USC-SIPI Report #444

Prosodic Substrate of Consonant and Tone Dynamics

By

Yoonjeong Lee

May 2018

Signal and Image Processing Institute
UNIVERSITY OF SOUTHERN CALIFORNIA
USC Viterbi School of Engineering
Department of Electrical Engineering-Systems 3740
McClintock Avenue, Suite 400
Los Angeles, CA 90089-2564 U.S.A.

THE PROSODIC SUBSTRATE OF CONSONANT AND TONE DYNAMICS

by

Yoonjeong Lee

A Dissertation Presented to the
FACULTY OF THE GRADUATE SCHOOL
UNIVERSITY OF SOUTHERN CALIFORNIA
In Partial Fulfillment of the
Requirements for the Degree
DOCTOR OF PHILOSOPHY
(LINGUISTICS)

May 2018

Dedication

This dissertation is dedicated to the memory of my beloved brother, *Little Mouse*.

Acknowledgments

This dissertation would not have been possible without the guidance and support from many people.

First, I extend my deepest gratitude to my co-chairs Louis Goldstein and Dani Byrd for their mentorship and advice in guiding me through this process. They have invaluably and tirelessly contributed to my growth as a scientist. Their insightful comments and encouragement have incented me to sharpen my linguistic perspective and improved this dissertation. I have always admired the breadth and depth of their knowledge, and they will be forever my role models. I am deeply indebted to my advisor Louis who has been always inspiring and a great friend for all these years. I am especially grateful to Dani for being such an enthusiastic mentor and having unending confidence in me. I am also very thankful to Louis and Dani for giving me so many invaluable opportunities to lead research projects. I am very humbled to have had such brilliant and supportive advisors. I cannot thank them enough.

For helpful comments and suggestions, I would like to thank my other committee member Khalil Iskarous. I am also honored to have Sandy Eckel on my committee; her engagement in this project and vast knowledge of statistics methods have helped me greatly. I am additionally grateful to the qualifying exam committee of this dissertation at its earlier stage. I have benefited from interacting with Rachel Walker, Karen Jesney, Shrikanth Narayanan, and Krishna Nayak.

In addition to my committee, I have received incredibly useful help from many scholars and friends at USC. The MRI data collection necessary for this project would not have been possible without the help of the SPAN MRI acquisition team. I am very thankful to my team, Asterios Toutios, Yongwan Lim, Colin Vaz, Zisis Iason Skordilis, and Tanner Sorensen. I would also like to thank the previous members, Jangwon Kim, Sajan Goud Lingala, and Yinghua Zhu. Many thanks to Reed Blaylock and Adam Lammert for providing me the Matlab codes for the MRI vocal tract analysis employed in this dissertation. I am happy to have overlapped with Miran Oh, my caring collaborator. The cool vertical larynx movement analysis has been possible thanks to her part on our larynx dynamics project.

I would like to thank Elsi Kaiser and Louis Goldstein for their valuable inputs on our ‘aɪskrɪm’ paper. I am also thankful to my coauthors of the prosodic convergence project, Dani

Byrd, Louis Goldstein, Samantha Gordon Danner, Sungbok Lee, and Benjamin Parrell, for the work they put in the project and their support. Thanks to everyone else in the USC Phon group and SPAN who have helped me both intellectually and in life outside of school over the years. I am thankful to Caitlin Smith, Samantha Gordon Danner, Hayeun Jang, Monica Do, Miran Oh, Saurov Syed, Brian Hsu, Ana Besserman, Mairym Llorens Monteserin, Chorong Kang, Mythili Menon, Thomas Borer, Ulrike Steindl, Maury Lander-Portnoy, Reed Blaylock, Sarah Harper, and all my friends who I've overlapped with, for the friendships that made my time at USC more than just a memorable experience. I am so grateful to have been a part of these groups full of intelligent and supportive minds.

Finally, I would like to thank my family for their love and support. My parents have been always supportive and encouraging along the way. My coolest parents-in-law, Christie Lewis and Boe Geib, have always believed in me. Special thanks to my brother-in-law, Aaron Geib, for providing me the pounds of Sulawesi coffee that fueled my early mornings and late nights. I thank my puppy goose, Snoops, for all the countless warm hugs she has given me. My final thanks and all my love go to Chad Geib who has always been my most relentless supporter. I would never have gotten here without his unending inspiration and love.

Support for this dissertation was provided by funding from NIH DC03172 to Dani Byrd, the project: "Tonal Placement - The Interaction of Qualitative and Quantitative Factors: TOPIQQ," subcontract D-71631-Z-600-145002301 from the University of Cologne, under funding from the Volkswagen Foundation, and NIH/NIDCD DC007124 to Shrikanth Narayanan.

Thank you all.

Table of Contents

Abstract

This dissertation investigates the complex interaction in the prosodic dynamics of consonant and tone, probing an essential role of phrasal prosody in spoken language production. The overarching hypothesis is that the local phonetic organization of a consonant system is regulated and shaped by the language's prosodic structure. The test language used to investigate this hypothesis is contemporary Seoul Korean. The two empirical studies presented here examine the segmental and tonal sensitivity to the unique phrasal prosodic system of this language, in which relatively fixed phrase tone patterns are co-active with its segmental tone patterns. Our systematic analysis of the global tonal structure demonstrates its interaction with the local phonetic distinctions of contrastive categories in this language—specifically, its three-way voiceless stop contrast. The acoustic and articulatory investigations in this dissertation provide an explanation for how phonological factors combine to shape the phrasal tone realization; these studies systematically illuminate the patterning of phonetic information for sequences containing varying consonant types [tense/lax] placed across several phrasal positions. Overall, both general cross-speaker patterns and individual speaker-specific patterns suggest that constraints for preserving paradigmatic and syntagmatic contrasts are simultaneously present and active in the phonology of younger speakers of Seoul Korean. The articulatory study with the real-time magnetic resonance imaging technique provides novel evidence for (a) articulatory mechanisms that express consonantal tenseness and tone and (b) the interplay between different phonological structures that deploy these mechanisms. Finally, this dissertation sheds light on some salient issues in sound change using the current findings as unique examples of variation and sound change in progress. The observed patterns of an ongoing tonogenic sound change are systematically influenced by higher-level phrasal prosodic contexts. Moreover, our results regarding the progression of the sound change across prosodic contexts suggest a further interaction of phrasal prosody with lexical word boundary in terms of information reorganization. Our findings have implications for an understanding of the complex role that prosodic conditioning can play in sound change. Taken together, this dissertation provides novel evidence for the seamless integration of segmental and suprasegmental phonological structure and contributes to our understanding of the complex orchestration of articulatory gestures as they are woven into the prosodic substrate of spoken language.

1. Introduction

The goal directed actions that produce language are intrinsically embedded in and structured by the cognitive system of language. This use of structured action to encode information is fundamental to human language. This dissertation strives to contribute to our understanding of this phonological structuring.

Speaking involves complex motor tasks that often require a great deal of precision in positioning articulators along the vocal tract over the course of time. As speech unfolds, the timing and amplitude patterns of articulatory actions, or gestures, are systematically governed not only by contrastive properties of word-internal, i.e. lexical, units but also by informational aspects of the speaker's plan, which are collectively referred to as phrase-level prosody. Language users masterfully coordinate these motor behaviors, weaving them into a multiscale structure that serves to communicate with other speakers in their language community.

This dissertation addresses a challenge in understanding phonological structure that gives rise to surface variations in the speech signal. Specifically, it considers how speakers integrate prosodic information with speech gestures while planning and producing words and phrases. The work presented here adopts the theoretical stance that understanding spoken language is not a function of transformations (or translation) of representational structures (traditionally described with symbols), but rather is better understood by deploying mathematical tools to characterize the deployment of cognitive language representation[1]. Specifically, dynamical representations of language and its overt realization allow for the explicit recognition that language unfolds in time.

There are many examples studied in phonology where some apparently categorical process is mimicked by a more constrained gradient or variable process. This challenges most traditional views of phonology (organization of cognitive units) and phonetics (physical properties of speech) that assume there is no formal link between the two because they are distinct components of grammar. For example, to review a now frequent example from

---

[1] Such dynamical applications to language are found: e.g., Browman and Goldstein 1985; Browman and Goldstein 1989; Byrd and Saltzman 1998, 2003; Elman 1995; Fowler, Rubin, Remez, and Turvey 1980; Gafos and Benuš 2006; Iskarous 2016, 2017; Jordan 1986; Kelso, Vatikiotis-Bateson, Saltzman, and Kay 1985; Roon and Gafos 2016; Saltzman, Nam, Krivokapić, and Goldstein 2008; Saltzman and Munhall 1989; Sorensen and Gafos 2016; Tilsen 2016; Tuller, Case, Ding, and Kelso 1994; Vatikiotis-Bateson and Kelso 1993; etc.

Browman and Goldstein (1990), the production of the final consonant "*t*" of "*perfect*" followed by the word "*memory*" can vary drastically during the act of talking, which would be conventionally analyzed as an example of *grammatical* alternation between the presence of the unit "*t*" and its deletion "*Ø*." That is, when each word is enunciated as part of a word list ("*..., Perfect, Memory, ...*"), "*t*" is clearly audible; in contrast, when the two words are said as part of a fluent phrase ("*…perfect memory…*"), the "*t*" is often *not* audible. However, the conventional theories largely overlook the fact that the *action* of tongue tip contact against the alveolar ridge for the "*t*" is still present even in the fluent version with no audible "*t*" (Browman and Goldstein 1990). An alternative view rejects a grammatical mediation between the phonological representation and its phonetic implementation provides a more principled account. In this approach, the difference between the list and fluent versions is due to variation in the gradient degree of temporal overlap between the tongue tip closure gesture for "*t*" and the following lip closure gesture for "*m*,[2]" and overlap is known to vary in a principled way during speaking (Browman and Goldstein 1990).

A viable framework for understanding spoken language production must, we argue, be able to describe not merely what processes manipulate phonological representations but also the nature of those representations themselves and how they unfold in time according to the phonological structuring of the grammar (i.e. temporal organization of speech). The dynamical systems approach pursed in Articulatory Phonology by Browman and Goldstein offers such an approach connecting underlying cognitive control and variability in performance. Fundamental to this framework is the postulation of dynamically modeled, abstract speech gestures that function simultaneously as units of information (contrast) and as units of action in speech production. Crucially, gestures are defined in terms of *tasks* within the vocal tract such as constrictions and releases of vocal tract subsystems. This grounding premise has insightfully handled a number of puzzles in the field, providing principled explanations for spoken language phenomena via the composition of gestures, manipulation of their dynamical parametrization, and the spatio-temporal overlap and coordination among gestures. Phenomena that have been addressed within this conceptualization include, for example, variabilities arising from speaking styles (Browman and Goldstein 1986), segmental composition, syllable structure (Browman and

---

[2] And possibly gradient realization of the "*t*" action itself.

Goldstein 1988, 2000), and, more recently, phrasal and accentual structure (Byrd and Saltzman 1998, 2003; Saltzman, Nam, Krivokapic, and Goldstein 2008).

It has become increasingly clear that speech planning processes must integrate prosody, or informational structuring of the component lexical units. Prosody, in a structural sense, refers to the organization of words into larger units such as feet and phrases and the marking of new or salient information within those groups. Structurally, the term prosody also is used to refer to rhythmic properties of these words and groups such as stress patterns. Qualitatively, the term prosody serves to refer to intonation, tone, amplitude, and local lengthening (duration) driven by phrasal, rhythmic, and/or prominence structure. Understanding the sensitivity of phonological units to prosodic structure has received much attention. Previous work on prosodically conditioned variability has largely focused on the speech planning process (Keating and Shattuck-Hufnagel 2000; Krivokapić 2007), phrase edge spatiotemporal patterns and models of these effects using a prosodic gesture (Byrd and Saltzman 1998, 2003) or modulation gesture (Saltzman et al. 2008), boundary and prominence coordination (Katsika, Krivokapić, Mooshammer, Tiede, and Goldstein 2014), and, to a lesser extent, lexical tone gesture (Gao 2008).

This dissertation extends this work that has looked at prosodically conditioned variability to a consideration of the interaction of lexical (i.e. segmental) tone and phrasal (i.e. accentual) tone. The studies leverage contemporary Seoul Korean to tackle the intricacies of prosody/segment interaction, probing an essential role of phrasal prosody on consonant and tone interaction. Contemporary Seoul Korean offers as an excellent test bed for this inquiry given that its unique phrasal prosodic system exhibits relatively fixed phrase intonation patterns, calling for a new way of expressing the regularities between consonant type and tone. The investigation of this complex integration contributes to our understanding of prosodically conditioned variability not only by providing the phonetic description needed for the new pronunciation norms emerging in younger generations speakers of this language, but also because the observed tonal patterns provide new data for testing a gestural theory of intonation.

The overarching hypothesis of this dissertation is that the local phonetic organization of a consonant system is regulated and shaped by the language's prosodic structure. Through the investigation of the phrasal tone patterns that are co-active with its segmental tone patterns, our goal is to illuminate the relation between phonological contrast as encoded in articulatory gesture and the phrasal and accentual goals within which those gestures are embedded. Specifically, we examine how the global tonal structure demonstrates its interaction with the local phonetic distinctions of three-way voiceless stop contrast in Seoul Korean. The acoustic and articulatory investigations presented here provide an explanation for how phonological factors combine to shape the phrasal tone realization; these studies systematically expound the patterning of phonetic information for sequences containing varying consonant types [tense/lax] placed across several phrasal positions.

This dissertation also offers discussion on some salient issues in sound change. An ongoing sound change is an example of adaptation from one stable, cognitively effective state to another over generational time. What we document here is the current status and progression of an ongoing tonogenic sound change in younger speakers of this language. Using the current findings as unique examples of variation and sound change in progress, this dissertation elucidates the complex role that prosodic conditioning can play in sound change.

Taken together, this dissertation provides novel evidence for the seamless integration of segmental and suprasegmental phonological structure and contributes to our understanding of the complex orchestration of articulatory gestures as they are woven into the prosodic substrate of spoken language.

1.1. Outline of the dissertation

Chapter 2 examines the segmental and tonal sensitivity to phrasal prosodic structure in a language with non-flexible intonational patterns, i.e., the contemporary Seoul dialect of Korean. It investigates the phonetic information organization of the three-way voiceless stop contrast (lenis /p/, aspirated /p^h/, fortis /p*/) and the consonant-type effect on tones in various prosodic contexts in younger speakers of this language. This chapter uncovers the phonological factors that shape the Accentual Phrase tones, which ultimately accounts for how the local phonetic information organization is shaped or constrained in different prosodic locations. Implications for an intricate interplay between the paradigmatic contrast maintenance and syntagmatic tonal patterns are discussed.

Chapter 3 consists of an articulatory investigation of Seoul Korean laryngeal consonant and tone dynamics, utilizing the real-time magnetic resonance imaging technique. It tackles questions of what motor tasks are deployed for consonantal “tenseness” and tone gestures, and how they function within the phonological system. This chapter discusses (a) articulatory mechanisms that express tone and tenseness and (b) the interplay between different phonological structures that deploy these mechanisms. Both general cross-speaker patterns and individual speaker-specific patterns reveal an intricate interaction between the lexical tones attributable to the [tense/lax] contrast and the prosodic structure (phrase-level prosody).

Chapter 4 discusses diachronic considerations of the findings from the previous two chapters, and probes an essential role of phrasal prosody in consonant and tone interaction. It is shown how the observed patterns of an ongoing tonogenic sound change are systematically influenced by higher-level phrasal prosodic contexts. This chapter further discusses implications for understanding of the complex role that prosodic conditioning can play in sound change.

Chapter 5 presents an overall summary and discussion of the results in the previous chapters and some theoretical implications of this work, as well as directions for future reseasrch.

2. The role of prosodic structure in modulating consonant-tone interaction

2.1. Introduction

The study presented here examines the segmental and tonal sensitivity to phrasal prosodic structure in a language with non-flexible intonational patterns, i.e., the contemporary Seoul dialect of Korean. In younger-generation speakers of Seoul Korean, the phonetic properties (voice onset time [VOT] and fundamental frequency [f0]) of the three-way stop contrast (i.e. lenis in /**p**ul/ "*fire*", aspirated in /**pʰ**ul/ "*grass*", fortis (or tense) in /**p***ul/ "*horn.*") has been well documented in phrase-initial position. They have been described as produced with distinctive combinations of VOT and f0 values associated with the following vowel (e.g., Cho, Jun, and Ladefoged 2002; Cho and Keating 2001; Kang and Guion 2008; Lee and Jongman 2012; Lisker and Abramson 1964; Table 1).

Table 1. VOT and f0 patterns in the Seoul Korean 3-way voiceless stop contrast, lenis /**p**/, aspirated /**pʰ**/, fortis /**p***/, in phrase-initial syllables. Different color codes indicate the tenseness distinctions, '**LAX**' vs. '**TENSE**.'

	VOT	f0
#/**p**a/	long	low
#/**pʰ**a/		high
#/**p***a/	short	

It has long been posited that setting aside the aspiration degree difference, these three stops are further divided into two categories, 'lax' and 'tense,' characterized by different levels of "articulatory strength" (e.g., Cho, Jun, and Ladefoged 2002; Cho, Son, and Kim 2016; Cho and Keating 2001; Dart 1987; Han and Weitzman 1970; Hirose, Lee, and Ushijima 1974; Jun 1996; Kagaya 1974; Kim 1965; Kim, Honda, and Maeda 2005; Kim, Maeda, and Honda 2010; Lee and Jongman 2012; Son, Kim, and Cho 2012). Relatively weaker articulation is associated with the lenis stop production as compared to aspirated and fortis; this includes: relatively slower buildup rate of buccal and subglottal pressure, shorter duration for maintaining the increased pressure, less linguopalatal contact (in the case of coronal stops) or smaller lip muscle activity or less constriction (bilabial stops) for the occlusion, lower level of burst intensity and during the aspiration period, smaller airflow amount of airflow following release, weaker harmonic components and slower rate of vibration in the following voice onset, etc. As such, lenis (lax) stops behave differently from aspirated and fortis stops, which pattern together as tense stops in many phonetic measures.

Although the articulatory strength difference appears to differentiate the tense (aspirated, fortis) and lax (lenis) series, there is likely a more direct cause for the f0 distinction between the two categories. Hirose, Lee, and Ushijima's (1974) electromyography study reported the activities of the intrinsic phonatory laryngeal muscles including both tensor and adductor muscles during the production of these stops in Kyungsang Korean. Aspirated stops are produced with suppressed activity of the tensor muscle such as cricothyroid and vocalis muscles throughout the closure. This suppression is always followed by a steep increase in muscle activity after the release, which in turn gives rise to high f0 during the following vowel. With respect to fortis stops, there is a substantial increase in vocalis muscle activity immediately before the stop release. The increased tension (stiffening) of the vocal folds and constriction of the glottis during or immediately after the closure (laryngealization, Abramson and Lisker 1972) should be responsible for the short VOT and high f0 in the following vowel. In contrast, lenis stops do not show a sharp increase in tensor muscle activity before or after the stop release, resulting in lower variants of f0. See the column 'f0' in Table 1.

Having identified physiological causes responsible for the f0 difference between tense and lax stops, now let us turn to the VOT difference among the three stops in Seoul Korean. It has been demonstrated that differences in VOT can be used to index voicing and aspiration contrasts among initial oral stops in various languages (since Lisker and Abramson 1964). In other words, VOT has become a standard measure to differentiate various oral stops with different relative timing between laryngeal and oral gestures. The glottal width during the closure is larger for aspirated stops, intermediate for lenis stops, and narrower for fortis stops (e.g., aerodynmaic evidence: Cho et al. 2002; Dart 1987; Lee and Jongman 2012; articulaotry evidence: Kagaya 1974; Kim et al. 2005, 2010). The glottal width can be arguably correlated with the degree of glottal opening at the time of the release; therefore VOT is indicative of the aspiration degree. If this holds, aspirated stops will show the longest VOT values, and lenis stops will show intermediate VOT values. For fortis stops, having the vocal folds approximated well before the articulatory release will give rise to the shortest VOT values.

Previous studies conducted during the 1960s and 1970s reported that the distinction among the stop contrast in Seoul Korean was made *mainly* through VOT difference (Han and Weitzman 1970; Kim 1965; Lisker and Abramson 1964; Lisker and Abramson 1971). A fortis stop is easily distinguished from the other two varieties due to its particularly short VOT value. However, more recent studies found that lenis and aspirated stops are no longer distinguished solely by VOT values in either production or perception, but rather, f0 has come to play a central role (Kang and Guion 2008; Silva 2006; cf. Lee and Jongman 2012). Most recently, it has been shown that there is a complete loss of VOT distinction between lenis (previously intermediate in VOT) and aspirated (previously the longest VOT) stops in younger generations, born during or after the 1980s (Bang, Sonderegger, Kang, Clayards, and Yoon 2018; Kang 2014; 'VOT' column in Table 1). For example, the average VOT difference between lenis and aspirated stops is only 5 ms for younger generation speakers born in the 1980s, which used to be 30 ms for older generation speakers born in the 1930s. Along with the VOT mergers, Kang also reported that the tonal distinction among word-initial stops has become sharper over time (Table 1). Bang et al. (2018) further shows that this f0 distinction among phrase-initial stops propagates across words (with different frequencies) and vowel (height) contexts, providing evidence for a gradual tradeoff relation between VOT and f0. Both of these apparent time reports imply that f0 has come to function as a reliable indicator of the lexical contrast in this language.

Prior to this study, the phonetic properties (including VOT and f0) of the lax-tense stops has been tested only with word-initial stops (crucially in most cases, directly nested under a larger phrasal unit(s) such as an Accentual Phrase [AP] or further nested under an Intonational Phrase [IP]). Earlier studies had shown that the stop contrast in word-medial positions is preserved through intervocalic voicing of the lenis series (e.g., Han and Weitzman 1970; Kagaya 1974; Kim 1965). In a recent acoustic study with earlier generation Seoul and Chonnam Korean speakers (born in the 1970s), Jun (1994) showed interspeaker variability and segmental context-derived variability in the degree of intervocalic lenis undergoing voicing. This raises a possibility that changes might also be occurring in word-internal position in younger generations. Since Kang's (2014) finding, no study has yet examined the production of the stop contrast in *non*-initial position. *It remains unknown if the continuing phonetic organization reported in AP-initial position is associated with changes in phonetic realization of the contrast in AP-internal position.*

In Seoul Korean, the interaction between tone and segment is further intertwined with phrase-level prosody. This language exhibits an intonationally defined phrasal level, denoted as the Accentual Phrase (AP), in its prosodic structure. An AP has been described as being associated with a phrasal tonal sequence *T*HLH, where the initial tone (*T*) interacts with segmental quality since Jun (1993). Specifically, if the AP-initial segment is a tense consonant (fortis and aspirated stops), the initial tone is realized as a high (*H*) tone, whereas a low (*L*) tone is assigned in the case of lax consonants including lenis stops and the sonorants.

Jun (1993, 1996) reported that the f0 difference between tense and lax categories persisted throughout the entire vowel at the beginning of an AP but was only maintained briefly during the initial portion of the vowel in the middle of an AP. She concluded that the phonologization of f0 is limited to the AP-initial position, resulting from a boundary-induced strengthening to enhance [+/- stiff vocal cords], the basis of the tense and lax distinction. In the case of an AP starting with a tense consonant, she further observed an effect of the initial *H* (e.g., ***H***HLH) on the following tones (e.g., *H***HL**H), in such a way that the phrase-initial extra high f0 values are followed by high f0 values of the following syllables, up to the penultimate syllable of an AP. This f0 patterning throughout an AP was quantitatively confirmed, recently in Cho and Lee (2016). Using telephone number strings and natural word readings, they compared the mean f0 values of AP syllables in tense-initial (repeating sequences of *tense*-*lax*-, *tense*-*tense*-) and lax-initial (repeating sequences of *lax*-*lax*-, *lax*-*tense*-) APs that were 2-5 syllables long. However, neither the segmental contexts or the prosodic phrasing were thoroughly controlled[3]. Cho and Lee reported that the pitch range of an entire AP was higher in a tense consonant-initial AP than in a lax consonant (or vowel)-initial AP. They also found that this type of the segmentally induced tonal difference (tense vs. lax) was much smaller when the tense-lax contrast was manipulated AP-medially. The authors concluded that the extended effect of the AP-initial (tense

---

[3] The APs compared in their study (for both telephone number string and natural word readings) consist of onset consonants (sometimes onset vowels) varying in manner and place of articulation and any high vowels, occasionally with coda elements. For example, their 4-syllable natural word comparisons are made with the following target APs: [[**pʰɨ**.**ɾ**ʌ.**pʰ**o.**ɕɨ**]$_{pwd}$]$_{AP}$ (*tense-lax-tense-lax*) vs. [[**kʰ**ʌm.**pʰ**ju.**tʰ**ʌ.]$_{pwd}$[**ɕʰ**ɛk]$_{pwd}$]$_{AP}$ (*tense-tense-tense-tense*) vs. [[**u**.**ɾ**i.]$_{pwd}$[**k**o.**m**o]$_{pwd}$]$_{AP}$ (*lax-lax-lax-lax*) vs. [[**p**i.**h**eŋ.**k**i]$_{pwd}$[**pʰ**jo]$_{pwd}$]$_{AP}$ (*lax-tense-lax-tense*)). According to the authors, one of their 5-syllable AP conditions actually show two individual APs [[in.tʰʌ.nɛt]$_{pwd}$]$_{AP}$[[ɕʰiŋ.ku]$_{pwd}$]$_{AP}$, in which f0 values behave differently from those measured in the 5-syllable phone number string reading. As shown in the examples, the APs are formed by morpheme concatenations that were not carefully manipulated. This is worrying in terms of a potential difference in prosodic parsing and/or tone assignment. Moreover, both reading conditions are designed to induce similar phrasing of reading a word-list, indicating the target AP might have been immediately nested under a bigger phrase that influences the phrase-final tonal elements.

consonant-induced) *H* is potential evidence of a tonogenic sound change but no explicit discussion or further testing of this possibility is available in this work. We will return to this topic later, and particularly in Chapter 4.

The previous reports strongly suggest that the *temporal scope* of the segmentally triggered f0 difference varies as a function of prosodic position (AP-initial position > *non*-initial position) in this language. Prior to the present study, the temporal scope of the initial *T* in relation to the recently documented VOT mergers of the stop consonants has not been systematically investigated, nor has the consonant-type effect on f0 in *non*-initial position. Thus, it remains to be examined how the local phonetic organization of this system in younger Seoul Korean speakers is regulated and shaped by the prosodic structure.

This study probes an essential role of phrasal prosody in consonant and tone interaction in which constraints for preserving paradigmatic and syntagmatic contrasts may be simultaneously present and active in the phonology. In Seoul Korean, the tonal contrast between the tense and lax stops is paradigmatically enhanced in a prosodically strong, AP-initial position (e.g., Cho and Lee 2016; Cho and Jun 2000; Jun 1998). However, in AP-internal position we hypothesize that younger generation Seoul speakers exhibit an interplay between preserving the consonantally derived tonal contrast and realizing the global (syntagmatic) tonal patterns characterizing an AP.

A systematic analysis of the global tonal structure of the contemporary Seoul Korean AP will demonstrate its interaction with the local phonetic distinctions of the contrastive categories in this language. To identify what phonological factors shape the AP tones, we carried out an acoustic study that primarily examines regularities in phonetic information within tonal and segmental strings that vary by consonant type and phrasal position. Previous studies reported that the initial *H* triggered by a tense consonant was followed by high f0 values of the following syllables up to the penultimate phrasal (Jun 1993, 1996) or up to the final (Cho and Lee 2016) syllable of an AP. Motivated by these reports, we hypothesize that the temporal scope of the initial tone (*T* in *T*HLH) will be broad regardless of whether the AP initial consonant is a lax stop or tense stop. That is, we expect to see *both* initial *L* and *H* exerting themselves on the following syllables of the phrase. The existing Accentual Phrase model postulates that non-initial position is a tone-neutralizing context, impervious to consonant type. For example, the second syllable in a 4-syllable AP is assumed to carry a high tone as in *T***H**LH, regardless of the local consonantal context. One might view this as a cap imposed by the global tonal structure in the tone-neutralizing context on freely realizing the consonantally triggered local f0 difference. This leads to the prediction that the effect of the phrase-*initial T* may spread throughout an AP. And consequently that a much smaller, locally constrained scope of the f0 difference between the tense and lax may arise AP-internally due to the interaction between the tone-neutralizing context and the 'global' scope of the preceding *T*.

Jun (1998) also observed that the phrase-medial L (in *T*H**L**H) becomes lower, as the number of syllables within an AP increases. This alludes to a possibility that at some point the initial *T* effect will cease or diminish to insignificance. To uncover how long the temporal scope of initial *T* is, this study manipulates the number of syllables per AP, ranging from 3 to 5. In addition to probing the size of initial *T* effect, factoring this in will let us see how the global (overall) tonal pattern changes due to (potential) tonal undershoot when an AP is composed of fewer than the canonical 4 syllables.

Gathering evidence suggests that the phrase-initial f0 effect may in young Seoul speakers be a stabilized exponent of the phrasal prosody. In this study, we further hypothesize that the enlarged initial *L* and *H* difference has become categorical. If that is so, one would expect to see a sharp bimodal distribution of f0 values in AP-initial position, which should accompany the recently documented initial VOT mergers. The abovementioned study by Jun (1993) with earlier generation Seoul and Chonnam speakers reported that quite a few lenis tokens in word-medial position were voiced (75%, 90 out of 120). However, as laid out earlier, no prior work has fully investigated this in different prosodic positions. That is, it is unknown if the ongoing VOT merger and enhanced tonal contrast reported in AP-initial position for contemporary Seoul Korean are also associated with such phonetic realizations of the three voiceless stop contrasts in AP-*internal* position. The current study will test whether the *non*-initial lax and tense stops also exhibit somewhat enhanced tonal distinction, and whether this is accompanied by de-stabilization in the other phonetic dimension (i.e. intervocalic lenis voicing). If the f0 difference between non-initial lax and tense categories is constrained by phrasal prosody, we expect to see more gradient f0 distributions in AP-internal position along with a relatively clearer distinction in the other phonetic dimension.

A prosodic word can be nested directly under an AP, and an AP can be nested under a bigger phrasal unit such as an Intonational Phrase (IP) (Selkirk 1986). In Seoul Korean, both AP-and IP-initial positions are prosodically stronger than (phrase-medial) word-initial positions (but arguably IP-initial > AP-initial, (Cho and Keating 2001; Cho, Lee, and Kim 2011; Jun 1993). In this investigation of contemporary Seoul speakers' phonetic realization of the stop system in prosodic positions of different boundary strengths, we test whether the new sound pattern is still sensitive to these prosodic contexts and if so in what way.

2.2. Method

2.2.1. Speakers

Six native speakers of Seoul Korean, 3 males (M1, M2, M3) and 3 females (F1, F2, F3), participated in the experiment. They were born between 1980 and 1990, categorized as younger generations' Seoul dialect, and completed their undergraduate degrees in Seoul. All 6 speakers were pursuing graduate studies at the University of Southern California at the time of the recording.

2.2.2. Test materials and recording

Disyllabic target words (c1v1c2v2) were selected to elicit a range of different intonational patterns in different phonological contexts. As we expect to see different f0 patterns as a function of consonant contexts, the target syllable was designed to have one of the four bilabial stop consonants, NASAL /m/, LENIS /p/, FORTIS /p*/, and ASPIRATED /p^h/, as an onset followed by a low central vowel /a/. Throughout this paper, /m/ and /p/ are labeled LAX consonants, and /p*/ and /p^h/ are labeled TENSE consonants, since the two groups have been described as having different f0 characteristics, which this study also confirms.

To see if there is any effect of the position of the target syllable within a word, sets of words were selected based on the following criteria. Our primary criterion was to identify word candidates that were real disyllabic words containing the target syllable in a desired position-first or second syllable within a word. From these candidates, words with non-target syllables always starting with a lenis or a sonorant were selected, as the phonologically associated tone is supposed to be a low tone. Note that our phonologically optimal set of target words was not balanced in terms of equal word frequency or parts of speech. Given that each target was embedded in an AP, our speakers were asked to treat each target word as a noun that modifies the following noun within the same AP to resolve the unnaturalness potentially arising from syntactic or semantic incorrectness.

For the AP-INITIAL condition, the target syllable was placed at the beginning of a phrase-initial word such as in /**ma**.na/, /**pa**.ta/, /**p*a**.ta/ and /**pʰa**.pa/ (presented in Hangul in the experiments). For the AP-INTERNAL condition, the target syllable was the second syllable of a phrase-initial word preceded by a lenis- or nasal-initial syllable so that the preceding syllable always gets a low tone. This condition included words like /ka.**ma**/, /ɕa.**pa**/, /pa.**p*a**/ and /ma.**pʰa**/. See Table 2 for a glossary of target words.

Table 2. A list of target items used in this study. Target syllables are in bold.

	AP-INITIAL	gloss	AP-INTERNAL	gloss
NASAL	**ma**.na	“super natural power”	ka.**ma**	“kiln”
LENIS	**pa**.ta	“sea”	ɕa.**pa**	“Java”
FORTIS	**p*a**.ta	slang for “butter”	pa.**p*a**	“being busy”
ASPIRATED	**pʰa**.pa	nickname for “*Paris Baguette*” (a famous Korean bakery)	ma.**pʰa**	“stir-fry”

In order to form APs that vary in number of syllables (3-SYLL vs. 4-SYLL vs. 5-SYLL), each disyllabic target word was followed by a monosyllabic bound morpheme or formed a phrase with a polysyllabic constituent built through compound postpositions (noun + case marker). For example, the monosyllabic suffix /-man/ (post-positional suffix “only”) was used to form a tri-syllabic AP with the target word (c1v1c2v2-/man/). A 4-syllable accentual phrase consisted of the target word and /pap-ɨl/ (“*rice*” + accusative case marker; c1v1c2v2+/pa.pɨl/), and a 5-syllable AP was formed with the target word and /man.tu-ɾɨl/ (“*dumpling*” + accusative case marker; c1v1c2v2+/man.tu.ɾɨl/). Note that the first syllable of the attached morphemes always consisted of a bilabial stop in the lax category (/m/ or /p/) and the low central vowel /a/.

The target AP was then embedded in two different carrier sentences that allowed different phrasings for the IP-initial and IP-medial conditions. The target AP was always an object of the entire sentence followed by the main verb. See a) below.

a) Target words in carrier sentences

IP-initial: [[kɨ.a.i.tɨ.ɾɨn]AP [ʌn.ɕɛ.na]AP]IP, [[“__________”]**AP** [ɕo.a.hɛ.jo]AP]IP.
[[*the children*]AP [*always*]AP]IP, **[[target phrase]AP** [*like*]AP]IP.
IP-medial: [[t*al.ki.u.ju.wa]AP **[“__________”]AP** [kɨ.ɾjʌ.s*ʌ.jo]AP]IP.
[[*strawberry milk and*]]AP **[target phrase]AP** [*were drawn/sketched*]AP]IP.

For the IP-INITIAL condition, speakers naturally put a pause after the adverb, /ʌn.ɕɛ.na/ (“*always*”), without any specific instruction. For the IP-MEDIAL condition, an IP consisted of three APs, and the target AP was always positioned in the middle. In any case, the target AP was never placed at the end of an IP, so as to avoid further interactions between segmental/tonal strings and the big prosodic juncture (such as phrase-final lengthening and a low boundary tone [L%] assignment). In this study, the number of APs between IP-INITIAL and IP-MEDIAL conditions was not kept the same for the following reason. For the IP-INITIAL condition, the entire sentence length was already relatively long, as it contained several polysyllabic APs. Introducing an extra AP between the target AP (object) and the main verb may cause variability or undesired renditions of phrasing. There could be a possible confounding coming from the different numbers of APs forming an IP (two APs in the IP-INITIAL condition vs. three APs in the IP-MEDIAL condition). For example, if there is any tonal or durational effect as a function of boundary strengths, we expect to see longer duration, higher pitch, or greater pitch excursion IP-initially than IP-medially. If this effect is observed, it might have also interacted with the fact that an IP with fewer APs could potentially have longer syllable or segmental durations. However, it turns out that there is no effect of boundary strength at all on our measures.

Each speaker repeated each sentence 5 times in a randomized order. In total, 1440 tokens were collected and analyzed (6 speakers x 4 consonants [2 consonant types] x 2 boundaries x 3 different numbers of syllables per AP x 2 target syllable locations x 5 repetitions). A head-mounted microphone was used to record speakers.

2.2.3. Measurements

In this study, measurements included 1) stop closure duration measured in the IP-MEDIAL condition, 2) VOT of each stop consonant, 3) the number of lenis tokens exhibiting voicing (measured only for the AP-INTERNAL condition), 4) f0 maximum values during all the vowels in the target AP, and 5) f0 excursion between the target vowel and its adjacent vowel

(|v1–v2|). Acoustic analysis was carried out using Praat software (Boersma and Weenink 2018). The details of each measure are given below.

In the stop-initial target syllable /c1v1/, we measured stop closure duration and VOT values of the ASPIRATED /p^h/, LENIS /p/ (only when not voiced) and FORTIS /p*/ stops, excluding the nasal stop /m/ from these measures. Closure duration of the voiceless stops was measured from the cessation of the vowel of the preceding syllable ([wa]), to the beginning of the stop burst seen in the spectrogram, combined with a complete silence seen in the waveform. This measure was done only with the IP-MEDIAL tokens, in which there is no preceding phrasal pause that is not separable from the silence portion during the closure. VOT was measured from the beginning of the stop burst seen in the spectrogram, combined with the end point of a complete silence seen in the waveform, and measured up to the onset of laryngeal pulsing of the following vowel.

Jun (1994) categorized the voicing status of a Korean lenis stop into three: ‘completely voiceless’, ‘partially voiced’, and ‘completely voiced’ depending on how many vocal fold vibrations there were, which were determined by the number of voicing bars/pulses during the stop closure (0 vs. 1-2 vs. more bars continuing throughout the whole stop closure). In our study, lenis tokens in the phrase-INTERNAL condition, #/ɕa.**p**a/ are considered to be possibly subject to intervocalic lenis voicing. Interestingly, most speakers showed patterns of all or nothing (fully voiced vs. completely voiceless), except for Speaker M3 who showed all three categories defined by Jun. In this study, lenis tokens produced with any voicing bars during the stop closure were considered as *voiced*, and the number of voiced lenis tokens for each speaker was recorded. (These voiced tokens were excluded for the VOT analysis pooled across speakers in §2.3.2.2.)

To understand the global pitch patterning during the target AP comprehensively, f0 maximum values ($f0_{max}$) were measured from each vowel over the course of the entire AP. Each vowel was manually segmented and labeled on a Praat text grid. After confirming pitch contours were accurately overlaid on the vowels, f0 maximum values were automatically taken from the vowels using a Praat script. In some cases, the vowel was too breathy (when preceded by lenis or aspirated stops produced with strong aspiration), resulting in difficulties in measuring f0 values. Tokens with this issue were excluded from our measurement. The target f0 excursion was then calculated as a difference between f0 maximum values of the target vowel and its immediately adjacent vowel—i.e., |v1–v2|.

2.2.4. Significance testing

The values of each measure were analyzed using linear mixed effects models. All statistical analyses were made in *R* (*R* Development Core Team 2018) using the *lmer* function of the *lmerTest* package (Kuznetsova, Brockhoff, and Christensen 2016).

Four phonological factors were entered in the analysis as fixed effects: 1) Consonant Type (TENSE [ASPIRATED /p^h/ & FORTIS /p*/] vs. LAX [LENIS /p/ & NASAL /m/]); 2) Phrase Position (AP-INITIAL vs. -INTERNAL); 3) Prosodic Boundary (IP-INITIAL vs. IP-MEDIAL); 4) Syllable Number (3-SYLL vs. 4-SYLL vs. 5-SYLL APs). Separate models were fitted for different dependent variables, as follows.

For the stop closure duration (CD) and VOT measures, because it is important to see how each stop consonant is differentiated, TENSE and LAX categories were further divided into subcategories, ASPIRATED /p^h/ (TENSE), FORTIS /p*/ (TENSE), and LENIS /p/ (LAX), but excluding the NASAL /m/ (LAX) category. Note that there is no specific hypothesis or prediction regarding the effect of the number of syllables per phrase on these measures. Therefore, the VOT and stop closure duration models contained three factors, Stop Item (ASPIRATED VS. FORTIS VS. LENIS), Syllable Position (Phrase-INITIAL vs. -INTERNAL), and Prosodic Boundary (IP-INITIAL vs. IP-MEDIAL). Evaluating effects of the Stop Item predictor with the three levels was done by building 3 separate regression models, having everything else in the model remain the same, but differing only in which pair comparison the predictor included for each model (CD_1 or VOT_1: ASPIRATED VS. LENIS; CD_2 or VOT_2: ASPIRATED VS. FORTIS; CD_3 or VOT_3: LENIS VS. FORTIS).

According to the formal description of tonal patterns as a function of segmental quality, lenis and nasal consonants should pattern together as a lax category and aspirated and fortis consonants should pattern together as a tense category in terms of f0 targets—i.e., lower f0 values for the lax category and higher f0 values for the tense category. To confirm this, we first tested how $f0_{max}$ values in the target vowel (i.e. v1 or v2) are influenced by Consonant Item (ASPIRATED VS. FORTIS VS. LENIS VS. NASAL). After this confirmatory step, the levels between consonants were merged so that Consonant Type include two levels—TENSE (ASPIRATED and FORTIS pooled) and LAX (LENIS and NASAL pooled) categories. It is important to note that our focus is on *interaction* among predictor factors (especially between Consonant Type and Syllable Position), rather than a simple effect of each predictor.

As random effects, intercepts were fit for both speakers and items. In contrast to a more traditional approach with data aggregation and repeated-measures of analysis of variance, *lmer* allows controlling for the variance without the data aggregation. These random effects significantly improved model fit according to likelihood ratio tests that compare models with and without the effect (using the *anova* function in *R*, see Baayen 2008; $p<.05$ for all). This suggests that there was indeed variation across individual speakers and items. We resolved those idiosyncratic variations that are due to multiple responses per speaker and per item by adding random effects of speakers and items in our model. Random effect slopes that are also specific to predictor factors were not included because they prevented the model from converging given the unbalanced nature of observational data and the number of additional factors added to the model when including random slopes. For this reason, we believe that random intercepts represent the maximal random effects structure for our data. Visual inspection of residual plots did not reveal any obvious deviations from homoscedasticity or normality.

Each analysis started with a full model that included a continuous dependent variable (e.g., VOT, stop closure duration, f0 maximum values in each syllable, f0 excursion between v1 and v2), and speakers and items as random factors. As there was more than one predictor in the model, we reduced the predictor variable(s) from the full model incrementally to see whether the fuller or sparser model was the better fit. We compared the likelihood of two different models (the fuller model with the effect in question against the model without the predictor that is of interest) using Chi-square tests on the log-likelihood values (using Wilk's Theorem). Whenever there was a significant interaction between predictor factors, we ran post-hoc regressions by condition. For the post-hoc tests, *p*-values were obtained by using the Satterthwaite's (1946) approximate method.

For all tests, *p*-values less than or equal to .05 were considered significant. Some other specific details in the analysis of each dependent variable are included in each corresponding subsection of §2.3 for ease of reading.

2.3. Results

2.3.1. Intervocalic lenis stop voicing

Table 3 shows the number of lenis stop tokens that are produced as voiced in AP-internal, intervocalic position.

Table 3. Number of voiced LENIS tokens in the AP-INTERNAL condition (pooled across Prosodic Boundary & Syllable Number conditions).

Speaker ID	# of voiced lenis tokens
F1	7/30
F2	5/30
F3	14/30
M1	0/30
M2	22/30
M3	9/30

In total, only 57 tokens out of 180 are voiced (32 %). Interestingly, the number varies greatly by speakers, ranging from 0 to 22 out of 30. Speaker M1 does not show any voiced tokens at all.

2.3.2. Stop consonant duration and VOT measures

For the closure duration model, predictors included Stop Item (ASPIRATED vs. FORTIS vs. LENIS) and Phrase Position (AP-INITIAL vs. AP-INTERNAL). As the stop closure duration measure was made only with IP-medial tokens, the Prosodic Boundary predictor was excluded. For the VOT model, predictors included were Prosodic Boundary (IP-INITIAL vs. IP-MEDIAL), Phrase Position (AP-INITIAL vs. AP-INTERNAL) and Stop Item (ASPIRATED vs. FORTIS vs. LENIS). The numerical results of the linear mixed effects regressions can be found in Appendix 1Appendix 2.

2.3.2.1. Closure duration

In all three models, there is a significant Stop Item effect on closure duration (CD$_1$: $\chi^2(1)=7.88$, $p<.005$, CD$_2$: $\chi^2(1)=8.92$, $p<.005$, CD$_3$: $\chi^2(1)=9.72$, $p<.005$). Closure duration is longest for fortis stops, intermediate for aspirated stops, and shortest for lenis stops. A significant Phrase Position effect is only found with the ASPIRATED and FORTIS comparison (CD$_3$: $\chi^2(1)=9.06$, $p<.005$). Closure durations of these stops are longer in the AP-INTERNAL condition than in the AP-INITIAL condition. The models built with LENIS show a significant interaction term (CD1: $\chi^2(1)=12.34$, $p<.001$; CD2: $\chi^2(1)=13.41$, $p<.0005$), which is largely due to the opposite directionality of Phrase Position effect. Unlike aspirated or fortis stops, lenis stops are produced with significantly *shorter* closure duration in AP-INTERNAL position than in AP-INITIAL position. Figure 1 visually summarizes the results.

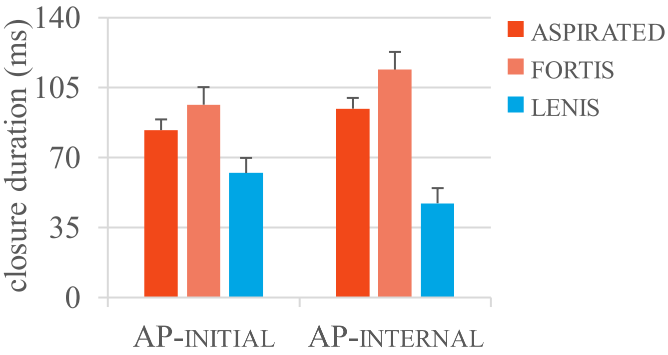

Figure 1. Stop Item and Phrase Position effects on stop closure duration (IP-medial conditions only; pooled across speakers & Syllable Number conditions).

2.3.2.2. VOT

The only significant main effect of Stop Item on VOT is found with the ASPIRATED and FORTIS pair (VOT3: $\chi^2(1)=8.74$, $p<.005$). VOT is longer for aspirated stops than for fortis stops. The only significant main effect of Phrase Position is found with ASPIRATED and LENIS pair (VOT1: $\chi^2(1)=9.82$, $p<.005$). For both stops, their VOT values are longer in the AP-INITIAL condition than in the AP-INTERNAL condition.

For all three comparisons, there is a significant interaction between predictors (VOT1: $\chi^2(1)=13.12$, $p<.001$, VOT2: $\chi^2(1)=6.84$, $p<.01$, VOT3: $\chi^2(1)=16.88$, $p<.001$). Post-hoc regressions reveal that VOTs for aspirated and lenis stops are not statistically different from each other in the AP-INITIAL condition (mean diff. 2 ms), confirming the initial VOT merger of these two categories. Note also that AP-*internal* lenis stops (tokens that do not undergo intervocalic voicing) and AP-internal fortis stops show similar mean VOT values (18 ms vs. 14 ms, respectively) although this difference is significant. Figure 2 shows the effects of Stop Item x Phrase Position on VOT.

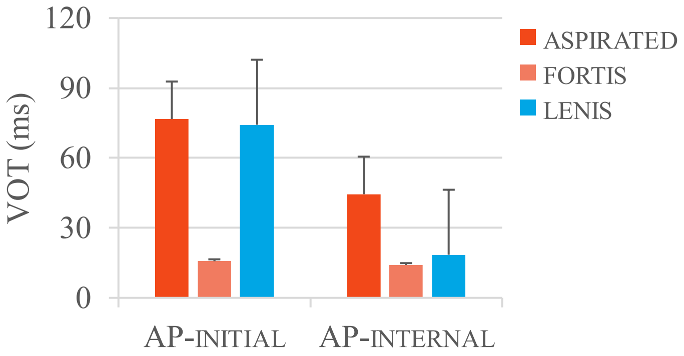

Figure 2. Stop Item and Phrase Position effects on VOT (pooled across speakers, Prosodic Boundary & Syllable Number conditions).

For all regression models, there is no effect of Prosodic Boundary (IP-INITIAL vs. IP-MEDIAL) at all, indicating that VOT values among stop consonants are not further modulated by different boundary strengths (i.e. IP-initial versus AP-initial [and IP-medial]).

2.3.3. F0 measures

To confirm that AP-initial aspirated and fortis stops actually pattern together, and AP-initial lenis and nasals stops pattern together, and further test whether the f0 patterning of these consonants continues in the AP-internal condition, we first present results of Consonant Item (ASPIRATED vs. FORTIS vs. LENIS vs. NASAL) and Phrase Position (AP-INITIAL vs. AP-INTERNAL) effects on f0 in the target vowel (v1 or v2 of an AP) (§2.3.3.1). In §2.3.3.2, a histogram analysis of f0 is presented. The histograms show distributions of the target vowel f0 values in the *z*-standardized f0 space in different Phrase Position conditions. In the subsequent subsections (§2.3.3.3-§2.3.3.4), we demonstrate how f0 values associated with the LAX (nasal and lenis stops) and TENSE (aspirated and fortis stops) categories are affected by other critical factors, Prosodic Boundary, Phrase Position, and Syllable Number.

2.3.3.1. F0 after different stops (pooled across Prosodic Boundary & Syllable Number conditions)

To assist readers' comprehension, the overall f0 patterning throughout an AP in different conditions is illustrated in Figure 3. Note that the statistical results of individual consonants reported here do not include Prosodic Boundary or Syllable Number effects, as largely similar effects are also observed with the grouped analysis results (TENSE vs. LAX categories) and therefore discussed in depth later in §2.3.3.3-§2.3.3.4. Linear mixed effects models constructed for the f0 in the target vowel (v1 or v2) had Phrase Position (AP-INITIAL vs. AP-INTERNAL) and Consonant Item (ASPIRATED vs. FORTIS vs. LENIS vs. NASAL) as predictors. As the Consonant Item predictor had four levels, 6 separate regression models with different pairwise comparisons were built ($f0_1$: ASPIRATED vs. FORTIS, $f0_2$: ASPIRATED vs. LENIS, $f0_3$: ASPIRATED vs. NASAL, $f0_4$: FORTIS vs. LENIS, $f0_5$: FORTIS vs. NASAL, $f0_6$: LENIS vs. NASAL).

For the LENIS and NASAL pair, there is no effect of Consonant Item, confirming that f0 values in the target vowels do not differ between lenis and nasal consonant conditions. This can be visually confirmed in Figure 3. The blue (LENIS) and light blue (NASAL) lines are almost completely overlaid on top of each other except one point (AP-INITIAL condition in 3-SYLL APs) in which the f0 value is slightly higher for LENIS compared to NASAL (*3a* in Figure 3). A significant effect of Phrase Position on the target f0 reveals that for both consonants, f0 values are higher in the AP-INTERNAL condition (v2) than in the AP-INITIAL condition (v1) ($\chi^2(1)=16.29$, $p<0001$, mean diff. 15 Hz).

Moreover, both lenis and nasal stops behave similarly in their relation to aspirated and fortis stops. In $f0_2$ (ASPIRATED vs. LENIS) and $f0_3$ (ASPIRATED vs. NASAL) models, there is a significant effect of Consonant Item ($f0_2$: $\chi^2(1)=4.03$, $p<05$, $f0_3$: $\chi^2(1)=4.11$, $p<05$). The target f0 values are higher for aspirated consonants compared to lenis (mean diff. 43 Hz) or nasal consonants (mean diff. 47 Hz). Although there is no main effect of Phrase Position in these models, the predictors interact with each other ($f0_2$: $\chi^2(1)=18.42$, $p<.0001$, $f0_3$: $\chi^2(1)=18.28$, $p<.0001$). Post-hoc analyses show that the Consonant Item effect is larger in the AP-INITIAL condition (*ASPIRATED > LENIS: mean diff. 74 Hz; *ASPIRATED > NASAL: mean diff. 81 Hz) than in the AP-INTERNAL condition (*ASPIRATED > LENIS: mean diff. 15 Hz; *ASPIRATED > NASAL: mean diff. 17 Hz).

In $f0_4$ (FORTIS vs. LENIS) and $f0_5$ (FORTIS vs. NASAL) models, there is no effect of Consonant Item or Phrase Position. However, there is a significant interaction term between predictors ($f0_4$: $\chi^2(1)=16.25$, $p<.0001$, $f0_5$: $\chi^2(1)=16.09$, $p<.0001$). This interaction also comes from the fact that the Consonant Item effect is larger in the AP-INITIAL condition (*FORTIS > LENIS: mean diff. 48 Hz; *FORTIS > NASAL: mean diff. 54 Hz) than in AP-internal position (*FORTIS > LENIS: mean diff. 8 Hz; *FORTIS > NASAL: mean diff. 10 Hz).

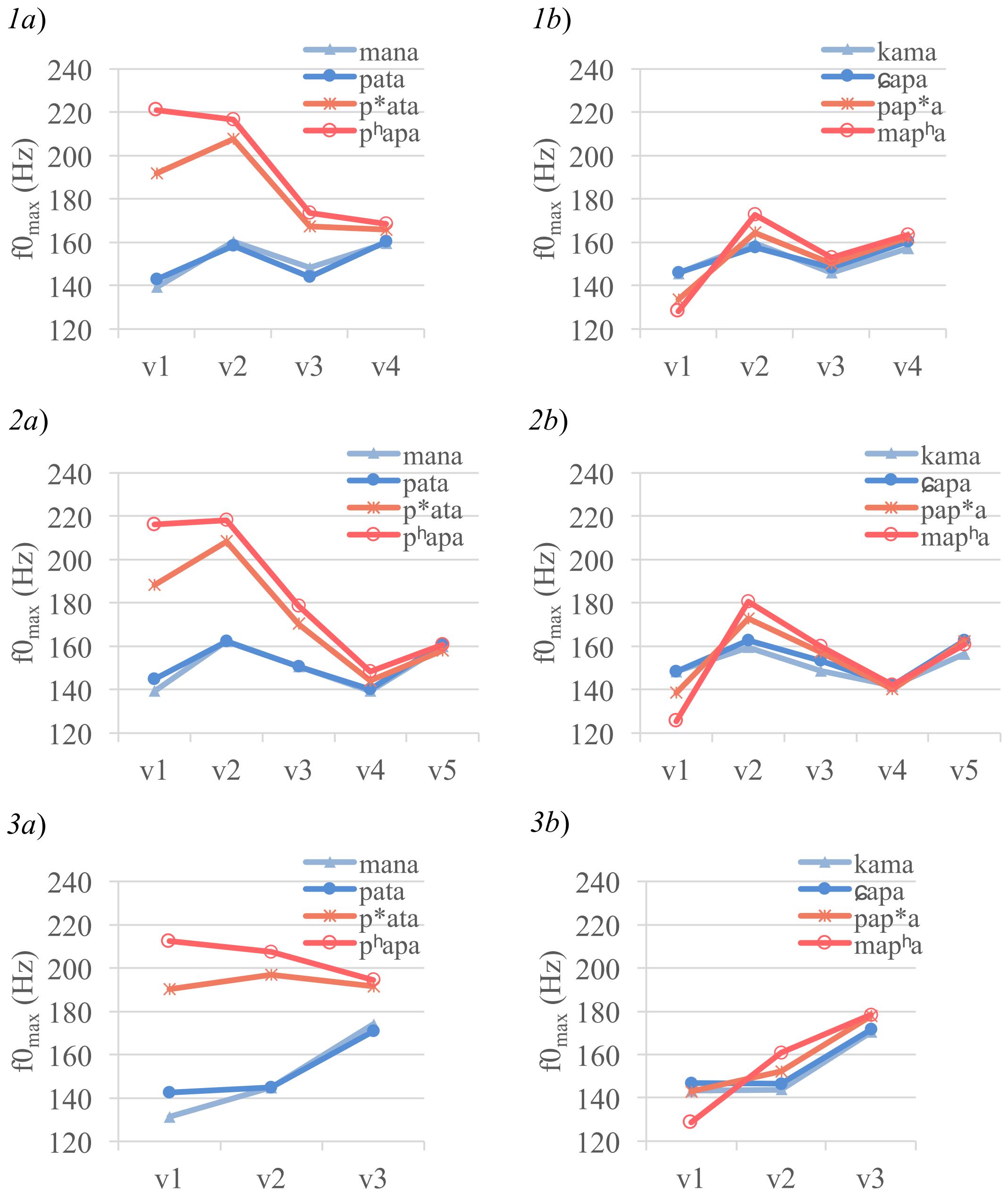

Figure 3. Mean values of $f0_{max}$ during each vowel taken from *1a*) AP-INITIAL condition in 4-SYLL APs *1b*) AP-INTERNAL condition in 4-SYLL APs, *2a*) AP-INITIAL condition in 5-SYLL APs, *2b*) AP-INTERNAL condition in 5-SYLL APs, *3a*) AP-INITIAL condition in 3-SYLL APs, *3b*) AP-INTERNAL condition in 3-syllable APs (pooled across speakers & prosodic boundaries). Different colors indicate different Consonant Item conditions (ASPIRATED /p^h/ in red, FORTIS /p*/ in orange, LENIS /p/ in blue, NASAL /m/ in light blue). Note that lines between one mean value (e.g., v1) to next (e.g., v2) are placed simply to illustrate the global pitch fluctuation of the whole AP, not to demonstrate any interpolating value.

Lastly, for the ASPIRATED and FORTIS pair, there is a significant effect of Phrase Position ($\chi^2(1)=10.73$, $p<.005$) and Consonant Item ($\chi^2(1)=5.28$, $p<.05$). The target f0 values are higher for aspirated stops than for fortis stops (mean diff. 16 Hz), as seen in Figure 3. As was the case with other consonants, phrase-initial aspirated and fortis stops are produced with higher f0 values when compared to phrase-internal ones (*initial aspirated > internal aspirated, mean diff. 46 Hz; *initial fortis > internal fortis, mean diff. 27 Hz). Phrase Position further interacts with Consonant Item ($\chi^2(1)=7.69$, $p<.01$). As shown in Figure 3, this interaction is attributable to the fact that the f0 difference coming from different Consonant Item conditions is larger in the AP-INITIAL condition (*ASPIRATED > FORTIS, mean diff. 27 Hz) than in the AP-INTERNAL condition (*ASPIRATED > FORTIS, mean diff. 7 Hz).

In sum, the results confirm that nasal and lenis pattern together with respect to f0 and that their f0 values are significantly lower than aspirated and fortis stops. The results also show that f0 values are higher for aspirated stops than for fortis stops. However, there seems to be no direct evidence of whether these two stops pattern together. The f0 distribution analysis presented in the following subsection will provide evidence for this.

2.3.3.2. F0 distribution (histogram analysis)

For the histogram analysis, each speaker's target f0 values were *z*-transformed to minimize variations coming from individual or gender differences. Then, histogram bin sizes were determined by square-root choice, which takes the square root of the number of observed points. Figure 4 shows the distribution of our speakers' *z*-standardized values of maximum f0 during the target vowel in different phrase positions (v1 or v2 in an AP).

In line with no significant f0 difference reported in the previous subsection, f0 values associated with nasal and lenis stops involve one mode in the distribution (blue bars). Although there is a significant f0 difference between aspirated and fortis stops (*ASPIRATED > FORTIS), these stop consonants form a single mode (red bars), which further suggests they indeed pattern together in terms of f0.

In the AP-INITIAL condition, the distribution is bi-modal (Figure 4 left panel). F0 values are clustered in a high range for TENSE (aspirated and fortis) consonants and clustered in a low range for LAX (lenis and nasal) consonants. In contrast, this type of f0 clustering is *not* observed in the AP-INTERNAL condition, as indicated by the overlapping, unimodal distribution of the different stop categories.

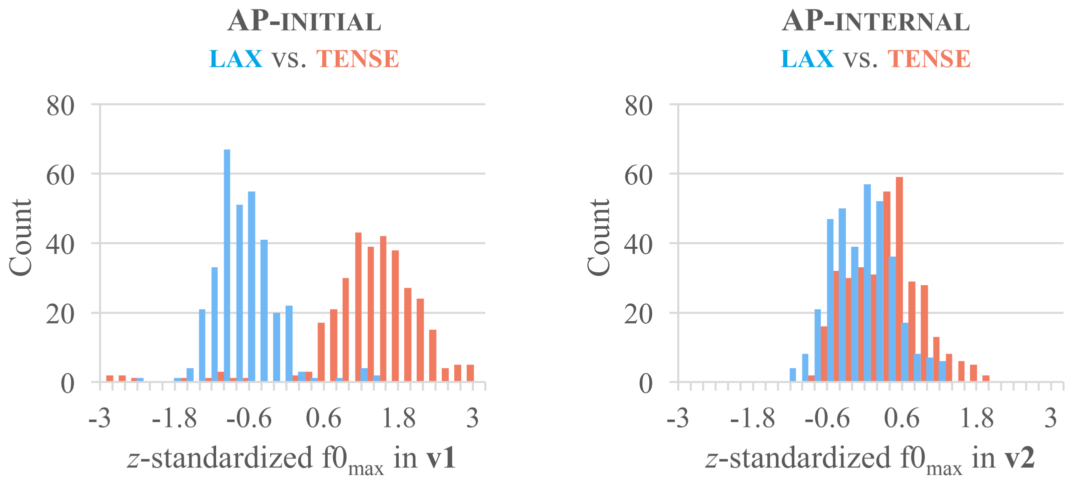

Figure 4. *Z*-standardized values of maximum f0 during the target vowel (v1 or v2) in AP-INITIAL (left panel) and in AP-INTERNAL (right panel) positions (pooled across speakers, syllable numbers & prosodic boundaries). LAX (LENIS /p/ and NASAL /m/ pooled) and TENSE (ASPIRATED /p^h/ and FORTIS /p*/ pooled) consonants are displayed in different colors, blue and red, respectively.

2.3.3.3. Overview of phonological context effects on f0

Figure 5 shows the mean values and 95 % confidence bands of maximum f0 during vowels in APs with different syllable numbers (3-SYLL vs. 4-SYLL vs. 5-SYLL APs). Other than expected differences in absolute values, there is no different patterning of f0 as a function of gender, speaker, or prosodic boundary strength. The figure represents the average values of f0 pooled across these factors.

Overall, from inspection of the graphs, there are systematic patterns of f0 values in the AP. The canonical *L*HLH pattern often transcribed in previous studies is observed in 4-syllable APs that start with a lax consonant (refer to the blue lines in AP-initial lax condition, *1a*, and both red and blue lines in AP-internal condition, *2a* in Figure 5). A similar pattern is observed with 5-syllable APs (*2b*, *3b*), but this time f0 values of v3 show an intermediate value between v2 and v4, rather than differing systematically from its neighbors.

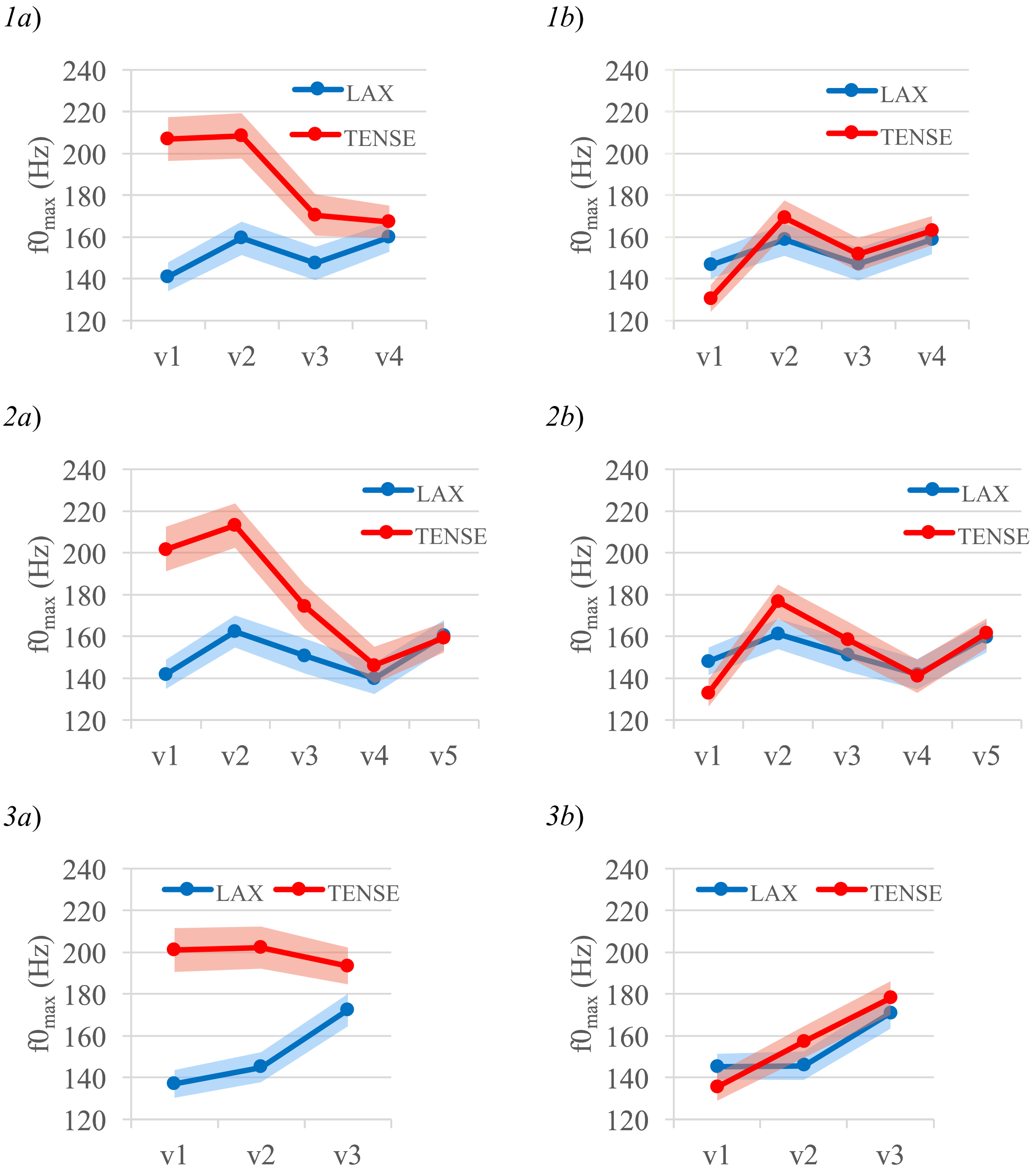

Figure 5. Mean values (presented as red and blue dots) and 95 % confidence bands of $f0_{max}$ during each vowel taken from *1a*) AP-INITIAL condition in 4-SYLLABLE APs *1b*) AP-INTERNAL condition in 4-SYLL APs, *2a*) AP-INITIAL condition in 5-SYLL APs, *2b*) AP-INTERNAL condition in 5-SYLL APs, *3a*) AP-INITIAL condition in 3-SYLL APs and *3b*) AP-INTERNAL condition in 3-SYLL APs (pooled across speakers & prosodic boundaries). The blue color scheme indicates the condition, in which the target consonant is a LAX category (NASAL /m/ & LENIS /p/ pooled together); the red color scheme indicates the TENSE category (ASPIRATED /p^h/ & FORTIS /p*/ pooled). Note that lines between one mean value (e.g., v1) to next (e.g., v2) are placed to simply illustrate the global pitch fluctuation of the whole AP, not to demonstrate any interpolating value.

The graphs show a clear tonal difference between tense versus lax categories: overall, higher f0 values are associated with TENSE than with LAX, whether AP-initial (v1s of APs in the left panel) or AP-internal (v2s of APs in the right panel). As illustrated in the figures (Figure 5 *1a*, *2a*, *3a*), when a phrase starts (AP-INITIAL condition) with a syllable that has a TENSE target consonant, f0 values are much higher than in the LAX condition. For the AP-INTERNAL condition, the confidence bands for the f0 values between TENSE and LAX of the second vowel (v2) are overlapping, indicating that the consonantally triggered f0 difference in the AP-INTERNAL condition is marginal.

The f0 value of the preceding syllable in the phrase seems to be a major determinant of the f0 value of the current syllable. In the case of the phrase-initial syllable, where there is no preceding syllable to refer to, the f0 values are largely dependent on the consonant type; higher f0 values for the tense category versus lower f0 values for the lax category. F0 values of the second vowel, especially in tense-initial APs, unlike f0 values of the later, non-adjacent vowels, are largely comparable to those of the preceding vowel (Figure 5 left panel). When an AP starts with a lax consonant, the overall f0 values for that phrase appear to be set in a speaker's low or default register/range. In contrast, when an AP starts with a tense consonant, it starts with a substantially high tone, and this high tone seems to be sustained over two syllables, or even a bit increased for the second vowel. In any case, after the first two syllables, the overall pitch starts falling towards the final syllable. Regardless of the starting point (which register the phrase resides in), f0 values of the penultimate syllable are dropped down to the lower register of the speaker if there are enough tone bearing units (setting aside the 3-syllable condition at the moment). This is the case for all 4- and 5-syllable APs (*1a*-*2b*), except for 4-syllable APs starting with tense consonants (red line in Figure 5, *1a*). The f0 values for the penult (v3) in the 4-syllable AP-INITIAL TENSE condition are not quite as low as the values for the other penults in the other conditions.

Notably, there is a clear difference in f0 excursion between v1 and v2 (calculated as $|f0_{v1}–f0_{v2}|$) in the AP-INTERNAL condition (Figure 5: *1b*, *2b*, *3b*). The f0 of the preceding syllable (v1) is lower when the second syllable is TENSE compared to LAX. This effect is not observed in the AP-INITIAL condition (Figure 5: *1a*, *2a*, *3a*).

The most noticeable f0 difference in terms of the number of syllables in the phrase is found with the phrase-final f0. In 4- or 5-syllable APs, the invariant phrase-final tone is observed at 160 Hz, which is an average value of 6 speakers (individual patterning confirmed across speakers; each speaker has an invariable value for his or her AP-final tone). For the 3-syllable APs, the phrase-final f0 is not at the value observed with APs with more syllables. Instead, the final f0 values seem to be influenced by the preceding syllables' f0 values—i.e., higher f0 for TENSE-initial APs than for LAX-initial APs.

These descriptive observations are statistically confirmed and further analyzed in the following subsection.

2.3.3.4. F0 analyses

For the f0 models, the dependent variable was the f0 value for each vowel: (A) f0 in the target vowel (AP-INITIAL, v1 or AP-INTERNAL, v2) whose Consonant Type is manipulated, (B) f0 in the phrase-medial (intermediate) vowel(s) (v3 of 4-syllable APs vs. v3 or v4 of 5-syllable APs), and (C) f0 in the final vowel (v3 of 3-syllable APs vs. v4 of 4-syllable APs vs. v5 of 5-syllable APs). In addition, (D) the f0 difference between v1 and v2 was used as the dependent variable for the f0 excursion model. Predictors included were Syllable Number (3-SYLL vs. 4-SYLL vs. 5-SYLL APs), Prosodic Boundary (IP-INITIAL vs. IP-MEDIAL), Phrase Position (AP-INITIAL vs. AP-INTERNAL), and Consonant Type (TENSE vs. LAX). As the Syllable Number predictor had three levels, 3 separate regression models were built ($f0_1$: 3-SYLL vs. 4-SYLL; $f0_2$: 3-SYLL vs. 5-SYLL; $f0_3$: 4-SYLL vs. 5-SYLL). The results tables of the f0 regression models, except phrase-medial f0 models, can be found in Appendix 3Appendix 4Appendix 5.

A. Target vowel f0 measures

For all three f0 models, there is a significant main effect of Consonant Type ($f0_1$: $\chi^2(1)=7.34, p<.01$, $f0_2$: $\chi^2(1)=7.9, p<.01$, $f0_3$: $\chi^2(1)=8.02, p<.01$). Higher f0 values are associated with tense consonants than with lax consonants (*TENSE > LAX, mean diff. 37 Hz for all three models). There is no main effect of Phrase Position. However, there is a significant interaction between Phrase Position and Consonant Type ($f0_1$: $\chi^2(1)=14.92, p<.001$, $f0_2$: $\chi^2(1)=12.11$, $p<.001$, $f0_3$: $\chi^2(1)=13.81, p<.001$). Post-hoc regressions reveal that this interaction stems from two facts. The effect of Consonant Type on f0 is greater in the AP-INITIAL condition (*TENSE > LAX, $f0_1$: mean diff. 65 Hz, $f0_2$: mean diff. 62 Hz, $f0_3$: mean diff. 63 Hz) than in AP-INTERNAL position (*TENSE > LAX, $f0_1$: mean diff. 11 Hz, $f0_2$: mean diff. 14 Hz, $f0_3$: mean diff. 13 Hz). This interaction is also attributable to the opposite direction of Phrase Position effects found in TENSE versus LAX conditions. While f0 values in TENSE are higher in the AP-INITIAL condition than in the AP-INTERNAL condition, f0 values in LAX are higher in the AP-INTERNAL condition than in the AP-INITIAL condition. This can be confirmed in Figure 5; compare figures on the left side (*1a*, *2a*, *3a*) with figures on the right side (*1b*, *2b*, *3b*).

A main effect of Syllable Number is found with two models ($f0_1$: $\chi^2(1)=43.86$, $f0_2$: $\chi^2(1)=62.62$, $p<.001$ for both). F0 values are generally higher in APs with 4- or 5-syllables than in APs with 3 syllables. However, Syllable Number shows a significant interaction with Phrase Position ($f0_1$: $\chi^2(1)=24.15$, $p<.0001$, $f0_2$: $\chi^2(1)=44.91$, $p<.0001$). Post-hoc analyses reveal that the effect of Syllable Number is observed only in the AP-INTERNAL condition: the second vowel of 3-syllable APs is produced with lower f0 values than in 4- or 5-syllable APs (*3-SYLL < 4-SYLL, mean diff. 13 Hz; *3-SYLL < 5-SYLL, mean diff. 18 Hz). This can be visually confirmed in Figure 5; compare *3b* with *1b* and *2b*, but note that this 2-way interaction term is defined with both consonant types pooled together. There is no further significant interaction term with this predictor.

Prosodic Boundary shows a main effect on f0 (*IP-INITIAL > IP-MEDIAL, mean diff. 6-7 Hz, $f0_1$: $\chi^2(1)=21.12$, $p<.0001$, $f0_2$: $\chi^2(1)=15.88$, $p<.05$, $f0_3$: $\chi^2(1)=44.1$, $p<.0001$). As this study mainly focuses on VOT and f0 modulation as a function of Consonant Type, an emphasis should be placed on a significant interaction term of one predictor with Consonant Type, not its main effect. That said, for all regression models, Prosodic Boundary does not interact with Consonant Type at all.

For all f0 models, there is no significant 3- or 4-way interaction between predictors.

B. F0 in phrase-medial vowels (4-SYLL & 5-SYLL APs only)

As shown in the global f0 patterns in Figure 5, the f0 in the phrase-medial vowels seems to be largely similar to the f0 of the preceding syllable in which it is determined by the consonant quality. To test how f0 in the phrase-medial vowel of 4-syll and 5-syllable APs is affected by the critical factors, f0 values in v3s were compared (penult in v1v2**v3**v4 vs. antepenult v1v2**v3**v4v5), and f0 values in v3 of 4-syllable APs were compared with those in v4 of 5-syllable APs (penults in v1v2**v3**v4 vs. v1v2v3**v4**v5). The predictors included were

Consonant Type, Phrase Position, and Syllable Number. Again, we pay close attention to the between-predictor interaction terms.

For the v3 f0 comparison, all three predictors show a significant effect (Syllable Number: $\chi^2(1)=11.53$, $p<.001$; Consonant Type: $\chi^2(1)=9.31$, $p<.005$; Phrase Position: $\chi^2(1)=4.47$, $p<.05$). In general, f0 values of v3 are higher in 5-syllable APs than in 4-syllable APs (*antepenult in 5-SYLL AP > penult in 4-SYLL AP, mean diff. 5 Hz), higher in the AP-INITIAL condition than in the AP-INTERNAL condition (*after the initial consonants > after the second consonants, diff. 9 Hz), and higher in the TENSE consonant condition than in the LAX consonant condition (*tense-following > lax-following, mean diff. 15 Hz). Syllable Number does not interact with the other predictors. However, there is a significant Consonant Type by Phrase Position interaction ($\chi^2(1)=11.9$, $p<.0001$, Figure 6). The interaction is due in part to the fact that the effect of Phrase Position on v3's f0 is significant only in the TENSE condition (*after initial tense > after internal tense, mean diff. 17 Hz), not in the LAX condition. This interaction is also attributable to the fact that the f0 difference between v3 after the lax-syllable and v3 after the tense-syllable is larger when the target (tense or lax) syllable is in the AP-INITIAL condition (mean diff. 23 Hz) than in the AP-INTERNAL condition (mean diff. 6 Hz).

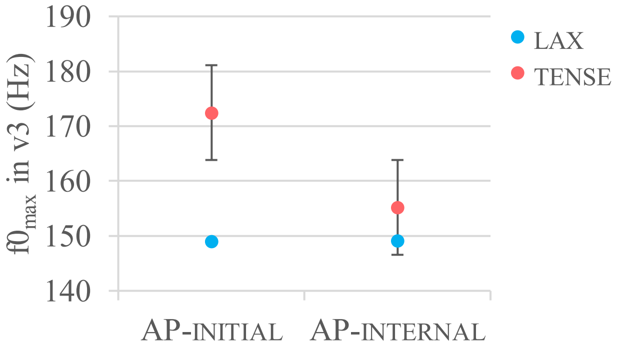

Figure 6. Consonant Type and Phrase Position effects on f0 in v3 (pooled across speakers & Syllable Number conditions).

There is no further interaction between predictors.

For the f0 comparison between penultimate vowels (v3 of 4-syllable APs vs. v4 of 5-syllable APs), again, all three factors show a significant main effect (Syllable Number: $\chi^2(1)=139.37$, $p<.0001$; Consonant Type: $\chi^2(1)=8.24$, $p<.005$; Phrase Position: $\chi^2(1)=4.81$,

$p<.05$). F0 values are higher in the 4-syllable APs than in the 5-syllable APs (mean diff. 12 Hz), higher in the TENSE condition than in the LAX condition (mean diff. 8 Hz), and higher in the AP-INITIAL condition than in the AP-INTERNAL condition (mean diff. 6 Hz). There is a significant interaction term between Consonant Type and Syllable Number ($\chi^2(1)=23.62$, $p<.0001$), which is due in part to the fact that the Consonant Type effect is significant only in v3 of 4-syllable APs (mean diff. 14 Hz), not in v4 of 5-syllable APs (mean diff. 3 Hz) (See Figure 7).

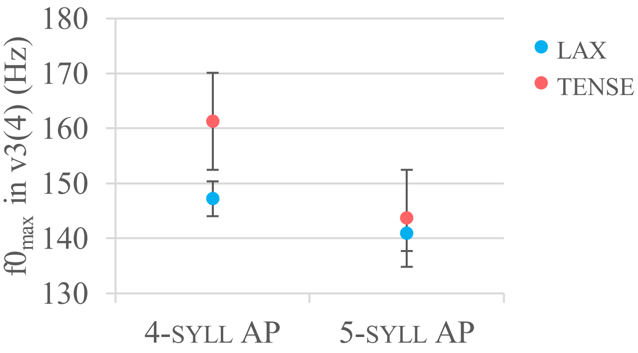

Figure 7. Consonant Type and Syllable Number effects on f0 in the penultimate vowels in v1v2**v3**v4 and v1v2v3**v4**v5 (pooled across speakers & Phrase Position conditions).

Consonant Type further interacts with Phrase Position ($\chi^2(1)=9.46$, $p<.005$). As was the case with the v3 comparison, this interaction is attributable to the fact that the Consonant Type effect is larger when the target consonant contrast is in AP-initial position (mean diff. 18 Hz) than in AP-internal position (mean diff. 6 Hz) (Figure 8).

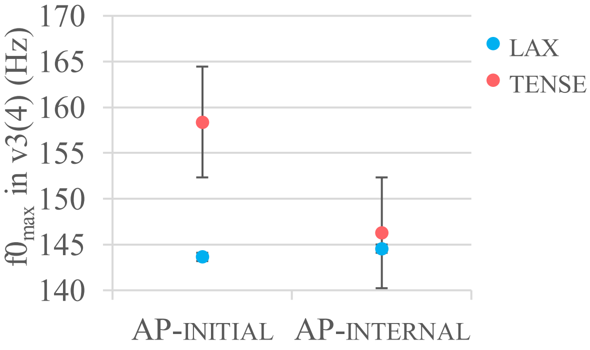

Figure 8. Consonant Type and Phrase Position effects on f0 in the penultimate vowels in v1v2**v3**v4 & v1v2v3**v4**v5 (pooled across speakers, 4-SYLL & 5-SYLL AP conditions).

Lastly, there is a 3-way interaction among predictors ($\chi^2(1)=8.88$, $p<.005$). Figure 9 illustrates this effect. There is a robust consonant-type effect on f0 in the AP-medial vowel when the target consonant is at the beginning of a 4-syllable AP (leftmost lax-tense contrast in the figure), and this initial effect is larger in 4-syllable APs than 5-syllable APs (compare the first and third lax-tense contrasts). Again, the post-hoc regressions show that the consonant effect on f0 is not significant any more in v4 of 5-syllable APs, when the target consonant contrast is in the AP-INTERNAL condition (rightmost lax-tense contrast in the figure).

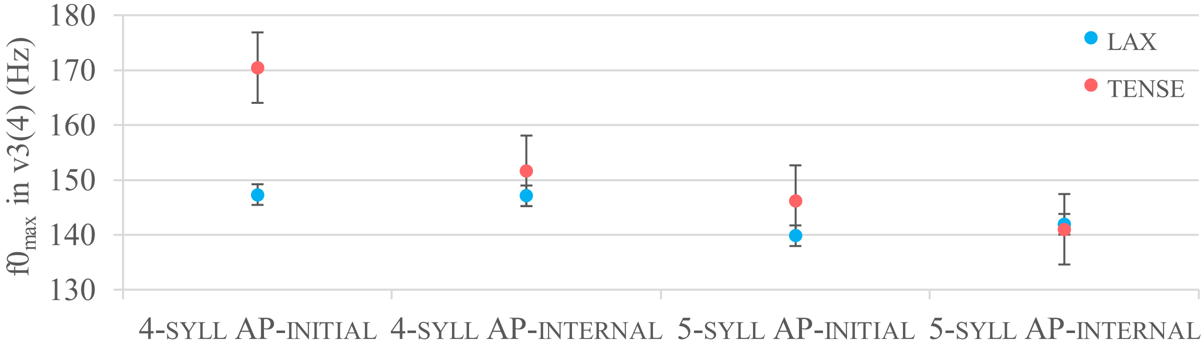

Figure 9. Consonant Type, Syllable Number and Phrase Position effects on f0 of the penultimate vowels in v1v2**v3**v4 and v1v2v3**v4**v5 (pooled across speakers & Prosodic Boundary conditions).

In sum, the results show that the tense-lax effect on the f0 of AP-medial vowels is larger when the target consonant is in the first syllable than in the second of an AP, and this effect can extend up to the third syllable of an AP, but not much farther.

C. Phrase-final f0

For the phrase-final f0, Syllable Number seems to play a main role. As shown in Figure 5, f0 values during the phrase-final vowels (v4 in 4-syllable APs, v5 in 5-syllable APs, v3 in 3-syllable APs) also vary by Consonant Type and Phrase Position. As was the case with the target f0 models, 3 separate models were constructed ($f0_1$: 3-SYLL vs. 4-SYLL, $f0_2$: 3-SYLL vs. 5-SYLL, $f0_3$: 4-SYLL vs. 5-SYLL) were constructed.

A significant main effect of Consonant Type is found with all three models ($f0_1$: $\chi^2(1)=7.34$, $p<.0005$; $f0_2$: $\chi^2(1)=11.15$, $p<.001$; $f0_3$: $\chi^2(1)=4.83$, $p<.05$). The final f0 is higher in the TENSE condition than in the LAX condition ($f0_1$: mean diff. 10 Hz, $f0_2$: mean diff. 7 Hz, $f0_3$: mean diff. 3 Hz). Two models show a significant main effect of Phrase Position ($f0_1$: $\chi^2(1)=6.82$, $p<.01$, $f0_2$: $\chi^2(1)=6.95$, $p<.01$). The final f0 is higher when the target tense-lax contrast is in the AP-INITIAL condition than in the AP-INTERNAL condition ($f0_1$: mean diff. 6 Hz, $f0_2$: mean diff. 4 Hz). In the 3-SYLL versus 4-SYLL pair, these two factors interact with each other ($\chi^2(1)= 8.06$, $p<.005$), showing a larger consonant type effect in the AP-INITIAL condition (mean diff. 10 Hz) than in the AP-INTERNAL condition (mean diff. 4 Hz).

There is a significant main effect of Syllable Number in $f0_1$ ($\chi^2(1)=176.61$, $p<.0001$) and $f0_2$ ($\chi^2(1)=217.07$, $p<.0001$), but no effect in $f0_3$: 3-syllable AP-final tones are higher than 4- or 5-syllable AP-final tones (*3-SYLL > 4-SYLL: mean diff. 16 Hz, *3-SYLL > 5-SYLL: mean diff. 18 Hz); the final f0 values in most conditions of 4- and 5-syllable APs are not significantly different from each other; see Figure 5. All three models show a significant interaction between Syllable Number and Consonant Type ($f0_1$: $\chi^2(1)=13.21$, $p<.0005$, $f0_2$: $\chi^2(1)=40.31$, $p<.0005$, $f0_3$: $\chi^2(1)=7.59$, $p<.01$). This interaction is due mainly to the fact that the effect of the tense-lax difference is larger in 3-syllable APs (mean diff. 10 Hz) than in 4-sllable (mean diff. 6 Hz) or 5-syllable APs (mean diff. <1). See Figure 10.

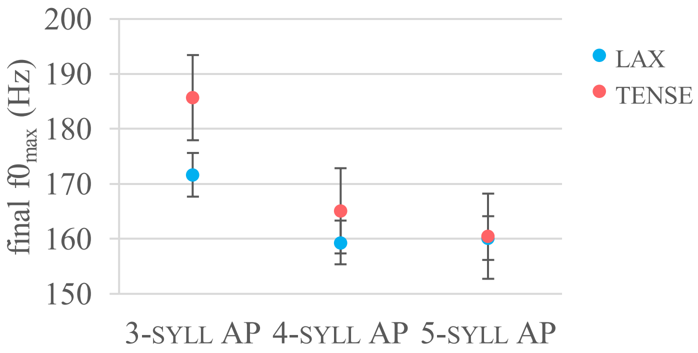

Figure 10. Consonant Type and Syllable Number effects on the AP-final f0 (pooled across speakers, Phrase Position & Prosodic Boundary conditions).

For $f0_1$ and $f0_2$ models (3-SYLL vs. 4- or 5-SYLL), there is a significant 3-way interaction between predictor factors ($f0_1$: $\chi^2(1)=4.96$, $p<.05$, $f0_2$: $\chi^2(1)=12.7$, $p<.0005$). This interaction is due mainly to the fact that the effect of Consonant Type that is larger in the AP-INITIAL condition than in the AP-INTERNAL condition and decreases incrementally from APs with

3 syllables to APs with more syllables. The tense-lax effect on the final f0 disappears in 5-syllable APs. These patterns are shown in Figure 11.

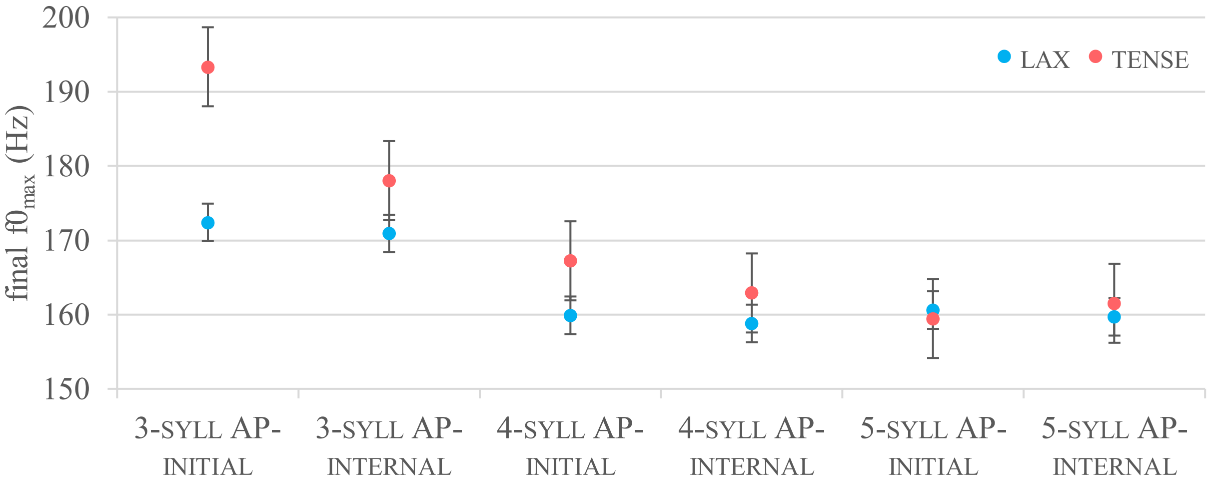

Figure 11. Consonant Type, Syllable Number and Phrase Position effects on the phrase-final f0 (pooled across speakers & Prosodic Boundary conditions).

D. F0 excursion (difference between f0 values in v1 and v2)

Recall that the f0 excursion was calculated as the difference between f0 in v1 and f0 in v2. The f0 excursion model was built based on the results from target syllable f0 models in (A). As there was no meaningful effect of Prosodic Boundary or Syllable Number on the target f0 values, the f0 excursion model only contained Phrase Position and Consonant Type predictors.

Results indicate that although there are no main effects of the predictors, there is a significant interaction ($\chi^2(1)=9.5$, $p<.005$) (Figure 12). The interaction is partly due to the unequal effect of Phrase Position in different Consonant Type conditions. The f0 excursion values do not change significantly within the lax category (mean diff. <1 Hz). However, there is a significant effect of Phrase Position in the TENSE condition (*AP-INITIAL < AP-MEDIAL, mean diff. 22 Hz). This interaction is also attributable to the opposite direction of the Consonant Type effect in different Phrase Position conditions. In the AP-INITIAL condition, the f0 excursion is smaller in the TENSE condition than in the LAX condition (mean diff. 6 Hz). However, the f0 excursion in the AP-INTERNAL condition is larger for TENSE than for LAX, (mean diff. 16 Hz). Compare this result with the Consonant Type effect on the target vowel f0 that was smaller AP-*internally* than AP-initially.

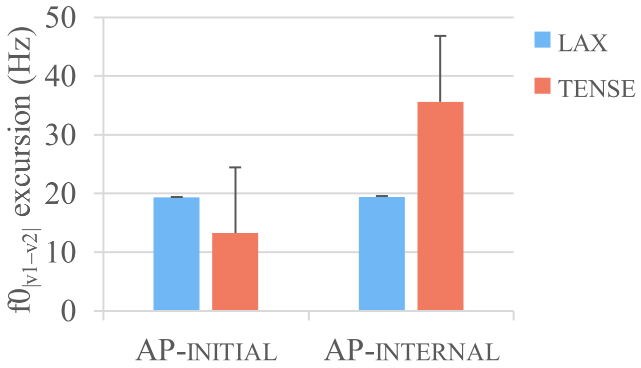

Figure 12. Consonant Type and Phrase Position effects on $f0_{|v1-v2|}$ excursion (pooled across speakers, Prosodic Boundary & Syllable Number conditions).

2.3.3.5. Information organization of the 3-way stop contrast in individual speakers

This subsection visualizes the findings of the phonetic information organization in the 3-way stop contrast in individual speakers. Figure 13 shows each speaker's f0 and VOT distributions of the three-stop contrast (LENIS VS. ASPIRATED VS. FORTIS) in different Phrase Position conditions (AP-INITIAL VS. AP-INTERNAL).

Overall, the three initial stops are distinct from each other in terms of their f0 x VOT combinations (columns 1, 3 in Figure 13). For all speakers, there are three separate clusters in the f0 and VOT space. Nevertheless, there exists inter-speaker variability in terms of how close these clusters are. Speaker F2 shows the classic reported tonogenic pattern with clearly separated, non-overlapping stop clusters: there is a VOT merger between aspirated and lenis stops and a clear 2-way tonal distinction between tense and lax. Speakers F1 and F3 show a weaker case of the initial VOT merger, resulting in a rather crowded f0 x VOT space. The aspirated and fortis clusters are formed immediately adjacent to each other, and VOT values of the lenis stops seem to partially overlap with either the fortis (Speaker F1) or the aspirated (Speaker F3) stops. The f0 distinction between tense and lax is yet clearly made. Speaker M1 is the only speaker who does not show the initial VOT merger or the accompanying tonal distinction. Rather, this speaker shows a clear 3-way distinction in both f0 and VOT, mainly heightening the aspiration of the aspirated stop category. Among the speakers, this speaker shows the longest VOT values for the aspirated stops in both positions. F0 is high for the aspirated stops, intermediate for the fortis stops, and low for the lenis stops. VOT is long for the aspirated, intermediate for the lenis, and short for the fortis stops. Speaker M2 shows a 2-way tonal distinction, but accompanied by a 3-way VOT distinction. However, the VOT values of the lenis and aspirated are still partially overlapping. (NB: The two initial lenis tokens of this speaker that seem to belong to the aspirated stop cluster are IP-initial stops.) Both Speakers M1 and M2, who show a 3-way VOT distinction, show that the clusters of the three stops are fairly proximate in the f0 dimension. Speaker M3's patterns are similar to Speaker F2. This speaker also shows the initial VOT merger and 2-way tonal distinction, and the category clusters are clearly separated.

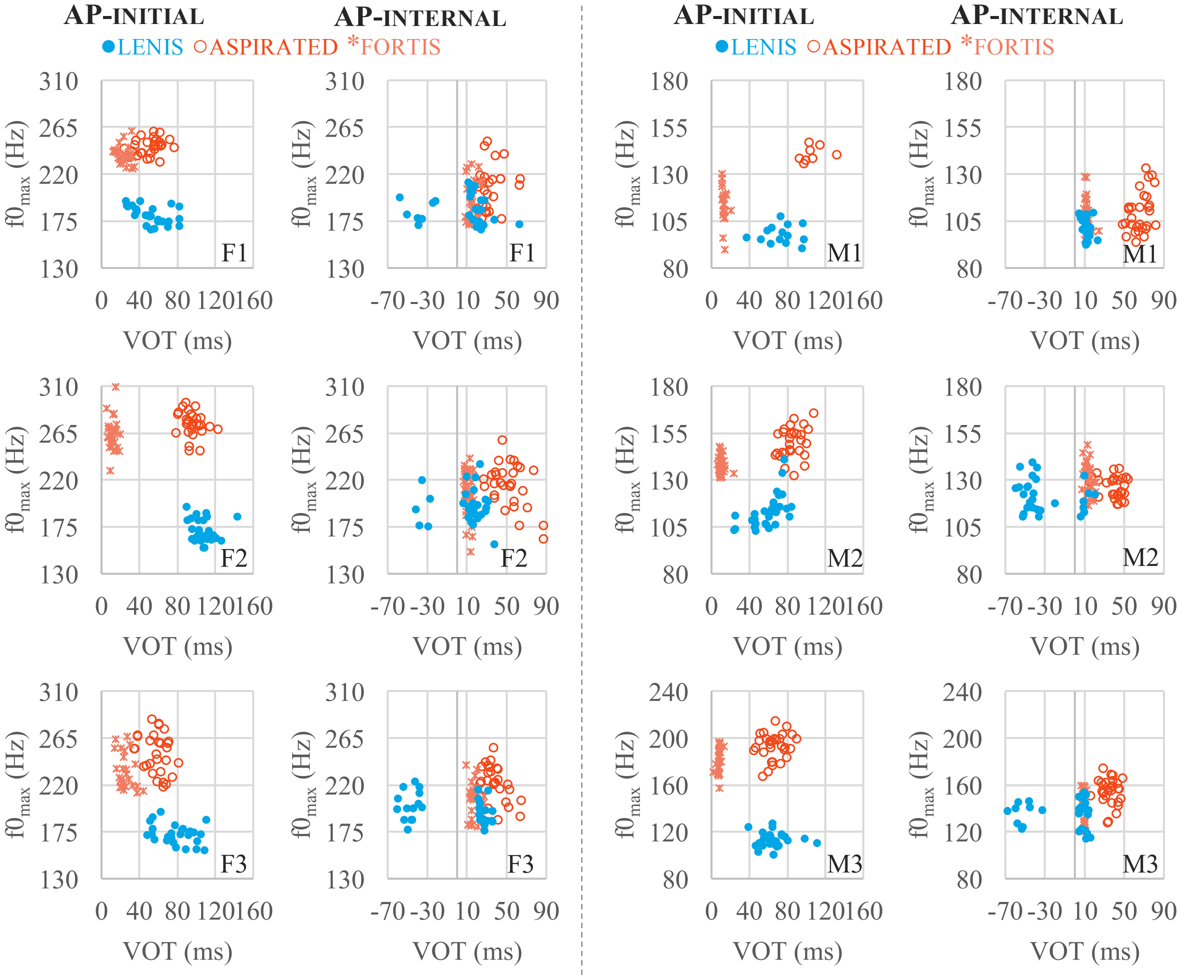

Figure 13. VOT and f0 distributions for the stop tokens in different phrase positions (AP-INITIAL vs. AP-INTERNAL) of individual speakers (F1, F2 F3, M1, M2, M3) (pooled across Syllable Number & Prosodic Boundary conditions). Red empty circles indicate aspirated stops; blue filled circles indicate lenis stops; orange asterisks indicate fortis stops. Note that the $f0_{max}$ axis scale is differently set between female and male speakers; Speaker M3's is set independently from the other two male speakers due to his higher f0 values.

In AP-internal position, unlike AP-initial stops, the three stops show largely overlapping stop clusters in terms of the f0 x VOT space (columns 2, 4 in Figure 13). However, all speakers, except Speakers F1 and F3, show that the aspirated stops do not overlap with the other stops in the f0 and VOT space, which is attributable to their longish VOT values. The most clear AP-internal VOT distinction between the aspirated and the other stops is made by Speaker M1, who does not show any voiced lenis tokens. For all speakers, the unvoiced intervocalic lenis tokens show overlapping VOT values with the AP-internal fortis stops, and the f0 values between tense and lax are largely overlapping. The least overlapping f0 values between tense and lax are observed in Speakers M1 and M2. Recall that the f0 histogram analysis in Figure 4 shows exactly these patterns: a non-overlapping bi-modal distribution for the AP-initial tense and lax consonants, and largely overlapping distributions of the AP-internal tense and lax.

The three AP-internal stops are not distinctively distributed in terms of the combination of VOT and f0. Figure 14 plots the AP-internal stops in terms of VOT and closure duration from each individual speaker. For all speakers, the stop categories form distinctive clusters. These clusters are clearly separated for Speakers F2, M1, and M3. Although there are some overlapping tokens of aspirated and fortis AP-internal stops for the other speakers, the unvoiced intervocalic lenis tokens never overlap with the other two, due to their particularly short closure durations.

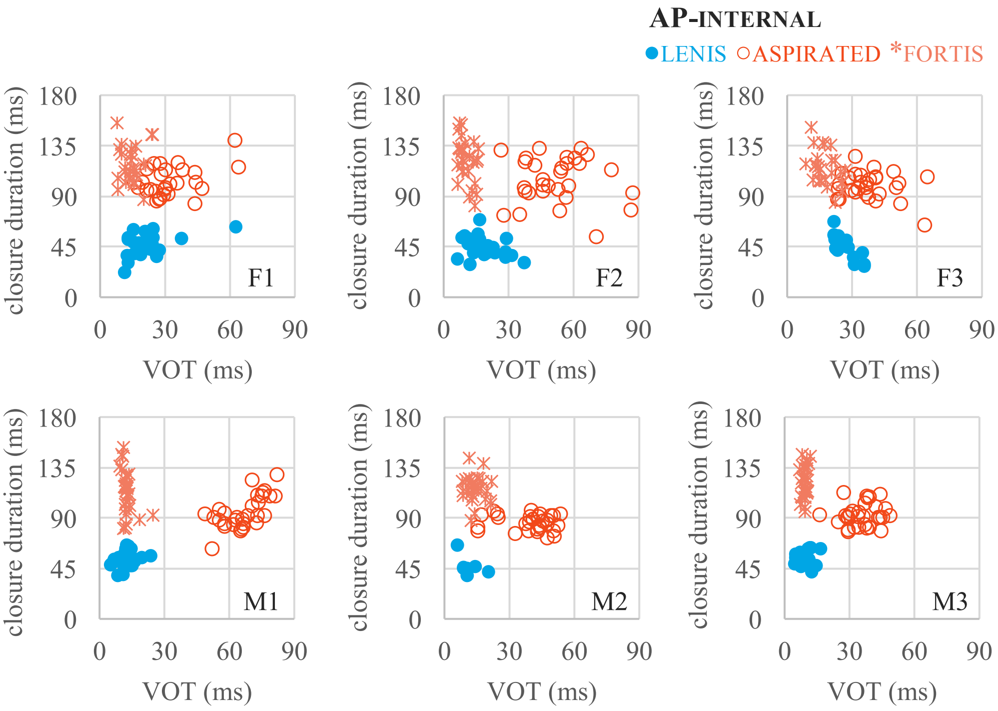

Figure 14. VOT and stop closure duration distributions for AP-INTERNAL stops of individual speakers (F1, F2, F3, M1, M2, M3) (pooled across Syllable Number & Prosodic Boundary conditions). Red empty circles indicate aspirated stops; blue filled circles indicate unvoiced lenis stops; orange asterisks indicate fortis stops.

2.4. Discussion

This study investigates how the phonetic information of consonant and tone is organized in speakers' production of contemporary Seoul Korean. As this language exhibits a phrasal level phonology—an Accentual Phrase (AP) that interacts with local (segmental and tonal) properties—we hypothesized an essential role of phrasal prosody in terms of the information organization associated with the newly emerging stop system. We found supporting evidence for this assumption. The observed patterns of consonant and tone interaction are systematically influenced by higher-level phrasal prosodic contexts. In what follows, we will discuss what phonological factors regulate Seoul Korean AP tones in the new system (§2.4.1) and then discuss how the local consonant contrast information plays out through an intricate interaction of segmental properties with phrasal prosody (§0). In Chapter 4, we will return in greater detail to the process of this tonogenic sound change.

2.4.1. Phonological factors shaping Accentual Phrase tones in Seoul Korean

The existing tonal model for Seoul Korean states that the underlying tonal sequence *T*HLH of an AP can surface as various tonal patterns, with more variations in APs with fewer syllables (Jun 1993, 1998, 2000). While various surface patterns may not be entirely predictable from the underlying tones for many reasons, our goal is to identify what factors shape AP tones in the new system of Korean stops. For the patterns observed in the APs tested in this study, readers are referred back to Figure 5.

Let us start with the results for APs of 4- and 5-syllables. Overall, the f0 patterns of these APs conform to canonical tonal patterns of a Seoul Korean AP, *T*HLH. The results show that the initial *T* (*L* or *H*) is determined by consonantal quality. The initial f0 of an AP is higher for tense (aspirated and fortis) consonants than for lax (lenis and nasal) consonants. Our results also indicate that the second syllable of an AP carries a phrasally defined underlying high tone as in *T***H**LH. The f0 is higher in the second syllable than in the initial tone. Exceptionally, when an AP starts with a tense-triggered *H* tone, the following tone is only marginally higher. This might arise from a ceiling effect, as the initial *H* tone already seems to be in speakers' high(est) f0 range. A medial low tone in the model (*T*H**L**H) is generally observed in the penultimate syllable of an AP. In a 5-syllable AP, the f0 of the antepenultimate syllable is intermediate between the values of the medial H and L tones, which has been viewed as tonal interpolation in the model (Jun 1993, 1996). Finally, the results indicate that there may be an invariant tonal target designated for the AP-final tone. In most cases, a high tone, rising from the previous low tone on the penultimate syllable, is employed to demarcate the end of the phrasal boundary. This is consistent with the previous report that the most common AP-final tone is a rising tone (e.g., Jun 1993). One case that does not show a rising tone is the 4-syllable AP starting with a tense consonant. The f0 of its penultimate syllable is largely similar to the final f0. Nevertheless, the f0 value of the phrase-final vowel is unquestionably similar across different AP conditions.

As expected, the tonal patterns of 3-syllable APs do not completely comply with the surface patterns observed in APs with more syllables. This surface variation in APs with fewer syllables has been accounted for by a possible tonal undershoot or truncation. That is, the phrase-medial tones of the underlying tonal sequence *T*HLH of an AP may be truncated in APs with fewer syllables than four (Cho and Flemming 2015; Jun 1993). Our results show that the f0 of the medial vowel (v2) of the 3-syllable AP does not behave similarly to the f0 of any of the medial vowels in 4- or 5- syllable APs, though the initial and final tones seem to behave similarly to the case of APs with more syllables and the same triggering effect of consonant type on f0 is observed in the initial syllable. Most 3-syllable APs also show an invariant tone at the phrase end, if a bit higher than in longer APs. As is the case with the 4-syllable AP, when the 3-syllable AP starts with a tense consonant, the final f0 is not higher than the preceding syllable's f0, but, again, largely similar to the f0 of the preceding syllable. There is no evidence of whether either of the underlying medial tones (HL) has surfaced or not, or evidence of which of the two gets truncated if tonal undershoot is occurring.

Based on our results, we propose that the underlying tonal shape of an AP is the alternating LH sequence. This seems to surface faithfully as *L*HLH, when a 4-syllable AP starts with a lax onset. However, factors including segmental makeup and number of syllables per AP further shape the surface tonal patterns in a systematic way. 3-syllable APs show the most deviate tonal patterns from the canonical *L*HLH pattern, but the observed patterns are still not totally random. This suggests that there exist phonological biases in the system that change the surface patterns both in magnitude and form. In the remainder of this section, we expound on these previously undiscussed biases.

The results show that the consonantly triggered f0 is manifested differently in different prosodic locations. In both AP-initial and AP-internal positions, there is a significant tonal difference between tense and lax categories. Higher f0 values are associated with tense stops than with lax stops. However, our results show an asymmetric effect arising in different phrasal positions. The consonantally induced f0 difference is substantially bigger in AP-initial position than in AP-internal position (cf. Cho and Lee 2016). Our study further assesses this asymmetry by looking at f0 distributions in each position. In AP-initial position, there is a non-overlapping bi-modal distribution of f0 values: a cluster in a high range for tense stops and another cluster in a low range for lax stops. An analysis of individual speakers' f0 and VOT space (Figure 13) indicates that the consonantally derived f0 clustering functions so as to maintain or enhance the three-way stop contrast. In contrast, this type of f0 clustering is *not observed at all* in the AP-internal condition, as indicated by the overlapping unimodal distribution of different stop categories. Taken together, our results suggest that the consonant type effect on f0 is *categorical* in AP-initial position but *quantitative* (gradient) in AP-internal position.

This positional asymmetry is also found in the temporal scope of the consonantally induced f0 difference. The temporal scope of the initial *T* was evaluated through the f0 comparisons made between the non-initial syllables in *L*-initial and *H*-initial APs. The results show that the consonantal effect on f0 has a broad scope AP-initially but a small scope AP-internally. In AP-initial position, *both* initial *H* and initial *L* of an AP (*H* or *L* in ***T***HLH) exert an influence on the following syllable. The initial *H* triggered by a tense stop is followed by high f0 values of the following syllables, up to the third syllable of an AP. This pattern is found across APs with different numbers of syllables. In other words, the first, second, and third syllables of 3-, 4- and 5-syllable APs are produced with higher f0 values when the initial syllable of their AP starts with a tense consonant than when the initial syllable starts with a lax consonant. The initial *L* triggered by a lax stop also seems to constrain the f0 values of the following syllable(s) (at least up to the second syllable). This is indicated by the fact that the *non*-initial (second syllable) high tone associated with a tense stop following the initial *L* (*L***H**LH) is never as high as either the initial *H* or the later H syllables (***H*****H**LH).

In AP-internal position, the scope of the consonantally derived tonal difference is limited to the target syllable and its preceding syllable at best. Non-initial position has been considered to be a tone-neutralizing context regardless of consonant types (e.g., the second syllable always carrying a high tone as in *T***H**LH). However, our results show a small but significant f0 difference on the AP-internal target syllable (*L***H**LH). In addition, the tonal difference between categories seems to be preserved through f0 excursion, calculated as the difference between f0 values of the AP-initial and -second syllables (***L*****H**LH). The f0 excursion is bigger when the target syllable starts with a tense consonant than when the target syllable starts with a lax consonant, which is achieved by lowering the f0 of the initial syllable.

As discussed in Jun (1996), the stabilization (phonologization) of the consonantally triggered f0 has been viewed as an AP-initial specific phenomenon, that is, the laryngeal contrast [tense] is enhanced in a prosodically salient position. Our finding is only partly consistent with this claim. A large, categorical effect of consonant type on the f0 is found in AP-initial position, but a small, gradient effect is also observed in AP-internal position. In accentual phrase-initial position, one might postulate that the presence of a robust laryngeal gesture results from the overlap in time with π-gesture (local clock slowing in the vicinity of prosodic juncture [Byrd and Saltzman 1998, 2003]) or μ-gesture (spatial and temporal modulation due to accentuation

[Saltzman, Nam, Krivokapić, and Goldstein 2008]). That said, the temporal scope results call for an additional explanation. It remains unclear why the scope of both initial *Tones*, *L* and *H*, spans over multiple syllables.

The positional asymmetry has a further implication for the finding that the f0 value of a *preceding* syllable is a major determinant of the f0 value of the current syllable. In the 'default' case of the phrase-initial syllable, where there is no preceding syllable to refer to, activation of the initial *H* versus *L* tone is solely controlled by whether the initial consonant in the syllable is lexically tense or lax (much higher f0 values after the phrase-initial tense versus lower f0 values after the phrase-initial lax). This initial f0 setup appears to substantially affect the subsequent syllables. Within an AP, the f0 of a non-initial syllable is shown to be largely similar to the f0 of its previous syllable. This suggests that the value of f0 is determined by the bias to become (or remain) similar to that of the preceding syllable. That said, in AP-internal position, small f0 differences due to the weaker tense versus lax bias are still observed, although, tense versus lax does not exclusively select the tone. When an AP starts with a tense consonant, it starts with a substantially high tone, and this high f0 trend or register is sustained over multiple syllables. When an AP starts with a lax consonant, the overall f0 values for that phrase appear to be set in a speaker's low register. This explains why the f0 of the second syllable triggered by a tense consonant following the initial *L* in a lax-initial AP is never as high as the initial *H* and the following medial tones in a tense-initial AP.

We have identified phonological biases—the consonant type, f0 of the preceding syllable, and invariant phrase-final tone value—that explain most surface tonal patterns. The variability arising from the number of syllables per AP can be accounted for by assuming an alternating LH as the basic AP tonal pattern. In the case of the 4- or 5- syllable APs, the first LH sequence is associated with the first two syllables and the second LH sequence is associated with the last two syllables. The biases we have established are responsible for the surface variation. (The third syllable of the 5-syllable AP is therefore assigned with no tone, but the value is still determined by the f0 bias of the preceding syllable.) We have established that the bias coming from the f0 of a preceding syllable is quite strong. The consonant type effect on the second syllable is consistently constrained by the f0 of the initial syllable. Moreover, there is evidence that shows that the underlying tone is not fully realized due to the constraint coming from the f0 of the previous syllable. For example, in the 4-syllable AP starting with a tense consonant, the f0 of the penultimate syllable does not quite reach the 'unadulterated' low tone (as in *H*H**L**H), owing to the fairly high f0 of its preceding syllable. For the troublesome three-syllable AP variations, we posit a single set of LH spanning over the entire AP, which leaves the second syllable with no unique tone. In this situation, the biases in the system shape the overall tonal patterns (*3a*-*3b* in Figure 5). When the 3-sylalble AP starts with a tense consonant, the phrase-final tone fails to reach the invariant or target f0 value due to the bias created by the preceding high tones[4].

Overall, our results illuminate previously undocumented aspects of the consonant-type effect on tones in the Accentual Phrase of contemporary Seoul Korean. The consonant-type effect on f0 is manifested differently in different prosodic locations. The effect is categorical in phrase-initial position, but gradient in phrase-internal position. To explain the positional asymmetry, and the other previously unaccounted for variability, we have identified phonological constraints/biases in the system that affect the underlying LH(LH) tonal shape of an AP: 1) the f0 of the previous syllable, 2) consonant type, and 3) the invariant phrase-final tone. These biases work together synergistically, giving rise to the surface variability. Finally, the locally constrained effect of consonant type in phrase-internal position is accompanied by the f0$_{|v1-v2|}$ excursion difference. In the following subsection, we further discuss how these phrasally determined overall tonal patterns interact with the information organization among consonant contrasts for younger generation Seoul speakers.

Prosodic structure modulating local phonetic organization of consonants

In contemporary Seoul Korean, previous studies have reported that younger generation speakers show an enhanced f0 distinction between phrase-initial voiceless stop consonants, aspirated and lenis stops along with VOT mergers (Bang et al. 2018; Kang and Guion 2008; Kang 2014; Lee and Jongman 2012; Silva 2006). While no prior study has yet looked at such trade-off relation in *non*-initial stops, this study examines the three-way stop contrast of interest

---

[4] Our finding regarding the temporal scope of the initial *H* is a bit different from Cho and Lee (2016), who reported initial *Hs* affecting the tones within an entire AP, even a 5-syllable AP. The discrepancy comes from different experimental materials. Our target APs were designed to have lax onsets throughout an AP except for the target syllable that was experimentally manipulated (i.e. AP-initial condition). In contrast, their APs were designed to have alternating tonal sequences such as low-low (*lax-lax*), low-high (*lax-tense*), high-low (*tense-lax*) or high-high (*tense-tense*) tones. For example, the f0 values of the non-initial syllables including the final syllable in *h*-l-h-l-h or *h*-h-h-h-h were consistently higher than those in *l*-l-l-l-l or *l*-h-l-h-l. This indicates that the effect of the initial *T* might have been confounded with the effect coming from the f0 of the preceding syllable.

in both prosodic phrase initial and medial positions, testing the hypothesis that stabilization of local phonetic distinctions would be differently conditioned by phrasal position.

Our results show an unequal effect of prosodic location on the information reorganization. As discussed in detail in the previous subsection (§2.4.1), the effect of consonant type on the f0 of the following vowel is asymmetric in different prosodic positions. The consonantally triggered f0 difference is substantial in AP-initial position (tense >> lax), whereas it is marginally significant in AP-internal position (tense > lax). Although the f0 distribution analysis indicates that aspirated and fortis stops pattern together as a tense category, there are further breakdowns. In both positions, the f0 is slightly higher after an aspirated stop than after a fortis stop, which is in line with previous findings (e.g., Kang 2014; Silva 2006). Along with this finding, all speakers except one show that there are initial VOT mergers between the aspirated and lenis stops, confirming the recent findings (Bang et al. 2018; Kang 2014). In initial position, for all speakers the three stops are distinctive from each other in terms of the VOT and f0 combination (See Table 4 for a summary).

Table 4. 3-way voiceless stop contrast, lenis /**p**/, aspirated /**pʰ**/, fortis /**p***/, in *AP-initial* syllables. Different color codes indicate the tenseness distinctions, '**LAX**' vs. '**TENSE**.'

	VOT	f0
#/**p**a/	long	low
#/**pʰ**a/		much higher
#/**p***a/	short	

However, the individual speaker analysis further indicates that the f0 and VOT distribution patterns among speakers are not uniform. For the speakers who show a completely overlapping range of VOT values between the initial lenis and aspirated stops, the tonal distinction is sharp between tense (aspirated, fortis) and lax (lenis). For the speakers who show only partially overlapping portions of the initial VOTs, leading to a three-way distribution of the initial VOTs (aspirated > lenis >> fortis), there is a three-way distinction in the f0 values (aspirated > fortis > lenis), rather than a two-way distinction (tense >> lax). This suggests that the information organization might be adaptive, favorably serving to maintain or enhance the contrast, rather than making every phonetic aspect maximally distinctive. This type of adaptive (re)organization is also found in AP-internal position.

A clear contrast among medial stops is made chiefly through VOT and closure duration, not f0. In both prosodic positions, stop closure durations are shortest for lenis /p/ stops, intermediate for aspirated /pʰ/ stops, and longest for fortis /p*/ stops. While the lenis stops are produced with shorter closure duration in AP-internal position than in AP-initial position, the tense stops, both aspirated and fortis, are produced with longer closure duration in AP-*internal* position than in AP-initial position. Our results are comparable to previous findings[5] (e.g., Cho and Keating 2001; Han 1996; Jun 1994; Martin 1982; Oh and Johnson 1997; Yu 1989).

In AP-internal position, the occurrence of intervocalic lenis voicing (/apa/ → /aba/) is substantially reduced for our speakers (i.e. younger generations), compared to the data from the older generation reported in Jun (1993; 75 % voiced lenis tokens). Intervocalic lenis voicing has been shown to be gradient and subject to contextual variability such as segmental/prosodic environments and speech rate (Jun 1993, 1994). In our study, the frequency of lenis voicing varied greatly by speaker, ranging from 0 to 22 voiced out of 30, which again suggests that the information organization of the stop system varies individually, perhaps indicating a differential penetration of an ongoing sound change, as will be discussed in Chapter 4. Despite the individual difference, only 32 % of the lenis tokens underwent intervocalic voicing (57 out of 180). This shift led to a new case of VOT mergers.

The AP-internal aspirated stop always has the longest VOT, which made it distinct from the other two varieties. However, the AP-internal fortis and lenis (measured from tokens that did not undergo intervocalic voicing) stops show similar though still statistically differentiable mean VOT values. This near-merger of VOT is concurrent with the marginally significant f0 distinction (tense > lax). This is exactly the case in which the local f0 difference between non-initial lax and tense categories is conditioned by phrasal prosody. As the f0 of the non-initial syllable is greatly constrained by the f0 of the previous syllable, it cannot freely achieve the tonal distinction among the stop consonants. This is illustrated in the f0 and VOT distribution analyses (Figure 12). For most speakers, the AP-internal fortis and unvoiced lenis are not distinctive, providing no strong support for the role of f0 in maintaining contrast among the internal stops. Two speakers, however, show evidence for a possible role of f0 in retaining the contrast. Speaker M1 who does not produce a single voiced lenis token shows the least

---

[5] Note that Korean tense stops have long been said to be geminated in word-medial position. Therefore, this finding is not new.

overlapping f0 values between the AP-internal tense and lax stops. Moreover, Speaker M2 who shows the most voiced lenis tokens, shows the unvoiced lenis tokens associated with lower variants of f0 compared to the voiced-lenis, fortis, and aspirated stops. Crucially, we found that the tonal distinction between initial tense and lax is additionally supported by the f0 difference between adjacent syllables ($\text{f0}_{|v1\text{-}v2|}$ = tense > lax). This is achieved by lowering f0 of the lax-initial syllable at the beginning of an AP, in which a tonal realization is fairly free from phrasal constraints of the system. Taken together, our findings of prosodic conditioning in information (re)organization have implications for an intricate interplay between the paradigmatic contrast maintenance (tense vs. lax) and syntagmatic tonal patterns. Table 5 summarizes the newly emerging phonetic system of the three-way contrast among the AP-internal stops.

Table 5. 3-way voiceless stop contrast, unvoiced lenis /**p**/, fortis /**p***/, aspirated /**pʰ**/, in *AP-internal* syllables for contemporary Seoul Korean. Different color codes indicate the tenseness distinctions, '**LAX**' vs. '**TENSE**.'

	VOT	closure duration	f0	$\text{f0}_{|v1\text{-}v2|}$ excursion
/a**p**a/	short	short	low	small
/a**p***a/		long	higher	larger
/a**pʰ**a/	long			

Finally, this study tests whether there are the effects of boundary strengths on the phonetic properties of the stops. The results show no effect of prosodic boundaries on the phonetic organization. A lack of effect of boundary strength on VOT or f0 may simply indicate no effect of an IP. However, previous studies have shown that stronger articulation is associated with IP-initial position as compared to (IP-medial) AP-initial position (e.g., linguopalatal contact and VOT, IP-initially > AP-initially, Cho and Keating 2001), which suggests that IP-initial position may be the originally triggering context of the VOT merger. Alternatively, there is a possibility that the distinction in different boundary strengths is still made in qualities of the articulatory closure, as this data was not available for this study. Regardless, as discussed above, the extension of initial f0 effect to phrase-internal contexts is systematically conditioned by phrasal prosody that regulates the surface tonal patterns.

2.5. Summary and conclusions

In this work, we investigate the phonetic information organization in contemporary Seoul Korean of the three-way voiceless stop contrast (lenis /p/, aspirated /p^h/, fortis /p*/) and the consonant-type effect on tones in novel prosodic contexts that have not been examined previously (AP-initial vs. AP-internal).

In sum, the consonant type effect on f0 (tense > lax) is *asymmetric* in different locations within an AP. The effect is substantially larger in AP-initial position (tense >> lax) than in AP-internal position (tense > lax). There is a non-overlapping bi-modal distribution of f0 values in AP-initial position, dividing tense- and lax-induced f0 values into two discrete modes. This type of tonal grouping is not observed in AP-internal position. The results suggest that the consonant type effect on f0 is *categorical* in AP-initial position, compared to the *gradient* effect in AP-internal position. The global f0 patterns are dramatically different between an AP starting with a lax consonant versus an AP starting with a tense consonant. While a lax-initial AP consistently shows the canonical *L*HLH pattern, a tense-initial AP shows the non-canonical surface patterns such as *H*HHL or *H*HLL. When the consonant type is manipulated phrase-internally, the resulting f0 difference is locally constrained, conforming to the overall intonation pattern of the entire phrase. The results suggest that the temporal scope of the consonant effect on f0 is broad in AP-initial position, but small in AP-internal position. In AP-initial position, *both* initial *H* and initial *L* of an AP exert an influence on the following syllables, up to the third syllable of an AP. In AP-internal position, the scope of the consonantally derived tonal difference is rather limited to the target syllable and its preceding syllable at best.

These asymmetric f0 differences found in different prosodic positions may function to maintain or augment contrast among stop categories that exhibit VOT mergers and near-mergers. Our results confirm the recently reported word-initial VOT merger between aspirated /p^h/ and lenis /p/ stops (Bang et al. 2018; Kang 2014; Lee and Jongman,2012). We found the near-merger of VOT between the word-*internal* lenis /p/ and fortis /p*/ stops, arising from the substantially reduced occurrence of intervocalic lenis voicing in younger generation speakers. The small effect of the consonant type on f0 in phrase-internal position is augmented by f0 excursion between adjacent syllables, but closure duration seems to play a main role in this position. Our results show that the phonetic organization of consonants, particularly with regard to f0, is largely constrained by phrasal prosody, suggesting an intricate interplay between the paradigmatic contrast maintenance and syntagmatic tonal patterns.

This study also identifies phonological factors that shape the AP tones. The positional asymmetry suggests that the f0 value of the *preceding syllable* is a major determinant of the f0 value of the current syllable within the Korean AP. We propose that the underlying tonal pattern of an AP is a (repeating) LH(LH) sequence and that there exist phonological biases in the system that shape the surface tonal patterns. The identified biases are a) the f0 of the preceding syllable, b) consonant type, and c) an invariant phrase-final tone. Based on the temporal scope analyses, we assume the bias coming from the previous syllable to be stronger than the bias coming from the consonant type. This can also account for how the local phonetic information organization is shaped or constrained in different prosodic locations.

Our results show continuous effects of consonant type and tone that parallel a categorical shift in phonological forms, L versus H tonal contrast. While no formal analysis of these patterns is currently available, the observed parallelism of categorical and gradient effects can be modeled by adopting a dynamical grammar framework by Gafos and Benuš (2006). In this grammar, constraints/biases can select certain preferred states along a physical phonetic dimension. In the case of Seoul Korean tones, there are two categorical attractor states along the f0 continuum, low and high f0. With the phonological biases identified above, this grammar can predict the actual distribution of data. If the dynamical system has two distinct attractor states (e.g., L vs. H), a bimodal distribution of values along the dimension is predicted, one mode corresponding to each contrasting category. In order to account for the context-determined selection of modes (e.g., the prosodic control variable observed in this study), the dynamical system can be biased in the direction of one or another of the modes. In the case of the phrase-initial position, the consonant type biasing factor is the strongest, as there is no bias coming from a preceding syllable. Therefore, the [lax] versus [tense] factor selects L versus H tone in initial position, functioning as a phonological contrast. In the case of the phrase-*internal* position, however, a bias coming from consonant type is simultaneously present with a stronger bias (the f0 of the preceding syllable). In this case, the [tense] versus [lax] biases can still function to shift f0 values quantitatively in the presence of a stronger bias.

Acknowledgments

This work was supported by NIH DC03172 (Dani Byrd) and the project: “Tonal Placement - The Interaction of Qualitative and Quantitative Factors: TOPIQQ,” subcontract D-71631-Z-600-145002301 from the University of Cologne, under funding from the Volkswagen Foundation.

3. Seoul Korean laryngeal consonant and tone dynamics

3.1. Introduction

The phonological features [lax] and [tense] of stop consonants have generally been characterized by different "articulatory strengths." Previous studies with the 3-way voiceless stops (i.e. lenis, fortis, aspirated) in Korean have established that lenis stops behave differently from the other two varieties, aspirated and fortis stops in many phonetic measures (e.g., Cho, Jun, and Ladefoged 2002; Cho, Son, and Kim 2016; Cho and Keating 2001; Dart 1987; Han and Weitzman 1970; Hirose, Lee, and Ushijima 1974; Jun 1996; Kagaya 1974; Kim 1965; Kim, Honda, and Maeda 2005; Kim, Maeda, and Honda 2010; Lee and Jongman 2012; Son, Kim, and Cho 2012). When compared to the tense stops, relatively weaker articulation is associated with the lenis stop production, which includes: relatively slower buildup rate of buccal and subglottal pressure, shorter duration for maintaining the increased pressure, less linguopalatal contact (in the case of coronal stops) or smaller lip muscle activity (bilabial stops) for the occlusion, smaller degree and occlusion duration for the lip closing constriction, lower level of burst intensity and during the aspiration period, smaller amount of airflow following release, weaker harmonic components and slower rate of vibration in the following voice onset, etc.

The segmental "tenseness" has also been known to trigger distinctive behaviors of fundamental frequency (f0) during the following vowel (e.g., in Seoul Korean, Cho and Lee 2016; Jun 1993, 1996; Kang and Guion 2008; Kang 2014; Silva 2006), implying its systematic relation with tone gestures. In the previous chapter, we provide further new insights into the relation between the local phonetic organization of segmental contrast and the phrasal tone patterns in the phonological system of the contemporary Seoul dialect of Korean. We examine how the 3-way stop contrast (LENIS /p/, FORTIS /p*/, ASPIRATED /p^h/) is phonetically realized in various prosodic positions within an Accentual Phrase (AP) for 6 younger-generation speakers (born 1980-1990). Our results show that the local phonetic organization systematically interacts with prosodic structure. Fig. 1 summarizes the main findings regarding segmentally triggered f0 behaviors.

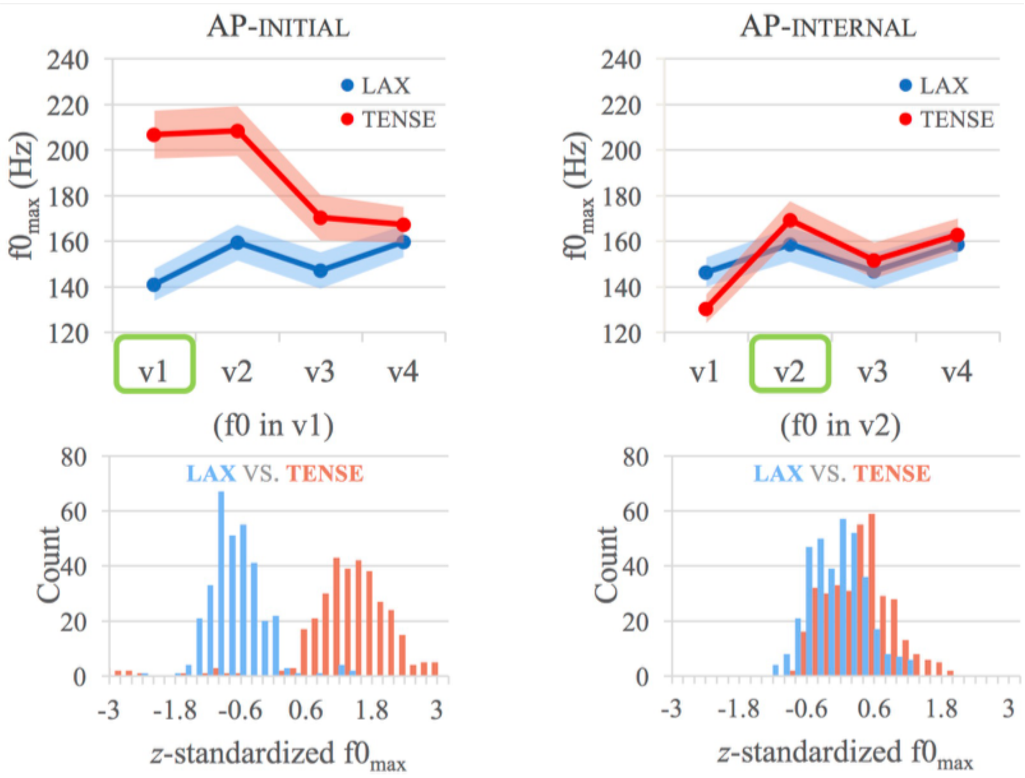

Figure 1. Consonant effect on f0 in different phrase positions.

The consonant effect on f0 is *categorical* in AP-initial position (**σ**-σ-σ-σ), compared to exhibiting a *gradient* effect in AP-internal position (σ-**σ**-σ-σ). We confirm a large f0 difference between lax (nasal, lenis) and tense (fortis, aspirated) stops in initial position, showing virtually no overlap in the distribution of f0 values between the two categories (left panel of Figure 1). In AP-internal position, we find a significant though small f0 difference between LAX and TENSE stops (overlapping distributions; right panel of Fig. 1). Moreover, this positional asymmetry is also found in the temporal scope of the consonantally induced f0 difference (initial *T*). The consonantal effect on f0 has a broad scope AP-initially—*T* exerting itself on syllables 2 and 3—but a locally constrained scope AP-internally. The key observation is that there is a substantial similarity between the f0 in the current syllable and the f0 of the previous syllable, even when they alternate by 40 Hz or so. We argue that several phonological constraints work together to produce the surface patterns: the f0 of the previous syllable, the consonant type (TENSE VS. LAX), and an oscillatory tone pattern (LHLH).

Based on these novel findings of the complex interaction in the prosodic dynamics of consonant and tone, in this chapter, we aim to shed light on what motor tasks are deployed for tone and segmental “tenseness” gestures, and how they function within the phonological system. In general, the vocal fold lengthening caused by cricothyroid (CT) muscle actions is known to be responsible for a high pitch. A low pitch is generally thought to result from a decrease in vertical tension produced by sterno-hyoid muscle action that lowers the entire larynx (Ohala 1972, *inter alia*). As such, contrastive tones (H vs. L) are controlled by discretely different articulatory mechanisms.

Prior electromyography and cine-magnetic resonance imaging (MRI) studies with older generation speakers of Kyungsang and Seoul Korean suggest that CT muscle activities and vertical larynx movements may be crucial for producing the 3-way stop contrast. In an electromyography study, Hirose, Lee, and Ushijimas (1974) examined the activities of the intrinsic phonatory laryngeal muscles including both tensor and adductor muscles during the production of the 3-way Korean stops in the Kyungsang dialect. Aspirated stops are produced with suppressed activity of the tensor muscle such as cricothyroid and vocalis muscles throughout the closure. This suppression is always followed by a steep increase in muscle activity after the release, which in turn gives rise to high f0 during the following vowel. With respect to fortis stops, there is a substantial increase in vocalis muscle activity immediately before the stop release. The increased tension (stiffening) of the vocal folds and constriction of the glottis during or immediately after the closure (laryngealization, Abramson and Lisker 1972) should be responsible for high f0 in the following vowel. In contrast, lenis stops do not show a sharp increase in tensor muscle activity before or after the stop release, resulting in lower variants of f0. Recent stroboscopic-cine MRI studies with two middle-aged Seoul Korean speakers (Kim, Honda, and Maeda 2005; Kim, Maeda, and Honda 2010) show that there are differences in vertical larynx movement when producing the 3-way contrast *both* in word-initial and word-internal position (fortis >/= aspirated > lenis).

These previous reports point to a possibility that both larynx height and vocal fold tension (lengthening of thyro-arytenoids caused by CT muscle actions) may play a role in contrastive tones in Korean. However, the speakers who participated in the abovementioned articulatory studies are categorized as earlier generations speakers of (non-)Seoul Korean who do not exhibit the ongoing tonogenetic sound change in Seoul Korean. Prior to this study, the change in f0 patterns has not been systematically investigated. Moreover, previous studies have used the cine-MRI technique that relies on composite images made from 128-256 individual repetitions of each individual utterance and is therefore not optimal for capturing the temporal and spatial variability intrinsic to speech articulation. This chapter employs the real-time MRI technique developed by the SPAN group at USC that provides real-time data on time-varying changes in vocal tract shaping. With this method, we imaged both supra-laryngeal and laryngeal articulations during the production of the AP in younger generation speakers.

As discussed above, in the newly emerging phonetic system of Seoul Korean stops, the consonant effect on f0 is categorical in AP-initial position but gradient in AP-internal position. This chapter serves as a testing ground for how the asymmetric positional effect on f0 is expressed by *controlled* actions of laryngeal articulators. Two competing hypotheses are (a) that the categorical versus gradient f0 differences arise from a single pitch-raising gesture that varies dynamically across prosodic contexts or (b) that they result from qualitatively different types of pitch gestures—specifically, larynx raising and stretched vocal folds due to CT muscle activity.

The possibility of different types of pitch gestures is motivated by a recent finding in Cantonese, in which there are four different levels of lexical tones (high level falling, mid-high level, mid-low level, low level falling). Nissenbaum (2008, 2010) reports that in Cantonese the f0 values of the two mid tones in running speech are not distinct. His cine-MRI evidence nonetheless shows that tones in different registers are associated with reliably different larynx heights. Each of the upper and lower extreme tones is produced with combination of the high larynx and stretched folds or with low larynx and short folds, respectively. Crucially, although the mid tones merge in f0 space in connected speech, they are nevertheless produced with distinct gesture combinations. Different registers (upper vs. lower) are associated with different larynx heights (high vs. low), and finer distinctions within each register are associated with vocal fold tension differences (stretched vs. shortened). Therefore, Nissenbaum argues that different combinations of larynx height and vocal fold tension can account for the observed variability. In other words, the two articulatory actions in the larynx are contrastive, yielding the observed f0 variations via their combination.

In this chapter, the scope of investigation and discussion is limited to the vertical larynx movement measure, leaving vocal fold tension to future research. For the observed positional asymmetry, we hypothesize that the categorical effect of consonants on the AP-initial f0 difference might be due to a large *larynx raising (or lowering) gesture*, which essentially constrains the overall pitch range (register) for the entire phrase. Based on a large temporal scope of the initial larynx raising or lowering gesture, we hypothesize that there may be a smaller (supporting Hypothesis a) or no difference at all (supporting Hypothesis b) in larynx height between tense and lax stops phrase-internally. If there is no larynx height difference between the phrase-internal stops, the significant but small f0 difference may be yielded by some other articulatory mechanism such as vocal fold tension. The previous chapter confirms the frequently observed AP-final high tone (e.g., Jun 1993), and further shows that there may be an invariant tonal target at the end of an AP. This chapter also tests what articulatory mechanism is responsible for the phrase-final high, and how the tone gesture further interacts with the consonant effect in phrase-final position.

The real-time MRI technique employed will also allow for the quantification of time-varying changes in the oral constriction gestures as well as in the larynx upwards/downwards motions, which will provide further information regarding the lax versus tense distinction and the inter-gestural coordination between tone and segmental gestures. Previous magnetometer studies (Cho et al. 2016; Son et al. 2012) found that the tense and lax distinction among the Seoul Korean stops are manifested most consistently by constriction degree (fortis /p*/ > aspirated /p^h/ > lenis /p/) and also reliably by constriction (occlusion) duration (/p*/, /p^h/ > /p/) across prosodic positions. This study will test what supra-laryngeal kinematic characteristics are associated with the tense-lax distinction (e.g., longer constriction formation duration and greater constriction degree for tense stops), and whether the consonantal constriction goals interact with phrasal prosody. This investigation will contribute to an understanding of the articulatory mechanisms that express f0 for different phonological structures, and of the tone gestures during voiceless oral gestures.

3.2. Method

3.2.1. Speakers

Three native speakers of Seoul Korean, two females (Speakers S1 & S2) and one male (Speaker S3), participated in the experiment. All three speakers were born in 1990, categorized as younger generations' Seoul dialect, and completed their undergraduate degrees in Seoul. They were pursuing graduate studies at the University of Southern California at the time of the recording. Two of them (Speakers S2 & S3) participated in the acoustic study reported in the previous chapter (F1 & M3, respectively), in which they exhibited the younger generation's speech characteristics. None of them reported any history of speech or hearing impairment.

3.2.2. Test materials

The test phrases were designed to have two disyllabic words ($[c1v1c2v2\#c3v3c4v4]_{AP}$), forming quadrisyllabic APs. Each target AP was designed to have a target syllable (CV) composed of one of the four bilabial stop consonants—nasal /m/, lenis /p/, aspirated /p^h/, and fortis /p*/—as an onset, followed by a high front vowel /i/ in Korean. As established in the previous chapter, the nasal and lenis stops pattern together as LAX consonants, and the fortis and aspirated stops pattern together as TENSE consonants in terms of their f0 characteristics. Before the experiment, a pilot MRI study with Speaker S3 was conducted to check if a newly developed centroid tracking method for the larynx up-and-down movements would perform well for this dataset (see the subsection §3.2.4.2. for details about this method). In this pilot run, the vowel /a/ was found to contribute to erroneous data points, as during the production of /a/, a portion of the supra-laryngeal articulators such as the tongue root or epiglottis often invade the rectangular region in which the quantification of vertical larynx movements is made. For this reason, /i/ was selected for the experimental stimuli.

The quadrisyllabic target APs were built using a morpheme concatenation of two disyllabic words. Sets of two words that form a quadrisyllabic AP were selected to examine the effect of prosodic positions within an AP. The primary criterion for selecting words was to find real words with the consonantal contexts of interest. The disyllabic words contained the target CV, the first or second syllable within a word. Non-target syllables within the AP were composed of either a lax or tense onset consonant (not necessarily a bilabial stop). Note that having a phonologically optimal set of quadrisyllabic target phrases put a restriction on balancing word frequency or finding an actual real-word compound. In particular, some of the pseudo-compound sequences with the tense-initial syllables (e.g., /p*i/, /p^hi/) were exceptionally low frequency words. Our speakers were asked to treat each target AP as a compound (proper) noun phrase that modifies the following AP, another noun phrase, to resolve the unnaturalness potentially arising from the low frequency or semantic incorrectness.

First, two types of pseudo-word APs were constructed to understand how the overall accentual phrasal tones are produced when the consonant type (LAX or TENSE) does not vary throughout the phrase (Table 1). In the ALL-LAX condition, all four syllables had a lax onset stop: [mi.mi.pi.pi]AP and [pi.pi.mi.mi]AP. In the ALL-TENSE condition, all four syllables had a tense onset stop: [p*i.p*i.p^{h}i.p^{h}i]AP and [p^{h}i.p^{h}i.p*i.p*i]AP. This controlled set of stimuli also allows for examining the supra-laryngeal articulatory kinematics for different bilabial stops (NASAL, LENIS, FORTIS, and ASPIRATED). All test materials were presented in Hangul in the experiments.

Table 1. ALL-TENSE VS. ALL-LAX conditions. Expected accentual tone patterns are listed in parentheses.

σ1-σ2-σ3-σ4	Consonant	Target AP
LAX-LAX-LAX-LAX (*LOW*-HIGH-LOW-HIGH)	NASAL-NASAL-LENIS-LENIS	mi.mi#pi.pi
	LENIS-LENIS-NASAL-NASAL	pi.pi#mi.mi
TENSE-TENSE-TENSE-TENSE (*HIGH*-HIGH-LOW-HIGH)	FORTIS-FORTIS-ASPIRATED-ASPIRATED	p*i.p*i#p^{h}i.p^{h}i
	ASPIRATED-ASPIRATED-FORTIS-FORTIS	p^{h}i.p^{h}i#p*i.p*i

Additionally, quadrisyllabic APs with varying compositions and positions of LAX- and TENSE-initial syllables were constructed to test the consonant effect in different phrase positions within an AP. One of the comparisons was made to test (and to confirm) the consonant effect on tone in AP-initial position and AP-internal position. For the AP-INITIAL condition, the target syllable with varying consonants was placed at the beginning of a phrase, and the rest of the syllables had a lax consonant (**LAX**-LAX-LAX-LAX VS. **TENSE**-LAX-LAX-LAX). These phrases included [**mi**.ni.pi.ni]AP, [**pi**.ɾi.ki.ɕi]AP, [**p*i**.mi.pi.ni]AP, and [**p^{h}i**.ti.pi.ɾi]AP. For the AP-INTERNAL condition, the target syllable was the second syllable of a phrase, and the rest of the syllables had a lax consonant (LAX-**LAX**-LAX-LAX VS. LAX-**TENSE**-LAX-LAX). This condition included [ki.**mi**.pi.pi]AP, [mi.**pi**.ki.ɕi]AP, [mi.**p*i**.pi.ni]AP, and [ti.**p^{h}i**.ki.ɕi]AP.

In order to test how the phrase-medial high (*T***H**LH) or low (*T*H**L**H) accentual tones and phrase-final high accentual tones (*T*HL**H**) are realized depending on the consonantally triggered AP-initial tone (*T*: L for LAX or H for TENSE), we further compare the LAX-INITIAL APs with the TENSE-INITIAL APs with varying onset consonants. In one condition, the effect of the AP-initial tone on the tone of the immediately following syllable was tested using the following items: LAX-**LAX**-LAX-LAX: [ki.**mi**.pi.pi]AP, [mi.**pi**.ki.ɕi]AP; LAX-**TENSE**-LAX-LAX: [mi.**p*i**.pi.ni]AP, [ti.**p^{h}i**.ki.ɕi]AP; TENSE-**LAX**-LAX-LAX: [t^{h}i.**mi**.pi.pi]AP, [t^{h}i.**pi**.ki.ɕi]AP; and TENSE-**TENSE**-LAX-LAX: [p*i.**p*i**.ki.ɕi]AP, [hi.**p^{h}i**.ki.ɕi]AP. In the other condition, the effect of the AP-initial tone on the tone of the *third* syllable was tested: LAX-LAX-**LAX**-LAX: [ki.ɕi.**mi**.ti]AP, [mi.ni.**pi**.ni]AP; LAX-LAX-

TENSE-LAX: [mi.ni.**p*i**.mi]AP, [pi.ɾi.**pʰi**.ti]AP; TENSE-LAX-**LAX**-LAX: [tʰi.pi.**mi**.ti]AP, [p*i.mi.**pi**.ni]AP; and TENSE-LAX-**TENSE**-LAX: [p*i.ki.**p*i**.mi]AP, [tʰi.pi.**pʰi**.ti]AP. Finally, the LAX-INITIAL and TENSE-INITIAL APs with varying final consonants were built to test if there is any consonant effect on the production of the "invariant" AP-final tone: LAX-LAX-LAX-**LAX**: [mi.ni.ki.**mi**]AP, [mi.ni.pi.**pi**]AP; LAX-LAX-LAX-**TENSE**: [mi.ni.mi.**p*i**]AP, [pi.ɾi.ki.**pʰi**]AP; TENSE-LAX-LAX-**LAX**: [pʰi.ɕi.ki.**mi**]AP, [tʰi.mi.pi.**pi**]AP; and TENSE-LAX-LAX-**TENSE**: [p*i.ki.mi.**p*i**]AP, [tʰi.pi.ki.**pʰi**]AP.

Table 2 provides the glossary for target phrases in different consonant *x* prosodic conditions.

Table 2. The glossary of target items used in this study. Target syllables are in bold. Expected accentual tone patterns are listed in parentheses.

σ1-σ2-σ3-σ4	Consonant	Target AP	Gloss Word 1	Gloss Word 2
LAX-LAX-LAX-LAX (***LOW***-HIGH-LOW-HIGH)	NASAL	**mi**.ni#pi.ni	“*miniature*”	“*beanie*”
	LENIS	**pi**.ɾi#ki.ɕi	“*corruption*”	“*base*”
TENSE-LAX-LAX-LAX (***HIGH***-HIGH-LOW-HIGH)	FORTIS	**p*i**.mi#pi.ni	“*Ppimi*-Prop. N”	“*beanie*”
	ASPIRATED	**pʰi**.ti#pi.ɾi	“*producer*”	“*corruption*”
LAX-**LAX**-LAX-LAX (*LOW*-**HIGH**-LOW-HIGH)	NASAL	ki.**mi**#pi.pi	“*freckles*”	“*blemish balm*”
	LENIS	mi.**pi**#ki.ɕi	“*incompleteness*”	“*base*”
LAX-**TENSE**-LAX-LAX (LOW-**HIGH**-LOW-HIGH)	FORTIS	mi.**p*i**#pi.ni	“*Mippi*-Prop. N”	“*beanie*”
	ASPIRATED	ti.**pʰi**#ki.ɕi	“*display*”	“*base*”
TENSE-**LAX**-LAX-LAX (*HIGH*-**HIGH**-LOW-HIGH)	NASAL	tʰi.**mi**#pi.pi	“*Timmy*-Prop. N”	“*blemish balm*”
	LENIS	tʰi.**pi**#ki.ɕi	“*television*”	“*base*”
TENSE-**TENSE**-LAX-LAX (*HIGH*-**HIGH**-LOW-HIGH)	FORTIS	p*i.**p*i**#ki.ɕi	“*pager*”	“*base*”
	ASPIRATED	hi.**pʰi**#ki.ɕi	“*hippie*”	“*base*”
LAX-LAX-**LAX**-LAX (*LOW*-HIGH-**LOW**-HIGH)	NASAL	ki.ɕi#**mi**.ti	“*base*”	“*MIDI*”
	LENIS	mi.ni#**pi**.ni	“*miniature*”	“*beanie*”
LAX-LAX-**TENSE**-LAX (*LOW*-HIGH-**LOW**-HIGH)	FORTIS	mi.ni#**p*i**.mi	“*miniature*”	“*Ppimi*-Prop. N”
	ASPIRATED	pi.ɾi#**pʰi**.ti	“*corruption*”	“*producer*”
TENSE-LAX-**LAX**-LAX (*HIGH*-HIGH-**LOW**-HIGH)	NASAL	tʰi.pi#**mi**.ti	“*television*”	“*MIDI*”
	LENIS	p*i.mi#**pi**.ni	“*Ppimi*-Prop. N”	“*beanie*”
TENSE-LAX-**TENSE**-LAX (*HIGH*-HIGH-**LOW**-HIGH)	FORTIS	p*i.ki#**p*i**.mi	“*Ppiki*-Prop. N”	“*Ppimi*-Prop. N”
	ASPIRATED	tʰi.pi#**pʰi**.ti	“*television*”	“*producer*”
LAX-LAX-LAX-**LAX** (*LOW*-HIGH-LOW-**HIGH**)	NASAL	mi.ni#ki.**mi**	“*miniature*”	“*freckle*”
	LENIS	mi.ni#pi.**pi**	“*miniature*”	“*blemish balm*”
LAX-LAX-LAX-**TENSE** (*LOW*-HIGH-LOW-**HIGH**)	FORTIS	mi.ni#mi.**p*i**	“*miniature*”	“*Mippi*-Prop. N”
	ASPIRATED	pi.ɾi#ki.**pʰi**	“*corruption*”	“*avoidance*”
TENSE-LAX-LAX-**LAX** (*HIGH*-HIGH-LOW-**HIGH**)	NASAL	pʰi.ɕi#ki.**mi**	“*blackhead*”	“*freckle*”
	LENIS	tʰi.mi#pi.**pi**	“*Timmy*-Prop. N”	“*blemish balm*”
TENSE-LAX-LAX-**TENSE** (*HIGH*-HIGH-LOW-**HIGH**)	FORTIS	p*i.ki#mi.**p*i**	“*Ppiki*-Prop. N”	“*Mippi*-Prop. N”
	ASPIRATED	tʰi.pi#ki.**pʰi**	“*television*”	“*avoidance*”

Each quadrisyllabic target AP was placed in the middle of a carrier sentence, which is one Intonational Phrase (IP) consisted of four APs. As shown in b), the target AP was treated as a movie name that modifies the following noun phrase, /pitio-ɾɨl/ (“*video*” + accusative case marker).

b) The carrier frame:

[[sʌn.sæŋ.ni.mi]AP **[“__________”]AP** [pi.ti.o.ɾɨl]AP [pil.lim.ni.ta]AP]IP.
The teacher **target AP (PROPER NOUN)** *video*-ACCUSATIVE *is renting.*

The flanking syllables of the target AP consisted of a bilabial lax stop (/m/ or /p/) and the high front vowel /i/. In order to induce natural prosodic phrasing even with a potentially unnatural target phrase in a sentence, a frame sentence with some quadrisyllabic name of a real movie (e.g., [hæ.ɾi.pʰo.tʰʌ]$_{\text{AP}}$ “*Harry Potter*”) was presented at the beginning of each experimental block.

Each speaker repeated each sentence 6 times in a randomized order. In total, 576 tokens were collected and analyzed ([3 speakers x 2 consonant type x 2 items x 6 repetitions] + [3 speakers x 4 consonants x 7 prosodic positions x 6 repetitions], Tables 1-2).

3.2.3. Real-time MRI data and audio acquisition

MR image and audio data were acquired at Los Angeles County Hospital using an MRI protocol developed for research on speech production (Narayanan, Nayak, Lee, Sethy, and Byrd 2004). During scans, speakers laid supine with the head restrained in a still position. The test sentences were presented on a back-projection screen, which the speakers could read from within the scanner using a mirror. Each sentence was presented one at a time. A 13-interleaf spiral sequence was used (TR = 6.004 ms, field of view = 200 x 200 mm, flip angle = 15^{o}). For the spiral sequence, a 5 mm slice located in the mid-sagittal plane of the vocal tract was scanned with a resolution of 68 x 68 pixels with 2.9 mm width. The acquired videos were reconstructed with a 2-frame sliding window giving an effective frame rate of 83.3 frames/s (one frame = 0.012 s). This high frame rate is enabled by constrained reconstruction (Lingala et al. 2017). Audio was simultaneously recorded at a sampling frequency of 20 kHz inside the MRI scanner while subjects were imaged.

3.2.4. MRI data analysis

3.2.4.1. Region-of-interest (ROI) analysis for supra-laryngeal constriction formation

Articulatory kinematic data were obtained from the MR images by using mean pixel intensity values within localized regions-of-analysis (ROIs) of the vocal tract (Blaylock 2017; Lammert, Ramanarayanan, Proctor, and Narayanan 2013). Changes in pixel intensity values of a particular pixel over time signify localized changes in tissue density. Lower intensity values correspond to the absence of tissue, while higher values indicate that some speech articulators are present at that particular point. Therefore, in any given region of the vocal tract, increased mean pixel intensity value in a region reflects the presence of a vocal tract constriction, as the articulator (such as the tongue or lips) impinges on the ROI.

In this study, circles with a radius of 3 pixels were placed in the image plane along the length of the vocal tract from the larynx to the lower lip. The origins of these ROIs were placed along the vocal tract *midline* determined by finding the connected sequence of pixels that exhibit the highest standard deviation of pixel intensity across frames. Fig. 2a shows an example standard deviation image of a video for Speaker S1, and Fig. 2b shows the automatically calculated midline (red rectangles indicate pixels constituting the midline).

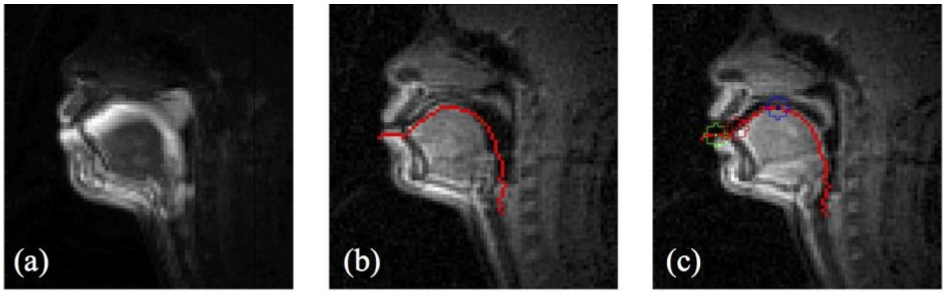

Figure 2. Representative images for ROI analysis for Speaker S1.

Then, the location of the circle along the midline appropriate for each constriction gesture was manually determined. Although the target consonant was always a bilabial stop consonant, the neighboring non-target consonant articulation was also important in order to assess the overall coordination between the supra-laryngeal gesture and larynx movement. Therefore, local regions that cover lower lip closing, tongue tip constriction, and tongue body constriction were employed (Fig. 2c). In this chapter, we only report the lip closing kinematics for a target bilabial stop consonant (/m/, /p/, /p*/, /p^h/).

Each region in the image plane was carefully selected by visually confirming the following. First, the active articulator creating the consonant of interest reliably was present in the region in every frame. Second, an appropriate origin pixel for the region was selected manually, by inspecting the vocal tract morphology in the mean images and articulator movement during the entire videos. The origin pixel for each circular region was chosen with reference to edge pixels of the circle that overlap with the edge of the (passive) anatomical landmarks where a relevant speech articulator makes a constriction (e.g., lower lip touching the upper lip, tongue tip against alveolar-ridge, or tongue body raising toward the center of the palate). The process allowed for 2-3 candidate regions immediately adjacent to each other. At a frame of maximum supra-laryngeal constriction, *displacement* values measured in each candidate region were compared, and the region with the largest displacement value was selected. (Displacement values were obtained through the procedure described later in this subsection. Movement displacement was calculated as the difference in mean pixel intensity between the points of gestural onset and constriction maximum in the intensity contours [see Fig. 3].) As speakers had different vocal tract shapes, the circular ROIs were defined on a by-subject basis. For an example of the manually selected region placement, see Fig. 2c.

Once a region was defined, a constriction time function was obtained by averaging the intensity values of pixels within the region for each frame. This kind of averaging is useful to reduce noise substantially as compared to the signal from an individual pixel, and, more crucially, to estimate the speech articulator motion in each region by measuring the average tissue density in the selected region. (The reader is referred to Lammert et al. (2013) for more detailed discussion.) Change in mean pixel intensity value in a region over time reflect the formation and release of a constriction as the articulator (such as the tongue or lips) moves into and out of the ROI.

To minimize noise or random intensity fluctuations, all resulting signals were smoothed by a locally weighted linear regression (e.g., Lammert, Goldstein, and Iskarous 2010). The weighting function used was a Gaussian kernel K with a standard deviation of h samples. Here, the kernel width parameter was $h = .9$ samples. As samples lying more than $3h$ from the center of the kernel in either direction receive weights near zero, this gives a smoothing window width of roughly 90 ms given the sampling period of 12 ms.

Articulatory analysis was carried out by using MView (algorithm by Mark Tiede at Haskins Laboratory). Temporal landmarks for the lips, tongue tip, and tongue body of the stops were algorithmically identified using the velocity of a manually located measurement window on the mean-pixel-intensity-based contours. These landmarks included the time points of movement onset, target achievement, constriction maximum, constriction offset, and gestural offset. Movement onset is defined as the point where the velocity first crosses the +/–10 percent threshold of the first peak velocity. Target achievement (constriction onset) is defined by locating the point at which the velocity falls below the same percent threshold. Constriction maximum is defined by identifying the zero-crossing point in the velocity signals. Constriction offset is defined as the first threshold-crossing point before the second (release) peak velocity. Movement offset is defined as the point where the velocity falls below the same +/–10 percent threshold at the right edge of the window. The landmark identification may be more easily understood with reference to the schematic in Fig. 3.

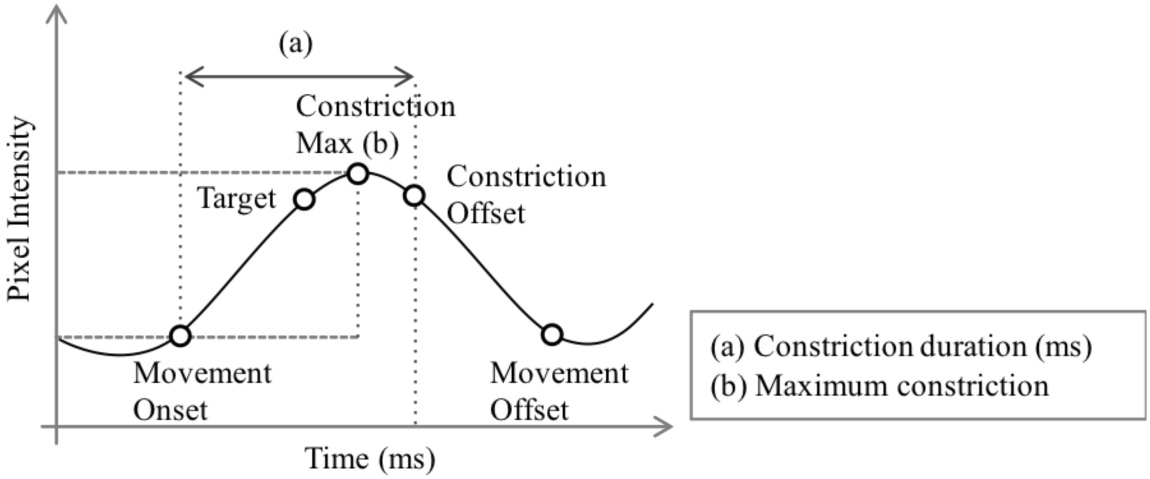

Figure 3. Temporal landmarks identified using the velocity function.

Based on the landmarks defined above, temporal and magnitude measures were both derived. Temporal measures included constriction (formation) duration defined as the time between movement onset and constriction offset ((a) in Fig. 3). For the magnitude measure, a mean pixel intensity value was measured at the point of constriction maximum ((b) in Fig. 3) to capture extreme constriction of the compressed tissue. Note that one of the speakers (Speaker S3) shows a largely reduced lip constriction gesture for some phrase-penultimate (4 tokens out of 6) and -final nasals (3 out of 6) in [pi.pi.**mi.mi**]$_{\text{AP}}$. The lip constriction measures for these nasals are not available.

3.2.4.2. Centroid tracking method

Although the above ROI technique is well suited for examining the oral constriction formations in real-time MRI data, it is not optimal for quantifying the direction and magnitude of movements of an articulatory structure that is not engaged in forming a constriction, such as upwards or downwards movement of the larynx. This study uses a method that finds the intensity *centroid(s)* of a selected ROI in the image (Oh, Toutios, Byrd, and Narayanan 2017; Tilsen et al. 2016). Intensity centroids are spatial positions, which are different from aggregating measures of pixel intensity (as done in the above ROI technique). An intensity centroid, the intensity-weighted average of an object in a certain selected region, represents the mean spatial location of tissue in that region. This method is well suited for tracking the vertical aspect of larynx movement because the centroid values reflect the mean location of tissue within a region and therefore the position of the articulator whose tissue is represented, and because changes in the centroid value over time indicate the direction and magnitude of articulator. When the selected region corresponds to the area of the image around the larynx, the vertical position of the centroid can be interpreted as larynx height.

In this study, we used a revised version of the Matlab code used in Oh et al. (2017) to estimate vertical movements of the larynx. This code tracks the time-varying pixel intensity centroid of a manually selected rectangular ROI for the larynx. A fixed ROI appropriate for each subject was defined based on the subject's cervical vertebra location (Fig. 4a). Specifically, a rectangular region was placed between the midline of their 2nd cervical vertebra (C-2) and the midline of C-4 for Speakers S1 and S3, and a region was placed between the top line of C-3 and the midline of C-4 for Speaker S2. The right side of a larynx region is always aligned against the outline of the pharyngeal wall. The size of the ROI appropriate for each speaker varied by their vocal tract shape and size (width x height in pixels: 4 x 15 for Speaker S1; 4 x 12 for Speaker S2; 4 x 16 for Speaker S3).

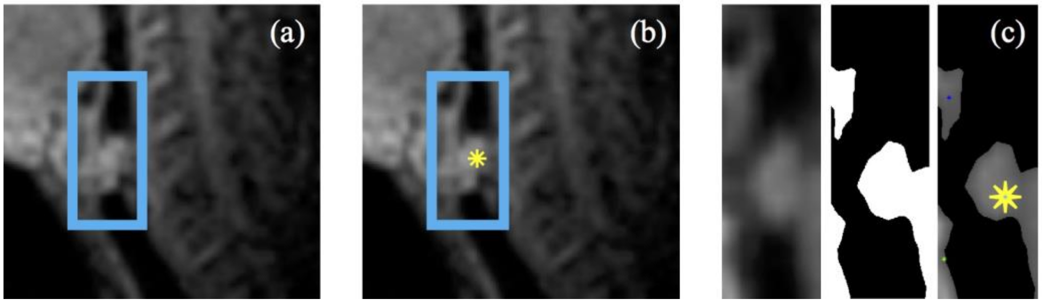

Figure 4. Example preprocessing steps of centroid tracking analysis.

Within a specified region, there can be other articulator objects in addition to the larynx object of interest (e.g., tongue root or epiglottis). In order to capture only the object of interest and remove centroid weighting generated by any other objects in the ROI, some preprocessing had to be undertaken. First, a *seed* was selected anywhere on the larynx object (indicated by a yellow asterisk [*] in Fig. 4b). Based on each pixel's intensity values, a binary matrix was obtained by assigning 1 to pixels brighter than the threshold intensity and 0 otherwise. The threshold intensity is calculated based on the mean and the standard deviation across the region with a 90% confidence interval ($Z > 0.8225$). The flood-fill algorithm on this binary matrix was employed to get connected components in the ROI. Then, the intensity-weighted centroid of each connected component was calculated, and a centroid that is closest to the seed was tracked as the centroid of the first frame. This step serves to ensure the continuity of tracking the laryngeal structure from frame to frame. From the following frame onward, the closest centroid from the previous frame's centroid was automatically selected as the current centroid of a given frame. Fig. 4c depicts the final preprocessing step.

In order to reduce noise and faulty intensity fluctuations, the larynx height trajectories were smoothed by loess smoothing (i.e. a locally weighted scatter plot smooth method) using a quadratic polynomial regression model with a local span of 50 data points.

3.2.5. F0 measurements

F0 analysis was carried out using Praat (Boersma and Weenink 2018). A speaker-specific pitch range was used. For both female speakers, the pitch range was set to 100-600 Hz. For the male speaker, the range was set between 75-300 Hz. Then, Praat's built-in autocorrelation tracking algorithm was used to automatically track pitch during any voiced intervals of the recorded speech. In order to match up with the frame rate of the MR images, f0 values at every 0.012 s were obtained by interpolation. In general, this method worked successfully, but there were some cases in which it would track data points that are not f0 but its higher harmonics. For all three speakers, these erroneous f0 points were trimmed using the median absolute deviation (MAD) measure. For the female speakers, data points that were more than two absolute deviations from the median of their f0 values were excluded from the analysis. For the male speaker, 3 MADs was the appropriate trimming threshold.

3.2.6. F0 maximum and the corresponding larynx height

For the AP tonal pattern analysis, each target vowel was manually segmented and labeled on a Praat text grid, and the time points, at which each f0 maximum value ($f0_{max}$) occurs, were automatically taken from the vowels using a Praat script. Then, the f0 values and the corresponding larynx height centroid values at f0 peaks were coded.

3.2.7. Significance testing

Statistical evaluation on the critical factors was made for each speaker independently, using *R* (*R* Development Core Team 2018). In order to first assess the general relationship between the variables, f0 and vertical larynx movement, a Pearson product-moment correlation coefficient was computed.

For the values of $f0_{max}$ and the corresponding vertical centroid of the larynx during the target CV, we conducted multiple sets of an ANOVA with different combinations of the critical factors: Consonant Type (LAX [NASAL vs. LENIS pooled] vs. TENSE [FORTIS vs. ASPIRATED pooled]), Phrase Position (AP-INITIAL vs. AP-INTERNAL), Initial Tone (LAX-INITIAL vs. TENSE-INITIAL), and Prosodic Position (AP-SECOND vs. AP-THIRD vs. AP-FINAL) factors. When significant variation among conditions was detected, post-hoc comparisons using the Tukey's HSD were conducted on all possible pairwise contrasts. First, a two-way ANOVA with Consonant Type and Phrase Position was conducted, comparing $f0_{max}$ and larynx height at $f0_{max}$ of consonants in the following phrase positions: **LAX**-LAX-LAX-LAX vs. **TENSE**-LAX-LAX-LAX vs. LAX-**LAX**-LAX-LAX vs. LAX-**TENSE**-LAX-LAX. (This was done also to confirm the effects reported in the previous chapter.) Then, we ran a three-way ANOVA testing the effects of Consonant Type, Initial Tone, and Prosodic Position on the variables of AP-tones. The conditions included in this analysis are (also see Table 2): LAX-**LAX**-LAX-LAX, LAX-LAX-**LAX**-LAX, LAX-LAX-LAX-**LAX**, LAX-**TENSE**-LAX-LAX, LAX-LAX-**TENSE**-LAX, & LAX-LAX-LAX-**TENSE** for LAX-INITIAL APs; and TENSE-**LAX**-LAX-LAX, TENSE-LAX-**LAX**-LAX, TENSE-LAX-LAX-**LAX**, TENSE-**TENSE**-LAX-LAX, TENSE-LAX-**TENSE**-LAX, & TENSE-LAX-LAX-**TENSE** for TENSE-INITIAL APs.

For the lip closing kinematic measures (i.e. constriction duration and maximum constriction), a one-way ANOVA with Consonant (NASAL vs. LENIS vs. FORTIS vs. ASPIRATED) was conducted on the AP-initial stops, in which the initial consonant was always precede by [mi] in the preceding AP. The items from Table 2 that were used in this analysis included: [**mi**.ni.pi.ni]AP, [**mi**.pi.ki.ɕi]AP for NASAL; [**pi**.ɾi.ki.ɕi]AP, [**pi**.ɾi.ki.pʰi]AP for LENIS; [**p*i**.mi.pi.ni]AP, [**p*i**.ki.mi.p*i]AP for FORTIS; [**pʰi**.ti.pi.ɾi]AP, [**pʰi**.ɕi.ki.mi]AP for ASPIRATED. In addition, a two-way ANOVA with Consonant (NASAL vs. LENIS vs. FORTIS vs. ASPIRATED) and Prosodic Position (AP-INITIAL vs. AP-SECOND vs. AP-THIRD vs. AP-FINAL) factors was additionally performed on bilabial consonants in the ALL-LAX and ALL-TENSE conditions to examine the possible interaction between factors (Table 2).

In all cases, *p*-values less than .05 were considered significant.

3.3. Results

3.3.1. Tonal measures

3.3.1.1. Correlation between f0 and larynx centroid vertical movement

For all three speakers, there is a strong positive correlation between f0 and larynx height (S1: *r*=.769; S2: *r*=.685; S3: *r*=.832; all at *p*<.05). As shown in Fig. 5, lower f0 values are associated with lower positions of the larynx, and higher f0 values are associated with higher positions of the larynx.

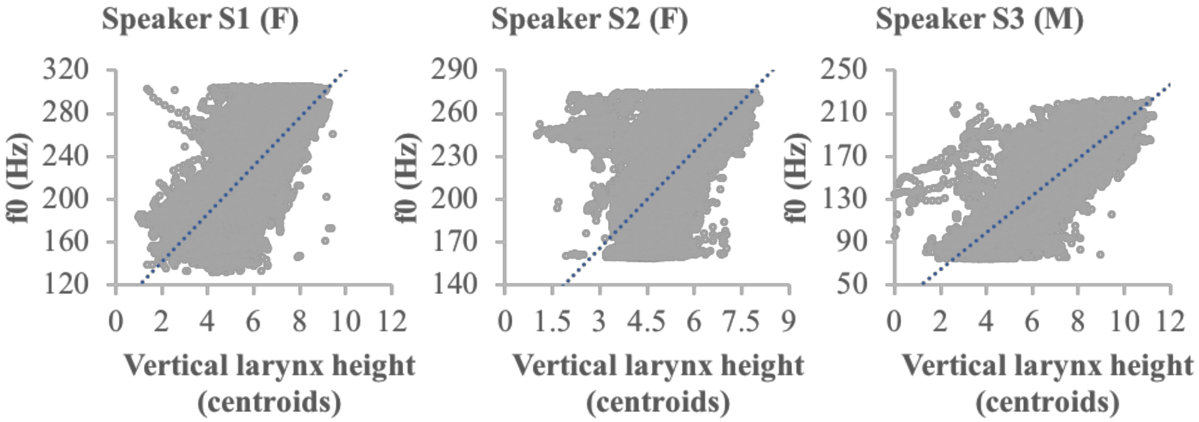

Figure 5. Correlation between f0 and vertical larynx height.

3.3.1.2. F0 and larynx height at f0 maximum point

Consonant Type effect on AP-initial vs. AP-internal (second) tones: To help visualize overall phrase tone patterns as well as the subsequent analyses on the consonant effect in different prosodic positions, Fig. 6 shows graphically each individual speaker's results. Each panel shows the mean values and 95 % confidence bands of maximum f0 during vowels in the quadrisyllabic APs. The panels on the left side (in reference to the dotted gray line in the middle) of Fig. 6 show the experiment condition in which the Consonant Type manipulation (TENSE vs. LAX) was made at the beginning of an AP (i.e. AP-INITIAL condition). The right side of Fig. 6 shows the AP-INTERNAL condition, in which the Consonant Type (TENSE vs. LAX) factor was manipulated in the phrase-second syllable (LAX otherwise).

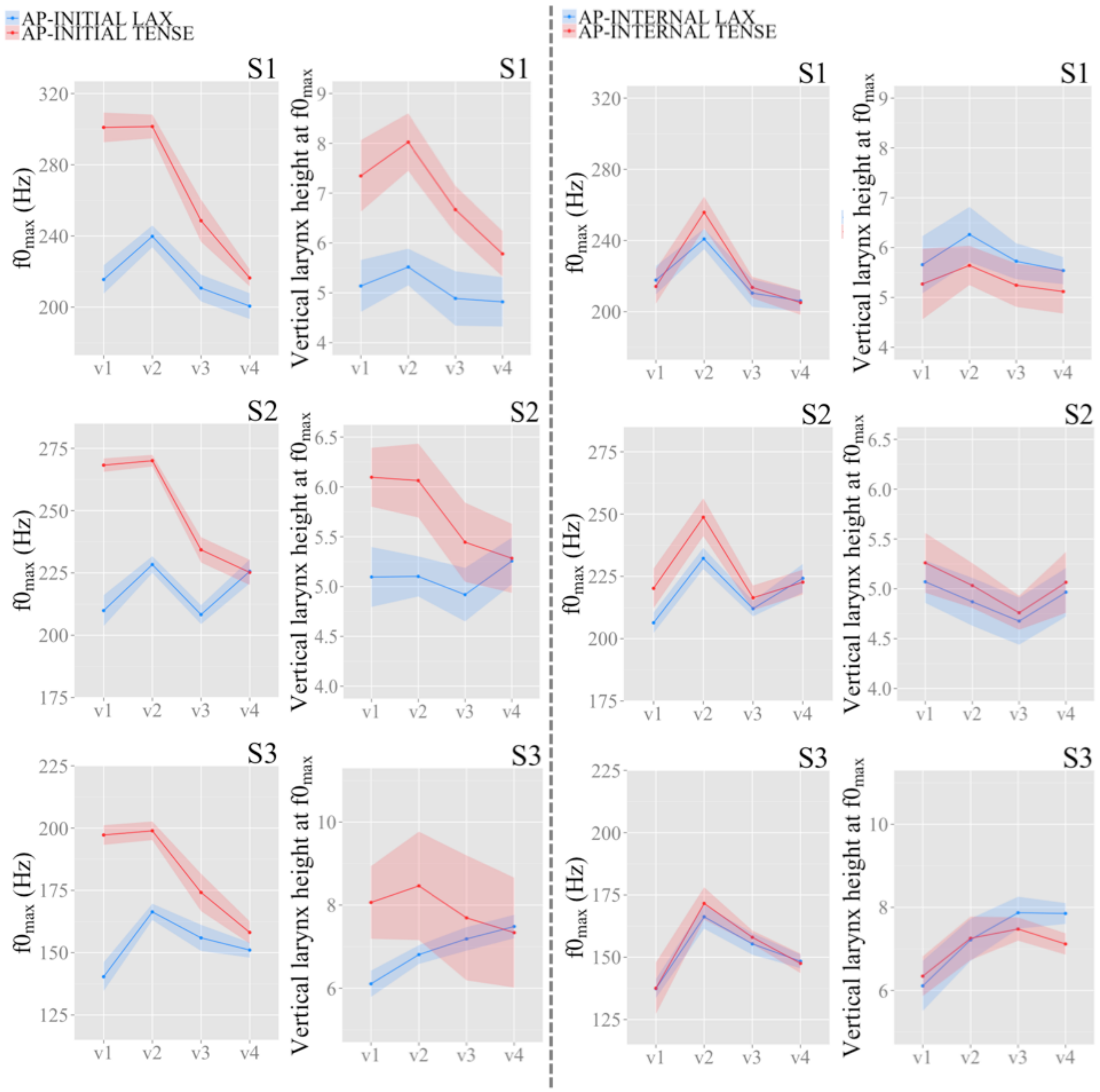

Figure 6. Mean values (presented as red and blue dots) and 95% confidence bands of $f0_{max}$ and vertical larynx height at $f0_{max}$ during each AP vowel. Relative to the dotted gray line down the center of the figure, the panels on the left side show the AP-INITIAL condition (LAX-INITIAL [blue] and TENSE-INITIAL [red] in v1); the right side panels show the AP-INTERNAL condition with LAX-INITIAL APs—i.e., Consonant Type (LAX [blue] vs. TENSE [red]) was manipulated in the AP-SECOND syllable (v2; all LAX elsewhere). Each row of panels represents an individual speaker's data (top: Speaker S1, middle: Speaker S2, & bottom: Speaker S3). Note that lines between one mean value (e.g., v1) to next (e.g., v2) are placed simply to illustrate the global pitch fluctuation of the whole AP, not to demonstrate any interpolating value.

Some observations can first be made from the figure above. Overall, speakers show similar phrase tonal ($f0_{max}$) patterns, particularly Speakers S1 and S3. The graphs show a clear tonal difference between TENSE versus LAX categories in many conditions: overall, higher f0 values are associated with TENSE than with LAX, whether AP-INITIAL (v1s of APs in the first column) or AP-INTERNAL (v2s of APs in the third column), except for Speaker S3. When a phrase starts (AP-INITIAL condition) with a syllable with a TENSE consonant, f0 values of the initial vowel (v1) are much higher than in the LAX condition. This f0 difference in the initial vowel seems to affect the f0 values through the following vowels of the phrase. For all three speakers, higher f0 values are associated with the phrase-internal syllables in the AP-INITIAL TENSE condition (v2, v3) when compared to the internal syllables in the AP-INITIAL LAX condition. For Speakers S1 and S3, the effect of the AP-initial consonant on f0 of non-initial syllables is also observed in the final syllable (v4). Speaker S2 does not show this pattern, presumably due to the high phrase-final f0. The canonical LHLH is observed only in Speaker S2's APs that start with a lax consonant (refer to the blue lines in AP-INITIAL LAX condition and both red and blue lines in AP-INTERNAL condition). For the AP-INTERNAL condition, the confidence bands for the f0 values between TENSE and LAX of the second vowel (v2) are overlapping completely for Speakers S3 and partially for Speaker S1, indicating that the consonantally triggered f0 difference is marginal.

Overall, the larynx height results conform to the f0 behaviors. The vertical larynx positions are higher for the AP-INITIAL TENSE condition (red dots for v1 the column 2) as compared to the AP-INITIAL LAX condition (blue dots for v1 the column 2). Importantly, larynx height does not seem to be affected by the local consonant type in the AP-INTERNAL conditions at all (v2 in the column 4). These are the phrase positions in which there is no consonantally triggered f0 difference for Speakers S1 and S3. Interestingly, the significant consonant effect on f0 found for Speaker S2 is not observed in her larynx height measure.

Interestingly, there seems to be *one single larynx raising/lowering* movement associated with an AP, based on a visual inspection of the time series data (the time series data analyses forthcoming). Fig. 7 shows a single vertical larynx movement during a lax-initial target AP $[\text{mi.mi.pi.pi}]_{\text{AP}}$ production.

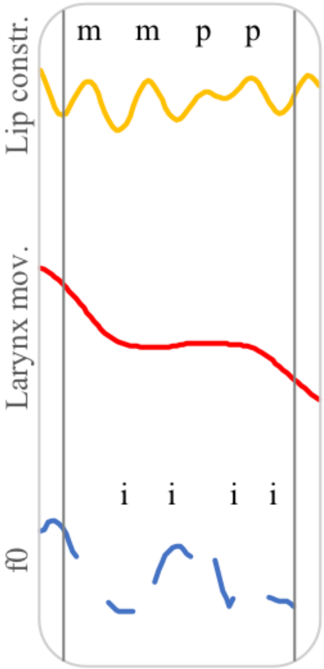

Figure 7. Example time functions of Speaker S1's lip constriction displacement, vertical larynx movement, and f0 during [mi.mi.pi.pi]AP.

The larynx comes down relatively continuously from an initial high position for the preceding tense-initial syllable of the AP '[sʌn.sæŋ.ni.mi]AP,' then exhibits a plateau until about the final lip constriction, and then finally descends for the subsequent lax-initial AP '[pi.ti.o.ɾɨl]AP.' Note that the spatiotemporal characteristic of the vertical larynx movement (e.g., the temporal location of the movement's peak, its displacement, the temporal size of a plateau, etc.) are likely to vary as a function of various factors well beyond the scope of this study. That said, the larynx height centroid values seem to reflect well the overall shape of larynx (vertical) action.

For all three speakers, the peak of the larynx raising movement occurs at the second syllable (v2) of the TENSE-INITIAL APs. Speaker S1 shows a similar pattern of the larynx vertical movement in the LAX-INITIAL APs. Speaker S3 shows the peak of larynx raising towards the latter part of the AP (v3 or v4) in the LAX-INITIAL APs. For Speaker S2, some larynx *lowering* movement is observed with LAX-INITIAL APs, with maximum lowering occurring during the third syllable.

A two-way ANOVA with Consonant Type (TENSE VS. LAX) and Phrase Position (AP-INITIAL VS. AP-INTERNAL) factors on the $f0_{max}$ and the corresponding larynx height values of the target syllables confirms the above inspection of the graphs (compare v1 values [AP-INITIAL] in the left half figure panel with v2 values [AP-INTERNAL] in the right half figure panel in Fig. 6). All three speakers show a significant main effect of Consonant Type on both $f0_{max}$ (S1: $F(1,44)=158.25$; S2: $F(1,44)=176.69$; S3: $F(1,44)=130.91$; all at $p<.001$) and on the corresponding larynx height (S1: $F(1,44)=7.76$, $p<.01$; S2: $F(1,44)=18.23$, $p<.001$; S3: $F(1,44)=10.94$, $p<.01$). F0 peaks are higher for the TENSE stops than for the LAX stops. The tense consonants are associated with higher positions of the larynx, and the lax consonants are associated with lower positions of the larynx. For Speakers S1 and S3, there are also some significant main effects of Phrase Position on these measures (S1: *AP-INITIAL $f0_{max}$ > AP-INTERNAL $f0_{max}$; $F(1,44)=6.19$, $p<.05$; S3: *AP-INITIAL larynx height > AP-INTERNAL larynx height; $F(1,44)=22.37$, $p<.01$). Crucially, however, all three speakers show a significant interaction between Consonant Type and Phrase Position (S1: $F(1,44)=78.09$, $p<.001$ for $f0_{max}$; $F(1,44)=78.09$, $p<.001$ for larynx height; S2: $F(1,44)= 24.64$, $p<.001$ for $f0_{max}$; $F(1,44)= 55.15$, $p<.001$ for larynx height; S3: $F(1,44)= 89.7$, $p<.001$ for $f0_{max}$; $F(1,44)=10.6$, $p<.01$ for larynx height). For both f0 and larynx height measures, the interaction is due mainly to the fact that there is an asymmetric effect of consonant type in different prosodic positions. The Consonant Type effect on f0 is more robust in AP-initial position (*TENSE > LAX, for all speakers) than in AP-internal position (*TENSE > LAX for Speaker S2; only marginal difference for Speaker S1 [$p=0.053$]; TENSE = LAX for Speaker S3). For larynx height, the Consonant Type effect is only found in the AP-INITIAL condition (*TENSE > LAX), not in the AP-INTERNAL condition (TENSE = LAX).

Phrasal (Tone [T]) and local Consonant Type effects on the following AP-tones: After confirming the strong consonant-type effect on f0 and larynx height in phrase-initial syllables (v1), we now turn to a consideration of the consonant and tone dynamics in *non*-initial syllables. Figs. 8-10 show the effects of the AP-initial tone (triggered by consonant type) and local manipulation of consonant type in AP-internal (v2, v3) and AP-final syllables (v4). (Note that the right half of Fig. 6 [AP-INTERNAL condition] is repeated as the left half of Fig. 8 [AP-SECOND condition].) A three-way ANOVA with Initial Tone (LAX-INITIAL VS. TENSE-INITIAL), Consonant

Type (LAX vs. TENSE), and Prosodic Position (AP-SECOND vs. AP-THIRD vs. AP-FINAL) reveals complex interactions of these factors in f0 and larynx height behaviors.

First, all three speakers show significant main effects of all three factors on $f0_{max}$. Overall, f0 values are higher for TENSE consonants than for LAX consonants (S1: $F(1,132)=99.68$; S2: $F(1,132)=115.91$; S3: $F(1,132)=39.72$, all at $p<.001$), f0 values are higher in the TENSE-INITIAL APs than in the LAX-INITIAL APs (S1: $F(1,132)=139.38$; S2: $F(1,132)=109.93$; S3: $F(1,132)=122.06$, all at $p<.001$), and f0 values are highest in the AP-SECOND vowel, intermediate for AP-THIRD vowels, and lowest for the AP-FINAL vowels (S1: $F(2,132)=135.2$; S2: $F(2,132)=73.91$; S3: $F(2,132)=65.41$, all at $p<.001$). For the larynx height measure, speakers show somewhat varied results. All three speakers have a significant main effect of Initial Tone on the vertical larynx position (S1: $F(1,132)=58.42$; S2: $F(1,132)=31.63$; S3: $F(1,132)=50.83$, all at $p<.001$). The higher larynx height is associated with the TENSE-INITIAL APs compared to the LAX-INITIAL APs. Speakers S1 and S2 show a significant main effect of Prosodic Position on the vertical larynx positions (S1: $F(2,132)=37.82$, $p<.001$; S2: $F(2,132)=10.4$, $p<.001$), which, as was the case with f0, was the highest for the AP-SECOND vowel, intermediate for AP-THIRD vowels, and the lowest for the AP-FINAL vowels. A significant main effect of Consonant Type on this measure is found only with Speaker S2 (*TENSE > LAX, $F(1,132)=4.45$, $p<.05$).

The complex f0 and the larynx height patterns shown in Figs. 8-10 can be explained by different interactions among factors. For all three speakers, there is a significant Initial Tone x Prosodic Position interaction in both measures (S1: $F(2,132)=16.75$, $p<.001$ for $f0_{max}$, $F(2,132)=6.08$, $p<.001$ for larynx height; S2: $F(2,132)=17.88$, $p<.001$ for $f0_{max}$, $F(2,132)=14.07$, $p<.01$ for larynx height; S3: $F(2,132)=26.08$, $p<.001$ for $f0_{max}$, $F(2,132)=4.09$, $p<.05$ for larynx height). This interaction term may indicate the temporal scope of a consonantally triggered phrase-initial tone effect that occurred for all speakers on AP-SECOND (v2) and AP-THIRD (v3) vowels but not on AP-FINAL vowels (v4). Specifically, phrase-medially both f0 and the corresponding vertical larynx position are higher for the TENSE-INITIAL APs than for the LAX-INITIAL APs, but this (register) difference disappears phrase-finally. This can be visually confirmed in Figs. 8-10, comparing the left side figure panels (LAX-INITIAL APs) with the right side figure panels (TENSE-INITIAL APs).

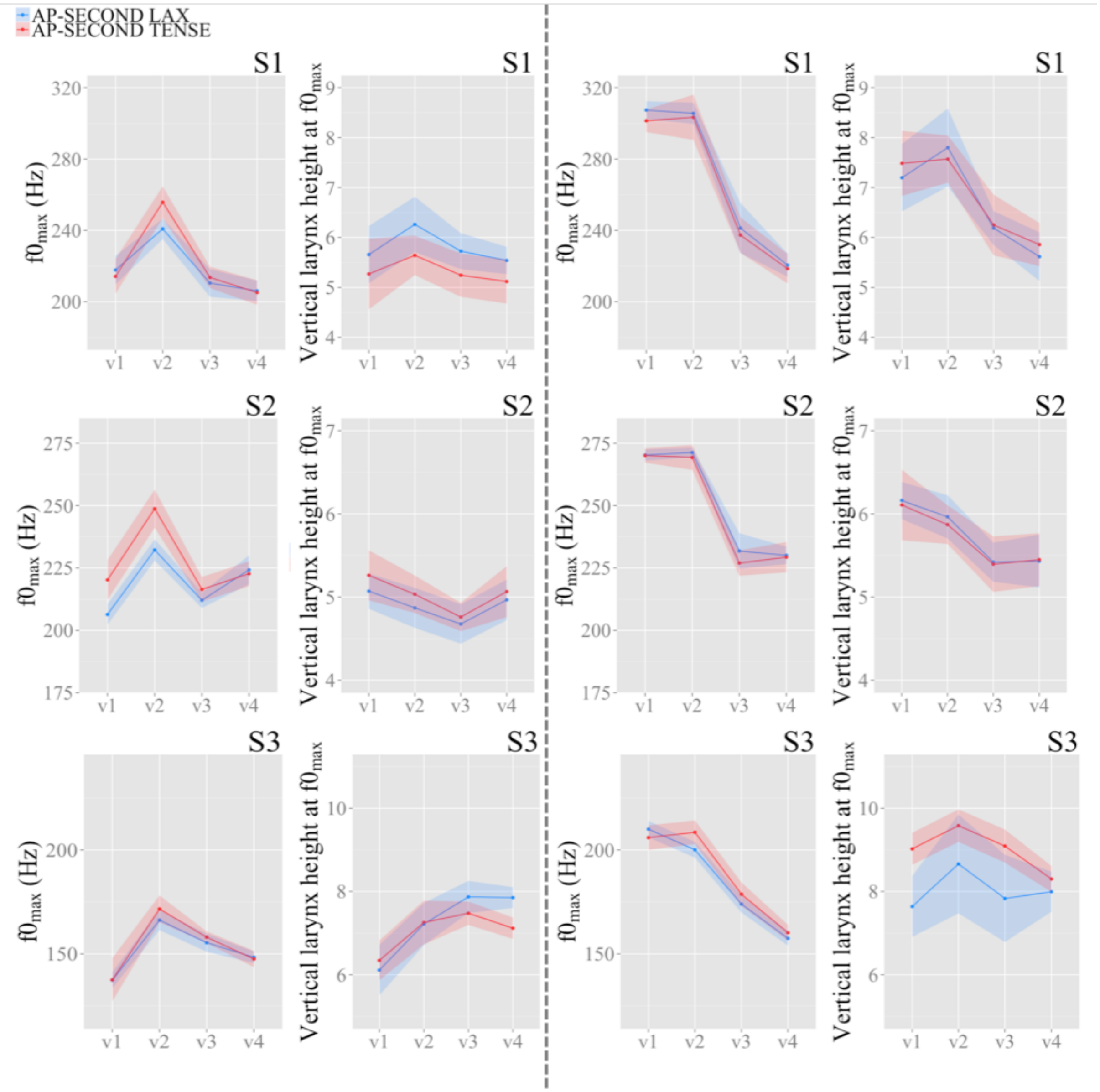

Figure 8. Mean values (presented as red and blue dots) and 95% confidence bands of $f0_{max}$ and vertical larynx height at $f0_{max}$ during each AP vowel in the AP-SECOND condition. Consonant Type (LAX [blue] vs. TENSE [red]) was manipulated in the second syllable (v2). Relative to the dotted gray line down the center of the figure, the left side panels show the LAX-INITIAL condition and the right side panels show TENSE-INITIAL condition. Each row of panels represents an individual speaker's data (top: Speaker S1, middle: Speaker S2, & bottom: Speaker S3); Note that lines between one mean value (e.g., v1) to next (e.g., v2) are placed to simply illustrate the global pitch fluctuation of the whole AP, not to demonstrate any interpolating value.

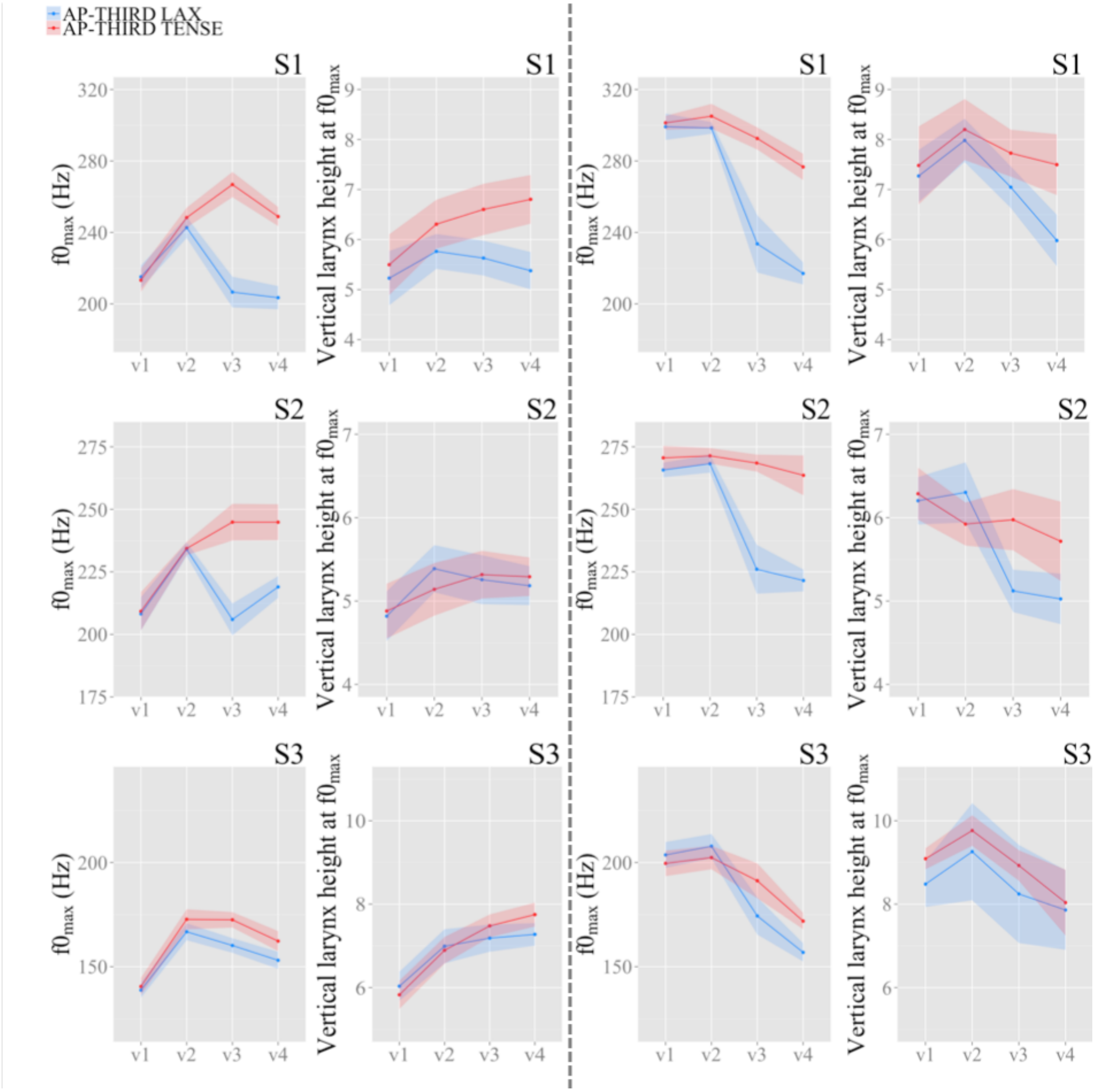

Figure 9. Mean values (presented as red and blue dots) and 95% confidence bands of $f0_{max}$ and vertical larynx height at $f0_{max}$ during each AP vowel in the AP-THIRD condition. Consonant Type (LAX [blue] vs. TENSE [red]) was manipulated in the phrase-penultimate syllable (v3). Relative to the dotted gray line down the center of the figure, the left side panels show the LAX-INITIAL condition and the right side panels show TENSE-INITIAL condition. Each row of figure panels represents an individual speaker's data (top: Speaker S1, middle: Speaker S2, & bottom: Speaker S3); Note that lines between one mean value (e.g., v1) to next (e.g., v2) are placed simply to illustrate the global pitch fluctuation of the whole AP, not to demonstrate any interpolating value.

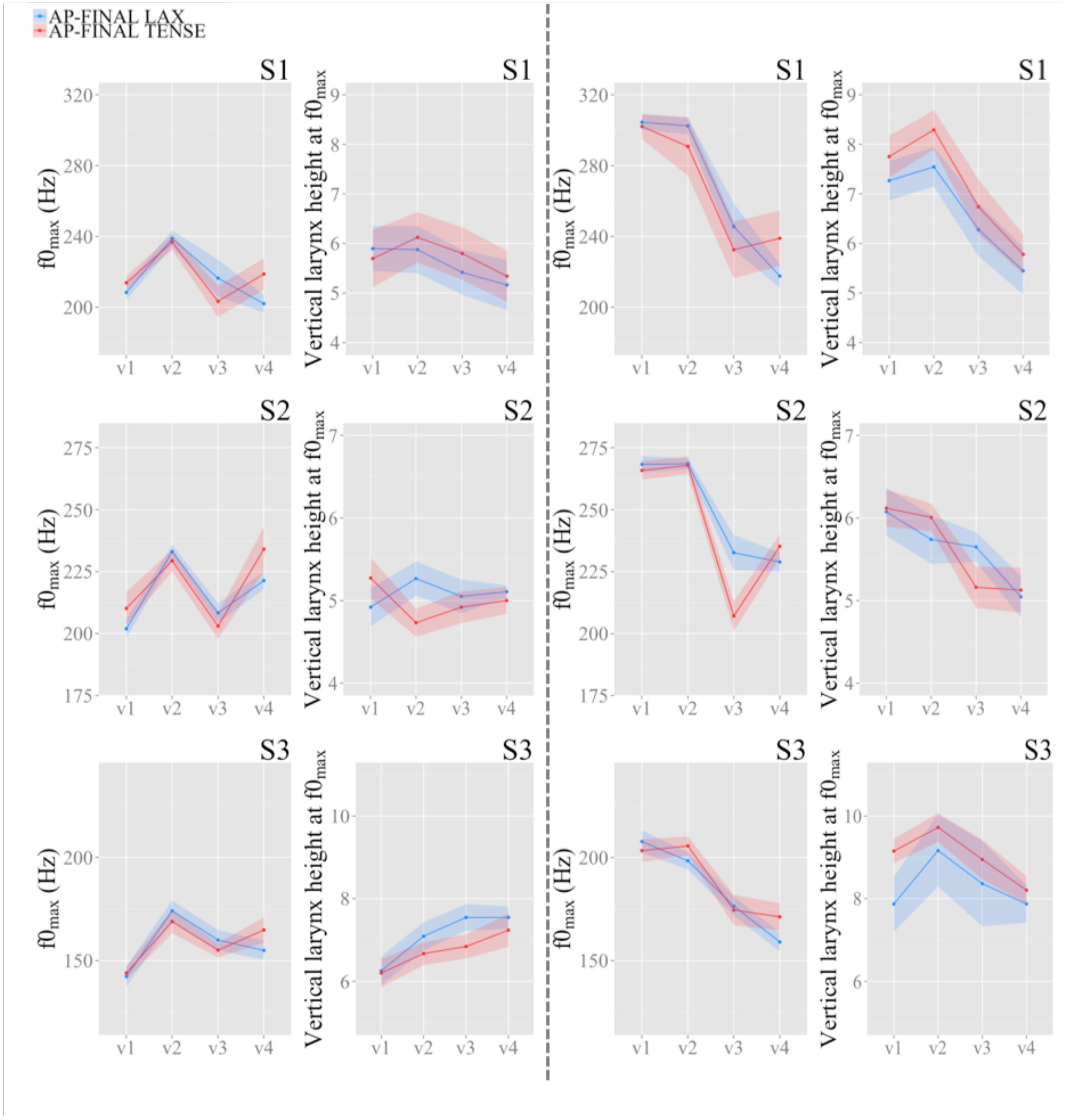

Figure 10. Mean values (presented as red and blue dots) and 95% confidence bands of $f0_{max}$ and vertical larynx height at $f0_{max}$ during each AP vowel in the AP-FINAL condition. Consonant Type (LAX [blue] vs. TENSE [red]) was manipulated in the phrase-final syllable (v4). Relative to the dotted gray line down the center of the figure, the left side panels show the LAX-INITIAL condition and the right side panels show TENSE-INITIAL condition. Each line of figure panels represents an individual speaker's data (top: Speaker S1, middle: Speaker S2, & bottom: Speaker S3); Note that lines between one mean value (e.g., v1) to next (e.g., v2) are placed simply to illustrate the global pitch fluctuation of the whole AP, not to demonstrate any interpolating value.

For Speakers S1 and S2, there is a significant Consonant Type and Prosodic Position interaction on both f0 and larynx height (S1: *F*(2,132)=32, *p*<.001 for f0; *F*(2,132)=6.08, *p*<.01 for larynx height; S2: *F*(2,132)=36.73, *p*<.001 for f0; *F*(2,132)=3.89, *p*<.05). This interaction reveals an asymmetric effect of the local consonant type manipulation in different prosodic positions (see Figs. 8-10 for these two speakers). There is no consonant effect on f0 and larynx height in the AP-SECOND condition (Fig. 8 the top and middle rows, but cf. the leftmost figure panel for S2). On the other hand, in the AP-THIRD condition both f0 and larynx height are higher for tense than for lax (Fig. 9 the top and middle rows, but cf. the second figure panel of S2). For the AP-FINAL condition, both speakers show a significant Consonant effect on f0 (*TENSE > LAX) but do no longer exhibit the same effect on the larynx height measure (Fig. 10 the top and middle rows).

The two exceptions in Speaker S2's data that do not conform to the overall patterns in Figs. 8-10 are accounted for by a significant 3-way interaction between factors (f0: *F*(2,132)=3.2, *p*<.05; larynx height: *F*(2,132)=4.16, *p*<.05). There is indeed a significant Consonant Type effect on f0 in AP-second syllables but only when an AP starts with a lax consonant (*TENSE > LAX, Fig. 8 the leftmost figure panel for S2), not a tense consonant (TENSE = LAX). The interaction further reveals that a significant consonantally induced larynx height difference in the AP-THIRD syllable is observed only in the TENSE-INITIAL APs (*TENSE > LAX), but not in the LAX-INITIAL APs (compare v3s of the second and rightmost figure panels for Speaker S2 in Fig. 9). Finally, Speaker S2 shows a significant two-way interaction between Initial Tone and Consonant type on f0 (*F*(2,132)=4, *p*<.05), due to the fact that the tense versus lax difference is larger in the LAX-INITIAL APs (*TENSE > LAX, mean diff. 23 Hz) than in the TENSE-INITIAL APs (*TENSE > LAX, mean diff. 16 Hz).

3.3.2. Lip closing kinematic measures

3.3.2.1. Phrase-initial lip closing kinematics in /mi/#/Ci/

For the AP-initial consonants, speakers show different lip closing kinematic patterns. Speaker S1 shows a significant main effect of Consonant in both measures (constriction duration: *F*(3,44)=8.14, constriction maximum: *F*(3,44)=7.8, all at *p*<.001). A post-hoc Tukey test indicates that the AP-initial FORTIS is produced with the longer constriction duration and greater constriction maximum compared to the AP-initial NASAL, LENIS, and ASPIRATED stops (*FORTIS > NASAL, LENIS, ASPIRATED; see Fig. 11).

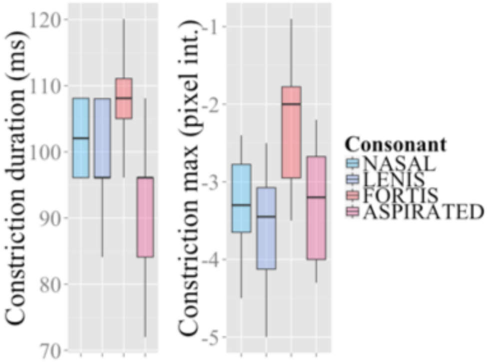

Figure 11. Main effect of Consonant (NASAL, LENIS, FORTIS, and ASPIRATED) on constriction duration and maximum constriction (Speaker S1).

For Speaker S2, a significant main effect of Consonant is found with the constriction maximum (*F*(3,44)=3.18, *p*<.05). In AP-initial position, the fortis stop shows a greater maximum constriction value compared to the lenis stop (*FORTIS > LENIS; see Figure 12).

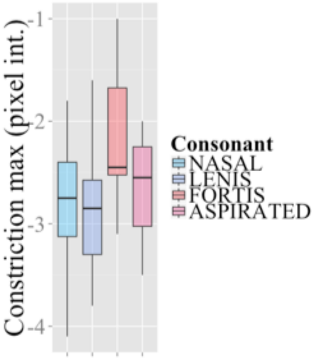

Figure 12. Main effect of Consonant (nasal, lenis, fortis, and aspirated) on maximum constriction (Speaker S2).

There is no Consonant effect on lip closing constriction duration for this speaker (*F*(3,44)=2.45, *p*=.076). Speaker S3 does not show any significant differences in the initial consonant kinematics.

3.3.2.2. Lip closing kinematics in ALL-LAX and ALL-TENSE APs

Constriction degree: For all three speakers, the LAX versus TENSE distinction is most consistently made with the constriction degree measure—i.e., constriction maximum value. Fig. 13 shows the main effect of Consonant on this measure found with the three speakers.

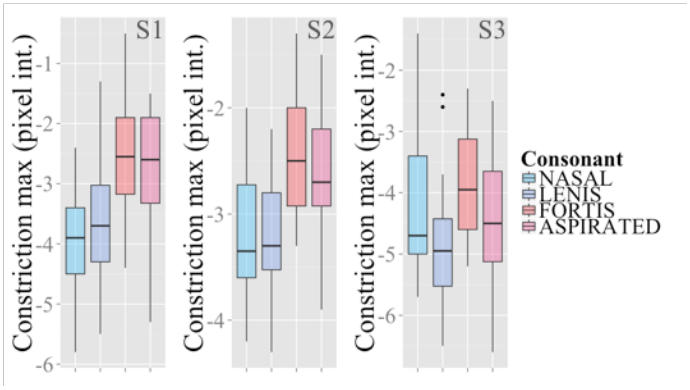

Figure 13. Main effect of Consonant (NASAL, LENIS, FORTIS, and ASPIRATED) on maximum constriction (Speakers S1, S2, & S3).

Speaker S1 shows a significant main effect of Consonant (*F*(3,80)=12.47, *p*<.001). Both tense consonants are produced with greater constriction than the lax consonants (*FORTIS, ASPIRATED > NASAL, LENIS, the left panel of Fig. 13). There is no significant main effect of Prosodic Position (*F*(3,80)=2.64, *p*=.055) or interaction between factors (*F*(9,80)<1). For Speaker S2, both factors show significant main effects (Consonant: *F*(3,80)=13.41, *p*<.001; Prosodic Position: *F*(3,80)=6.8, *p*<.001). As was the case with Speaker S1, both FORTIS and ASPIRATED stops are associated with the greater constriction when compared to the NASAL and LENIS stops (the middle panel of Fig. 13). The positional effect on the constriction maximum shows that the AP-INITIAL and AP-SECOND stops are produced with the greater constriction than the AP-THIRD and AP-FINAL stops (the left panel of Fig. 14). No interaction between factors is found (*F*(9,80)=1.39, *p*=.21).

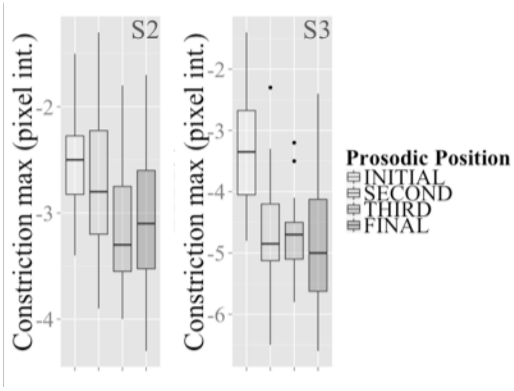

Figure 14. Main effect of Prosodic Position (AP-INITIAL, AP-SECOND, AP-THIRD, and AP-FINAL) on maximum constriction (Speakers S2 & S3).

Speaker S3 also shows a significant main effect of both factors (Consonant: *F*(3,73)=3.87, *p*<.05; Prosodic Position: *F*(3,73)=10.6, *p*<.001). The fortis stop is more constricted than the lenis stop (the right panel of Fig. 13), and the AP-initial stops are more constricted than the stops in the other phrase positions (the right panel of Fig. 14). There is no interaction between factors (F(9,73)<1).

Constriction duration: For the constriction duration measure, there is weak evidence for LAX and TENSE distinction. Speaker S1 shows significant main effects of Consonant and of Prosodic Position (*F*(3,80)=3.13, *p*<.05; *F*(3,80)=6.78, *p*<.001). The FORTIS stop is associated with the longer duration when compared to the NASAL stop (the left panel of Fig. 15).

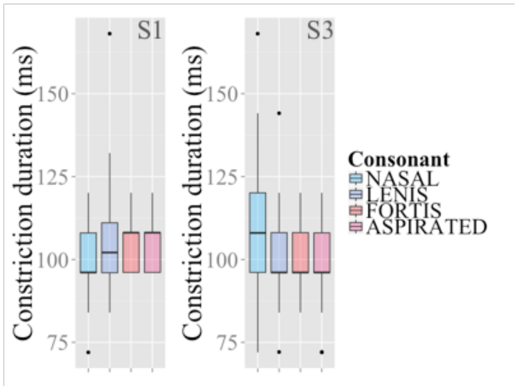

Figure 15. Main effect of Consonant (NASAL, LENIS, FORTIS, and ASPIRATED) on constriction duration (Speakers S1 & S3).

For this speaker, the longer constriction duration is observed in the phrase-internal stops compared to the phrase-edge stops (*AP-SECOND, AP-THIRD > AP-INITIAL, AP-FINAL; see the left panel of Fig. 16).

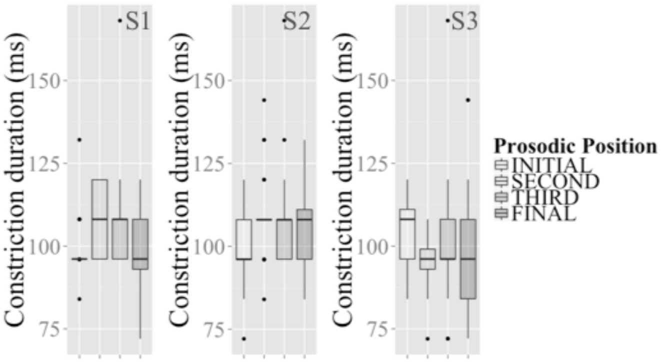

Figure 16. Main effect of Prosodic Position (AP-INITIAL, AP-SECOND, AP-THIRD, and AP-FINAL) on constriction duration (Speakers S1, S2, & S3).

A significant interaction between factors (*F*(9,80)=2.88, *p*<.01) for Speaker S1 further reveals that the Prosodic Position effect is only observed with the LENIS stop (*AP-SECOND LENIS > AP-FINAL LENIS, figure not given). Speaker S3 shows a significant main effect of both factors (Consonant: *F*(3,73)=2.93, *p*<.05; Prosodic Position: *F*(3,80)=3.24, *p*<.05). For this speaker, the constriction duration is longer for the NASAL stop than for the ASPIRATED stop (the right panel of Fig. 15). The AP-INITIAL stops are longer than the AP-SECOND stops (the right panel of Fig. 16). A significant interaction between factors (*F*(9,73)=4.77, *p*<.001) reveals that the NASAL stop is longer than the FORTIS stop in AP-penultimate position, and is longer than the ASPIRATED stop in AP-final position (figure not given). Recall that this speaker has several missing data points for the nasals in these phrase positions, leaving only few, perhaps atypical, nasal tokens. For Speaker S2, there is no Consonant effect (*F*(3,80)=1.77, *p*=.16), but there is a significant main effect of Prosodic Position (*F*(3,80)=3.72, *p*<.05). For this speaker, the AP-SECOND stops are longer than with the AP-INITIAL stops (the middle panel of Fig. 16). There is no interaction between factors (*F*(9,80)=1.46, *p*=.18).

3.4. Discussion

This study of consonant and tone dynamics employing the real-time MRI technique presents novel findings of an intricate interaction between the lexical tones of [tense/lax] and the prosodic structure—i.e., phrase-level prosody. The goal of the study was to investigate what motor tasks are deployed for consonantal "tenseness" and tone gestures, and how they function within the phonological system. In order to address these questions, we examine Seoul Korean, as segmental tone and non-flexible phrase tone patterns are both present in this language.

This investigation provides novel evidence for (a) articulatory mechanisms that express f0 and tenseness and (b) the interplay between different phonological structures that deploy these mechanisms. The primary finding is that while the individuals show somewhat varying patterns in the phonetic measures (i.e. f0, the vertical larynx motions, and the lip closing kinematics), nevertheless, several crucial points emerge from both general (cross-speaker) patterns and individual speaker-specific patterns, illuminating an intricate interplay between the lexical contrast maintenance and syntagmatic tonal patterns.

As shown in the previous chapter, in the newly emerging phonetic system of Seoul Korean stops, the consonant (tense/lax) effect on f0 is categorical in AP-initial position but gradient in AP-internal position. Given the asymmetric positional effect on f0, we entertained two competing hypotheses about the responsible articulatory gestures. Hypothesis a postulated that the categorical versus gradient f0 differences arise from a single f0 task, whose major contributing articulatory system is vertical laryngeal movement. In contrast Hypothesis b postulated that this prosodic asymmetry results from two different types of gestural (laryngeal) tasks—specifically, a tenseness gesture or task, whose main articulator action is larynx raising and a phrase tone f0 gesture or task, whose main articulator action is stretched vocal folds.

The results demonstrate that all three speakers show a strong positive correlation between the f0 and vertical larynx position, suggesting that the vertical larynx movement is certainly engaged in some tonal manipulation. At face value, this may seem consistent with the previous reports on cine-MRI evidence for a tense and lax distinction in larynx height, showing a significant difference in both word-initial and word-internal stops (tense > lax, Kim et al. 2005, 2010). However, our results further reveal that there is much more going on beyond this simple relation.

For all three speakers, the consonant-type effect on f0 and a corresponding larynx height difference (i.e. tense > lax) is confirmed in Accentual Phrase-initial position. This suggests that the segmental tenseness that is expressed by f0 may result from the larynx height manipulation in this prosodic position. This phrase-initial tenseness carried out in the laryngeal and f0 settings is observed to affect tonal aspects in the following syllables of the AP. The second and penultimate syllables of the AP are produced with the higher f0 *and* higher larynx position in a tense-initial quadrisyllabic AP than in a lax-initial quadrisyllabic AP. This finding confirms a broad temporal scope of the AP-initial tone (*T*).

The study also provides evidence that the consonantally triggered initial *T* effect on f0 may be achieved by a *more global* vertical larynx position adjustment. Visual inspection of the vertical larynx movement indicates one single raising or lowering larynx movement integrated for one single AP. The categorical AP-initial effect of consonant type on f0 (tense-initial AP > lax-initial AP) may be viewed as a tonal *register* difference, similar to the finding of the role of larynx height in expressing the register difference in Cantonese tones (Nissenbaum 2008, 2010).

In non-initial positions across speakers, f0 and larynx height patterns do not always match. However, speakers show variability in their individual prosodic patterns. One of the speakers (S3) manifests the consonant-type effect on both f0 and larynx height only in AP-initial position but not in AP-medial or AP-final positions. In contrast, the other two speakers (S1 & S2) show a more complex interaction between phonological factors in these measures. For these speakers, the consonant-type effect on f0 (tense > lax) is also asymmetric in different phrase-*medial* positions (AP-second vs. AP-penultimate positions). In general, the tense versus lax distinction is not reflected in the f0 and larynx height measures in the AP-second syllable. (Recall, however, that one exception comes from Speaker S2, indeed exhibiting a significant consonant-type effect on f0 [tense > lax], but not on the larynx height [tense = lax].) In AP-penultimate position, both speakers show the consonantally triggered f0 difference in both f0 *and* the larynx height measures (AP-third tense > AP-third lax). This AP-internal prosodic asymmetry can possibly be due to the presence of the word boundary before the penult that coincides with this prosodic position in these stimuli. The fact that there is a strong effect on both f0 and larynx measure indicates that this position behaves like an 'intermediate' or weak AP boundary. (Perhaps even suggestive of a recursion in AP structure yielding nesting [Byrd 2002].) This suggests that there may be an active vertical larynx movement expressing tenseness that is specifically found in word-initial position. As shown in the top two rows in Fig. 9, there seems to be an active larynx raising movement (for lax-initial APs) or maintenance of the larynx posture (for tense-initial APs). The degree of executing this task or the 'strength' of this gesture may vary as a function of phrase position. The only exception to this is the second figure panel of Speaker S2—the lax-initial APs, in which the phrase-internal word-initial consonant-type effect on larynx height (active larynx raising) is not observed. This suggests that for this speaker the larynx raising movement that expresses the word-*initial* consonantal tenseness may be suppressed by the low phrasal register setting.

It is further worth noting that both Speakers S1 and S2 have a significant consonant-type effect on f0 (tense > lax) in AP-final position but do not exhibit the same effect on their corresponding larynx height measure (Fig. 10 the top and middle rows). Similar patterns are also observed in the AP-internal conditions for Speaker S2, exhibiting a significant consonant-type effect on f0 (tense > lax), but *not* on the larynx height (tense = lax) (see the second figure panels of S2 in Figs. 8-9). These results may indicate that some other articulatory maneuver, possibly the vocal fold stretching, must play out to account for the f0 variations, perhaps supporting Hypothesis b. Alternatively, it might be simply the case that since the larynx height data are noisier than f0 (overall wider confidence bands for larynx height than f0 in figures), the actual difference in the vertical larynx movement is comparably very small and not detectable with our current statistical power. Future work on vocal fold tension may elucidate these possibilities.

In the previous chapter, we have identified several phonological biases that synergistically work together to shape the AP tones: 1) the underlying LH(LH) tonal shape of an AP, 2) the f0 of the previous syllable, 3) consonant type, and 4) the invariant phrase-final tone. By examining more exhaustive test materials, in this chapter we confirm the consonant type effect and its temporal scope through an AP (i.e. the bias coming from the previous syllable). However, we obtain some mixed results for the oscillatory pattern and the invariant tonal target designated for the AP-final tone. The LHLH phrasal patterns are observed only in some conditions of Speaker S2, not with Speakers S1 or S3 at all. (Note also that this speaker occasionally shows larynx lowering for lax-initial APs, which is left for future study on phonological factors shaping spatio-temporal characteristics of vertical larynx movement plateaus.) Overall, the phrase-final high tone is not observed (cf. consonant effect on f0 in this position; cf. also Jun 1993). The absence of the consistency here is perhaps due to the specificity in the stimuli design. In order to make the phonologically controlled target phrase as natural as possible, it was designed to form a single syntactic object phrase with the subsequent phrase. The subsequent phrase—i.e., the phrase following the target AP—was always lax-initial and thus may involve some active lowering of the larynx, perhaps suppressing the preceding AP-final tone from reaching its target. To what degree the final syllable "constraint" is actually due to the following AP remains to be clarified.

Finally, the results of this study also provide evidence for the lax versus tense distinction in the lip articulatory kinematics (nasal vs. lenis vs. fortis vs. aspirated). Overall, there is the consistent consonant-type effect on constriction degree in the ALL-LAX and ALL-TENSE APs. The distinction between tense and lax is most consistently made with the constriction maximum measure. Inter-speaker variability is also observed here. For two speakers (Speakers S1 & S2), both fortis and aspirated stops are associated with greater constriction when compared to the nasal and lenis stops. For Speaker S3, a significant consonant effect is found with fortis and lenis stops (fortis stop is more constricted than the lenis stop). For the constriction duration measure, there is only weak evidence for lax and tense distinction found with one speaker (fortis > nasal for Speaker S1). These results suggest that the tense and lax may have different target constriction degrees, which is partially in line with the previous findings (Cho et al. 2016; Son et al. 2012).

Similar but less consistent results are obtained for the tense-lax distinction for AP-initial lip kinematics (in /mi/#/**C**i/). Speaker S1 shows the consonant distinction in both kinematic measures (fortis > nasal, lenis, aspirated). For Speaker S2, a significant AP-initial consonant effect is found with the constriction maximum measure (fortis > lenis). Speaker S3 does not show any consonant effect. Null or weak effect of consonant on the initial kinematics may be due in part to certain AP positional effect on overall production of the stops. For Speakers S2 and S3, consonants are more constricted at the beginning of an AP (S2: AP-initial, AP-second stops < AP-penultimate, AP-final; S3: AP-initial stops < AP-second, AP-penultimate, AP-final). In AP-initial position, the lax consonants may have been produced more strongly thereby blurring the tense versus lax distinction in that position. In any case, the tenseness distinction is clearly made in the tonal regime in initial position.

3.5. Conclusions

In sum, this study of consonant and tone dynamics utilizing the real-time MRI technique examines an intricate interaction between the lexical tones of [tense/lax] and the prosodic structure—i.e., phrase-level prosody. Using as a test bed Seoul Korean, in which segmental and relatively fixed phrase tone patterns are both present in the prosodic system, we provide new evidence for what motor tasks are deployed for consonantal “tenseness” and tone gestures, and how they function within the phonological system. The findings provide novel evidence for (a) articulatory mechanisms that express f0 and tenseness and (b) the interplay between different phonological structures that deploy these mechanisms. Both general (cross-speaker) patterns and individual speaker-specific patterns serve to illuminate an intricate interplay between the lexical contrast maintenance and syntagmatic tonal patterns.

Acknowledgements

This work was supported by NIH/NIDCD DC007124 (Shrikanth Narayanan).

4. Prosodic conditioning in sound change

4.1. Introduction

Speech is tremendously variable, but at the same time, the variation is often lawfully structured. The parallelism observed between such synchronic speech variation and diachronic change in sound patterns has led to phonetic discoveries of importance to sound change (Ohala 1974, et seq.). Naturally occurring phonetic precursors of sound change can be conditioned by physiological and auditory properties of speech. Sound change may be driven by listeners' misinterpretation or imperfect learning of coarticulated variants (the hypo-corrective sound change [Ohala 1993]; 'choice' in evolutionary phonology [Blevins 2006]), or by listeners' unnormalized, but correct, perception of the variants in a hypo-speech situation where the intelligibility demands are low (application of hyper- and hypo-articulation [H&H] theory to sound change [Lindblom 1990; Lindblom, Guion, Hura, Moon, and Willerman 1994]). Alternatively, it can also be the case that listeners may be consistently sensitive to, or actively use, the informative acoustic consequences of coarticulated variants (Beddor 2009). Although different theoretical accounts of sound change hypothesize different sources or triggering mechanisms for diachronic language change, the sources of sound change claimed in these theories are fundamentally grounded in (co)articulatory variation in the speech signal.

The goal of this chapter is to shed light on some salient issues in sound change, using the data presented in Chapters 2 and 3, as unique examples of variation and sound change in *progress*, focusing on identifying the intrinsic variation in speech across phonological contexts with an eye to understanding how the phonetic information of sound changing categories is (being) structured in the speakers' production. This dissertation targets one of the most widely attested sound change patterns, the development of phonological tone—'tonogenesis' following Matisoff (1973)—which is rarely observed in progress. Phonologically contrastive tones arise from unintentionally distorted effects on the intrinsic fundamental frequency (f0) perturbation owing to segmentally grounded variation (Edkins 1864; Haudricourt 1954, 1961; Hyman 1973; Matisoff 1973; Ohala 1973, *inter alia*). In general, a voiceless oral obstruent produces a higher variant of f0 on the following vowel, whereas its voiced counterpart produces a lower variant of f0 on the following vowel (e.g., Hombert 1978). One of the apparent physiological causes of this effect comes from the role of longitudinal vocal fold tension, which has been experimentally confirmed by electromyography studies (Löfqvist, Baer, McGarr, and Story 1989; Dixit and Macneilage 1980). These studies have shown that longitudinal tension in the larynx plays a role in controlling consonant voicing. Voiceless obstruents are produced with a higher level of contraction activity of the cricothyroid muscles in the larynx, which creates increased longitudinal tension that may extinguish voicing, together with abduction. This increased longitudinal tension due to cricothyroid activity is directly involved in f0 regulation, particularly for the production of high tones, whether lexical or intonational. In contrast, such increase in longitudinal tension is absent in the production of voiced obstruents.

As sketched in Table 1, the concomitant f0 perturbation in the vowel following a consonant is predicted originally as an automatic consequence of the difference in longitudinal vocal fold tension. The intrinsic consonantal effect on f0 then eventually becomes part of the pronunciation norm, i.e., *phonologized*. Here, the term 'phonologization' is interpreted as an explicitly controlled property of non-contrastive low-level phonetic modifications. The f0 reinterpretation is subsequently accompanied by the de-phonologization (often referred to as 'transphonologization' [Hagège and Haudricourt 1978]) of voicing—that is, voicing is neutralized as the voiced obstruent undergoes devoicing (Hyman 1976), as illustrated in the 'contrastive effect' column in Table 1. This *reanalysis* by the learner is often construed as voicing becoming redundant information, while previously redundant f0 newly obtains quasi-phonological status.

Table 1. A sketch of a phonologization account of tonogenesis (after Hyman 1976).

voicing contrast	redundant effect	contrastive effect
/**p**á/	/**p**á/	/p**á**/
/**b**á/	/**b**ǎ/	/p**ǎ**/

Although many issues regarding this type of word-level (phonemic) phonologization have been extensively investigated in the literature, phonologization in connection with high-level prosodic aspects of the linguistic structure has been largely overlooked. Speakers' phonological knowledge regarding the speech system of their language extends well beyond simply distinguishing meanings. When speakers produce words embedded in larger phrasal structures, prosodically appropriate adjustments are made. There has been little work that explicitly discusses the possible role of prosodic structure in sound change in the context of its progression (but cf. recent work by Cole and colleagues).

Acknowledging that sound change is based on the systematic variation in speech and that there is a pervasive effect of the global *prosodic* structure on low-level articulatory behavior, Cole and colleagues have argued that some examples of sound change are, or at least can be, initially developed or 'seeded' in prosodically salient positions such as a stressed or accented vowel or preceding a phrase boundary. Cole, Hualde, Blasingame, and Mo (2010) tested if prosodic prominence interacts with vowel shift, that is, whether vowels in prominent positions are more advanced along the direction of the shift in acoustic space. In the Chicago dialect of American English, lexically stressed vowels such as /ɛ/ as in "*pet*" and /ʌ/ as in "*cut*" have become more retracted since the late 1970s (Labov 1994). Cole et al. (2010) found that younger generations of this dialect exhibit no further retracting but rather consistent *lowering* of these vowels under prominence, perhaps from a larger jaw opening associated with a higher-level prominence. The vowels are lowered more under contrastive focus, intermediate under broad focus, and less in post-focal position. They argue that the vowel shift might appear first in the most prominent positions and then spread to positions of lesser prominence, further speculating that the lowering effect of prominence following the completion of one kind of shift (i.e., retracting) might be a new change arising for this vowel system.

Cole and Hualde (2013) make a parallel argument for *boundary-conditioned* sound changes. In German, for example, obstruents are devoiced at the end of a word (see example (I)). Hock (1991) analyzes devoicing of final obstruents as an assimilation to following pause in utterance-final position. Devoicing then may be extended to obstruents in word-final position through an analogical process of leveling that does not differentiate contexts any longer.

(I) Examples of German final obstruent devoicing

Ra[d] > *Ra*[t]	'wheel'
das Ra[d] > *das Ra*[t]	'the wheel'
das Ra[d] *ist* > *das Ra*[t] *ist*	'the wheel is'

As illustrated in Table 2, Cole and Hualde generally agree with Hock's analysis on the German final obstruent devoicing as sound change involving two distinct processes, i.e., sound change in the original triggering context (followed by silent pause) and its analogical extension in the other contexts in which the trigger is absent. In other words, the extension (generalization) of the phonetic modification going beyond its originating contexts is considered as the first step of the phonologization process. However, they argue for a more gradual process of extension, assuming an intermediate stage of sound change conditioned by phrasal prosody—that is, the context of final obstruent devoicing extends gradually from being adjacent to silent pause, to being adjacent to phrasal boundary (whether or not followed by silence), to being adjacent to word boundary.

Table 2. Gradual extension processes of German final devoicing (Cole and Hualde 2013).

	sound change	analogy (I)	analogy (II)
obstruent > -voice	/__]SILENCE	/__]PHRASE	/__]WORD

These studies by Cole and colleagues nicely serve to illuminate a role of prosody in sound change, but there is still much to be learned. Although Cole and Hualde (2013) have reviewed many cases of boundary-conditioned sound change in various languages, the cases were limited only to the left edge of a phrase boundary. Prosodic boundary-induced phonetic modulations can be observed in the neighborhood of *both* edges of a prosodic boundary (i.e. finally and initially in a prosodic domain), resulting from a prosodic (π-) gesture that operates on the temporal unfolding of an utterance *in the vicinity of* the phrase boundary (Byrd and Saltzman 1998, 2003). In parallel with prominent syllables or phrase-final material, the initial position of a phrase boundary is also prosodically salient. This has been referred to variously as strengthening and/or lengthening and has been observed in both acoustic and articulatory domains (e.g., Byrd, Kaun, Narayanan, and Saltzman 2000; Cho and Keating 2001; Fougeron and Keating 1997; Pierrehumbert and Talkin 1992). We believe that a systematic investigation of examples of sound change in which the global phrasal structure interacts with information restructuring in sound-changing categories will provide a clearer picture for assessing prosodic conditioning in sound change.

The studies presented in this dissertation explicitly examine a case of sound change that has been argued to be limited to only the beginning of larger prosodic phrasal contexts, having initial strengthening as a driving force (see e.g. Chapters 2-3). This newly emerging stop system of the contemporary Seoul dialect of Korean offers an excellent test case for examining prosodically conditioned tonogenesis in progress, as the observed patterns of an ongoing tonogenic sound change are systematically influenced by higher-level phrasal prosodic contexts.

In Seoul Korean, a phonetic organization in the 3-way stop contrast (i.e. lenis in /**p**ul/ "*fire*", aspirated in /$\mathbf{p}^{\mathbf{h}}$ul/ "*grass*", fortis (or tense) in /$\mathbf{p*}$ul/ "*horn*") in phrase-initial position has been described as produced with distinctive combinations of voice onset time (VOT) and f0 values associated with the following vowel (e.g., Cho, Jun, and Ladefoged 2002; Cho and Keating 2001; Kang and Guion 2008; Lee and Jongman 2012; Lisker and Abramson 1964). And it has been consistently documented that the f0 reanalysis is active in younger generation speakers, that is, those born during or after the 1980s (Bang, Sonderegger, Kang, Clayards, and Yoon 2018; Kang 2014; Lee and Jongman 2012). There is a complete loss of VOT distinction between AP-initial lenis (previously intermediate in VOT) and aspirated (previously the longest VOT) stops. Compare 'VOT' columns under 'old system' vs. 'new system' in Table 3. Along with the VOT mergers, the tonal distinction among phrase-initial stops has become sharper over time (compare old vs. new 'f0' columns in Table 3).

Table 3. VOT and f0 reanalysis in the Seoul Korean 3-way voiceless stop contrast, lenis /**p**/, aspirated /$\mathbf{p}^{\mathbf{h}}$/, fortis /$\mathbf{p*}$/, in phrase-initial syllables. Different color codes indicate the tenseness distinctions, '**LAX**' vs. '**TENSE**,' and yellow shading indicates where changes have occurred.

	old system		new system	
	VOT	f0	VOT	f0
#/pa/	mid	low	long	low
#/p^ha/	long	slightly higher		much higher
#/p*a/	short		short	

What we observe in phrase-initial position is the reduction of informativeness along one dimension (i.e. initial VOT mergers between lenis [lax] and aspirated [tense] stops), accompanied by the enhancement of a previously redundant dimension (i.e. f0 difference between lax and tense categories). A frequently adopted interpretation for such 'trade-off' reorganization is that the change serves to maintain (or enhance) the phonological contrast (Keyser and Stevens 2001, among others). The enhancement of f0 in the Korean stop series has been viewed as an AP-initial specific phonologization of f0, consequent to an initial strengthening mechanism that enhances the laryngeal contrast [tense] in prosodically salient position (Bang et al. 2018; Cho and Lee 2016; Jun 1993; Kirby 2013).

Adding further complexity to the picture, Seoul Korean exhibits an intonationally defined phrasal level, referred to as the Accentual Phrase (AP), that has been described as being associated with an alternating phrasal tonal sequence *T*HLH. Crucially, the initial tone (*T*) interacts with segmental quality (Jun 1993). This initial *T* is H if the AP-initial segment is a tense consonant (including fortis and aspirated stops) and L elsewhere. The non-initial phrase positions have been reported to be insensitive to the segmental quality (cf. or less sensitive; see Cho and Lee 2016).

Prior to this dissertation's investigation (Chapters 2-3), no study examined the younger generations' production of the stop contrast in *non*-initial position. In earlier-generation speakers, the contrast between fortis (shortest VOT) and lenis stops in phrase-*internal* position is maintained through intervocalic lenis voicing (/a**p**a/ often becomes [a**b**a]; e.g., Han and Weitzman 1970; Kagaya 1974; Kim 1965). However, it was unknown how or if the phonetic realization of the contrast in AP-*internal* position has changed as the VOT merger and enhanced f0 difference has played out in AP-initial position. This prosodic context (i.e. AP-internal) is intriguing to look at because it is the exact context where phrasal prosody may exhibit its intricate interaction with the lexical tone (i.e. the consonantally triggered f0). Our investigation tested the hypothesis that the local f0 reanalysis is conditioned by the global phrasal tonal patterns, which entails that the phrasal prosody can also affect the realization of phonetic information (i.e. VOT, voicing, closure duration) that has the trade-off relation with f0.

As discussed in Cole and Hualde (2013), if the new change, that is, the augmented tonal distinction between lax and tense, is observed beyond its originating contexts—extending from phrasally prominent syllables to less prominent syllables—it might indicate that a phonologization process has occurred and progressed. Bang et al. (2018) shows that the VOT and f0 organization among Intonational Phrase-initial stops propagates across words (having different frequencies) and vowel (height) contexts. Such evidence for a gradual extension of the f0 reanalysis further supports that f0 has come to function as a reliable indicator of the lexical contrast, specifically in the younger generation's dialect of this language. If the reorganization is shown not to be strictly restricted in phrase-initial position, this sound change is potentially extending from an original strengthening environment (initial position) to a prosodically weak(er) environment through leveling. A magnitude difference in the information reinterpretation in different phrasal units (Intonational Phrase [IP]-initial AP vs. IP-medial AP) may be an indication of the future progress of a gradual extension of sound change (from phrasally prominent syllables to less prominent syllables). No difference may indicate a completed process of extension, as the new sound pattern no longer differentiates these prosodic contexts. The study in Chapter 2 tests exactly this possibility. In what follows we summarize the main findings of the previous chapters and discuss their implications for the nature of ongoing tonogenic sound change in contemporary Seoul Korean.

4.2. The newly emerging stop system in contemporary Seoul Korean

This dissertation probes an essential role of phrasal prosody in consonant and tone interaction. In the two previous chapters, we show how the f0 of the vowel following LAX (NASAL /m/, LENIS /p/) and TENSE (FORTIS /p*/, ASPIRATED /p^h/) stops is organized in different prosodic positions for younger generation speakers (born 1980-1990). We find that the consonant effect on f0 is *categorical* in AP-INITIAL position, compared to exhibiting a *gradient* effect in AP-INTERNAL position. AP-INITIAL LAX and TENSE stops show virtually no overlap in the distribution of f0 values between the two categories (Fig. 1 left panel). In AP-INTERNAL position, we find a significant though small f0 difference (overlapping distributions; Fig. 1 right panel).

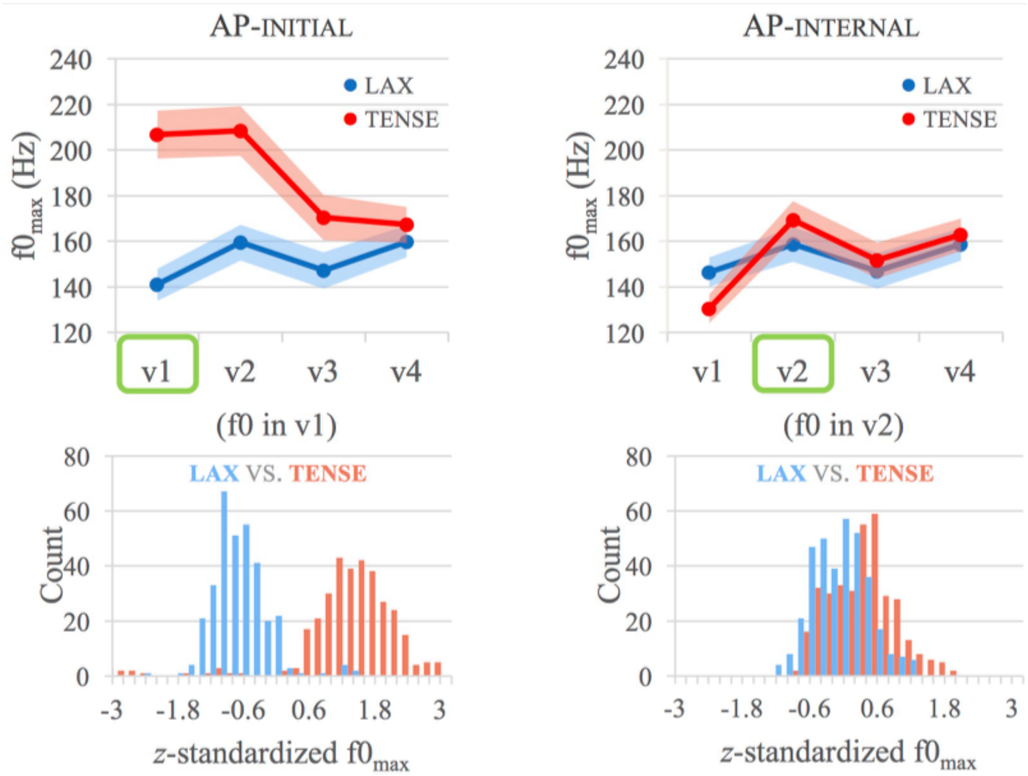

Figure 1. Consonant-type (LAX vs. tense) effect on AP-INITIAL (v1) vs. AP-INTERNAL (v2) f0.

These asymmetric f0 differences found in different prosodic positions may function to maintain or augment contrast among stop categories that exhibit VOT mergers and near-mergers. Our results confirm the recently reported AP-initial VOT merger between aspirated /p^h/ and lenis /p/ stops (Bang et al. 2018; Kang 2014; Lee and Jongman 2012). We found the near-merger of VOT between the AP-*internal* lenis /p/ and fortis /p*/ stops, arising from the substantially reduced occurrence of intervocalic lenis voicing in younger generation speakers. See Fig. 2.

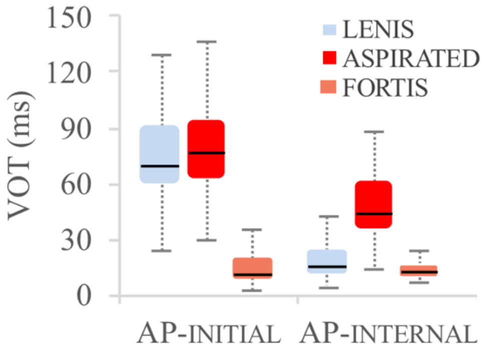

Figure 2. VOT values for AP-INITIAL vs. AP-INTERNAL LENIS /p/, ASPIRATED /p^h/, FORTIS /p*/.

As discussed at the outset of the chapter, the phonologization (stabilization) process of one phonetic property often goes hand in hand with the de-phonologization of another. Our finding of an unequal effect of prosodic location on the information reorganization reveals that stabilization of phonetic distinctions is differently conditioned by phrasal position. In phrase-initial position, the three stops are distinctive from each other in terms of the VOT and f0 combination (See Table 4 for a summary).

Table 4. 3-way voiceless stop contrast, lenis /**p**/, aspirated /**pʰ**/, fortis /**p***/, in *AP-initial* syllables. Different color codes indicate the tenseness distinctions, '**LAX**' vs. '**TENSE**,' and yellow shading indicates where changes have been made.

	VOT	f0
#/pa/	long	low
#/pʰa/		much higher
#/p*a/	short	

In AP-internal position, the occurrence of intervocalic lenis voicing (/apa/ → /aba/) is substantially reduced for our speakers (i.e. younger generation), compared to the data from the older generation reported in Jun (1993). This shift led to a new case of VOT merger. Along with the near merger of VOT, the small effect of the consonant type on f0 in phrase-internal position is accompanied by a large closure duration difference in this position. Table 5 summarizes how the phonetic properties of the AP-internal three-way contrast are organized in younger Seoul Korean speakers.

Table 5. 3-way voiceless stop contrast, unvoiced lenis /**p**/, fortis /**p***/, aspirated /**pʰ**/, in *AP-internal* syllables contemporary Seoul Korean. Different color codes indicate the tenseness distinctions, '**LAX**' vs. '**TENSE**.'

	VOT	closure duration	f0
/apa/	short	short	low
/ap*a/		long	higher
/apʰa/	long		

Our findings suggest an adaptive extension of (re)organization in phonetic properties of the stop system. It is worth highlighting that the f0 reanalysis among speakers is not uniform in both prosodic positions. Across speakers a systematic inverse relation is observed between the magnitude of the VOT difference and the accompanying f0 difference in AP-initial position. This type of relation is not observed between the initial VOT difference and the amount of the internal closure voicing, suggesting that the change is not proceeding exactly in parallel in AP-initial and internal positions. (This is discussed in detail in Chapter 2). That said, the information reanalysis can be considered to be adaptive, favorably serving to maintain or enhance the contrast, rather than making every phonetic aspect maximally distinctive. The frequency of lenis voicing also varies greatly by speaker, which again suggests that the progression of sound change is adaptive.

Finally, this dissertation identifies phonological factors that shape the AP tones. The positional asymmetry suggests that the f0 value of the *preceding syllable* is a major determinant of the f0 value of the current syllable within the Korean AP. We propose that the underlying tonal pattern of an AP is a (repeating) LH(LH) sequence and that there exist phonological biases in the system that shape the surface tonal patterns: a) the f0 of the preceding syllable, b) consonant type, and c) an invariant phrase-final tone. Based on the temporal scope of the effects observed, we assume the bias coming from the previous syllable to be stronger than the bias coming from the consonant type. These identified biases can also account for how the information reorganization in different prosodic locations is shaped or constrained. Taken together, our findings of prosodic conditioning in information (re)organization have implications for an intricate interplay between paradigmatic consonant contrast maintenance and syntagmatic tonal patterns.

4.3. Extension of consonantally triggered tone across prosodic positions

Prosodic conditioning, including prominence- and boundary-related conditioning (e.g., Cole et al. 2010; Cole and Hualde 2013), in sound change have implications for understanding prosodic variation as a source of sound change. The work presented in this dissertation extends this idea to the context of sound change in which constraints for preserving paradigmatic and syntagmatic contrasts may be simultaneously active in the phonology. In Seoul Korean, the tonal contrast between the tense and lax stops is paradigmatically enhanced in a prosodically strong, AP-initial position (Bang et al. 2018; Cho and Jun 2000; Jun 1998; Kang 2014). However, as our findings show, in AP-internal position there is an interplay between preserving the consonantally derived tonal contrast and realizing the global (syntagmatic) tonal patterns characterizing an AP. Such findings suggest that the local phonetic (re)organization is systematically modulated by phrase-level prosody, illuminating the role of prosodic conditioning in sound change.

Chapter 2 also tests whether there is evidence for gradual extension in sound change by looking at the effects of boundary strength on phonetic properties of the contemporary Seoul Korean stops. The results show no effect of prosodic boundaries on the information reanalysis. No effect of boundary strength on VOT or f0 may simply indicate no effect of an IP. However, previous studies have shown that stronger articulation is associated with IP-initial position as compared to (IP-medial) AP-initial position (e.g., linguopalatal contact and VOT, IP-initially > AP-initially, Cho and Keating 2001), which suggests that IP-initial position may be the originally triggering context of the VOT merger. A lack of prosodic boundary effects suggests that the analogical extension from the original context, IP-initial position (strengthening context) to a non-originating context, AP-initial position (also strengthening context but perhaps with a smaller magnitude) may have been completed, with no distinction any longer made by younger generation speakers between different phrase-levels. Our results parallel the previous findings of gradual extension in sound change in phrase-final position (Cole and Hualde 2013). We speculate that the f0 reanalysis has become a stable phonetic mark of phrase-initial position in this language (IP-initial > AP-initial). However, there is a possibility that the distinction in different boundary strengths is still made in qualities of the articulatory closure, as this data was not available for this study. Regardless, the extension of f0 reanalysis within an AP is systematically conditioned by phrasal prosody that regulates the surface tonal patterns.

Further supporting evidence of the extension comes from a significant consonant-type effect on f0 *and* larynx height in AP-internal word-*initial* position (i.e. AP-third tense > AP-third lax in the penult of a quadrisyllabic AP; see Chapter 3), which seems to behave like an ‘intermediate’ or weak AP boundary. According to the phrasal oscillatory tone pattern *T*H**L**H, the penult, just like the AP-second syllable, is also supposed to be a tone-neutralizing context. However, the penult being a word-initial syllable creates another type of prosodic asymmetry (AP-medial word-internal vs. AP-medial word-initial).

This brings to mind many examples of sound change, which are highly productive in languages, showing certain phonetic properties that emerge spontaneously at IP boundaries being generalized to occur at "word" boundaries. For example, the case of word-final devoicing has been extensively studied in the experimental literature in utterance-final position (e.g., in German, Cole and Hualde 2013; Hock 1991; in Russian, Hock 1991; Hyman 1978, Steriade, 1997, *inter alia*). A word-final voiced obstruent is often realized as voiceless in phrase-medial position as well as phrase-final position, although only the latter is a phonetic context for devoicing. This type of shift in the domain of a distribution pattern from a more inclusive prosodic domain (e.g., an utterance or IP) to less inclusive one (e.g., an utterance- or phrase-medial word) can be viewed as a type of 'overphonologization,' in the sense that generalization is applied to the domains without the phonetic underpinnings ("boundary narrowing" by Hyman 1978; "domain generalization" by Myers and Padgett 2014).

Hualde (2013) argues that overphonologization is a natural diachronic process and therefore should be equally relevant to word-*initial* generalizations. An example of initial boundary narrowing is found in blocking of the application of a phonological process in western Romance. In Judeo-Spanish, the intervocalic stop within a word is changed to a fricative, as in *sa**b**er* > *sa**v**er* "*to know*." Hualde argues for suppression of lenition in word-initial position; a word-initial voiced stop is preserved even when intervocalic, as in *la **b**oka* "*the mouth,*" due to overphonologization of an utterance-initial requirement (e.g., utterance-initial ***b**oka* "*mouth*") to (phrase-medial) word-initial position.

In the case of contemporary Seoul Korean, some syntagmatic constraints are imposed by phrasal prosody, which is evident from the reorganization of phonetic properties of stops in the AP-second (word-medial) syllables. While the phrasal constraint is present, our findings also reveal that the phrase-initial consonant-type effect on tone has been extended to phrase-medial *word*-initial position, in which this phrasal constraint seems to be overridden.

4.4. Conclusions

This study investigates how the phonetic information of sound changing categories (lenis /p/, aspirated /p^h/, fortis /p*/) is (re)organized in speakers' production of contemporary Seoul Korean. As this language exhibits a phrasal level phonology—an Accentual Phrase (AP) that interacts with local (segmental and tonal) properties—we hypothesized an essential role of phrasal prosody in terms of the information restructuring associated with the contemporary Seoul Korean stop system. We found supporting evidence for this hypothesis. The observed patterns of an ongoing tonogenic sound change are systematically influenced by higher-level phrasal prosodic contexts. Moreover, our results of the progression of the sound change across prosodic contexts suggest a further interaction of phrasal prosody with lexical word boundary in terms of information reorganization. Taken together, our findings have implications for enhancing the field's understanding of the complex role that prosodic conditioning can play in sound change.

Acknowledgments

This work was supported by NIH DC03172 (Dani Byrd) and the project: "Tonal Placement - The Interaction of Qualitative and Quantitative Factors: TOPIQQ," subcontract D-71631-Z-600-145002301 from the University of Cologne, under funding from the Volkswagen Foundation.

5. Summary and conclusions

This dissertation primarily considers how speakers integrate prosodic information with speech gestures while planning and producing words and phrases. It investigates the complex interaction in the prosodic dynamics of consonant and tone, probing an essential role of phrasal prosody in spoken language production. The investigation of this complex integration contributes to our understanding of prosodically conditioned variability not only by providing the phonetic description needed for the new pronunciation norms emerging in younger generations speakers of this language, but also because the observed tonal patterns provide new data for testing a gestural theory of intonation.

This work consists of two empirical, acoustic and articulatory, studies of contemporary Seoul Korean (Chapters 2-3). The results suggest that the local phonetic organization of the consonant system is systematically modulated by the language's prosodic structure. The unique phrasal prosodic system of this language (i.e. non-flexible phrasal tone pattern *T*HLH for the Accentual Phrase [AP]; e.g., Jun 1993) serves as an excellent testing ground for the questions addressed in this work: (a) how the local phonetic organization (i.e. fundamental frequency [f0], voicing, voice onset time, [VOT], closure duration) of the contrastive stop system—lenis /p/ [lax], aspirated /p^h/ [tense], fortis /p*/ [tense]—is modulated as a function of accentual phrasal positions, (b) what motor tasks are deployed for tone and segmental "tenseness" gestures, and how they function within the phonological system, (c) what phonological factors shape the AP tones, and (d) what role phrasal prosody plays in sound change.

The acoustic and articulatory investigations in this dissertation provide an explanation for how phonological factors combine to shape the phrasal tone realization. And they systematically illuminate the patterning of phonetic information for sequences containing varying consonant types [tense/lax] placed across several phrasal positions. The consonant type effect on f0 (tense > lax) is *asymmetric* in different locations within an AP. The effect is substantially larger in AP-initial position (tense >> lax) than in AP-internal position (i.e. in the second syllable of *T***H**LH, a tone-neutralizing context) position (tense > lax). There is a non-overlapping bi-modal distribution of f0 values in AP-initial position, dividing tense- and lax-induced f0 values into two discrete modes. This type of tonal grouping is not observed in AP-internal position. The results suggest that the consonant type effect on f0 is *categorical* in AP-initial position, while *gradient* in AP-internal position.

Given the asymmetric positional effect on f0, we entertained two competing hypotheses about the responsible articulatory gestures or tasks. Hypothesis (*a*) postulated that the categorical versus gradient f0 differences arise from a single f0 task, whose major contributing articulatory system is vertical laryngeal movement. In contrast Hypothesis (*b*) postulated that the prosodic asymmetry results from two different types of gestural (laryngeal) tasks—specifically, a tenseness gesture or task, whose main articulator action is larynx raising, and a phrase tone f0 gesture or task, whose main articulator action is stretched vocal folds.

The results demonstrate that there is a strong positive correlation between the f0 and corresponding vertical larynx position, suggesting that the vertical larynx movement is certainly engaged in some tonal manipulation. At face value, this may seem consistent with the previous reports of cine-MRI evidence for a tense and lax distinction in larynx height, showing a significant difference in both word-initial and word-internal stops (tense > lax, Kim, Honda, and Maeda, 2005; Kim, Maeda, and Honda 2010). However, our results further reveal that there is much more going on beyond this simple relation.

The global f0 *and* larynx height patterns are dramatically different between an AP starting with a lax consonant versus an AP starting with a tense consonant (tense > lax). This suggests that the segmental tenseness that is expressed by f0 may result from the larynx height manipulation in AP-initial position. While a lax-initial AP consistently shows the canonical *L*HLH pattern, a tense-initial AP shows the non-canonical surface patterns such as *H*HHL or *H*HLL. When the consonant type is manipulated phrase-internally, the resulting f0 and larynx height differences are locally constrained, conforming to the overall pattern of the entire phrase. The results suggest that the temporal scope of the consonant effect on f0 is broad in AP-initial position, but small in AP-internal position. In AP-initial position, *both* initial *H* and initial *L* of an AP exert an influence on the following syllables, in fact up to the third syllable of the AP. The articulatory findings provide novel evidence that this consonantally triggered initial *T* effect on f0 may be achieved by a *more global* vertical larynx position adjustment, expressing a tonal *register* difference (Nissenbaum 2008, 2010).

In AP-internal (second syllable) position, the scope of the consonantally derived tonal difference is rather limited to the target syllable and its preceding syllable at best. It is worth recalling that there is a small but significant consonant-type effect on f0 (tense > lax) in this prosodic position but no effect on the corresponding larynx height measure. Similar patterns are also observed in the AP-final conditions, exhibiting a significant consonant-type effect on f0 (tense > lax), but not on the larynx height (tense = lax). These results likely indicate that some other articulatory maneuver, possibly vocal fold stretching, must play out to account for the f0 variations, supporting Hypothesis (*b*). Alternatively, it might be simply the case that since the larynx height data are noisier than f0, the actual difference in the vertical larynx movement is comparably very small and not detectable with our current statistical power. Future work on vocal fold tension may elucidate these possibilities.

In addition, a significant consonant-type effect on f0 *and* larynx height is observed in AP-internal word-*initial* position (i.e. AP-third position tense > AP-third position lax in the penult of a quadrisyllabic AP), which seems to behave like an 'intermediate' or weak AP boundary. According to the phrasal oscillatory tone pattern *T*H**L**H, the penult, just like the AP-second syllable, is also supposed to be a tone-neutralizing context. However, the penult being a word-initial syllable (in this study) in fact creates another type of prosodic asymmetry (AP-medial word-internal vs. AP-medial word-initial). There appears to be an active larynx raising movement (for lax-initial APs) or maintenance of the larynx posture (for tense-initial APs) in this position, suggesting that there may be an active vertical larynx movement expressing tenseness that is specifically found in word-initial position. While the phrasal constraint is present, our findings also reveal that the phrase-initial consonant-type effect on tone extends to phrase-medial *word*-initial position, in which the phrasal constraint seems to be overridden.

This dissertation also sheds light on some salient issues in sound change using the current findings as unique examples of variation and sound change in progress. Overall, the observed patterns of an ongoing tonogenic sound change in both general cross-speaker patterns and individual speaker-specific patterns are systematically influenced by higher-level phrasal prosodic contexts. Along with the positional asymmetry in consonant-type effect on tone, we confirm the recently reported AP-initial VOT merger between aspirated /p^h/ and lenis /p/ stops (Bang, Sonderegger, Kang, Clayards, and Yoon 2018; Kang 2014; Lee and Jongman 2012). Additionally, we found a near-merger of VOT between the AP-*internal* lenis /p/ and fortis /p*/ stops, arising from the substantially reduced occurrence of intervocalic lenis voicing in younger generation speakers. An adaptive extension of (re)organization in phonetic properties of the stop system is observed. Across speakers a systematic inverse relation is observed between the magnitude of the VOT difference and the accompanying f0 difference in AP-initial position. This type of relation is not observed between the initial VOT difference and the amount of the internal closure voicing, suggesting that the change is not proceeding exactly in parallel in AP-initial and internal positions.

Moreover, our results of the progression of the sound change across prosodic contexts suggest a further interaction of phrasal prosody with lexical word boundary in terms of information reorganization. These findings suggest that constraints for preserving paradigmatic and syntagmatic contrasts are simultaneously present and active in the phonology of younger speakers of Seoul Korean. This dissertation has taken a step toward enhancing the field's understanding of the complex role that prosodic conditioning can play in sound change.

Based on the results, this dissertation identifies how phonological factors combine to shape the AP tones. The positional asymmetry suggests that the f0 value of the *preceding syllable* is a major determinant of the f0 value of the current syllable within the Korean AP. We propose that the underlying tonal pattern of an AP is an oscillating LH(LH) sequence and that there exist phonological biases in the system that shape the surface tonal patterns; these are: a) the f0 of the preceding syllable, b) consonant type, c) an invariant phrase-final tone, and d) lexical-word boundary. Based on the temporal scope of the effects observed, we assume the bias coming from the previous syllable to be stronger than the bias coming from the consonant type. These identified biases can also account for how the information (re)organization in different prosodic locations is shaped or constrained.

A dynamical approach is well suited for this project because ongoing sound change can be understood as an example of adaptation from one stable, cognitively effective state to another, as such showing the evolution in the intricate interaction playing out over generational time. This dissertation employs a dynamical model to capture the quality of the interaction between phrasal prosody and consonant contrast realization (see Chapter 2). An applicable complementary approach is offered by the dynamical grammar framework of Gafos and Benuš (2006), as within this approach the parallelism between quantitative and qualitative (categorical) effects can emerge in a principled way. In this grammar, the possible contrastive values of a property such as f0 are modeled as the attractors of a nonlinear dynamical system. Phonological constraints are modeled as biases that can shift the probability of the occurrence of the two modes, and simultaneously shift their modal values.

In the case of Seoul Korean tones, there are two categorical attractor states, low and high f0 along the f0 continuum (under the assumption that f0 is the state space). With the above identified phonological constraints, this grammar successfully predicts the actual data distribution. If the dynamical system has two distinct attractor states (e.g., L vs. H), a bimodal distribution of values along the dimension is predicted, one mode corresponding to each contrasting category. In order to account for the context-determined selection of modes (e.g., the prosodic control variable in this study), the dynamical system can be biased in the direction of one or another of the modes. In the case of the phrase-initial position, the consonant-type biasing factor is the strongest, as there is no bias coming from a preceding syllable. Therefore, the [tense] versus [lax] factor selects H versus L tone in initial position, functioning as a phonological contrast. In the case of the phrase-internal position, however, a bias coming from consonant type is simultaneously present with a stronger bias—the f0 of the preceding syllable. In this case, the [tense] versus [lax] biases can still function to shift f0 values quantitatively in the presence of the stronger bias. This initial modeling result supports the central claim that a unified, non-dualistic theoretical model that combines, rather than maps between, phonetics (physical) and phonology (cognitive) can capture and clarify surface prosodic variation. Future modeling work should integrate into this picture the findings of the two articulatory mechanisms that are responsible for surface tonal patterns (i.e. larynx raising for tenseness gesture and stretched vocal folds for a phrase tone modulation gesture).

Finally, this dissertation invites another theory to account for the observed tonal patterns that exhibit the interaction between global and local structures. Goldsmith (1994) extended the dynamic computational model developed by Goldsmith and Larson (1990) to metrical phonology to show how the global structure of prominence profile emerges from local interactions between syllables. This application has been used to model alternating patterns of stress or sonority profile in language, showing how successive grammatical units enter into gradient relationships with their neighbors. It will be illuminating to employ the application of this dynamical model for modeling the current findings, given that the Seoul Korean AP exhibits an oscillating sequence of tones (e.g., most commonly *L*HLH) that further interact with the phonological biases that this dissertation has identified—i.e., neighboring (preceding) tone, consonant type, and lexical-word boundary.

In sum, building on the insights of previous work on prosodically conditioned variability, this dissertation extends the examination of prosodic conditioning to a consideration of the interaction of lexical (i.e. segmental) tone and phrasal (i.e. accentual) tone. Both acoustic and articulatory studies presented here illuminate an intricate interaction between phrase-level prosody and the local phonetic reorganization in the newly emerging phonetic system of Seoul Korean stops. Our findings of prosodic conditioning in information (re)organization have implications for an intricate interplay between paradigmatic consonant contrast maintenance and syntagmatic tonal patterns. Taken together, this dissertation provides novel evidence for the seamless integration of segmental and suprasegmental phonological structure and contributes to our understanding of the complex orchestration of articulatory gestures as they are woven into the prosodic substrate of spoken language.

Appendices

Once we reduced down to an ideal model from a full model, in all cases, the absolute value of the *t*-statistic that exceeds 2 was considered significant (Baayen 2008).

Appendix 1: Closure duration models.

CD$_1$ Intercept: AP-INITIAL, ASPIRATED			
Factor level	Estimate (ms)	Standard Error	*t*-value
(Intercept)	88.74	2.8	29.91*
Phrase Position: AP-INTERNAL	10.74	2.4	4.48*
Stop Item: LENIS	-21.39	2.42	-8.85*
AP-INTERNAL x LENIS	-26.11	3.29	-7.94*
CD$_2$ Intercept: AP-INITIAL, FORTIS			
Factor level	Estimate (ms)	Standard Error	*t*-value
(Intercept)	96.04	3.37	28.49*
Phrase Position: AP-INTERNAL	17.91	2.57	6.97*
Stop Item: LENIS	-33.73	2.58	-13.09*
AP-INTERNAL x LENIS	-33.35	3.5	-9.54*
CD$_3$ Intercept: AP-INITIAL, ASPIRATED			
Factor level	Estimate (ms)	Standard Error	*t*-value
(Intercept)	81.98	4.16	19.72*
Phrase Position: AP-INTERNAL	14.26	2.09	6.83*
Stop Item: FORTIS	15.89	2.09	7.61*

Appendix 2: VOT models.

VOT$_1$

Intercept: AP-INITIAL, ASPIRATED

Factor level	Estimate (ms)	Standard Error	*t*-value
(Intercept)	76.71	4.36	17.59
Phrase Position: AP-INTERNAL	-32.43	1.85	-17.50*
Stop Item: LENIS	-2.82	1.85	-1.52
AP-INTERNAL x LENIS	-23.90	2.77	-8.64*

VOT$_2$

Intercept: AP-INITIAL, FORTIS

Factor level	Estimate (ms)	Standard Error	*t*-value
(Intercept)	15.63	2.55	6.12
Phrase Position: AP-INTERNAL	-1.26	1.59	-0.79
Stop Item: LENIS	58.46	1.58	36.90*
AP-INTERNAL x LENIS	-54.11	2.37	-22.87*

VOT$_3$

Intercept: AP-INITIAL, ASPIRATED

Factor level	Estimate (ms)	Standard Error	*t*-value
(Intercept)	76.65	2.68	28.65
Phrase Position: AP-INTERNAL	-32.28	1.55	-20.82*
Stop Item: FORTIS	-61.10	1.55	-39.41*
Intercept: AP-INITIAL, ASPIRATED	31.05	2.20	14.12*

Appendix 3: Target syllable f0 models.

$f0_1$

Intercept: 4-SYLL, AP-INITIAL, LAX

Factor level	Estimate (Hz)	Standard Error	*t*-value
(Intercept)	137.91	17.24	8.00
# of Syll: 3-SYLL	-1.00	2.25	-0.44
Phrase Position: AP-INTERNAL	21.39	5.35	4.00*
Consonant Type: TENSE	63.99	5.36	11.93*
3-SYLL x AP-INTERNAL	-11.56	3.14	-3.69*
3-SYLL x TENSE	-1.76	3.19	-0.55
AP-INTERNAL x TENSE	-52.53	7.56	-6.95*
3-SYLL x AP-INTERNAL x TENSE	1.15	4.44	0.26

$f0_2$

Intercept: 5-SYLL, AP-INITIAL, LAX

Factor level	Estimate (Hz)	Standard Error	*t*-value
(Intercept)	140.47	17.37	8.09
# of Syll: 3-SYLL	-3.54	2.31	-1.54
Phrase Position: AP-INTERNAL	20.95	5.68	3.69*
Consonant Type: TENSE	58.05	5.70	10.19*
3-SYLL x AP-INTERNAL	-11.14	3.20	-3.48*
3-SYLL x TENSE	4.22	3.25	1.30
AP-INTERNAL x TENSE	-42.93	8.03	-5.35*
3-SYLL x AP-INTERNAL x TENSE	-8.46	4.52	-1.87

$f0_3$

Intercept: 5-SYLL, AP-INITIAL, LAX

Factor level	Estimate (Hz)	Standard Error	*t*-value
(Intercept)	140.35	17.76	7.90
# of Syll: 4-SYLL	-2.53	2.25	-1.12
Phrase Position: AP-INTERNAL	21.11	5.37	3.93*
Consonant Type: TENSE	58.01	5.39	10.77*
4-SYLL x AP-INTERNAL	0.50	3.13	0.16
4-SYLL x TENSE	5.91	3.18	1.86
AP-INTERNAL x TENSE	-42.98	7.59	-5.66*
4-SYLL x AP-INTERNAL x TENSE	-9.44	4.42	-2.14*

Appendix 4: AP-final f0 models.

$f0_1$

Intercept: 4-SYLL, AP-INITIAL, LAX

Factor level	Estimate (Hz)	Standard Error	*t*-value
(Intercept)	160.89	15.90	10.12
# of Syll: 3-SYLL	12.60	2.24	5.62*
Phrase Position: AP-INTERNAL	-0.61	2.25	-0.27
Consonant Type: TENSE	6.82	2.23	3.05*
3-SYLL x AP-INTERNAL	-1.74	3.17	-0.55
3-SYLL x TENSE	13.16	3.16	4.16*
AP-INTERNAL x TENSE	-2.48	3.17	-0.78
3-SYLL x AP-INTERNAL x TENSE	-9.98	4.48	-2.23*

$f0_2$

Intercept: 5-SYLL, AP-INITIAL, LAX

Factor level	Estimate (Hz)	Standard Error	*t*-value
(Intercept)	162.37	15.69	10.35
# of Syll: 3-SYLL	11.14	2.24	4.98*
Phrase Position: AP-INTERNAL	-1.72	2.24	-0.77
Consonant Type: TENSE	-2.43	2.23	-1.09
3-SYLL x AP-INTERNAL	-0.69	3.16	-0.22
3-SYLL x TENSE	22.43	3.15	7.12*
AP-INTERNAL x TENSE	3.45	3.16	1.09
3-SYLL x AP-INTERNAL x TENSE	-15.94	4.46	-3.58*

$f0_3$

Intercept: 5-SYLL, AP-INITIAL, LAX

Factor level	Estimate (Hz)	Standard Error	*t*-value
(Intercept)	162.14	14.83	10.94
# of Syll: 4-SYLL	-1.23	2.19	-0.56
Phrase Position: AP-INTERNAL	-1.85	2.19	-0.85
Consonant Type: TENSE	-2.20	2.17	-1.01
4-SYLL x AP-INTERNAL	1.23	3.09	0.40
4-SYLL x TENSE	8.99	3.07	2.93*
AP-INTERNAL x TENSE	3.37	3.08	1.10
4-SYLL x AP-INTERNAL x TENSE	-6.01	4.35	-1.38

Appendix 5: F0 excursion (v2-v1) model.

Intercept: AP-INITIAL, LAX			
Factor level	Estimate (Hz)	Standard Error	*t*-value
(Intercept)	18.82	4.15	4.54
Phrase Position: AP-INTERNAL	-0.89	3.71	-0.24
Consonant Type: TENSE	-6.39	3.70	-1.73
AP-INTERNAL x TENSE	23.14	5.25	4.41*

References

Abramson, A. S., & Lisker, L. (1972). Voice timing in Korean stops. In *Proceedings of the 5th Internacional Congress of Phonetic Sciences* (pp. 439–446).

Baayen, R. H. (2008). *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R*. Cambridge: Cambridge University Press.

Bang, H. Y., Sonderegger, M., Kang, Y., Clayards, M., & Yoon, T. J. (2018). The emergence, progress, and impact of sound change in progress in Seoul Korean: Implications for mechanisms of tonogenesis. *Journal of Phonetics*, *66*, 120–144.

Beddor, P. S. (2009). A coarticulatory path to sound change. *Language*, *85*(4), 785–821.

Blaylock, R. (2017). Matlab code: rtMRI-vocal-tract-analysis. Retrieved from https://github.com/reedblaylock/rtMRI-vocal-tract-analysis/.

Blevins, J. (2006). A theoretical synopsis of Evolutionary Phonology. *Theoretical Linguistics*.

Boersma, P., & Weenink, D. (2018). Praat: Doing phonetics by computer [Computer program]. Version 6.0.37 retrieved 3 February 2018 from http://www.praat.org/.

Browman, C., & Goldstein, L. (1985). Dynamic modeling of phonetic structure. In *Phonetic Linguistics: Essays in Honor of Peter Ladefoged*. Academic Press, New York.

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

Browman, C., & Goldstein, L. M. (1986). Towards an Articulatory Phonology. *Phonology Yearbook*, *3*(1986), 219–252.

Browman, C. P., & Goldstein, L. (1988). Some notes on syllable structure in articulatory phonology. *Phonetica*, *45*(2–4), 140–55.

Browman, C. P., & Goldstein, L. (1990). Tiers in articulatory phonology, with some implications for casual speech. *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech*, 341–376.

Browman, C. P., & Goldstein, L. (2000). Competing constraints on intergestural coordination and self-organization of phonological structures. *Bulletin de La Communication Parlee*, *5*, 25–34.

Byrd, D. (2002). Commentary: Relating prosody and dynamic action units. In *Laboratory Phonology*. New Haven, Conneticut.

Byrd, D., Kaun, A., Narayanan, S., & Saltzman, E. (2000). Phrasal signatures in articulation. *Papers in Laboratory Phonology V: Acquisition and the Lexicon*, 70–87.

Byrd, D., & Saltzman, E. (1998). Intragestural dynamics of multiple prosodic boundaries. *Journal of Phonetics*, *26*(2), 173–199.

Byrd, D., & Saltzman, E. (2003). The elastic phrase: Modeling the dynamics of boundary-adjacent lengthening. *Journal of Phonetics*, *31*(2), 149–180.

Cho, H., & Flemming, E. (2015). Compression and truncation of the Seoul Korean LHLH Accentual Phrase. *Studies in Phonetics, Phonology and Morphology*, *21*(2), 359–382.

Cho, S., & Lee, Y. (2016). The effect of the consonant-induced pitch on Seoul Korean intonation. *Linguistic Research*, *33*(2), 299–317.

Cho, T., & Jun, S. (2000). Domain-initial strenghtening as featural enhancement: Aerodynamic evidence from Korean. *Chicago Linguistics Society.*

Cho, T., Jun, S.-A., & Ladefoged, P. (2002). Acoustic and aerodynamic correlates of Korean stops and fricatives. *Journal of Phonetics*, *30*(2), 193–228.

Cho, T., & Keating, P. A. (2001). Ariculatory and acoustic studies on domain-intial strengthening in Korean. *Journal of Phonetics*, *29*, 155–190.

Cho, T., Lee, Y., & Kim, S. (2011). Communicatively driven versus prosodically driven hyper-articulation in Korean. *Journal of Phonetics*, *39*(3), 344–361.

Cho, T., Son, M., & Kim, S. (2016). Articulatory reflexes of the three-way contrast in labial stops and kinematic evidence for domain-initial strengthening in Korean. *Journal of the International Phonetic Association*, *46*(2), 129–155.

Cole, J., & Hualde, J. I. (2013). Prosodic structure in sound change. In S.-F. Chen & B. Slade (Eds.), *Grammatica et verba Glamor and verve* (pp. 28–45). Ann Arbor: Beech Stave Press.

Cole, J., Hualde, J. I., Blasingame, M., & Mo, Y. (2010). Shifting Chicago vowels: Prosody and sound change. In *Proceedings of Speech Prosody*.

Dart, S. N. (1987). An aerodynamic study of Korean stop consonants: measurements and modeling. *Journal of the Acoustical Society of America*, *81*(1), 138–47.

Dixit, P. R., & Macneilage, P. F. (1980). Cricothyroid activity and control of voicing in hindi stops and affricates. *Phonetica*, *37*, 397–406.

Edkins, J. (1864). *A Grammar of the Chinese Colloquial Language Commonly Called the*

Mandarin Dialect. Shanghai: Presbyterian Mission Press.

Elman, J. L. (1995). Language as a Dynamical System. *Mind as Motion: Explorations in the Dynamics of Cognition*, 195–225.

Fougeron, C., & Keating, P. A. (1997). Articulatory strengthening at edges of prosodic domains. *Journal of the Acoustical Society of America*, *101*(6), 3728–40.

Fowler, C. A., Rubin, P., Remez, R., & Turvey, M. (1980). Implications for speech production of a general theory of action. In B. Butterworth (Ed.), *Language Production*. Academic Press, New York.

Gafos, A. I., & Benuš, S. (2006). Dynamics of phonological cognition. *Cognitive Science*, *30*(5), 905–943.

Gao, M. (2008). *Tonal Alignment in Mandarin Chinese: An Articulatory Phonology Account*. PhD, Yale University.

Goldsmith, J. (1994). A Dynamic Computational Theory of Accent Systems. *Perspectives in Phonology*, 1–28.

Goldsmith, J., & Larson, G. (1990). Local Modeling and Syllabification. *Papers from the 26th Annual Regional Meeting of the Chicago Linguistic Society*.

Hagège, C., & Haudricourt, A.-G. (1978). *La Phonologie Panchronique*. Paris: Presses Universitaires de France.

Han, J.-I. (1996). *The Phonetics and Phonology of Tense and Plain Consonants in Korean*. PhD dissertation, Cornell University.

Han, M., & Weitzman, R. S. (1970). Acoustic features of Korean /P, T, K/, /p, t, k/, and /ph, th, kh/. *Phonetica*, *22*, 112–128.

Haudricourt, A. G. (1954). De l’origine des tons en Vietnamien. *Journal Asiatique*, *242*, 69–82.

Haudricourt, A. G. (1961). Bipartition et tripartition des systèmes de tons dans quelques langues d’Extrême-Orient. *Bulletin de La Société de Linguistique de Paris*, *56*(1), 163–180.

Hirose, H., Lee, C. Y., & Ushijima, T. (1974). Laryngeal control in Korean stop production. *Journal of Phonetics*, *2*(2), 145–152.

Hock, H. H. (1991). *Principles of Historical Linguistics* (2nd ed.). Berlin: de Gruyter.

Hombert, J.-M. (1978). Development of tones from vowel height? In V. A. Fromkin (Ed.), *Tone: A Linguistic Survey* (pp. 77–111). Academic Press, New York.

Hualde, J. I. (2013). Intervocalic lenition and word-boundary effects: Evidence from Judeo-

Spanish. *Diachronica*, *30*(2), 232–266.

Hyman, L. (1978). Tone and/or accent. In D. J. Napoli (Ed.), *Elements of Tone, Stress and Intonation* (pp. 1–20). Washington, D.C.: Georgetown University Press.

Hyman, L. M. (1973). The role of consonant types in natural tonal assimilations. In L. M. Hyman (Ed.), *Consonant Types and Tone, Southern California Occasional Papers In Linguistics* (pp. 151–179). University of Southern California.

Hyman, L. M. (1976). Phonologization. In A. Juilland (Ed.), *Linguistic studies presented to Joseph H. Greenberg*. Saratoga: Anma Libri.

Iskarous, K. (2016). Compatible dynamical models of environmental, sensory, and perceptual systems. *Ecological Psychology*, *28*(4), 295–311.

Iskarous, K. (2017). The relation between the continuous and the discrete: A note on the first principles of speech dynamics. *Journal of Phonetics*, *64*, 8–20.

Jordan, M. I. (1986). *Serial order: A parallel distributed approach. ICS Report 8604*.

Jun, S.-A. (1993). *The Phonetics and Phonology of Korean Prosody*. PhD dissertation, The Ohio State University.

Jun, S.-A. (1994). The status of the lenis stop voicing rule in Korean. In Young-Key Kim-Renaud (Ed.), *Theoretical Issues in Korean Linguistics*. CSLI Publications for Standford Linguistics Society.

Jun, S.-A. (1996). Influence of microprosody on macroprosody: A case of phrase initial strengthening. *University of California Working Papers in Phonetics*, *92*, 97–116.

Jun, S.-A. (1998). The Accentual Phrase in the Korean prosodic hierarchy. *Phonology*, *15*, 189–226.

Jun, S.-A. (2000). K-Tobi labelling conventions. *UCLA Working Papers in Phonetics*, *99*, 149–173.

Kagaya, R. (1974). A fiberscopic and acoustic study of the Korean stops, affricates and fricatives. *Journal of Phonetics*, *2*(2), 161–180.

Kang, K.-H., & Guion, S. G. (2008). Clear speech production of Korean stops: Changing phonetic targets and enhancement strategies. *Journal of the Acoustical Society of America*, *124*(6), 3909–17.

Kang, Y. (2014). Voice onset time merger and development of tonal contrast in Seoul Korean stops: A corpus study. *Journal of Phonetics*, *45*(1), 76–90.

Katsika, A., Krivokapić, J., Mooshammer, C., Tiede, M., & Goldstein, L. (2014). The coordination of boundary tones and its interaction with prominence. *Journal of Phonetics*, *44*(1), 62–82. http://doi.org/10.1016/j.wocn.2014.03.003

Keating, P., & Shattuck-Hufnagel, S. (2000). A prosodic view of word form encoding for speech production. *UCLA Working Papers in Phonetics*, *101*(1999), 112–156.

Kelso, J. a, Vatikiotis-Bateson, E., Saltzman, E. L., & Kay, B. (1985). A qualitative dynamic analysis of reiterant speech production: Phase portraits, kinematics, and dynamic modeling. *Journal of the Acoustical Society of America*, *77*(1), 266–80.

Keyser, S. J., & Stevens, K. N. (2001). Enhancement revisited. In *Ken Hale: A life in languages* (pp. 271–291). Cambridge, MA: MIT Press.

Kim, C.-W. (1965). On the anatomy of the tensity feature in stop classification (with special reference to Korean stops). *Word Journal of the International Linguistic Association*, *21*(3), 339–359.

Kim, H., Honda, K., & Maeda, S. (2005). Stroboscopic-cine MRI study of the phasing between the tongue and the larynx in the Korean three-way phonation contrast. *Journal of Phonetics*, *33*(1), 1–26.

Kim, H., Maeda, S., & Honda, K. (2010). Invariant articulatory bases of the features [tense] and [spread glottis] in Korean plosives: New stroboscopic cine-MRI data. *Journal of Phonetics*, *38*(1), 90–108.

Kirby, J. P. (2013). The role of probabilistic enhancement in phonologization. In A. C. L. Yu (Ed.), *Origins of Sound Change: Approaches to Phonologization* (pp. 1–30). Oxford University Press.

Krivokapić, J. (2007). Prosodic Planning: Effects of Phrasal Length and Complexity on Pause Duration. *Journal of Phonetics*, *35*(2), 162–179. http://doi.org/10.1016/j.wocn.2006.04.001

Krivokapić, J. (2014). Gestural coordination at prosodic boundaries and its role for prosodic structure and speech planning processes. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369:*, 20130397.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2016). lmerTest: Tests for random and fixed effects for linear mixed effect models. *R Package Version*. Retrieved from http://cran.r-project.org/package=lmerTest.

Labov, W. (1994). *Principles of Linguistic Change, Vol. I: Internal Factors*. Oxford: Blackwell.

Lammert, A. C., Goldstein, L., & Iskarous, K. (2010). Locally-weighted regression for estimating the forward kinematics of a geometric vocal tract model. *Learning*, 1604–1607.

Lammert, A., Ramanarayanan, V., Proctor, M., & Narayanan, S. (2013). Vocal tract cross-distance estimation from real-time MRI using region-of-interest analysis. In *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH* (pp. 959–962).

Lee, H., & Jongman, A. (2012). Effects of tone on the three-way laryngeal distinction in Korean: An acoustic and aerodynamic comparison of the Seoul and South Kyungsang dialects. *Journal of the International Phonetic Association*, *42*(2), 145–169.

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. In W. J. Hardcastle & A. Marchal (Eds.), *Speech Production and Speech Modelling* (pp. 403–439). Kluwer Academic, Dordrecht.

Lindblom, B., Guion, S., Hura, S., Moon, S.-J., & Willerman, R. (1994). Is sound change adaptive? *Rivista Di Linguistica*, *7*(1), 5–37.

Lingala, S. G., Zhu, Y., Kim, Y.-C., Toutios, A., Narayanan, S., & Nayak, K. S. (2017). A fast and flexible MRI system for the study of dynamic vocal tract shaping. *Magnetic Resonance in Medicine*, *77*(1), 112–125.

Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word Journal of the International Linguistic Association*, *20*(3), 384–422.

Lisker, L., & Abramson, A. S. (1971). Distinctive features and laryngeal control. *Language*, *47*, 767–785.

Löfqvist, A., Baer, T., McGarr, N. S., & Story, R. S. (1989). The cricothyroid muscle in voicing control. *Journal of the Acoustical Society of America*, *85*(3), 1314–1321.

Martin, S. (1982). Features, markedness, and order in Korean Phonology. In I. Yang (Ed.), *Linguistics in the Morning Calm* (pp. 601–618). Seoul: Hanshin.

Matisoff, J. A. (1973). Tonogenesis in Southeast Asia. In L. M. Hyman (Ed.), *Consonant Types and Tone. Southern California Occasional Papers in Linguistics* (Vol. 1, pp. 73–95). University of Southern California, Los Angeles.

Myers, S., & Padgett, J. (2014). Domain generalization in artificial language learning. *Phonology*, *31*(3), 399–433.

Narayanan, S., Nayak, K., Lee, S., Sethy, A., & Byrd, D. (2004). An approach to real-time magnetic resonance imaging for speech production. *Journal of the Acoustical Society of America*, *115*(4), 1771–1776.

Nissenbaum, J. (2008). Tone and register as articulatory parameters: Evidence from Cantonese. *Poster Presented at the Annual Meeting of the Canadian Linguistic Association*.

Nissenbaum, J. (2010). Tones in Cantonese: Articulatory vs. acoustic representation. *Talk given at the 18th Annual Meeting of the International Association of Chinese Linguistics*.

Oh, M., & Johnson, K. (1997). A phonetic study of Korean intervocalic laryngeal consonants. *Journal of Speech Sciences*, *1*, 83–102.

Oh, M., Toutios, A., Byrd, D., & Narayanan, S. S. (2017). Tracking larynx movement in real-time MRI data. *Journal of the Acoustical Society of America*, *142*(4), 2579.

Ohala, J. (1993). The phonetics of sound change. In C. Jones (Ed.), *Historical Linguistics: Problems and Perspectives* (pp. 235–278). Longman Academic, London.

Ohala, J. J. (1972). How is pitch lowered? *Journal of the Acoustical Society of America*, *52*(1A), 124.

Ohala, J. J. (1973). The Physiology of Tone. In *Consonant Types and Tone, Southern California Occasional Papers In Linguistics* (pp. 2–14). University of Southern California, Los Angeles.

Ohala, J. J. (1974). Experimental historical phonology. In J. M. Anderson & C. Jones (Eds.), *Historical Linguistics: Problems and Perspectivesinguistics II* (pp. 353–389). North-Holland, Amsterdam.

Pierrehumbert, J., & Talkin, D. (1992). Lenition of /h/ and glottal stop. *Papers in Laboratory Phonology II: Gesture, Segment, Prosody*, 90–117.

R Core Team. (2018). R Development Core Team. *R: A Language and Environment for Statistical Computing*. http://doi.org/http://www.R-project.org.

Roon, K. D., & Gafos, A. I. (2016). Perceiving while producing: Modeling the dynamics of phonological planning. *Journal of Memory and Language*, *89*, 222–243.

Saltzman, E. L., Nam, H., Krivokapic, J., & Goldstein, L. (2008). A task-dynamic toolkit for modeling the effects of prosodic structure on articulation. *Proceedings of the 4th*

International Conference on Speech Prosody, 175–184.

Saltzman, E., & Munhall, K. (1989). A dynamical approach to gestural paterning in speech production. *Ecological Psychology*, *1*(4), 333–382.

Satterthwaite, F. E. (1946). An Approximate Distribution of Estimates of Variance Components. *Biometrics Bulletin*, *2*(6), 110.

Selkirk, E. (1986). On derived domains in sentence phonology. *Phonology*, *3*, 371.

Silva, D. J. (2006). Acoustic evidence for the emergence of tonal contrast in contemporary Korean. *Phonology*, *23*(2), 287–308.

Son, M., Kim, S., & Cho, T. (2012). Supralaryngeal articulatory signatures of three-way contrastive labial stops in Korean. *Journal of Phonetics*, *40*(1), 92–108.

Sorensen, T., & Gafos, A. (2016). The gesture as an autonomous nonlinear dynamical system. *Ecological Psychology*, *28*(4), 188–215.

Steriade, D. (1997). *Phonetics in Phonology : The Case of Laryngeal Neutralization*. University of California, Los Angeles.

Tilsen, S. (2016). Selection and coordination: The articulatory basis for the emergence of phonological structure. *Journal of Phonetics*, *55*, 53–77.

Tilsen, S., Spincemaille, P., Xu, B., Doerschuk, P., Luh, W. M., Feldman, E., & Wang, Y. (2016). Anticipatory posturing of the vocal tract reveals dissociation of speech movement plans from linguistic units. *PLoS ONE*, *11*(1).

Tuller, B., Case, P., Ding, M., & Kelso, J. A. S. (1994). The nonlinear dynamics of speech categorization. *Journal of Experimental Psychology: Human Perception and Performance*, *20*(1), 3–16.

Vatikiotis-Bateson, E., & Kelso, J. A. S. (1993). Rhythm type and articulatory dynamics in English, French and Japanese. *Journal of Phonetics*, *21*(3), 231–265.

Yu, J. (1989). A study on Korean glottalized tense sound and aspirated. *Hankul*, *203*, 25–48.