the association line indicates stressed syllables; H* refers to a High tone as the nuclear pitch accent; L-, a phrase tone at the end of an intermediate phrase (ip); L% and H%, the boundary tones at the end of an IP.

Changing the phrasing of an utterance can change the meaning conveyed. An example is shown below.

(1) a. When you make hollandaise slowly, it curdles.

b. When you make hollandaise, slowly it curdles.

The two sentences have the same segmental makeup but only are distinguished in the prosodic phrasal organization. In (1a), the pause is inserted after *slowly* whereas in (1b) it is inserted before the word *slowly*. Depending on which prosodic phrase *slowly* was grouped, the target verb phrase it modifies differs. In (1a), it is the action *making hollandaise* that is slow, while in (1b), it is the action *curdling* that is slow.

Research on prosody and phonetic correlates of prosody in the past several decades has primarily adopted the structural view of linguistic prosody starting from (Pierrehumbert, 1980). Pierrehumbert views prosody as tone and intonation arranged in a structured way to express both the locations of the head and the boundary of prosodic constituents. Other scholars later developed this framework and became what was then called *autosegmental-metrical theory* of prosody (Ladd, 2008). The current study also adopts the structural view of linguistic prosody.

Before further discussion, note that prosody in this dissertation should not be confused with another related yet different concept "intonation". According to Ladd (2008, p. 4), "The use of suprasegmental phonetic features to convey postlexical or sentence-level pragmatic meanings in a linguistically structured way." Most typical suprasegmental phonetic features used include duration and pitch. Intonation, therefore, is not merely structural but also functional.

(2) a. He ate the ↘ cake.

b. He ↘ ate the cake.

c. He ate the ↗ cake?

d. ↗ He ate the cake?

The direction of the arrows in the example above indicates falling or rising pitch contour. The position of the arrows indicates where the falling or rising begins. The four sentences have identical phonological segmental makeups but differ hugely in the meaning they convey without changing the phrasing of the utterance. (2a) is the most neutral way of describing a past event in English. (2b) indicate that the person who ate the cake is *he* instead of anyone else. (2c) is a question that aims to confirm if it is true that "he ate the cake." (2d) is trying to seek key information from the interlocutor that if it is *he* who ate the cake.

The current study focuses on prosodic phrasing in speech. Specifically, I aim to investigate the influence of boundaries marking the phrases on the dynamics of the first two formants.

1.2 Tautosyllabic vowel sequences (TVS)

Tautosyllabic vowel sequences are defined as those sequences that contain heterogeneous vowels and are not interrupted by any non-vocalic segments or any prosodic boundaries. This includes phonemic diphthongs (e.g., /aɪ/, /aʊ/, /oʊ/, etc.) in languages like English or Chinese, and non-phonemic vowel sequences co-occurring in single syllables (e.g., /ue/, /ae/, and /ao/, etc.) in Japanese and Estonian. In this study, they are both considered tautosyllabic vowel sequences. Traditionally, diphthongs are those TVS with only two vocalic targets. And diphthongs that start with a lower vowel and end with a higher vowel are called closing diphthongs or falling diphthongs, while the opposite transitional movement (from high target to low target) is called opening diphthongs or rising diphthongs. In this dissertation, I will investigate only those TVS with a closing or falling vocalic contour, i.e., a vowel sequence that consists of a lower initial vowel and a higher final vowel, e.g., /ai, au/. For notational convenience, TVS in English will be transcribed without using IPA symbols for lax vowels [ʊ, ɪ].

The phonotactic constraints on possible TVS in both English (Wells & Wells, 1982) and Chinese (Duanmu, 2007; Lee & Zee, 2003) suggest that the two languages have phonemic diphthongs in their inventories. For instance, the only possible TVS in American English usually include /aɪ, aʊ, oʊ, oɪ, eɪ/ (Wells & Wells, 1982), in Standard Chinese they are /ai, au, ou, ei/ and /ie, ia, ye, uo, ua/ (Lee & Zee, 2003). In both languages, all other vowel sequences must occur across certain prosodic boundaries (e.g., syllable, word, etc.). For instance, /ue/ cannot occur within one single syllable, either in English or Chinese. Contrarily, Japanese vowels are organized according to the mora structure such that, in principle, each vowel of different qualities makes up a mora in the syllable. All possible combinations of the five phonemic vowels (/i, e, a, o, u/) are attested if prosodic boundaries are not considered. However, within existing native or Sino-Japanese lexical items, only combinations of /a, i, u, o/ + /i, e, o/ are found (Labrune, 2012), even though Japanese has a wider range of possible TVS. There is an open empirical question as to how the dynamics of mora-based vowel sequences in Japanese compare with those found in English and Mandarin. Might speakers produce these sequences differently at different types of prosodic boundaries?

Phonetic research, especially those aimed at exploring theoretical issues, has almost exclusively focused on monophthongal productions, leaving TVS under-studied. The majority of existing studies on diphthong or vowel sequences within a syllable to date either focused on the difference between diphthongs and monophthongs/hiatus/semivowel-vowel sequences (Aguilar, 1999; Chitoran, 2002; Chitoran & Hualde, 2007; Gubian et al., 2015; Hualde & Prieto, 2002), or phonetic descriptions of diphthongs in a certain language if there is any (Elvin et al., 2016; Emerich, 2012; Mayr & Davies, 2011; Xia & Hu, 2016). Little has been done exploring how dynamic vowel targets are implemented in prosody. A rare example was found in Marin (2007), where they explored the acoustic property and perception of two diphthongal vowel sequences in Romanian, i.e., /ea/ and /oa/, and compared the gestural organization of diphthongs to that of vowel hiatus/glide vowel sequences from

the perspective of Articulatory Phonology. Marin (2007) found that cross-boundary unstressed vowel sequences (i.e., /e#a/ and /o#a/) can give rise to near-diphthongal sequences. However, although the stimuli in its perception study were simulated using TADA, a computational application based on Task-Dynamic Model and Articulatory Phonology, it did not analyze the dynamic production or acoustic data directly. This leaves room for future studies to look into the dynamics of TVS.

The target of TVS is dynamic, unlike monophthongs. Potentially it contains three subcomponents; the target of the initial vowel, the transition from the initial vowel to the final vowel, and the target of the final vowel. The literature has been focused on syllables with static vocalic targets. Previous studies on prosodic modulation on speech productions either looked into the production of consonantal gestures in syllables with simple structure (CV), leaving out vowel sequences (Beckman & Edwards, 1992; Berkovits, 1994; Byrd & Saltzman, 1998, 2003; Edwards et al., 1991) or included vowel sequence in the stimuli but did not analyze the production of the vocalic contour (Byrd & Choi, 2010; Krivokapić, 2007). Phonetic research on diphthongs also mainly focused on how it is produced and how to represent them properly based on the articulatory data without mentioning how that could interact with linguistic prosody (Hsieh, 2017; Hu, 2013; B. Kim et al., 2019). There is thus an empirical gap in the previous study of phonetics-prosody interface that vowel sequences were almost entirely left out from the discussion. It is, therefore interesting to see how the target of TVS is implemented in phonetics and how the variation in the implementation interacts with prosody.

Is there a way to allow us to study how variability is structured and motivated through modeling the movement? These characteristics of TVS posit a theoretical question to modeling speech production. In theoretical phonetic research on prosodic modulation on speech production, Byrd and Saltzman (2003) used a computational modeling method called "Task-Dynamic Model" proposed in E. L. Saltzman (1991), E. Saltzman and Byrd (2000), and E. L. Saltzman and Munhall (1989) based on the theoretical framework Ar-

ticulatory Phonology (Browman & Goldstein, 1989, 1995; Browman & Goldstein, 1986; Goldstein & Fowler, 2003, etc.)[1]. Articulatory phonology treats gestures as the units of phonological representation and lexical contrasts on the one hand and the units of actual speech production on the other. Gestures are abstract linguistic tasks or goals, such as creating a constriction using the lips and jaw. Gestures are considered abstract and discrete units but activated and coordinated with each other continuously in running speech. Task-Dynamic Model, therefore, aims to bridge this duality of discreteness and continuity by quantitatively modeling combinatorial articulatory gestures in the speech production system. Task-Dynamic Model models the goal-oriented gestures as critically damped point attractor systems defined in the space of vocal tract constriction tasks (E. L. Saltzman & Munhall, 1989). From this point of view, Byrd and Saltzman (2003) argued that prosodic strengthening due to the presence of boundaries is the consequence of the activation of π-gesture. π-gesture works as a clock slowing-down mechanism that is coactive with lexically specified constriction gestures but does not have specified dynamics of vocal tract constriction action of its own. Like other constriction gestures, the π-gesture has an inherent interval of activation, which wanes and waxes. Under the model of π-gesture, prosodic lengthening and strengthening observed in the past research were interpreted as the consequence of slowed-down gestures within the interval of π-gesture activation.

The Task-Dynamic Model assumes that static articulatory gestures attract articulators to the target region at distinct time points in the utterance. We model movement from one target to the next. For example, in Articulatory Phonology and Task-Dynamic Model, the production of the English word "pop /pɑp/" is analyzed as first activating the lip constriction gesture and tongue body lowering gesture simultaneously in the onset. Then, the lip constriction is released, and glottal constriction is activated to initiate vocal fold vibration to produce the vowel. At the end of the word, the glottis is open again, and the lip constriction gesture is activated to produce the coda [p]. However, diphthongs are inherently

[1]For a recent detailed overview, please see Byrd and Krivokapić (2021).

challenging to these models because they do not have single targets; by some accounts, the dynamics are the production targets. Previous studies have found that diphthongal TVS, like those in English, are not merely two simple vowels that co-occur. Gay (1968) found that for a diphthong /ai/, the final F2 value is significantly lowered in fast speech. An earlier EMA study on Ningbo Chinese (Hu, 2013) found that in closing (e.g., /ai, au/) and rising diphthongs (e.g., /ia, ie, ua, ue/), the dynamics of tongue positions differ. In the closing diphthongs, both the formant and the tongue position of the initial vowel target are more stable than those of the final vowel target. However, both vowel targets' formants and tongue positions are stable in the opening diphthongs. This result suggests that some vowel sequences inherently have complex targets, especially those closing ones. Therefore, the tongue movement for the vowel sequences may be the goal of gestural control. Studying how vowel sequences are produced in the interplay with prosodic boundaries thus can enrich our knowledge about speech production.

In this study, for my purpose of considering how the phonetic implementation of TVS is influenced by linguistic prosody, I define prosody as "informational structure in the language above the level of individual lexical entry" following Byrd and Krivokapić (2021). I will use Beckman (1996, 1997)'s conventional terms to describe prosodic structure.

1.3 Background

Many experimental studies have shown that fine-grained details in speech production reflect higher-order linguistic prosodic structure. The granularity of phonetic variation contributes to the grammar of a language rather than being a mere computational output from phonological encoding (c.f. "phonetic knowledge" in Kingston & Diehl, 1994). Cho (2016) notes that "it has now become a 'norm' that an understanding of the linguistic sound system can never be completed without referring to the phonetics–prosody interface - the interaction of sounds and sound patterns with prosodic structure in the grammatical system

of the language." Researchers have explored the issue since the late 1980s by referring to prosody's delimitative and culminative functions. The influence of delimitative function on speech production includes pre- or post-boundary strengthening/lengthening. Pre-boundary lengthening affects the temporal signature of the segments and lengthens the prosodic constituents or gestures that occur in the vicinity of a prosodic boundary (see overviews in Cho, 2015, 2016; Fletcher, 2010). I will mainly review the literature on pre-boundary lengthening/strengthening in the following.

1.3.1 Pre-boundary lengthening

Speech sounds and prosodic constituents tend to be longer in intonational phrase-final and utterance-final positions than when the same sound or constituent is uttered in non-final or phrase-medial positions. This durational phenomenon has been referred to as final lengthening (Berkovits, 1993; Cambier-Langeveld, 1997; Jang & Katsika, 2020; Shattuck-Hufnagel & Turk, 1998), prepausal lengthening (Klatt, 1973, 1975), domain-final lengthening, or preboundary lengthening (Cho et al., 2013; Gussenhoven & Rietveld, 1992) in the literature. I will use the term pre-boundary lengthening in this dissertation.

Pre-boundary lengthening refers to the lengthening effect in both the articulation and the acoustics of segmental or suprasegmental productions that occur at the boundary of prosodic constituents: speech sounds or sound patterns are produced with longer durations or gestures at final positions than at medial positions in a phrase. Pre-boundary lengthening is considered universal and has been repeatedly found in many languages. It has been observed in English (Beckman & Edwards, 1990; Byrd, 2000; Byrd & Riggs, 2008; Edwards et al., 1991; Klatt, 1975; Lehiste, 1976; Lehiste, 1973; Oller, 1973; Turk & Shattuck-Hufnagel, 2007), French (Fletcher, 1987; S.-A. Jun & Fougeron, 2002; Tabain, 2003), Japanese (Kaiki et al., 1992; Seo et al., 2019; Sugahara, 2005), Mandarin Chinese (Cao, 2004; Duanmu, 1996; Li, 2015; Liu & Li, 2003; Tseng, 2014; Y. Yang & Wang, 2002), German (Kohler, 1983; Kuzla et al., 2007), and Dutch (Cambier-Langeveld, 1997,

1999). In speech perception, pre-boundary lengthening is also crucial for the detection of upcoming boundaries in speech (J. Kim, 2020; Steffman, 2020; White, 2014; White et al., 2020). The scope of the final lengthening is not limited to the final word/syllable in many languages. The effect can seek forward to lengthen syllables before the syllable immediately precedes the boundary lengthening the penultimate syllables (Berkovits, 1994; Cambier-Langeveld, 1999; Kohler, 1983; Kuzla et al., 2007; Turk & Shattuck-Hufnagel, 2007).

However, this lengthening effect, although universal, exhibits structured variance in the phonetic detail when another aspect of linguistic structure also needs to be encoded in the local phonetic context, alternating the extent of the lengthening effect. First, pre-boundary lengthening interacts with post-lexical accentuation signaled by nuclear pitch accents in languages like English and Germany. Cho et al. (2013) and S. Kim et al. (2017) found that pre-boundary lengthening has a greater magnitude in English when the target words are less prominent (de-accented). Pre-boundary lengthening also interacts with the lexical pitch accent. In Japanese, Seo et al. (2019) found that when a disyllabic word in Japanese has a pitch accent on the initial syllable, it suppresses the effect of pre-boundary lengthening: the final syllable of the initially accented words lengthens less than those of the unaccented words probably to preserve the prominence of the accented syllable. In addition, pre-boundary lengthening interacts with the phonemic vowel quantity contrast in some languages, such as Finnish (Nakai et al., 2009; Nakai et al., 2012). It was found that the lengthening of long vowels is suppressed when they are adjacent to another long vowel (Nakai et al., 2009), and vowels were lengthened less when they were next to a syllable containing a long vowel than when they were next to a syllable containing a short vowel (Nakai et al., 2012) in Finnish. A similar trend was reported for Creek, too (Johnson & Martin, 2001). These results suggest that prosodic lengthening may be universal but implemented in a way specific to the phonological system of a language.

The majority of the studies above interpret the different effect sizes in the pre-boundary

lengthening to be related to the "prosodic strength", i.e., the higher the position in the prosodic hierarchy, the greater the lengthening effect is (Beckman & Edwards, 1994; Byrd et al., 2006; Byrd & Saltzman, 2003; Fougeron & Keating, 1997; Ladd & Campbell, 1991; Tabain, 2003; Tabain & Perrier, 2005, 2007; Turk & Shattuck-Hufnagel, 2000, 2007). However, results also showed that the size of the lengthening effect might not strictly correlate with the position of the boundary in the prosodic structure. This means that while a lengthening effect is confirmed between the lowest and highest boundaries in the structure (e.g., no boundary vs. utterance-final), there might be little to no difference between boundaries closer in the hierarchy of prosodic structure. For example, Byrd and Saltzman (1998) found that the three speakers could only distinguish two to three phrasing levels despite five in the experimental design (no boundary, word, list, vocative, utterance). Klatt (1975) and Umeda (1975) found little sign in a connected speech that the syllables in the sentence-final syllables are longer than any other phrase-final syllables. Similar results were presented in Wightman et al. (1992) that the duration of final syllables at the right edge of full intonational phrases, intonation phrases followed by pauses, and sentences did not vary significantly. This further suggests that the prosodic lengthening as a consequence of the activation of π-gesture is not categorical but rather probabilistic. The boundaries of larger prosodic constituent at higher positions in the prosodic structure is more likely to induce a longer duration.

Proposals have been made over the past few decades to account for the pre-boundary lengthening/strengthening effect. Earlier hypotheses include B. Lindblom (1968) and B. E. F. Lindblom (1975) that utterance duration is a reflection of a generative constraint that depends on the size of the chunk of speech. Each piece of speech is planned with a phrase buffer. Hence a pre-boundary lengthening is the consequence of the natural deceleration in speech articulation towards the end of a chunk. Other later studies support the similar view that the pre-boundary lengthening is due to supralaryngeal declination throughout an utterance (Berkovits, 1994; Tabain, 2003; Vayra & Fowler, 1992). Berkovits (1993,

1994) examined that the degree of lengthening increases progressively from the beginning to the end of phrase final disyllabic words, suggesting a gradual temporal declension. This trend was also observed in Japanese data reported in Seo et al. (2019). This, however, indicates that pre-boundary lengthening is merely due to physiological or biological constraints directly imposed on the phonetic sound patterns. As discussed above, the fact that the pre-boundary effect could interact with such a wide range of linguistic phenomena cross-linguistically indicates that speakers can have control over the degree of lengthening and find a way to solve the conflicts between the tendency to lengthen the final constituents before a boundary and the tendency to preserve some other linguistic contrast (vowel quantity or intensity contrast) or functions (prominence, accentuation, pitch accent).

However, to date, very few studies have tried directly comparing languages with different basic prosodic organizational units: mora and syllable, although research has been carried out separately on both types of languages. Language. In Shepherd (2008) on Japanese pre-boundary lengthening, the author found that the lengthening effect is confined to the last mora before the prosodic boundary. This suggests that in Japanese, the interval of π-gesture activation is rather narrow and does not reach too far away from the boundary. If Japanese pre-boundary lengthening has its scope limited to the final mora, then the effect of lengthening should be less than those that could be found in Chinese and English since mora is not a unit in the two languages. Syllable being the direct target of lengthening, the effect in Chinese and English should be larger than that in Japanese.

1.3.2 Strategies of strengthening

The pre-boundary effect does not only include the durational effect but can modulate the amplitude of the articulatory gestures or the acoustic quality of the vowels. Articulatory studies have shown that the pre-boundary lengthening is associated with spatially larger, longer, and slower closing gestures (Byrd et al., 2006; Byrd & Saltzman, 1998; Cho, 2006; de Jong et al., 1993; Fougeron & Keating, 1997; Harrington et al., 2015; Tabain & Perrier,

2005, 2007).

Two hypotheses have been made in the previous studies concerning the strengthening effect: SONORITY EXPANSION and HYPERARTICULATION. SONORITY EXPANSION hypothesis claims that the intrinsic sonority of the vowel is enhanced in prosodically more prominent positions to boost the syntagmatic vowel-consonant contrast. In speech production, this is usually correlated with wider jaw opening or acoustically with a higher F1 value, as the F1 value is inversely correlated with tongue height and mouth opening. In a corpus phonetics study, Mo et al. (2009) investigated the correlation between perceptual prominence and the first two formants of vowels in English. They found that perceptually prominent vowels tend to have higher F1 values regardless of the vowel height, indicating that sonority expansion is used in the phonetic encoding of prominence.

On the other hand, the HYPERARTICULATION hypothesis, based on B. Lindblom (1990)'s H&H Model, claims that prominence involves enhancing the contrastive features of the speech sounds. Since hyperarticulated sounds/gestures are in contrast to their hypoarticulated version, hyperarticulation is paradigmatic and largely correlates with lingual movements. This implies that if a sound is [+front], after hyperarticulation, its tongue position will be even more front (Cho, 2005; Harrington et al., 2000) or F2 will be raised. In acoustics, it is reflected as high vowels tend to be hyperarticulated with lower F1, front vowels with higher F2, and back vowels with lower F2. In a word, hyperarticulated vowels tend to be more peripheral as compared to their hypoarticulated counterparts. An illustration of the difference between SONORITY EXPANSION and HYPERARTICULATION is shown in figure 1.3.

In an articulatory investigation of English /a, i/ uttered at prosodic boundaries, Cho (2005) concluded that the pre-boundary strengthening enhances phonological features and positional strength that may license phonological contrast, making final vowels more peripheral than their counterparts in medial positions. However, mixed results were reported in the acoustic domain that lengthened vowels might not be more peripheral. Johnson

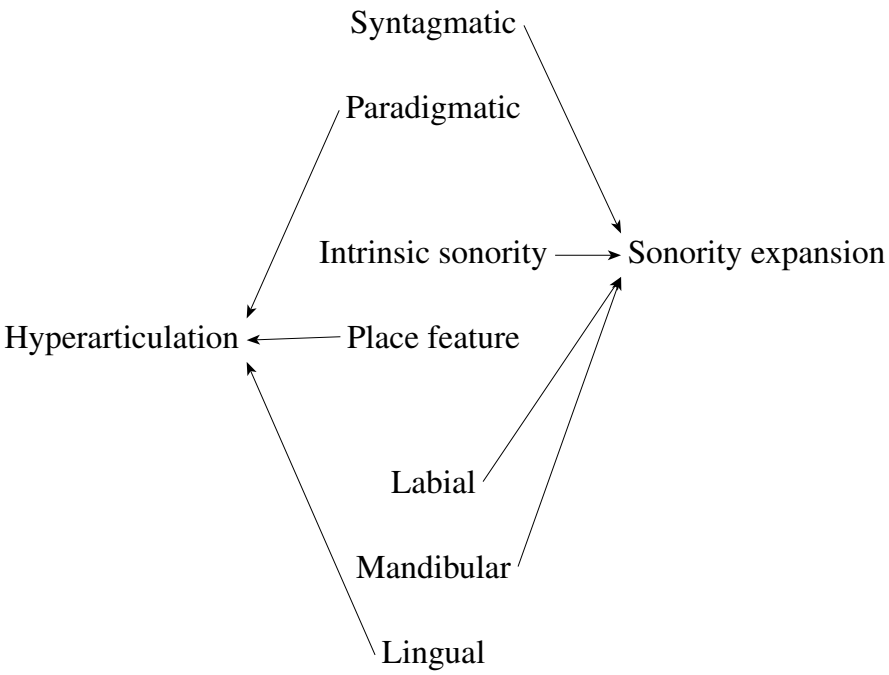

Figure 1.3: An illustration of the difference between *sonority expansion* and *hyperarticulation*.

and Martin (2001) showed that in Creek, lengthened vowels are centralized despite being lengthened at the final positions. Nord (1986) found similar trends in Swedish that duration alone does not determine vowel quality. In the current study, I will compare the overall F1 and F2 curves over time to examine if TVS exhibits sonority expansion or hyperarticulation as the strategy of prosodic strengthening. The exciting question is whether and how the initial or final target in a TVS is strengthened.

Following the sonority expansion hypothesis, if the SONORITY EXPANSION hypothesis accounts for the prosodic modulation on TVS as well, it should be found that the F1 that corresponds to how open the mouth increases in the entire course of the TVS when it occurs at higher prosodic boundaries such as intonational phrase-final positions. On the other hand, several predictions can be made should the HYUPERARTICULATION hypothesis account for the strategies used in preboundary-lengthening. For the TVS to be hyperarticulated, different portions might be hyperarticulated differently as TVS inherently involves different vowel qualities. For instance, for falling diphthongs /ai, au/ to be hyperarticulated,

/a-/ as the onset should be hyperarticulated to have a lower tongue position and more open mouth. This feature should correlate with a higher F1 since F1 is inversely correlated with tongue height and mouth openness. On the other hand, in the offset of the TVS, the high vowels should be produced higher with a lower F1. The final /-i/ should be more front with a higher F2, whereas the final /-u/ should be more back with a lower F2.

In an articulatory study on pre-boundary lengthening, Another issue concerning the strategies of prosodic modulation in the pre-boundary gestures is which articulatory parameter is used in speech production. Based on Task-Dynamic Model, Cho (2002, 2006) proposed four articulatory parameters that could account for possible articulatory strategies used by speakers to adjust their speech production according to the local context.

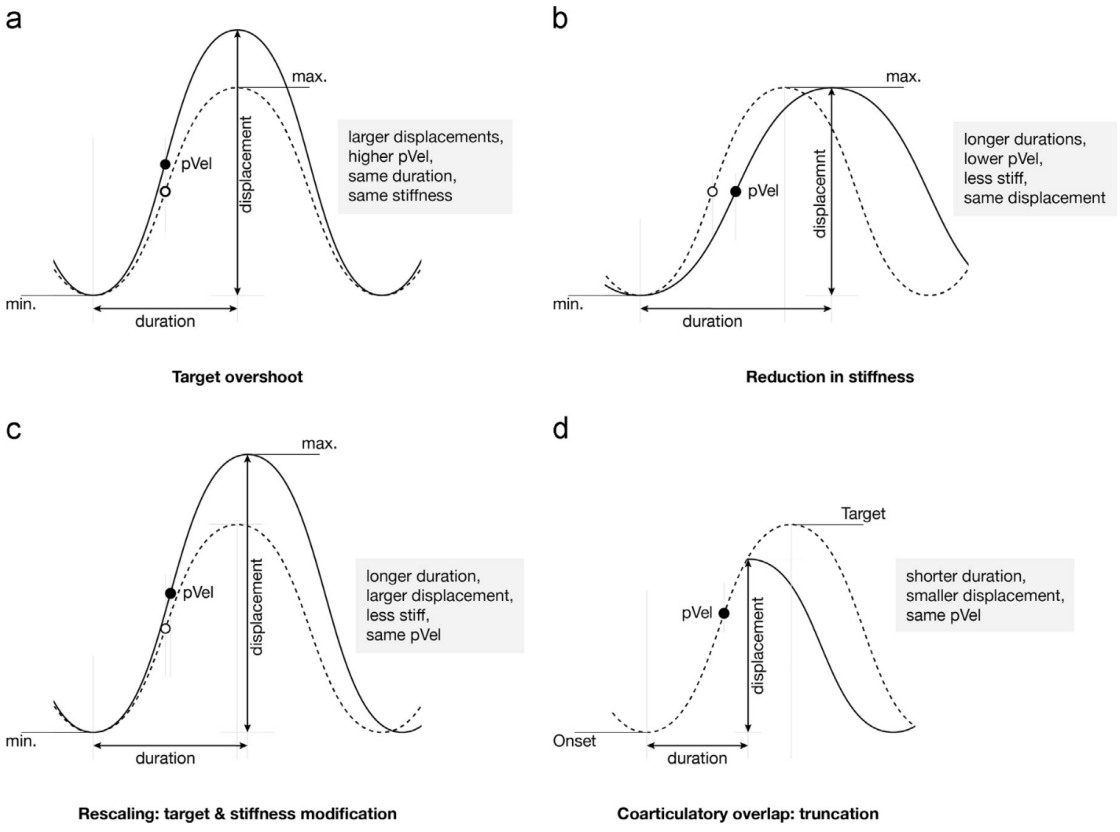

Figure 1.4: Possible articulatory strategies that might be used by speakers (figure taken from Mücke & Grice, 2014).

(3) Articulatory strategies

a. *Target overshoot* involves increasing the distance that articulators travel to produce a certain gesture while the duration of the movement remains the same. The velocity is thus higher, while the stiffness does not change.

b. *Reduction in stiffness* slows the entire articulatory movement but does not necessarily increase the distance the articulators need to travel.

c. *Rescaling* is to increase the "size" of the articulatory gesture entirely with a larger distance, lower stiffness, and longer movement duration, but the peak velocity does not vary.

d. *Truncation* is due to coarticulation that the gesture is cut-off at the end for the articulators to start producing the following gesture earlier. It is observed as shorter duration, smaller distance but the same peak velocity.

As is shown in figure 5.1, The four parameters are based on four kinematic measures: the maximum displacement, the total duration of movement, the peak velocity of the movement, and a derived measure of stiffness: the ratio of peak velocity and maximum displacement.

Although these parameters were proposed based on Articulatory Phonology and Task-Dynamic Model to account for articulatory movement in space, it has shown possible extension to acoustic analysis as well, the idea of measuring the dynamic change in the acoustic domain is not new to phoneticians. The PENTA model proposed by Prom-On et al. (2014) and Xu et al. (2016) is a vital application that integrates communicative functions on the one hand and articulatory mechanisms on the other. Like Articulatory Phonology, the PENTA model also tries to bridge higher-order low-dimensional communicative functions such as focus or interrogative questions and continuous moving trajectories in speech production such as F0 movement. Xu and Prom-on (2019) exhibit a successful analysis of kinematic measurements of formant data. In this study, I will apply this approach of kinematic analysis and take kinematic measures (displacement (the total length of the trajectory in vowel space), movement duration, peak velocity (the maximal rate of change of the trajectory),

and stiffness (the maximal rate of change and trajectory length ratio)) to analyze the dynamics of trajectory movement in the F1/F2 vowel space. Therefore I assume that the parameters illustrated above should also apply to the acoustic domain in analyzing acoustic data such as the formants, especially when the segment involves a transition of formants during its production.

Previous studies have shown that pre-boundary modulation frequently uses *stiffness reduction* as the strategy, which means that gestures are produced with longer time, slower movement, and smaller stiffness but not larger displacement (Beckman & Edwards, 1992, 1994; Cho, 2002; Edwards et al., 1991). However, bigger displacement at the boundaries was not unfound in previous studies (Cho, 2002; Krivokapić, 2007). According to articulatory studies (Byrd & Saltzman, 2003; Edwards et al., 1991), this is a natural consequence of the activation of π-gesture at the boundaries. The effect of the π-gesture is to slow the time course of the constriction gestures that are active simultaneously as the π-gesture. A slower course of activation leads to gestures being temporally longer (Krivokapić, 2020).

If the idea of π-gesture on prosodic modulation can be applied to the acoustic domain, we should expect that the primary strategies used in modulating the trajectory movement in vowel space would be *stiffness reduction* as well. The trajectory movement in the vowel space should exhibit smaller peak velocity, longer duration, and smaller stiffness at higher prosodic boundaries.

1.4 Research questions and hypotheses

In the current study, I aim to compare the contours of TVS under different prosodic conditions in Chinese, English, and Japanese. The first research question is

(4) **Research question 1**: are the TVSs monophonemic or biphonemic

(5) If the TVS is monophonemic, formants in the onset of the TVS should be less affected by those in the offset of the TVS.

(6) If the TVS is biphonemic, formants in the onset and offset should be both affected by the prosody.

If a TVS is monophonemic, then the presence of offglide is more salient than the offset vocalic target. Therefore, as long as formant gliding is present, reaching the offset target may not be crucial to the production of TVS. Thus, the offset should exhibit more variation than the onset in a TVS in various prosodic contexts. However, if a TVS is biphonemic, both the targets in the onset and offset are salient to prosodic modulation. Therefore, the prosodic modulation should both target formants in onset and offset.

Two major research questions can be raised concerning the pattern of pre-boundary lengthening.

(7) **Research question 2**: Is the pre-boundary lengthening sensitive to the presence of mora in Japanese compared to the lack of mora as a prosodic unit in Chinese and English?

(8) **Research question 3**: Are the effects of the pre-boundary lengthening on monophthongs and TVS the same?

The hypotheses concerning the two questions can be formulated as below.

(9) If the pre-boundary lengthening is sensitive to the basic unit of prosodic organization, with the presence of mora, Japanese TVS should exhibit less lengthening than those of Chinese and English.

(10) If the effect of pre-boundary lengthening holds the same on monophthongs and TVS, the lengthening effect should be the same across different nuclei types.

The research questions about the strategy of pre-boundary prosodic modulation are:

(11) **Research question 4**: Does pre-boundary lengthening modulate the TVS by SONORITY EXPANSION or HYPERARTICULATION?

(12) **Research question 5**: Which strategy was used in the prosodic modulation of trajectory movement in the vowel space by the TVS?

The hypotheses hence are formulated as follows:

(13) Regarding sonority expansion and hyperarticulation

a. If the lengthening involves sonority expansion, then the overall F1 should be higher in TVS at higher boundaries in the prosodic hierarchy.

b. If the lengthening involves hyperarticulation, then the F2 of the final vowels should be higher if it ends with a high front vowel or lower if the TVS ends with a high back vowel.

(14) Regarding strategies

a. If *target overshoot* is utilized by the speakers, then the total trajectory length of the TVS moved in the F1/F2 vowel space should be larger, and the maximal rate of change of the trajectory should be higher. At the same time, the duration and stiffness should remain the same.

b. If the boundary-related prosodic strengthening involves *stiffness reduction*, then the stiffness and trajectory length stay constant. At the same time, the duration is longer, and the maximal rate of change is slower.

c. If the target of the TVS is *rescaled* at higher boundaries, longer trajectory length, higher maximal rate of change, and longer duration should be observed while stiffness stays the same.

d. If the production of TVS is *truncated* at lower boundaries, the trajectory length should be smaller, the duration shorter, but the stiffness and maximal rate of change should not change.

Chapter 2

Methods

2.1 Experiment

2.1.1 Stimuli

The target TVSs include /ai, au, ou/ for Chinese and English and /ai, ae, au/ for Japanese. All TVSs are embedded in word-final syllables with a stop consonant as the syllable onset. The onsets of the target syllables vary across three places of articulations: labial, alveolar, and velar. Table 2.1 shows the target words in Chinese.

Table 2.1: Target words in Chinese.

	/ai/	/au/	/ou/
Bilabial	/li˩˧pai˥˩/	/li˩˧pau˥˩/	/li˩˧p^{h}ou˥/
Alveolar	/li˩˧tai˥˩/	/li˩˧tau˥˩/	/li˩˧tou˥˩/
Velar	/li˩˧kai˥˩/	/li˩˧kau˥˩/	/li˩˧kou˥˩/

In the Chinese stimuli, all target syllables carry a high falling tone (the fourth tone: ˥˩) except for /p^{h}ou/ which carries a ˥ tone because the high falling tone does not occur with the syllable. The Chinese target words are all constructed as disyllabic person names, with the first syllable held constant being [˩˧li]. The advantage of constructing all target words as

person names is to ensure that the target syllable, i.e., the second syllable in the disyllabic words, is the prominent syllable in a disyllabic root. This is because although Standard Chinese does not have obligatory lexical stress as other typical stress languages (Hyman, 2006, 2009), researchers did find evidence that even in disyllabic words that do not have toneless syllables (i.e., syllables with the so-called "neutral tone"), syllables do differ in the word-level prominence in disyllabic words (L. Yang, 2011). /li˧/ was chosen as the first syllable because it is one of the most common family names among Chinese. It should be clear enough to the participants that the target word is a person's name without causing too much confusion during the experiment. All Chinese target words are embedded in the three carrier sentences in table 2.2.

Table 2.2: Chinese stimuli.

Position	Carrier sentence	Meaning
word-final	zhè gè bāo nǐ qù qǐng [TARGET WORD] ná yí xià.	You can go ask [TARGET WORD] to carry this bag for a moment.
list-final	jīn tiān lái kāi huì de rén yǒu [TARGET WORD], né gǔ hé gāo yǔ.	Today, people who came to the meeting include [TARGET WORD], Ni Gu, and Gao Yu
IP-final	zhè wèi jiù shì nǐ yào zhǎo de [TARGET WORD]. nǐmen ké yǐ hǎohao liáoliao.	This is [TARGET WORD] who you've been looking for. You guys can talk.

English stimuli are exhibited in table 2.3, there is a gap in the target words in that there is no target word for the syllable /tau/ or /dau/ since no common monosyllabic words consist of this syllable alone.

Table 2.3: Target words in English.

	/ai/	/au/	/ou/
Bilabial	pie	bow (v.)	bow (n.)
Alveolar	tie	-	toe
Velar	guy	cow	Go

Japanese target words are shown in table 2.4.

Table 2.4: Target words in Japanese.

	/ai/	/ae/	/au/
Bilabial	/pai/ 'pie'	/dekibae/ 'workmanship'	/bau/ 'Bau (a pet name)'
Alveolar	/kitai/ 'expectation'	/kotae/ 'answer'	/tau/ 'τ'
Velar	/rikai/ 'understanding'	/okikae/ 'replacement'	/gau/ 'Gau (a place name)'

While constructing Japanese target words, several compromises were made. First, the Japanese TVS does not include the sequence of /ou/ as it is phonologically not available in the Japanese sound system. Historically, /ou/ monophthongized in modern Japanese to /oː/, therefore another possible TVS /ae/ was added to Japanese stimuli. Secondly, several multisyllabic words were used to create more common words that should be well known to Japanese speakers.

Since both English and Japanese target words vary much more than those for Chinese stimuli, to make the sentence as natural as possible, the carrier sentences were designed for each target word to fit the lexical meanings best. Please see the full English and Japanese stimuli sets in the appendix A.

Words with monophthongal nuclei and the same set of onsets were also created for each language to compare the lengthening in monophthongs and TVS. The monophthongs used in English are /i, ɔ/. /ɔ/ was chosen over /ɑ/ because /ɑ/ is a lax vowel in English and does not occur in word-final open syllables. The monophthongs used in Chinese and Japanese are /i, a, u/. In Chinese, since the syllable /ki/ is not allowed in phonology, it is not included as a filler. The monophthongal nuclei in Japanese were created as bimoraic long vowels to make maximally comparable the comparison of durations between monophthongs and TVS.

In addition, 20 unrelated words embedded in the same carrier sentences were added as fillers for each language. The words with monophthongal nuclei are shown in table 2.5,

2.6, and 2.7.

Table 2.5: Chinese fillers.

	/i/	/a/	/u/
Bilabial	/li˩˥pi˥˩/	/li˩˥pa˥˩	/li˩˥pu˥˩/
Alveolar	/li˩˥ti˥˩/	/li˩˥ta˥˩/	/li˩˥tu˥˩/
Velar	-	/li˩˥k^{h}a˥/	/li˩˥ku˥˩/

Table 2.6: English fillers.

	/i/	/ɔ/
Bilabial	bee	paw
Alveolar	tea	-
Velar	key	cough

Table 2.7: Japanese fillers.

	/i/	/a/	/u/
Bilabial	/kabiː/ (a pet name)	/baː/ ‘bar’	/jahuː/ ‘Yahoo’
Alveolar	/tɕiː/ (a pet name)	/koNpjutaː/ ‘computer’	/sotsuː/ ‘communication’
Velar	/kiː/ ‘key’	/kaː/ ‘car’	/kakuː/ ‘fiction’

Target words are embedded in three different prosodic contexts: word-final, list-final, and intonational phrase-final (IP-final) to elicit three different prosodic boundaries. These are the three most frequently elicited prosodic contexts in the literature (Byrd, 2000; Byrd & Saltzman, 1998; Cho, 2002; Krivokapić, 2007; Tabain, 2003; Tabain & Perrier, 2005, to name a few) and are considered to correspond to three different types of boundaries: word boundary, boundary intermediate or minor phrases (ip boundary), and boundary of intonational or major phrases (IP boundary). In the list-final positions, all target words are positioned as the first member of a list that has three members in total. The IP-final positions

are not utterance-final but followed by another intonational phrase. In the Japanese stimuli, however, there are monomoraic phrase-final morphemes following the target words. They are the nominative particle /-ga/ following target words in word-final positions and the copula /-da/ following words in IP-final positions.

A total of 147 trials (42 target trials for English, 51 for Chinese, and 54 for Japanese) were created. Twenty unrelated words of each language were embedded in the same carrier sentences as random fillers. All trials were randomized upon being presented to the participants. Each trial was repeated seven times for each speaker.

2.1.2 Speakers

Fourteen speakers were recruited for the English experiment (8 males and 6 females). English and Chinese speakers are all in their 20s and are college or graduate students from the University at Buffalo, SUNY. All English speakers are from the areas near Buffalo and Rochester and have lived there for their entire lifetime. The English speakers all spoke Western New York (WNY) variety of American English. WNY English, like other English varieties of Inland North cities, is characterized by several phonological characteristics that target its monophthongs. For instance, the vowel chain shift of [ɛə] ← /æ/ ← /ɑ/ ← /ɔ/ ← /ʌ/ ← /ɛ/ (Labov et al., 2008). However, WNY English does not differ much from General American English in its diphthong inventory nor the pronunciation of diphthongs.

Twelve Chinese speakers (7 males and 5 females) were recruited from Buffalo. They are all Chinese international students in their 20s studying at the University at Buffalo, SUNY. Ten of the Chinese speakers are from Shanghai, with one speaker from Shanxi province and another one from the city of Tianjin. All Chinese speakers are native speakers of Standard Mandarin Chinese. The mandarin spoken in Shanghai is slightly different from Standard Chinese in the merger of coda nasals: [-n] and [ŋ]. While the local Sinitic language Shanghai Wu does not possess diphthongs (Y. Chen & Gussenhoven, 2015), the Mandarin variety spoken in Shanghai does not differ from Standard Chinese concerning

diphthongs. It should not interfere with the production of vowel sequences in the experiment by Chinese speakers.

Twelve Japanese speakers (2 males and 10 females) are Japanese nationals currently living in Buffalo. Japanese speakers' ages range from the late 20s to 50s. Some of the speakers were from the Kansai area. They speak the Kansai dialect of Japanese and the Tokyo accent as the standard variety. They were instructed not to speak the Kansai dialect during the experiment.

None of the speakers reported any hearing or speaking impairment or difficulties. Each speaker was paid 10 dollars for their participation in the experiment.

2.1.3 Recording procedure

The recordings were collected in a sound-treated sound booth of the Department of Linguistics, University at Buffalo, using a Denon DN-700R professional recorder and Røde NT1A Shock Mount condenser microphone. The microphone was mounted at the left front of the speaker at a distance of 10cm from the mouth.

Actions were taken to protect the participant from Covid-19. First of all, only one person was able to enter the sound booth each time. Between any two uses, the sound booth stayed open and was ventilated for at least 20 minutes. All surfaces in the room were disinfected using Metrex CavaCide all-purpose cleaner before and after each use.

The stimuli were presented to the speakers on a computer screen in the free software OpenSesame. Chinese stimuli were presented in Chinese characters and Japanese inspirations in Japanese writing conventions (Kana and Chinese Characters when necessary). All trials were randomized upon presentation. Before the experiment, each speaker was given a training session to get used to the experiment's setting and task. The entire experiment includes eight blocks, with the first as the practice block. Participants were allowed to take a break between each block. Since only one person was allowed to sit in the sound booth, speakers were instructed to press keys on the keyboard to proceed after each trial

by themselves. A total of 147 trials (42 target trials for English, 51 for Chinese, and 54 for Japanese) were created.

Speakers produced seven repetitions of each trial. A total of 6070 English trials ((42 targets + 20 fillers) $\times$ 7 repetitions $\times$ 14 speakers), 5964 Chinese trials ((51 targets + 20 fillers) $\times$ 7 repetitions $\times$ 12 speakers) and 6216 Japanese trials ((54 targets + 20 fillers) $\times$ 7 repetitions $\times$ 12 speakers) were collected.

2.2 Data processing and measurements

The author labeled all recordings in PRAAT. The boundaries of the target TVS and the syllable were labeled. The boundaries were labeled by visually examining the presence and absence of F1. The beginning boundary of the syllable is where the F1 from the previous vowel disappears, and the beginning boundary of the TVS is where F1 reappears in the spectrogram. The end boundary of the syllable and rime is identified by the disappearance of the visible F1 again in the spectrogram. Figure 2.1 gives an example of data labeling.

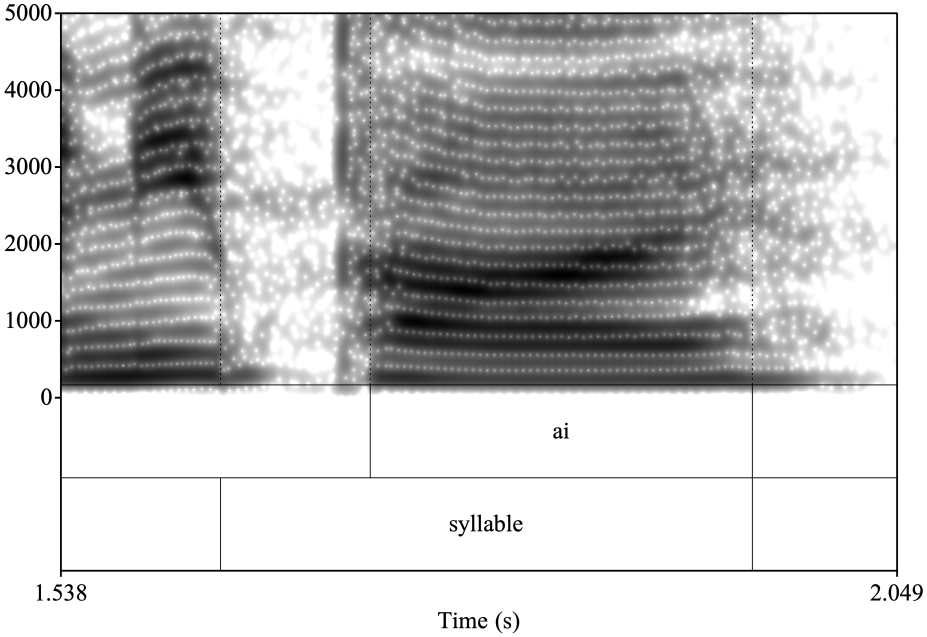

Figure 2.1: An example of data labeling in PRAAT.

The duration of the entire syllable and the rime of the syllable was measured alongside the F1 and F2 of the target TVS. F1 and F2 were obtained from 30 equidistant intervals in the TVS using a PRAAT script written by the author. The PRAAT script uses the algorithm by Burg to track formants based on LPC coefficients as described in Childers (1978). The time of the midpoints of each interval was also measured. The procedure of formant extraction used the seeding method of formant extraction as proposed in W.-R. Chen et al. (2019). The seeding method extracts more accurate formant values by setting different reference values for different target vocalic segments and genders. Since my data consists of TVS with formant excursions, the seeding method was further developed by setting different formant references in the first, medial, and final 33% of the TVS. The reference values were taken from previous acoustic studies on vowels in the three languages (for Chinese (Howie, 1976), English (J. Hillenbrand et al., 1995), and Japanese (Hara, 2015)). The reference values for the medial 33% are the mean of the F1/F2 reference values in the initial and final 33% of the TVS. The ceiling frequency range also varies for genders, with 4500Hz for male and 5500Hz for female speakers. The number of formants to track was 4 for males and 5 for females. The reference values of the first two formants for each language are shown in table 2.8.

The extracted formant data were further cleaned to avoid extreme outliers due to the algorithmic errors of formant tracking. To achieve this, the data were first grouped by Language, Gender, and TVS. In each group, the F1 and F2 values that are two standard deviations from the mean were removed for each time point. The formant data were then normalized to correct the noise in the measured formants caused by different vocal tract lengths of speakers using the ΔF method described in Johnson (2020).

(15) formant normalization:

a. $\Delta F = \frac{1}{mn}\sum_j^m \sum_i^n [\frac{F_{ij}}{i-0.5}]$ where i = formant number, j = token number

b. $F'_{ij} = \frac{F_{ij}}{\Delta F}$

The author visually checked the normalized formants distribution to ensure that the

TVS	Gender	Initial		Medial		Final	
		F1	F2	F1	F2	F1	F2
ai	M	670	1400	520	1900	550	2300
au	M	680	940	520	880	500	830
ou	M	500	990	421	800	334	690
ai	F	710	1400	650	2200	600	2400
au	F	710	1600	630	970	600	900
ou	F	600	1100	500	900	300	750

(a) Chinese

TVS	Gender	Initial		Medial		Final	
		F1	F2	F1	F2	F1	F2
ai	M	760	1100	650	1800	420	2300
au	M	700	1000	550	890	420	800
ou	M	550	980	410	800	390	750
ai	F	485	919	375	1314	265	1709
au	F	493	1132	396	955	299	775
ou	F	311	915	260	803	208	691

(b) English

TVS	Gender	Initial		Medial		Final	
		F1	F2	F1	F2	F1	F2
ai	M	775	1163	519	1713	263	2263
au	M	775	1163	569	1231.5	363	1300
ae	M	775	1163	625.5	1626	476	2089
ai	F	485	919	405	1822	325	2725
au	F	493	1132	434	1403.5	375	1675
ae	F	888	1363	685.5	1840	483	2317

(c) Japanese

Table 2.8: Reference formant values for formant tracking.

extreme outliers were removed as much as possible. Normalized F2 values greater than 1.9 after the 20th time point of /au, au/, normalized F1 values greater than 1.0, and normalized F2 values over 2.6 after the 20th datapoint of /ai/ were removed from English data. F2 values that are over 2.0 and F1 values that are over 0.8 of /ou/ were removed from English data. The results of each step of data cleaning (filtering, normalization, final cleaning) are

shown below in figure 2.2 (green dots are F2s and red dots are F1s).

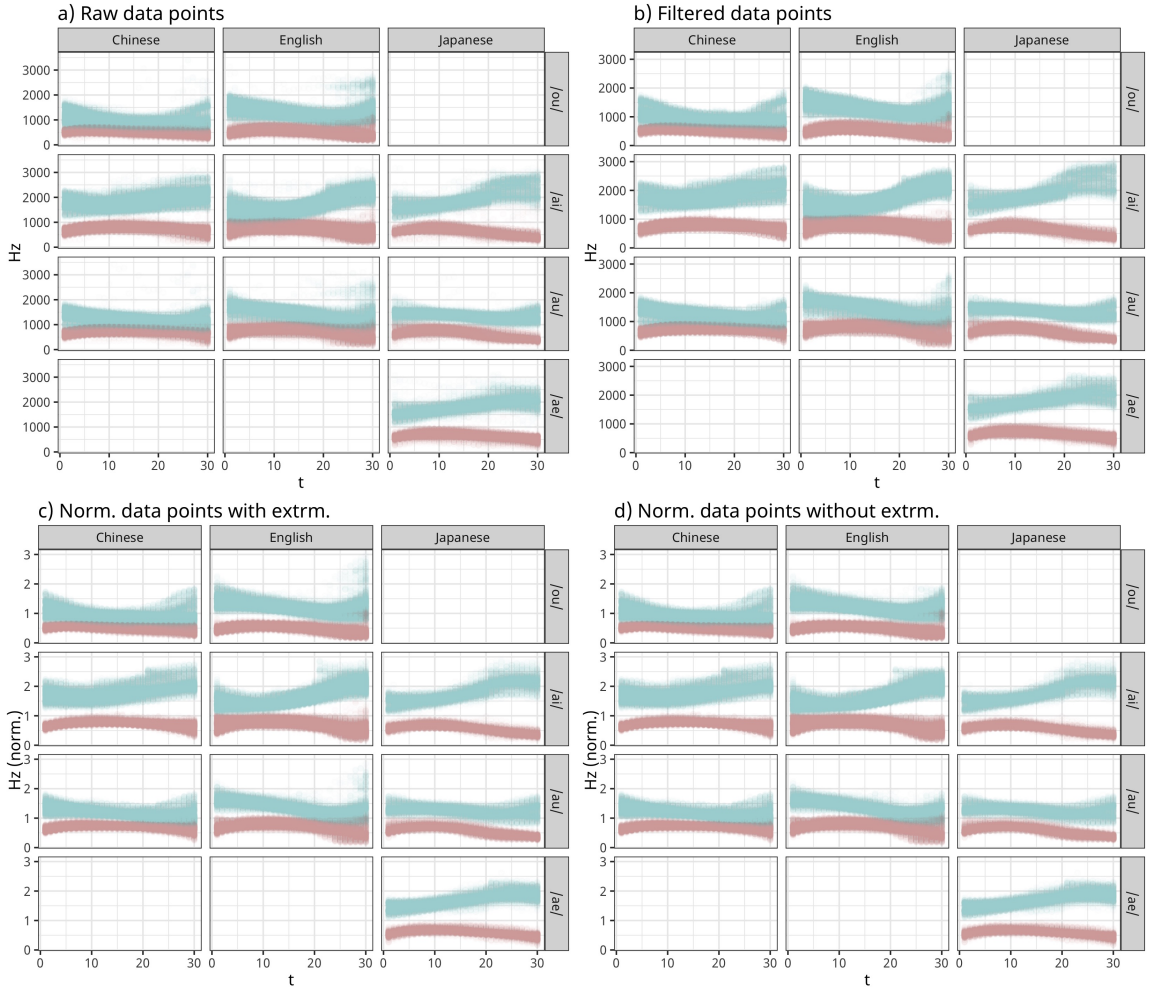

Figure 2.2: The four steps of data cleaning.

Formant data were further smoothed by third-degree polynomial regression to obtain kinematic measurements (total trajectory length (TTL), maximal rate of change of trajectory (max ROC), trajectory duration, and maximal rate of change and total trajectory length ratio (RL ratio)) of each token. This is to remove formant jump-up and -downs. This step is necessary since if there are abrupt formant jumps in the data, then the curves are not smooth for some of the tokens. This would result in high leverage values in formant kinematic measures such as TTL and max ROC. Spline smoothing was specifically not chosen here because although spline smooth is powerful in modeling non-linear relations (Perperoglou et al., 2019), with so few data points in each token, even when the number of knots

(the turning points in the distribution of the data) is as low as 3, it tends to cause overfitting. The degree of polynomial regression was set to three because, in a few studies on the acoustics of vowel formants, a third-degree cubic polynomial regression was found to best capture the curves of formant movements (Flego & Forrest, 2021; Van Der Harst et al., 2014).

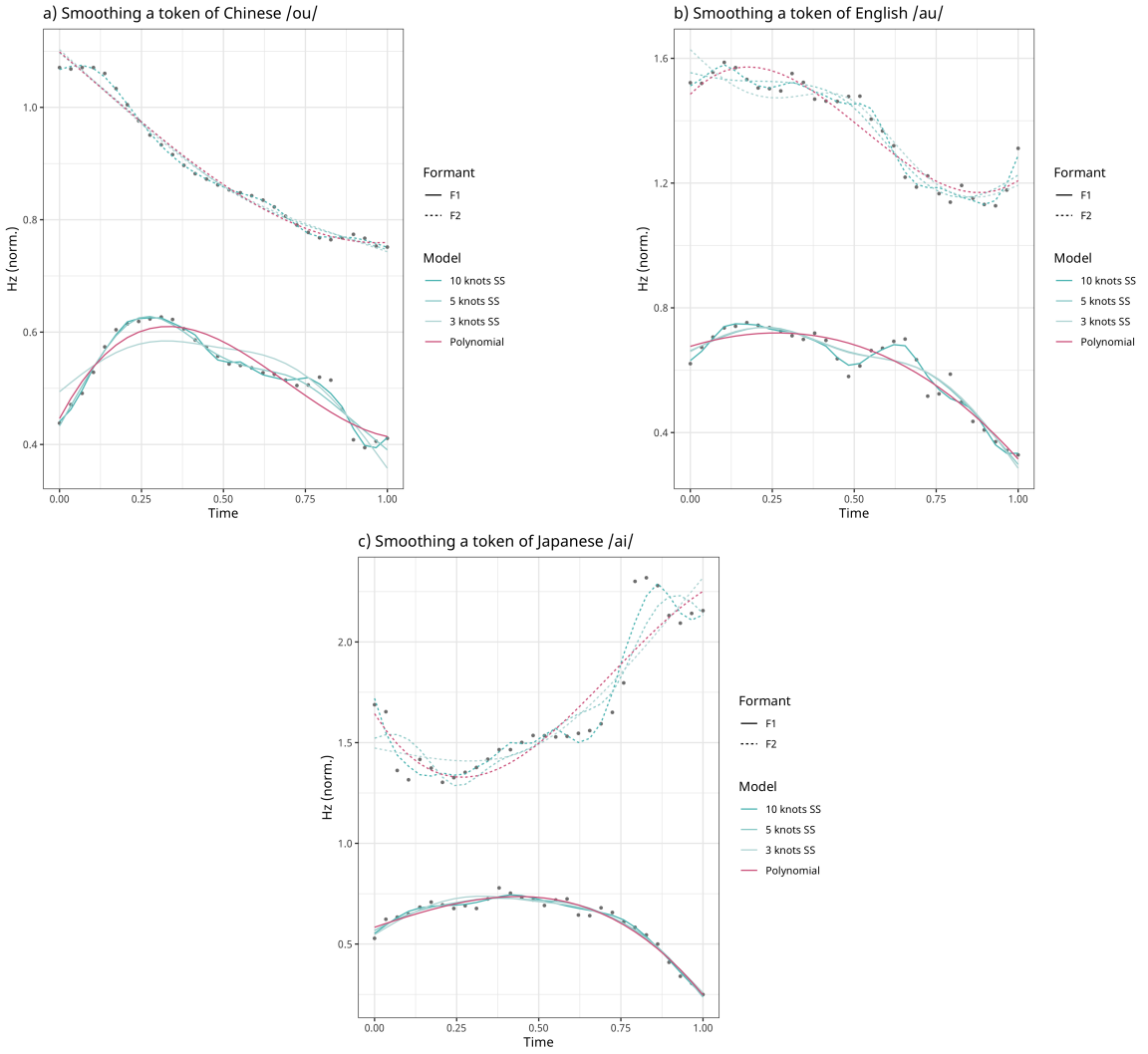

Figure 2.3: A Comparison of spline smoothing and polynomial smoothing (third order).

As is easily seen from the three examples of smoothing in figure 2.3, spline smoothing tends to 'follow' the raw data points and is too wiggly even with only 3 knots. Therefore, formant data smoothing was done with third-degree polynomial fitting. The result and a comparison to raw data are shown in figure 2.4.

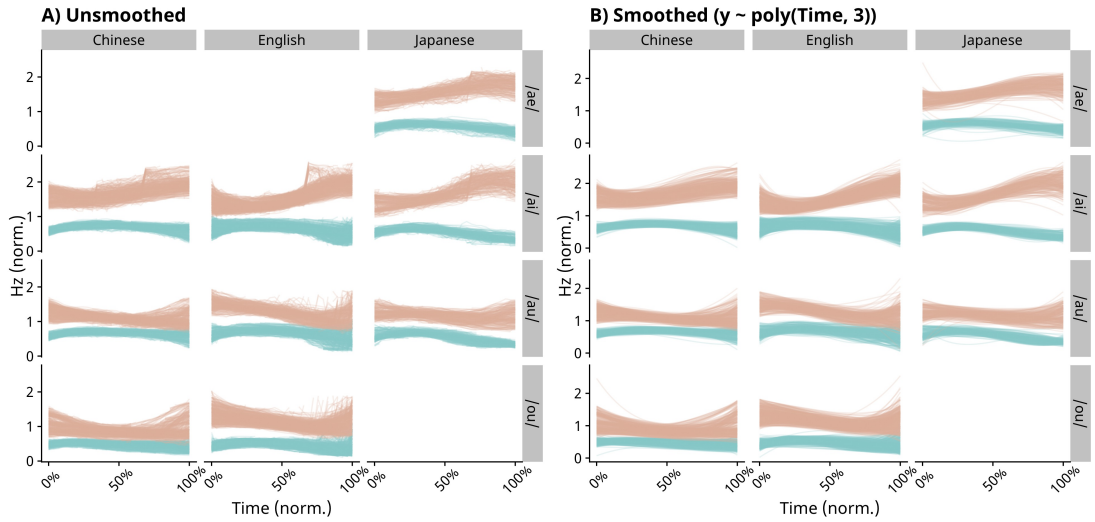

Figure 2.4: A comparison of unsmoothed and smoothed formant values.

Before obtaining the kinematic measurements, I define an interval within each segment of TVS as the "proper interval of TVS movement". Let's look at the trajectories of the TVS in the F1F2 vowel space in the three languages. Note that at the beginning and the end of TVS, C-V and V-C coarticulation have a considerable impact on the formant trajectories (see figure 2.5). Especially for TVSs that end with a high back vowel, due to the influence of the following coronal consonants across the boundary, the F2 is raised after reaching the minimum.

The true interval of TVS movement is defined in different ways depending on if the TVS ends with a high front vowel or a high back vowel. The true interval of TVS movement for /au, ou/ begins from the F1 maxima and ends with the F2 minima. However, the true interval for /ai, ae/starts from F1 maximum too but ends with F2 maxima. This is the interval where the tongue moves from the lowest to the frontest or most backward position. Compared to previous studies that arbitrarily cut off the first and last 20% of the vowel (Akpanglo-Nartey, 2020; Fox & Jacewicz, 2009; J. Hillenbrand et al., 1995; J. M. Hillenbrand et al., 2001), this approach is more accurate than and faithful in determining the actual contour of TVS movement. After dropping data from outside the proper interval of TVS, the data look like below in figure 2.6.

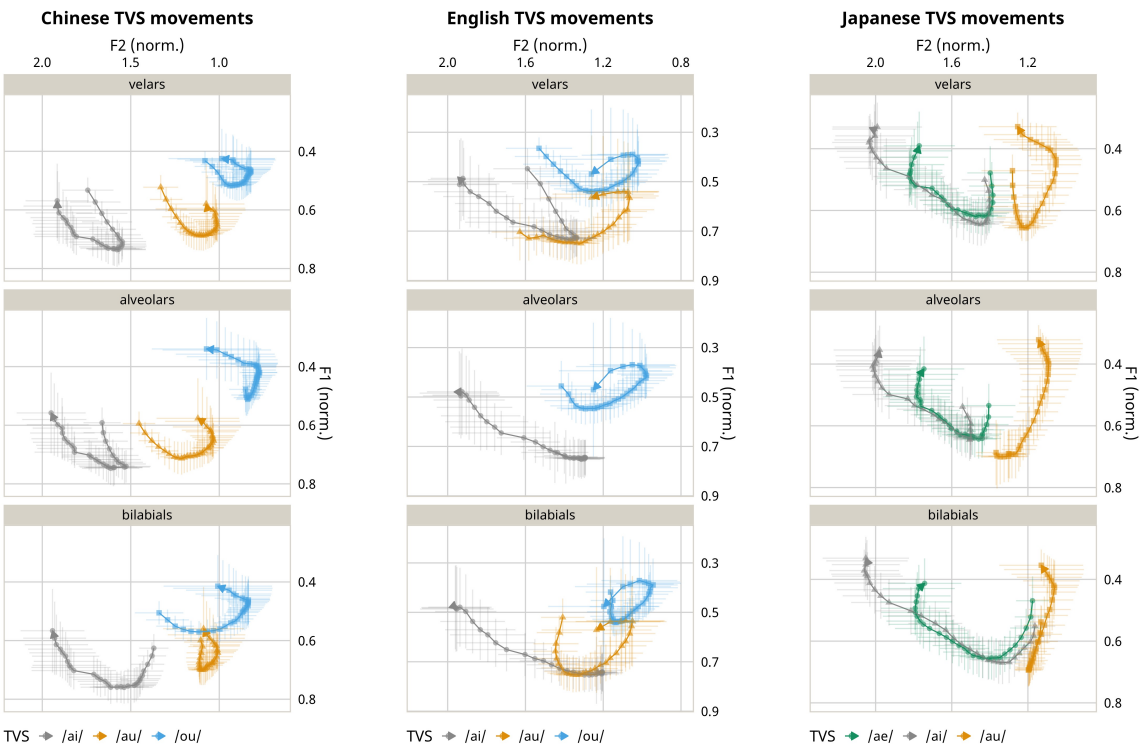

Figure 2.5: Average trajectories of TVS movement in the vowel space following different onsets (based on unsmoothed normalized data). The grey, yellow, and blue trajectories are /ai, au, ou/ respectively, and the green trajectories are /ae/.

Based on the smoothed values from polynomial regression and time information of the proper interval of TVS movement, vowel space displacement (Disp), interval duration (ID), and velocity of movement in the vowel space between each adjacent time point were calculated first.

(16) a. $\text{Displacement}_i = \sqrt{(F1_{i+1} - F1_i)^2 + (F2_{i+1} - F2_i)^2}$

b. $\text{ID}_i = \text{t}_{i+1} - \text{t}_i$

c. $\text{Vel}_i = \text{Disp}_i / \text{ID}_i$

Then the total true displacement, total true duration and peak velocity (pVel) were calculated as follows:

(17) a. Total disp = $\sum_1^n Displacement_i$

b. Total dur = $\sum_1^n ID_i$

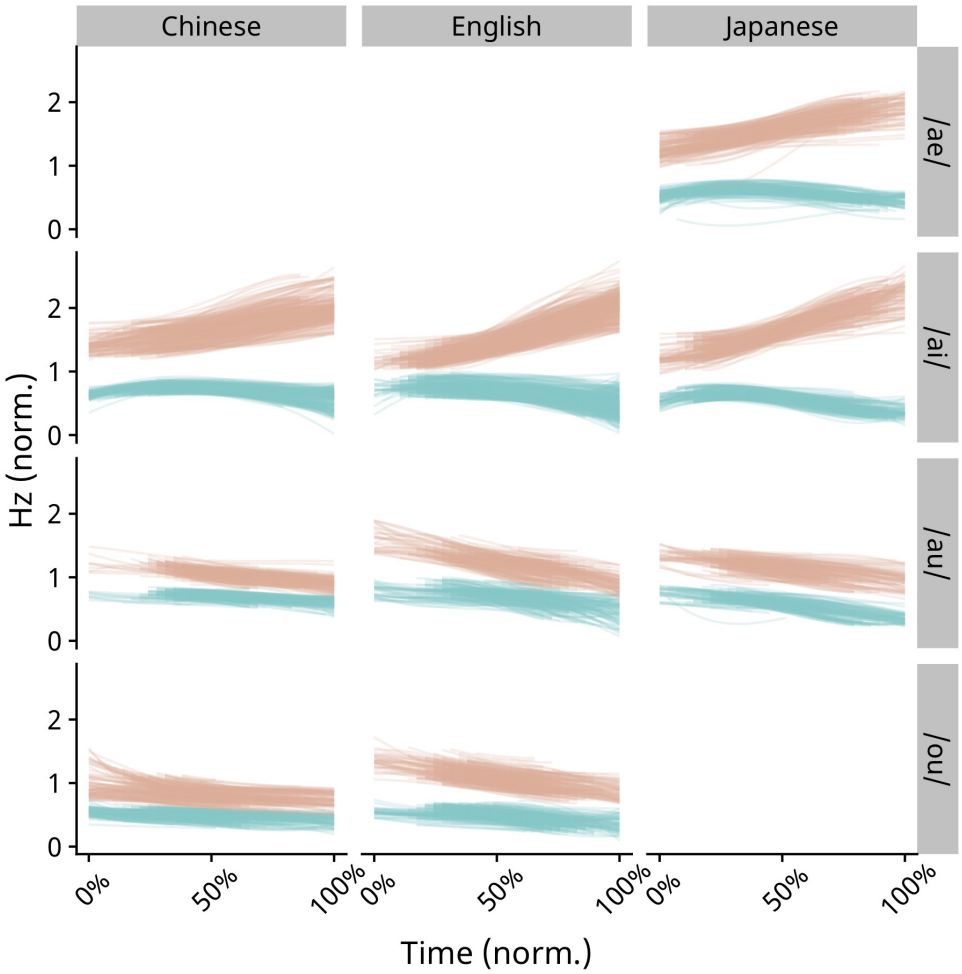

Figure 2.6: Formant curves after getting the proper intervals of TVS movement.

c. pVel = $\max(V_i)$

Further, the indication of stiffness of the trajectory movement: the ratio of peak velocity and total displacement was also calculated.

(18) Stiffness = $\frac{Totaldisp}{Totaldur}$

2.3 Data analysis

The overall contours of the F1 and F2 of each TVS in the three languages will be analyzed by using GAMM (generalized additive mixed-effect model) based on the `mgcv` package

(Wood & Wood, 2015) in R (Team, 2022). Analyzing non-linear time-series phonetic data using GAMM has had successful applications in articulatory data analysis (Tomaschek et al., 2018), intra-oral pressure analysis (Brandt & Simpson, 2021), f0 analysis (Sun & Shih, 2021), and formant analysis (Hualde et al., 2021).

The kinematic measures will be submitted to LMEM (linear mixed-effect model) to examine if they vary across the prosodic conditions in different languages (for /ai, au/ in Chinese, English, and Japanese, /ou/ in Chinese and English, and /ae/ in Japanese).

2.4 Procedure of GAM anaylsis

GAM models were built separately for normalized F1 and F2 values of each TVS in the three languages. 18 GAM models were built for the nine target TVS (/ai, au, ou/ in Chinese and English, /ai, ae, au/ in Japanese). In each model, prosodic `Position` were included as the fixed effect along with the factor smooths as the by-`Speaker` random effect.

I used treatment contrast coding for the fixed effect of `Position`, with "word-final" as the reference level. Autocorrelation in time-series measurement was also corrected using the residuals at Lag 1.

2.4.1 Tensor product interaction between `Time` and `Block`

Additional models also with the tensor product smooth interactions between `Block` (repetitions of individual trials) and `Time` were created as well. These models essentially model non-linear interactions between `Block` and `Time` by allowing the coefficients underlying the smooth for time to vary non-linearly depending on the value of `Block` (ranging from 1 to 7). The interaction was added to examine if the speakers 'ease' their production as they got used to the experiment format. Speakers may have eased their articulation as they record more trials during the recording. However, the results below show that neither the by-`Block` smoothing nor the result of tensor product interaction introduced much variation

to the data as shown in figure 2.7, 2.8, and 2.9.

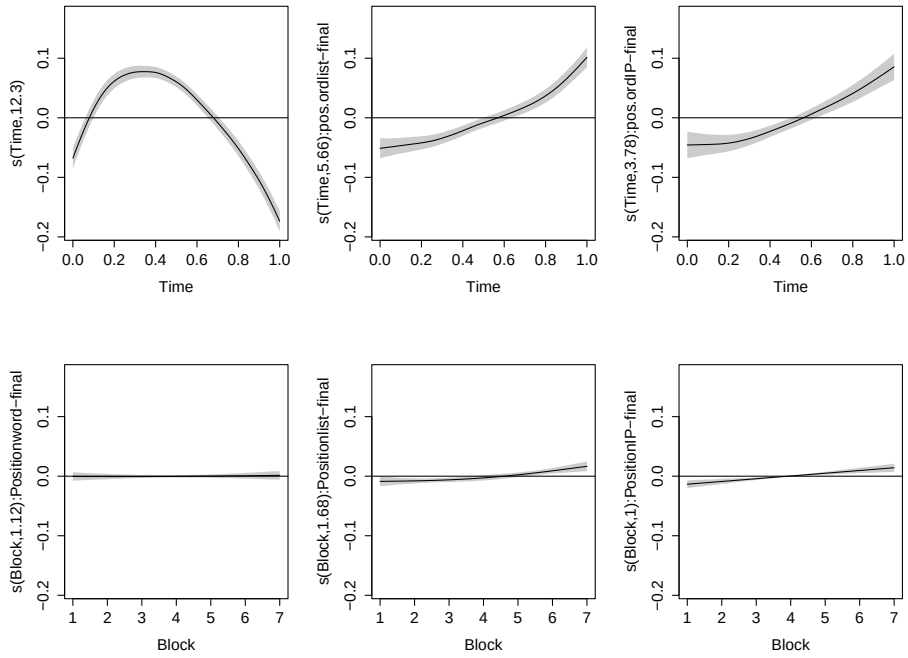

Figure 2.7: Visualizing the partial effect of prosodic `Position` and `Block`. Note that the top-middle and top-right figures show the smoothing of the differences between list-final and IP-final positions.

In figure 2.7, the smoothed F1 of Chinese /ai/ is displaced. The top three figures show the F1 change across time points, while the lower three show the F1 change across experimental blocks. It is shown that the F1 changes over time: it increases in the first 30% of the vowel and then decreases until the end. The two figures on the top panel's right show the difference between the two other prosodic contexts and word-final position. It is first lower in the first half and then higher in the second half of Chinese /ai/. The bottom left figure shows that the F1 value of Chinese /ai/ does not change across blocks in word-final positions, while the bottom middle and right show that in the list- and IP-final positions. F1 of Chinese /ai/ does increase a little, and the change is very linear: F1 increases to a minor degree across blocks in list-final and IP-final positions.

Figure 2.8 show the three-dimensional visualization of tensor product interaction. In the upper three figures, the x-axis represents time points and the y-axis experimental blocks.

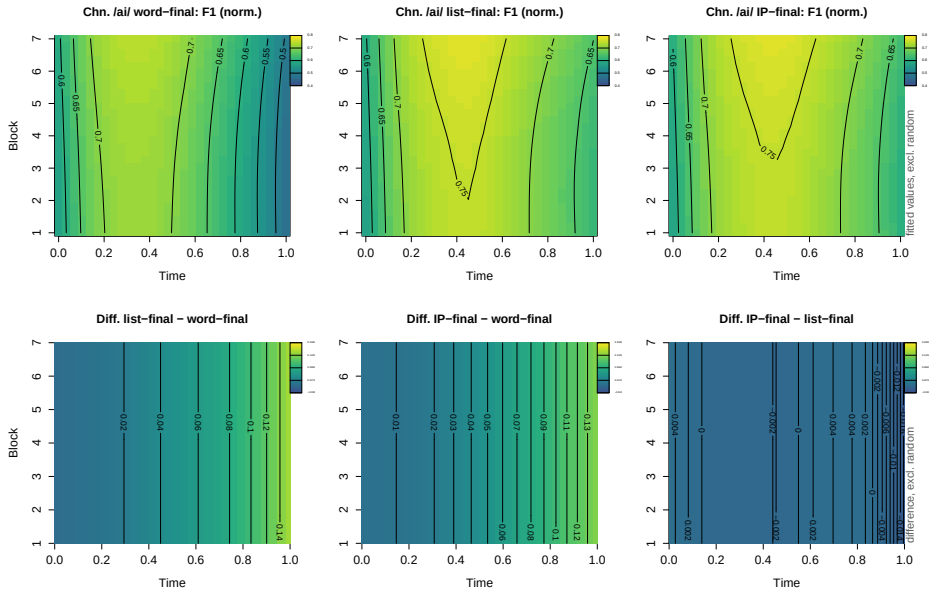

Figure 2.8: Contour plots visualizing the non-linear interactions of between `Time` and `Block` for the prosodic `Positions` on the top row, and their differences (bottom row).

Note that the F1 value in all three prosodic conditions increases slightly. The difference smooths in the lower three figures show that it is either not affected by the block at all or is only slightly affected.

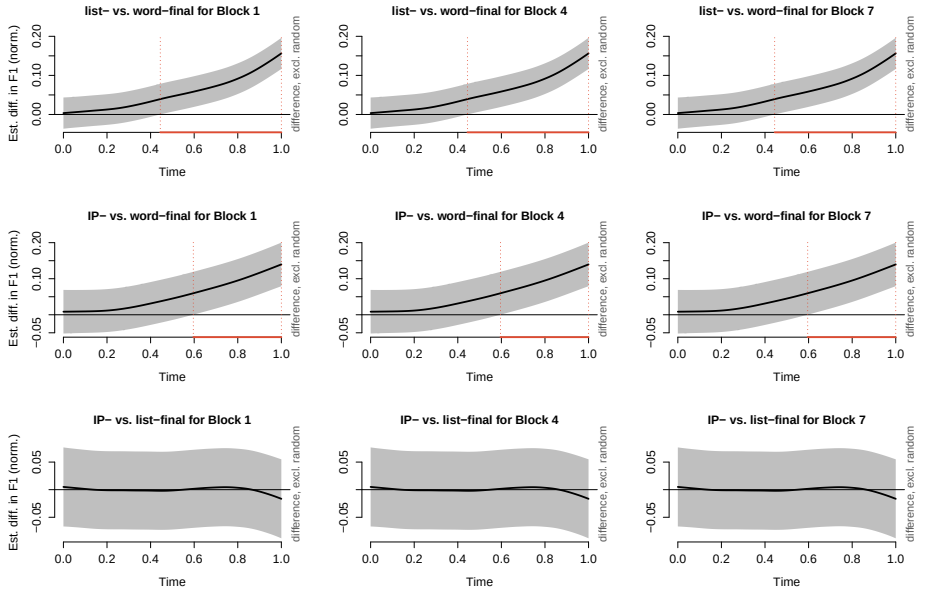

Figure 2.9: The corresponding non-linear pattern over time for `Block` 1 (left-panel), 4 (middle panel), and 7 (right panel).

Figure 2.9 showed the F1 value across time in block 1 (left panel), 4 (middle panel), and 7 (right panel). The difference between list-final and word-final, IP-final and word-final, and IP-final and list-final positions are shown in the upper, middle, and bottom panels. Note that the difference in F1 among the three blocks is almost indiscernible.

In sum, both figure 2.8 and 2.9 showed that the tensor product interaction does not seem to affect either the estimated smooth of dependent value in the three prosodic positions or the difference between them. The tensor product interaction, however, was kept in all the models as the models that included the tensor product interaction term did improve the r-squared value of and deviance explained by the models. This is illustrated below in table 2.9. Therefore in the final model specification, the tensor product interactions were dropped.

2.4.2 Specification of random effect

Furthermore, the variation in F1 and F2 contours due to the inter-speaker variability is also considered. This is accomplished by including factor smooths in the model to represent random effects. These factor smooths are a non-linear alternative to random intercepts and random slopes in a linear mixed-effects regression model. They are essential for considering the structural variability associated with individual speakers.

In an overview of GAM analysis, Wieling (2018) and Sóskuthy (2021) summarized that there are multiple ways of adding random effect smooths to a GAM model:

(19) a. `md1` (Random intercept only):

```
s(random, bs="re")
```

b. `md2` (Random intercept + slopes):

```
s(random, bs="re") + s(random, by=fixed, bs="re")
```

c. `md3` (Non-linear random effect):

```
s(random, fixed, bs="re") + s(x, random, bs="fs", m=1)
```

Table 2.9: The comparison of R-squared values and deviance explained between models with and without a tensor product interaction.

			Without tensor product interaction		With tensor product interaction	
			R-squared	Deviance explained	R-squared	Deviance explained
Chinese	/ai/	F1	0.62	0.504	0.63	0.514
		F2	0.696	0.599	0.699	0.602
	/au/	F1	0.549	0.436	0.561	0.448
		F2	0.585	0.502	0.586	0.503
	/ou/	F1	0.499	0.459	0.498	0.457
		F2	0.57	0.431	0.571	0.432
English	/ai/	F1	0.571	0.501	0.575	0.503
		F2	0.834	0.704	0.835	0.706
	/au/	F1	0.545	0.499	0.549	0.501
		F2	0.739	0.668	0.741	0.671
	/ou/	F1	0.518	0.52	0.535	0.524
		F2	0.618	0.548	0.622	0.551
Japanese	/ai/	F1	0.766	0.668	0.766	0.667
		F2	0.796	0.73	0.797	0.731
	/au/	F1	0.771	0.689	0.771	0.689
		F2	0.62	0.541	0.62	0.542
	/ae/	F1	0.574	0.488	0.575	0.491
		F2	0.749	0.664	0.755	0.67

d. `md4` (Individual variability over time ("Item-by-Effect" in Sóskuthy (2021))):

```
s(x, random, by=fixed, bs="fs", m=1)
```

e. `md5` (Random reference/difference smooths (ordered factor smoothing)):

```
s(x, random, bs="fs", m=1) + s(x, random, by=fixed.ord, bs="fs",
m=1)
```

In the schema of model specifications above, `x` represents the variable the response variable (dependent variable) is smoothed over. `random` represents the random effect that one wants to include in the model. Usually they are variables such as `Speaker` or `Trial`. `fixed` represents the fixed effects in a model that one wants to model linearly as a paramet-

ric term in a GAM model. In this study, my random effect is `Speaker`, and my fixed effect prosodic `Position`. The `x` variable that will be smoothed against is the 30 `Time` points from which the F1 and F2 values were extracted.

In 2.4.2, the first model `md1` 'Random intercept only', the random effect only includes the random intercept per speaker, but the non-linear patterns from different speakers are assumed to be the same, and there is no influence on the non-linear pattern from the fixed effect(s). In `md2`, `Fixed` is added to the second smooth term as the by-speaker random slope. However, this does not account for the by-speaker non-linear variability across time. This is considered in `md3` in that the by-speaker variability in the fixed linear effect is modeled along with the random non-linear pattern. `md4` specifies what was called "Item-by-Effect" random smooth in Sóskuthy (2021) by adding the by-speaker random effect on both linear and non-linear patterns. This is a "so-called factor smooth (hence the `"fs"` basis) which models a (potentially) non-linear difference over time concerning the general time pattern for each of the speakers" (Wieling, 2018). The last random effect specification was recommended in Sóskuthy (2021). The model still models the factorial smooth of the random effect. But what is different is that given that this random structure separates each level of the hierarchical variable into a reference smooth (`s(x, random, bs="fs", m=1)`) and a difference smooth (`s(x, random, by=fixed.ord, bs="fs", m=1`), it should be able to produce shrinkage for both and thereby produce well-calibrated type I error rates and power.

The first three ways of specifying random effects in a GAM model are not recommended since they fail to capture the linear and non-linear variability in the data and are prone to committing Type I error. However, the fourth model was also overly conservative in that its power is low according to Sóskuthy (2021). This is because "models do not recognize the connection between contours representing different levels of the fixed effect variable within the same speaker: they treat them as if they were completely independent. In other words, there is no shrinkage on the random within-speaker differences between

groups. This results in a situation where some of the systematic variations that should be captured by the fixed effect terms in the model (i.e., the parametric term and the smooth difference term) are, instead, hijacked by the random effect term" (Sóskuthy, 2021, p.15). Model comparison through `compareML()` function and `AIC()` function both revealed that the five ways of random effect specification have progressively lower AIC value as the structure becomes more and more complicated (AIC: `md5` $<$ `md4` $<$ `md3` $<$ `md2` $<$ `md1`). The estimated smooths and differences of the five models are shown in figure 2.10. In this research, all the random effects in GAM models were specified using the "Random reference/difference smooths" approach.

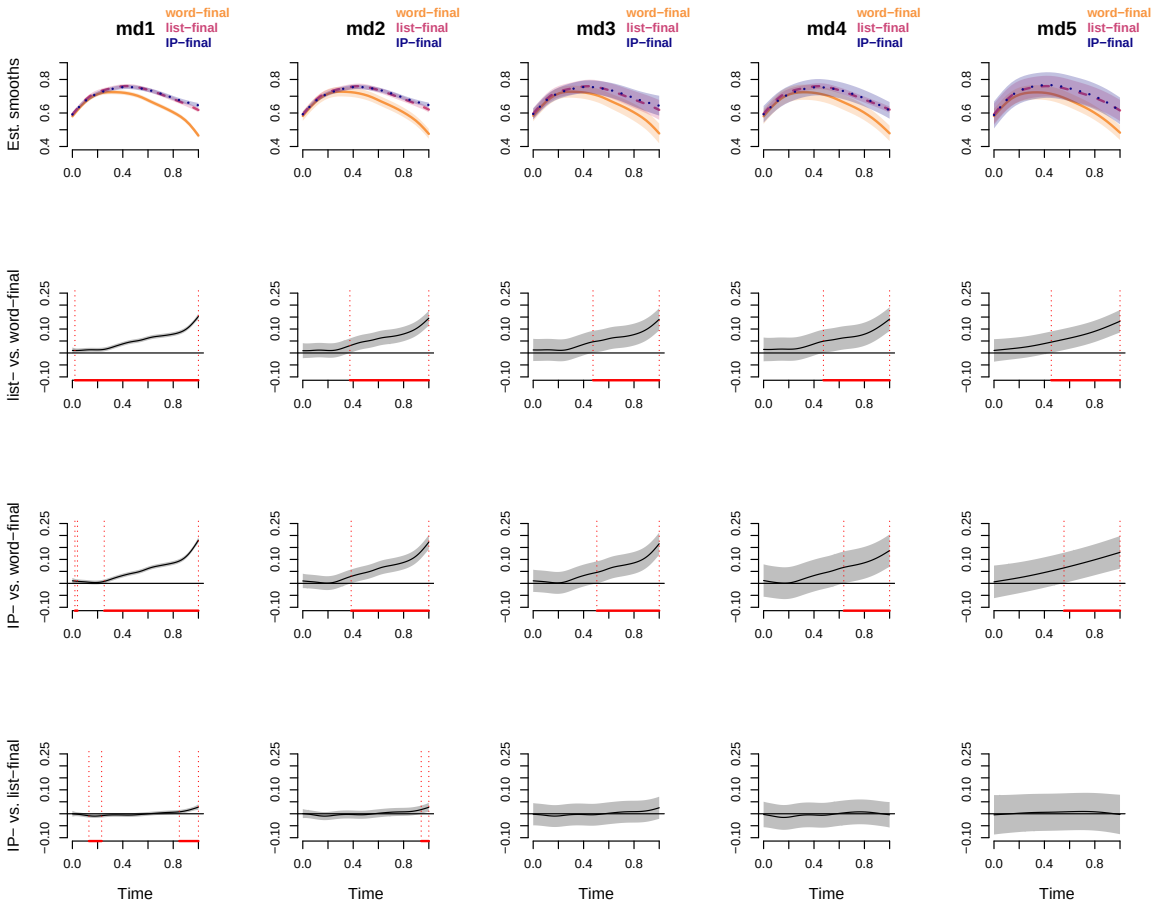

Figure 2.10: An illustration of different structures of random effects. The data is the normalized F1 data of Chinese /ai/. The graphs in the first row show the summed effects of prosodic position on the curves of F1 of Chinese /ai/. The following three rows show the difference among the three prosodic positions: "word-final", "list-final", and "IP-final". All the smooths in the figure were plotted with the random effect excluded.

The adjusted R-squared values and the deviance explained by each model are listed in table 2.10.

Table 2.10: Adjusted R squared value and deviance explained by each model.

	R-sq. (adjusted)	Deviance explained
md1	0.496	47.8%
md2	0.54	50.5%
md3	0.587	53.5%
md4	0.64	57.2%
md5	0.64	57.2%

Since the fifth model was both conservative and accounts for more variation in the data, all the models in the current study were guilt using the fifth approach: Random reference/difference smooths.

2.4.3 AR1 correction of autocorrelation in time-series data

The residuals of a generalized additive mixed model (GAM) need to be independent. However, when analyzing time-series data like formant contours which are smooth and slow-moving, the residuals will generally be autocorrelated. This means that the residuals at a time `t` will be correlated with the residuals at time `t+1`. If this autocorrelation is not corrected, the p-values will be anti-conservative. I used the `bam()` function of the `mgcv` package to take into account the autocorrelation of residuals, thus allowing me to make a more reliable assessment of the model fit and the associated p-values. The uncorrected and corrected autocorrelations in the model are shown below in figure 2.11.

2.4.4 Scaled-t distribution

After model fitting, I conducted model criticism using the `gam.check()` function of the `mgcv` package. Figure 2.12 shows the diagnostic graphs produced by this function.

The upper four graphs of figure 2.12 reveal that the residuals show a problematic non-normal distribution, which almost certainly will affect the estimates and p-values of the

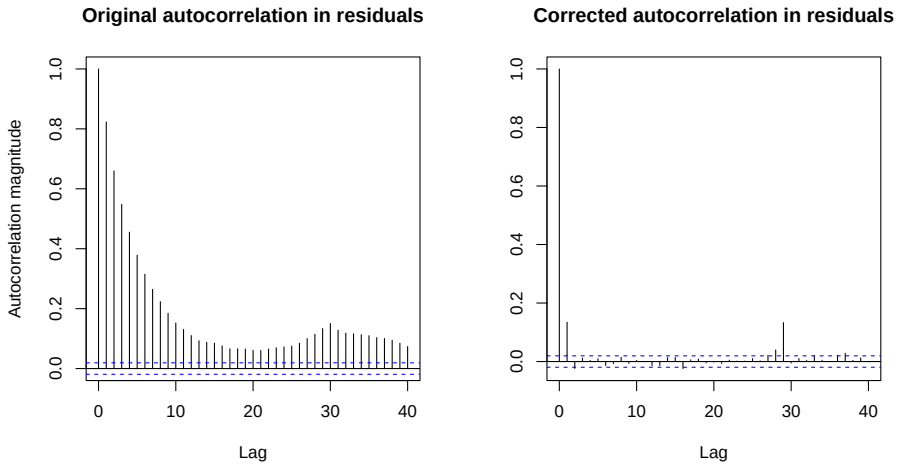

Figure 2.11: Autocorrelation in the residuals. Left: without correction; right: after correction.

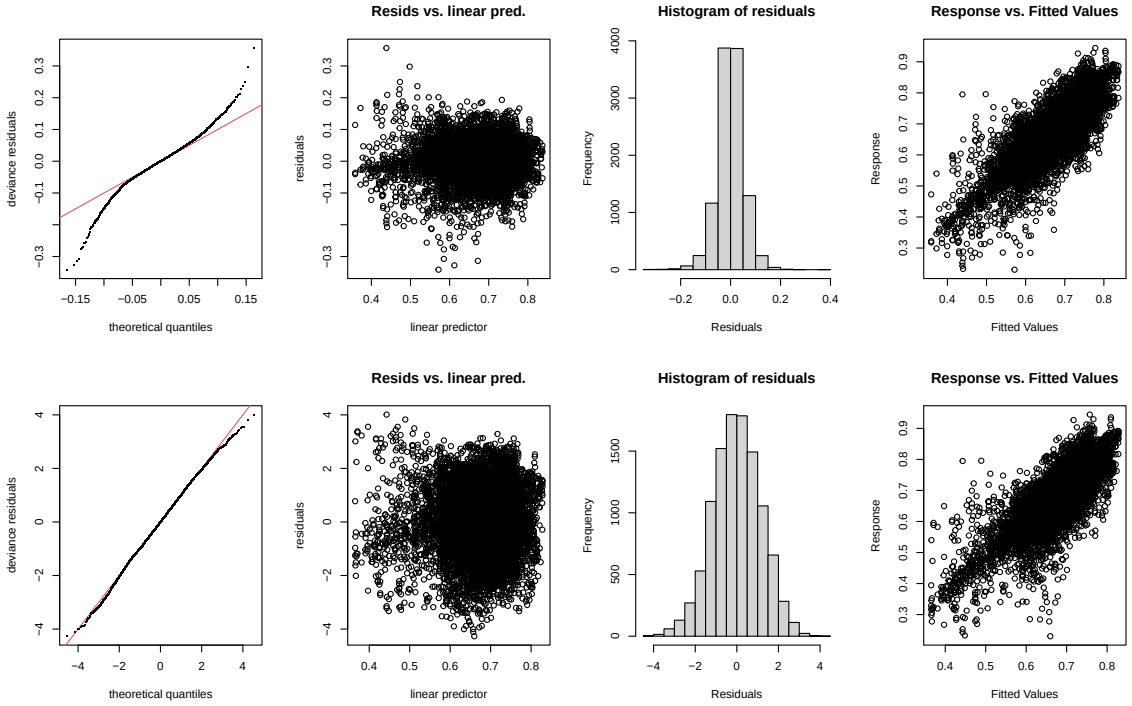

Figure 2.12: Comparing the residuals of models fitted with a Gaussian and scaled-t distributions.

model. The upper Q-Q plot shows that the data distribution is highly tailed. In addition, the second and fourth graphs on the top also indicate a heteroscedasticity issue of the model fitted in Gaussian distribution. Given that the pattern of the residuals resembles that of

a normal distribution with heavier tails, I refitted all the models using the scaled-t family for heavy-tailed data following the procedure introduced in Wieling (2018). The resulting model summary and the associated patterns are reasonably similar to the Gaussian model, and this procedure resulted in improved characteristics of the residuals (see the lower four graphs in 2.12). All results reported in this paper are based on the models fitted using the scaled-t family after model criticism.

2.4.5 Test of significance

There are mainly three ways of testing the significance between curves in GAM (Van Rij, 2016).

(20) a. Model comparison

b. Summary inspection

c. Visualization

Model comparison

The goal of model comparison is to test if a particular variable contributes to the variance in the data. Model comparison is through using the `compareML()` and `AIC()` functions to compare the AIC values of two models that only differ in the inclusion of the variable of which one wants to test the significance. The one that includes the variable is the complete model, which does not have the nested model. The function `compareML()` compares two models based on the minimized smoothing parameter selection score specified in the model and performs a χ^2 test on the difference in scores and the difference in degrees of freedom. In principle, the model with the lower AIC value should be preferred. The schema for specifying the full and the nested model is shown below.

(21) a. Full model:

```
y ~ var + s(x, var) + s(x, random, by=var, bs="fs", m=1
```

b. Nested model:

`y ~ s(x) + s(x, random, bs = "re", m=1)`

There are several drawbacks to the model comparison approach. First, the two models can only be fitted with maximum likelihood estimation to make the comparison meaningful. Although less prone to local minima, running requires substantially more time. However, there is a way to get around the computational cost issue due to maximum likelihood estimation. According to Sóskuthy (2021, p.9), the alternative relies on the fact that `REML` and `fREML` can be used for model comparison when the models only differ in their random effects. It is possible to replace the parametric difference term with a random intercept (`s(fixed, bs="re")` instead of `fixed`) and to place an additional null space penalty on the smooth difference term (by setting the select parameter of `bam()` to the value `TRUE`), effectively turning both of them into random effects. Model comparison can then be performed using models estimated via fREML.

Another drawback is that while the model comparison tells us if the difference due to the critical variable is significant or not, it does not tell us how it affects the non-linear pattern or the direction of the effect; hence the result is hard to interpret. Therefore it needs to be complemented with a visual inspection of the result to determine where the difference comes from.

Wald test

An alternative way to test the significance of the parametric term (the linear fixed effect) is through the function `anova()`. For a single GAM object, Wald tests of the significance of each parametric and smooth term are performed. An example of testing the significance in the Chinese /ai/ normalized F1 data through using `anova()` is given below.

```
Parametric Terms:
        df    F p-value
pos.ord  2 5.69 0.00339
```

```
Approximate significance of smooth terms:
                                      edf  Ref.df        F  p-value
s(Time)                            12.303  13.355   60.681  < 2e-16
s(Time):pos.ordlist-final           5.662   7.019   21.702  < 2e-16
s(Time):pos.ordIP-final             3.784   4.408   13.520  < 2e-16
s(Time,Speaker)                    92.973 150.000    4.556  < 2e-16
s(Time,Speaker):pos.ordlist-final  64.169 178.000   11.125  0.52355
s(Time,Speaker):pos.ordIP-final   100.361 178.000 6297.462  < 2e-16
s(Block):Positionword-final         1.123   1.228    0.145  0.86895
s(Block):Positionlist-final         1.682   1.898   10.298  0.00038
s(Block):PositionIP-final           1.000   1.000   17.604 2.75e-05
ti(Time,Block):Positionword-final   5.176   7.405    0.991  0.42684
ti(Time,Block):Positionlist-final   1.647   2.086    0.527  0.56530
ti(Time,Block):PositionIP-final     4.190   6.075    0.892  0.49960
```

Clearly there is quite strong evidence that the linear fixed effect `pos.ord` (prosodic position) matters ($F = 5.69$ ($p < .05$)). The intercepts of F1 in the three prosodic conditions are significantly different. This can be further explored by plotting the parametric term as in figure 2.13.

The figure demonstrated that the F1 in the word-final position is lower than those in the list- and IP-final positions. However, the difference between the latter two prosodic contexts was negligible.

Summary inspection

Inspecting the summary of the model is another way to test the non-linear significance of the linear parameters (fixed effects). This approach requires GAMs with so-called difference terms (binary or ordered factor difference smooth). If this difference smooth is found

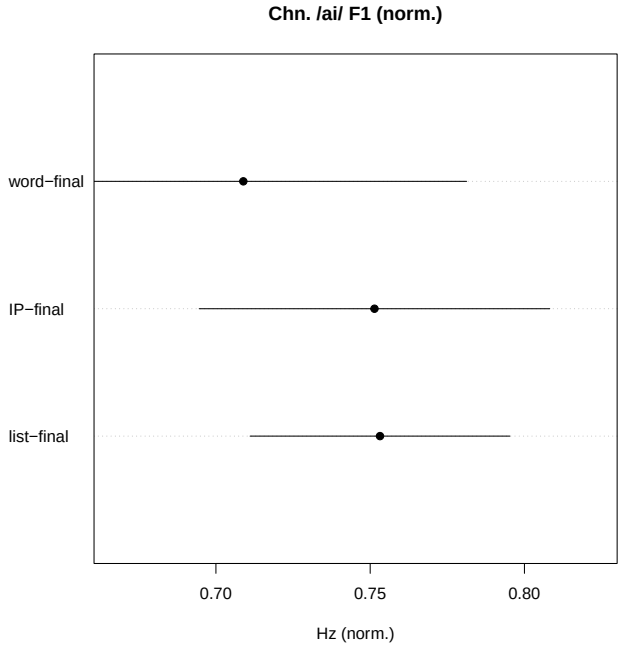

Figure 2.13: The summed effect of `pos.ord` of F1 in Chinese /ai/.

to be significant, the additional complexity of distinguishing two or more levels in a parametric variable (the fixed effects) is required. A disadvantage of this method is that the difference smooth simultaneously includes the non-linear and the intercept difference between the two levels. It may be desirable to separate these since I am interested in assessing the nature of the difference in the formants among the three prosodic positions. It is also essential to keep in mind that each distinct binary predictor may only occur exactly once in the model specification. I adopted the ordered factor difference smooth to model the difference. The schema for specifying the model is as below:

Table 2.11: Ordered factor difference smooth

	Terms	Purpose
`y` $\sim$	`fixed.ordered +`	# *parametric term*
	`s(x) +`	# *reference smooth*
	`s(x, by=fixed.ordered)`	# *difference smooth*

An example of a model summary is given in Table 2.12. The summary of the model of F1 of Chinese /ai/ included a factorial smooth treating `Speaker` as the random effect,

a smooth over `Block` and a tensor product interaction of `Time` and `Block` over prosodic `Position`. It is clearly shown in the result that the smooth terms `s(Time):pos.ordlist-final` (the difference between "list-final" and "word-final") and `s(Time):pos.ordIP-final` (the difference between "IP-final" and "word-final") are both significantly different from 0. This indicates that the inclusion of prosodic position in the model is necessary for modeling the F1 curve of Chinese /ai/.

Table 2.12: The summary of F1 of Chinese /ai/
($***$: $p < .005$; $**$: $p < .01$; $*$: $p < .05$; .: $p < .1$)

A. parametric coefficients	Estimate	Std. Error	t-value	p-value	Sig.
(Intercept)	0.6842	0.0235	29.1403	< 0.0001	$***$
pos.ord.L	0.0383	0.0209	1.8319	0.0670	.
pos.ord.Q	-0.0226	0.0197	-1.1466	0.2516	
B. smooth terms	edf	Ref.df	F-value	p-value	
s(Time)	12.3029	13.3552	60.6805	< 0.0001	$***$
s(Time):pos.ordlist-final	5.6620	7.0189	21.7025	< 0.0001	$***$
s(Time):pos.ordIP-final	3.7839	4.4081	13.5202	< 0.0001	$***$
s(Time,Speaker)	92.9734	150.0000	4.5557	< 0.0001	$***$
s(Time,Speaker):pos.ordlist-final	64.1686	178.0000	11.1248	0.5235	
s(Time,Speaker):pos.ordIP-final	100.3605	178.0000	6297.4624	< 0.0001	$***$
s(Block):Positionword-final	1.1227	1.2283	0.1453	0.8689	
s(Block):Positionlist-final	1.6817	1.8982	10.2983	0.0004	$***$
s(Block):PositionIP-final	1.0001	1.0003	17.6037	< 0.0001	$***$

The partial effects of the difference smooths are shown in figure 2.14.

Visualization

The final method of testing the fixed parametric term's significance uses visual methods based on confidence intervals.

The advantage of visualizing the difference modeled by GAM is that it allows the researchers to see specifically from which point the curves from the two groups differ. Visual methods are crucial to interpreting the output of GAMs. They offer an ideal tool for exploratory investigations just as other methods that model the non-linear relation between two variables, such as SSANOVA, GCA (Growth Curve Analysis), and FDA (Functional

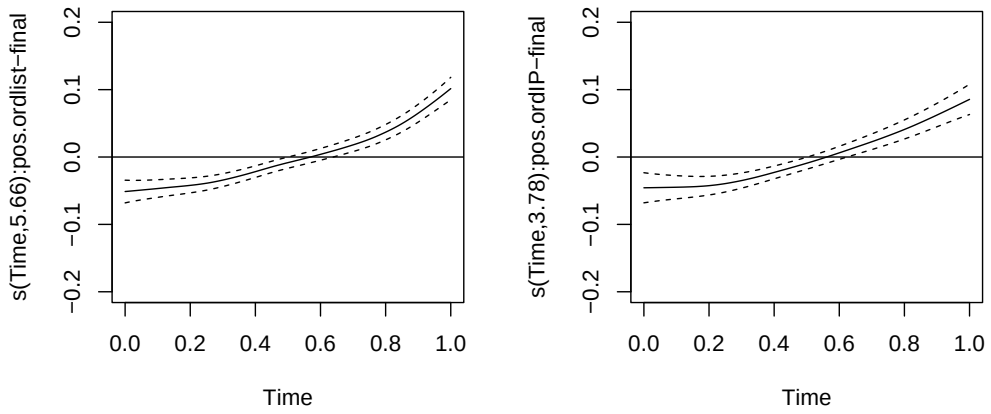

Figure 2.14: Visualization of the ordered factor difference smooth (partial effect) of the model shown in 2.12.

Data Analysis). When used in this way, they are an essential complement to the significance testing methods above. However, visual methods have drawbacks, too, as the significant difference in a specific region does not necessarily imply that the curves of the two groups are significantly different overall. When the researcher does not have a hypothesis about where the difference should be expected, interpreting significance from the visualization alone can be dangerous.

The three methods of significance testing should complement each other. They will be used in conjunction in the following sections when reporting the result of the GAMs. The by-speaker random effect and the tensor product interaction between `Time` and `Block` will not be reported in the subsequent sections.

2.4.6 Final specification of models

The final model of normalized F1/F2 is as below:

The full model to be used in model comparison specifies the parametric term in the final model as a random effect: `s(pos.ord, bs="re")`. As mentioned above, `pos.ord` is the ordered factor treating "word-final" as the reference level. The nested model removes

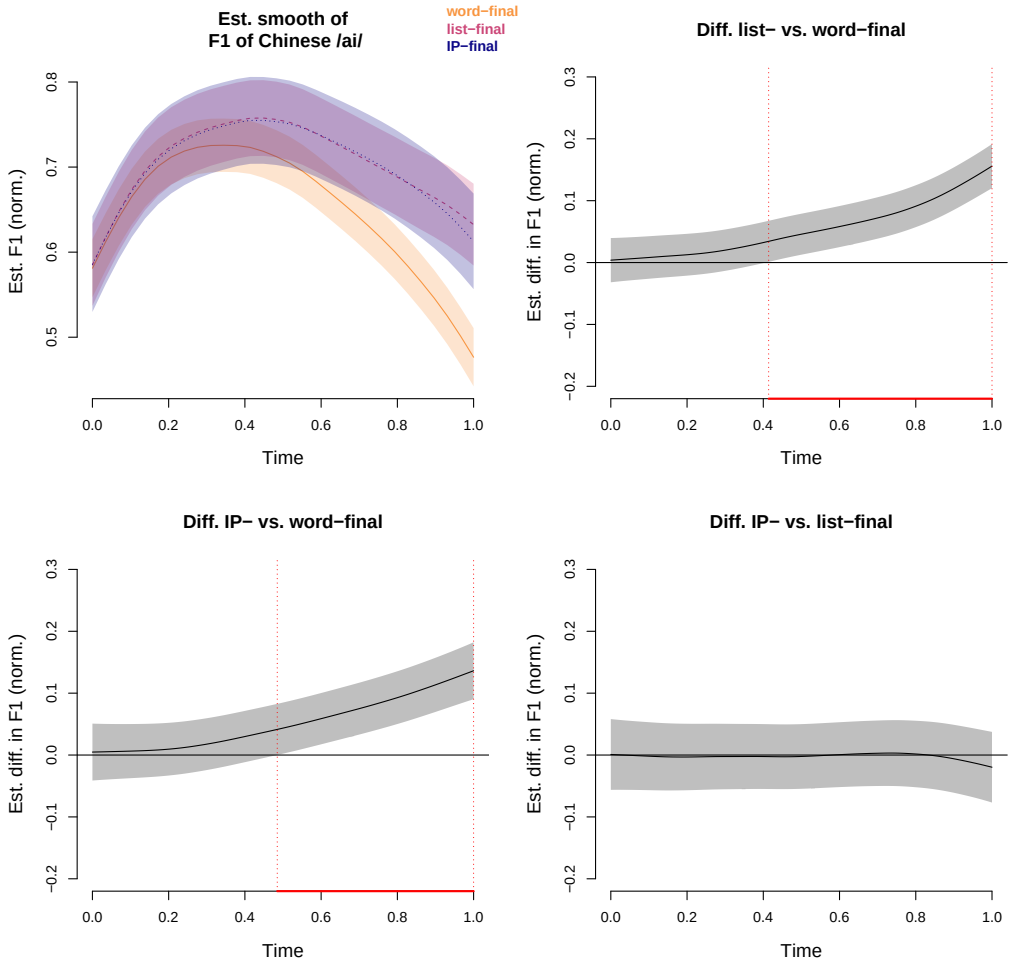

Figure 2.15: Top left: Model predictions for the two groups of contours with 95% pointwise confidence intervals. Top right, bottom left and right: The estimated difference among "IP-final", "list-final", "word-final" with the associated 95% pointwise confidence interval. The highlighted area indicates where the confidence interval excludes zero.

Table 2.13: The final model specification of F1/F2 of TVS.

	Terms	Purposes
`Formant (norm.)` ∼	`pos.ord +`	*parametric term*
	`s(Time, k=25) +`	*reference smooth*
	`s(Time, by=pos.ord, k=30) +`	*difference smooth*
	`s(Time, Speaker, bs="fs", m=1, k=25) +`	*random reference smooth*
	`s(Time, Speaker, by=pos.ord, bs="fs", m=1, k=30) +`	*random difference smooth*

all the smooth terms that included `pos.ord`. The numbers of basis functions (specified as `k=...`) in all the models were large enough to model the data after examination by using `gam.check()` function.

When F1 and F2 are mentioned in the subsequent subsections, they refer to the normalized F1 and F2 corrected for each speaker's vocal tract length. The plots of the difference smooths and the tensor product interaction is in the appendix.

Chapter 3

Pre-boundary lengthening

This chapter reports the results of the analysis on the duration of the vocalic segments, including both TVS and monophthongs. The analysis consists of two components: the analysis based on raw duration and the analysis based on the percentage of lengthening compared to word-final context. The percentage analysis was included because the raw duration is influenced by language-specific intrinsic duration of segments. The speech rhythm differs across languages. Both vowels and consonants can vary to a great extent even though they are transcribed using the same set of IPA symbols (For a detailed overview of research on speech rhythm, please see Fletcher (2010, section 3). The percentage of lengthening compared to the duration in the word-final context can better capture how much extra articulatory effort was made in pre-boundary lengthening for different languages.

Each of the analyses will be further divided into two subparts: 1. statistical analysis will be performed to examine if the pre-boundary lengthening effect differs for the two different segment types: monophthongs and vowel sequences; 2. further statistical analyses will be performed for each language to explore how the magnitude of pre-boundary lengthening differs for different vocalic segments.

3.1 Results of raw duration

Linear mixed-effect models were built to examine the statistical significance. For the overall comparison between monophthongs and TVS, a model was built treating the raw duration of the vowels as the dependent variable, `Position`, (vowel) `Type`, `Language` as fixed effects, and `Speaker` as random effects. `Position`, `Type` are treated as within-speaker fixed effects while `Language` as between-speaker fixed effects. The maximally complex model specification that managed to converge has uncorrelated random intercepts and random slopes. For the by-language models, the `Position`, `Segment` are treated as fixed effects, and `Speaker` as the random intercept, except for the model for Chinese segment durations in which the `Position` was treated as a within-speaker fixed effect, which its random slope did not correlate with the random intercept.

3.1.1 Overall difference of durations between monophthongs and TVS

Figure 3.1 exhibits the overall durations of both monophthongs and TVS in different prosodic contexts. Table 3.1 shows the mean values and standard deviations of each vocalic segment in each prosodic context across the languages.

Table 3.1: The mean values (ms) and standard deviations of durations of monophthongs and TVS.

Language	Position	Mean (Mono)	Mean (TVS)	Sd. (Mono)	Sd. (TVS)
Chinese	word-final	150.25	160.38	43.23	37.45
	list-final	176.33	185.64	36.36	37.56
	IP-final	156.30	168.08	34.34	33.63
English	word-final	164.42	192.04	43.79	46.90
	list-final	231.08	279.04	44.35	47.68
	IP-final	209.31	252.26	52.06	54.29
Japanese	word-final	100.48	135.15	27.46	24.36
	list-final	155.28	188.71	32.40	32.99
	IP-final	110.88	142.65	22.19	19.20

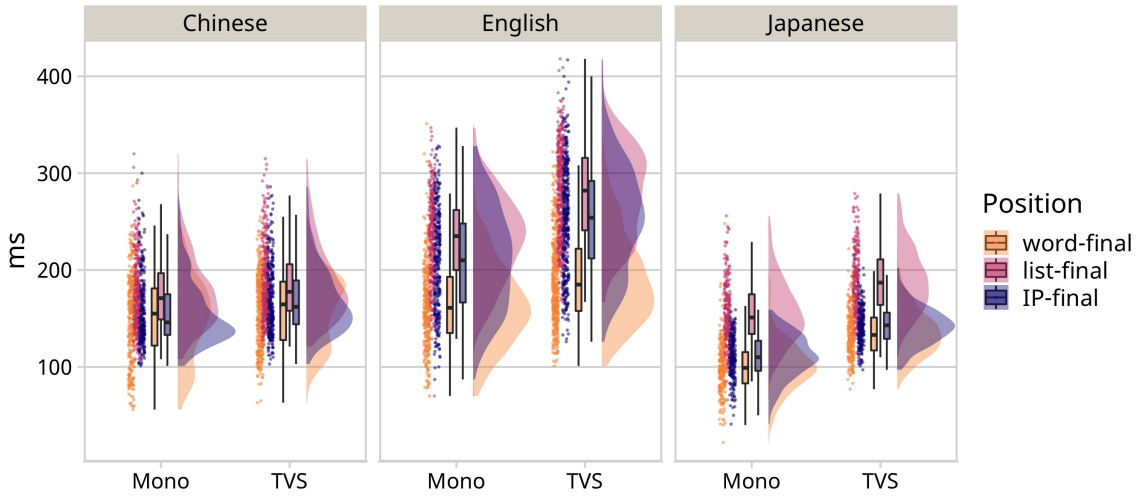

Figure 3.1: The durations of monophthongs (Mono) and TVS in different prosodic contexts.

From the figure and the table, we can observe that English vocalic segments are the longest while Japanese ones are the shortest. The TVS is longer than monophthongs in three languages. With regard to the influence of prosody on segmental durations, the three languages exhibit the same trends for both segment types: vowels that occurred in word-final positions were produced with the shortest duration, while those in the list-final positions had the longest duration. The difference in durations induced by prosody seems larger in English and Japanese than in Chinese.

Statistical analysis confirmed that there is a main effect of `Position` ($F(2, 14.80) = 73.81$, $p < 0.005$), `Language` ($F(2, 15.00) = 17.76$, $p < 0.005$), and segment `Type` ($F(1, 14.73) = 148.55$, $p < 0.005$). The interactions between `Language` and `Position` ($F(4, 14.81) = 5.21$, $p < 0.01$), and between `Langauge` and `Type` ($F(2, 14.78) = 17.24$, $p < 0.005$). The interaction of `Position` and `Type` and three-way interaction failed to reach significance, suggesting that the overall lengthening effect measured on raw duration is not different between monophthongs and TVS. The post-hoc Tukey-adjusted comparisons are shown below.

Figure 3.2 demonstrates that the differences associated with different prosodic contexts in Chinese were not significant. In English, the duration in list-final conditions is longer

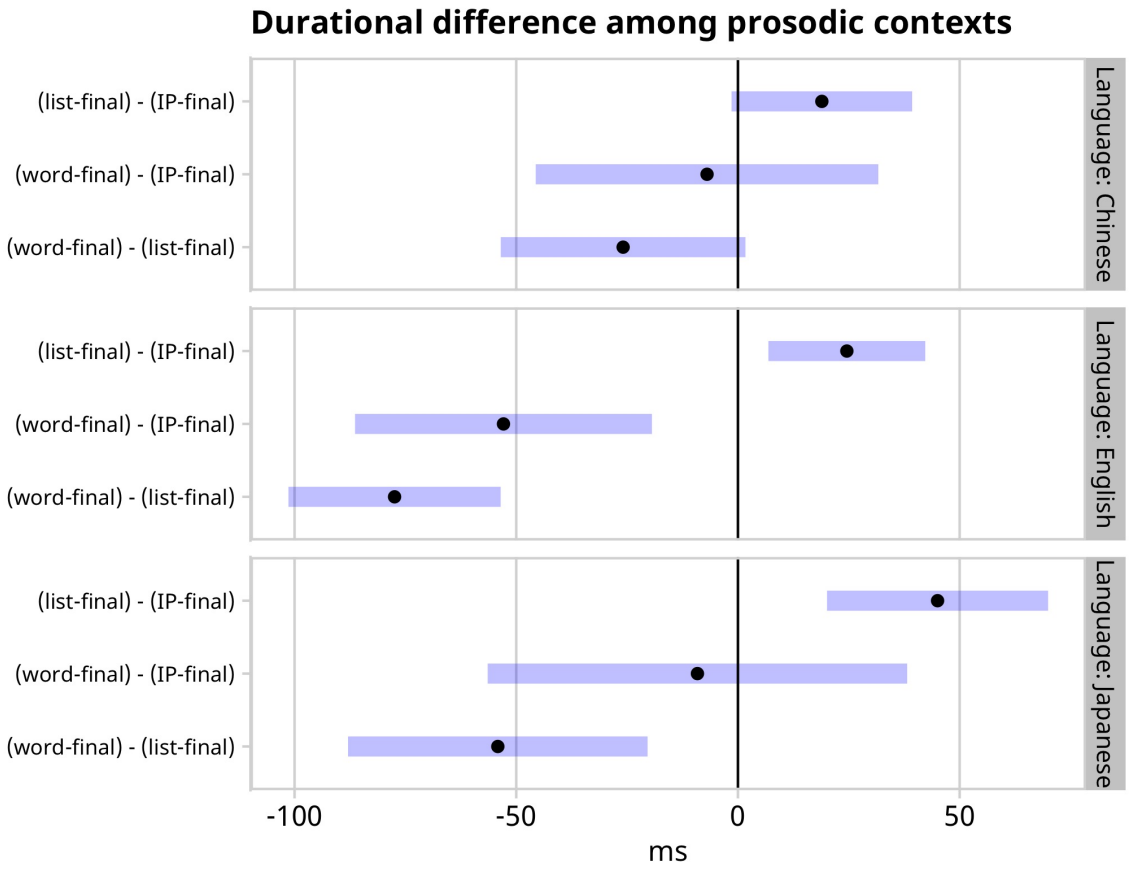

Figure 3.2: The post-hoc comparison among prosodic contexts by language.

than in IP-final conditions, and the word-final condition was produced with the shortest duration.

Further 3.3 shows that monophthongs are shorter than TVS in all three languages but much shorter in English and Japanese than in Chinese.

3.1.2 Durational effect by segment

Figure 3.4 exhibits the distribution of durations of each vocalic segment in the three languages. In all three languages for all segments, there seems to be a lengthening effect in comparing word-final and list-final positions. The duration in list-final positions is slightly longer than in IP-final positions. The lengthening effect did not seem to depend too much on the vocalic segments. The mean duration and standard deviations of each segment are

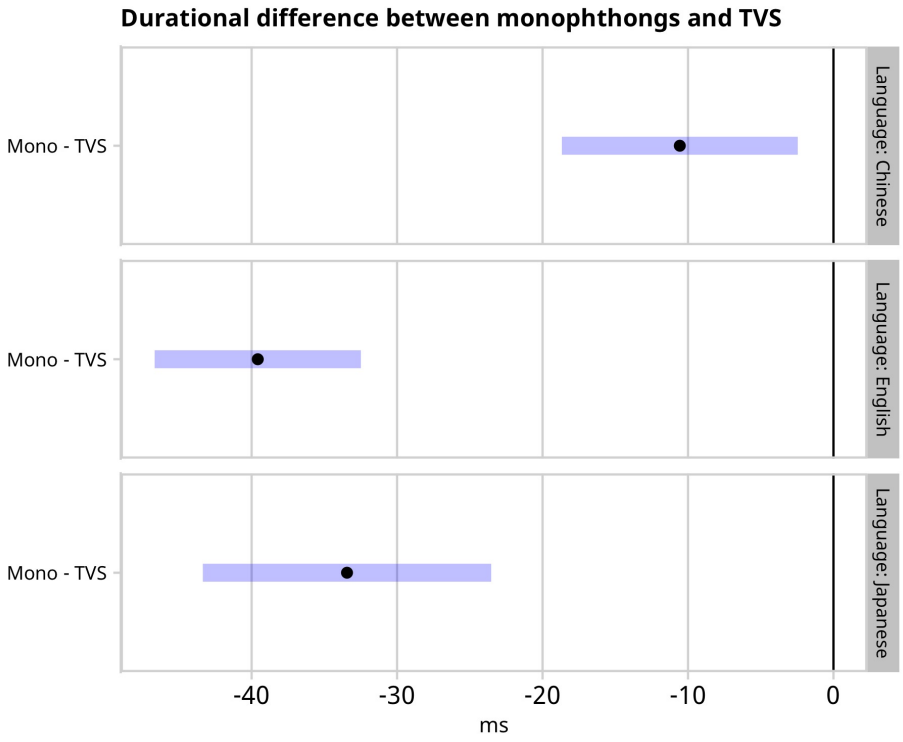

Figure 3.3: The post-hoc comparison between segment types by language.

given in table 3.2, 3.3, and 3.4.

The statistical analysis results show that `Position`, `Segment` and the interaction between them are all significant in three languages. The result is demonstrated in table 3.5.

The post-hoc Tukey-adjusted comparisons are shown below. In Chinese, the overall trend is that segments in the list-final positions had the longest durations, whereas word-final durations were the shortest. Figure 3.5 exhibits that none of the vocalic segments showed significant differences between word-final and IP-final positions in Chinese. The difference between word-final and list-final positions was significant except for /ai, a/. List-final positions were also produced with longer durations than IP-final positions, except in /ai/. The difference induced by prosody seems to be slightly larger in /i, u, ou/ than in /a, ai, au/.

Figure 3.6 shows the pattern of how English vowels are affected by the prosody. The patterns in English showed the same trend as that in Chinese but were much more signifi-

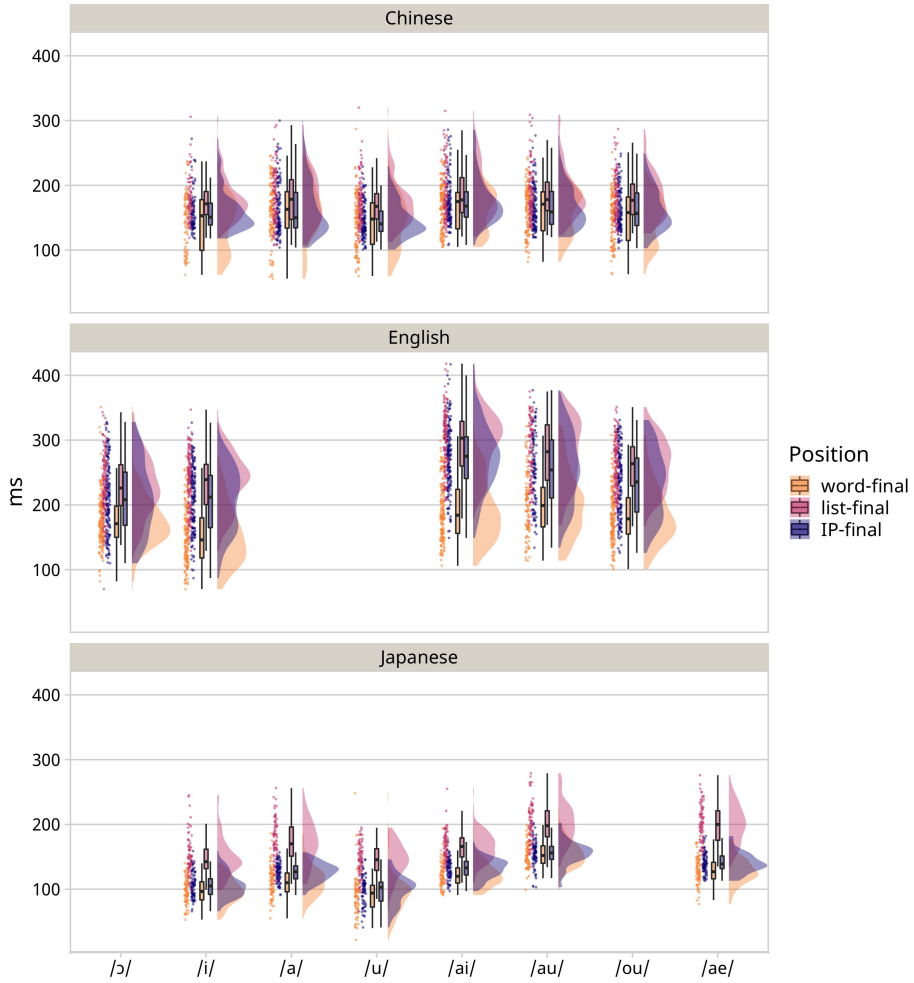

Figure 3.4: The durations of monophthongs (Mono) and TVS in different prosodic contexts.

cant. List-final segments were the longest. IP-final segments were shorter than the list-final ones but longer than word-final ones. The result is significant in all vocalic segments in English. The most influenced segment in English was /ai/, while the least influenced segment was /ɔ/.

Figure 3.7 shows the post-hoc comparison of segment duration in Japanese. The trend that list-final vowels have the longest duration and the word-final vowels have the shortest is also present in Japanese. However, the difference between word-final and IP-final positions was only significant for three segments, i.e., /i, u, au/.

In sum, the lengthening effect on the vocalic segments showed the same trend in the

Table 3.2: Chinese segment durations (mean values and standard deviations).

Segment	Position	Mean	Sd.
/i/	word-final	143.45	42.57
/i/	list-final	177.36	33.09
/i/	IP-final	159.60	30.82
/a/	word-final	161.36	43.64
/a/	list-final	180.01	39.80
/a/	IP-final	164.16	40.60
/u/	word-final	143.97	41.34
/u/	list-final	172.01	34.73
/u/	IP-final	146.30	26.74
/ai/	word-final	167.41	36.20
/ai/	list-final	188.37	40.59
/ai/	IP-final	175.30	35.31
/au/	word-final	163.02	36.41
/au/	list-final	186.38	37.44
/au/	IP-final	166.16	32.42
/ou/	word-final	150.37	37.97
/ou/	list-final	182.08	34.35
/ou/	IP-final	162.72	32.02

Table 3.3: English segment durations (mean values and standard deviations).

Segment	Position	Mean	Sd.
/ɔ/	word-final	176.02	40.08
/ɔ/	list-final	227.92	46.47
/ɔ/	IP-final	210.22	54.74
/i/	word-final	152.81	44.38
/i/	list-final	234.28	42.01
/i/	IP-final	208.39	49.37
/ai/	word-final	192.96	47.45
/ai/	list-final	295.98	46.67
/ai/	IP-final	272.01	50.29
/au/	word-final	201.61	49.59
/au/	list-final	280.45	49.79
/au/	IP-final	254.44	54.04
/ou/	word-final	184.67	43.37
/ou/	list-final	260.85	40.42
/ou/	IP-final	231.07	50.72

Table 3.4: Japanese segment durations (mean values and standard deviations).

Segment	Position	Mean	Sd.
/i/	word-final	97.74	19.84
/i/	list-final	148.88	30.32
/i/	IP-final	105.96	18.89
/a/	word-final	111.80	23.64
/a/	list-final	172.41	33.23
/a/	IP-final	126.64	14.50
/u/	word-final	91.87	33.35
/u/	list-final	144.40	26.27
/u/	IP-final	98.71	22.55
/ai/	word-final	122.58	19.18
/ai/	list-final	165.39	23.11
/ai/	IP-final	132.40	15.28
/au/	word-final	154.70	20.65
/au/	list-final	201.59	31.53
/au/	IP-final	154.89	19.67
/ae/	word-final	128.63	20.38
/ae/	list-final	199.72	30.39
/ae/	IP-final	140.70	15.38

Table 3.5: The statistical result of by-language analysis of segment duration.

	Chinese	English	Japanese
`Position`	F(2, 5.02) = 24.51***	F(2, 2269) = 748.73***	F(2, 1455) = 740.63***
`Segment`	F(5, 2076.02) = 41.93***	F(4, 2269) = 171.71***	F(5, 1455) = 216.12***
`Position:Segment`	F(10, 2076.04) = 2.67**	F(8, 2269) = 10.75***	F(10, 1455) = 4.8708***

(***: $p < .005$; **: $p < .01$; *: $p < .05$.)

three languages: list-final vowels are longer than word-final vowels. The difference between IP-final and word-final vowels depends on the segment and language. IP-final vowels are longer in all English segments, a part of Japanese segments (/a, ae, ai/) but none in Chinese. Cross-linguistically, lengthening measured in raw duration is more evident in English than in Japanese and Chinese. Lengthening in Japanese is also more evident than in Chinese.

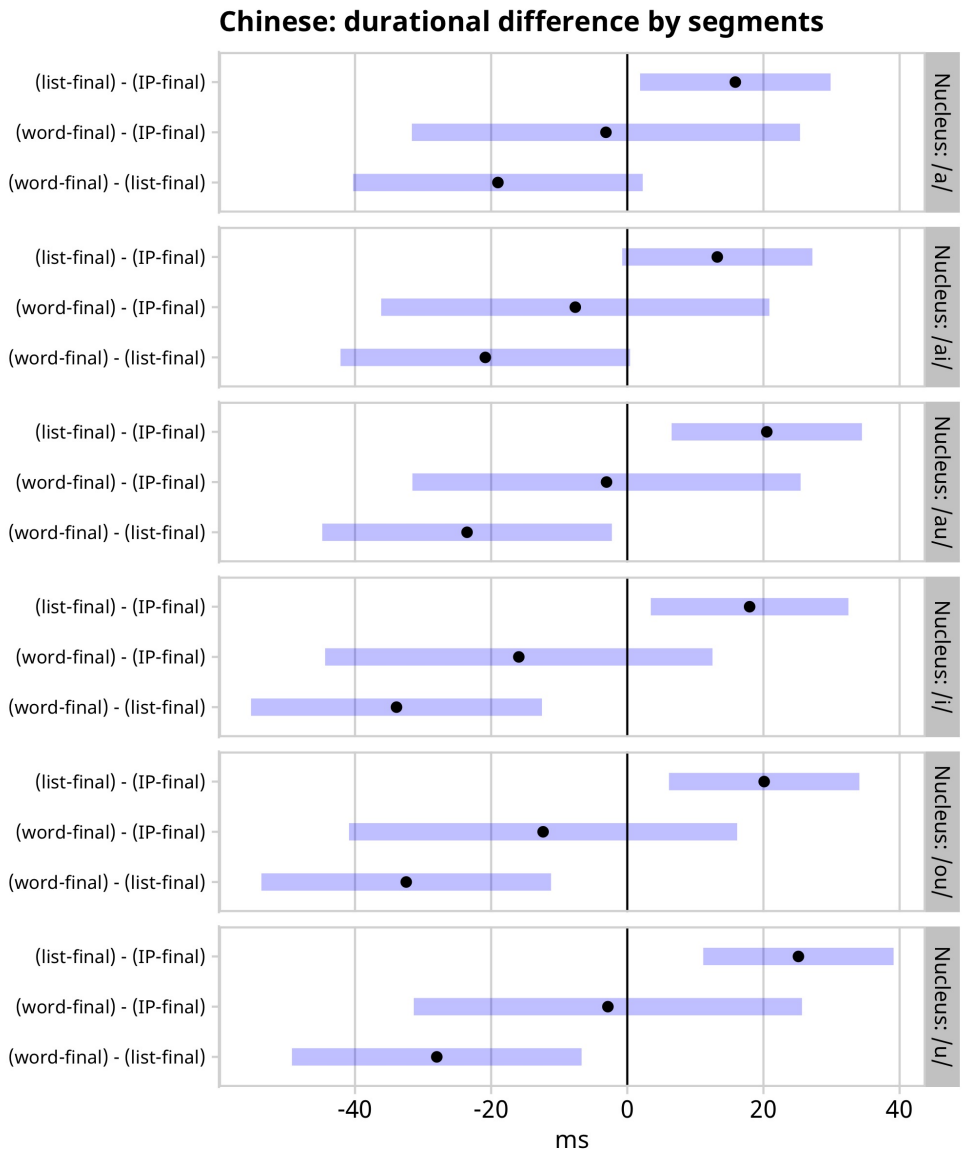

Figure 3.5: The post-hoc comparison among prosodic contexts in Chinese.

3.2 Results of lengthening percentage

As mentioned above, because languages may have intrinsic segment durations, the lengthening effect measured in raw duration might not reflect the nature of efforts made in pre-boundary lengthening by speakers well. Therefore, in this section, I will analyze the lengthening percentage by comparing list-final/IP-final positions to word-final positions.

With regard to the statistical analysis of the percentage data, the percentage of length-

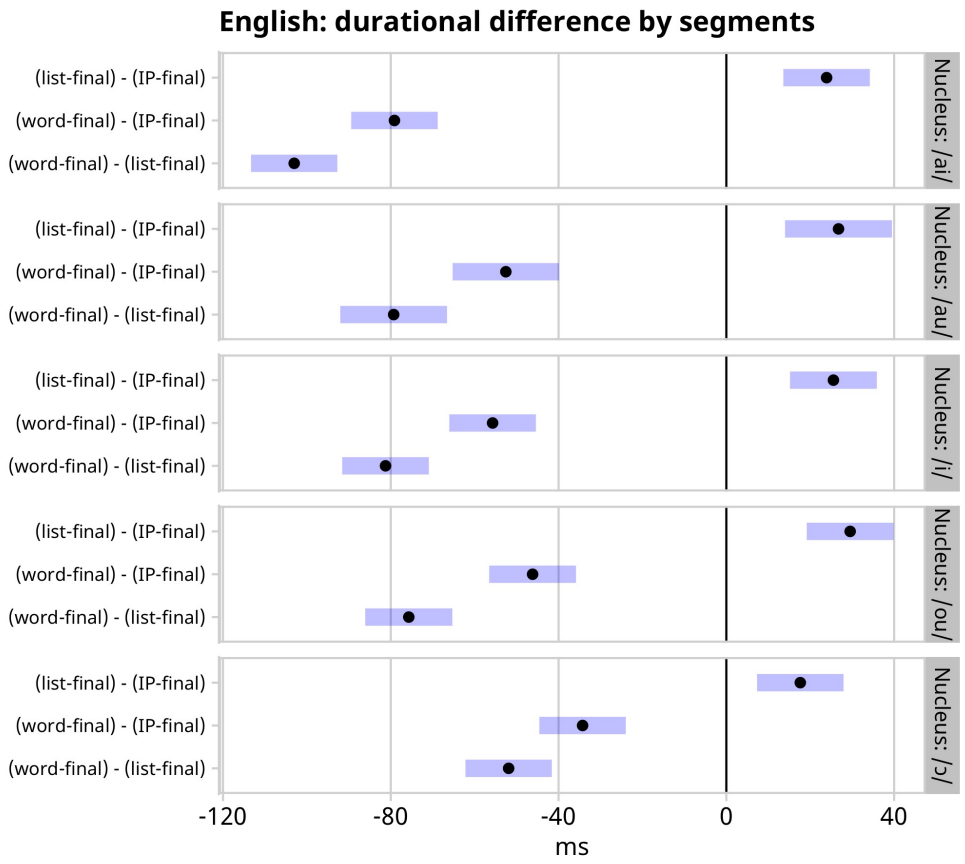

Figure 3.6: The post-hoc comparison among prosodic contexts in English.

ening was calculated using the equation below.

(22) Lengthening percentage = (duration − duration in word-final position)/(duration in word-final position)

The maximally converging linear mixed-effect models for percentage analysis only included `Speaker` as the random slope.

3.2.1 Overall difference between monophthongs and TVS

Figure 3.8 shows the percentage of lengthening in monophthongs and TVS in Chinese, English, and Japanese. It is shown that percentage-wise, the lengthening is more evident in list-final positions than in IP-final positions. The lengthening effects comparing IP-final to word-final positions in Chinese and Japanese are much smaller than in English. Table 3.6

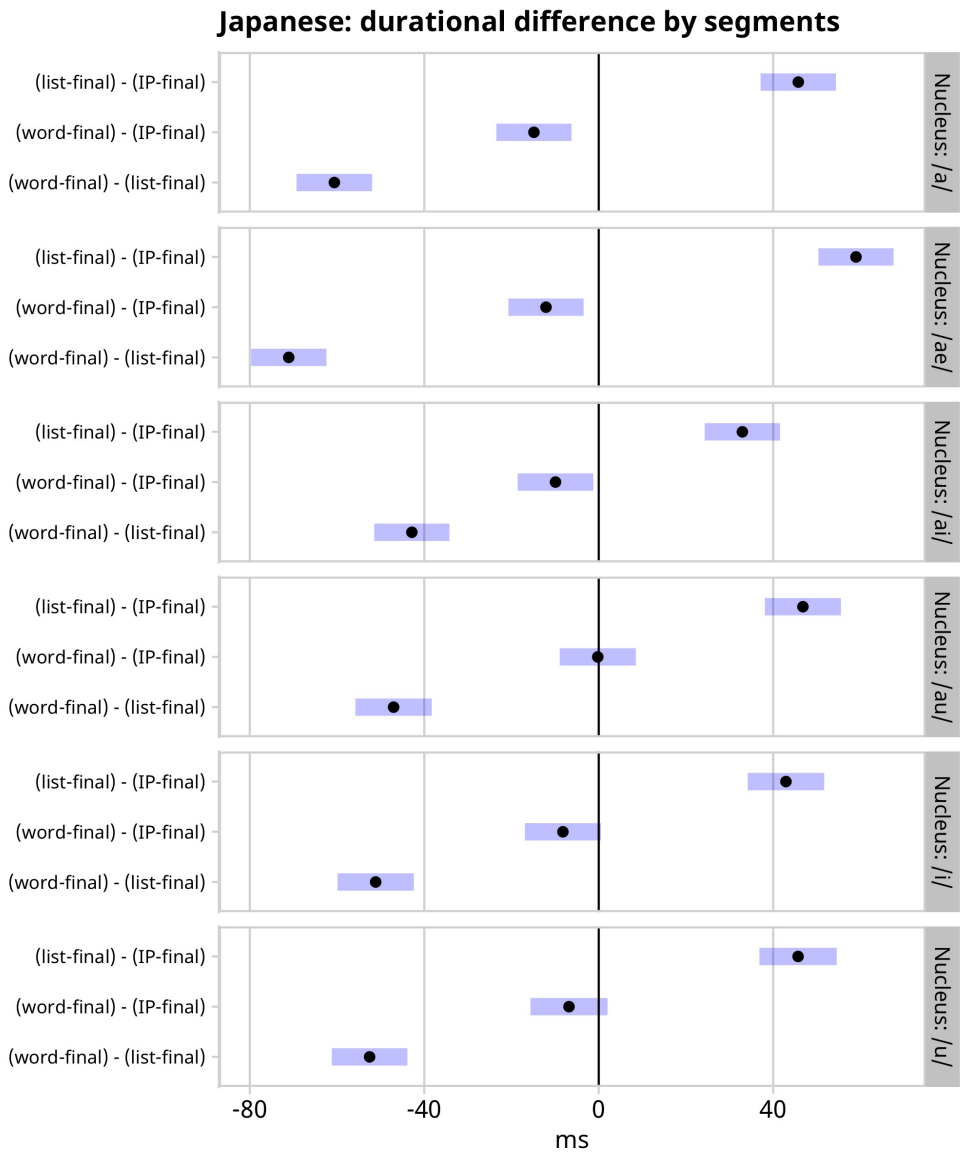

Figure 3.7: The post-hoc comparison among prosodic contexts in Japanese.

shows the summary for the three languages.

The statistical analysis showed that there are main effects of `Position` ($F(1, 3847) = 528.25$, $p < 0.005$) and segment `Type` ($F(1, 3847) = 27.08$, $p < 0.005$) but not of `Language` ($F(2, 15) = 2.47$). Note that the main effect of `Language` found in raw duration data is absent in percentage data. All three two-way interactions were statistically significant (`Position:Language` ($F(2, 3847) = 75.65$, $p < 0.005$), `Position:Type` ($F(1, 3847) = 7.30$, $p < 0.01$); `Language:Type` ($F(2, 3847) = 24.1689$, $p < 0.005$)). The three-way inter-

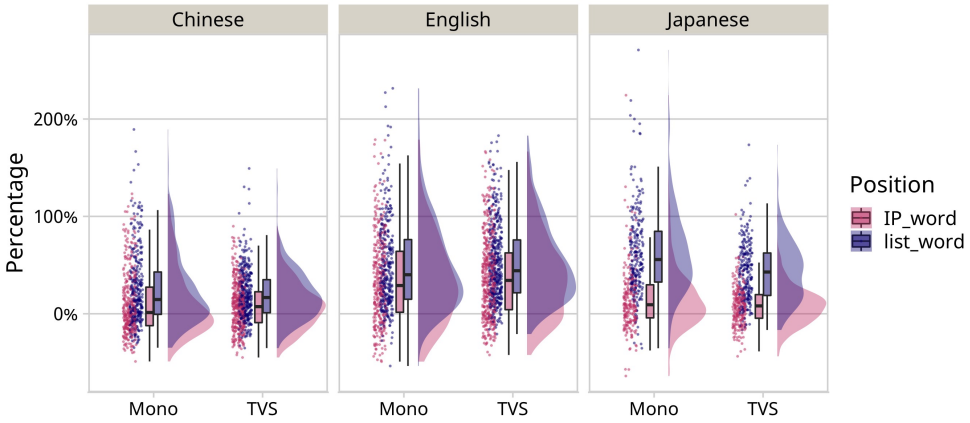

Figure 3.8: The percentage of lengthening by segment type.

Table 3.6: The summary of the percentage of lengthening by segment type.

Language	Position	Mean (Mono)	Mean (TVS)	Sd. (Mono)	Sd. (TVS)
Chinese	IP-word	0.11	0.08	0.33	0.24
	list-word	0.26	0.20	0.37	0.26
English	IP-word	0.35	0.38	0.45	0.42
	list-word	0.49	0.52	0.46	0.40
Japanese	IP-word	0.15	0.08	0.33	0.20
	list-word	0.63	0.43	0.45	0.31

action reached significance ($F(2, 3847) = 3.57$, $p < .05$) as well. The three-way interaction visualized in figure 3.9.

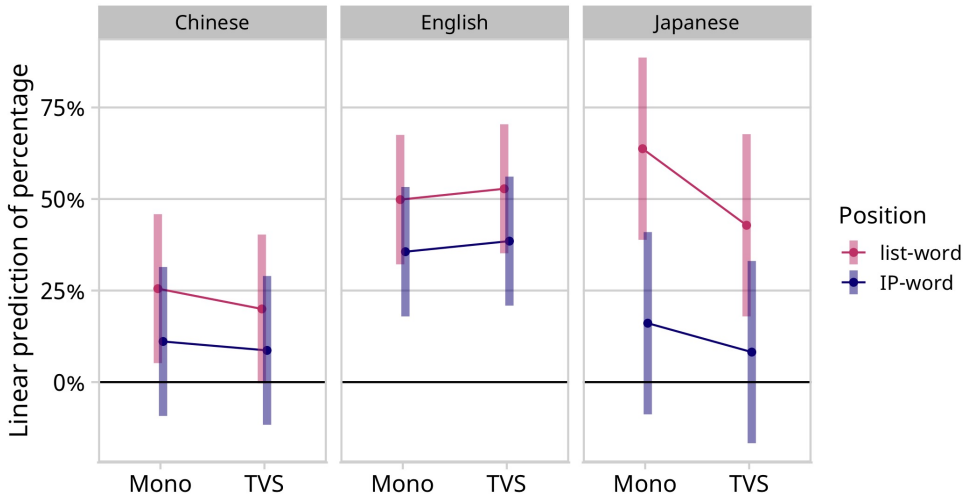

Figure 3.9: The percentage of lengthening by segment type.

Since the lower bound of the 95% confidence is lower than 0%, the lengthening due to the IP boundary is insignificant in Chinese and Japanese. The Japanese vowels are much more lengthened in the list-final contexts than in the IP-final context. Also, the monophthongs in Chinese and Japanese seemed more lengthened than the TVS, while the segment type difference is insignificant in English. Table 3.7 shows the result of the post-hoc comparison between segment types in three languages.

Table 3.7: The post-hoc comparison of segment types across languages.

contrast	estimate	SE	df	t.ratio	p.value
Language = Chinese					
Mono - TVS	0.0399	0.0162	3847.00	2.464	0.0138
Language = English					
Mono - TVS	-0.0292	0.0156	3847.01	-1.873	0.0611
Language = Japanese					
Mono - TVS	0.1440	0.0194	3847.06	7.415	<.0001

Results are averaged over the levels of: Position
Degrees-of-freedom method: kenward-roger

Therefore, the percentage data showed a different trend than in raw durations that the lengthening is different for monophthongs and TVS. And the pattern is that monophthongs are more lengthened in Chinese and Japanese.

3.2.2 Percentage of lengthening by segments

I will further analyze the difference in lengthening percentage in different segments. The summary of the statistical result is shown in table 3.8.

Table 3.8: The statistical result of by-language analysis of lengthening percentage.

	Chinese	English	Japanese
`Position`	F(1, 1375) = 91.99***	F(1, 1501) = 76.77***	F(1, 949) = 415.05***
`Segment`	F(1, 1375) = 12.24***	F(4, 1501) = 49.88***	F(5, 1949) = 18.06***
`Position:Segment`	F(10, 2076.04) = 1.38	F(4, 1501) = 0.61	F(5, 949) = 4.34***

(***: $p < .005$; **: $p < .01$; *: $p < .05$.)

The main effects of `Position` and `Segment` were confirmed for all three languages. However, the lengthening measured as percentages only showed significant interactions between `Position` and `Segment` in Japanese. This indicates the difference between the lengthening of IP-final positions compared to word-final positions is not different from the lengthening in list-final positions compared to word-final positions in both Chinese and English. The linear predictions of lengthening percentage in the three languages are shown in figures 3.10, 3.11, and 3.12. In these three figures, the lengthening is significant if the confidence intervals do not cross 0%.

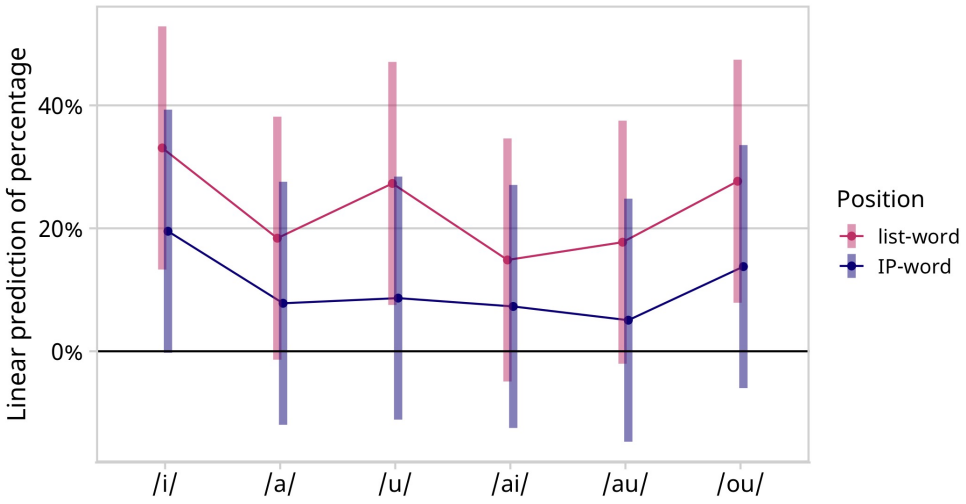

Figure 3.10: The lengthening percentages in Chinese by segment.

Chinese /a, ai, au/ are not significantly lengthened in the list- or IP-final positions compared to their word-final versions. Chinese /i, u, ou/ are lengthened in list-final positions significantly, but the lengthening of /i, u, ou/ is insignificant in IP-final positions. The Chinese monophthongs and diphthongs are more lengthened if they do not involve low vocalic targets. Chinese percentage data are summarized in table 3.9.

The data in English is rather different. The only insignificant lengthening is /ɔ/ in IP-final positions. All other lengthenings were significant, especially for /i/ and /ai/. The average lengthening percentages of /i, ai/ are over 60% compared to the word-final counterparts. Lengthening in English is more evident when the vowel or diphthong involves

Table 3.9: Summary of Chinese percentage data by segment.

Nucleus	Position	Mean	Sd
/i/	list-word	0.33	0.40
/i/	IP-word	0.19	0.36
/a/	list-word	0.19	0.37
/a/	IP-word	0.08	0.34
/u/	list-word	0.27	0.36
/u/	IP-word	0.08	0.30
/ai/	list-word	0.15	0.21
/ai/	IP-word	0.07	0.22
/au/	list-word	0.18	0.24
/au/	IP-word	0.05	0.22
/ou/	list-word	0.28	0.31
/ou/	IP-word	0.14	0.28

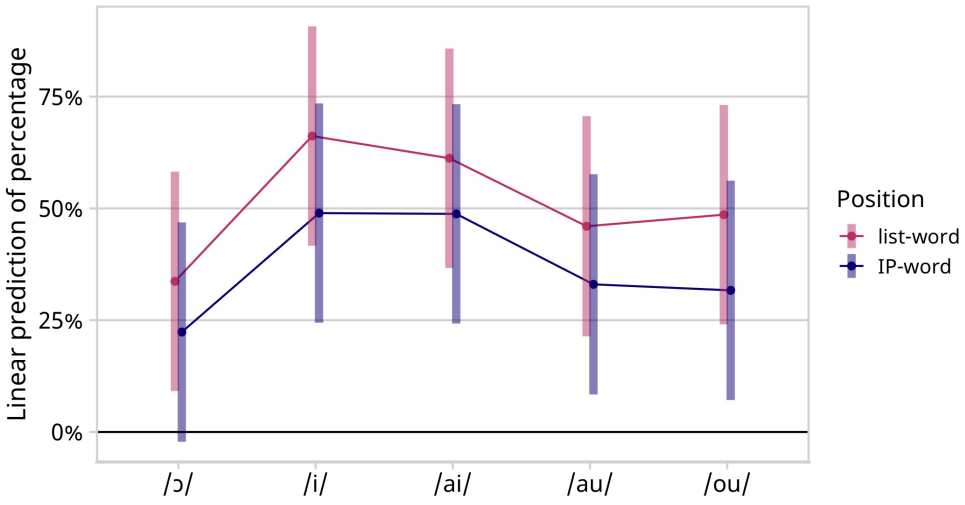

Figure 3.11: The lengthening percentages in English by segment.

reaching a high front vocalic target. English data are summarized in table 3.10.

In Japanese, the lengthening is more significant in monophthongs than in TVS. IP-final lengthening in Japanese was insignificant for /i, ai, au, ae/. The three TVS are only lengthened in the list-final positions, even though the extent of lengthening is much less compared to the monophthongs in the list-final positions. The Japanese percentage data are summarized in table tab:sumJpnPercNuc.

Cross-linguistically, the lengthening effects in list-final positions in English and Japanese

Table 3.10: Summary of English percentage data by segment.

Nucleus	Position	Mean	Sd
/ɔ/	list-word	0.33	0.30
/ɔ/	IP-word	0.22	0.30
/i/	list-word	0.66	0.53
/i/	IP-word	0.49	0.52
/ai/	list-word	0.61	0.41
/ai/	IP-word	0.49	0.41
/au/	list-word	0.46	0.37
/au/	IP-word	0.33	0.41
/ou/	list-word	0.48	0.39
/ou/	IP-word	0.31	0.41

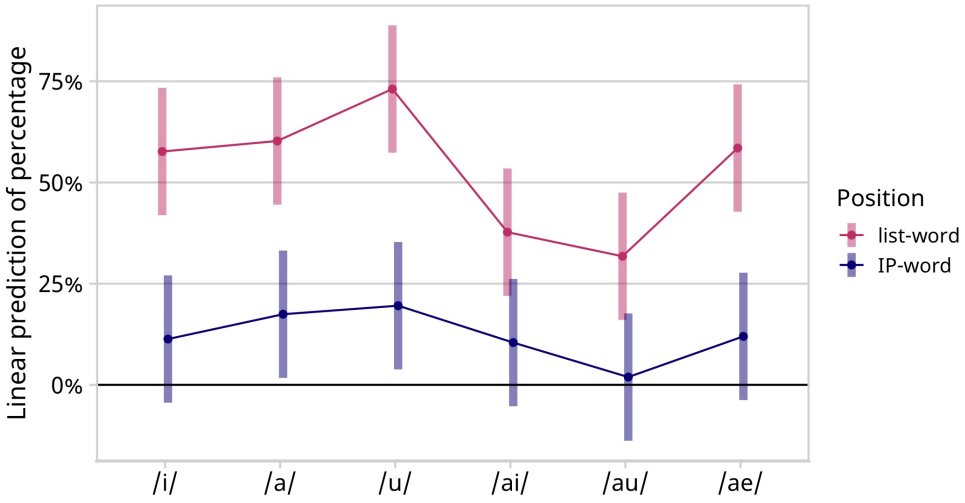

Figure 3.12: The lengthening percentages in Japanese by segment.

are comparable to each other, ranging from 30% to 75%, much larger than that in Chinese ranging from 15% to around 30%. The lengthening in the IP-final positions is much less evident in Chinese and Japanese than in English. Overall, English showed more lengthening than Chinese and Japanese. Lengthening in percentages in Japanese is also larger than in Chinese.

Table 3.11: Summary of Japanese percentage data by segment.

Nucleus	Position	Mean	Sd
/i/	list-word	0.57	0.38
/i/	IP-word	0.11	0.26
/a/	list-word	0.60	0.43
/a/	IP-word	0.17	0.21
/u/	list-word	0.73	0.51
/u/	IP-word	0.18	0.46
/ai/	list-word	0.38	0.26
/ai/	IP-word	0.10	0.20
/au/	list-word	0.32	0.26
/au/	IP-word	0.02	0.20
/ae/	list-word	0.58	0.34
/ae/	IP-word	0.12	0.18

3.3 Discussion

The pre-boundary lengthening effect was confirmed in raw duration and percentage data for all three languages. Although the magnitude depends on the specific segment, the result was significant in at least one data type. However, several noticeable phenomena found need further discussion.

First, it was found that the lengthening was most evident in English but least evident in Chinese, with Japanese lying in between. This is, however, different than the predictions I made in section 1.4 that Japanese should exhibit the least amount of lengthening should its prosody be organized based on mora. That English exhibited the most lengthening among the three is probably because it is the language with the least phonological or phonetic constraints on the segment duration. The possible reason Chinese vowels showed the least amount of lengthening in both raw duration and percentage is probably because the fourth tone used to create the target syllable imposed a ceiling effect on the range of durational variability in speech (Howie, 1976). Although Chinese is a tonal language and the four lexical tones are primarily distinguished in pitch (f0), duration (J. Yang et al., 2017) and amplitude contour (Whalen & Xu, 1992). This intrinsic short duration of Tone 4 in Chinese

might have prohibited some preboundary lengthening in the syllables at boundaries. One may wonder if it is the overall short duration in general that has influenced the preboundary lengthening in Chinese. But this should not be the case. As shown in previous sections, the Japanese vowels are shorter than Chinese. Regardlessly, Japanese vowels were lengthened more than Japanese vowels. Therefore the small amount of lengthening observed in Chinese could not be attributed to the short durations of the vocalic segments in general. Further study is needed to examine the pre-boundary lengthening effect in syllables with different lexical tones in Chinese. In sum, the phonology of basic prosodic units did not seem to predict the differences in the patterns of preboundary lengthening in this study.

Second, segments in the IP-final position in Japanese are not as lengthened as in list-final positions. This is probably because the target syllables at IP boundaries in Japanese are followed by a monosyllabic copula /-da/. As pointed out in previous research (Campbell, 1992; Shepherd, 2008), the pre-boundary lengthening effect is largely confined to the last mora of a prosodic constituent, suggesting that the interval of π-gesture activation is narrower than most languages wherein the boundary-related lengthening have been investigated. Even though some segments still showed significant lengthening in IP-final positions compared to word-final positions in Japanese, such as /a, ai, ae/ in duration data, and /a, u/ in percentage data.

Third, After examining the percentage data, monophthongs are lengthened more than TVS at higher boundaries in Chinese and Japanese. Suppose we interpret that the percentage data could measure speakers' effort in lengthening each segment. In that case, this discrepancy between the result of raw duration and percentage of lengthening on different segments indicates that speakers make different efforts in lengthening monophthongs and TVS to signal prosodic boundaries. Two scenarios are conceivable in explaining this difference. First is that monophthongs are too short in duration. Therefore it needs to be produced much longer than non-final positions to make prosodic boundaries more salient for the speaker. This is a perception-oriented strategy. The second possibility is that the

possible amount of lengthening has a ceiling effect. One can only lengthen segments to a certain degree. The segment duration usually does not exceed a certain threshold for the speakers to save energy in speech production. This is a production-oriented strategy. Without further data, the two accounts cannot be easily teased apart. Further research is needed to investigate the two possible theoretical scenarios.

In conclusion, the pre-boundary lengthening effect was confirmed for all three languages, although the magnitude varies cross-linguistically. The observations presented in this chapter do not fully support the hypothesis that segments in Japanese should be lengthened less than in Chinese and English because mora plays a vital role in its prosodic organization. It is true that when the segment is not directly adjacent to a prosodic boundary, the effect is largely attenuated as in IP-final positions. However, when the segment is directly adjacent to a boundary, like those in list-final positions, the lengthening effect is comparable to that in English, in which mora is not considered a prosodic unit. In addition, the effect of tone-intrinsic duration on the segment in Chinese might have constrained preboundary lengthening. Further research awaits to fully understand the interplay of tone, segment, and prosodic lengthening. The results also indicate that TVS is not as lengthened as monophthongs.

Chapter 4

Analysis of formant excursions

In this chapter, I will analyze the formant excursions, focusing on the interval from 20% to 80% of the TVS. The analysis will be performed using Generalized Additive Mixed-effect Models.

4.1 Chinese

The overall distribution of Chinese /ai, au, ou/ formants is shown in figure 4.1.

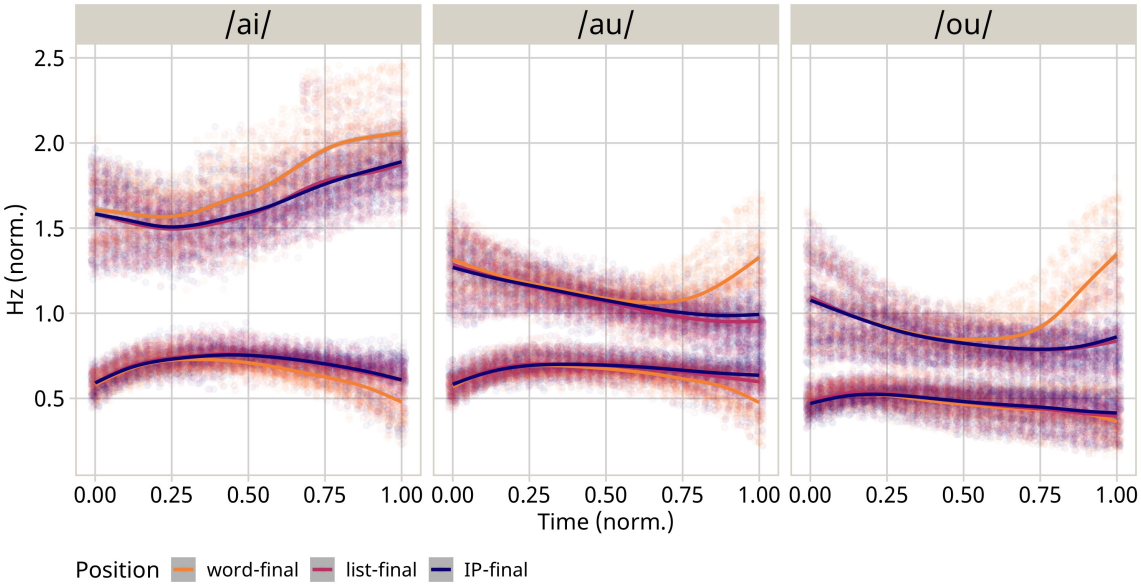

Figure 4.1: The formant excursion of Chinese TVSs.

The most significant effect of prosodic position is the F2 raising and F1 lowering effect in the TVSs. The F2 overall is higher and F1 lower in the word-final positions than in the list-final and IP-final positions. Significantly, in /ai/, the F2 is entirely raised compared to that in list-final and IP-final positions. The F1 lowering effect does not seem as significant in /ou/ as in /ai/ and /au/. The results of the analyses will be reported separately for each TVS. Summary tables of the models are found in the appendix.

4.1.1 /ai/

The GAM analysis revealed that the full models of F1 and F2 both had lower AIC than the nested model (The AIC difference between `full` and `nested` models for F1 is -4224.59, for F2 -4063.47). This indicates that adding prosodic position as a factor did improve the model fit. The results of Wald tests on the parametric terms showed that prosodic `Position` has a significant effect on the intercepts of both F1 ($F = 5.69$, $p < 0.005$) and F2 ($F = 28.9$, $p < 0.005$). Figure 4.2 shows that overall, F1 is lower, and F2 is higher in word-final positions than in the other two positions. The F2 is higher in word-final positions than in the other two prosodic contexts. The F1 showed the opposite trend, which is lower in word-final positions.

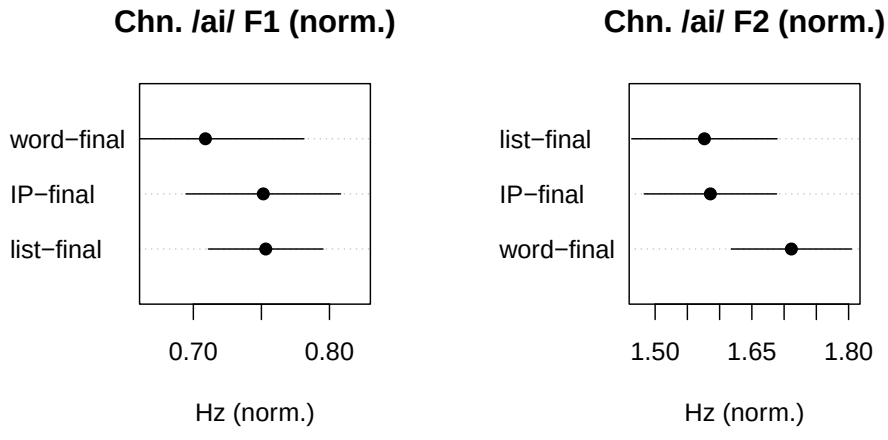

Figure 4.2: Difference in the intercepts of F1 and F2 of Chinese /ai/.

Figure 4.3 shows the estimated smooths of F1 and F2 of Chinese /ai/.

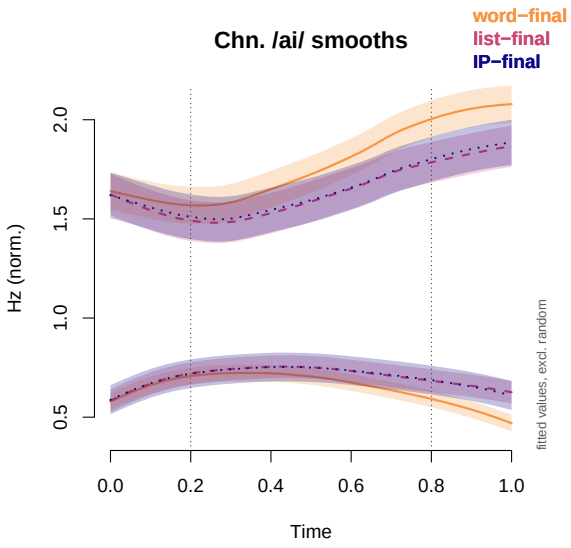

Figure 4.3: Non-linear smooths (summed effects) of F1 and F2 of Chinese /ai/ for 'word-final' (orange), 'list-final' (purple) and 'IP-final' (navy) positions. The pointwise 95% confidence intervals are shown by shade. The vertical lines show the boundary of the 20% and 80% into the vowel.

From the beginning (0%) to the point slightly later than the 20% line, the F2 falls while the F1 rises in /ai/. This suggests that the tongue first moves downward and backward a little bit to reach the target for the initial /a/. Following are a prolonged increase in F2 and a decrease in F1 until the end. F1 and F2 change slightly (steeper slopes) in word-final positions than in list-final and IP-final positions. The formant excursions showed little or no difference between the list-final and IP-final positions.

The estimated difference smooths are shown in 4.4. The difference in the formants between the word-final position and the other two prosodic contexts is confirmed in the graph. The upper three graphs demonstrate that the F2 is higher in word-final positions than in list- and IP-final positions in Chinese /ai/for the most part in the TVS. However, the difference between the list- and IP-final position on F2 is insignificant. As for F1, the trend is the opposite: F1 is lower in word-final positions. However, the difference in F1 (starting from later than 40% and 80%) emerges later than in F2 (starting from around 20%).

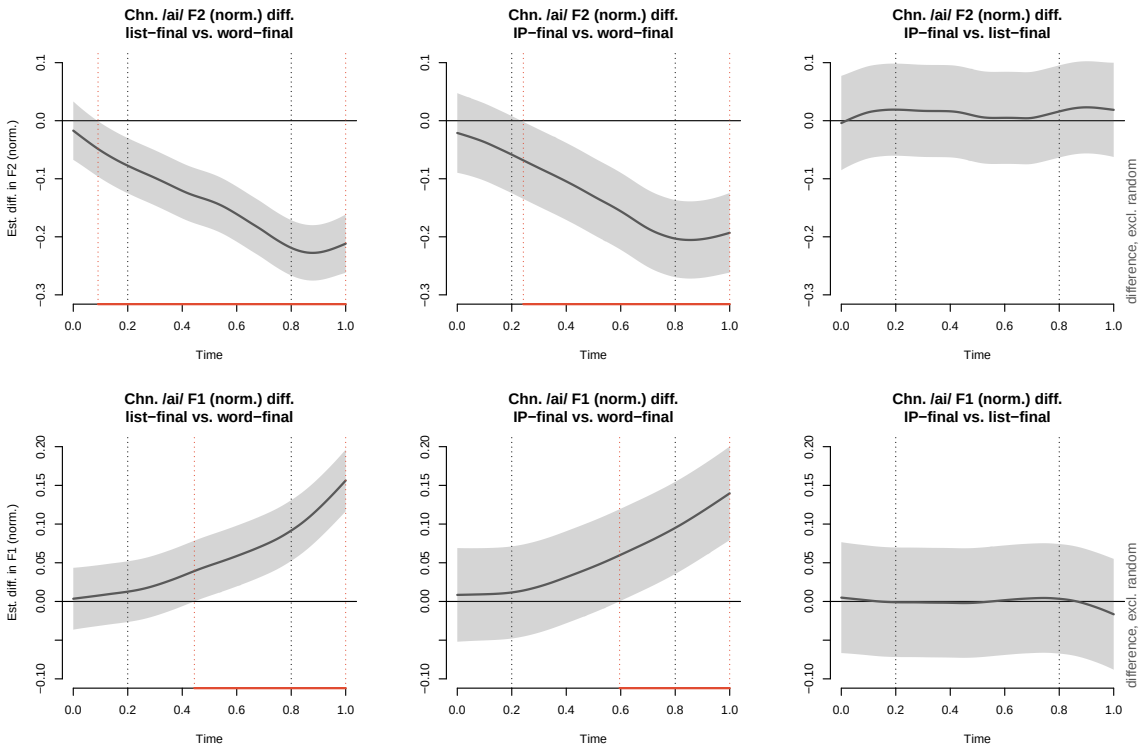

Figure 4.4: Difference between the three smooths comparing 'word-final', 'list-final', and 'IP-final'. The upper three graphs show the difference in F2 and the lower three graphs F1. The pointwise 95%-confidence interval is shown by a shade. When the shaded confidence band does not overlap with the horizontal line 'y=0' (i.e., the value is significantly different from zero), this is indicated by a red line on the x-axis (and vertical dotted lines).

The result suggests that Chinese /ai/ moves toward a higher front position in the vowel space (as F1 and F2 are inversely correlated with tongue height and frontness) in word-final positions. This is probably due to the anticipatory coarticulation with the following coronal consonant in the speech material. The word-final position undergoes more coarticulation than the other two positions. When produced in list-final and IP-final positions, which are also pre-pausal, the TVS is more resistant to anticipatory coarticulatory influence. The time normalized F1 and F2 movements are almost identical in list-final and IP-final positions.

However, the F1 is not so affected in the first half of Chinese /ai/, suggesting that the tongue height for the initial target of Chinese /ai/ is less affected by the prosodic position than tongue frontness.

4.1.2 /au/

As for Chinese /au/, the result of the model comparison shows that the full models of F1 and F2 both had lower AIC than the nested model. The AIC difference between `full` and `nested` for F1 model is -2648.80, for F2 model -4063.47. The result of Wald tests on the parametric term also showed that prosodic `Position` has a significant effect on the intercepts of both F1 ($F = 5.228$, $p < 0.01$) and F2 ($F = 19.82$, $p < 0.005$). This main effect of `Position` on the overall F1 and F2 is illustrated in 4.5.

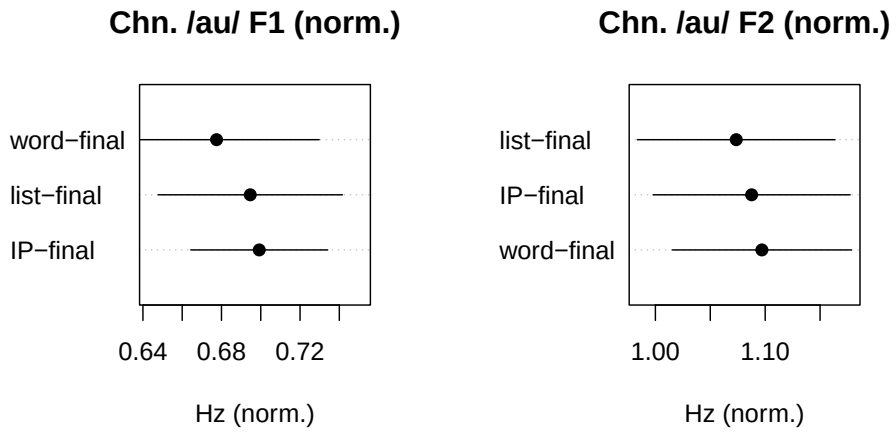

Figure 4.5: Difference in the intercepts of F1 and F2 of Chinese /au/.

F1 is lower, and F2 is higher in word-final positions than in the other two prosodic positions. This is the same trend seen in Chinese /ai/ to a lesser extent. The estimated smooths of F1 and F2 in Chinese /au/ are shown in figure 4.6.

It is shown in figure 4.6 that when /au/occurs in word-final positions, the F2 starts to rise after it reaches the minima around 60-70% into the vowel. F2 in list-final and IP-final positions do not exhibit this final raising. The overall F1 curve reaches its maximum and slowly decreases in the rest of the TVS. F1 lowers more in word-final positions. The estimated difference graphs in figure 4.7 demonstrate that the significant difference between the word-final position and the other two positions lies in the last 40% of Chinese /au/. Again, no significant differences between list- and IP-final positions were found.

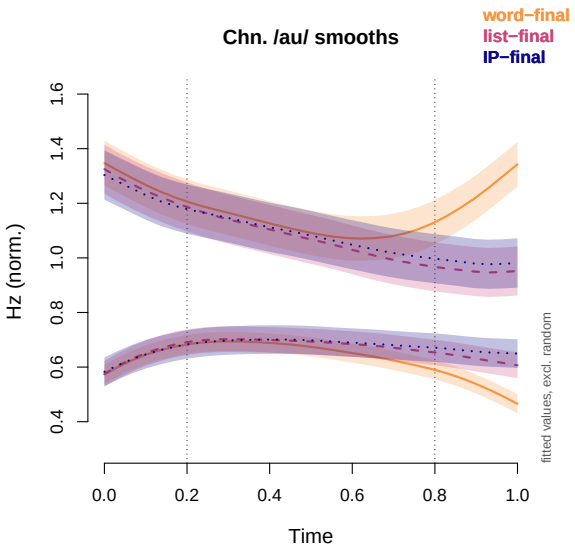

Figure 4.6: Non-linear smooths (summed effects) of F1 and F2 of Chinese /au/.

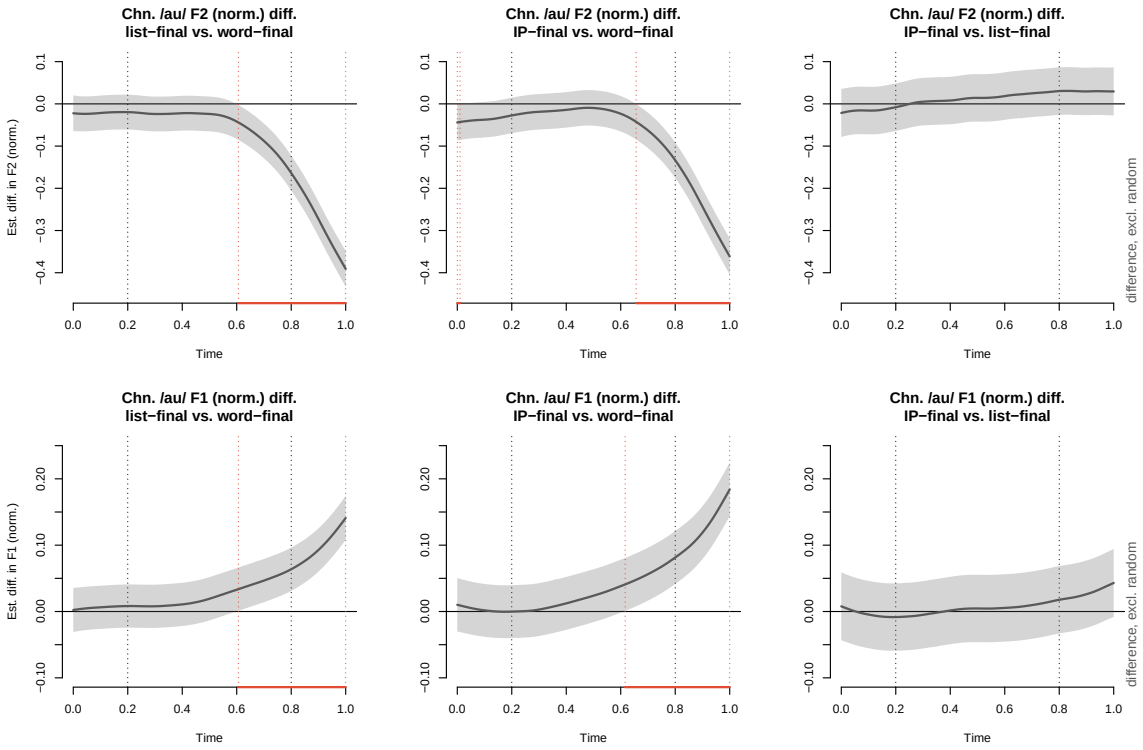

Figure 4.7: Difference between the three smooths comparing 'word-final', 'list-final', and 'IP-final' in Chinese /au/.

The difference smooth graph shows that F2 and F1 difference between the prosodic positions starts to emerge after 60% into /au/. As it approaches the end, the difference

becomes larger and larger. The difference between the list-final and IP-final positions was insignificant during the entire course of the TVS. This result suggests that the tongue moves higher and forward due to the anticipatory coarticulation when producing Chinese /au/, resulting in a higher F2 and a lower F1 in the last 40% in the TVS in the word-final position. However, in the first 60% of the vowel, the formant excursions were identical among the prosodic conditions. The onset of the coarticulatory effect on the formants is not as early as in /ai/. This might be because the second target of /au/ is incongruent with the following coronal sound. The production of /au/ requires the tongue to move upward and backward (low F1 and F2) after producing /a/, while a coronal sound requires the tongue to move upward and forward (high F2 and low F1). If the tongue begins to move up and forward too early, it will hinder the identification of the high back target in the second half of /au/. Henceforth, the coarticulatory effect must be suppressed to some extent to maintain its underlying articulatory and acoustic target. The late onset of the coarticulation on F1 and F2 can be seen as a compromise between the two incongruent tongue movements: one backward and the other forward.

4.1.3 /ou/

For Chinese /ou/, the difference in F1 due to the influence of prosodic positions was not as significant as in /ai, au/. The AIC difference between `full` and `nested` for F1 model is -435.93, for F2 model -4629.19. However, the result of Wald tests on the parametric terms showed that the effect of prosodic `Position` on the intercept of F1 did not reach significance ($F = 1.095$, $p > 0.05$), whereas it did for F2 ($F = 37.16$, $p < 0.005$). This effect is illustrated in figure 4.8.

Although the average F1 (the intercept) seems lower than in list- and IP-final positions, this difference was not found significant. The estimated smooths are shown in figure 4.6. The F2 raising starts from around 60% into the vowel in the word-final positions, whereas the difference in F1 seems negligible.

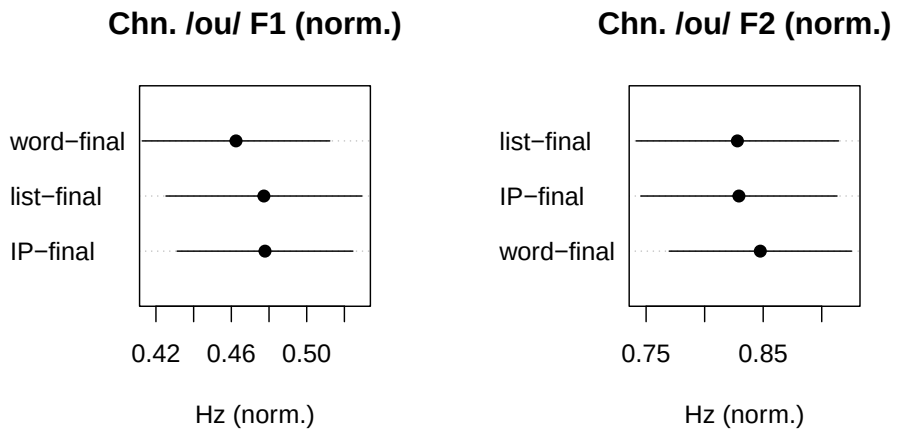

Figure 4.8: Difference in the intercepts of F1 and F2 of Chinese /ou/.

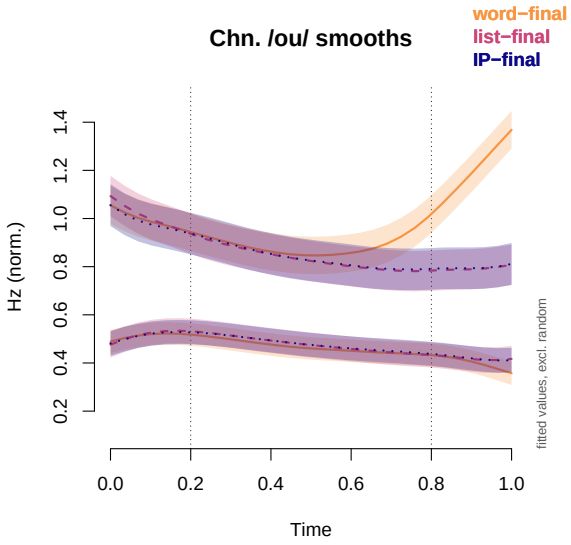

Figure 4.9: Non-linear smooths (summed effects) of F1 and F2 of Chinese /ou/.

When the /ou/ occurs in word-final positions, the F2 rises when it reaches the minimum value at around 60% into the vowel. At the same time, F1 did not show a significant change in word-final positions than in the list- or IP-final positions. The estimated difference graphs in figure 4.10 show that the difference between the word-final position and the other two positions lies in the last 40% of Chinese /ou/. In F1, there is only a minor difference at the end of /ou/, outside the 80% boundary.

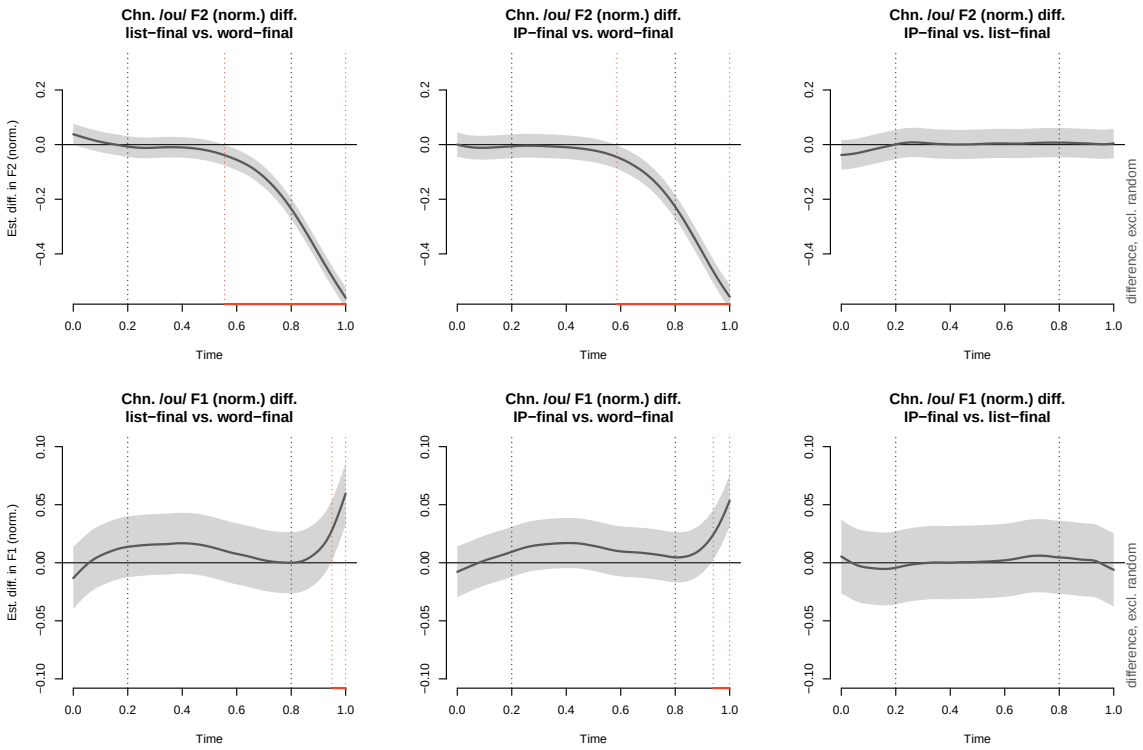

Figure 4.10: Difference between the three smooths comparing 'word-final', 'list-final', and 'IP-final' in Chinese /ou/.

4.1.4 Interim discussion: Chinese TVS and prosody

The results of GAM analysis on Chinese TVS suggest that prosodic context influences the acoustics of the vowel sequences /ai, au, ou/. Overall, the most significant differences were found in the last 60% of the vowel and the movement of F2. For all of the three TVSs in Chinese, F2 was more affected than F1, especially for /ou/; the difference in F1 was negligible. In the central 60% of the vowel, a difference in F1 movement was not found. The general pattern is that, under the influence of the following coronal consonants, the F2 tends to be higher. At the same time, F1 is lower in word-final positions than in the other two prosodic positions approaching the end of TVS. This is the result of anticipatory coarticulation. As the tongue starts to move upward and forward in the last half of the TVS, the F2 is raised (tongue moving forward), and the F1 is lowered (tongue moving upward). The first half of the TVS was only affected in Chinese /ai/. The difference in F2 of /ai/ starts

to emerge as early as before 20% into the vowel. The F1 difference in /ai/ also begins to appear in the first half of the vowel, earlier than /au, ou/. The difference between list-final and IP-final positions was not significant.

In the word-final positions, the difference emerged earlier in /ai/ than /au, ou/ because the following sound environment with a coronal consonant is congruent with /ai/ yet incongruent with /au, ou/. In the production of /ai/, the tongue moves upward and forward to reach the target area in the last half of the TVS. This is due to the gestural requirement of producing a coronal sound. The tongue tip needs to move into the high front region of the oral cavity. This early movement into the following coronal consonant results in higher F2 and lower F1 from the first half of /ai/. The production of /au, ou/, however, requires the tongue body to first move back into the dorsal position before continuing to move forward to produce the coronal sound. If the tongue starts to move for the following consonant too early, then the phonetic identity of /au, ou/ would be obscured. It may hinder the recognition of the speech sound by the listeners. This can be broadly seen as a consequence of contrast preserving, a listener-oriented speech production strategy.

However, there is an alternative approach to account for the patterns observed. In the π-gesture model proposed in Byrd and Saltzman (1998, 2003), gestures coordinated in time are produced with shorter temporal intervals of gestural activation at lower boundaries than at higher boundaries due to the influence of π-gesture that slows down the clock of gestures. This is called gestural blending at lower boundaries. For example, Krivokapić (2007) found that the consonant gestures are coordinated closer to each other in the vicinity of lower boundaries. In this regard, gestural blending at lower boundaries can also account for the patterns in the word-final position. The gesture of moving the tongue back to reach the final /-u/ target and the gesture of moving the tongue forward to make contact for the coronal sound following the target TVS were produced with greater blending, resulting in the tongue fronting gesture starting earlier in word-final positions than in the other two prosodic contexts. Consequentially, the F2 rises at the end of /au, ou/. The earlier F2

raising in /ai/ then can be seen as the consequence of even more gestural blending because the gestural movement for /-i/ in /ai/ and the following coronal consonants are spatially congruent with each other. Therefore, the tongue can start moving toward the high front region in the oral cavity, consequently raising F2 earlier in time.

Regarding the difference in the F1 patterns between /au/ and /ou/, gestural blending can also account for it. In the course of /au/ and /ou/, the tongue does not end in the same position. While their second target vowels are transcribed the same as /u/ in a broad transcription or [ʊ] in a narrow transcription, the final target of /ou/ is higher than /au/ in space. Hence /ou/ is produced with a higher tongue position and a lower F1 than /au/ because the gesture of producing /o-/ is spatially more congruent with the gesture of the following coronal consonant than that of /a-/. This effect is more evident in the word-final positions. There is more gestural blending in word-final positions.

4.2 English

The overall distribution of English /ai, au, ou/ formants is shown in figure 4.1.

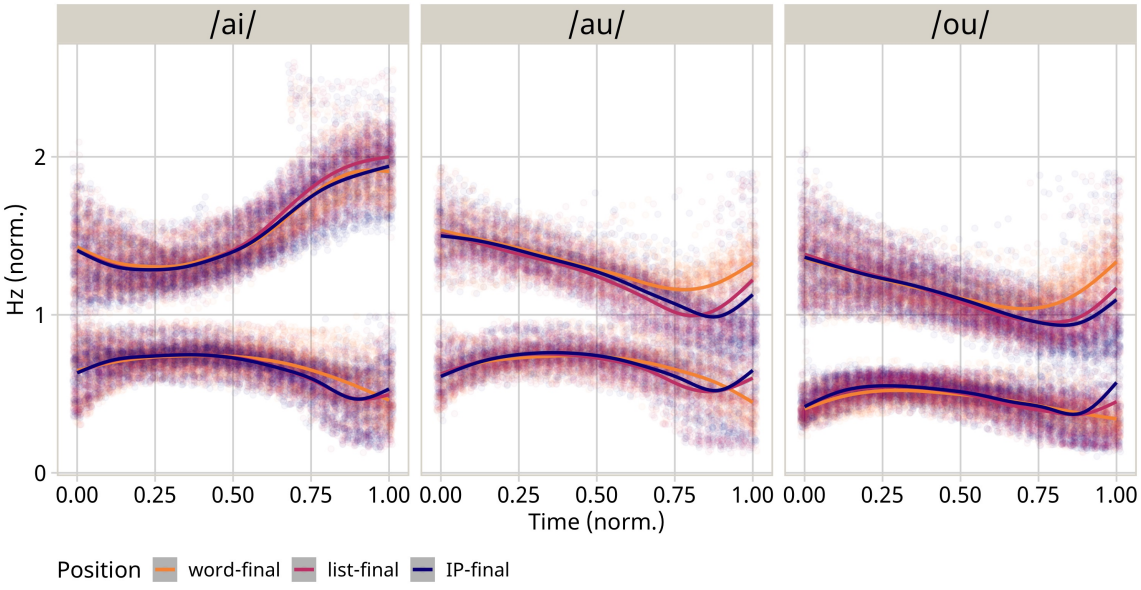

Figure 4.11: The formant excursion of English TVSs.

The effect of prosodic position in English is not as large as in Chinese. The formant curves of English /ai/ seem to be overlapped with each other. There seem to be minor effects in /au, ou/ at the end of the vowel sequences. Such subtle non-linear difference is precisely what GAM is good at modeling. The result to be reported below will show that there are minor effects of prosodic modulation in all three TVS in English.

4.2.1 /ai/

While the model comparison suggest that the `full` models of F1 and F2 have lower AIC values (F1 model difference: -3309.43, F2 model difference: -2666.37), the results of Wald tests on the parametric terms showed that prosodic `Position` did not have any significant effects on the intercepts of either F1 ($F = 0.608$, $p > 0.05$) or F2 ($F = 0.77$, $p > 0.05$).

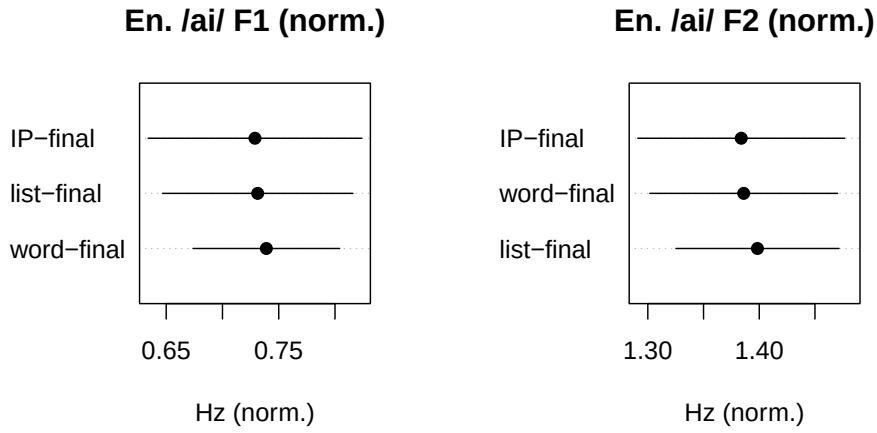

Figure 4.12: Difference in the intercepts of F1 and F2 of English /ai/.

The average of F1 and F2 seems not to distinguish from each other in different positions in figure 4.12.

From the estimated smooths in figure 4.13, there is hardly any noticeable difference in the formant excursions from the three prosodic positions for English /ai/. It seems that F2 is higher at the end in list-final positions, but F1 is higher in word-final positions. The estimated difference graphs in figure 4.14 show that the difference was primarily found in

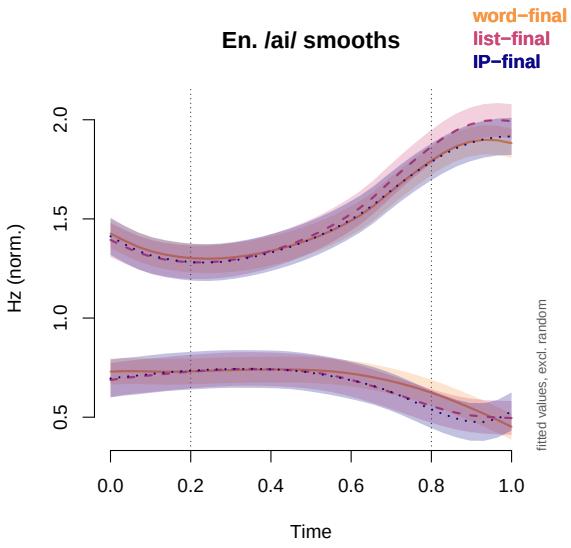

Figure 4.13: Non-linear smooths (summed effects) of F1 and F2 of English /ai/.

F2 in list-final positions.

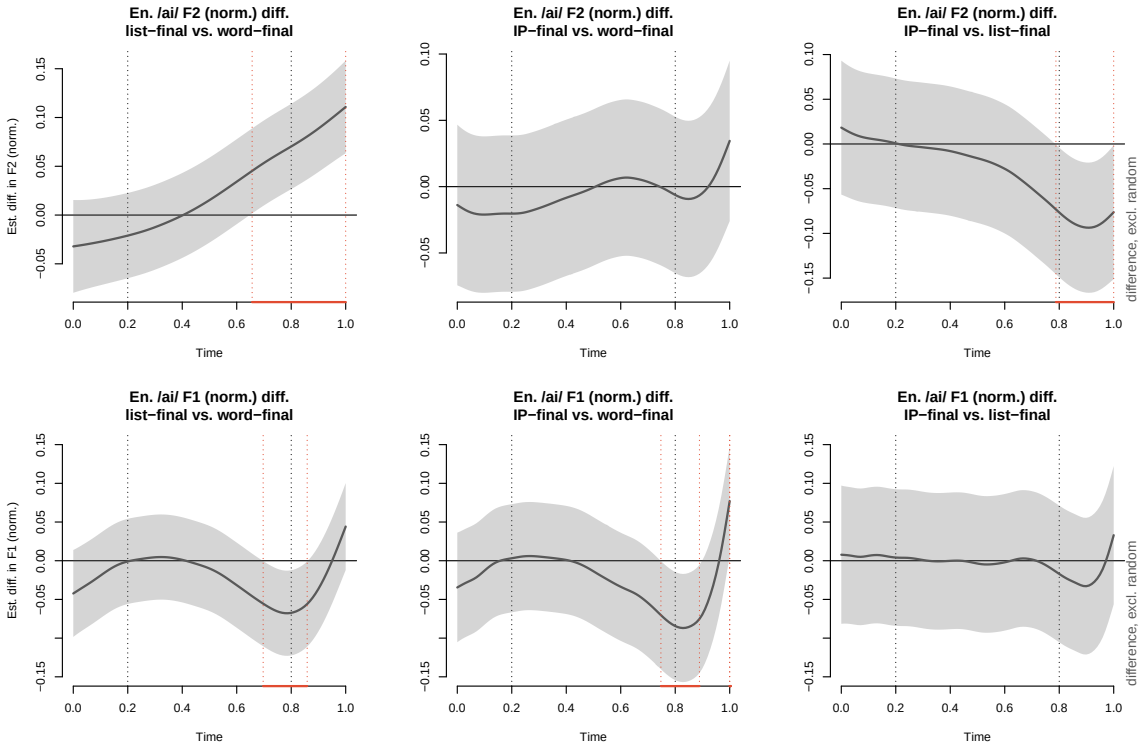

Figure 4.14: Difference between the three smooths comparing 'word-final', 'list-final', and 'IP-final'. The upper three graphs show the difference in F2 and the lower three graphs F1.

The figure confirms that the F2 is higher in list-final positions than in word-final and

IP-final positions. F2 is higher in list-final positions in the last 35% comparing list-final to word-final positions and in the final 20% comparing list-final with IP-final positions. The difference between the word-final and IP-final positions on F2 is not significant. F1, however, is higher in word-final positions for a short span of around 80% into the vowel but does not last until the very end. There was no significant difference in the F1 curves between IP-final and list-final positions.

Overall, the difference induced by prosodic contexts in English /ai/ is much smaller than that in Chinese. The F1 was barely lowered in word-final positions, and the F2 was raised only at the end of the vowel in list-final positions. This result suggests that the following coronal consonant did not influence the vowel sequence in any of the prosodic contexts, including word-final positions.

4.2.2 /au/

For English /au/, model comparisons suggest that the `full` models of F1 and F2 have lower AIC values (F1 model difference: -2666.37, F2 model difference: -4365.45), indicating that prosody does influence both the formants. The results of Wald tests on the parametric terms showed that prosodic `Position` did not have any significant effects on the intercept of F1 ($F = 0.023$, $p > 0.05$), but the intercept of F2 was affected ($F = 16.34$, $p < 0.005$).

The average of F1 seems not to differ for prosodic positions in figure 4.15, whereas F2 is higher in word-final and IP-final positions than list-final positions.

The estimated smooths in figure 4.16 show that in the word-final positions, the F2 is higher in the last half of /au/. There is a relatively small raising effect of F2 in the final 20% of the vowel between IP-final and list-final positions. In the last 30% of the vowel sequence, the F1 curve shows a smooth lowering trajectory in the word-final position, whereas in the list- and IP-final positions, it shows a more complicated movement with a dipping-raising trajectory. The estimated difference graphs in figure 4.17 show that the difference was primarily found in F2 in list-final positions.

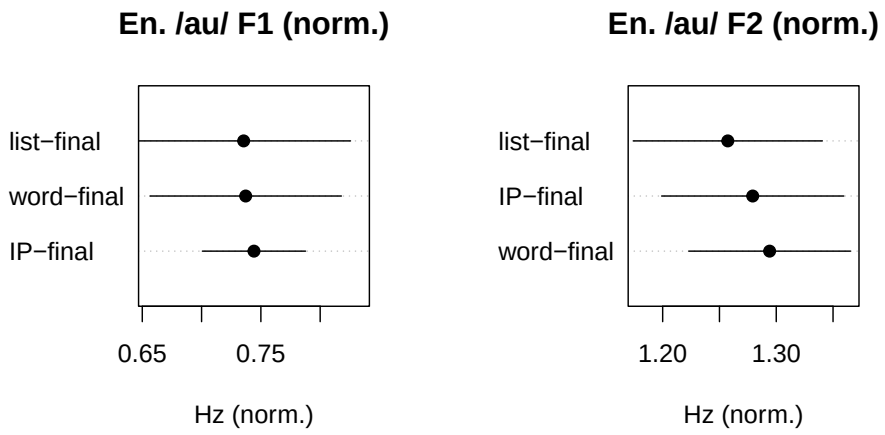

Figure 4.15: Difference in the intercepts of F1 and F2 of English /au/.

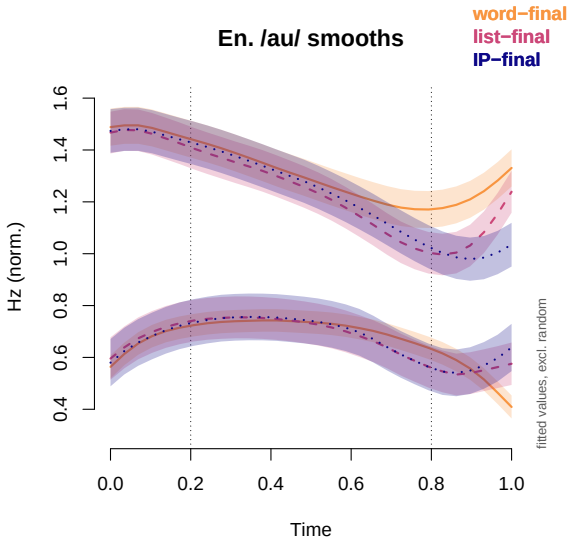

Figure 4.16: Non-linear smooths (summed effects) of F1 and F2 of English /au/.

The upper three graphs show that the F2 is higher in word-final positions than in list-final and IP-final positions in English /ai/. This difference extends well beyond the last 20% of the vowel sequence (the final 50% comparing list-final and word-final positions and the final approximately 40% comparing IP-final to word-final positions). Although there is a significant difference between IP-final and list-final positions, it did not happen in the vowel sequence's central 60% (the 20% to 80% interval). The prosodic context showed little effect on the F1 of English /au/. F1 is lower in word-final positions than the other two

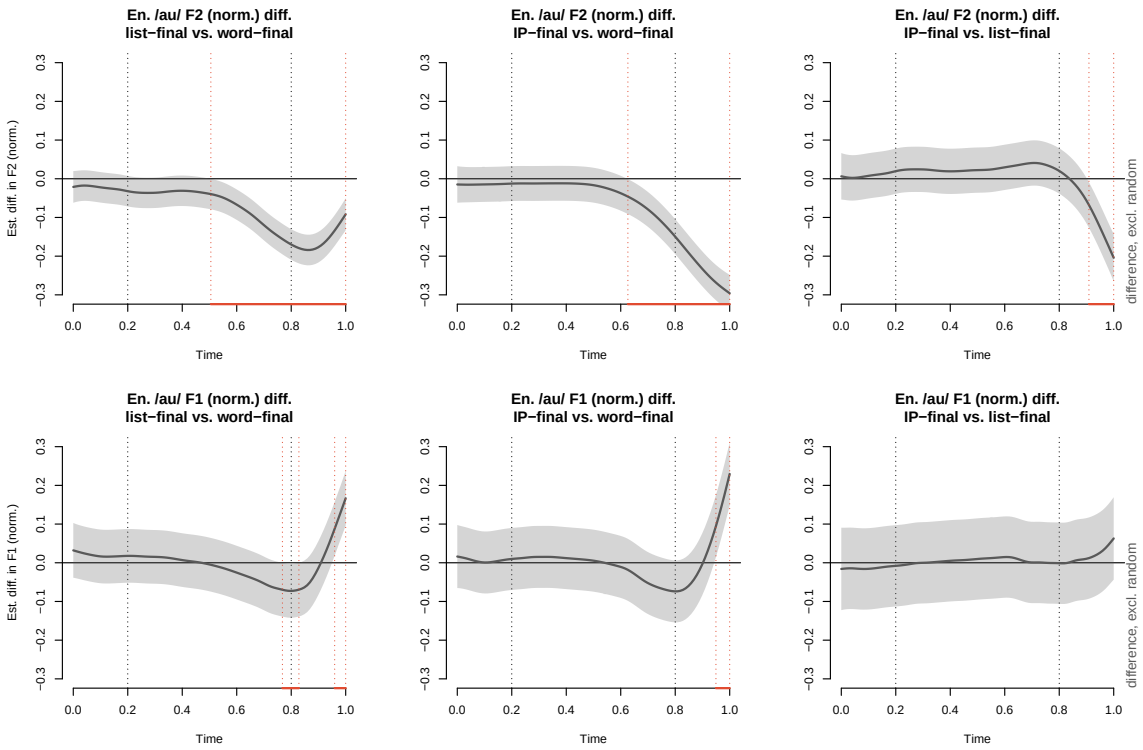

Figure 4.17: Difference between the three smooths comparing 'word-final', 'list-final', and 'IP-final'. The upper three graphs show the difference in F2 and the lower three graphs F1.

contexts, but it only lasted for a concise duration.

The results above show that English /au/ moves higher and more front in the vowel space when in word-final positions. The TVS is less resistant to anticipatory coarticulation when produced in word-final positions due to the following consonant. The time normalized F1 and F2 movements are almost identical in the list- and final positions.

4.2.3 /ou/

For English /ou/, model comparisons suggest that the `full` models of F1 and F2 have lower AIC values (F1 model difference: -4828.26, F2 model difference: -3833.24), indicating that the prosody does influence both of the formants. The results of Wald tests on the parametric terms revealed that neither the intercept of F1 nor that of F2 was influenced by the prosodic `Position` (F1: $F = 0.225$, $p > 0.05$; F2: $F = 1.885$, $p > 0.05$). The intercepts of F1 and F2

in each of the prosodic contexts are shown in figure 4.18.

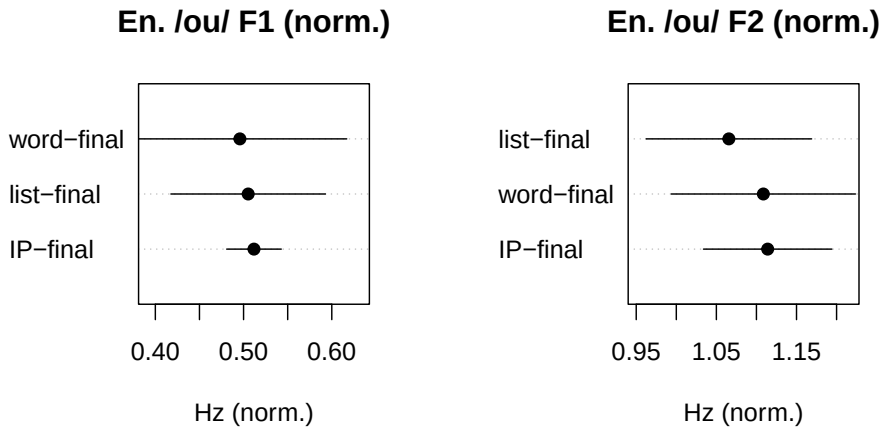

Figure 4.18: Difference in the intercepts of F1 and F2 of English /au/.

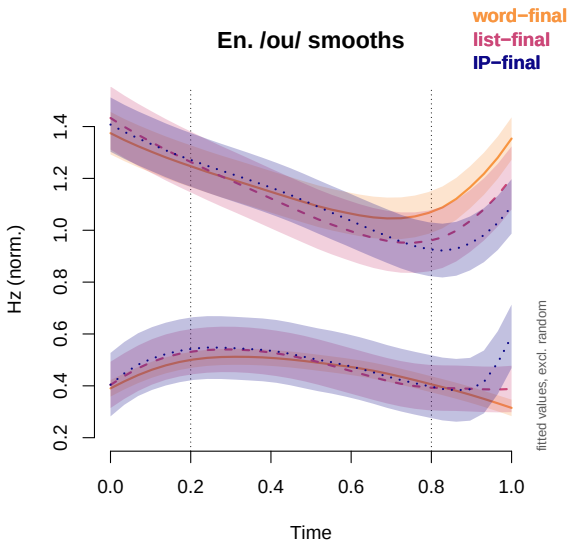

Figure 4.19: Non-linear smooths (summed effects) of F1 and F2 of English /ou/.

The estimated smooths in figure 4.19 show that the formant excursions of English /ou/ in the three prosodic contexts are almost indistinguishable. The only difference seems to come from the last 20% of the vowel, in that F2 in the word-final position is higher while F1 in the list-final position is higher at the end.

In the difference smooths in figure 4.20, the F2 results demonstrated that the F2 is

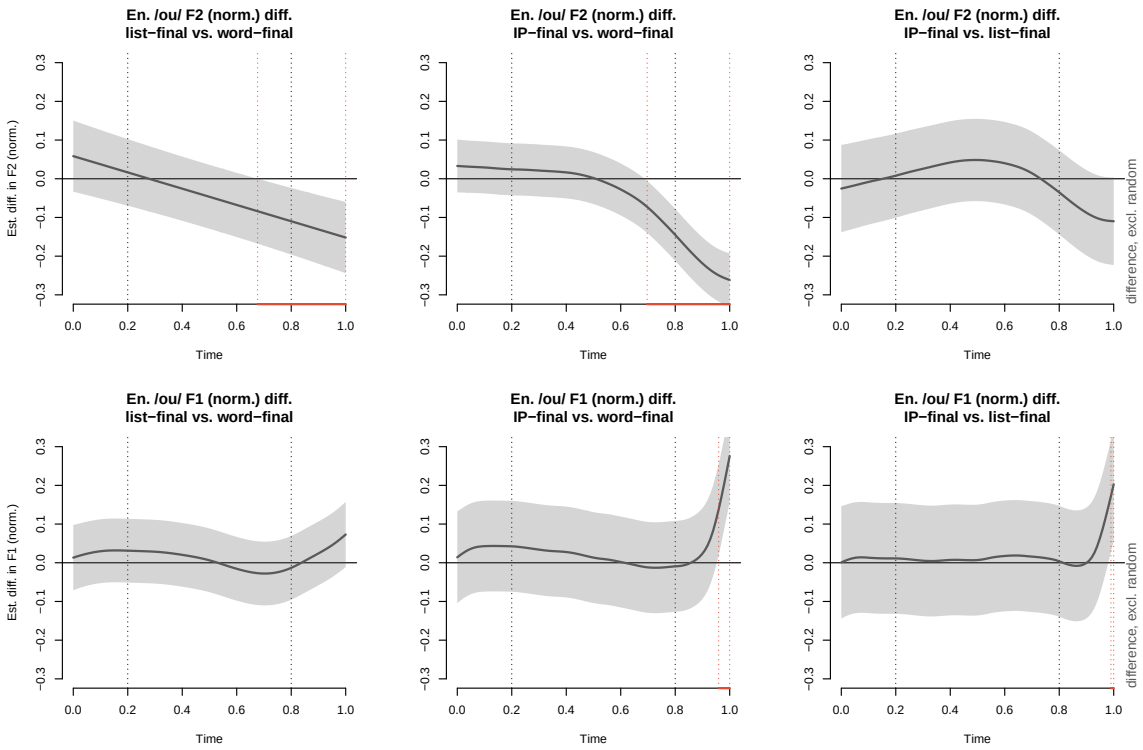

Figure 4.20: Difference between the three smooths comparing 'word-final', 'list-final', and 'IP-final'. The upper three graphs show the difference in F2 and the lower three graphs F1.

higher in word-final positions than in list-final and IP-final positions. The difference extends from around 70% of the vowel to the end of the vowel. No difference was found in the comparison of IP-final and list-final positions. As for F1, although there are some differences in the comparisons of the IP-final position and list-/word-final positions, they only exist in the vowel's last part. This difference is not meaningful because of two reasons. First, it did not affect the central part of /ou/ (the 20%-80% interval). Second, the measures of formants might not have been accurate at the boundary of the TVS in the two pre-pausal contexts since the speakers may have used non-model phonation such as creaky voice or breathy voice. The F1 and F2 are so close that the LPC algorithm used in formant extraction could not track the formants well.

From the results presented above, similar to the result in Chinese TVS, English /ou/ also moves more front in the vowel space in the final portion when in word-final positions. This is also due to the influence of the following coronal consonants. The TVS is less resistant

to anticipatory coarticulation when produced in word-final positions due to the following consonant. The time normalized F1 and F2 movements are almost identical in the list- and final positions.

4.2.4 Interim discussion: English TVS and prosody

The influence of prosodic context on English TVS is not as large as in Chinese. The difference in the formant excursions induced by the prosodic positions is much smaller than that in Chinese. The differences all needed to be confirmed in the difference smooth graphs. This is a sharp contrast between English and Chinese. However, similarities emerge after taking a closer look at the difference in formants in the three prosodic contexts. First, similar to Chinese /ai, au, ou/, F2 is more influenced by the prosody, implying that the tongue frontness is more easily affected by the prosody rather than the tongue height. Secondly, the second half of TVS is more likely to be influenced by the position than the first half. This result is partially in line with previous studies like Gay (1968) in which variations of diphthong production under different speech rates were found in the variability of the final F2 target instead of the initial F2 target or F1. Thirdly, most differences were found in the comparisons that involve word-final position. The difference between list-final and IP-final positions was either insignificant or negligible. The only exception is the F2 of English /ai/, where the differences in the comparisons involving list-final positions were significant while that between word-final and IP-final positions were not. This trend suggests that, in acoustics, the time-normalized formant excursions are more susceptible if the target vowel is immediately before a pause. If there is a pause after the vowel sequence, the formants would have more time to reach the underlying target. Suppose another sound follows the vowel sequence. In that case, the anticipatory coarticulatory effect is likely to cut off the TVS, and production of the following sound would take place earlier, leading to greater anticipatory coarticulation.

4.3 Japanese

The overall distribution of formants of English /ae, ai, au/ is shown in figure 4.21.

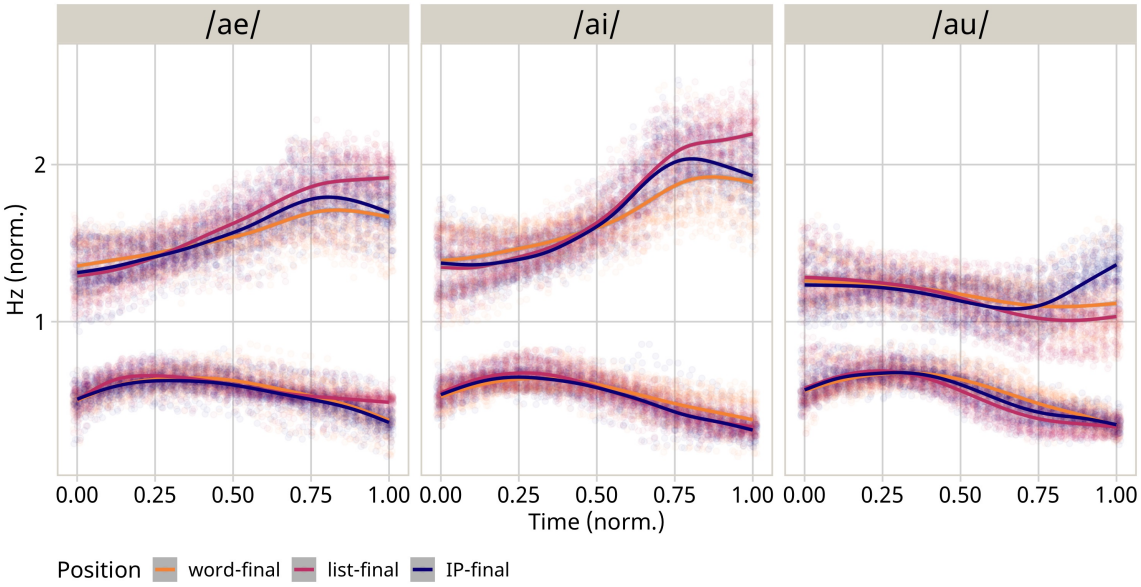

Figure 4.21: The formant excursion of Japanese TVSs.

It looks like prosodic context influences the movement of the formants, especially F2. the F2 ends lower in /ae/ than in /ai/. In /ae, ai/, the F2 is highest in list-final positions rather than in word-final positions, but in /au/, the highest final F2 is found in IP-final positions.

4.3.1 /ae/

The model comparisons suggest that the `full` models of F1 and F2 have lower AIC values (F1 model difference: -1396.68, F2 model difference: -1855.65), indicating that prosody does influence both of the formants. The results of Wald tests on the parametric terms revealed that the prosody did not significantly affect the intercept of F1 ($F = 0.994$, $p > 0.05$), but the intercept of F2 was ($F = 10.87$, $p < 0.005$). The intercepts of F1 and F2 in each of the prosodic contexts are shown in figure 4.22.

The estimated smooths in figure 4.23 show that the F2 movements of /ae/reach higher in list-final positions than in the other two contexts. The difference in F1 is not so noticeable

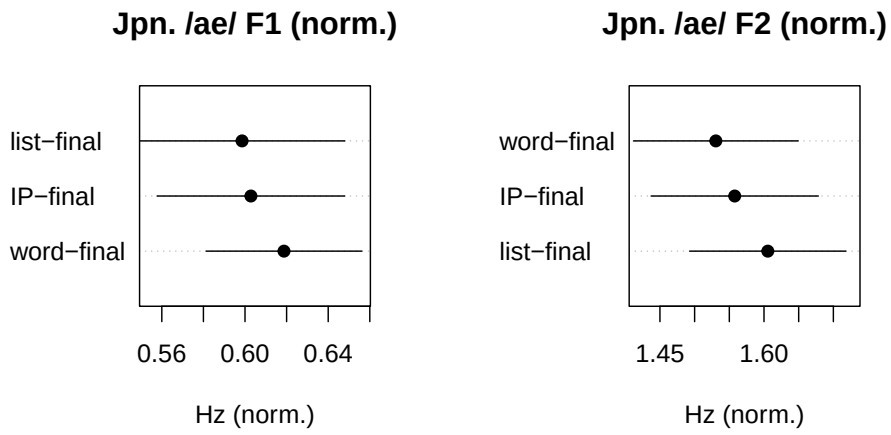

Figure 4.22: Difference in the intercepts of F1 and F2 of English /au/.

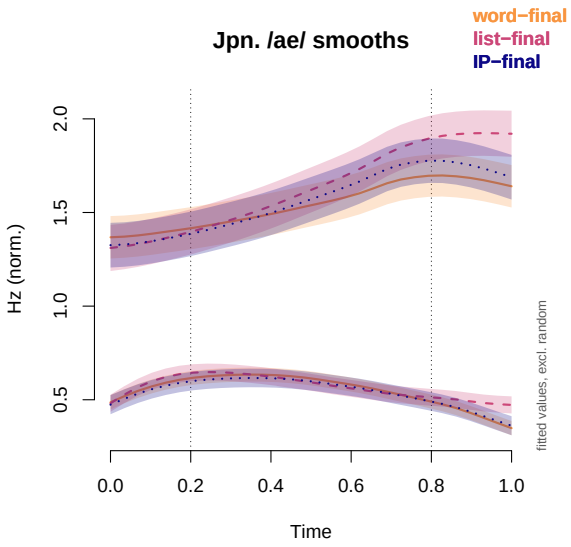

Figure 4.23: Non-linear smooths (summed effects) of F1 and F2 of Japanese /ae/.

from the estimated overall smooth graph.

In the difference smooths in figure 4.24, the F2 results demonstrated that the F2 is lower in word-final positions than in list-final and IP-final positions. The difference extends from around 40% of the vowel to the end of the vowel in comparing list-final and word-final positions and 55% in comparing IP-final and word-final positions. There is also a difference between IP-final and list-final positions in that F2 is lower than in list-final positions. In F1, the difference was found in the comparisons involving the list-final position. Comparing

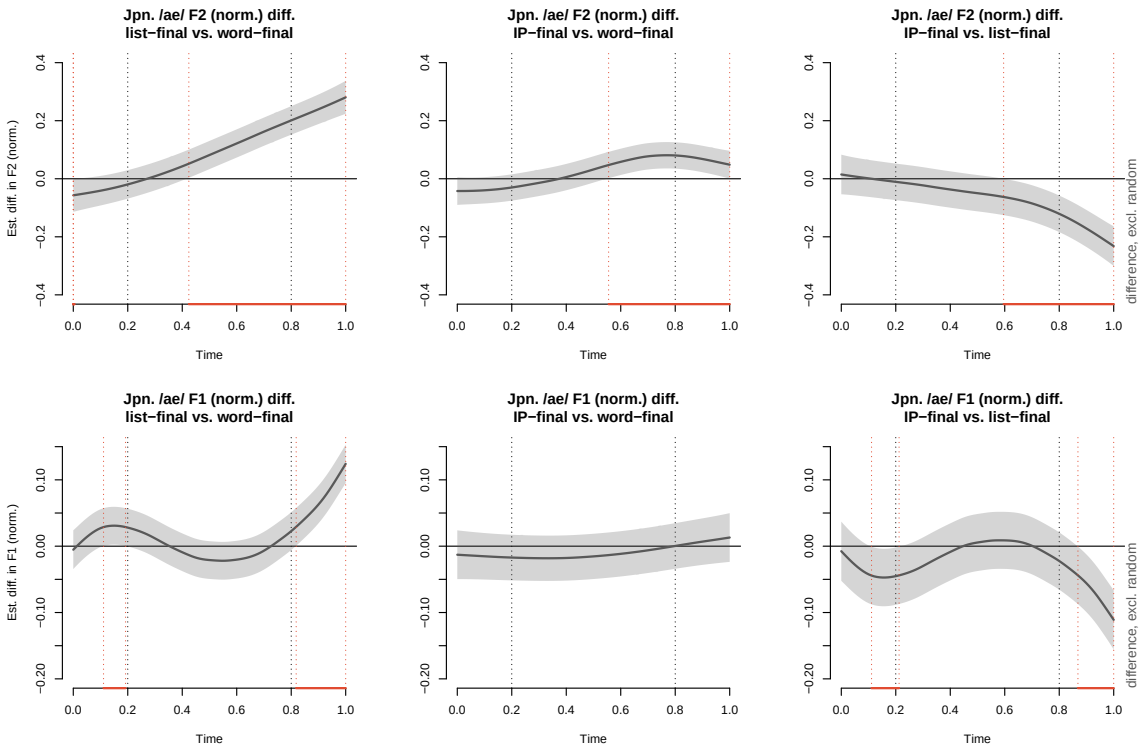

Figure 4.24: Difference between the three smooths comparing 'word-final', 'list-final', and 'IP-final'. The upper three graphs show the difference in F2 and the lower three graphs F1.

list-final to word-final positions, the F1 is higher in the last 20% but lower when comparing IP-final to list-final positions. No difference was found in the F1 difference smooths between IP-final and word-final positions.

From the results presented above, the most formant excursion in Japanese /ae/was found in list-final positions, especially for F2 in the last 40% of the vowel sequence. This means that the final tongue position of Japanese /ae/ is more fronted in the list-final position than the other two prosodic contexts. On the other hand, tongue height was not so influenced. This might be because /ae/ in the list-final position was produced with the longest duration (see the analysis in chapter 3). Most of the difference in the F1 was found outside the 20%-80% interval, which was not so meaningful compared to the movement within the 20%-80% boundary.

4.3.2 /ai/

The model comparison results suggest that the `full` models of F1 and F2 captured more variance in the data (F1 model difference: -1539.60, F2 model difference: -1745.78). The results of Wald tests on the parametric terms showed that prosodic `Position` did not have a significant effect on the intercepts of either F1 (F = 0.558, $p > 0.05$) while it did affect F2 (F = 3.039, $p < 0.05$). The intercepts of formants of Japanese /ai/ in different prosodic contexts are exhibited in 4.25.

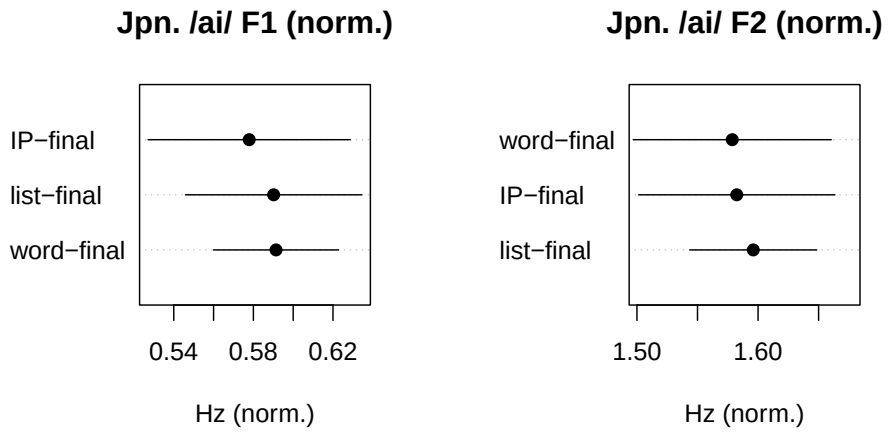

Figure 4.25: Difference in the intercepts of F1 and F2 of English /ai/.

From the estimated smooths in figure 4.26, there is hardly any noticeable difference in the F1 movements from the three prosodic positions for Japanese /ai/. F1 seems to be higher in the last 60%. In the onset of the /ai/, F2 appears to be higher in word-final positions, whereas in the offset, F2 is higher in list-final positions. The estimated difference graphs in figure 4.27 show that significant differences were found in most of the comparisons of the formants.

The figure confirms that the F2 is higher at the end in list-final positions. F2 is higher in list-final positions in the last 45% comparing list-final to word-final positions and in the final 20% comparing list-final to IP-final positions. The difference between word-final and IP-final position also lasted in the final 40% of /ai/. Different from English and Chinese,

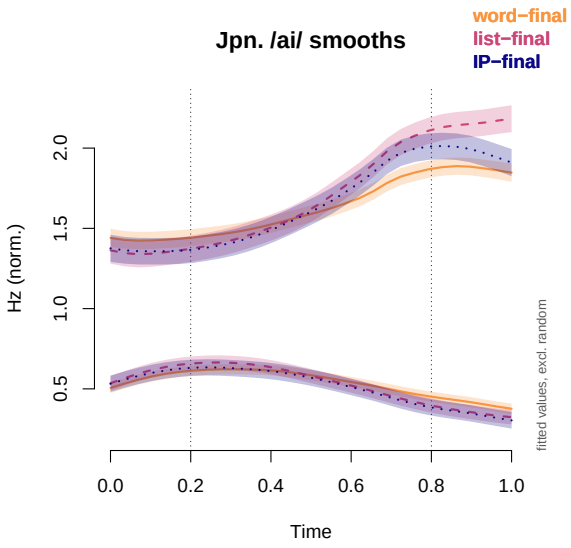

Figure 4.26: Non-linear smooths (summed effects) of F1 and F2 of English /ai/.

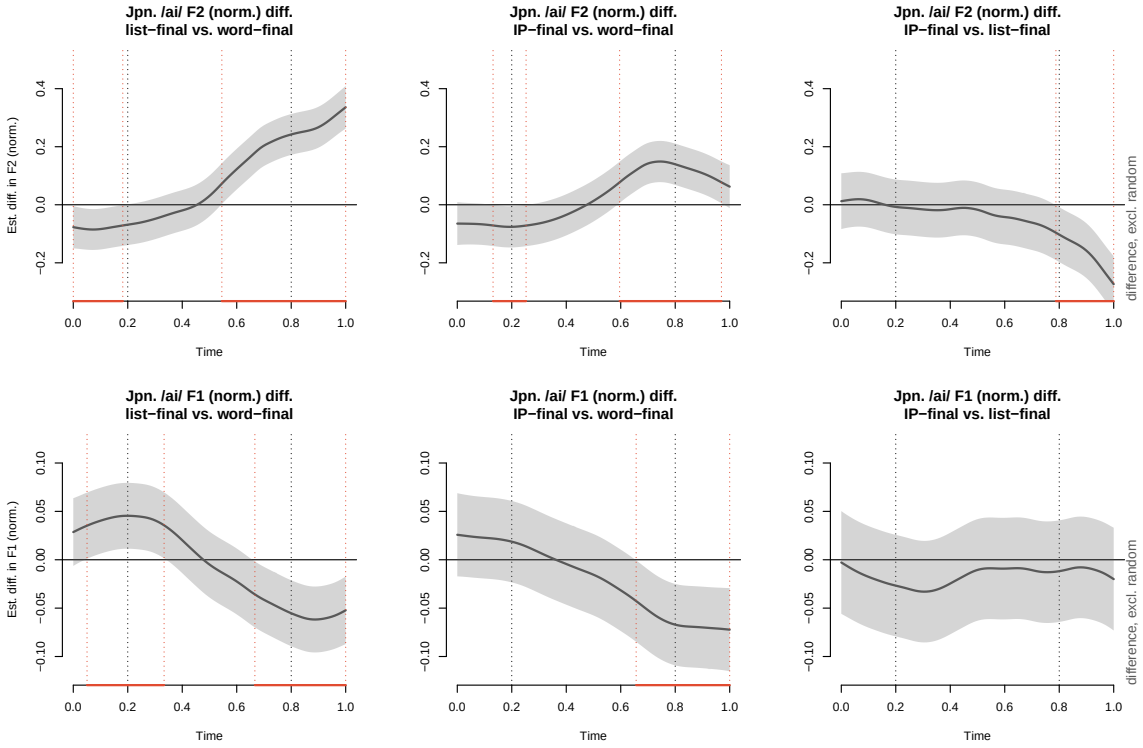

Figure 4.27: Difference between the three smooths comparing 'word-final', 'list-final', and 'IP-final'. The upper three graphs show the difference in F2 and the lower three graphs F1.

there is also some difference at the beginning of the vowel sequence. The F2 is higher in word-final positions at the onset. As for F1, the same trend in F2 was also observed:

the difference was found from both the onset and offset of the vowel when comparing list-final to word-final positions: F1 is higher in the onset while lower in the offset in list-final positions than word-final position. The difference between F1 curves in IP-final and list-final positions was insignificant.

The result combined suggests that upon producing Japanese /ai/, the tongue is more fronted and higher in the onset, but less fronted and lower in the offset in the word-final position than in IP-final and list-final positions. The differences were not trivial to the production of the vowel sequence since most of the difference either extends into or starts from the central 60%.

4.3.3 /au/

The results of model comparisons suggest that the `full` models of F1 and F2 have lower AIC values (F1 model difference: -1368.89, F2 model difference: -2891.56), indicating that prosody does influence both the formants. The results of Wald tests on the parametric terms showed that prosodic `Position` did not have any significant effects on the intercept of F1 ($F = 7.235$, $p < 0.005$), but the intercept of F2 was affected ($F = 2.128$, $p > 0.05$). The difference in intercepts of F1 and F2 in prosodic contexts are shown below.

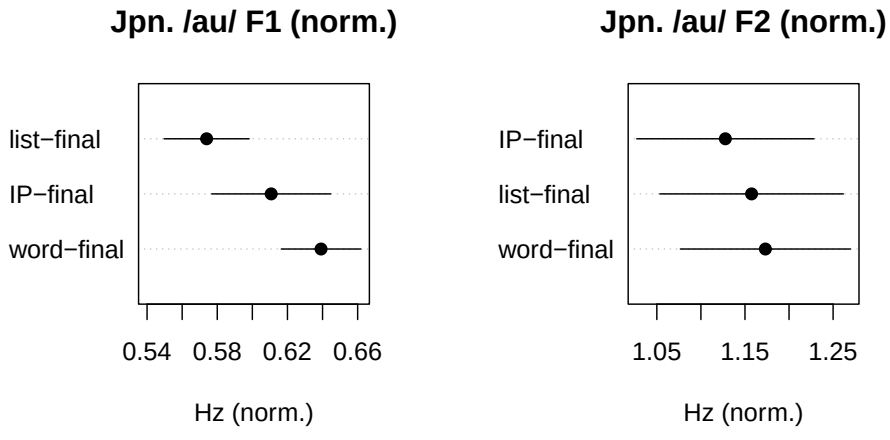

Figure 4.28: Difference in the intercepts of F1 and F2 of English /au/.

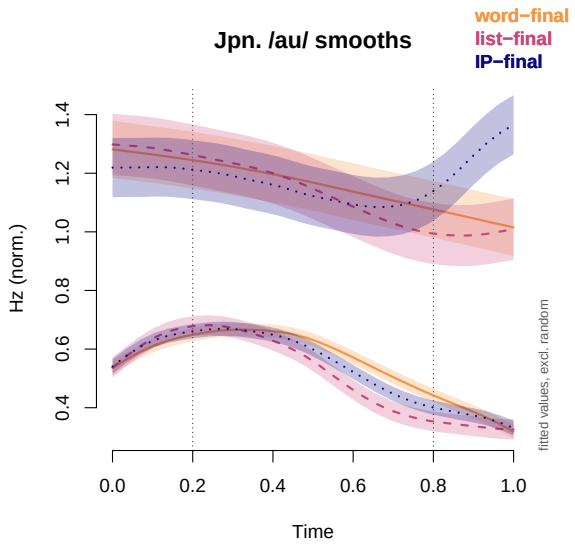

Figure 4.29: Non-linear smooths (summed effects) of F1 and F2 of English /au/.

The estimated smooths in figure 4.29 show that in the IP-final positions, the F2 of Japanese /au/ is higher in approximately the last 40% of the vowel sequence. There is a relatively small raising effect of F1 in the final 50% of the vowel. The F1 is lowest in list-final positions and highest in word-final positions.

The upper three graphs show that the F2 is higher in IP-final positions than in list-final and IP-final positions in Japanese /au/. A difference in the onset was also found when comparing F2 in IP-final and list-final positions and IP-final to word-final positions: the F2 is lower for IP-final positions than in the other two contexts. The F1 difference smooths confirmed the trend seen in the overall smooths. F1 is lower in list-final and IP-final positions but lower when comparing IP-final to list-final positions.

4.3.4 Interim discussion: Japanese TVS and prosody

A prosodic influence on both F1 and F2 of Japanese TVS was confirmed. The general pattern of the prosodic influence on TVS in Japanese is the one found in Chinese and English as well: F2 is more influenced than F1, and the offset of F2 is more influenced than in the onset (see figures, 4.23, 4.26, and 4.29).

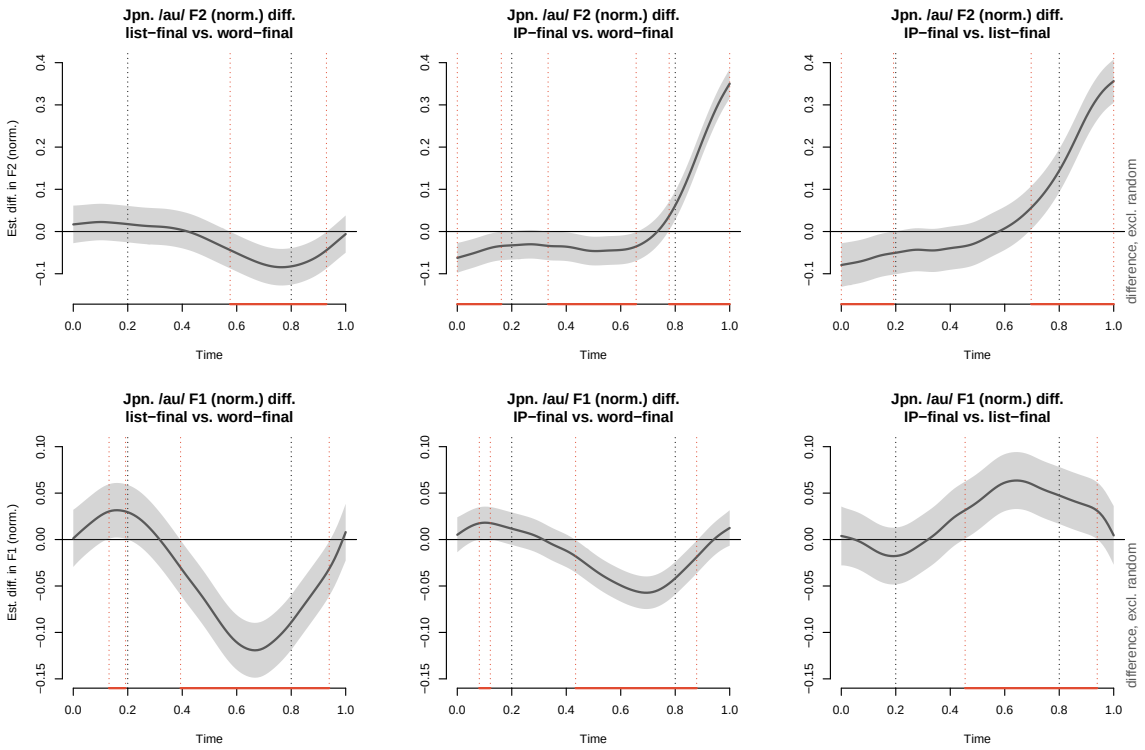

Figure 4.30: Difference between the three smooths comparing 'word-final', 'list-final', and 'IP-final'. The upper three graphs show the difference in F2 and the lower three graphs F1.

There are, however, some noticeable differences in Japanese data as well. First, the offset and the onset were both influenced. The onset F1 is higher, and the onset F2 is lower for all three TVS in Japanese. This suggests that, unlike in Chinese/English, the first vowel in the TVS is also a target of prosodic modulation in Japanese. Secondly, Japanese TVS behave differently in different prosodic contexts. In the two TVS that end with a high front target (/ae, ai/), the list-final position showed the highest F2 offset than the IP-final and word-final positions. This is probably because the IP-final positions in Japanese stimuli are not immediately before a boundary. The target words in IP-final positions in Japanese are always followed by a copula /-da/. These results may hint that the underlying target for /ai, ae/ in Japanese is maximally realized in list-final positions when there are no speech sounds following the target word. On the other hand, in Japanese /au/, the IP-final position raised the F2 at the offset of the vowel, whereas the difference between the list-final and word-final position is much smaller. This is probably because, in the word-final positions,

the following sound is a topic marker /-wa/, starting with a labio-velar sound that has a lower effect on the F2 of the preceding vowel.

4.4 Discussion

4.4.1 Prosodic effect on formant excursions

The GAM analysis examined how formant excursions of various TVS in Chinese, English, and Japanese vary at different boundaries in the prosodic structure. The summaries of the results are presented in table 4.1, 4.2, and 4.3.

Table 4.1: Summary of the results of Chinese GAM

		list-final vs. word-final		IP-final vs. word-final		IP-final vs. list-final	
		20-50%	50-80%	20-50%	50-80%	20-50%	50-80%
/ai/	F2	lower	lower	lower	lower	-	-
	F1	higher	higher	-	higher	-	-
/au/	F2	-	lower	-	lower	-	-
	F1	-	higher	-	higher	-	-
/ou/	F2	-	lower	-	lower	-	-
	F1	-	-	-	-	-	-

Table 4.2: Summary of the results of English GAM

		list-final vs. word-final		IP-final vs. word-final		IP-final vs. list-final	
		20-50%	50-80%	20-50%	50-80%	20-50%	50-80%
/ai/	F2	-	higher	-	-	-	-
	F1	-	lower	-	lower	-	-
/au/	F2	-	lower	-	lower	-	-
	F1	-	lower	-	-	-	-
/ou/	F2	-	lower	-	lower	-	-
	F1	-	-	-	-	-	-

Table 4.3: Summary of results of Japanese GAM

		list-final vs. word-final		IP-final vs. word-final		IP-final vs. list-final	
		20-50%	50-80%	20-50%	50-80%	20-50%	50-80%
/ae/	F2	higher	higher	-	higher	-	lower
	F1	-	-	-	-	-	-
/ai/	F2	-	higher	lower	higher	-	-
	F1	higher	lower	-	lower	-	-
/au/	F2	-	lower	-	lower	-	-
	F1	higher	lower	-	lower	-	-

The results clearly show that prosody overall does influence the acoustics of TVS in all three languages. A shared property across the three languages is that there is almost no difference in either F1 or F2 in comparing IP-final and list-final positions except in Japanese. This is because the local sound context of IP-final and list-final positions in Japanese are different from each other. In Chinese/English, TVS was not followed by any other sounds within the same prosodic phrase in both IP-final and list-final positions. But in Japanese, the TVS is followed by a monomoraic copula /-da/. This /-da/ hypothetically blocked the lengthening of the IP-final vowel sequences and led to more anticipatory coarticulation in the TVS.

Most differences induced by prosodic context were attested from the comparisons that involve word-final positions. Namely, the speakers only distinguish two categories regarding pre-boundary lengthening, word-final position, and non-word-final position. Speakers only distinguishing two or three different categories in lengthening is not new in the literature of research on the phonetics-prosody interface. For instance, in Krivokapić (2007), although the stimuli involved six different prosodic contexts, the speakers could only distinguish 2 or 3 levels of prosodic lengthening categories.

However, depending on the language and the TVS, the boundary effect on formant excursion differs. First, the impact on the first 20-50% of the TVS differs across the three languages. While the boundary effect induces no difference in English, there is a significant

difference in F1 and F2 in the first 20-50% of Chinese /ai/. Moreover, the first 20-50% of all the Japanese TVS were influenced when comparing word boundaries to ip and IP boundaries. This may indicate whether the prosodic boundary strengthening affects the initial target of a TVS because of the nature of the segment and specific language. That prosodic effect differs for segments is also a finding reported in the literature (Bombien et al., 2013; Cho, 2004). It was found in Cho (2004) that vowel-to-vowel coarticulation depends on the order of the adjacent vowels. While /a#i/ sequence showed a duration-independent effect of coarticulatory resistance, the strength of prosodic boundaries did not influence /i#a/ coarticulation. Also, in Bombien et al. (2013), it was reported that different onset consonant clusters in German are affected differently by the boundaries. It was pointed out that /kl/ and /kn/ were affected by prosodic boundaries and stress, while the categorical difference in coarticulatory gestural overlap between /kl/ and /kn/ was maintained.

This might be accounted for by referring to "phonetic knowledge" (Kingston & Diehl, 1994). "Phonetic knowledge" refers to the phenomenon that 'phonetic implementation is not automatically determined by constraints reflecting articulatory contingencies' (Kingston & Diehl, 1994, p. 423). In different languages, speakers may have different fine-grained encodings in phonetics, even for the same phonological category. Kingston and Diehl (1994) mentioned that even though the VOT of the voiceless consonant was shortened in word-initial /sC/ sequences in English, the onset f0 of the vowel is still the same as those that are not preceded by /s/, suggesting that onset f0 lowering is not an automatic phonetic by-product of the duration of VOT. Cho and Ladefoged (1999) mentioned that languages implement VOT for aspirated/unaspirated consonants very differently despite that a universal trend is found simultaneously.

My results above suggest that the prosodic strengthening effect on the TVS is not automatic either. While Chinese TVS showed a lowered F2 and raised F1 at the end of all the TVS in the offset, comparing ip- and IP-boundaries to word boundaries, Japanese showed an opposite direction of formant modulation: higher F2 but lower F1 for /ae, ai/. For En-

glish TVS, most of the effect was realized as a lowering effect on both formants. If the prosodic effect on the segment is only due to automatic computation from phonology to phonetics, the same trend should have been found across all TVS and languages. Each language has its non-automatic encoding of prosodic information in producing segments adjacent to a prosodic boundary.

The Chinese pattern showed that the TVSs in the word-final position are more susceptible to the coarticulatory effects due to the following coronal consonant that raises the F2 and lowers the F1. The result in English is less consistent in that F2 in the word-final positions is lower in the offset for /ai/ but higher for /au, ou/. In Japanese, F2 is lowered in word-final positions in /ae, ai/, and in both word-final and list-final positions in /au/. Interestingly, F2 in /ai/ and /ae/ (Japanese) showed the opposite direction of modulation: higher in Chinese but lower in English and Japanese. This result suggests two mechanisms behind the modulation of F2 movement in the offset across languages. One possibility is that in Chinese, the underlying F2 target of /-i/ in /ai/ is lower than the F2 value due to the influence of the following coronal consonant, while in English and Japanese, it is the opposite. The F2 target of /-i/ is higher in the list-final position than the F2 in the word-final position, indicating that the tongue is in a higher and more front position in the list-final position. Henceforth in Chinese, it is due to vowel-consonant coarticulation. In English and Japanese, however, It is probably the consequence of target down-scaling due to time pressure in word-final positions. It is probable that in word-final positions, the vowels are hypoarticulated since English, and Japanese TVS did not have enough time to reach the target.

To conclude, although prosodic boundaries influence the formant excursions of TVS in all three languages, the phonetic detail is by no means automatically determined. It instead supports the view that pre-boundary strengthening is under the speaker's control. It must be specified in a linguistic description of the phonetics–prosody interface as part of the phonetic grammar of the language (Cho & Ladefoged, 1999; Keating, 1984, 1990; Kingston &

Diehl, 1994). Pre-boundary lengthening effect is both universal and also language-specific.

4.4.2 Sonority expansion or hyperarticulation?

Regarding whether the strengthening strategy used sonority expansion or hyperarticulation, the result of the GAM analysis is not conclusive. Sonority expansion predicts that the TVS is produced with a more open mouth and lower tongue position in higher prosodic boundaries, resulting in higher F1. Henceforth if the sonority of the target vocalic segment is expanded, we should expect raised F1 somewhere in the vowel sequence. From the result presented above, one can hardly say that this hypothesis was born out. F1 is barely influenced in the three languages in the first half of the TVS, meaning that the prosodic modulation on formants probably does not target the first half of the vowel or the first vocalic target /a/ of the vowel sequence. In Chinese /ai/, it does seem that F1 of /ai/ and /au/ was raised at the end at higher prosodic boundaries (list-final and IP-final positions), but it is accompanied by a lowering of F2 as well. As I argued above, this should be seen as a coarticulatory effect instead of sonority expansion. Therefore, in the acoustic domain, sonority expansion cannot account for the prosodic variation in the data.

The Hyperarticulation hypothesis claims that the distinctive feature is enhanced in more prominent contexts, such as at higher prosodic boundaries. This predicts that front vowels are produced more front (higher F2) and back vowels more back (lower F2). If this were the case, we should expect to observe that at the offset of all the TVS discussed in this study, at a higher prosodic boundary (list-final or IP-final positions), if it ends with a /-i/, the F2 should be higher. But if it ends with a /-u/, the F2 should be lowered in the offset of TVS. In TVS that ended with a front vowel /ai, ae/, it was confirmed that F2 was indeed higher at higher prosodic boundaries (list-final or IP-final positions) in English and Japanese, indicating that TVSs are hyperarticulated when produced in the vicinity of a higher prosodic boundary. When TVS ends with a high back vowel (/au, ou/), F2 is always lower at the end at higher boundaries in all languages. This indicates that higher prosodic

boundaries hyperarticulate the TVS.

However, the hyperarticulation hypothesis does not account for all the reported patterns in this chapter. For instance, in Chinese /ai/, the F2 is higher at lower prosodic boundaries (word-final positions). The difference is probably due to the different phonological statuses of TVS in different languages. Chinese /ai/ is implemented phonetically with less formant excursions than those in English and Japanese. Therefore when it is more adjacent to the following coronal consonant in the experiment, it is more coarticulated with the tongue raising and fronting gesture, leading to a higher F2 (more front tongue position) and a lower F1 (higher tongue position). If we account for this directly from the view of the hyperarticulation hypothesis, then it is the /ai/ in the word-final position that was hyperarticulated, contrary to the theoretical prediction that higher boundary (IP-/list-final positions) should induce hyperarticulated speech sounds. Instead, the Chinese data of /ai/ prosodic modulation should be accounted for as coarticulatory resistance. Speech sounds at boundaries of larger prosodic constituents (i.e., intermediate, phrase, or intonational phrase) are more resistant to coarticulation. This could be interpreted as a kind of "hyperarticulation" as well, but it is not considered the hyperarticulation defined in this study.

In sum, prosodic boundaries impact the formant excursions of TVS in the three languages, and the mechanism of the prosodic modulation on formant excursion is local hyperarticulation that enhances the distinctive features of the segments.

4.5 Conclusion of GAM analysis

This chapter used Generalized Additive Mixed-effect Modeling to analyze the formant excursions of TVS in Chinese, English, and Japanese. Prosodic boundaries affect the phonetic implementation of the TVS, although the detail of the effect differs across languages. Through the prosodic effect on formant excursion, we also learned that the phonological representations of TVS are probably different across languages. English and Chinese TVS

form multi-target single phonemes, while Japanese ones are vowel phoneme strings that occur in the same syllable.

However, since the GAM analysis was performed on time-normalized data, I did not consider actual duration to explain the variation. Crucially, GAM analysis did not answer the research question about what strategies were used in prosodic strengthening for TVS. Therefore, in the next chapter, I will analyze the data by referring to the duration-sensitive kinematic measures to explore more about the TVS in different prosodic contexts.

Chapter 5

Kinematic analysis of vowel space movement

In this chapter, I will report the result of statistical analysis on the four kinematic measures of formants movement in the vowel space as introduced in section 2.2. The four kinematic measures are total displacement, total duration, peak velocity, and stiffness (the peak velocity and displacement ratio) of the true diphthongal movement (the movement from the maxima of F1 to the minima/maxima of F2). The method of calculation of the four kinematic measures is repeated below. Note that the duration here is calculated based on the measured time of the middle point of each of the thirty time points. This is, therefore, different than the duration data presented in chapter 3.

(23) Calculating displacement, duration, and velocity between the data points.

a. Displacement$_i$ = $\sqrt{(F1_{i+1} - F1_i)^2 + (F2_{i+1} - F2_i)^2}$

b. ID$_i$ = t$_{i+1}$−t$_i$ (Interval duration)

c. Vel$_i$ = Disp$_i$/ID$_i$ (Velocity between each two data points)

(24) a. Total disp = $\sum_1^n Displacement_i$ (Total trajectory displacement)

b. Total dur = $\sum_1^n ID_i$ (Total duration)

c. pVel = $\max(V_i)$ (peak velocity)

d. Stiffness = $\frac{Totaldisp}{Totaldur}$

This analysis aims to quantitatively examine how prosodic boundaries influence the dynamic movement of the vowel sequences in the vowel space. As already outlined in the section 2.2, during the acoustic interval of a TVS, not all the formant excursions are to produce the target of the TVS. Therefore the kinematic measures were all taken from the true movement interval, which is defined as starting from the maxima of F1 to the minima (for /au, ou/) or the maxima (for /ai, ae/) of F2. This is to remove the formant excursions at the beginning and the end of TVS due to coarticulation. In the articulatory domain, previous studies have found that several of the four measures are correlated with the strength of the prosodic boundary. The possible strategies are shown again below in figure 5.1 (the same as figure 1.4).

Linear mixed-effect models were built for each kinematic measure of each vowel sequence (/ai, au, ae, ou/) separately treating prosodic POSITION and LANGUAGE as the fixed effects, and SPEAKER as the random effect. The prosodic POSITION was also included as a within-subject variable whereas LANGUAGE as a between-subject variable.

5.1 /ai/

All models of /ai/ kinematic analysis included `Speaker` as the random effect with correlated intercept and slope. The results are reported below.

Displacement

Figure 5.2 shows the overall distribution of displacement. The trajectory movement seems the longest in list-final positions and shortest in word-final positions in English and Japanese, but longest in word-final positions in Chinese. /ai/ moved longer in the vowel space in English and Japanese than in Chinese. In all three languages, the displacement in list-final

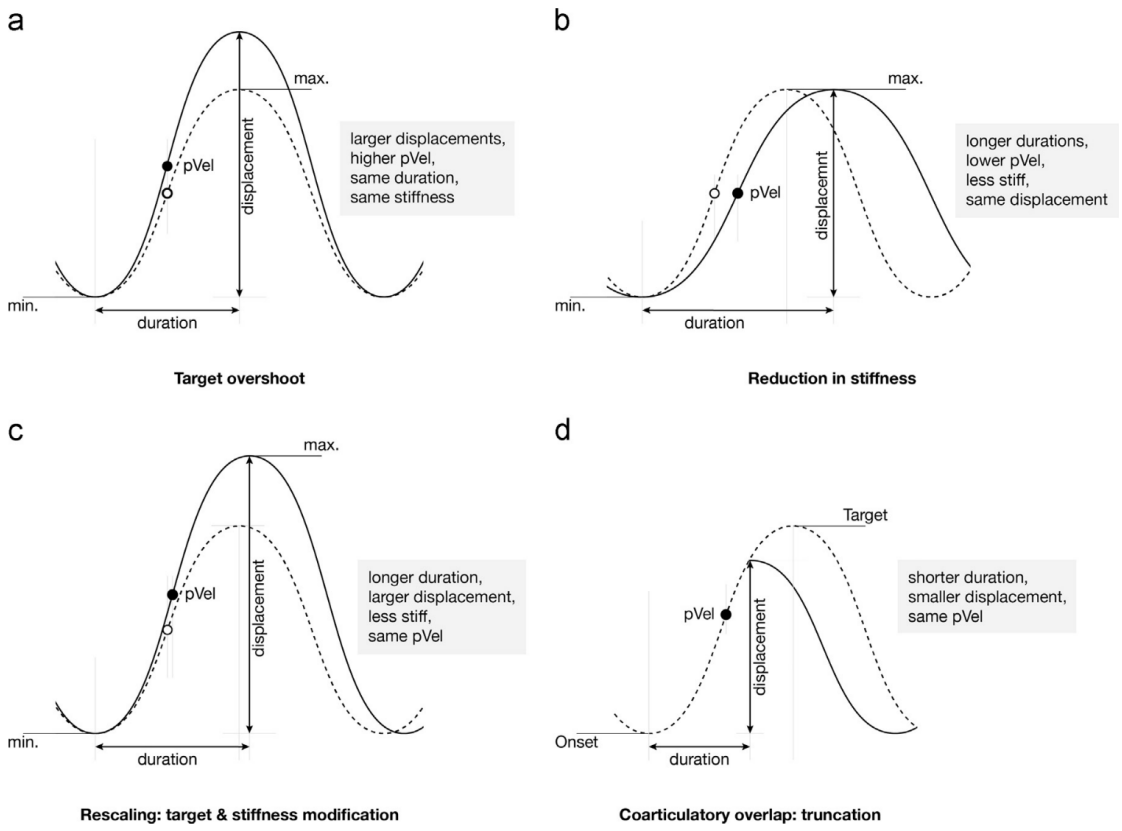

Figure 5.1: Possible articulatory strategies that might be used by speakers (figure taken from Mücke & Grice, 2014).

is longer than in IP-final positions. The statistical analysis showed that `Position` ($F(2, 15.25) = 8.62$, $p < 0.005$), `Language` ($F(2, 15) = 7.98$, $p < 0.005$), and the interaction between them ($F(4, 15.26) = 5.30$, $p < 0.01$) were all significant. Tukey-adjusted post-hoc comparison of /ai/ displacement is shown in figure 5.3 and table 5.1. The red arrows in figure 5.3 show the heterogeneous groups in each level (If one group's arrow overlaps with another group's, the difference is not significant). The blue bars on the right indicate the 95% confidence intervals in both figures. The contrast is significant if the x=0 intercept does not lie within the confidence interval.

Figure 5.3 and table 5.1 together show that the only significant contrasts were in Japanese. In Japanese, /ai/ was produced with longer displacement in the vowel space in the list-final position than in the other two contexts.

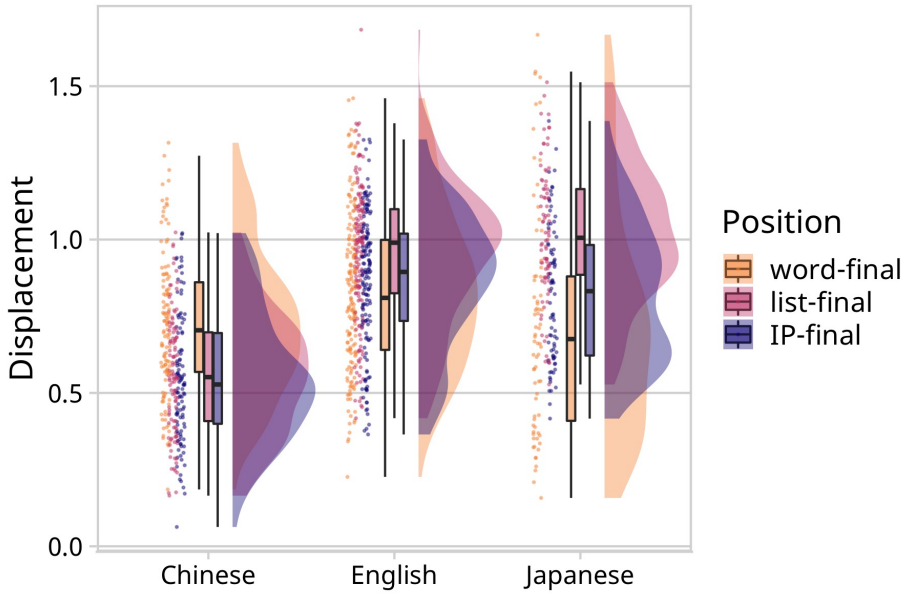

Figure 5.2: The distribution of displacement of /ai/ trajectories.

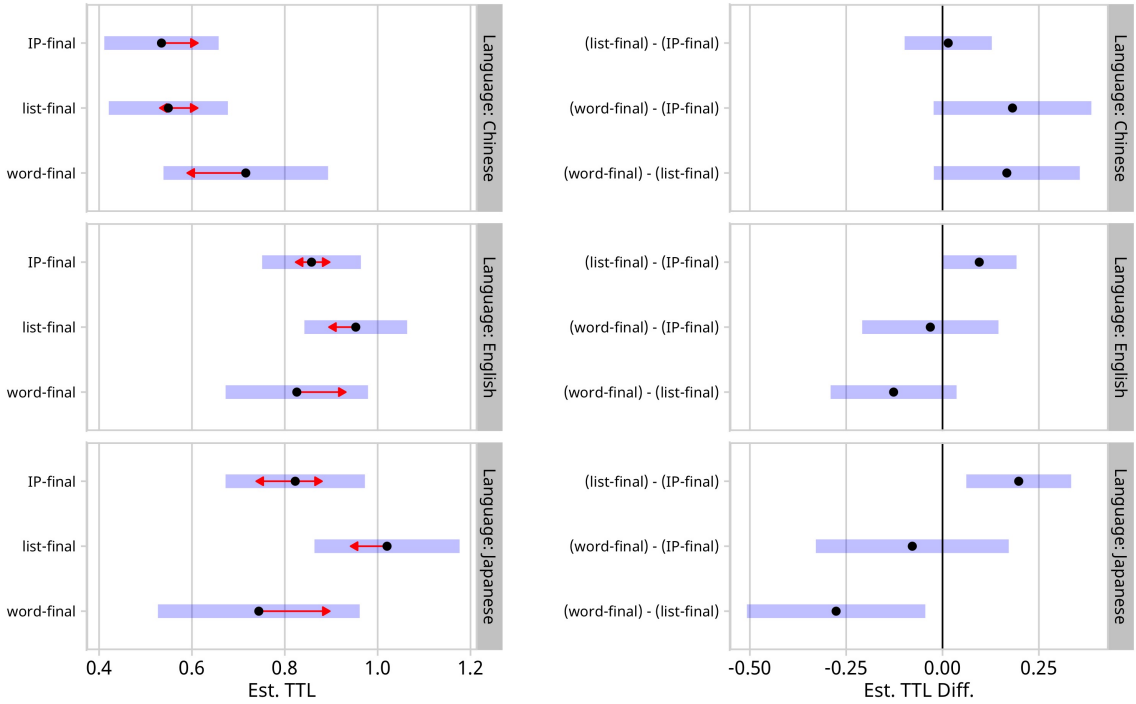

Figure 5.3: The estimated marginal means (left) and the estimated difference of displacement between pairwise contrasts (right) for /ai/.

Duration

The patterns in the proper moving duration of /ai/ trajectory shown in figure 5.4 look rather different in the three languages. There is hardly any difference among the three prosodic

Table 5.1: Post-hoc pairwise comparisons of displacement for /ai/.

contrast	estimate	SE	df	t.ratio	p.value
Language = **Chinese**					
(word-final) - (list-final)	0.1668	0.0730	15.19	2.285	0.0887
(word-final) - (IP-final)	0.1816	0.0789	15.12	2.302	0.0861
(list-final) - (IP-final)	0.0148	0.0437	15.61	0.338	0.9393
Language = **English**					
(word-final) - (list-final)	-0.1270	0.0629	14.86	-2.019	0.1420
(word-final) - (IP-final)	-0.0317	0.0681	14.92	-0.465	0.8883
(list-final) - (IP-final)	0.0953	0.0372	14.81	2.561	0.0539
Language = **Japanese**					
(word-final) - (list-final)	-0.2762	0.0891	15.00	-3.099	0.0189
(word-final) - (IP-final)	-0.0785	0.0964	14.99	-0.814	0.7002
(list-final) - (IP-final)	0.1977	0.0521	14.24	3.795	0.0051

Degrees-of-freedom method: kenward-roger

P value adjustment: tukey method for comparing a family of 3 estimates

contexts in Chinese data, whereas English TVS were produced with the longest moving duration in list-final and IP-final positions. In Japanese, TVS were produced with the longest moving duration only in list-final positions. Japanese TVS IP-final positions were longer than those in word-final positions.

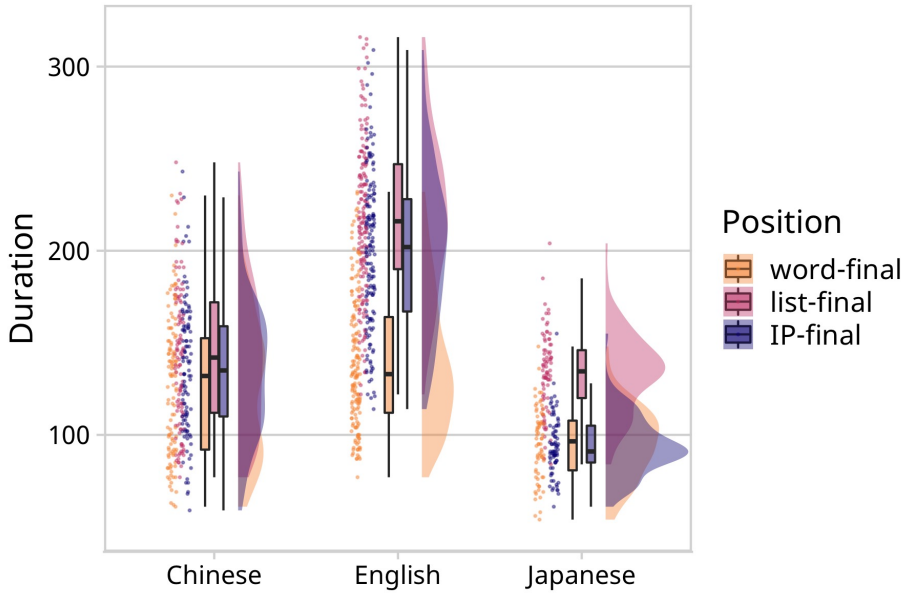

Figure 5.4: The distribution of moving duration of /ai/ trajectories.

The statistical results confirmed the main effect of `Position` ($F(2, 14.99) = 37.01$, $p < 0.005$), `Language` ($F(2, 14.99) = 19.02$, $p < 0.005$), and the interaction between them ($F(4, 15.01) = 8.04$, $p < 0.01$) on moving duration. Results of post-hoc comparison of duration are shown in figure 5.5 and table 5.2.

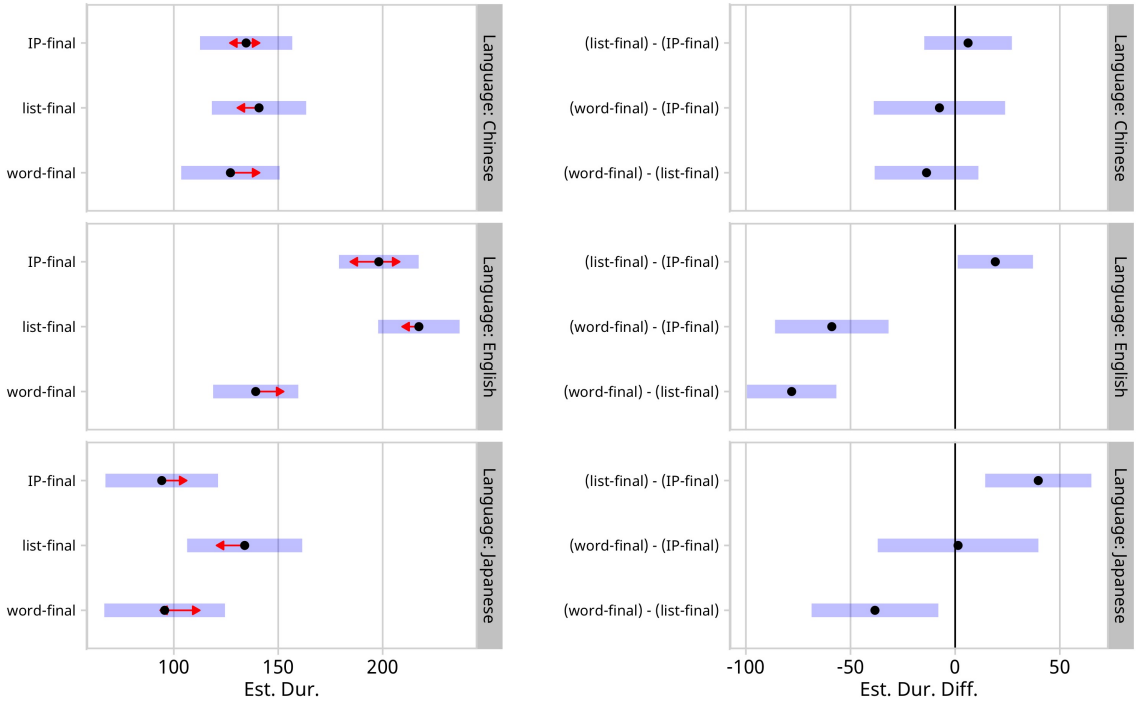

Figure 5.5: The estimated marginal means (left) and the estimated difference of duration between pairwise contrasts (right) for /ai/.

It is shown that the duration of /ai/ did not differ significantly in Chinese at all. In English, the longest duration was found in the list-final position and the shortest in the word-final position, with the IP-final position in between. In Japanese, The list-final position /ai/ was produced with the longest duration, but the difference between IP-final and word-final positions was insignificant.

Peak velocity

The distribution of peak velocity is demonstrated in figure 5.6. The peak velocity seems to be largest in the word-final position in Chinese and English. Japanese word-final po-

Table 5.2: Post-hoc pairwise comparisons of duration for /ai/.

contrast	estimate	SE	df	t.ratio	p.value
Language = **Chinese**					
(word-final) - (list-final)	-13.6832	9.5574	15.23	-1.432	0.3501
(word-final) - (IP-final)	-7.5149	12.0929	15.11	-0.621	0.8108
(list-final) - (IP-final)	6.1682	8.0841	15.42	0.763	0.7305
Language = **English**					
(word-final) - (list-final)	-78.1239	8.2201	14.83	-9.504	<.0001
(word-final) - (IP-final)	-58.9417	10.4386	14.92	-5.647	0.0001
(list-final) - (IP-final)	19.1822	6.9258	14.87	2.770	0.0362
Language = **Japanese**					
(word-final) - (list-final)	-38.3239	11.6582	15.00	-3.287	0.0130
(word-final) - (IP-final)	1.3815	14.7780	14.99	0.093	0.9952
(list-final) - (IP-final)	39.7054	9.7341	14.52	4.079	0.0028

Degrees-of-freedom method: kenward-roger

P value adjustment: tukey method for comparing a family of 3 estimates

sition, however, was associated with the most negligible peak velocity, showing a slower movement in word-final positions.

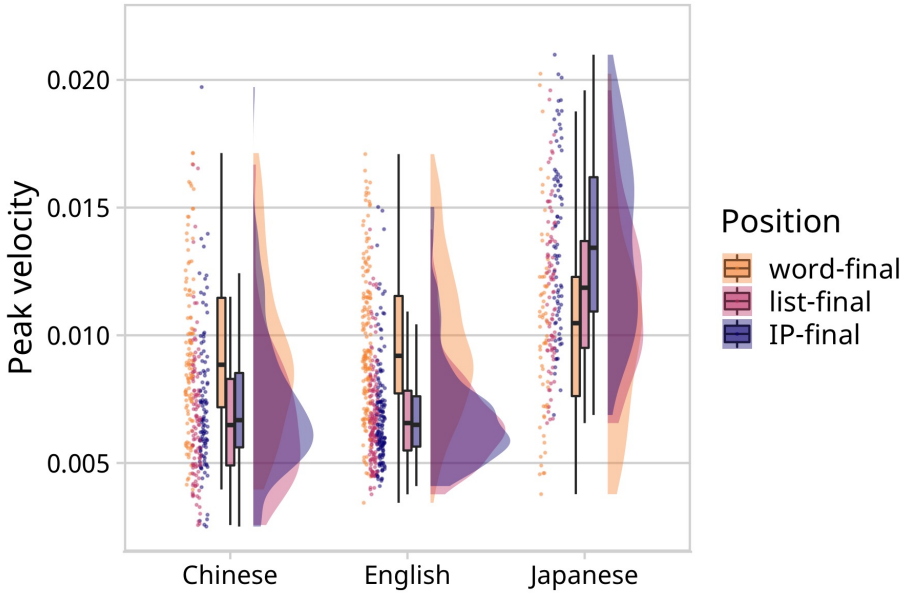

Figure 5.6: The distribution of peak velocity of /ai/ trajectories.

Main effects of `Position` ($F(2, 15.1) = 8.07$, $p < 0.005$), `Language` ($F(2, 15.1) = 20.8$, $p < 0.005$) are significant on peak velocity. But the interaction did not reach signif-

icance (F(2, 15.12) = 0.96). The post-hoc comparison is displayed in figure 5.7 and table tab:aiVelPost. In English, /ai/ trajectory moved significantly faster in the word-final position than in the other two contexts. The differences in the peak velocity were insignificant, neither in Chinese nor in Japanese.

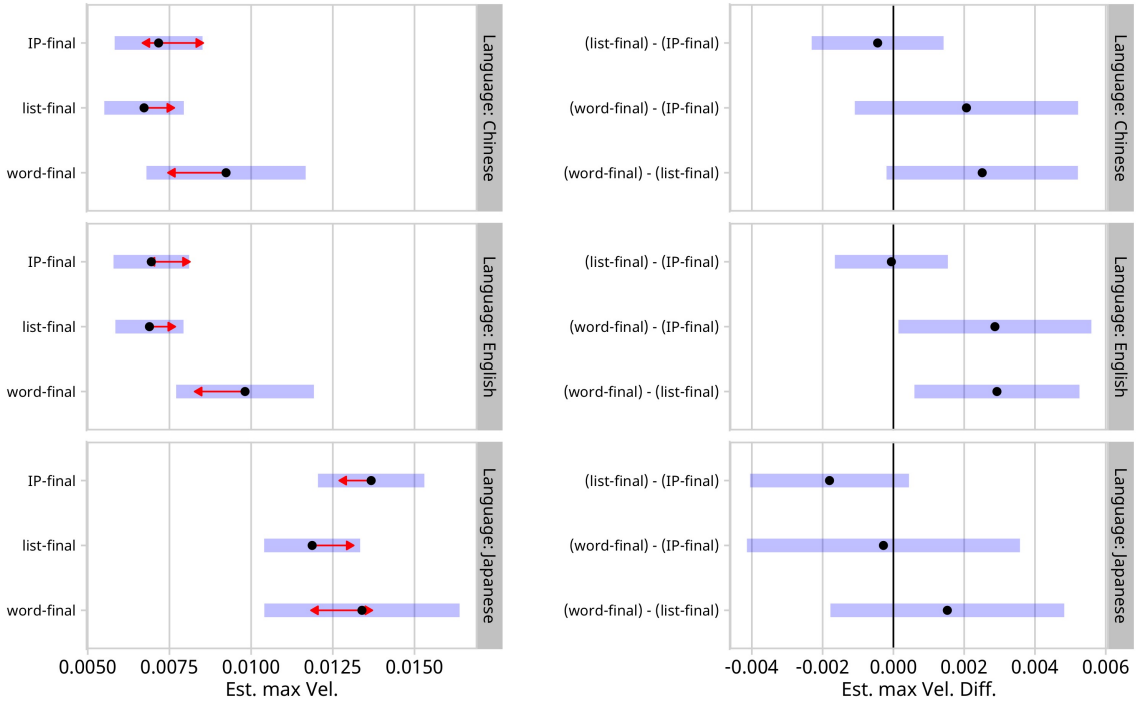

Figure 5.7: The estimated marginal means (left) and the estimated difference of peak velocity between pairwise contrasts (right) for /ai/.

Stiffness

Figure 5.8 displays the stiffness across languages. The trend observed in the figure shows that in Chinese, the difference is negligible, whereas, in English and Japanese, stiffness is larger in word-final positions than in list-final positions. The stiffness in English patterns with that in list-final positions, while in Japanese, it patterns with that in word-final positions. This is probably because the Japanese /ai/ trajectory is produced with a comparable duration in the word-final and IP-final positions.

Main effects of `Position` ($F(2, 14.94) = 57.91$, $p < .005$), `Language` ($F(2, 14.86)$

Table 5.3: Post-hoc pairwise comparisons of peak velocity for /ai/.

contrast	estimate	SE	df	t.ratio	p.value
Language = **Chinese**					
(word-final) - (list-final)	0.0025	0.0010	15.23	2.407	0.0709
(word-final) - (IP-final)	0.0021	0.0012	15.12	1.698	0.2380
(list-final) - (IP-final)	-0.0004	0.0007	15.58	-0.618	0.8124
Language = **English**					
(word-final) - (list-final)	0.0029	0.0009	14.83	3.262	0.0139
(word-final) - (IP-final)	0.0029	0.0010	14.91	2.735	0.0386
(list-final) - (IP-final)	-0.0001	0.0006	14.82	-0.093	0.9952
Language = **Japanese**					
(word-final) - (list-final)	0.0015	0.0013	15.01	1.198	0.4724
(word-final) - (IP-final)	-0.0003	0.0015	14.99	-0.190	0.9804
(list-final) - (IP-final)	-0.0018	0.0009	14.30	-2.099	0.1252

Degrees-of-freedom method: kenward-roger

P value adjustment: tukey method for comparing a family of 3 estimates

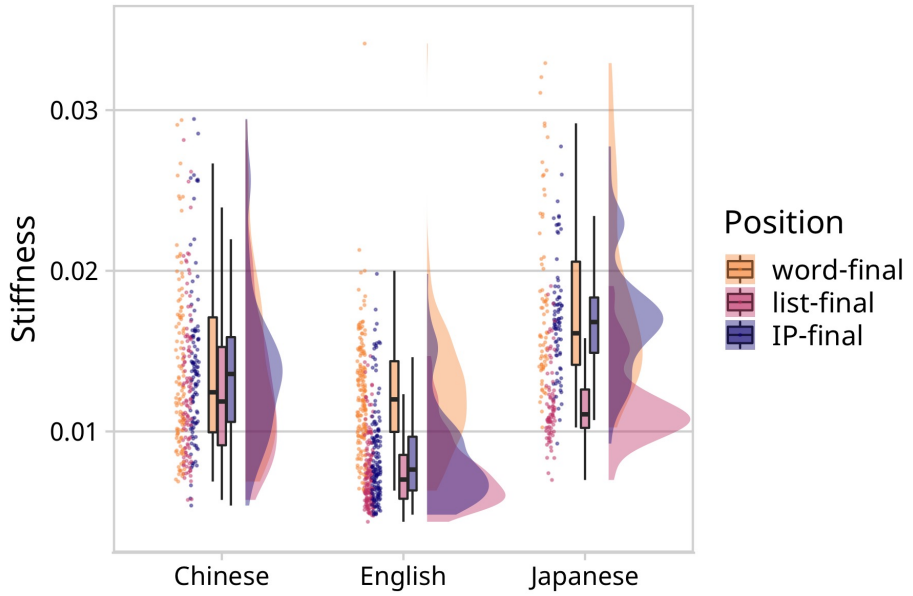

Figure 5.8: The distribution of stiffness of /ai/ trajectories.

= 20.51, $p < .005$) were significant for stiffness as well as the interaction between ($F(4, 14.96) = 10.47$, $p < .005$). The post-hoc comparison result in figure 5.9 and table 5.4 demonstrates that /ai/ trajectory in English had significantly larger stiffness in the word-final position than in both list- and IP-final positions. /ai/ trajectory also had larger stiffness

in word-final and IP-final positions than in list-final positions.

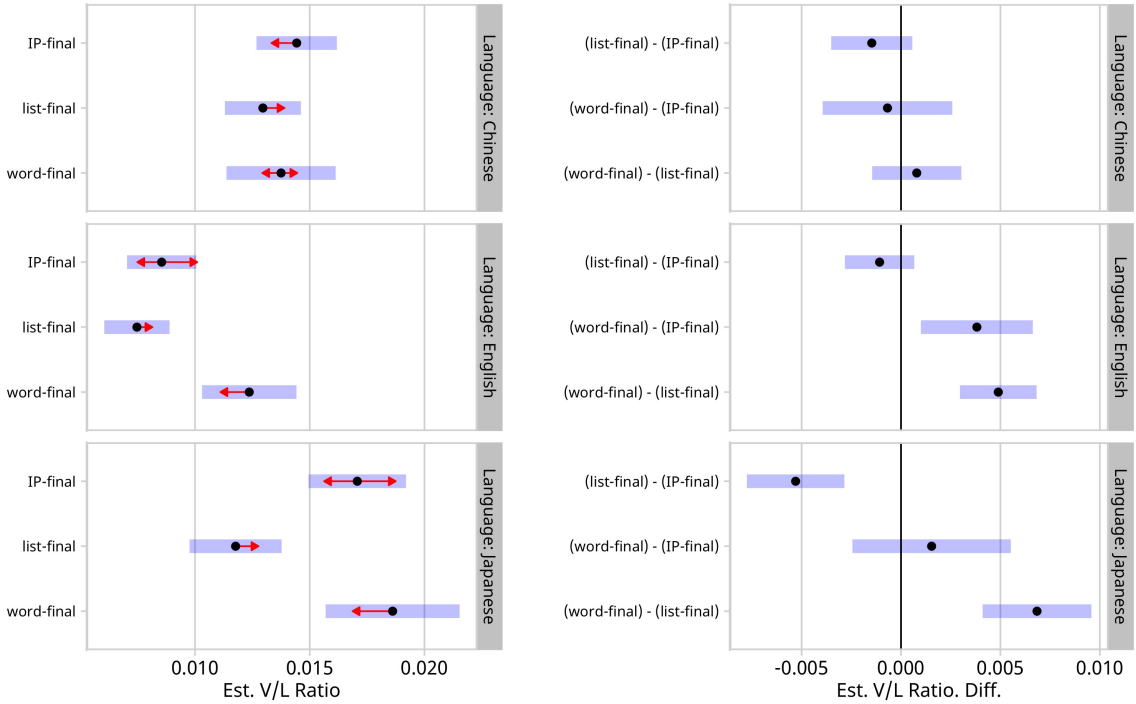

Figure 5.9: The estimated marginal means (left) and the estimated difference of stiffness between pairwise contrasts (right) for /ai/.

Table 5.4: Post-hoc pairwise comparisons of stiffness for /ai/.

contrast	estimate	SE	df	t.ratio	p.value
Language = **Chinese**					
(word-final) - (list-final)	0.0008	0.0009	15.45	0.913	0.6403
(word-final) - (IP-final)	-0.0007	0.0013	15.16	-0.542	0.8522
(list-final) - (IP-final)	-0.0015	0.0008	15.61	-1.867	0.1813
Language = **English**					
(word-final) - (list-final)	0.0049	0.0007	14.66	6.611	<.0001
(word-final) - (IP-final)	0.0038	0.0011	14.88	3.525	0.0082
(list-final) - (IP-final)	-0.0011	0.0007	14.82	-1.607	0.2736
Language = **Japanese**					
(word-final) - (list-final)	0.0068	0.0011	15.02	6.495	<.0001
(word-final) - (IP-final)	0.0015	0.0015	14.98	1.004	0.5857
(list-final) - (IP-final)	-0.0053	0.0009	14.21	-5.651	0.0002

Degrees-of-freedom method: kenward-roger

P value adjustment: tukey method for comparing a family of 3 estimates

Summary of /ai/ kinematics

Table 5.5 summarizes the findings in the analysis of /ai/ kinematics.

Table 5.5: The summary of kinematic analysis of /ai/ trajectory movement.

	word- vs. list-final	word- vs. IP-final	list- vs. IP-final
Chinese			
Displacement	n.s.	n.s.	n.s.
Duration	n.s.	n.s.	n.s.
Peak velocity	n.s.	n.s.	n.s.
Stiffness	n.s.	n.s.	n.s.
English			
Displacement	n.s.	n.s.	n.s.
Duration	shorter	shorter	longer
Peak velocity	faster	faster	n.s.
Stiffness	larger	larger	n.s.
Japanese			
Displacement	smaller	n.s.	larger
Duration	shorter	n.s.	longer
Peak velocity	n.s.	n.s.	n.s.
Stiffness	larger	n.s.	smaller

It is rather apparent from the summary table that /ai/ in the three languages underwent quite different prosodic modulation in terms of the kinematics of the movement. During the proper vowel sequence interval, /ai/ showed no differences in the movement in the vowel space whatsoever in Chinese.

In contrast, the movement of /ai/ in the vowel space was indeed influenced in English and Japanese, although the strategies used were somewhat different in the two languages. In English, at higher prosodic boundaries (list-final and IP-final positions), /ai/ trajectory showed longer duration, with faster peak velocity and smaller stiffness but not necessarily larger displacement. This is in line with the articulatory findings reported in the literature (Beckman & Edwards, 1992; Byrd, 2000; Byrd & Saltzman, 1998; Edwards et al., 1991).

Japanese, however, utilizes a somewhat different strategy that the movement is rescaled: the movement takes longer to reach a larger target displacement with larger prosody when

lengthened in the list-final positions. Another noteworthy phenomenon is that in Japanese, IP-final TVS has the same dynamical profile as those in word-final positions.

5.2 /au/

The linear mixed-effect models of /au/ peak velocity and stiffness only included `Speaker` as the random intercept as more complex models failed to converge. Otherwise, `Speaker` was included as the random effect with correlated intercepts and slopes.

Displacement

Figure 5.10 displays the displacement of /au/ across languages. The displacement is smaller in the word-final position than in the other two positions in both Chinese and English. In Japanese, the range of displacement in the word-final position is rather wide, while the list-final position showed longer displacement than in IP-final positions.

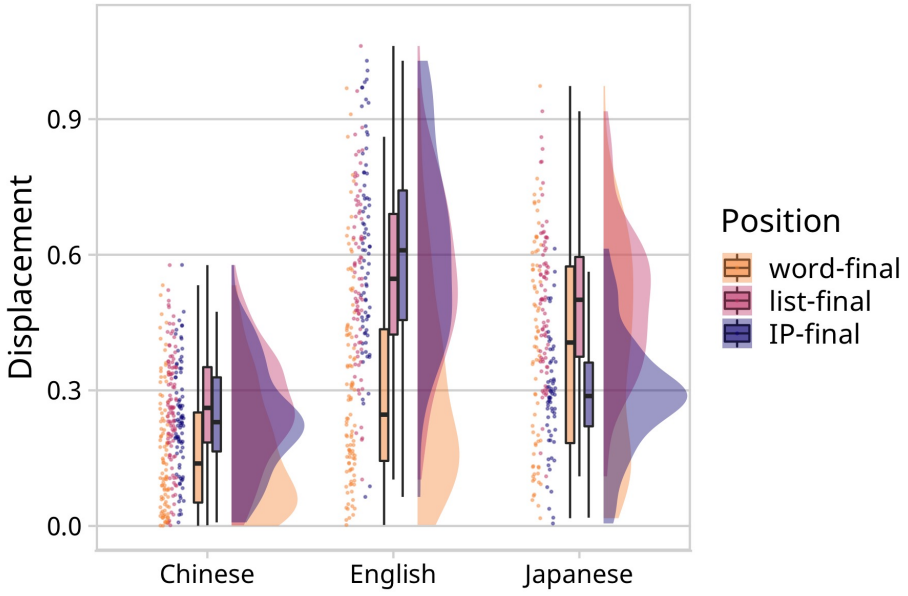

Figure 5.10: The distribution of displacement of /au/ trajectories.

Statistical analysis showed that the main effects of `Position` ($F(2, 14.18) = 24.66$, $p < 0.005$) and `Langauage` $F(2, 14.98) = 19.68$, $p < 0.005$, as well as their interaction

F(4, 14.31) = 3.35, $p < 0.05$, were all significant. Tukey-adjusted post-hoc comparison is displayed in figure 5.11 and table 5.6.

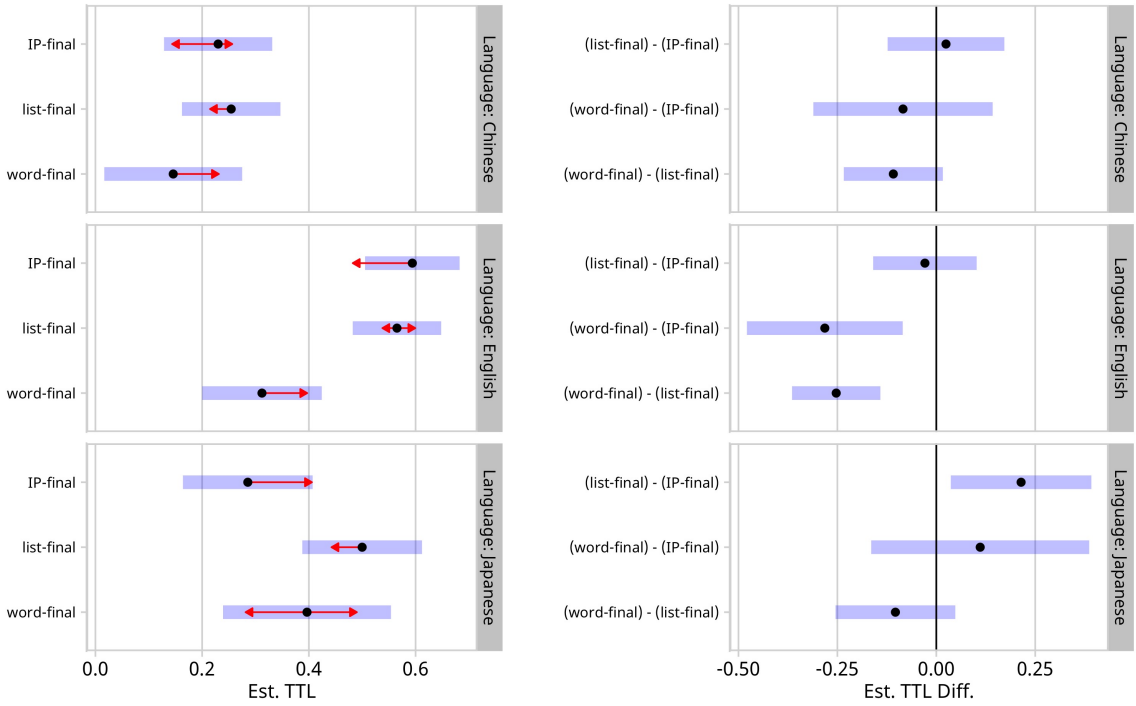

Figure 5.11: The estimated marginal means (left) and the estimated difference of displacement between pairwise contrasts (right) for /au/.

The post-hoc comparison reveals that the displacement of Chinese /au/ did not differ across the prosodic contexts. In English, the difference came from the comparisons that involved word-final position. /au/ in the word-final position moved for shorter distances in the vowel space than in the other two prosodic contexts. In Japanese, the displacement is only significantly larger when comparing the list-final position to the IP-final position.

Duration

The distribution of duration in different prosodic contexts across the three languages is demonstrated in figure 5.12. The pattern is similar to that in the displacement in that the word-final position in both Chinese and English was associated with shorter duration. In contrast, the list- and IP-final positions were associated with longer durations. The Japanese

Table 5.6: Post-hoc pairwise comparisons of displacement for /au/.

contrast	estimate	SE	df	t.ratio	p.value
Language = **Chinese**					
(word-final) - (list-final)	-0.1086	0.0479	14.12	-2.269	0.0935
(word-final) - (IP-final)	-0.0842	0.0872	14.89	-0.966	0.6089
(list-final) - (IP-final)	0.0245	0.0564	14.10	0.434	0.9021
Language = **English**					
(word-final) - (list-final)	-0.2530	0.0435	16.75	-5.822	0.0001
(word-final) - (IP-final)	-0.2818	0.0760	15.35	-3.708	0.0054
(list-final) - (IP-final)	-0.0288	0.0509	16.59	-0.566	0.8397
Language = **Japanese**					
(word-final) - (list-final)	-0.1033	0.0574	13.10	-1.801	0.2075
(word-final) - (IP-final)	0.1109	0.1055	14.30	1.051	0.5579
(list-final) - (IP-final)	0.2143	0.0674	13.16	3.181	0.0183

Degrees-of-freedom method: kenward-roger

P value adjustment: tukey method for comparing a family of 3 estimates

showed a different pattern that the shortest duration was found in the IP-final position but the longest in the list-final position.

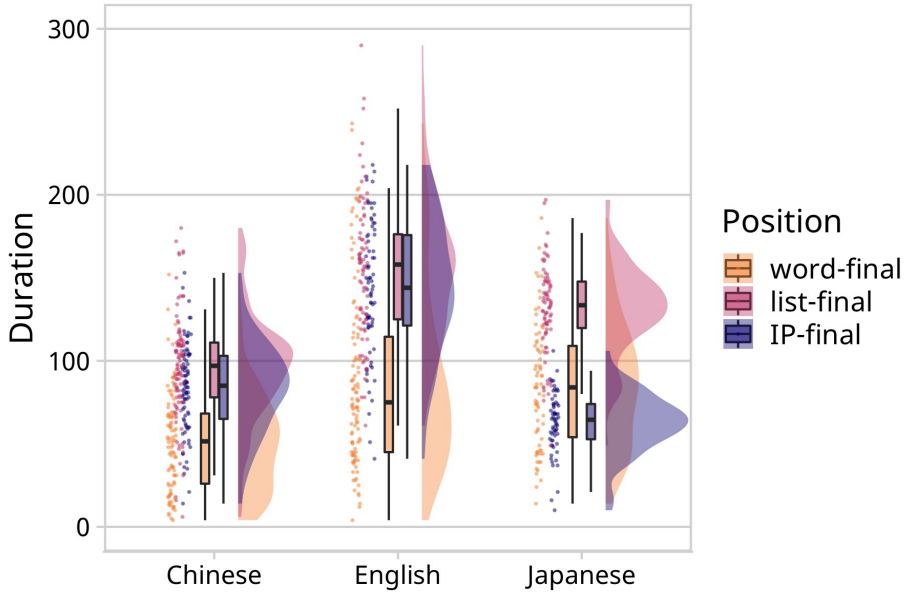

Figure 5.12: The distribution of moving duration of /au/ trajectories.

Main effects of `Position` ($F(2, 13.21) = 91.29$, $p < .005$) and `Langauge` ($F(2, 14.57) = 10.90$, $p < .005$) were confirmed significant. The interaction between them was also

significant (F(4, 13.31) = 8.07, $p < .005$). The post-hoc comparison in figure 5.13 and table 5.7 demonstrates that in Chinese, the only significant difference was between word-final and list-final positions with the a longer in the list-final position. Both comparisons involving the word-final position in English were significant, but IP- and list-final positions showed no difference. In Japanese, both the word-final and the IP-final positions were shorter than in the list-final positions, while there is no difference between word-final and IP-final positions.

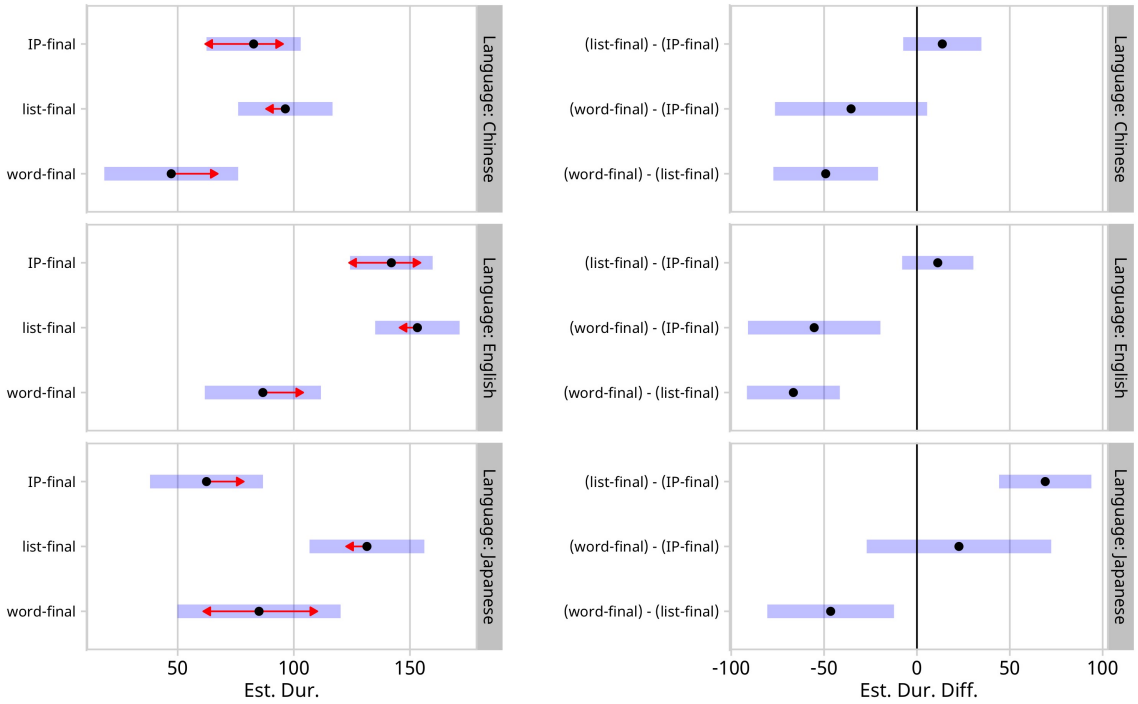

Figure 5.13: The estimated marginal means (left) and the estimated difference of duration between pairwise contrasts (right) for /au/.

Peak velocity

Peak velocity of /au/ trajectories is shown in figure 5.14. The distribution shows that the peak velocity largely overlaps each other in different prosodic positions.

Statistical analysis still shows significant results for the main effects of `Position` (F(2, 738.37) = 6.99, $p < .005$) and `Language` (F(2, 15.51) = 12.11, $p < .005$), as well as their

Table 5.7: Post-hoc pairwise comparisons of duration for /au/.

contrast	estimate	SE	df	t.ratio	p.value
Language = **Chinese**					
(word-final) - (list-final)	-49.1279	10.7925	14.26	-4.552	0.0012
(word-final) - (IP-final)	-35.4704	15.7463	14.78	-2.253	0.0948
(list-final) - (IP-final)	13.6576	7.9645	12.88	1.715	0.2372
Language = **English**					
(word-final) - (list-final)	-66.5350	9.7309	16.61	-6.838	<.0001
(word-final) - (IP-final)	-55.3262	13.7786	15.52	-4.015	0.0029
(list-final) - (IP-final)	11.2089	7.5397	18.68	1.487	0.3196
Language = **Japanese**					
(word-final) - (list-final)	-46.4577	12.9613	13.33	-3.584	0.0084
(word-final) - (IP-final)	22.6189	18.9924	14.08	1.191	0.4774
(list-final) - (IP-final)	69.0766	9.2821	11.70	7.442	<.0001

Degrees-of-freedom method: kenward-roger

P value adjustment: tukey method for comparing a family of 3 estimates

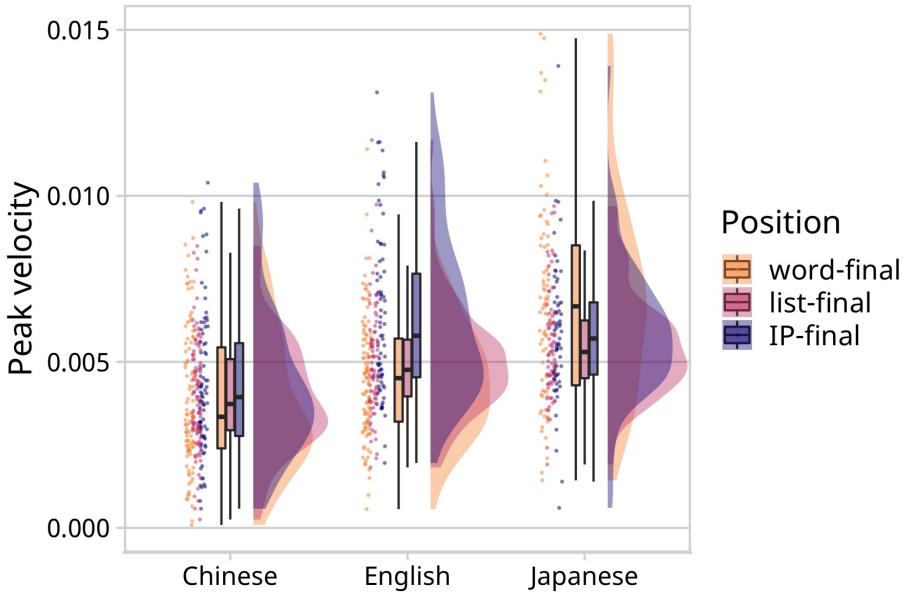

Figure 5.14: The distribution of peak velocity of /au/ trajectories.

interaction (F(4, 738.49) = 12.37, p < .005) on the peak velocity of /au/ trajectory. Post-hoc comparison are summarized in figure 5.15 and table 5.8. Chinese data did not show any difference in the prosodic contexts. English IP-final position showed faster movement with a larger peak velocity than the other two positions. Japanese /au/ was faster in the

word-final position than the list-final and the IP-final positions.

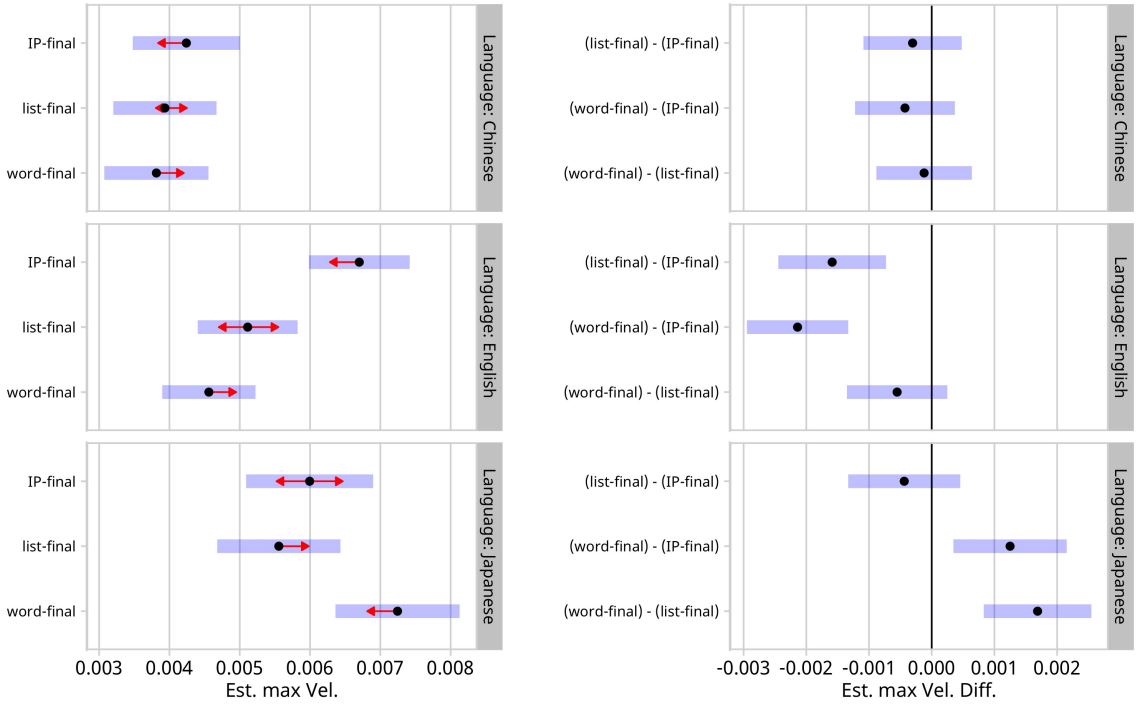

Figure 5.15: The estimated marginal means (left) and the estimated difference of peak velocity between pairwise contrasts (right) for /au/.

Table 5.8: Post-hoc pairwise comparisons of peak velocity for /au/.

contrast	estimate	SE	df	t.ratio	p.value
Language = **Chinese**					
(word-final) - (list-final)	-0.0001	0.0003	739.31	-0.374	0.9258
(word-final) - (IP-final)	-0.0004	0.0003	740.14	-1.259	0.4189
(list-final) - (IP-final)	-0.0003	0.0003	737.77	-0.915	0.6312
Language = **English**					
(word-final) - (list-final)	-0.0006	0.0003	743.77	-1.619	0.2384
(word-final) - (IP-final)	-0.0021	0.0003	739.64	-6.219	<.0001
(list-final) - (IP-final)	-0.0016	0.0004	738.95	-4.342	<.0001
Language = **Japanese**					
(word-final) - (list-final)	0.0017	0.0004	732.27	4.622	<.0001
(word-final) - (IP-final)	0.0013	0.0004	733.10	3.248	0.0035
(list-final) - (IP-final)	-0.0004	0.0004	733.44	-1.153	0.4821

Degrees-of-freedom method: kenward-roger

P value adjustment: tukey method for comparing a family of 3 estimates

Stiffness

Stiffness of /au/ trajectory is shown in figure 5.16. The larger stiffness is associated with the word-final position in English and Chinese and both the word-final and the IP-final position in Japanese.

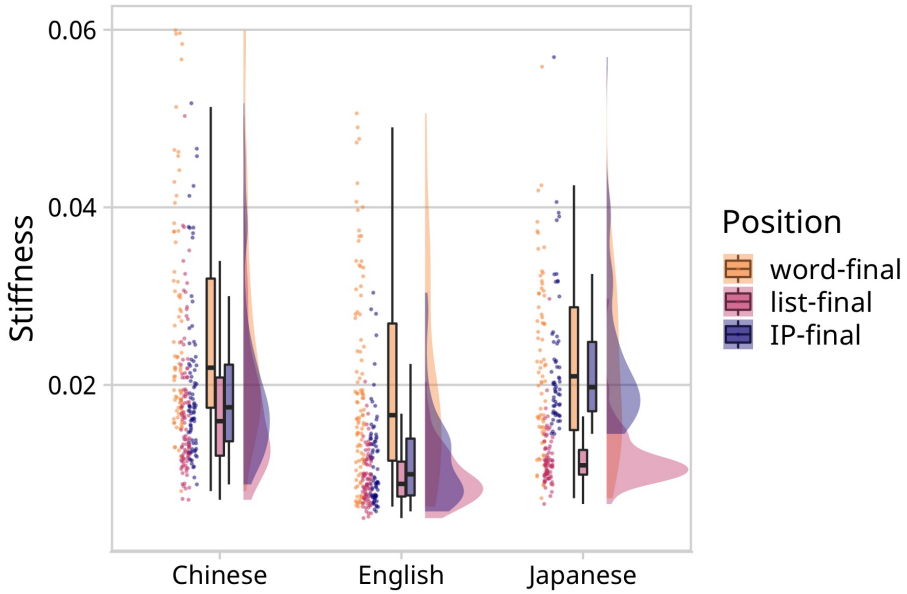

Figure 5.16: The distribution of stiffness of /ai/ trajectories.

Statistical analysis confirms the significant main effects of `Position` ($F(2, 734.49) = 54.37$, $p < .005$) and `Language` ($F(2, 14.51) = 4.62$, $p < .05$), as well as the interaction ($F(4, 734.56) = 8.05$, $p < .005$). Post-hoc comparison in figure 5.17 and table 5.9 revealed that the stiffness of Chinese /au/ trajectory was larger in the word-final position than both two other contexts. The same result was confirmed in English as well. In Japanese, the stiffness of /au/ is larger in both the word-final and the IP-final positions.

Summary of /au/ kinematics

The significant results and differences in prosodic modulation pattern on the kinematic measures are summarized in table 5.5.

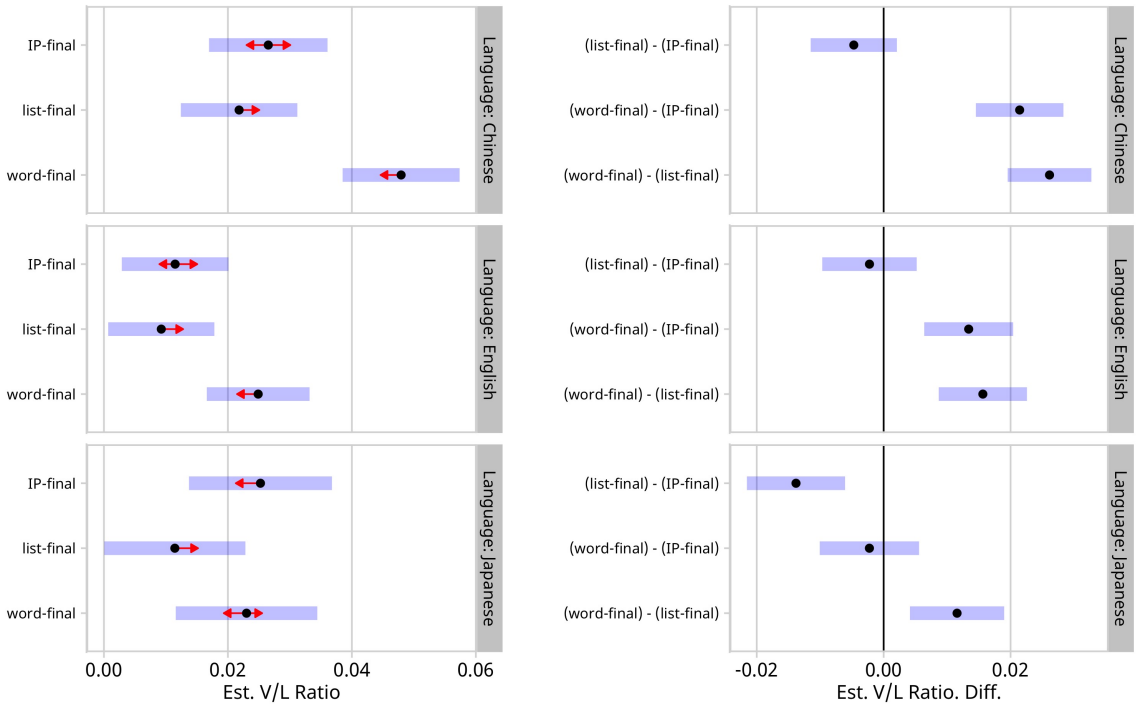

Figure 5.17: The estimated marginal means (left) and the estimated difference of stiffness between pairwise contrasts (right) for /au/.

Table 5.9: Post-hoc pairwise comparisons of stiffness for /au/.

contrast	estimate	SE	df	t.ratio	p.value
Language = **Chinese**					
(word-final) - (list-final)	0.0261	0.0028	735.34	9.306	<.0001
(word-final) - (IP-final)	0.0214	0.0029	735.91	7.297	<.0001
(list-final) - (IP-final)	-0.0047	0.0029	734.59	-1.627	0.2347
Language = **English**					
(word-final) - (list-final)	0.0156	0.0030	738.54	5.281	<.0001
(word-final) - (IP-final)	0.0134	0.0030	735.73	4.492	<.0001
(list-final) - (IP-final)	-0.0022	0.0032	735.46	-0.705	0.7606
Language = **Japanese**					
(word-final) - (list-final)	0.0116	0.0032	732.09	3.657	0.0008
(word-final) - (IP-final)	-0.0022	0.0033	732.41	-0.674	0.7785
(list-final) - (IP-final)	-0.0138	0.0033	732.54	-4.194	0.0001

Degrees-of-freedom method: kenward-roger

P value adjustment: tukey method for comparing a family of 3 estimates

Speakers from the three languages also used different strategies in prosodic modulation for /au/. In Chinese, /au/ was only minimally affected by prosodic boundaries: in word-final

Table 5.10: The summary of kinematic analysis of /au/ trajectory movement.

	word- vs. list-final	word- vs. IP-final	list- vs. IP-final
Chinese			
Displacement	n.s.	n.s.	n.s.
Duration	shorter	n.s.	n.s.
Peak velocity	n.s.	n.s.	n.s.
Stiffness	larger	larger	n.s.
English			
Displacement	smaller	smaller	n.s.
Duration	shorter	shorter	n.s.
Peak velocity	n.s.	slower	slower
Stiffness	larger	larger	n.s.
Japanese			
Displacement	n.s.	n.s.	larger
Duration	shorter	n.s.	longer
Peak velocity	faster	faster	n.s.
Stiffness	larger	n.s.	smaller

positions, /au/ was produced with shorter durations and larger stiffness without changing its displacement or peak velocity. This is very close to the strategy of *stiffness reduction* at higher prosodic boundaries, except that peak velocity was unaffected.

In English, different from /ai/, the movement of /au/ in the vowel space was modulated by a different strategy. At higher prosodic boundaries (list-final and IP-final positions), /au/ trajectory showed larger displacement, longer duration, faster movement, and smaller stiffness. This is close to *target rescaling* except that when comparing IP-final positions to the other two positions, the peak velocity was also faster with faster peak velocity and smaller stiffness but not necessarily with larger displacement. This contrasts with findings of boundary-related prosodic strengthening, whereby the stiffness modulation is how speech production is modulated.

Japanese speakers used another strategy for prosodic modulation, depending on the prosodic context. When comparing word-final positions to list-final positions, the duration is shorter, the peak velocity is faster, and the stiffness is larger, whereas the displacement

is not larger. This conforms to the strategy of *stiffness changing*. On the other hand, when comparing list-final positions to IP-final positions, the displacement is larger, the duration is longer, and the stiffness is smaller without changing its peak velocity. This seems to be *target rescaling*. The movement of /au/ is *rescaled* in Japanese. When lengthened in the list-final positions, reaching a larger target displacement with larger prosody takes longer. Again, Japanese /au/ was produced with few significant kinematic differences between word-final and IP-final positions.

5.3 /ou/

Similar to the analysis of /ou/ kinematics, the linear mixed-effect models of /ou/ peak velocity and stiffness only included `Speaker` as the random intercept as more complex models failed to converge. Otherwise, `Speaker` was included as the random effect with correlated intercepts and slopes.

Displacement

The displacement distribution for /ou/ in Chinese and English is displayed in figure 5.18. For the models of peak velocity and stiffness of /ou/, SPEAKER was only included as the random intercept as any more complicated models failed to converge.

Statistical results found that the main effects of `Position` ($F(2, 10.72) = 13.32$, $p < .005$) and `Language` ($F(1, 12.03) = 5.49$, $p < .05$) were both significant on /ou/ displacement but the interaction ($F(2, 10.72) = 2.96$) failed to reach significance. The result of Tukey-adjusted post-hoc comparison can be found in figure 5.19 and table 5.11. The results show that the displacement of /ou/ did not vary for prosodic contexts in

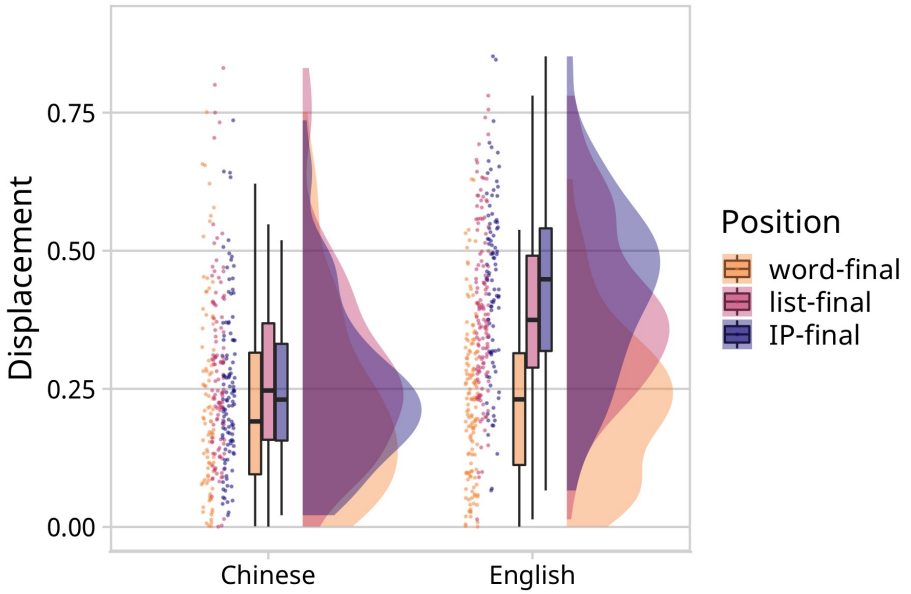

Figure 5.18: The distribution of displacement of /ou/ trajectories.

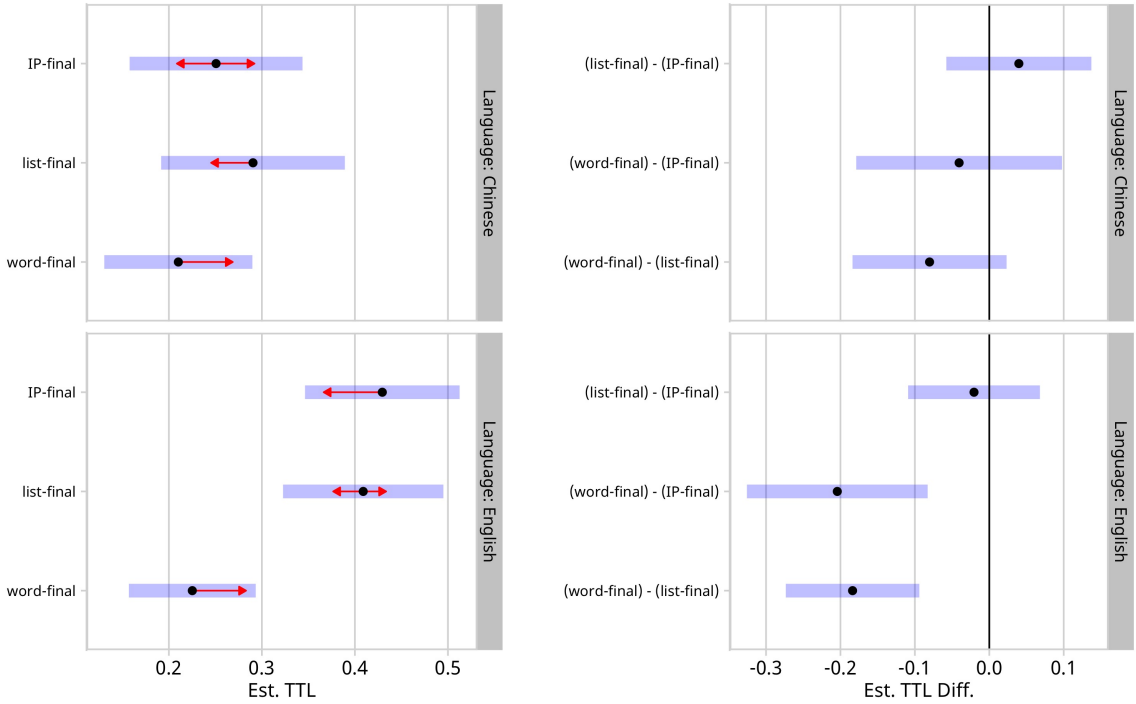

Figure 5.19: The estimated marginal means (left) and the estimated difference of displacement between pairwise contrasts (right) for /ou/.

Duration

Duration of /ou/ is demonstrated in figure 5.20. The trend is the same in Chinese and English in that the word-final position is associated with shorter durations than in the other

Table 5.11: Post-hoc pairwise comparisons of displacement for /ou/.

contrast	estimate	SE	df	t.ratio	p.value
Language = **Chinese**					
(word-final) - (list-final)	-0.0803	0.0387	11.86	-2.072	0.1380
(word-final) - (IP-final)	-0.0406	0.0515	11.45	-0.788	0.7176
(list-final) - (IP-final)	0.0397	0.0358	10.48	1.109	0.5294
Language = **English**					
(word-final) - (list-final)	-0.1837	0.0336	11.96	-5.465	0.0004
(word-final) - (IP-final)	-0.2043	0.0457	12.38	-4.469	0.0019
(list-final) - (IP-final)	-0.0205	0.0335	12.87	-0.613	0.8156

Degrees-of-freedom method: kenward-roger

P value adjustment: tukey method for comparing a family of 3 estimates

two prosodic contexts. In Chinese, the duration in the IP-final position also seems slightly shorter than in the list-final positions.

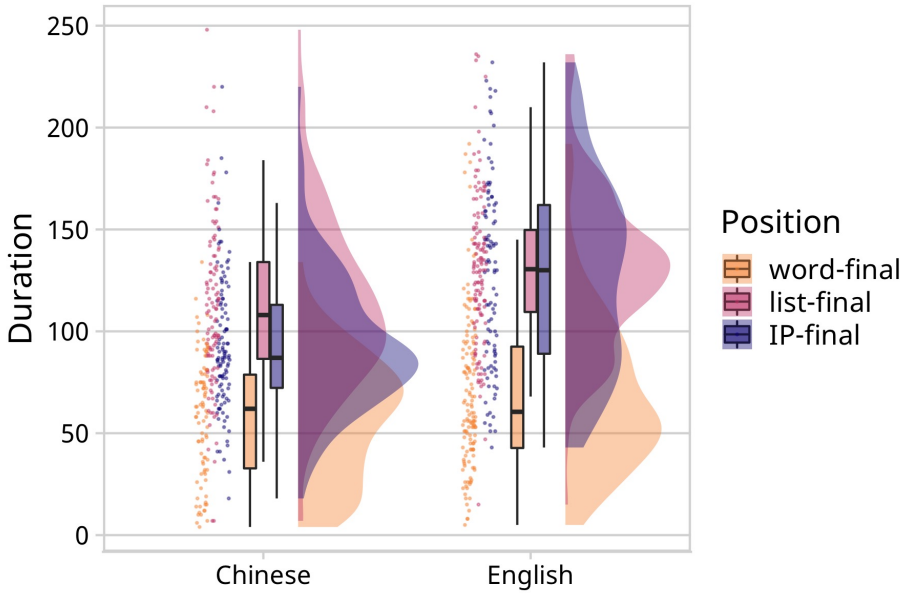

Figure 5.20: The distribution of moving duration of /ou/ trajectories.

Statistical analysis showed that the main effects of `Position` ($F(2, 11.49) = 13.31$, $p < .005$) and `Language` ($F(1, 12.03) = 8.13$, $p < .05$) were both significant on /ou/ moving duration but the interaction ($F(2, 10.71) = 2.96$) was insignificant. Post-hoc comparison in figure 5.21 and table 5.12 demonstrates that /ou/ is shorter in the word-final position than in

the list-final position in Chinese. It is also shorter in the word-final position than both the list-final and the IP-final positions. The durational difference between IP-final and list-final positions was insignificant in both languages.

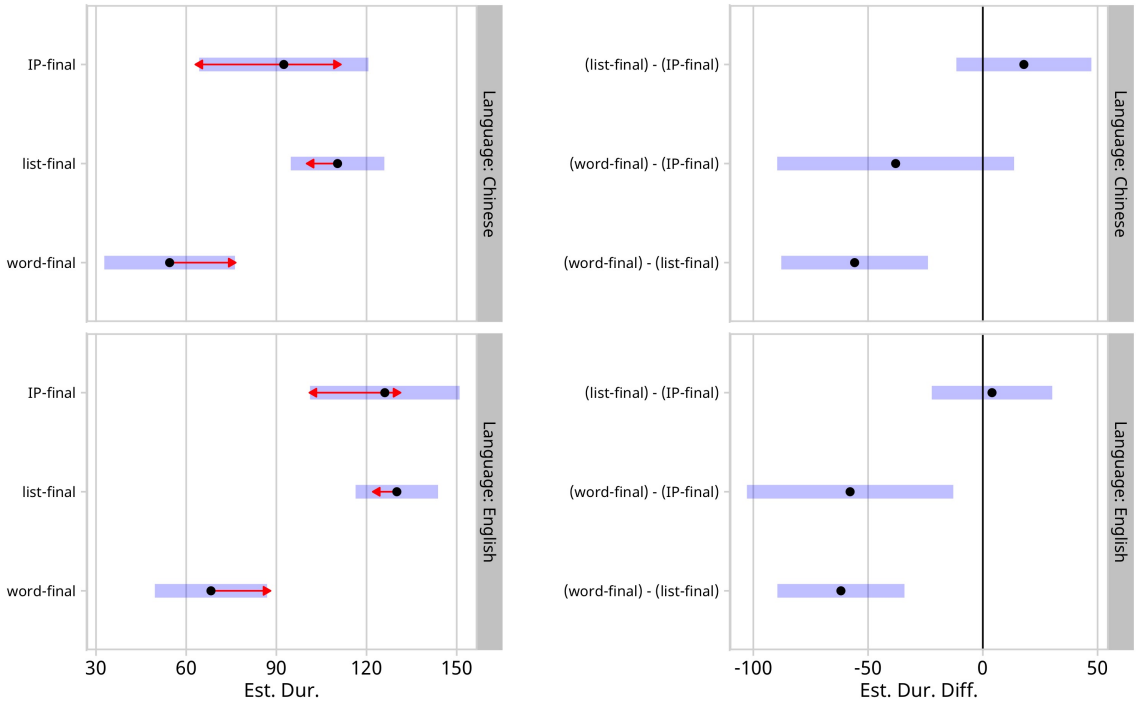

Figure 5.21: The estimated marginal means (left) and the estimated difference of duration between pairwise contrasts (right) for /ou/.

Table 5.12: Post-hoc pairwise comparisons of duration for /ou/.

contrast	estimate	SE	df	t.ratio	p.value
Language = **Chinese**					
(word-final) - (list-final)	-55.8673	11.9771	11.97	-4.665	0.0015
(word-final) - (IP-final)	-37.9864	19.2899	11.74	-1.969	0.1632
(list-final) - (IP-final)	17.8809	10.8740	10.89	1.644	0.2694
Language = **English**					
(word-final) - (list-final)	-61.8344	10.3761	11.97	-5.959	0.0002
(word-final) - (IP-final)	-57.8332	16.8891	12.19	-3.424	0.0127
(list-final) - (IP-final)	4.0012	9.9121	12.69	0.404	0.9147

Degrees-of-freedom method: kenward-roger

P value adjustment: tukey method for comparing a family of 3 estimates

Peak velocity

The overall distribution of peak velocity in Chinese and English is displayed in figure 5.22. It looks slightly faster in Chinese but not quite different in English.

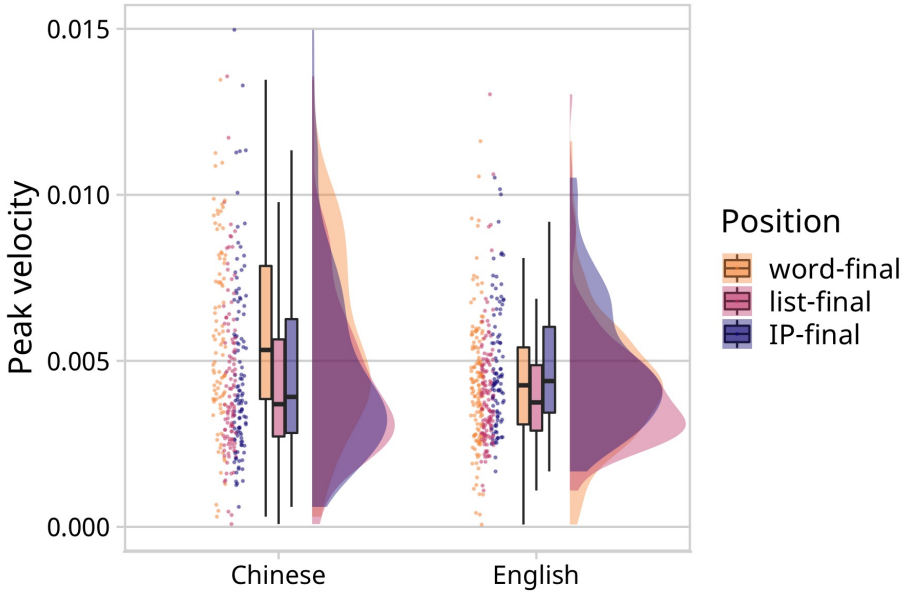

Figure 5.22: The distribution of peak velocity of /ou/ trajectories.

Statistical analysis showed that neither of the main effects of `Position` ($F(2, 666.19) = 2.39$) and `Language` ($F(1, 11.99) = 0.97$) reached significance. However, the interaction ($F(2, 666.19) = 3.97$, $p < 0.05$) was significant. Post-hoc comparison in figure 5.23 and table 5.14 showed that the only significant comparison was the larger peak velocity in the word-final position than in the list-final position.

Stiffness

The distribution of stiffness of /ou/ is illustrated in figure 5.24. The stiffness seemed larger in the word-final position in both languages. The IP-final position also had a slightly larger stiffness than the list-final position.

Statistical results show that the the main effects of `Position` ($F(2, 664.74) = 76.93$, $p <$ .005) and `Language` ($F(1, 11.99) = 0.97$, $p < .05$) as well as the interaction ($F(2, 664.74) =$

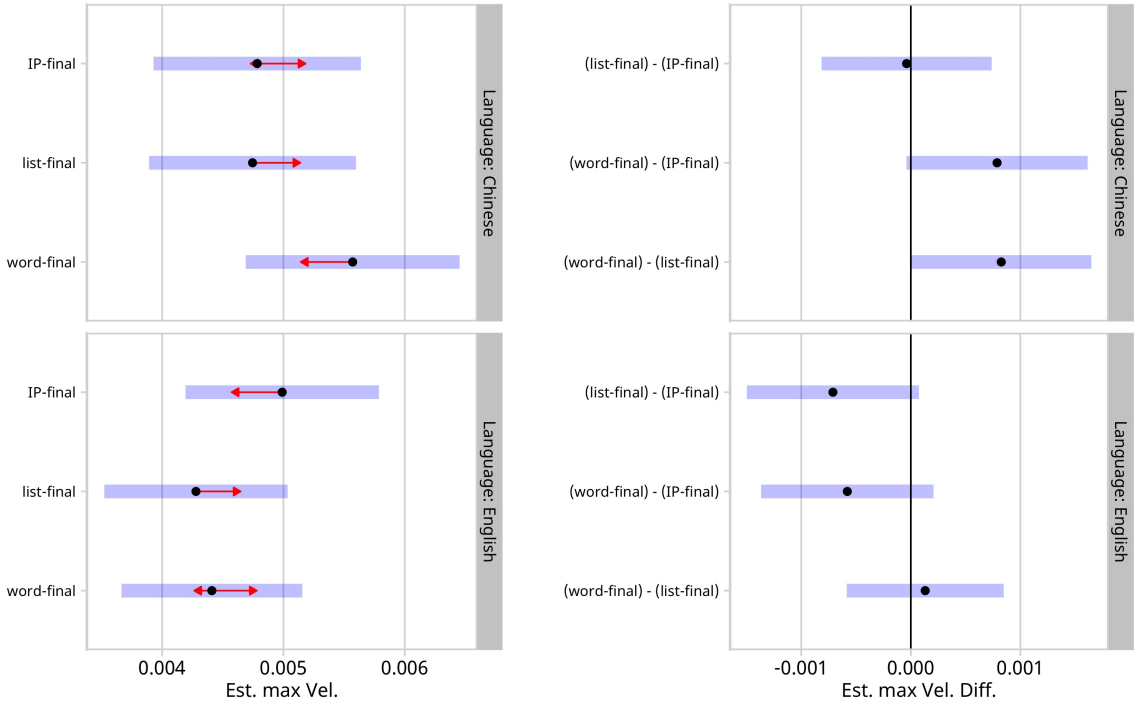

Figure 5.23: The estimated marginal means (left) and the estimated difference of peak velocity between pairwise contrasts (right) for /ou/.

Table 5.13: Post-hoc pairwise comparisons of peak velocity for /ou/.

contrast	estimate	SE	df	t.ratio	p.value
Language = **Chinese**					
(word-final) - (list-final)	0.0008	0.0004	663.25	2.357	0.0490
(word-final) - (IP-final)	0.0008	0.0004	663.77	2.236	0.0660
(list-final) - (IP-final)	-0.0000	0.0003	661.09	-0.116	0.9926
Language = **English**					
(word-final) - (list-final)	0.0001	0.0003	669.54	0.430	0.9031
(word-final) - (IP-final)	-0.0006	0.0003	672.07	-1.730	0.1948
(list-final) - (IP-final)	-0.0007	0.0003	666.29	-2.126	0.0853

Degrees-of-freedom method: kenward-roger

P value adjustment: tukey method for comparing a family of 3 estimates

3.16, $p < .05$) were all significant. Figure 5.25 and table 5.14 summarizes the results of the post-hoc comparison. In both languages, the stiffness is larger in word-final positions than in the other two prosodic contexts. The difference between the IP-final and the list-final positions was insignificant.

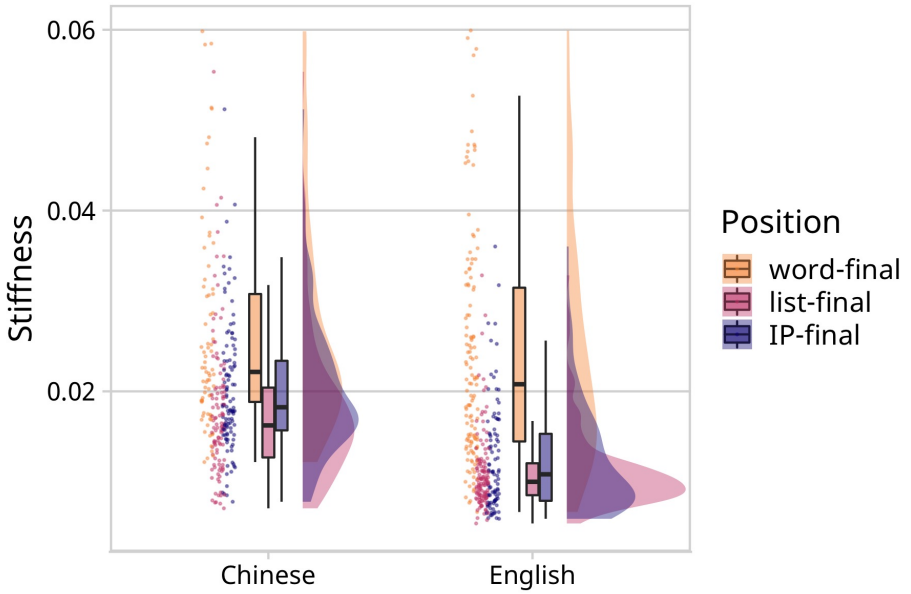

Figure 5.24: The distribution of stiffness of /ou/ trajectories.

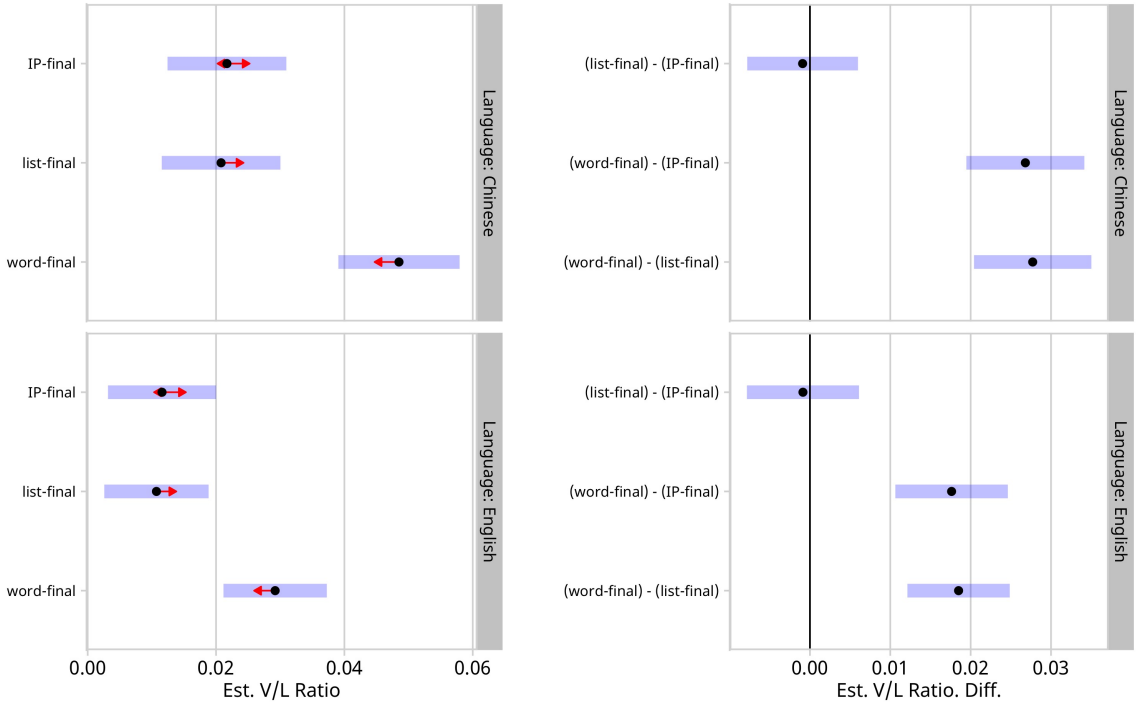

Figure 5.25: The estimated marginal means (left) and the estimated difference of stiffness between pairwise contrasts (right) for /ou/.

Summary of /ou/ kinematics

Based on the results presented above for /ou/ in Chinese and English, a summary table of significant results can be made below in table 5.15.

Table 5.14: Post-hoc pairwise comparisons of stiffness for /ou/.

contrast	estimate	SE	df	t.ratio	p.value
Language = **Chinese**					
(word-final) - (list-final)	0.0277	0.0031	662.45	8.916	<.0001
(word-final) - (IP-final)	0.0268	0.0031	662.80	8.578	<.0001
(list-final) - (IP-final)	-0.0009	0.0029	661.05	-0.311	0.9482
Language = **English**					
(word-final) - (list-final)	0.0185	0.0027	667.29	6.823	<.0001
(word-final) - (IP-final)	0.0176	0.0030	669.67	5.916	<.0001
(list-final) - (IP-final)	-0.0009	0.0030	664.55	-0.293	0.9538

Degrees-of-freedom method: kenward-roger

P value adjustment: tukey method for comparing a family of 3 estimates

Table 5.15: The summary of kinematic analysis of /ou/ trajectory movement.

	word- vs. list-final	word- vs. IP-final	list- vs. IP-final
Chinese			
Displacement	n.s.	n.s.	n.s.
Duration	shorter	n.s.	n.s.
Peak velocity	faster	n.s.	n.s.
Stiffness	larger	larger	n.s.
English			
Displacement	smaller	smaller	n.s.
Duration	shorter	shorter	n.s.
Peak velocity	n.s.	n.s.	n.s.
Stiffness	larger	larger	n.s.

From the table, we can see no differences in the kinematic measures in Chinese and English. The difference between word-final and IP-final positions in Chinese was also fragile: only stiffness was increased in word-final positions. Comparing word-final to list-final positions, we can see that /ou/ was produced with shorter duration, faster movement, and larger stiffness without changing its displacement. This is the strategy of *stiffness reduction* as seen in English /ai/ production.

In contrast, the prosodic modulation of English /ou/ does not utilize *stiffness reduction*. The vowel sequence was produced with larger displacement, longer duration, and smaller stiffness. This is *target rescaling* that increases the scale of the target entirely but does not

change the stiffness of the movement.

5.4 /ae/

The models for analyzing kinematic measures of /ae/ movement only included SPEAKER as a random intercept since more complex models all resulted in singular fits.

Displacement

Figure 5.26 shows the distribution of displacement of /ae/ trajectory in Japanese. The pattern is similar to other Japanese TVS. /ae/ moved the longest distance in the list-final positions than in the word-final and IP-final positions. The difference between word-final and IP-final positions is small.

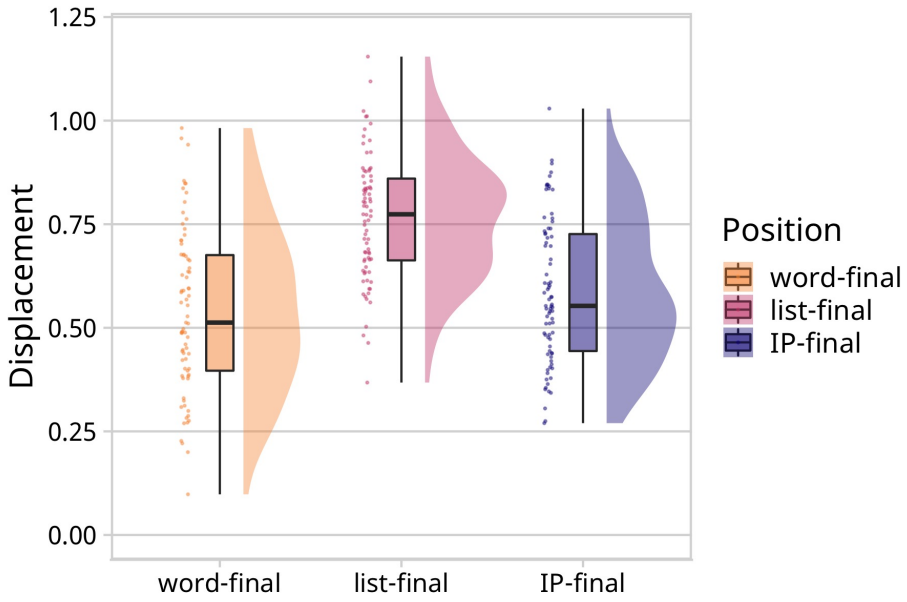

Figure 5.26: The distribution of displacement of /ae/ trajectories.

Statistical results show that the main effect of `Position` is significant ($F(2, 232.06) = 35.13$, $p < .005$). Post-hoc comparison demonstrated in figure 5.27 and table 5.16. It confirmed the observation: displacement is larger in the list-final position than in the other two conditions.

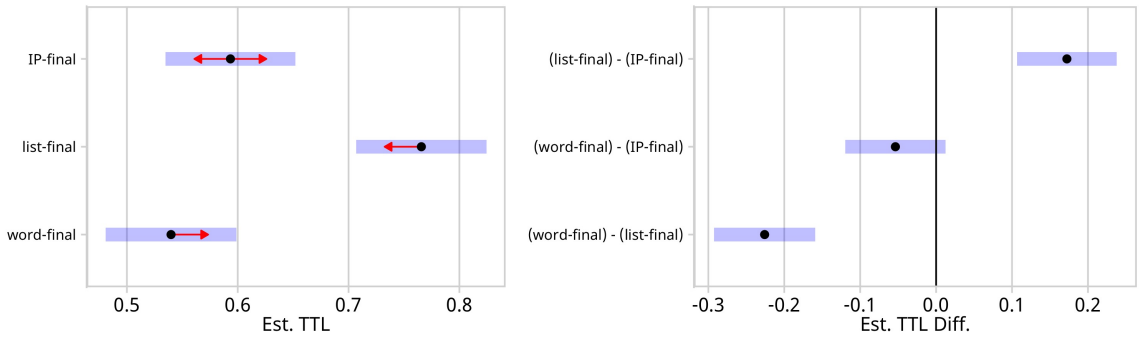

Figure 5.27: The estimated marginal means (left) and the estimated difference of displacement between pairwise contrasts (right) for /ae/.

Table 5.16: Post-hoc pairwise comparisons of displacement for /ae/.

contrast	estimate	SE	df	t.ratio	p.value
(word-final) - (list-final)	-0.2259	0.0283	232.02	-7.994	<.0001
(word-final) - (IP-final)	-0.0536	0.0280	232.05	-1.914	0.1369
(list-final) - (IP-final)	0.1723	0.0278	232.05	6.192	<.0001

Degrees-of-freedom method: kenward-roger

P value adjustment: tukey method for comparing a family of 3 estimates

Duration

The duration distribution of /ae/ follows the same pattern: the list-final position is produced with a longer duration.

The effect of `Position` is indeed statistically significant (F(2, 232.05) = 213.51, p ¡ .005). This is shown in figure 5.29 and table 5.17.

Table 5.17: Post-hoc pairwise comparisons of duration for /ae/.

contrast	estimate	SE	df	t.ratio	p.value
(word-final) - (list-final)	-70.3360	3.8092	232.01	-18.465	<.0001
(word-final) - (IP-final)	-5.5137	3.7751	232.02	-1.461	0.3119
(list-final) - (IP-final)	64.8223	3.7503	232.02	17.284	<.0001

Degrees-of-freedom method: kenward-roger

P value adjustment: tukey method for comparing a family of 3 estimates

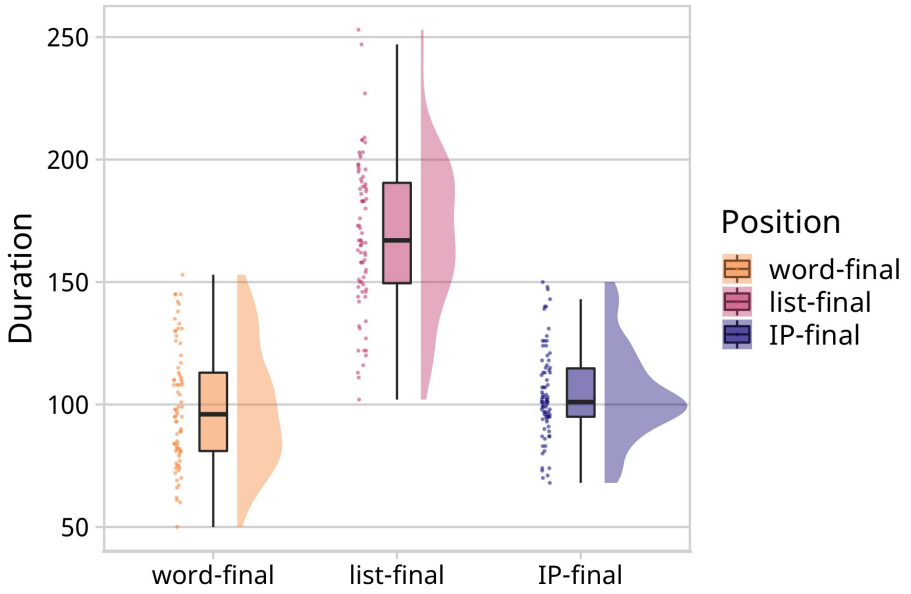

Figure 5.28: The distribution of moving duration of /ae/ trajectories.

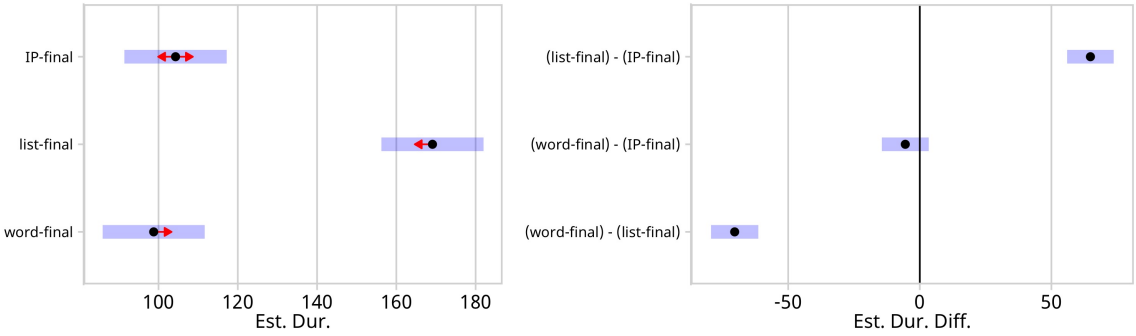

Figure 5.29: The estimated marginal means (left) and the estimated difference of duration between pairwise contrasts (right) for /ae/.

Peak velocity

Peak velocity data as displayed in figure 5.30 showed that the trajectory moved faster in word-final and IP-final positions.

This result is statistically confirmed. The main effect of `Position` ($F(2, 232.036) = 6.26$, $p < .005$) is significant. Peak velocity is larger in both word-final and IP-final positions. Figure 5.31 and table 5.19 illustrates the result of post-hoc comparison.

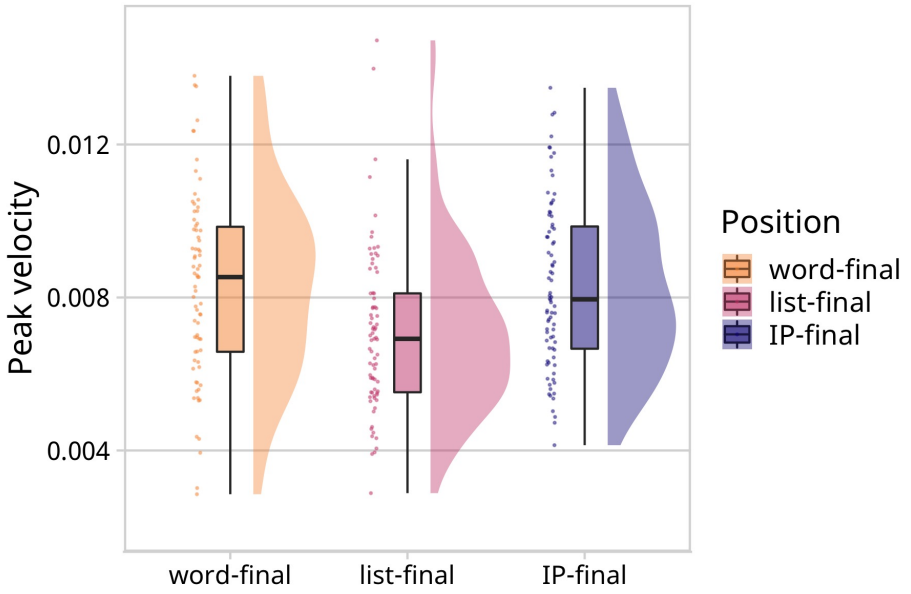

Figure 5.30: The distribution of peak velocity of /ae/ trajectories.

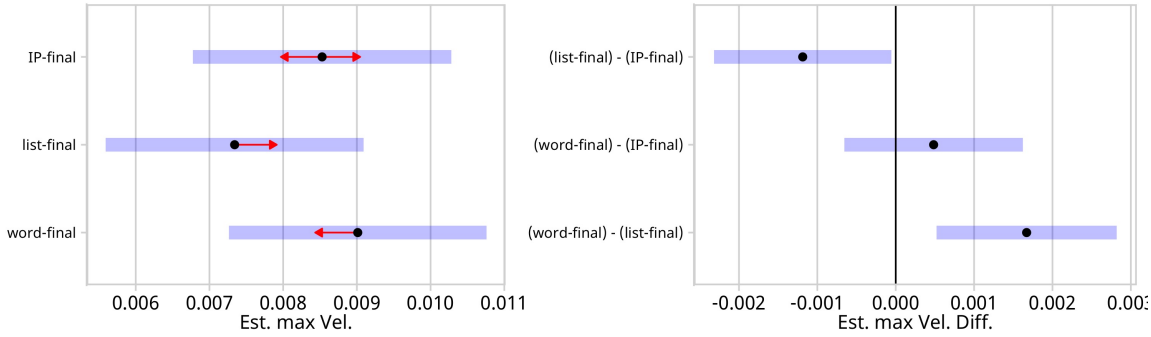

Figure 5.31: The estimated marginal means (left) and the estimated difference of peak velocity between pairwise contrasts (right) for /ae/.

Table 5.18: Post-hoc pairwise comparisons of peak velocity for /ae/.

contrast	estimate	SE	df	t.ratio	p.value
(word-final) - (list-final)	0.0017	0.0005	232.01	3.428	0.0021
(word-final) - (IP-final)	0.0005	0.0005	232.02	1.001	0.5767
(list-final) - (IP-final)	-0.0012	0.0005	232.02	-2.474	0.0373

Degrees-of-freedom method: kenward-roger

P value adjustment: tukey method for comparing a family of 3 estimates

Stiffness

Stiffness data of /ae/ shows the same trend observed in peak velocity. /ae/ in the word-final and the IP-final positions were produced with larger stiffness than in the list-final position.

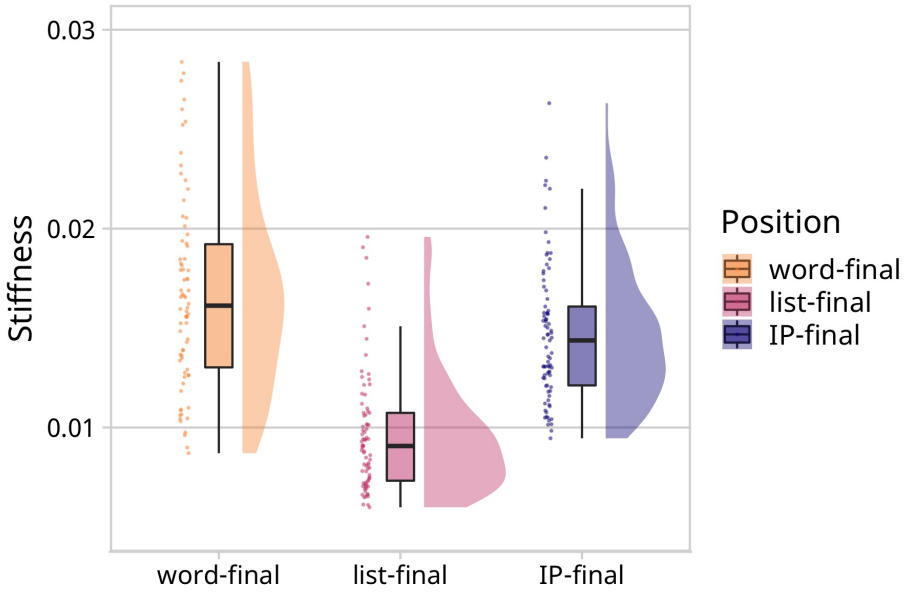

Figure 5.32: The distribution of stiffness of /ae/ trajectories.

Statistical analysis shows that this result is significant (`Position`: $F(2, 232.05) = 58.61$), $p < .005$). The result of post-hoc comparison is displayed in figure 5.33 and table 5.19.

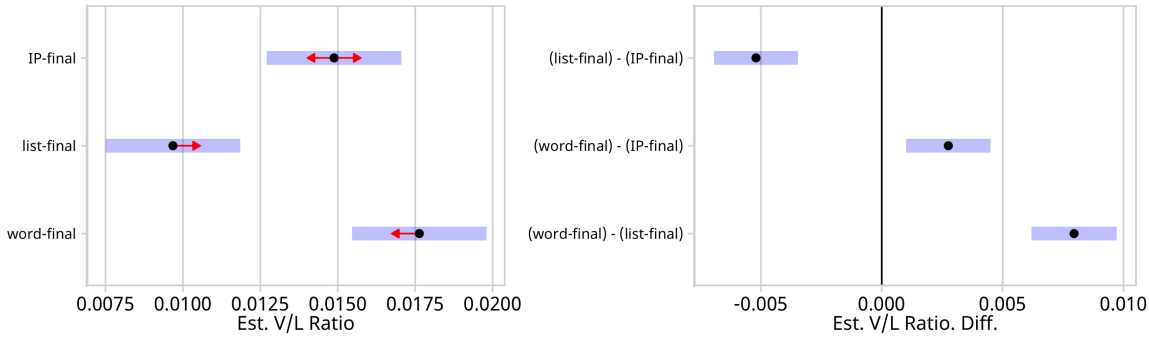

Figure 5.33: The estimated marginal means (left) and the estimated difference of stiffness between pairwise contrasts (right) for /ae/.

Summary of /ae/ kinematics

It can be seen from table 5.20 that /ae/ underwent very similar prosodic modulation as other vowel sequences in Japanese. When it occurred in word-final and IP-final positions, it was produced with smaller displacement, shorter duration, faster movement, and larger

Table 5.19: Post-hoc pairwise comparisons of stiffness for /ae/.

contrast	estimate	SE	df	t.ratio	p.value
(word-final) - (list-final)	0.0080	0.0007	232.01	10.639	<.0001
(word-final) - (IP-final)	0.0028	0.0007	232.03	3.714	0.0007
(list-final) - (IP-final)	-0.0052	0.0007	232.02	-7.068	<.0001

Degrees-of-freedom method: kenward-roger

P value adjustment: tukey method for comparing a family of 3 estimates

stiffness. From how displacement, duration, and stiffness were affected, it is probable that the movement of /ae/ in the vowel space also underwent *target rescaling*. The only difference is that the peak velocity was also faster when /ae/ occurred in lower prosodic boundaries.

Table 5.20: The summary of kinematic analysis of /ae/ trajectory movement.

	word- vs. list-final	word- vs. IP-final	list- vs. IP-final
Displacement	smaller	n.s.	larger
Duration	shorter	n.s.	longer
Peak velocity	faster	n.s.	slower
Stiffness	larger	larger	smaller

5.5 Discussion

The kinematic measures of formant movement in the vowel space revealed several interesting observations on how tautosyllabic vowel sequences are acoustically affected by the prosodic structure. To summarize the results, the strategies used are shown in table 5.21.

5.5.1 Patterns of pre-boundary strengthening

First, the pattern of prosodic modulation seemed different among the three languages. In Chinese, my data showed that /ai/ was not affected much in all the prosodic positions. /au, ou/ showed prosodic modulation on their vowel space movement only when comparing word-final and list-final positions. In English, the three TVSs were modulated by the

Table 5.21: Summary of strategies used for the movement of TVS in the vowel space.

	word-final vs. list-final	word-final vs. IP-final	list-final vs. IP-final
Chinese			
/ai/			
/au/	stiffness adjustment		
/ou/	stiffness adjustment		
English			
/ai/	stiffness adjustment	stiffness adjustment	
/au/	target rescaling	complex	
/ou/	target rescaling	target rescaling	
Japanese			
/ai/	target rescaling		target rescaling
/au/	stiffness adjustment		target rescaling
/ae/	complex		complex

prosodic positions in the comparisons involving word-final position, although the type of strategy used differs across the TVSs. In Japanese, all TVSs were affected by the prosodic contexts, and the difference came from the comparison involving list-final position, the prosodic context where TVS had the longest duration. Unlike in Chinese and English, Japanese TVSs were produced with comparable duration in word-final and IP-final positions. These results indicate that the influence of prosodic boundaries on the production of individual speech sounds is a recurrent cross-linguistic phenomenon, although the specific phonetic encodings differ. A viable assumption is that the effect can be attributed to low-level universal and automatic phonetic implementation (B. Lindblom, 1968). Studies also have suggested that pre-boundary prosodic strengthening is supralaryngeal declination throughout an utterance (Berkovits, 1994; Fowler, 1988; Tabain, 2003). Especially Berkovits (1993, 1994) found that in Hebrew, the pre-boundary lengthening progressively from the beginning to the end of a phrase-final disyllabic word. This also seems to be the case in the three languages examined here that prosodic boundary did seem to have modulated the movement of TVS in the vowel space.

But should the pre-boundary strengthening effect be considered simply stemming from

physiological and biomechanical constraints imposed on the human speech production system? Based on what has been discussed in the literature (Cho, 2015), there is ample evidence that the seemingly physiologically determined slowing-down effect is under speakers' control: many languages show language-specific granular effects which interact with other linguistic factors such as lexical stress (e.g., English (Cho et al., 2013; Shattuck-Hufnagel & Turk, 1998)), mora (e.g., Japanese (Campbell, 1992, 1999; Seo et al., 2019)), and vowel quantity (e.g., Finish (Nakai et al., 2009; Nakai et al., 2012)). For example, Turk and Shattuck-Hufnagel (2007) showed that the domain of pre-boundary lengthening in English was not confined to the final syllable but also included the stressed non-final antepenultimate syllable, indicating that pre-boundary lengthening may skip the intervening penultimate unstressed syllable and be extended to a stressed antepenultimate syllable. In Finnish, the magnitude of pre-boundary lengthening was larger in long vowels than in short vowels, suggesting that the final lengthening was attenuated on the short vowels to preserve the phonemic length contrast (Nakai et al., 2009). My data also showed that prosodic strengthening could not be simply attributed to biomechanical constraints. For instance, the movements of TVS in the vowel space examined in the current study are not always affected by prosodic strengthening. In Chinese, the boundary strength did not affect the closing diphthong /ai/. None of the four kinematic measures showed significant differences in various prosodic contexts.

Moreover, the differences between prosodic contexts where the TVSs had comparable durations were insignificant. Those contexts include list-final and IP-final positions in English and Chinese and word-final and IP-final positions in Japanese. The result is particularly interesting in Japanese that regardless of the strength of the boundary, the profile of TVS movement in word-final and IP-final positions cannot be teased from each other. Vowel sequences in these two contexts in Japanese showed a striking similarity in their kinematic profiles. This is probably due to TVSs having similar durations of proper movement in Japanese. Duration plays a vital role in determining how much displacement the

TVSs travel within the vowel space. The Pearson correlation tests for each TVS in each language show that the duration and the displacement of the movement of TVS are highly correlated. The result is shown in figure 5.34.

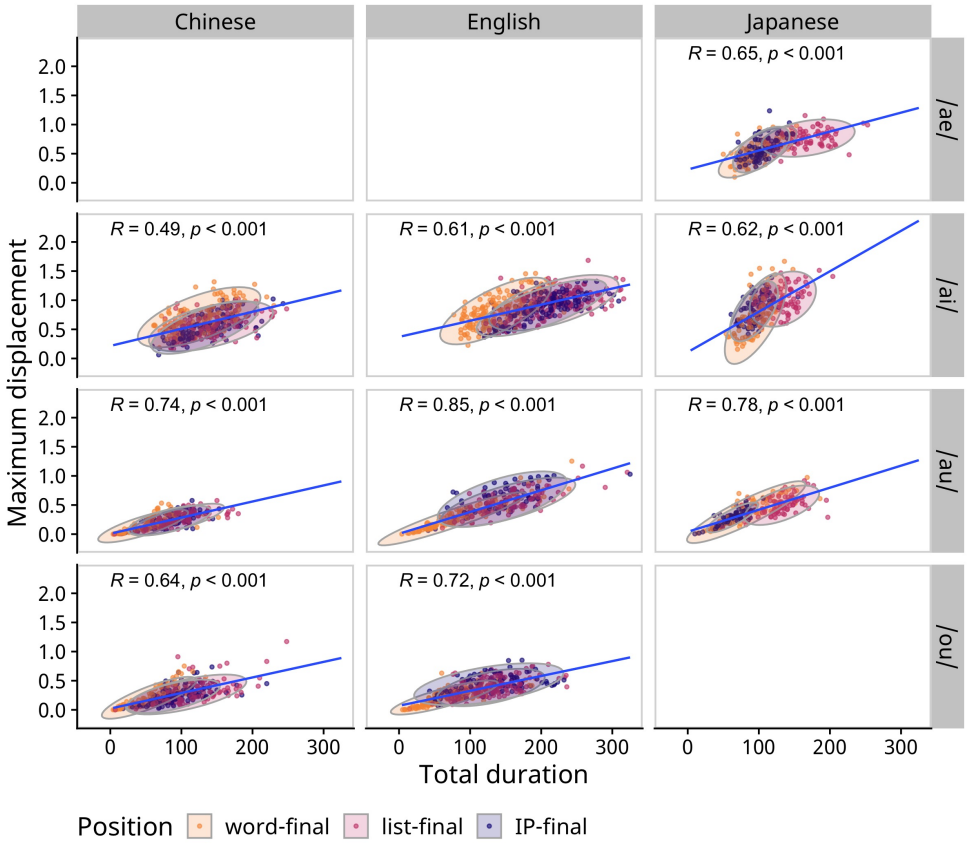

Figure 5.34: The correlation between proper duration and displacement of TVS.

This figure displayed both the cross-linguistically universality and language-specific phonetic encodings of pre-boundary strengthening. It is cross-linguistically universal in that as the duration of the TVSs increases, they travel longer in the vowel space. The correlation is significant in all cases, as shown in figure 5.34. It is language-specific in that the duration did not increase isomorphically with the strength of the prosodic boundaries in the prosodic structure. Japanese showed the longest duration for TVS in list-final positions but shorter in word-final and IP-final positions. Since target TVSs were followed by a phrase-final particle (/-ga/ for word-final positions and /-da/ for IP-final positions), this result is in line with previous research. Shepherd (2008) noted that the pre-boundary lengthening

in Japanese is primarily confined to the last mora preceding a boundary. This explains the variation we saw only in the context where the target TVS was not followed by any particles. Therefore, a viable hypothesis is that the modulation in the movement in Japanese that we observed in the current study might not directly stem from prosodic strengthening. The other possibility is that the movement itself is also constrained by duration, i.e., the time pressure that was available for the segment. When there is not enough time to reach the target, even though the boundary is higher in the prosodic structure, the segment will display phonetic properties that resemble those at lower boundaries in the prosodic hierarchy. That time pressure plays a crucial role in governing the formant excursion was also discussed in Xu and Prom-on (2019). However, further investigation is needed to fully understand the mechanism and the scope of prosodic strengthening in Japanese.

5.5.2 Strategies used in pre-boundary strengthening

In terms of what strategies were used by the languages to strengthen the TVSs preceding a boundary prosodically, it differs across the languages, the TVSs, and the prosodic contexts. The strategies that were used in the pre-boundary prosodic strengthening were not always the same. Primarily two strategies were used among the three languages: *stiffness adjustment* and *target rescaling*. In general, TVSs at a higher boundary were produced with either a reduction in stiffness that affected the duration and velocity or a rescaled target that affected the displacement, duration, and stiffness.

In Chinese, the prosodic strengthening used the same modulation strategy across prosodic contexts and TVS: *stiffness adjustment*. /au, ou/ in Chinese were produced with longer duration, slower peak velocity, and smaller stiffness (the peak velocity and the displacement ratio) when uttered at higher prosodic boundaries. Japanese /au/ and English /ai/ was also strengthened by using *stiffness adjustment*. This is in accordance with the π-gesture model (Byrd & Saltzman, 1998, 2003) that the π-gesture activated at boundaries acts as a tempo slowing down mechanism for gestures. Gestures produced under the scope of π-gesture are

longer in duration and slower in peak velocity but not necessarily with larger displacement. This result indicates that the π-gesture model can account for articulatory and acoustic movement in the F1/F2 plane. The lower peak velocity and smaller stiffness of the movement at higher boundary has been reported in other studies as well (Beckman & Edwards, 1990; Edwards et al., 1991)

However, /au, ou/ in English and /ai, au/ in Japanese used different strategies: *target rescaling*, which means they were produced with larger displacement, longer duration, and less stiffness at higher prosodic boundaries. Japanese TVS movement in vowel space also primarily used *target rescaling*. /ai, au/ were all produced with larger displacement, longer duration, and less stiffness when occurring in the list-final position than in the word-final or IP-final position. The peak velocity was not significantly affected by the prosodic boundaries. This result contrasts with some of the findings in the literature on boundary-related prosodic strengthening in the articulatory domain. Many studies have found that gestures are realized with a longer moving duration, slower peak velocity, and smaller stiffness but not with larger displacement or spatial magnitude when they are prosodically lengthened (Beckman & Edwards, 1990, 1992; Byrd & Saltzman, 1998, 2003; Edwards et al., 1991).

On the contrary, this result is in accordance with studies on postboundary or domain-initial strengthening (Bombien et al., 2013; Cho & Keating, 2001; Fougeron & Keating, 1997; Georgeton & Fougeron, 2014; Georgeton et al., 2016) and accentual strengthening (). In the postboundary articulation, it was found that the spatial displacement of the gestures was enhanced to signal the prosodic boundary. The different target of prosodic strengthening is one of the properties that distinguish pre-boundary strengthening and post-boundary strengthening (Cho, 2001, 2016; Fletcher, 2010). Fougeron (1999) suggested an explanation from a physiological perspective of spatial displacement enhancement of post-boundary strengthening, claiming that it is ascribable to the "articulatory force" associated with the postboundary positions. It is the energy necessary to realize all the muscular

effort involved in producing a consonant. This extra muscular effort results in a larger displacement. However, this explanation is hardly conceivable for pre-boundary strengthening. In pre-boundary strengthening, the physiological basis of prosodic is considered to be an articulatory declination or relaxation toward the end of a prosodic phrase (Berkovits, 1994; Fougeron & Keating, 1997; Tabain, 2003, etc.). B. E. F. Lindblom (1975) hypothesized that upcoming speech constituents are planned and stored in a planning unit: buffer. Pre-boundary strengthening reflects a general tendency to decelerate towards the end of a chunk (a prosodic phrase) with no other units to be produced further within the same buffer. Therefore, it is difficult to account for the *target rescaling* at the end of a phrase observed in the current study.

To date, no research discovered target rescaling would be the primary pre-boundary strengthening strategy. This result is even more surprising because, in the pre-boundary strengthening research mentioned above, a significant difference was found in peak velocity between lower and higher boundaries: gestures move slower at higher than lower boundaries. My study thus suggests that stiffness reduction might not be the only possibility of pre-boundary strengthening in the acoustic domain. Other possibilities also exist. Future studies must examine the supralaryngeal movement of the tongue and jaw.

Moreover, it is hard in several cases to identify the strategy used since all four kinematic measures are influenced. They include English /au/ in IP-final positions, and Japanese /ae/. The TVS were produced with larger displacement, longer duration, larger peak velocity, and smaller stiffness at higher prosodic boundaries.

Finally, there was no prosodic strengthening of the vowel space movement in comparing two durationally similar prosodic contexts: list-final and IP-final positions in Chinese and English and word-final and IP-final positions in Japanese.

5.5.3 Summary of kinematic analysis

In the kinematic analysis, we found that the prosodic contexts affect TVSs in the three languages. In Chinese, although /ai/ was not influenced at all, /au, ou/ both showed stiffness reduction occurring at higher boundaries at the end of a list and an intonational phrase. They were produced with a longer duration, smaller peak velocity, and less stiffness but not necessarily a larger displacement. This result is in accordance with most studies on pre-boundary strengthening (Byrd & Krivokapić, 2021) and is also compatible with the π-gesture model. π-gestures activate at the prosodic boundaries as a tempo slowing mechanism that do not specify vocal tract variables and subsequently does not attract articulators to move to achieve gestural targets. Instead, informational gestures act vicariously on the constriction gestures with which they are coactive. The π-gesture model proposed that "prosodic variability can emerge from the interaction of lexically specified dynamics of constriction gestures with prosodic gestures that represent phrase boundaries via time-varying modulation" (Byrd & Krivokapić, 2021, p. 36).

In English and Japanese, the picture is slightly more complicated. *Stiffness adjustment* is no longer the only strategy used. *Target rescaling* also plays a vital role in modulating the speech sound to accommodate the prosodic needs. The underlying mechanism that led to target rescaling at the boundary is still unknown. One conceivable possibility is that the longer displacement in the vowel space serves as a possible cue to signal the prosodic boundaries as it makes the initial and final target of the vowel sequence more salient. Therefore, in other words, unlike TVSs, which only underwent stiffness reduction at higher boundaries, rescaled TVSs choose to syntagmatically contrast the initial and final vocalic target in the sequence by increasing the distance they travel in the vowel space, lengthening the duration of the movement, and reducing the stiffness of the system. This seems incompatible with the π-gesture, but the π-gesutre model did not rule out magnitude changing for pre-boundary strengthening. It may increase the spatial displacement as well

(Byrd, 2000).

This is probably why there are a few cases wherein it is hard to single out one strategy being used because when the gesture is slowed down and has more time to move, it may end up being hyperarticulated and therefore increasing the displacement, even though the actual underlying mechanism is only to slow down the movement. If π-gesture can also affect the displacement, it is possible to observe that all four measures are influenced at prosodic boundaries.

In sum, the π-gesture model remains the best model to account for the variation observed at prosodic boundaries. It is valuable for modeling articulatory movement and suitable for explaining the patterns seen in the acoustic domain.

Chapter 6

General Discussion

This dissertation has investigated how prosodic boundaries condition phonetic realizations of tautosyllabic vowel sequences (TVS) by examining the formant data in three kinds of prosodic contexts: word-final position (prosodic word boundary), list-final position (intermediate phrase boundary), and IP-final position (intonational phrase boundary). The primary goal of this dissertation was to understand how prosodic strengthening affects the dynamics of formant excursions that may arise from these prosodically strong locations as manifested in the formant movement through the vowel sequence and movement kinematics in the F1/F2 vowel space. Four TVS, /ai, ae, au, ae/ in three languages, Chinese, English, and Japanese, were examined. To accomplish the analysis, acoustic data were collected from 36 speakers (12 for Chinese, 14 for English, and 10 for Japanese). The analysis was performed by examining the excursion of F1 and F2 separately with Generalized Additive Models and then the proper movement of the TVS in the F1/F2 vowel space with third-order polynomial regression. In what follows, I will summarize the results of this dissertation, with some implications of the study.

6.1 Language-specific pre-boundary prosodic modulation

6.1.1 Language-specific pre-boundary lengthening effect

The results in chapter 3 showed that the languages differ regarding whether they lengthen the TVS and the lengthening pattern and whether the lengthening effect on TVS is different than that on monophthongs.

The TVS in Chinese were lengthened much less than those in English and Japanese. Not only were fewer segments influenced by pre-boundary lengthening in Chinese, but also the magnitude of lengthening was much less in Chinese compared to that in English and Japanese. English segments showed significant differences not only between word-final position and list-final position but also between word-final position and IP-final position. The lengthening from the word-final position to the IP-final position was less evident in Japanese than in English.

The monophthong was also lengthened in both Chinese and Japanese. It is yet unknown what led to this difference between monophthongs and TVS. Two theoretical scenarios are possible. First, monophthongs are lengthened more because they are shorter in duration than TVS. The pre-boundary lengthening, however, needs to be implemented such that the lengthened duration should exceed a certain threshold to make the prosodic boundary salient to the speaker. This is a listener-oriented or perception-oriented strategy. The other possibility is that the TVS are already long enough. Hence the speakers do not need to make it much longer to reach the threshold of salient prosodic lengthening. The speakers, therefore, may have been saving energy in their speech production. This is a speaker-oriented or production-oriented strategy. Without further experiment and research, the two theoretical possibilities cannot be teased apart. Further studies are needed to fully understand the difference between lengthening on monophthongs and TVS.

Furthermore, in terms of where the lengthening effect is located differs, Chinese and

English are more alike each other. Chinese and Japanese pre-boundary lengthening happens in list-final and IP-final positions compared to word-final positions. In contrast, in Japanese, the duration of TVS in IP-final positions was not significantly different from that in the word-final positions. This is probably due to Japanese stimuli; a monosyllabic particle follows the TVS in both word-final and IP-final positions. The pre-boundary lengthening effect was too constrained in Japanese (Shepherd, 2008) such that segments only one syllable away from the boundary did not increase their duration. Future studies must further investigate the scope and magnitude of pre-boundary lengthening in Japanese.

6.1.2 Sonority expansion or hyperarticulation

The results regarding the formant movements show that neither sonority expansion nor hyperarticulation can fully account for the variability induced by prosodic boundaries.

Sonority expansion predicts that the TVS is produced with a more open mouth and lower tongue position in higher prosodic boundaries, resulting in higher F1. GAM analysis on F1 showed that F1 is barely influenced in the three languages in the first half of the TVS, indicating that the prosodic modulation on formants probably does not target the first half of the vowel or the first vocalic target /a/ of the vowel sequence. In the last half of some TVSs, such as Chinese /ai/, although F1 of /ai/ and /au/ was raised at the end at higher prosodic boundaries (list-final and IP-final positions), it is accompanied by a lowering of F2. This should be considered the anticipatory coarticulation toward the following coronal consonant instead of sonority expansion due to prosodic boundaries.

The Hyperarticulation hypothesis claims that the distinctive feature is enhanced in more prominent contexts, such as at higher prosodic boundaries. This predicts that front vowels are produced more front (higher F2) and back vowels more back (lower F2). In TVSs with a final front vowel /ai, ae/, it was confirmed that F2 was indeed higher at higher prosodic boundaries in English and Japanese, indicating that TVSs are hyperarticulated when produced in the vicinity of a higher prosodic boundary. In Chinese /ai/, the F2 is

higher at lower prosodic boundaries (word-final positions). The difference is probably due to the different phonological statuses of TVS in different languages. Chinese /ai/ only need more minor formant movements than those in English and Japanese. When TVS ends with a high back vowel (/au, ou/), F2 is always lower at the end at higher boundaries in all languages. This indicates that higher prosodic boundaries hyperarticulate the TVS.

In sum, prosodic boundaries impact the formant movements of TVS in the three languages, and the mechanism is local hyperarticulation that enhances the distinctive features of the segments.

6.1.3 Strategies of modulating vowel space movement

In chapter 5, I analyzed the four kinematic measures of the movement of TVSs in the F1/F2 vowel space, intending to extend the kinematic analysis method used for articulatory data (Byrd & Krivokapić, 2021) to acoustic data.

The result showed that it is hard to pin down a single modulation strategy for any given TVS across languages or all TVS in a specific language. Although Chinese showed a consistent use of stiffness reduction that affects the duration, peak velocity, and stiffness of the movement at higher boundaries, the primary strategies used by English and Japanese were target rescaling, which lengthens the movement with a more significant displacement and smaller stiffness, but not with faster peak velocity. There were also cases where complex strategies were used to affect all four kinematic measures. In addition, Japanese showed a strengthening effect in list-final positions as opposed to word- and IP-final positions.

Despite the variability in the strategies used for specific TVS in different languages, a general tendency was still confirmed. All the modulated movements of TVS in the vowel space show longer duration and less stiffness (peak-velocity to displacement ratio) when the TVS gets lengthened. This is a result in line with the π-gesture model that claims that a non-constriction gesture activated at the prosodic boundaries slows down the tempo of the gestures close to a prosodic boundary, making the gesture move in longer duration with

less stiffness, and occasionally with more significant displacement as well (Cho, 2001).

The analysis also showed that it is possible to extend the dynamic analysis from the articulatory domain to the acoustic domain, and it can replicate the findings previous researchers have found in articulation.

6.2 Acoustic pre-boundary strengthening and its linguistic significance

This study has shown that the production of TVSs in the acoustic domain is prosodically strengthened regarding the formant excursions and formant movement in the vowel space. The difference came from boundaries at a higher position in the prosodic hierarchy introduced in chapter 1 in Chinese and English and in segments where the duration was lengthened. What is the linguistic function of prosodically-conditioned acoustic strengthening, and how is it different from an articulatory one? In the discussion of results in each chapter, I tried to answer this question. The prosodically-conditioned pre-boundary strengthening found for various TVSs in Chinese, English, and Japanese examined in this dissertation can be interpreted to maximize the salience of the prosodic constituency. Hypothetically, the variability observed at a prosodic boundary plays a vital cue both for the speaker and the listener. The physiological constraints imposed on human language speakers dictate that pauses in speech are necessary. The speakers must adjust their respiration and articulation in a connected running speech. Therefore the optimal location of such adjustment should coincide with the edges of the linguistic units. This is widely understood as the physiological and biomechanical basis of prosodic pre-boundary strengthening/lengthening (Fletcher, 2010; B. Lindblom, 1968). At the same time, listeners also need cues to detect the intended phrasing in the speech they hear (Krivokapić, 2007; Krivokapić & Byrd, 2012; Steffman, 2019a, 2019b; White et al., 2020).

Thus, one way to maximize the salience of a boundary is to raise phonetic clarity

(Cho, 2001). In the results I have presented, the phonetic clarity of TVS was increased by showing less coarticulatory formant excursions from the environment and lengthening the movement both temporarily and spatially of TVS in the vowel space. This then allows the speaker to produce the segment in a more discernible way. It is also more accessible for the listeners to reconstruct the underlying phonological representation of the speech sounds. Furthermore, coarticulatory resistance at higher boundaries or in longer segments can also be interpreted as adding to phonetic clarity, which would otherwise be obscured by contextual influence. However, my results also showed that the onset formant values were not a target of this phonetic clarity raising mechanism in English and Chinese, indicating that speakers from different languages pick different targets to make their speech clearer.

Previous studies also suggested that phonetic clarity was enhanced to maximize the linguistic contrast (Cho, 2001), both syntagmatically (structurally) and paradigmatically (lexically or phonemically). My study replicated this result too. The onset formant values of Japanese TVS and offset formant values, especially F2, were modulated so that the distinctive feature of the initial or final vowel target was enhanced. The TVSs were hyperarticulated probably to sound more prominent than the neighboring sounds that are not adjacent to a prosodic boundary. Also, it may serve as a cue to contrast the speech sounds being uttered to other sounds in the inventory of phonemes of the language. Both can be analyzed to increase the phonetic clarity of the speech sounds.

Chapter 7

Conclusion

In this dissertation, I have analyzed the production of tautosyllabic vowel sequences (TVS) using the Generalized Additive Model and Polynomial Regression. I attempted to provide a relatively comprehensive account of the effect of prosodic boundary on the acoustics of TVS in Chinese, English, and Japanese in English in terms of prosodically-conditioned strengthening. The literature on prosodic strengthening primarily focused on analyzing the articulation of the gestures around monophthongs. The prosodic strengthening of vowel sequences like diphthongs was an understudied area of research. My research looked at the production of vowel sequences in various prosodic conditions in languages with different prosodic organizations. The results I obtained from analyzing acoustics largely replicate the findings reported for the articulatory domain, indicating a close link between articulation and acoustics for prosodic strengthening.

This dissertation suggests that phonetic realization is systematically governed by higher-level prosodic hierarchy, and the prosodically-conditioned acoustic patterns could signal higher-level prosodic structures. Much remains to be done in terms of articulation of vowel sequences. For example, questions like the difference between a tong tip raising gesture for the coda vowel in a vowel sequence and a homorganic coda consonant following a monophthong remain unexplored, and further investigation awaits.

Nevertheless, it is hoped that this dissertation will contribute to the theory of the phonetics-prosody interface.

Appendix A

Experiment stimuli

Table A.1: English stimuli.

Position	Trial
IP-final	While taking a walk in the park, we noticed a tiny tie. Nat suggested that we go and have another look.
list-final	We noticed a tiny tie, nine slices of pizza, and an old film canister near the tent.
word-final	We noticed a tiny tie next to the tent.
IP-final	While taking a walk in the park, we noticed a tiny pie. Nat suggested that we go and have another look.
list-final	We noticed a tiny pie, nine slices of pizza, and an old film canister near the tent.
word-final	We noticed a tiny pie next to the tent.
IP-final	While taking a walk in the park, we noticed a tiny guy. Nat suggested that we go and have another look.
list-final	We noticed a tiny guy, nine slices of pizza, and an old film canister near the tent.
word-final	We noticed a tiny guy next to the tent.
IP-final	When he met the King, he only took a tiny bow. Nat thought that was very rude.
list-final	When they meet the king, they took a tiny bow, nod, step back.
word-final	He took a tiny bow next to the King.
IP-final	While taking a walk in the park, we noticed a tiny cow. Nat suggested that we go and have another look.
list-final	We noticed a tiny cow, nine slices of pizza, and an old film canister near the tent.
word-final	We noticed a tiny cow next to the tent.
IP-final	While taking a walk in the park, we noticed a container of Utah tea. Nat suggested that we go and have another look.
list-final	We noticed some Utah tea, nine slices of pizza, and an old film canister near the tent.
word-final	We noticed some Utah tea lying on the ground.
IP-final	While travelling out west, we noticed a flower with a Utah bee. Nat suggested that we go and have another look.
list-final	We noticed a Utah bee, nine slices of pizza, and an old film canister near the tent.
word-final	We noticed a Utah bee next to the tent.
IP-final	While taking a walk in the park, we noticed a long key. Nat suggested that we go and have another look.
list-final	We noticed a long key, nine slices of pizza, and an old film canister near the tent.
word-final	We noticed a long key lying on the ground.
IP-final	Before I went to Kentucky, I learned a bit about Kentucky law. Nat thought that wasn't necessary.
list-final	He seems to know Kentucky law, New York law and New Jersey law very well.
word-final	There will be a Kentucky law to curb drug abuse.
IP-final	While taking a walk in the park, we noticed a toy paw. Nat suggested that we go and have another look.
list-final	We noticed a toy paw, nine slices of pizza, and an old film canister near the tent.
word-final	We noticed a toy paw next to the tent.
IP-final	This week, I have a fever and a dry cough. Nat suggested that I go get a COVID-19 test.
list-final	I had a dry cough, nausea, and muscle pain a few days ago.
word-final	I started to have a dry cough two days ago.
IP-final	While taking a walk in the park, we noticed a toy toe. Nat suggested that we go and have another look.
list-final	We noticed a toy toe, nine slices of pizza, and an old film canister near the tent.
word-final	We noticed a toy toe next to the tent.
IP-final	While taking a walk in the park, we noticed a toy bow. Nat suggested that we go and have another look.
list-final	We noticed a toy bow, nine slices of pizza, and an old film canister near the tent.
word-final	We noticed a toy bow next to the tent.
IP-final	While taking a walk in the park, we noticed a set of mini Go. Nat suggested that we go and have another look.
list-final	We noticed a set of mini Go, nine slices of pizza, and an old film canister near the tent.
word-final	We noticed a set of mini Go near the tent.

Table A.2: Japanese stimuli.

Position	Trial
IP-final	tanaka san no ichiban jouzuna gakki wa gitaa da. hureau koto de sore wo sokushin su beki da.
list-final	tanaka san no sukina gakki to ieba gitaa, shamisen, piano nado ga agerareru.
word-final	tanaka san wa gitaa ga suki da.
IP-final	kore wa senshuu katta bakari no waipaa da. konnani hayaku kowareru to wa bikkuri shita.
list-final	sensyuu, kuruma no waipaa, shiitokussyon to hoiiru wo kounyuu shita.
word-final	mou shi-go nen gurai tsukatte kita node, atarashii waipaa ga hitsuyou da.
IP-final	kore wa juunen gurai aiyoo sitekita maikaa da. raigetsu ni wa haisya ni naru.
list-final	tanaka san ha maikaa, jibun no ie to goruhusetto wo motsu no wo jinsei no yume ni shite iru.
word-final	ima motte iru maikaa ni purasu suru dake de, syanai wo gureedoappu dekiru.
IP-final	sore wa mattaku setsumei ni wa naranai kotae da. mou ichido kotaete kudasai
list-final	eigo no ansaa to iu go wa, nihongo de kotae, kaitou, hentou nado to yakusareru.
word-final	kono mondai ga muzukashi sugite, nakanaka ii kotae ga mitsukaranai.
IP-final	sore wa ii dekibae da. shousan ni atai suru to omou.
list-final	ooku no hito wa shigoto no dekibae, douryou to no ningenkankei, kazoku to no kouryuu ni sutoresu wo kanjiru.
word-final	kondo ha dekibae ga warukunai ne.
IP-final	konkai mottomo juuyouna no wa IT gyoumu he no okikae da. sore wo tanakasan ni makasetai.
list-final	IT gyoumu he no okikae, shinki syain no koyou to, kenshuu teema no kettei wa kondo tanaka san ga tantou shite iru.
word-final	tanaka san no okage de, IT gyoumu heno okikae ga junchou ni susunde iru.
IP-final	ichiban uragiritaku nai no ha tanaka kantoku no kitai da. nanode zutto issyoukenmei ganbatte iru.
list-final	kantoku no kitai, nakama to no kizuna, kazoku kara no shiji ha
word-final	kokusai syakai kara no kitai ga takamatta.
IP-final	sore wa tanaka san ga tsukutte kureta piichi pai da. atsui uchi ni tabete kudasai.
list-final	tanakasan wa piichi pai, chokoreetomuusu, burauni nado wo yoku tukutte iru.
word-final	tanaka san no tsukutta piichi pai ga oishikatta.
IP-final	ibunkakouryuu ni daiji nano wa sougo rikai da. hureau koto de sore wo sokushin su beki da.
list-final	tagai no rikai, doujou to kyoukan de ibunkakouryuu wo susumeru
word-final	ibunka no kontekisuto ni taisuru rikai ga nakanaka susuma nai.
IP-final	sore wa girishia moji no tau da. raten moji no thii ni soutou suru.
list-final	koko ni shimesite aru tau, shiguma, pai wa minna girishia moji nano da.
word-final	girishia moji no tau ga raten moji no thii ni soutou suru.
IP-final	uchi no inu no namae wa bau da. kantan de yobiyasui kara so no namae wo tuketa no.
list-final	uchi no inu wa sorezore bau, sora, mugi to iu.
word-final	uchi no bau ga keiki wo tabetyatta.
IP-final	tanaka san to hajimete atta no wa oranda no gau da. sore ha mada samui toki datta.
list-final	konkai yooroppa ni itta toki, oranda no gau, napori to, berugii no hento wo otozureta.
word-final	kondo no sinpojiumu wa, oranda no gau de okonawareru yotei da.
IP-final	kono ko wa koneko no chii da. doubutsu byouin to jidousya wo kiratte iru.
list-final	musukoni koneko no chii, koguma no kabii to anpanman wo katte yatta.
word-final	koneko no chii ga michi ni mayotta.
IP-final	musukono hoshii no wa koguma no kabii da. kurisumasu no purezento ni shitai.
list-final	musuko ni koguma no kabii, neko no chii to anpanman wo katte yatta.
word-final	koguma no kabii ga hatsubai shita bakari desu.
IP-final	shihutokii no ueni aru no ha entaa kii da. sore wo osu to tugi no peeji ni susumeru.
list-final	entaa kii, shihuto kii, supeesubaa wa mottomo juuyou na kii to naru.
word-final	entaa kii ga kowareta node,
IP-final	ibunkakouryuu ni mottomo taisetsuna no wa enkatsu na ishi sotsuu da. sore ga dekinai to nanimo dekinai
list-final	shokuba de seikou suru ni wa, tanin to no ishi sotsuu, taimu maneijimento ya, rinki ouhen no shisei ga juuyou da.
word-final	kondai no taidan wa ishi sotsuu ga dekinai mama owatta.
IP-final	nihon de yoku tukawareru kensaku enjin wa yahoo da. amerika de wa guuguru ni naru.
list-final	kensaku enjin wa yahuu, guuguru to bingu nado ga aru.
word-final	kiiwaado wo kaete yahuu de mou ichido sagashimashou.
IP-final	kono sutoorii ha rekishi jiken ni motoduita kakuu da. shinjitsu de wa nai.
list-final	rekisi ni motoduita kakuu, syakai he no kansatsu, hakuryoku no aru byousya wa, ano sakka no seikou shita kii da to omou.
word-final	rekishi ni motoduita kakuu de kaita syousetsu wo, rekishi kakuu syousetsu to iu.

Table A.3: English translations of Japanese stimuli.

Table A.3.

Position	Meaning
IP-final	Tanaka is best at playing guitar. He has played for almost 20 years.
list-final	Speaking of Mr. Tanaka's favorite instruments, they are guitar, shamisen, and piano.
word-final	Mr. Tanaka is addicted to playing guitar.
IP-final	This is the wiper I bought just last week. I never thought it could be broken so quickly.
list-final	Last week, I bought a new wiper, seat pad, and wheels.
word-final	I have used this wiper for 4 or 5 years. Time to get a new one.
IP-final	This is my car that I have enjoyed driving for 20 years. I feel sad that it will have to be totaled next month.
list-final	Mr. Tanaka's life dreams are to get a car, a golf set, and own a house.
word-final	You can upgrade the cabin of your car by only investing a little bit more.
IP-final	This answer doesn't explain it at all. Please answer again.
list-final	The word answer in English is translated as 'kotae', 'kaitou', and 'hentou'.
word-final	This problem is so difficult. It's hard to find a good solution.
IP-final	The performance was really good. It deserves applause.
list-final	Many people feel stressed out worrying about their job performance, their relationship with their family and colleagues.
word-final	The performance this time was not bad.
IP-final	The most important task this time is changing the business to IT. I want Mr. Tanaka to take charge of it.
list-final	Mr. Tanaka took charge of changing business to IT, determining the topic of employee training, and hiring.
word-final	Thanks to Mr. Tanaka, changing the business to IT business was smoothly done.
IP-final	I don't want to fail Coach Tanaka's expectations. I will train as hard as I can.
list-final	I have been motivated by the coach's expectations, the bond with my friends, and the support from my family.
word-final	It bears a high expectation internationally.
IP-final	This is a peach pie made by Mr. Tanaka. Let's eat it while it's still warm.
list-final	Mr. Tanaka likes to make peach pie, chocolate mousse, and brownies.
word-final	The peach pie made by Mr. Tanaka was delicious.
IP-final	Mutual understanding is the most important thing in cross-cultural communication. People should deepen it by keeping in contact with each other
list-final	Mutual understanding, empathy, and compassion can improve cross-cultural communication.
word-final	It's pretty hard to fully understand the context in another culture.
IP-final	This is the Greek letter 'τ'. It corresponds to the Roman letter 't'.
list-final	The letters 'τ', 'π', 'σ' displayed here are all Greek letters.
word-final	The Greek letter 'τ' is just the Roman letter 't'.
IP-final	My dog's name is Bau. It's simple and easy to pronounce.
list-final	My dogs are called Bau, Sora, and Mugi.
word-final	My dog Bau ate the cake.
IP-final	I met with Mr. Tanaka in Gau, Netherland for the first time. It was cold back then.
list-final	I visited Gau (Netherland), Naples, and Gent during the trip to Europe this time.
word-final	This symposium will be held in Gau, Netherland.
IP-final	This is Chee the cat. He doesn't like vets and vehicles.
list-final	I bought Chee the cat, Kaby the bear, and Anpanman for my son.
word-final	The cat Chee got lost.
IP-final	My son wants a Kaby the bear. I'm planning to get one for him as a Christmas present.
list-final	I bought Kaby the bear, Chee the cat, and Anpanman for my son.
word-final	The toy, Kaby the bear, was just released.
IP-final	What's above shift key is the enter key. Press it to proceed.
list-final	Enter key, space bar, and backspace are the most important keys on a keyboard.
word-final	The enter key was broken. So I bought a new keyboard.
IP-final	The most important thing in cross-cultural communication is mutual understanding. Nothing would be achievable without it.
list-final	To succeed in a job, it's important to have good communication with your colleague, good time management skills, and the ability to be flexible.
word-final	The dialogue this time ended without a mutual understanding.
IP-final	The most popular search engine is Yahoo. It's Google in America.
list-final	As for search engines, there are Yahoo, Google, and Bing.
word-final	Change to another keyword and search again in Yahoo.
IP-final	This story is a fantasy inspired by historical events. It is not true.
list-final	Stories based on real history, spectacular depiction, and a keen observation of society are the keys to that author's success.
word-final	Alternate history novels are novels that are written based on real historical events.

Appendix B

Summary of GAMs

A. parametric coefficients	Estimate	Std. Error	t-value	p-value
(Intercept)	0.6842	0.0235	29.1403	< 0.0001
pos.ord.L	0.0383	0.0209	1.8319	0.0670
pos.ord.Q	-0.0226	0.0197	-1.1466	0.2516
B. smooth terms	edf	Ref.df	F-value	p-value
s(Time)	12.3029	13.3552	60.6805	< 0.0001
s(Time):pos.ordlist-final	5.6620	7.0189	21.7025	< 0.0001
s(Time):pos.ordIP-final	3.7839	4.4081	13.5202	< 0.0001
s(Time,Speaker)	92.9734	150.0000	4.5557	< 0.0001
s(Time,Speaker):pos.ordlist-final	64.1686	178.0000	11.1248	0.5235
s(Time,Speaker):pos.ordIP-final	100.3605	178.0000	6297.4624	< 0.0001
s(Block):Positionword-final	1.1227	1.2283	0.1453	0.8689
s(Block):Positionlist-final	1.6817	1.8982	10.2983	0.0004
s(Block):PositionIP-final	1.0001	1.0003	17.6037	< 0.0001

Table B.1: Summary of F1 model of Chinese /ai/.

A. parametric coefficients	Estimate	Std. Error	t-value	p-value
(Intercept)	1.6937	0.0468	36.1548	< 0.0001
pos.ord.L	-0.0895	0.0226	-3.9548	0.0001
pos.ord.Q	0.0625	0.0213	2.9367	0.0033
B. smooth terms	edf	Ref.df	F-value	p-value
s(Time)	10.7550	11.8328	32.2155	< 0.0001
s(Time):pos.ordlist-final	6.0897	7.6784	8.2896	< 0.0001
s(Time):pos.ordIP-final	4.9376	6.2765	9.3395	< 0.0001
s(Time,Speaker)	93.8689	150.0000	6.9771	< 0.0001
s(Time,Speaker):pos.ordlist-final	52.5698	178.0000	53.7276	0.0004
s(Time,Speaker):pos.ordIP-final	47.6093	178.0000	8.3104	0.2008
s(Block):Positionword-final	1.6768	1.8949	0.9612	0.3710
s(Block):Positionlist-final	1.0003	1.0006	0.6860	0.4077
s(Block):PositionIP-final	1.0004	1.0008	0.4067	0.5240

Table B.2: Summary of F2 of Chinese /ai/.

A. parametric coefficients	Estimate	Std. Error	t-value	p-value
(Intercept)	0.6545	0.0183	35.6822	< 0.0001
pos.ord.L	0.0310	0.0137	2.2694	0.0233
pos.ord.Q	-0.0118	0.0149	-0.7914	0.4287
B. smooth terms	edf	Ref.df	F-value	p-value
s(Time)	11.9384	12.8852	45.9826	< 0.0001
s(Time):pos.ordlist-final	7.6727	9.4852	17.4611	< 0.0001
s(Time):pos.ordIP-final	8.8495	10.7361	23.2650	< 0.0001
s(Time,Speaker)	100.6545	150.0000	6.8176	< 0.0001
s(Time,Speaker):pos.ordlist-final	65.3161	178.0000	1878.4628	< 0.0001
s(Time,Speaker):pos.ordIP-final	74.2225	178.0000	1093.6185	< 0.0001
s(Block):Positionword-final	1.2526	1.4413	14.5136	< 0.0001
s(Block):Positionlist-final	1.0004	1.0008	20.5491	< 0.0001
s(Block):PositionIP-final	1.2985	1.5075	4.0095	0.0191

Table B.3: Summary of F1 model of Chinese /au/.

A. parametric coefficients	Estimate	Std. Error	t-value	p-value
(Intercept)	1.1191	0.0413	27.1132	< 0.0001
pos.ord.L	-0.0576	0.0140	-4.1161	< 0.0001
pos.ord.Q	0.0420	0.0169	2.4901	0.0128
B. smooth terms	edf	Ref.df	F-value	p-value
s(Time)	11.4160	12.5873	48.8447	< 0.0001
s(Time):pos.ordlist-final	9.4918	11.7967	43.2784	< 0.0001
s(Time):pos.ordIP-final	10.0767	13.0465	50.6003	< 0.0001
s(Time,Speaker)	95.3938	149.0000	5.0217	< 0.0001
s(Time,Speaker):pos.ordlist-final	59.8851	178.0000	62.9951	0.0057
s(Time,Speaker):pos.ordIP-final	29.5611	178.0000	459.0368	< 0.0001
s(Block):Positionword-final	1.0002	1.0005	2.6437	0.1039
s(Block):Positionlist-final	1.0002	1.0005	0.1675	0.6825
s(Block):PositionIP-final	1.0003	1.0006	3.7594	0.0525

Table B.4: Summary of F2 model of Chinese /au/.

A. parametric coefficients	Estimate	Std. Error	t-value	p-value
(Intercept)	0.4692	0.0230	20.3863	< 0.0001
pos.ord.L	0.0085	0.0070	1.2231	0.2213
pos.ord.Q	-0.0043	0.0103	-0.4213	0.6736
B. smooth terms	edf	Ref.df	F-value	p-value
s(Time)	11.4378	12.6722	16.6047	< 0.0001
s(Time):pos.ordlist-final	11.4215	14.5572	6.7376	< 0.0001
s(Time):pos.ordIP-final	10.4784	14.0170	5.8967	< 0.0001
s(Time,Speaker)	96.7343	150.0000	5.3908	< 0.0001
s(Time,Speaker):pos.ordlist-final	37.5359	178.0000	138.1269	0.0030
s(Time,Speaker):pos.ordIP-final	4.0860	177.0000	28.4773	0.1506
s(Block):Positionword-final	1.0003	1.0006	0.6789	0.4101
s(Block):Positionlist-final	1.2001	1.3597	0.0856	0.7918
s(Block):PositionIP-final	1.0002	1.0004	3.0227	0.0821

Table B.5: Summary of F1 model of Chinese /ou/

A. parametric coefficients	Estimate	Std. Error	t-value	p-value
(Intercept)	0.9027	0.0394	22.9312	< 0.0001
pos.ord.L	-0.0830	0.0144	-5.7726	< 0.0001
pos.ord.Q	0.0468	0.0156	3.0067	0.0026
B. smooth terms	edf	Ref.df	F-value	p-value
s(Time)	13.8489	15.9106	95.3117	< 0.0001
s(Time):pos.ordlist-final	10.8313	14.3083	97.4830	< 0.0001
s(Time):pos.ordIP-final	11.5886	15.4105	100.2043	< 0.0001
s(Time,Speaker)	68.5661	149.0000	3.4026	< 0.0001
s(Time,Speaker):pos.ordlist-final	13.5971	178.0000	5.0621	0.2947
s(Time,Speaker):pos.ordIP-final	4.4327	177.0000	5.5567	0.2970
s(Block):Positionword-final	1.0002	1.0003	1.9951	0.1578
s(Block):Positionlist-final	1.0002	1.0005	5.0002	0.0253
s(Block):PositionIP-final	1.0002	1.0004	1.8790	0.1705

Table B.6: Summary of F2 model of Chinese /ou/.

A. parametric coefficients	Estimate	Std. Error	t-value	p-value
(Intercept)	0.6655	0.0340	19.5934	< 0.0001
pos.ord.L	-0.0164	0.0222	-0.7368	0.4613
pos.ord.Q	0.0075	0.0242	0.3105	0.7562
B. smooth terms	edf	Ref.df	F-value	p-value
s(Time)	7.6435	8.3736	27.9060	< 0.0001
s(Time):pos.ordlist-final	9.0811	10.4991	7.9337	< 0.0001
s(Time):pos.ordIP-final	13.2126	14.4711	9.2130	< 0.0001
s(Time,Speaker)	141.7993	200.0000	10.5573	< 0.0001
s(Time,Speaker):pos.ordlist-final	131.9065	238.0000	1315.6844	< 0.0001
s(Time,Speaker):pos.ordIP-final	162.2960	238.0000	190.5667	0.0002
s(Block):Positionword-final	1.0004	1.0008	5.6710	0.0172
s(Block):Positionlist-final	1.0002	1.0004	4.7118	0.0300
s(Block):PositionIP-final	1.0004	1.0007	0.2533	0.6148

Table B.7: Summary of F1 model of English /ai/.

A. parametric coefficients	Estimate	Std. Error	t-value	p-value
(Intercept)	0.6655	0.0340	19.5934	< 0.0001
pos.ord.L	-0.0164	0.0222	-0.7368	0.4613
pos.ord.Q	0.0075	0.0242	0.3105	0.7562
B. smooth terms	edf	Ref.df	F-value	p-value
s(Time)	7.6435	8.3736	27.9060	< 0.0001
s(Time):pos.ordlist-final	9.0811	10.4991	7.9337	< 0.0001
s(Time):pos.ordIP-final	13.2126	14.4711	9.2130	< 0.0001
s(Time,Speaker)	141.7993	200.0000	10.5573	< 0.0001
s(Time,Speaker):pos.ordlist-final	131.9065	238.0000	1315.6844	< 0.0001
s(Time,Speaker):pos.ordIP-final	162.2960	238.0000	190.5667	0.0002
s(Block):Positionword-final	1.0004	1.0008	5.6710	0.0172
s(Block):Positionlist-final	1.0002	1.0004	4.7118	0.0300
s(Block):PositionIP-final	1.0004	1.0007	0.2533	0.6148

Table B.8: Summary of F1 model of English /ai/.

A. parametric coefficients	Estimate	Std. Error	t-value	p-value
(Intercept)	1.5320	0.0367	41.6962	< 0.0001
pos.ord.L	-0.0040	0.0193	-0.2079	0.8353
pos.ord.Q	-0.0222	0.0196	-1.1331	0.2572
B. smooth terms	edf	Ref.df	F-value	p-value
s(Time)	14.5969	15.6007	55.7472	< 0.0001
s(Time):pos.ordlist-final	2.9819	3.5387	8.0549	< 0.0001
s(Time):pos.ordIP-final	7.2310	8.6487	2.0594	0.0301
s(Time,Speaker)	131.4188	199.0000	7.7253	< 0.0001
s(Time,Speaker):pos.ordlist-final	105.9788	238.0000	2.0450	0.5956
s(Time,Speaker):pos.ordIP-final	112.8047	238.0000	258.9170	< 0.0001
s(Block):Positionword-final	1.7352	1.9298	1.2991	0.2347
s(Block):Positionlist-final	1.0004	1.0007	1.8401	0.1748
s(Block):PositionIP-final	1.0004	1.0008	1.2116	0.2709

Table B.9: Summary of F2 model of English /ai/.

A. parametric coefficients	Estimate	Std. Error	t-value	p-value
(Intercept)	0.6663	0.0231	28.8024	< 0.0001
pos.ord.L	0.0049	0.0245	0.1993	0.8420
pos.ord.Q	0.0008	0.0279	0.0276	0.9780
B. smooth terms	edf	Ref.df	F-value	p-value
s(Time)	11.8654	13.0266	30.4550	< 0.0001
s(Time):pos.ordlist-final	11.7459	13.0852	11.3733	< 0.0001
s(Time):pos.ordIP-final	12.6592	13.8498	14.4793	< 0.0001
s(Time,Speaker)	130.3059	200.0000	5.6988	< 0.0001
s(Time,Speaker):pos.ordlist-final	156.9216	238.0000	750.6249	< 0.0001
s(Time,Speaker):pos.ordIP-final	167.3166	238.0000	554.3243	< 0.0001
s(Block):Positionword-final	1.0001	1.0003	3.0730	0.0796
s(Block):Positionlist-final	1.4911	1.7410	0.3607	0.5855
s(Block):PositionIP-final	1.2406	1.4232	0.1548	0.7008

Table B.10: Summary of F1 model of English /au/.

A. parametric coefficients	Estimate	Std. Error	t-value	p-value
(Intercept)	1.2708	0.0358	35.4657	< 0.0001
pos.ord.L	-0.0533	0.0143	-3.7373	0.0002
pos.ord.Q	0.0300	0.0160	1.8752	0.0608
B. smooth terms	edf	Ref.df	F-value	p-value
s(Time)	12.9027	14.4654	30.2395	< 0.0001
s(Time):pos.ordlist-final	11.3896	14.0937	9.1166	< 0.0001
s(Time):pos.ordIP-final	7.0176	8.6323	24.6422	< 0.0001
s(Time,Speaker)	109.6401	200.0000	8.2727	< 0.0001
s(Time,Speaker):pos.ordlist-final	70.0317	238.0000	36.9452	0.0001
s(Time,Speaker):pos.ordIP-final	85.3632	238.0000	8.9947	0.1110
s(Block):Positionword-final	1.0001	1.0002	3.4253	0.0642
s(Block):Positionlist-final	1.0003	1.0005	0.2195	0.6397
s(Block):PositionIP-final	1.9248	1.9943	20.0086	< 0.0001

Table B.11: Summary of F2 model of English /au/.

A. parametric coefficients	Estimate	Std. Error	t-value	p-value
(Intercept)	0.4693	0.0253	18.5636	< 0.0001
pos.ord.L	0.0212	0.0369	0.5745	0.5656
pos.ord.Q	0.0014	0.0378	0.0360	0.9713
B. smooth terms	edf	Ref.df	F-value	p-value
s(Time)	8.9775	10.1797	27.6612	< 0.0001
s(Time):pos.ordlist-final	6.3827	6.8448	3.8631	0.0004
s(Time):pos.ordIP-final	13.4664	13.8428	9.9670	< 0.0001
s(Time,Speaker)	115.7900	199.0000	4.9176	< 0.0001
s(Time,Speaker):pos.ordlist-final	188.1570	238.0000	22.0987	0.3412
s(Time,Speaker):pos.ordIP-final	207.1678	238.0000	4847.7147	< 0.0001
s(Block):Positionword-final	1.0016	1.0032	2.6428	0.1042
s(Block):Positionlist-final	1.9450	1.9969	9.8429	< 0.0001
s(Block):PositionIP-final	1.7649	1.9447	2.4839	0.0582

Table B.12: Summary of F1 model of English /ou/.

A. parametric coefficients	Estimate	Std. Error	t-value	p-value
(Intercept)	1.1450	0.0434	26.3571	< 0.0001
pos.ord.L	-0.0370	0.0231	-1.6043	0.1087
pos.ord.Q	0.0167	0.0368	0.4531	0.6505
B. smooth terms	edf	Ref.df	F-value	p-value
s(Time)	14.0762	15.8242	59.8960	< 0.0001
s(Time):pos.ordlist-final	1.0006	1.0008	26.1446	< 0.0001
s(Time):pos.ordIP-final	7.4979	9.4411	23.9567	< 0.0001
s(Time,Speaker)	92.1740	199.0000	4.7900	< 0.0001
s(Time,Speaker):pos.ordlist-final	143.3886	238.0000	816.2243	< 0.0001
s(Time,Speaker):pos.ordIP-final	69.0724	238.0000	1.8833	0.3802
s(Block):Positionword-final	1.8671	1.9823	11.5492	< 0.0001
s(Block):Positionlist-final	1.0002	1.0004	0.0005	0.9863
s(Block):PositionIP-final	1.0003	1.0006	12.9072	0.0003

Table B.13: Summary of F2 model of English /ou/.

A. parametric coefficients	Estimate	Std. Error	t-value	p-value
(Intercept)	0.5298	0.0154	34.4356	< 0.0001
pos.ord.L	-0.0148	0.0145	-1.0255	0.3052
pos.ord.Q	-0.0048	0.0153	-0.3137	0.7537
B. smooth terms	edf	Ref.df	F-value	p-value
s(Time)	10.5817	11.4123	28.9569	< 0.0001
s(Time):pos.ordlist-final	7.8372	10.0361	7.7699	< 0.0001
s(Time):pos.ordIP-final	6.0112	7.7324	7.5192	< 0.0001
s(Time,Speaker)	67.0356	99.0000	4.4743	< 0.0001
s(Time,Speaker):pos.ordlist-final	28.4220	118.0000	61.0826	0.0018
s(Time,Speaker):pos.ordIP-final	28.8095	118.0000	296.4288	< 0.0001
s(Block):Positionword-final	1.0003	1.0006	0.0178	0.8945
s(Block):Positionlist-final	1.7312	1.9277	1.9308	0.1945
s(Block):PositionIP-final	1.5310	1.7795	0.5951	0.5822

Table B.14: Summary of F1 model of Japanese /ai/.

A. parametric coefficients	Estimate	Std. Error	t-value	p-value
(Intercept)	1.6691	0.0238	69.9998	< 0.0001
pos.ord.L	0.0152	0.0229	0.6638	0.5069
pos.ord.Q	-0.0543	0.0287	-1.8916	0.0586
B. smooth terms	edf	Ref.df	F-value	p-value
s(Time)	12.3095	14.0246	18.3923	< 0.0001
s(Time):pos.ordlist-final	8.9619	11.7415	18.1880	< 0.0001
s(Time):pos.ordIP-final	8.1133	10.6009	7.0617	< 0.0001
s(Time,Speaker)	47.8481	100.0000	1.5938	< 0.0001
s(Time,Speaker):pos.ordlist-final	15.3428	118.0000	1.3010	0.1919
s(Time,Speaker):pos.ordIP-final	18.0804	118.0000	0.4720	0.5805
s(Block):Positionword-final	1.0001	1.0002	3.8076	0.0511
s(Block):Positionlist-final	1.7272	2.1276	0.5674	0.5217
s(Block):PositionIP-final	1.0001	1.0003	0.1209	0.7271

Table B.15: Summary of F2 model of Japanese /ai/.

A. parametric coefficients	Estimate	Std. Error	t-value	p-value
(Intercept)	0.5394	0.0067	80.4960	< 0.0001
pos.ord.L	-0.0104	0.0044	-2.3769	0.0175
pos.ord.Q	0.0250	0.0090	2.7711	0.0056
B. smooth terms	edf	Ref.df	F-value	p-value
s(Time)	11.1048	12.2264	44.4242	< 0.0001
s(Time):pos.ordlist-final	10.1058	12.0441	11.9057	< 0.0001
s(Time):pos.ordIP-final	8.5336	11.1809	6.3833	< 0.0001
s(Time,Speaker)	59.4645	99.0000	2.5529	< 0.0001
s(Time,Speaker):pos.ordlist-final	39.7154	99.0000	1.9121	0.7001
s(Time,Speaker):pos.ordIP-final	5.1661	99.0000	5.1010	0.1230
s(Block):Positionword-final	1.0005	1.0009	0.0108	0.9198
s(Block):Positionlist-final	1.0001	1.0003	0.0918	0.7621
s(Block):PositionIP-final	1.9397	2.3802	1.1317	0.3024

Table B.16: Summary of F1 model of Japanese /au/.

A. parametric coefficients	Estimate	Std. Error	t-value	p-value
(Intercept)	1.1606	0.0490	23.6931	< 0.0001
pos.ord.L	0.0162	0.0100	1.6268	0.1038
pos.ord.Q	0.0274	0.0170	1.6132	0.1067
B. smooth terms	edf	Ref.df	F-value	p-value
s(Time)	2.9758	3.4608	41.5072	< 0.0001
s(Time):pos.ordlist-final	8.9105	11.4808	7.0992	< 0.0001
s(Time):pos.ordIP-final	13.3389	16.3932	43.6769	< 0.0001
s(Time,Speaker)	61.7422	99.0000	5.4987	< 0.0001
s(Time,Speaker):pos.ordlist-final	16.7841	118.0000	17.1952	0.0344
s(Time,Speaker):pos.ordIP-final	28.0104	118.0000	0.5461	0.6460
s(Block):Positionword-final	1.0001	1.0002	0.0837	0.7726
s(Block):Positionlist-final	1.0006	1.0012	0.0104	0.9209
s(Block):PositionIP-final	1.0002	1.0003	1.3577	0.2440

Table B.17: Summary of F2 model of Japanese /au/.

A. parametric coefficients	Estimate	Std. Error	t-value	p-value
(Intercept)	0.5635	0.0176	32.0511	< 0.0001
pos.ord.L	-0.0067	0.0116	-0.5749	0.5654
pos.ord.Q	-0.0170	0.0121	-1.4010	0.1613
B. smooth terms	edf	Ref.df	F-value	p-value
s(Time)	11.3989	12.1491	43.7043	< 0.0001
s(Time):pos.ordlist-final	10.7425	13.3131	17.2777	< 0.0001
s(Time):pos.ordIP-final	2.4051	2.8548	1.9382	0.1348
s(Time,Speaker)	68.1816	100.0000	3.9370	< 0.0001
s(Time,Speaker):pos.ordlist-final	34.5657	118.0000	157.0557	< 0.0001
s(Time,Speaker):pos.ordIP-final	56.5415	118.0000	234.0140	< 0.0001
s(Block):Positionword-final	1.0001	1.0002	4.9689	0.0259
s(Block):Positionlist-final	1.5508	1.7979	1.0532	0.4407
s(Block):PositionIP-final	1.8943	1.9860	9.1710	0.0003

Table B.18: Summary of F1 model of Japanese /ae/.

A. parametric coefficients	Estimate	Std. Error	t-value	p-value
(Intercept)	0.5635	0.0176	32.0511	< 0.0001
pos.ord.L	-0.0067	0.0116	-0.5749	0.5654
pos.ord.Q	-0.0170	0.0121	-1.4010	0.1613
B. smooth terms	edf	Ref.df	F-value	p-value
s(Time)	11.3989	12.1491	43.7043	< 0.0001
s(Time):pos.ordlist-final	10.7425	13.3131	17.2777	< 0.0001
s(Time):pos.ordIP-final	2.4051	2.8548	1.9382	0.1348
s(Time,Speaker)	68.1816	100.0000	3.9370	< 0.0001
s(Time,Speaker):pos.ordlist-final	34.5657	118.0000	157.0557	< 0.0001
s(Time,Speaker):pos.ordIP-final	56.5415	118.0000	234.0140	< 0.0001
s(Block):Positionword-final	1.0001	1.0002	4.9689	0.0259
s(Block):Positionlist-final	1.5508	1.7979	1.0532	0.4407
s(Block):PositionIP-final	1.8943	1.9860	9.1710	0.0003

Table B.19: Summary of F1 model of Japanese /ae/.

A. parametric coefficients	Estimate	Std. Error	t-value	p-value
(Intercept)	1.5771	0.0563	28.0041	< 0.0001
pos.ord.L	0.0161	0.0149	1.0798	0.2803
pos.ord.Q	-0.0646	0.0174	-3.7050	0.0002
B. smooth terms	edf	Ref.df	F-value	p-value
s(Time)	10.2176	11.4362	14.3117	< 0.0001
s(Time):pos.ordlist-final	2.9788	3.4928	17.3877	< 0.0001
s(Time):pos.ordIP-final	5.2985	7.1360	6.9708	< 0.0001
s(Time,Speaker)	56.8316	100.0000	3.9149	< 0.0001
s(Time,Speaker):pos.ordlist-final	61.1225	118.0000	17.9908	0.0013
s(Time,Speaker):pos.ordIP-final	2.5120	118.0000	16.5581	0.0003
s(Block):Positionword-final	1.7285	2.1340	2.5520	0.0765
s(Block):Positionlist-final	1.0002	1.0004	0.0564	0.8127
s(Block):PositionIP-final	1.0001	1.0001	0.0876	0.7673

Table B.20: Summary of F2 model of Japanese /ae/.

Reference

Aguilar, L. (1999). Hiatus and diphthong: Acoustic cues and speech situation differences. *Speech Communication*, *28*(1), 57–74 (Page 9).

Akpanglo-Nartey, R. (2020). A study of format dynamics in ghanaian english diphthongs rebecca. *Journal of World Englishes and Educational Practices*, *2*(3), 1–21 (Page 35).

Beckman, M. E. (1986, January 31). *Stress and non-stress accent* [Publication Title: Stress and Non-Stress Accent]. DE GRUYTER. (Page 4).

Beckman, M. E. (1996). The parsing of prosody. *Language and Cognitive Processes*, *11*(1), 17–67 (Page 12).

Beckman, M. E. (1997). A typology of spontaneous speech. In Y. Sagisaka, N. Campbell, & N. Higuchi (Eds.), *Computing prosody: Computational models for processing spontaneous speech* (pp. 7–26). Springer US. (Page 12).

Beckman, M. E., & Edwards, J. (1990). Lengthenings and shortenings and the nature of prosodic constituency. *Between the grammar and physics of speech: Papers in laboratory phonology i* (pp. 152–178). (Pages 13, 146).

Beckman, M. E., & Edwards, J. (1992). Intonational categories and the articulatory control of duration [Publisher: IOS Press]. *Speech perception, production and linguistic structure*, 359 (Pages 10, 21, 118, 146).

Beckman, M. E., & Edwards, J. (1994). Articulatory evidence for differentiating stress categories. In P. A. Keating (Ed.), *Phonological structure and phonetic form: Papers in laboratory phonology III* (pp. 7–33). Cambridge University Press. (Pages 15, 21).

Berkovits, R. (1993). Progressive utterance-final lengthening in syllables with final fricatives. *Language and Speech*, *36*(1), 89–98 (Pages 13, 15, 142).

Berkovits, R. (1994). Durational effects in final lengthening, gapping, and contrastive stress [Publisher: SAGE PublicationsSage UK: London, England]. *Language and Speech*, *37*(3), 237–250 (Pages 10, 14, 15, 142, 147).

Bolinger, D. (1972). Accent is predictable (if you're a mind-reader). *Language*, *48*(3), 633 (Page 4).

Bolinger, D. L. (1958). A theory of pitch accent in english [Publisher: Routledge]. *Word*, *14*(2), 109–149 (Page 4).

Bombien, L., Mooshammer, C., & Hoole, P. (2013). Articulatory coordination in word-initial clusters of german [Publisher: Academic Press]. *Journal of Phonetics*, *41*(6), 546–561 (Pages 103, 146).

Brandt, E., & Simpson, A. P. (2021). The production of ejectives in german and georgian [Publisher: The Authors]. *Journal of Phonetics*, *89*, 101111 (Page 38).

Browman, C. P., & Goldstein, L. (1989). Articulatory gestures as phonological units. *Phonology*, *6*(2), 201–251 (Page 11).

Browman, C. P., & Goldstein, L. (1995). Dynamics and articulatory phonolgy. In R. F. Port & T. V. Gelder (Eds.), *Mind as motion: Explorations in the dynamics of cognition* (pp. 175–194). Bradford Books. (Page 11).

Browman, C. P., & Goldstein, L. M. (1986). Towards an articulatory phonology. *Phonology Yearbook*, *3*, 219–252 (Page 11).

Byrd, D. (2000). Articulatory vowel lengthening and coordination at phrasal junctures. *Phonetica*, *57*(1), 3–16 (Pages 1, 13, 27, 118, 149).

Byrd, D., & Choi, S. (2010). At the juncture of prosody , phonology , and phonetics — the interaction of phrasal and syllable structure in shaping the timing of consonant gestures [ISBN: 9783110224900]. *Laboratory Phonology*, *10*, 31–60 (Page 10).

Byrd, D., & Krivokapić, J. (2021, January 14). Cracking prosody in articulatory phonology [ISSN: 2333-9683]. In S. Shattuck-Hufnagel & J. Barnes (Eds.), *Annual review of linguistics* (pp. 31–53). MIT Press. (Pages 11, 12, 148, 153).

Byrd, D., Krivokapić, J., & Lee, S. (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 (Pages 15, 16).

Byrd, D., & Riggs, D. (2008). Locality interactions with prominence in determining the scope of phrasal lengthening. *Journal of the International Phonetic Association*, *38*(2), 187–202 (Page 13).

Byrd, D., & Saltzman, E. (1998). Intragestural dynamics of multiple prosodic boundaries [ISBN: 0147-5185 (Print)\r0147-5185 (Linking)]. *Journal of Phonetics*, *26*(2), 173–199 (Pages 1, 10, 15, 16, 27, 83, 118, 145, 146).

Byrd, D., & Saltzman, E. (2003). The elastic phrase: Modeling the dynamics of boundary-adjacent lengthening. *Journal of Phonetics*, *31*(2), 149–180 (Pages 10, 11, 15, 21, 83, 145, 146).

Cambier-Langeveld, T. (1997). The domain of final lengthening in the production of dutch. *Linguistics in the Netherlands*, *14*, 13–24 (Page 13).

Cambier-Langeveld, T. (1999). Temporal marking of accent and boundaries. *Linguistics in the Netherlands*, *16*(1), 13–25 (Pages 13, 14).

Campbell, N. (1992). Segmental elasticity and timing in japanese speech. *Speech perception, production and linguistic structure* (pp. 403–418). (Pages 72, 143).

Campbell, N. (1999). A study of japanese speech timing from the syllable perspective. *Journal of the Phonetic Society of Japan*, *3*(2), 29–39 (Page 143).

Cao, J. (2004). Restudy of segmental lengthening in mandarin chinese. *Speech Prosody*, 39–44 (Page 13).

Chen, W.-R., Whalen, D. H., & Shadle, C. H. (2019). F0-induced formant measurement errors result in biased variabilities. *The Journal of the Acoustical Society of America*, *145*(5), EL360–EL366 (Page 31).

Chen, Y., & Gussenhoven, C. (2015). Shanghai chinese [Publisher: Cambridge University Press]. *Journal of the International Phonetic Association*, *45*(3), 321–337 (Page 28).

Childers, D. G. (1978). *Modern spectrum analysis*. IEEE Computer Society Press. (Page 31).

Chitoran, I. (2002). A perception-production study of romanian diphthongs and glide-vowel sequences. *Journal of the International Phonetic Association*, *32*(2), 203–222 (Page 9).

Chitoran, I., & Hualde, J. I. (2007). From hiatus to diphthong: The evolution of vowel sequences in romance. *Phonology*, *24*(1), 37–75 (Page 9).

Cho, T. (2001). *Effects of prosody on articualtion in english* (Doctoral dissertation) [ISBN: 061239378X]. University of California, Los Angeles. (Pages 146, 154, 155).

Cho, T. (2002). *The effects of prosody on articulation in english*. Routledge. (Pages 19, 21, 27).

Cho, T. (2004). Prosodically conditioned strengthening and vowel-to-vowel coarticulation in english. *Journal of Phonetics*, *32*(2), 141–176 (Page 103).

Cho, T. (2005). Prosodic strengthening and featural enhancement: Evidence from acoustic and articulatory realizations of /a,i/ in english. *The Journal of the Acoustical Society of America*, *117*(6), 3867–78 (Page 17).

Cho, T. (2006). Manifestation of prosodic structure in articulation: Evidence from lip movement kinematics in english. *Laboratory phonology*, *8*, 519–548 (Pages 16, 19).

Cho, T. (2015). Language effects on timing at the segmental and suprasegmental levels [Publisher: Wiley Online Library]. *The handbook of speech production*, *505*, 529 (Pages 13, 143).

Cho, T. (2016). Prosodic boundary strengthening in the phonetics-prosody interface. *Language and Linguistics Compass*, *10*(3), 120–141 (Pages 4, 6, 12, 13, 146).

Cho, T., & Keating, P. A. (2001). Articulatory and acoustic studies on domain-initial strengthening in korean. *Journal of Phonetics*, *29*(2), 155–190 (Page 146).

Cho, T., Kim, J., & Kim, S. (2013). Preboundary lengthening and preaccentual shortening across syllables in a trisyllabic word in english [Publisher: Acoustical Society of America (ASA)]. *The Journal of the Acoustical Society of America*, *133*(5), 384–390 (Pages 13, 14, 143).

Cho, T., & Ladefoged, P. (1999). Variation and universals in VOT: Evidence from 18 languages. *Journal of Phonetics*, *27*(2), 207–229 (Pages 103, 104).

Cutler, A., Dahan, D., & van Donselaar, W. (1997). Prosody in the coinprehensioit of spoken laiigimge: A literature review. *Language and Speech*, *40*(2), 141–201 (Page 4).

de Jong, K., Beckman, M. E., & Edwards, J. (1993). The interplay between prosodic structure and coarticulation. [ISBN: 0023-8309 (Print)\r0023-8309 (Linking)]. *Language and speech*, *36 (Pt 2-*, 197–212 (Page 16).

de Jong, K. (1998). Stress-related variation in the articulation of coda alveolar stops: Flapping revisited. *Journal of Phonetics*, *26*(3), 283–310 (Page 5).

Duanmu, S. (1996). Pre-juncture lengthening and foot binarity [Publisher: Department of Linguistics, University of Illinois.]. *Studies in the Linguistic Sciences*, *26*(1), 95–115 (Page 13).

Duanmu, S. (2007). *The phonology of standard chinese*. OUP Oxford. (Page 9).

Edwards, J., Beckman, M. E., & Fletcher, J. (1991). The articulatory kinematics of final lengthening. *Journal of the Acoustical Society of America*, *89*(1), 369–382 (Pages 1, 10, 13, 21, 118, 146).

Elvin, J., Williams, D., & Escudero, P. (2016). Dynamic acoustic properties of monophthongs and diphthongs in western sydney australian english. *The Journal of the Acoustical Society of America*, *140*(1), 576–581 (Page 9).

Emerich, G. H. (2012). *The vietnamese vowel system* (Doctoral dissertation) [Volume: 632]. University of Pennsylvania. (Page 9).

Flego, S., & Forrest, J. (2021). Leveraging the temporal dynamics of anticipatory vowel-to-vowel coarticulation in linguistic prediction: A statistical modeling approach [Publisher: Elsevier Ltd]. *Journal of Phonetics*, *88*, 101093 (Page 34).

Fletcher, J. (1987). Some micro and macro effects of tempo change on timing in french [Publisher: Walter de Gruyter, Berlin/New York Berlin, New York] (Page 13).

Fletcher, J. (2010). *The prosody of speech: Timing and rhythm* [Publication Title: The Handbook of Phonetic Sciences: Second Edition]. (Pages 13, 55, 146, 154).

Fougeron, C. (1999). Prosodically conditioned articulatory variations: A review. *UCLA Working Papers in Phonetics*, *97*, 1–73 (Page 146).

Fougeron, C., & Keating, P. A. (1997). Articulatory strengthening at edges of prosodic domains [ISBN: 0001-4966 (Print)]. *The Journal of the Acoustical Society of America*, *101*(6), 3728–3740 (Pages 15, 16, 146, 147).

Fowler, C. (1988). Periodic dwindling of acoustic and articulatory variables in speech production. *Paw Review*, *3*, 10–13 (Page 142).

Fox, R. A., & Jacewicz, E. (2009). Cross-dialectal variation in formant dynamics of american english vowels [Publisher: Acoustical Society of America (ASA)]. *The Journal of the Acoustical Society of America*, *126*(5), 2603–2618 (Page 35).

Fry, D. B. (1965). The dependence of stress judgments on vowel formant structure. *Phonetic Sciences. 5th International Congress, Münster, August 1964: Proceedings*, 306–311 (Page 2).

Gay, T. (1968). Effect of speaking rate on vowel formant movements. *The Journal of the Acoustical Society of America*, *44*(6), 1570–1573 (Pages 12, 92).

Georgeton, L., & Fougeron, C. (2014). Domain-initial strengthening on french vowels and phonological contrasts: Evidence from lip articulation and spectral variation [Publisher: Elsevier]. *Journal of Phonetics*, *44*(1), 83–95 (Page 146).

Georgeton, L., Antolík, T. K., & Fougeron, C. (2016). Effect of domain initial strengthening on vowel height and backness contrasts in french: Acoustic and ultrasound data. *Journal of Speech, Language, and Hearing Research*, *59*(6), 1–15 (Page 146).

Goldstein, L., & Fowler, C. A. (2003). Articulatory phonology: A phonology for public language use [ISBN: 9783110178722]. *Phonetics and Phonology in Language Comprehension and Production: Differences and Similarities*, 159–207 (Page 11).

Gubian, M., Torreira, F., & Boves, L. (2015). Using functional data analysis for investigating multidimensional dynamic phonetic contrasts [Publisher: Academic Press ISBN: 0095-4470]. *Journal of Phonetics*, *49*, 16–40 (Page 9).

Gussenhoven, C., & Rietveld, A. (1992). Intonation contours, prosodic structure and preboundary lengthening. *Journal of Phonetics*, *20*(3), 283–303 (Page 13).

Hara, I. (2015). *An acoustic analysis of vowel sequences in japanese* (Doctoral dissertation). Newcastle University. (Page 31).

Harrington, J., Fletcher, J., & Beckman, M. E. (2000). Manner and place conflicts in the articulation of accent in australian english. *Papers in laboratory phonology v: Language acquisition and the lexicon* (pp. 40–51). (Page 17).

Harrington, J., Kleber, F., Reubold, U., & Siddins, J. (2015). The relationship between prosodic weakening and sound change: Evidence from the german tense/lax vowel contrast. *Laboratory Phonology*, *6*(1), 87–117 (Page 16).

Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of american english vowels. *Journal of the Acoustical Society of America*, *97*(5), 3099–3111 (Pages 31, 35).

Hillenbrand, J. M., Clark, M. J., & Nearey, T. M. (2001). Effects of consonant environment on vowel formant patterns. *The Journal of the Acoustical Society of America*, *109*(2), 748–763 (Page 35).

Howie, J. M. (1976). *Acoustical studies of mandarin vowels and tones* (Vol. 18). Cambridge University Press. (Pages 31, 71).

Hsieh, F.-y. (2017). *A gestural approach to the phonological representation of english diphthongs* (Doctoral dissertation) [Issue: May]. University of Southern California. (Page 10).

Hu, F. (2013). Falling diphthongs have a dynamic target while rising diphthongs have two targets: Acoustics and articulation of the diphthong production in ningbo chinese [Publisher: Acoustical Society of America (ASA)]. *The Journal of the Acoustical Society of America*, *134*(5), 4199–4199 (Pages 10, 12).

Hualde, J. I., Barlaz, M., & Luchkina, T. (2021). Acoustic differentiation of allophones of /ai/ in chicagoland english: Statistical comparison of formant trajectories. *Journal of the International Phonetic Association* (Page 38).

Hualde, J. I., & Prieto, M. (2002). On the diphthong/hiatus contrast in spanish: Some experimental results. *Linguistics*, *40*(378), 217–234 (Page 9).

Hyman, L. M. (2006). Word-prosodic typology. *Phonology*, *23*(2), 225–257 (Page 25).

Hyman, L. M. (2009). How (not) to do phonological typology: The case of pitch-accent. *Language Sciences*, *31*(2), 213–238 (Page 25).

Jang, J., & Katsika, A. (2020). The amount and scope of phrase-final lengthening in korean. *Speech Prosody 2020*, 270–274 (Page 13).

Johnson, K. (2020). The f method of vocal tract length normalization for vowels. *Laboratory Phonology*, *11*(1), 1–16 (Page 31).

Johnson, K., & Martin, J. (2001). Acoustic vowel reduction in creek: Effects of distinctive length and position in the word. *Phonetica*, *58*(1), 81–102 (Pages 14, 17).

Jun, S.-A., & Fougeron, C. (2002). Realizations of accentual phrase in french intonation [Publisher: Walter de Gruyter GmbH & Co. KG Berlin, Germany] (Page 13).

Jun, S.-A. (2005). *Prosodic typology: The phonology of intonation and phrasing*. OUP Oxford. (Page 4).

Jun, S.-a. (Ed.). (2014). *Prosodic typology II: The phonology of intonation and phrasing*. Oxford University Press. (Page 4).

Kaiki, N., Takeda, N., & Sagisaka, Y. (1992). Linguistic properties in the control of segmental duration for speech synthesis. *Talking Machines: Theories, Models, and Designs*, 255–264 (Page 13).

Keating, P. A. (1984). Phonetic and phonological representation of stop consonant voicing. *Language*, *60*(2), 286 (Page 104).

Keating, P. A. (1990). The window model of coarticulation: Articulatory evidence BT - papers in laboratory phonology i between the grammar and physics of speech [ISBN: 9780511627736]. *Papers in Laboratory Phonology I between the grammar and physics of speech*, (26), 451–470 (Page 104).

Kim, B., Tiede, M. K., & Whalen, D. H. (2019). Evidence for pivots in tongue movement for diphthongs. *ICPhS*, 2159–2163 (Page 10).

Kim, J. (2020). *Individual differences in the production and perception of prosodic boundaries in american english* (Doctoral dissertation) [Publication Title: The Journal of the Acoustical Society of America Volume: 148 Issue: 4 ISSN: 0001-4966]. The University of Michigan. (Page 14).

Kim, S., Jang, J., & Cho, T. (2017). Articulatory characteristics of preboundary lengthening in interaction with prominence on tri-syllabic words in american english [ISBN: 978-3-319-08397-1 978-3-319-08398-8]. *The Journal of the Acoustical Society of America*, *142*(4), EL362–EL368 (Page 14).

Kingston, J., & Diehl, R. L. (1994). Phonetic knowledge [Publisher: Linguistic Society of America ISBN: 0097-8507]. *Language*, *70*(3), 419 (Pages 12, 103, 104).

Klatt, D. H. (1973). Interaction between two factors that influence vowel duration [Publisher: Acoustical Society of America]. *The Journal of the Acoustical Society of America*, *54*(4), 1102–1104 (Page 13).

Klatt, D. H. (1975). Vowel lengthening is syntactically determined in a connected discourse [Publisher: Elsevier Masson SAS]. *Journal of Phonetics*, *3*(3), 129–140 (Pages 1, 13, 15).

Kohler, K. (1983). Prosodic boundary signals in german [Publisher: S. Karger AG]. *Phonetica*, *40*(2), 89–134 (Pages 13, 14).

Krivokapić, J. (2007). *The planning, production, and perception of prosodic structure* (Doctoral dissertation). University of Southern California. (Pages 4, 10, 21, 27, 83, 102, 154).

Krivokapić, J. (2012). Prosodic planning in speech production. In S. Fuchs, M. Weihrich, D. Pape, & P. Perrier (Eds.), *Speech planning and dynamics* (pp. 157–190). Peter Lang. (Page 4).

Krivokapić, J. (2020). Prosody in articulatory phonology. In S. Shattuck-Hufnagel & J. Barnes (Eds.), *Prosodic theory and practice*. MIT press. (Page 21).

Krivokapić, J., & Byrd, D. (2012). Prosodic boundary strength: An articulatory and perceptual study [Publisher: Academic Press]. *Journal of Phonetics*, *40*(3), 430–442 (Page 154).

Kuzla, C., Cho, T., & Ernestus, M. (2007). Prosodic strengthening of german fricatives in duration and assimilatory devoicing. *Journal of Phonetics*, *35*(3), 301–320 (Pages 13, 14).

Labov, W., Ash, S., & Boberg, C. (2008, July 14). *The atlas of north american english: Phonetics, phonology and sound change* [Google-Books-ID: tsPvynQtlMoC]. Walter de Gruyter. (Page 28).

Labrune, L. (2012). *The phonology of japanese* (Vol. 15) [Publication Title: The Phonology of Japanese Issue: Hoequist 1982 ISSN: 18255167]. Oxford University Press. (Page 9).

Ladd, D. R. (2008). *Intonational phonology*. Cambridge University Press. (Pages 2, 7).

Ladd, D. R., & Campbell, W. N. (1991). Theories of prosodic structure: Evidence from syllable duration. *Proceeding of the 12th International Congress of Phonetic Sciences*, 290–293 (Page 15).

Ladd, D. (1986). Intonational phrasing: The case for recursive prosodic structure. *Phonology Yearbook*, *3*, 311–340 (Page 4).

Lambrecht, K. (1994). *Information structure and sentence form*. Cambridge University Press. (Page 4).

Lee, W.-S., & Zee, E. (2003). Standard chinese (beijing) [Publisher: Cambridge University Press]. *Journal of the International Phonetic Association*, *33*(1), 109–112 (Page 9).

Lehiste, I. (1976). Isochrony reconsidered [ISBN: 1009982220290]. *Journal of Phonetics*, *10*(5), 253–263 (Page 13).

Lehiste, I. (1970). *Suprasegmentals*. MIT Press. (Page 2).

Lehiste, I. (1973). Rhythmic units and syntactic units in production and perception [Publisher: Acoustical Society of America (ASA)]. *The Journal of the Acoustical Society of America*, *54*(5), 1228–1234 (Page 13).

Li, Y. (2015). Prosodic boundaries effect on segment articulation in stardard chinese: An articulatory and acoustic study. *Journal of Chinese Linguistics*, *43*(1), 364–398 (Page 13).

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the h&h theory. *Speech production and speech modelling* (pp. 403–439). Springer Netherlands. (Page 17).

Lindblom, B. (1983). Economy of speech gestures. *The production of speech* (pp. 217–245). Springer. (Page 2).

Lindblom, B. (1968). Temporal organization of syllable production. *Speech Transmission Laboratory Quarterly Progress Status*, *9*(2), 1–5 (Pages 2, 15, 142, 154).

Lindblom, B. E. F. (1975). Some temporal regularities of spoken swedish. *Auditory analysis and perception of speech*. Stockholm. (Pages 15, 147).

Liu, Y., & Li, A. (2003). Cues of prosodic boundaries in chinese spontaneous speech [ISBN: 1876346485]. *International Conference of the Phonetic Sciences 15*, 1269–1272 (Page 13).

Marin, S. (2007). *Vowel to vowel coordination, diphthongs in articulatory phonology* (Doctoral dissertation). Yale University. (Pages 9, 10).

Mayr, R., & Davies, H. (2011). A cross-dialectal acoustic study of the monophthongs and diphthongs of welsh. *Journal of the International Phonetic Association*, *41*(1), 1–25 (Page 9).

Mo, Y., Cole, J., & Hasegawa-Johnson, M. (2009). Prosodic effects on vowel production: Evidence from formant structure. *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, 2535–2538 (Page 17).

Mücke, D., & Grice, M. (2014). The effect of focus marking on supralaryngeal articulation – is it mediated by accentuation? [Publisher: Elsevier ISBN: 0095-4470]. *Journal of Phonetics*, *44*(1), 47–61 (Pages 19, 110).

Nakai, S., Kunnari, S., Turk, A., Suomi, K., & Ylitalo, R. (2009). Utterance-final lengthening and quantity in northern finnish [ISBN: 0095-4470]. *Journal of Phonetics*, *37*(1), 29–45 (Pages 14, 143).

Nakai, S., Turk, A. E., Suomi, K., Granlund, S., Ylitalo, R., & Kunnari, S. (2012). Quantity constraints on the temporal implementation of phrasal prosody in northern finnish [Publisher: Elsevier]. *Journal of Phonetics*, *40*(6), 796–807 (Pages 14, 143).

Nespor, M., & Vogel, I. (1986). *Prosodic phonology*. Foris Publications, Dordrecht, Holland. (Page 3).

Nord, L. (1986). Quarterly progress and status report acoustic studies of vowel reduction in swedish. *Dept. for Speech, Music and Hearing Quaterly Progress and Status Rreport*, *27*(4), 19–36 (Page 18).

Oller, D. K. (1973). The effect of position in utterance on speech segment duration in english. *Journal of the Acoustical Society of America*, *54*(5), 1235–1247 (Pages 1, 13).

Perperoglou, A., Sauerbrei, W., Abrahamowicz, M., & Schmid, M. (2019). A review of spline function procedures in r [Publisher: BioMed Central Ltd.]. *BMC Medical Research Methodology*, *19*(1), 1–16 (Page 33).

Pierrehumbert, J. (1980). *The phonology and phonetics of english intonation* (Doctoral dissertation). MIT. (Pages 2, 7).

Prom-On, S., Birkholz, P., & Xu, Y. (2014). Identifying underlying articulatory targets of thai vowels from acoustic data based on an analysis-by-synthesis approach. *Eurasip Journal on Audio, Speech, and Music Processing*, *2014* (Page 20).

Saltzman, E. L. (1991). The task dynamic coordination in speech production. *Speech Motor Control and Stuttering*, 37–52 (Page 10).

Saltzman, E., & Byrd, D. (2000). Task-dynamics of gestural timing: Phase windows and multifrequency rhythms. *Human Movement Science*, *19*(4), 499–526 (Page 10).

Saltzman, E. L., & Munhall, K. G. (1989). A dynamical approach to gestural patterning in speech production [Publisher: Lawrence Erlbaum Associates, Inc. ISBN: 1040-7413]. *Ecological Psychology*, *1*(4), 333–382 (Pages 10, 11).

Selkirk, E. O. (1986). *Phonology and syntax: The relationship between sound and structure*. MIT press. (Page 3).

Selkirk, E. (1995). Sentence prosody: Intonation, stress, and phrasing. *The handbook of phonological theory* (pp. 550–569). (Page 5).

Seo, J., Kim, S., Kubozono, H., & Cho, T. (2019). 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 (Pages 13, 14, 16, 143).

Shattuck-Hufnagel, S., & Turk, A. (1998). The domain of phrase-final lengthening in english. *The Journal of the Acoustical Society of America*, *103*(5), 2889–2889 (Pages 13, 143).

Shattuck-Hufnagel, S., & Turk, A. E. (1996). A prosody tutorial for investigators of auditory sentence processing [Publisher: Kluwer Academic/Plenum Publishers]. *Journal of Psycholinguistic Research*, *25*(2), 193–247 (Page 3).

Shepherd, M. A. (2008). The scope and effects of preboundary prosodic lengthening in japanese. *USC Working Papers in Linguistics*, *4*, 1–14 (Pages 16, 72, 144, 152).

Sóskuthy, M. (2021). Evaluating generalised additive mixed modelling strategies for dynamic speech analysis. *Journal of Phonetics*, *84* (Pages 41–44, 48).

Steffman, J. (2019a). Intonational structure mediates speech rate normalization in the perception of segmental categories [Publisher: Academic Press]. *Journal of Phonetics*, *74*, 114–129 (Page 154).

Steffman, J. (2019b). Phrase-final lengthening modulates listeners' perception of vowel duration as a cue to coda stop voicing [Publisher: Acoustical Society of America]. *The Journal of the Acoustical Society of America*, *145*(6), EL560–EL566 (Page 154).

Steffman, J. (2020). *Prominence in vowel perception and spoken language processing* (Doctoral dissertation) [ISBN: 9789896540821 ISSN: 15209202]. University of California, Los Angeles. (Page 14).

Sugahara, M. (2005). Post-focus prosodic phrase boundaries in tokyo japanese: Asymmetric behavior of an f0 cue and domain-final lengthening. *Studia Linguistica*, *59*(2), 144–173 (Page 13).

Sun, Y., & Shih, C. (2021). Boundary-conditioned anticipatory tonal coarticulation in standard mandarin [Publisher: Elsevier Ltd]. *Journal of Phonetics*, *84* (Page 38).

Tabain, M. (2003). Effects of prosodic boundary on /aC/ sequences: Articulatory results [Publisher: Acoustical Society of America (ASA)]. *The Journal of the Acoustical Society of America*, *113*(5), 2834–2849 (Pages 13, 15, 27, 142, 147).

Tabain, M., & Perrier, P. (2005). Articulation and acoustics of /i/ in preboundary position in french [Publisher: Academic Press]. *Journal of Phonetics*, *33*(1), 77–100 (Pages 15, 16, 27).

Tabain, M., & Perrier, P. (2007). An articulatory and acoustic study of /u/ in preboundary position in french: The interaction of compensatory articulation, neutralization avoidance and featural enhancement [Publisher: Academic Press]. *Journal of Phonetics*, *35*(2), 135–161 (Pages 15, 17).

Team, R. C. (2022). R: A language and environment for statistical computing [Publisher: Vienna, Austria] (Page 38).

Tomaschek, F., Tucker, B. V., Fasiolo, M., & Baayen, R. H. (2018). Practice makes perfect: The consequences of lexical proficiency for articulation. *Linguistics Vanguard*, *4* (Page 38).

Tseng, S.-C. (2014). Chinese disyllabic words in conversation [Publisher: John Benjamins]. *Chinese Language and Discourse*, *5*(2), 231–251 (Page 13).

Turk, A. E., & Shattuck-Hufnagel, S. (2000). Word-boundary-related duration patterns in english. *Journal of Phonetics*, *28*(4), 397–440 (Page 15).

Turk, A. E., & Shattuck-Hufnagel, S. (2007). Multiple targets of phrase-final lengthening in american english words [ISBN: 0095-4470]. *Journal of Phonetics*, *35*(4), 445–472 (Pages 13–15, 143).

Umeda, N. (1975). Vowel duration in american english [Publisher: Acoustical Society of America]. *The Journal of the Acoustical Society of America*, *58*(2), 434–445 (Page 15).

Van Der Harst, S., Van De Velde, H., & Van Hout, R. (2014). Variation in standard dutch vowels: The impact of formant measurement methods on identifying the speaker's regional origin. *Language Variation and Change*, *26*(2), 247–272 (Page 34).

Van Rij, J. (2016). *Testing for significance*. Retrieved March 5, 2022, from https://cran.r-project.org/web/packages/itsadug/vignettes/test.html. (Page 47)

van Santen, J. P. H., & Shih, C. (2000). Suprasegmental and segmental timing models in mandarin chinese and american english [Publisher: Acoustical Society of America (ASA)]. *The Journal of the Acoustical Society of America*, *107*(2), 1012–1026 (Page 2).

Vayra, M., & Fowler, C. A. (1992). Declination of supralaryngeal gestures in spoken italian [Publisher: S. Karger AG]. *Phonetica*, *49*(1), 48–60 (Page 15).

Wells, J. C., & Wells, J. C. (1982). *Accents of english: Volume 1* (Vol. 1). Cambridge University Press. (Page 9).

Whalen, D. H., & Xu, Y. (1992). Information for mandarin tones in the amplitude contour and in brief segments [Publisher: Karger Publishers]. *Phonetica*, *49*(1), 25–47 (Page 71).

White, L. (2014). Communicative function and prosodic form in speech timing. *Speech Communication*, *63-64*, 38–54 (Page 14).

White, L., Benavides-Varela, S., & Mády, K. (2020). Are initial-consonant lengthening and final-vowel lengthening both universal word segmentation cues? [Publisher: Elsevier Ltd]. *Journal of Phonetics*, *81*, 100982 (Pages 14, 154).

Wieling, M. (2018). Analyzing dynamic phonetic data using generalized additive mixed modeling: A tutorial focusing on articulatory differences between l1 and l2 speakers of english [Publisher: Academic Press]. *Journal of Phonetics*, *70*, 86–116 (Pages 41, 43, 47).

Wightman, C. W., Shattuck-Hufnagel, S., Ostendorf, M., & Price, P. J. (1992). Segmental durations in the vicinity of prosodic phrase boundaries [Publisher: Acoustical Society of America]. *The Journal of the Acoustical Society of America*, *91*(3), 1707–1717 (Pages 1, 15).

Wood, S., & Wood, M. S. (2015). Package 'mgcv'. *R package version*, *1*(29), 729 (Page 38).

Xia, L., & Hu, F. (2016). Vowels and diphthongs in the taiyuan jin chinese dialect. *Proceedings of the Annual Conference of the International Speech Communication Association*, 993–997 (Page 9).

Xu, Y., Lee, A., Prom-On, S., & Liu, F. (2016). Explaining the PENTA model: A reply to arvaniti and ladd. *Phonology*, *32*(3), 505–535 (Page 20).

Xu, Y., & Prom-on, S. (2019). Economy of effort or maximum rate of information? exploring basic principles of articulatory dynamics. *Frontiers in Psychology*, *10*, 1–22 (Pages 20, 145).

Yang, J., Zhang, Y., Li, A., & Xu, L. (2017). On the duration of mandarin tones. *Interspeech 2017*, 1407–1411 (Page 71).

Yang, L. (2011). *Stress patterns of dissyllabic wrods in beijing dialect* (Doctoral dissertation). National University of Singapore. (Page 25).

Yang, Y., & Wang, B. (2002). Acoustic correlates of hierarchical prosodic boundaries in mandarin. *Speech Prosody 2002* (Page 13).