Journal of Phonetics 77 (2019) 100933

Contents lists available at ScienceDirect

Journal of Phonetics

journal homepage: www.elsevier.com/locate/Phonetics

Special Issue: Integrating Phonetics and Phonology, eds. Cangemi & Baumann

Syllable-internal corrective focus in Korean

Miran Oh *, Dani Byrd

Department of Linguistics, University of Southern California, Los Angeles, CA, USA

ARTICLE INFO

Article history:
Received 11 March 2018
Received in revised form 15 August 2019
Accepted 3 October 2019
Available online 11 November 2019

Keywords:
Prominence
Prosody
Focus
Articulatory Phonology
Syllable structure
Gestures
Scope

ABSTRACT

In producing linguistic prominence, certain linguistic elements are highlighted relative to others in a given domain; focus is an instance of prominence in which speakers highlight new or important information. This study investigates prominence modulation at the sub-syllable level using a corrective focus task, examining acoustic duration and pitch with particular attention to the gestural composition of Korean tense and lax consonants. The results indicate that focus effects are manifested with systematic variations depending on the gestural structures, i.e. consonants, active during the domain of a focus gesture, but that the patterns of focus modulation do not differ as a function of elicited focus positions within the syllable. The findings generally support the premise that the scope of the focus gesture is not (much) smaller than the interval of (CVC) syllable. Lastly, there is also some support for an interaction among prosodic gestures—focus gestures and pitch accentual gestures—at the phrase level. Overall, the current findings support the hypothesis that focus, implemented as a prosodic prominence gesture, modulates temporal characteristics of articulatory gestures, as well as possibly other prosodic gestures that are co-active in its the domain.

1. Introduction

In producing linguistic prominence, speakers highlight certain linguistic elements relative to others in a given domain. This relative highlighting may be rhythmic in nature—such as in the case of stress or accent—or informational—such as in the case of focus; and in fact these types of prominence may interact.

Placing focus on some portion of an utterance serves to highlight new or important information and/or to convey a contrast with given information. The phonetic realization of focal accent has been described as including higher or lower pitch (i.e., a pitch perturbation), longer duration, and/or greater intensity relative to the same non-focused word (e.g., Bolinger, 1961; Jun & Lee, 1998; Lee & Xu, 2010; Van Heuven, 1994). Other studies in this special issue focus on phrase- or word-level prominence (Cole et al., 2019; Smith, Erickson, & Savariaux, 2019; Zahner, Kutscheid, & Braun, 2019). However, little detail is known regarding how the effects of prominence may depend on the phonological nature of the unit focused or on its structural composition. In particular, how are focus effects exhibited and what is the scope of these effects when the focused region is pragmatically narrowed down to a segment-sized granularity? Is the scope of focus small enough to render distinguishable patterns as a function of where within a syllable focus is targeted, or do focus effects necessarily span over a syllable in its entirety? That is, we seek to address whether the effects of focus are sensitive to the focused unit's internal phonological composition.

It is well established that prominence is realized primarily on the nucleus of a focused syllabic domain (Baumann & Winter, 2018, Van Heuven, 1994). What is less investigated is whether a sub-syllable domain—specifically, segments—can be specifically targeted in the acoustic modulations that characterize focus or whether a syllable domain of focus can be marginally shifted or modulated by segmental focus. This study examines how segmental corrective focus is realized when placed on varying structural positions within a syllable. We use corrective focal prominence (setting aside other types of prominence such as lexical stress), as it generally has greater effects than other focus types (cf. broad and narrow focus) (Mücke & Grice, 2014). We probe the effects of prominence in Korean by examining stops (/n, t, t^h, t*/) in syllable onset and coda positions as well as the vowel in nucleus position [CaC].

Two aspects of sub-syllabic corrective focus effects are addressed. The first probes the minimal element that the effects

* Corresponding author.
E-mail address: miranoh@usc.edu (M. Oh).

https://doi.org/10.1016/j.wocn.2019.100933
0095-4470/

of focal prominence apply to, namely whether sub-syllable corrective focus is realized differentially as a function of syllable structure. Can a single sub-syllabic position be ‘picked out’ for focus, and if so, are there dependencies among the syllable positions for this focus realization? For example, given a gestural organization in which the production of onsets and their tautosyllabic vowel initiate in-phase (i.e., synchronous onsets), correctively focusing an onset consonant (**C**VC) might be predicted to yield lengthening of both that onset and the nuclear vowel as well (Goldstein et al., 2006, 2007, 2009; Mücke, Nam, Hermes, & Goldstein, 2012; Nam, Goldstein, & Saltzman, 2009). In contrast, corrective focus of a coda (CV**C**), which is understood to be timed sequentially with its tautosyllabic vowel (Goldstein, Nam, Saltzman, & Chitoran, 2009; Nam et al., 2009), might be predicted to exhibit acoustic durational effects only, or at least primarily, on the consonant and not on its preceding nuclear vowel. Alternatively to a finding of sub-syllabic targeted effects, equivalent lengthening across all syllable positions—irrespective of the particular segment in a syllable being correctively focused—could indicate that sub-syllable domains are not a viable domain for (at least, corrective) focus or, equivalently, that the scope of this focus may not be compressed to an interval smaller than the syllable. If a focus gesture cannot narrowly and specifically target a sub-syllable domain (i.e., onset, nucleus, and coda), the focus effects of segmental corrective focus will necessarily have a broader scope—such as the syllable or word—and its realization, regardless of which of the syllable’s segments is pragmatically corrected, is predicted to yield a comparable degree of lengthening throughout the syllable. Aside from the theory-specific representational accounts of prosodic events that we entertain below, an examination of temporal modulations at varying sub-syllabic focus positions informs as to whether and how phonological syllable structure influences the realization and scope of prosodic informational structure.

The second question investigated is how a focused segment’s intrinsic gestural characteristics are reflected in focus realization patterns. Given that Korean laryngeal consonants (lenis /t/, aspirated /t^h/, and fortis /t*/) contrast in pitch (f0) and durational values (stop closure and Voice Onset Time [VOT]) (Jun, 1998; Lee & Xu, 2010; Silva, 2006), correctively focusing each of these stops could show distinct acoustic patterns. Although Seoul Korean is not a lexical tone language, the f0 contrast between lax (lenis) and tense (aspirated and fortis) onset consonants on the following vowel is widely manifested, especially by the younger generation now in their early adulthood (Bang, Sonderegger, Kang, Clayards, & Yoon, 2018; Kang, 2014). Thus, the difference in the gestural composition of each segment could result in distinct output focus f0 patterns. For instance, if a pitch contrast between a tense stop and a lax stop—tense having high pitch on the following vowel and lax being associated with a lower pitch—is gesturally represented specifically by the presence of a H_{tenseC} pitch gesture,[1] a focus prosodic event (i.e., prosodic gesture) co-active with the stop would result in the f0 of a tense stop having its pitch raised as the focus gesture promotes the activation of the H_{tenseC} gesture, while the f0 of lax stops may not be much influenced (there being no pitch-related gesture to be affected by a focus gesture). Alternatively, if corrective focus has a pitch effect that is insensitive to (i.e., does not interact with) the tonal composition of specific segments, f0 perturbation due to focus would be predicted to occur in a like manner across each of the stop types and their tautosyllabic vowels.

Overall, the current study investigates whether and how syllable structure and segment-specific properties influence acoustic variability in the production of focal prominence as instantiated by a correction of one contrastive segment for another. The remainder of Section 1 presents a brief review of relevant theoretical and experimental literature and introduces the hypotheses examined in this study. Methods for the current experiment are presented in Section 2. Section 3 reports the results of the experiment; the first half presents syllable structure effects and the second half presents segmental composition effects. A discussion of the findings and implications follow in Sections 4 and 5.

1.1. Theoretical and experimental background

In the Articulatory Phonology framework for phonological representation, which we adopt, the primitive phonological units are gestures that control the speech effector subsystems (tract variables) (e.g., Browman & Goldstein, 1992; Goldstein, Byrd, & Saltzman, 2006). These abstract ‘task’ events are both action units and information units defined with inherent spatiotemporal specifications that serve as the basis for phonological contrast and are computationally implemented in speech production via a Task Dynamics (Saltzman & Munhall, 1989; Saltzman, 1995). Articulatory gestures are coordinated with one another in both language-specific and universal patterns that form the basis of syllable structure and, along with gestural parameter values, also contribute to phonological contrast. As reviewed in Katsika (2016), three types of gestures have been postulated within Articulatory Phonology: vocal tract constriction, (lexical and intonational) tone (Gao, 2008; Katsika, Krivokapić, Mooshammer, Tiede, & Goldstein, 2014; Mücke, Nam, Hermes, & Goldstein, 2012), and prosodic, with the last having been modeled as gestures (“prosodic” or π-gestures/μ-gestures) that induce local spatiotemporal modifications (e.g., Byrd & Saltzman, 2003; Krivokapić & Byrd, 2012; Nam, Saltzman, Krivokapić, & Goldstein, 2008; Saltzman, Nam, Krivokapić, & Goldstein, 2008).

In Articulatory Phonology’s dynamical systems approach to speech production, prosodic structures, such as prosodic gestures occurring at a phrasal juncture, have been encoded as vicarious events that act on articulatory gestures to produce local modulations of the spatiotemporal properties of all gestures that are active during a localized portion of an utterance (Byrd & Saltzman, 2003; Saltzman et al., 2008). The effects of prosodic events are hypothesized to extend in time and act in a transgestural fashion—meaning that the prosodic event or π-gesture modulates all of the gestures that are co-active during the ‘scope’ of this prosodic event. To date, much of the research incorporating prosodic structure into Articulatory

[1] Given that “H” is typically used to indicate lexical or intonational high tone gestures, we instead use the notation “H_{tenseC}” gesture to indicate a tense consonant’s structural pitch composition (which spans its co-active vowel production). We assume that only tense consonants have a pitch gesture based on Kang’s (2014) developmental corpus data in which there is a clear pitch distinction between tense/lax consonants, but this distinction lies in the tense consonant pitch raising, as no pitch lowering is observed with lax consonants (cf. Maran, 1973).

Phonology has centered on the gestural representation of phrase boundaries and their effects on articulation at phrase edges. Clock-slowing gestures, called π-gestures, have been postulated as instantiating phrase boundaries; these π-gestures slow the central clock controlling the pacing of the unfolding of the gestural constriction activation functions (Byrd & Saltzman, 2003). The outcome of such prosodic modulation at phrase edges includes local lengthening (which may also result automatically in larger displacement or spatial magnitude, as longer activation of a gesture gives more time for it to reach the desired target) and lessened overlap among the concurrent gestures (Byrd & Saltzman, 2003). The degree to which these prosodically driven variations are induced as a function of phrase boundaries is captured by the activation level or strength of the prosodic gesture (π-gesture) (Byrd & Saltzman, 2003). Many studies have found that the effects of boundary strengthening or lengthening are greater as phrase boundaries are aligned with higher level constituents in the prosodic hierarchy (Byrd, 2000; Cambier-Langeveld, 1997; Cho & Jun, 2000; Cho & Keating, 2001; Cho, Lee, & Kim, 2011; Fletcher, 1991; Keating, Cho, Fougeron, & Hsu, 2003; Krivokapić & Byrd, 2012). This can be interpreted as π-gestures having higher activation levels for larger prosodic constituents, so that, consequently, all gestures in the domain of a π-gesture slow down to a greater degree as the strength of the phrase boundary (i.e., disjuncture) gets bigger or occurs at a higher hierarchical level. The duration of a π-gesture, and consequently the constriction material with which it is co-active, is referred to as its domain or 'scope' (see e.g., Byrd & Riggs, 2008; Byrd, Krivokapić, & Lee, 2006).

The current study addresses focal prominence effects (see also Katsika et al., 2014) using some of the theoretical tools that have been developed in gestural accounts of phrasal prosody. The π-gesture approach has begun to be extended beyond phrase boundary effects to other aspects of prosodic structures such as focus, prominence, stress, and rhythmic structures (see e.g., Katsika, 2016; Nam et al., 2008; Saltzman et al., 2008). Saltzman et al. (2008) hypothesize a richer general modulation gesture (μ-gesture) that applies to gestural planning oscillators playing a role in speech rhythm, such as foot structure. μ-gestures are divided into two distinct types: temporal modulation gestures that modulate the clock rate by changing the frequency parameters of the gestural planning oscillators, and spatial modulation gestures that increase the spatial target parameters of the constriction gestures (Saltzman et al., 2008). μ-gestures allow multiple modulation gestures to overlap and compete with each other, thereby providing for interactions of multiple prosodic tasks such as phrasal structure and prominence (Katsika et al., 2014; Katsika, 2016; Saltzman et al., 2008) and for the interaction of various types of prominence, such as information focus and tonal accent.

The current study interprets acoustic findings on duration and pitch in terms of the gestural representation of focus, syllable structure, and segments, as well as phrasal accent. In considering the scope of focus effects, previous acoustic studies have suggested that focus effects, unlike boundary effects or stress, may not be reflected strictly in the acoustic properties of *individual segments*. For example, Cambier-Langeveld (1997) investigated duration of segments in Dutch to examine the amount of final lengthening as a function of two prosodic variables: prosodic depth (PW, PhP, IP, U) and focus distributions. While segmental durations were lengthened phrase-finally at utterance and intonational phrase boundaries, the additional placement of focus on the phrase-final target word had no significant effect on the duration of the segments in the phrase final word. On the contrary, other studies have proposed that focus does exhibit some segment-level interactions. Van Heuven (1994) study of English showed that pitch peak alignment tended to move away from the location of a focused segment. Another study on segmental focus in Arabic (De Jong & Zawaydeh, 2002) showed that while focus increases the durational disparity between vowels having a phonemic quantity contrast, it does not differentially affect the amount of lengthening associated with the voicing of the obstruent following the vowel, though focus does as expected increase vowel length in general. In a corpus study of vowel duration in English, De Jong (2004) found that "focus effects are mediated by stress such that increases in durational differences are localized largely in syllables which are primary stressed and accented." In addition, Botinis, Fourakis, and Prinou (1999) examined stress and focus effects on segmental durations in Greek. They concluded that focus effects in Greek were larger on vowels than on consonants since vowel durations were significantly different depending on both lexical stress and focus whereas consonant differences were affected only as a function of stress, not of focus. Cho, Kim, and Kim (2017) also found some small difference between segmental and lexical focus on nasal consonants, with the primary realization of focus always observed during the vowel. Mücke and Grice (2014) compared different degrees of prominence among broad, narrow, and contrastive focus structures and showed that words with contrastive focus had the highest degree of emphasis. Beyond these studies, however, little work is available examining focus effects at the segmental (sub-syllable) level in a granular way with respect to the phonological properties of segments and the specific syllable-internal position of the focused segments.

1.2. Hypotheses

The current study investigates at what granularity or scope speakers manifest prosodic prominence (focus) and the consequences of this prominence realization on the acoustic properties of consonants with differing underlying gestural and tonal composition. A specific position within a syllable is manipulated to be focused through eliciting corrective focus at the level of a segment: e.g., Q. *Did you say **n**ight?,* A: *No, I said **t**ight*. Or A: *No, I said ni**n**e*. Comparing the temporal and f0 characteristics that result from placing each syllable position in focus (relative to a control non-focused condition) can illuminate potential prosodic effects on sub-syllable domains. The study examines whether the temporal and f0 modulations produced by focus vary depending on which part of a syllable—onset, nucleus, or coda—is being targeted by focus. And, if there is a syllable position effect, whether the implementation of this corrective focus is evident on both onset consonant and its tautosyllabic vowel in a way different from coda consonants and the vowel.

The coupled oscillator model of syllable structure (Goldstein et al., 2006, 2007, 2009; Löfqvist & Gracco, 1999; Mücke et al., 2012; Nam et al., 2009) postulates that onset and coda represent two distinct coupling modes in a syllable: an onset consonant is coupled in-phase to the tautosyllabic vowel (i.e., synchronous timing relation), and a coda consonant and a tautosyllabic vowel are coordinated in the anti-phase mode (i.e., sequential production). If corrective focus is implemented as a μ-gesture with a sub-syllabic scope, it would modulate the spatiotemporal characteristics of an utterance in a manner sensitive to the coupling graph of the syllable structure. Specifically, the current experiment tests the following hypothesis: *a focus gesture active for syllable onset position modulates the clock rate during both the onset and the vowel gesture in a CVC syllable due to their synchronous timing; whereas correctively focusing coda will not exhibit temporal modulation of the preceding tautosyllabic vowel due to their sequential, as compared to synchronous, timing.* A 'softer' version of this hypothesis would predict a lesser effect on the vowel, rather than no effect, for coda focus, as the μ-gesture may be centered during the coda interval but nevertheless have a waxing activation that still partially overlaps with the final portion of the preceding vocalic period (Fig. 1a: left). The schemas in Fig. 1a illustrate activation intervals of each gesture and possible temporal scopes of focus.

Alternatively, the μ-gesture could have a scope roughly the size of the entire syllable or word (Fig. 1: right), regardless of what segment is pragmatically being corrected in the utterance. In this case, all the syllable's gestures would lengthen (though the gestures co-active during the strongest part of a μ-gesture would be predicted to lengthen most strongly).

The three-way laryngeal contrast in Korean (i.e., lenis, aspirated, and fortis stops) provides an excellent testbed for these hypotheses. The Korean stop consonants exhibit distinctive acoustic measures of VOT, f0, and stop closure duration. For lenis and aspirated stops, VOT is longer than for fortis stops (Kang & Guion, 2008; Lee & Jongman, 2012; Silva, 2006). In addition, in Seoul Korean, f0 values of vowels with aspirated and fortis stop onsets are much higher than those with a lenis stop onset (Cho & Lee, 2016; Jun, 1995; Kang & Guion, 2008; Kang, 2014; Silva, 2006), showing over 40 Hz (& up to 100 Hz) difference (Silva, 2006).[2] Stop closure duration is not much different among these three stops word-initially, though aspirated and fortis stop consonants have longer closure duration word-medially in Korean (Cho & Keating, 2001; Han, 1996).

In considering f0 patterning in Korean, it is vital to recognize the role of the prosodic phrase—Accentual Phrase (AP) and Intonational Phrase (IP) (composed of APs), because the AP carries a relatively inflexible pitch patterning. Seoul Korean exhibits a systematic and oscillating accentual pitch pattern (Jun, 1993; Lee, 2018) and does not impose stress on lexical items. Accentual Phrases in Seoul Korean with more than three syllables start with a Low tone, followed by an initial High

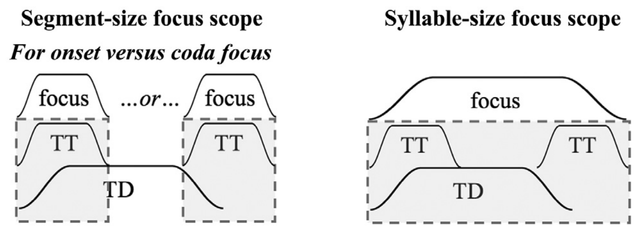


Fig. 1. Schematic gestural organizations for a syllable [tat] with onset focus or coda focus (left), or syllable/word focus (right) [TT indicates a Tongue Tip (consonant) gesture; TD indicates Tongue Dorsum (vocalic) gesture].

tone on the second syllable, and a Low minimum is on the penultimate syllable followed by the final High tone (e.g., LHLH). If the phrase contains more than four syllables, the remaining syllables that are not associated with any underlying tone are by default a Low tone (e.g., LHLLH; Jun, 1993). With lax initial words, tone in Accentual Phrases in Seoul Korean oscillates between Low and the High tone (e.g., LH, LLH/LHH, LHLH, etc.) (Jun, 1993, 1995, 1998) over the syllables of the Accentual Phrase. Words with tense initial onsets have the Accentual Phrase patterns: HH, HLH, HHLH. It follows that the f0 contrast between lax and tense stop onsets is revealed in the Accentual Phrase initial position but neutralized (although not fully, see Lee, 2018) within Accentual Phrases. The markings of prominence and phrasal boundaries, however, are not independent in Korean—for instance, inducing focus on a word can lead to de-phrasing of the upcoming speech materials, causing a focused word and a post-focal word to form a single Accentual Phrase (Jun, 1993, 2014). As we will see, however in Section 2.3.2, with the focused segment positioned in the third syllable in a tri-syllabic Accentual Phrase (LL**H**/LH**H**) no de-phrasing of following material occurs in our study.

Based on empirical studies on Korean stop consonant production (Cho & Jun, 2000; Cho, Jun, & Ladefoged, 2002; Dart, 1987; Hirose et al., 1974, 1983; Hong, Niimi, & Hirose, 1991; Jun, 1995; Kagaya, 1974; Kang & Guion, 2008; Kim et al., 2005, 2010; Kim, 1965; Kim, Maeda, Honda, & Crevier-Buchman, 2018; Lee & Jongman, 2012; Silva, 2006), the inherent spatiotemporal gestural organizations of syllables with each of the three stop onset consonants (/t, t^h, t*/) are suggested as in Fig. 2 (depicting a scope of focus as segment-sized). The stop closure duration is controlled by the activation of the tongue tip (TT) gesture; aspiration (VOT) reflects the temporal offset between oral closure release and the end of the glottal opening (GLO) gesture, and the articulation of the vowel is represented as the tongue dorsum (TD) gesture. A H_{tenseC} gesture, contributing to pitch of the co-active vowel gesture, is represented here as being active during tense consonants' gestural compositions.

The current experiment examines two hypotheses with regard to segmental composition and focus. First, it tests the hypothesis that *segmental corrective focus will produce a significant and systematic acoustic effect at a sub-syllabic level specific to the elicited locus of focus within the syllable*. And secondly, the study tests the hypothesis that *all and only the gestures co-active during the domain of a focus gesture are slowed, with the concomitant acoustic consequences* (e.g., more extreme active pitch modulation and longer duration).

[2] Electromyographic (EMG) studies on Korean stops indicate that the reactivation of the thyroarytenoid muscle activity is earlier and its peak is higher for tense stops than lax ones (Hirose et al., 1974, 1983; Hong et al., 1991). Given that f0 and air pressure are positively correlated (Lee & Jongman, 2012; Shipp & McGlone, 1971), aerodynamic studies on Korean stops also show a tense/lax distinction in f0 as the intraoral air pressure is higher and the duration of increased air pressure is longer in aspirated and fortis stops than in lenis stops (Cho et al., 2002; Dart, 1987; Kim, 1965; Kim et al., 2018). This finding is also reported in fiberscopic (Kagaya, 1971, 1974), aerodynamic (Kim, 1965), and stroboscopic cine-MRI studies (Kim et al., 2005, 2010) on Korean stops among others.

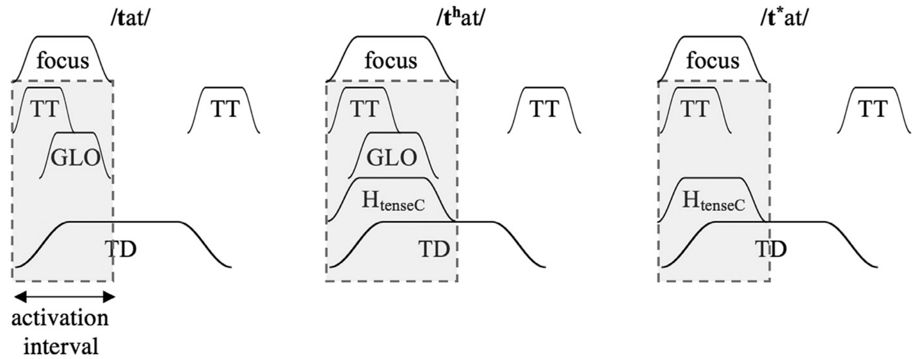


Fig. 2. Schematic gestural organizations of Korean stop onsets (/t, tʰ, t*/), depicting onset (segment) focus.

Specifically, it predicts that lenis and aspirated onsets (/t/ & /tʰ/), which have a glottal opening gesture for aspiration, will exhibit VOT lengthening under focus, whereas tense onsets (aspirated /tʰ/ & fortis /t*/) will have higher pitch values under focus due to the modulation of their underlying H_{tenseC} gesture.

2. Methods

2.1. Participants

Sixteen adult native Seoul Korean speakers (whose parents are also Korean native speakers) participated in this study (eight male; eight female). They were young adult graduate students in the United States from South Korea, with ages ranging from 23 to 31 years old and had no known hearing, vision, reading, or learning disabilities.

2.2. Materials

The pseudoword in the experiment functioned as a common noun for an object name and was a trisyllable with a LLH Accentual or Intonational Phrase pattern. The target word phrase may form a single Accentual Phrase (AP) or sometimes an Intonational Phrase (IP), perhaps more likely to form a larger phrase when focused (Jun & Kim, 2007). The target CVC syllable was embedded in the trisyllabic pseudoword of the form of two open syllables and a following closed syllable (CV.CV.$\underline{C_oV_nC_c}$), and each of the subscripted segments was correctively focused in individual trials. Therefore, there were three sub-syllable focus positions or locations: onset focus (C_o), nucleus focus (V_n), and coda focus (C_c). The target syllable was always the final syllable of the tri-syllabic word. The target word was followed by a tri-syllabic phrase beginning with a tense consonant, thus having a HLH AP or IP. Consonants focused in the CVC onset were alveolar oral and nasal stops (lenis /t/, aspirated /tʰ/, fortis /t*/, and /n/). Onset and coda focus were elicited by contrasting the stop consonants (/t, tʰ, t*/) with a nasal stop, and the nasal (/n/) with a lenis oral stop (/t/). The vowel placed in nucleus focus is /a/, which is the most common vowel in the language and is often used in names. Nucleus focus of /a/ was induced by contrasting it with /ʌ/. Consonants used for coda focus were the lenis alveolar stop (/t/) and an alveolar nasal (/n/). (Aspirated and fortis stops were not included in the coda focus items because these two laryngeal stops are neutralized to a lenis stop in the coda position.) A total of 11 target words occurred (Table 1).

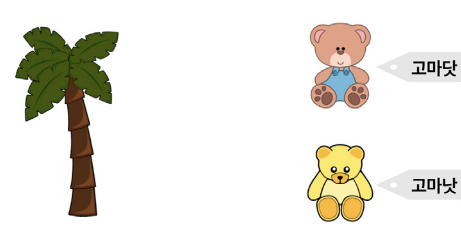
Table 1
Target words (trisyllabic pseudowords) in onset, nucleus, and coda focus positions.

Target syllable	Onset focus (/t, tʰ, t*, n/)	Nucleus focus (/a/)	Coda focus (/t, n/)
/tat/	koma**t**at	poma**t**at	ɕomata**t**
/tʰat/	ɕʌmat**ʰ**at	ɕumatʰ**a**t	–
/t*at/	pumat***at	nomat***a**t	–
/nat/	kuma**n**at	kʌman**a**t	–
/tan/	–	numat**a**n	pʌmata**n**

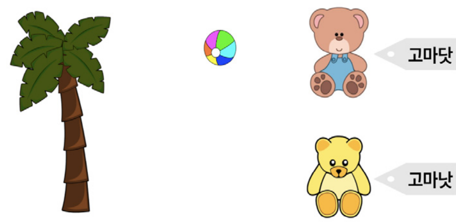
The target words were elicited in a carrier sentence. An interactive communication task was used in which the participant answered a question posed by the experimenter (the first author, who is a native speaker of Seoul Korean of similar age to the subjects).

Crucially, both a focus condition and a control post-focus condition were tested. In the *focus condition,* a segment in the target word was corrected by the participant in responding to a query by the experimenter. In *the post-focus condition,* a word produced two words *before* the target word was correctively focused upon the experimenter's prompt, causing the target syllable to be post-tonic and have no focus. An example for an onset focus elicitation and a corresponding post-focus condition is given in (1).

a. Onset focus (**bolded**: target segment, underlined: contrasting segment)



Q. Kong-i namu-wa komanat sai-e isseo?
ball-TOP tree-CONJ komanat between be-INT
Q [gloss]. Is the ball between the tree and the [komanat]?



A Ani, kong-i namu-wa koma**t**at sai-e isseo.
No, ball-TOP tree-CONJ koma**t**at between be-DEC
A [gloss]. No, the ball is between the tree and the [koma**t**at].

b. Post-focus condition (**bolded**: target word, <u>underlined</u>: contrasting word)

Q. <u>Tal</u>-i namu-wa komatat sai-e isseo?
<u>moon</u>-TOP tree-CONJ komatat between be-INT
Q [gloss]. Is <u>the moon</u> between the tree and the [komatat]?
A Ani, **kong**-i namu-wa komatat sai-e isseo.
No, **ball**-TOP tree-CONJ komatat between be-DEC
A [gloss]. No, the **ball** is between the tree and the [komatat].

Each participant produced each target word 7 times. In total, 154 syllables (11 target words, 2 focus conditions, and 7 repetitions) were collected for each subject. A total of 14 blocks were presented, blocks alternating between focus and post-focus conditions. Target sentences were randomized within block. Four lists were created with different order of alternation between the two focus conditions and by reversing the order of block lists; each participant was assigned to one of the four lists.

2.3. Recordings

In a sound-insulated room, participants were instructed to produce particular corrective sentences in response to questions posed aloud by the experimenter. Each question prompt took about 2.2–2.3 seconds to produce. The participants listened to the experimenter's question while seeing pictures with name tags on a computer screen. Participants were instructed to produce corresponding answers based on the information shown on the screen. Participants were to correct the experimenter's question, which always contained incorrect information. Their spoken responses each took around 2.6–2.8 seconds. For instance, the experimenter asked a question such as *"Is the moon between the tree and the komanat?"* and the participant answered *"No, the moon is between the tree and the koma**t**at."* The same target word in post-focus condition was elicited by having the participants correctively focus a preceding word, again based on graphic information shown on the computer screen. For instance, the experimenter asked a question such as *"Is the moon between the tree and the komatat?"* and the participant answered *"No, the **ball** is between the tree and the komatat."* Therefore, in focus conditions, a segment in the critical target word was correctively focused, and as a control the identical word was collected in a post-focus condition that had the focus earlier in the sentence. A desktop microphone was placed at a distance about 2 to 4 inches from the speakers, and the whole process was recorded using the Audacity software (Audacity Team, 2017). The audio recordings are provided as supplementary data (see Appendix B).

2.4. Measurements

Pitch, vowel duration, VOT, and stop closure duration were measured by analyzing recordings using the Praat software (Boersma & Weenink, 2017). The long data files were segmented into smaller sound files using Bion (2009) SegmentFile praat script. The annotation of each wav file on a TextGrid file was done with using a modified version of the TextGridMaker praat script (Crosswhite, 2013). The following measurement criteria were used:

- The stop closure duration of the onset consonant was measured from the abrupt amplitude drop and loss of energy in F2 and F3 of the preceding vowel to the release burst (for stops) or the start of the following vowel (for nasals).
- Voice Onset Time (VOT) was measured from the release burst of the stop to the beginning of periodic voicing and the start of the voice bar.
- The starting point of vowel duration was measured at the beginning of the periodicity (start of voicing). The endpoint for the vowel duration was determined by inspection of the following: a dramatic change in the waveform's amplitude, a loss of energy in the high formats (F2 and F3), and the start of aperiodicity.
- The maximum fundamental frequency (F0) was obtained within the interval of the nuclear vowel (labeled based on the criterion immediately above) for each target syllable. Praat pitch analysis setting used the following parameters: pitch floor as 75 Hz and pitch ceiling as 500 Hz for both men and women, time step as 0.01 seconds, and a pitch window threshold of 0.03.

The timepoints for the annotated labels in each set of stimuli were automatically extracted using a modified script from the source code by Lennes (2004). The script was used to calculate and obtain values from all labeled intervals. The measured acoustic values were onset consonant closure duration, VOT, vowel duration, and f0 maximum. Although the current study focuses on examining how maximum f0 values associated with tense/lax consonants change in varying focus productions, in depth analysis of f0 dynamics such as pitch peak alignment (Zahner et al., 2019) and f0 contour dynamics, while beyond the scope of the present work, would potentially provide a richer understanding of overall focus effects that span beyond the syllable-level. 7392 measurements were examined for f0, vowel duration, and onset stop closure (11 syllables $\times$ 2 focus conditions $\times$ 16 speakers $\times$ 7 repetitions $\times$ 3 acoustic values). 2016 measurements for target syllable VOT values (9 syllables $\times$ 2 conditions $\times$ 16 speakers $\times$ 7 reps) were obtained.[3] In instances where stop consonants were nasalized and/or voiced, there was no measurable VOT according to the criteria defined.[4] Three tokens from one speaker and two tokens from another speaker were omitted due to a misproduction. Except for these five cases, no outliers or other tokens were omitted. A total of 8723 datapoints were used for data analysis.

[3] The potential adaptation of the task with alternating focus and post-focus blocks was validated using a linear mixed effects model with repetitions as a fixed effect and subjects and items as random effects. For all measurements, there was no significant effect of repetitions (all at $p > .01$), suggesting that participants did not adapt to the given task over the course of the experiment.

[4] Out of 336 tokens, 17 items (5%) had no measurable VOT under focus. In post-focus conditions, 92 out of 336 items (27.4%) were voiced/nasalized.

Based on prior listening to pilot materials, the word containing a focused segment was expected to form its own prosodic phrase without triggering pre- or post-focal de-phrasing. The underlying phrases of the target word and its neighboring words in the absence of focus deploy a {LLH}{LLH}{HLH} tonal pattern (curly brackets indicate phrase boundaries). In the experimentally collected productions, the observed pitch contours from the recorded utterances of all speakers verified that there was no pre- and post-focal de-phrasing due to focal prominence, exhibiting the expected tri-phrasal {LLH}**{LLH}** {HLH} pattern. That is, the post-focal word maintained the underlying HLH pattern rather than a de-phrased LLH pattern [e.g., {LLH}{HLH} → {LHLLLH}] that would have occurred if de-phrasing had happened. Since the tri-syllabic word is in contrastive focus, a re-structuring of the phrasing would have resulted in a pitch peak placed on the second syllable (Jun & Lee, 1998), but as focus in this study is induced on the third syllable by segmental contrast, the expected natural re-phrasing might not be applicable. Fig. 3 shows two example f0 contours for focus and post-focus material. In the focus condition (indicated in blue), the target word is followed by a pause and the upcoming material begins with a relatively high tone, which correlates with the expected accentual pattern of the tense-initial phrase. This pitch excursion together with a pause indicates that the target word is IP-final (Jun, 1998). In the post-focus condition (in black), no pause is exhibited but the upcoming material again has a relatively high tone, which would not be the case if it had been de-phrased.

2.5. Data analysis

The values of each dependent measure were analyzed using the *lmer* function in the *lme4* package (Bates, Maechler, Bolker, & Walker, 2015) in R (R Core Team, 2017). Linear mixed-effects regression analyses were performed in two ways: i) to test for the syllable structure focus effect, focus condition (focus vs. post-focus), instructed focused syllable position (onset vs. nucleus vs. coda), and their interactions were chosen as fixed effects and ii) to examine the segment compositional effect of focus, focus condition (focus vs. post-focus), focused segment (/t, t^h, t*/ or /n, t, t^h, t*/), and the interaction of the two were selected as fixed effects. All models included subjects and items as random effects. The statistical significance testing of the main effects and the interaction was conducted by likelihood ratio tests, comparing the full model with the effect in question against the reduced model without the effect in question (Bolker et al., 2009; Pinheiro & Bates, 2000). For instance, a model with both the focus condition and the syllable position effects is compared with a model without the focus effect to test whether the inclusion of focus in the model yields a better estimation of the data. When the effect in question had more than two levels and was found to be a significant contributor in the model, we used the Satterthwaite approximation method incorporated in the *lmerTest* package (Kuznetsova, Brockhoff, & Christensen, 2013) in R to perform statistical analyses on multiple levels within a single model. The level of statistical significance for all tests was set as $p < .01$. R syntax for each statistical model is given in Appendix A.

3. Results

3.1. Focus and syllable structure effects

The first hypothesis predicts that onset and nucleus focus effects (as compared to the control post-focus condition) would exhibit similar patterns to one another, differing from the pattern elicited by focus in the coda position (given the coupling relations of CV [in-phase] versus VC [anti-phase]). However, if the domain of focus is larger—i.e. syllable-sized (regardless of the instructed segmental focus position)—the lengthening patterns would not be predicted to be sensitive to which particular sub-syllabic unit is elicited as contrastively focused. Linear mixed models fitted by maximum likelihood were compared to test the effects of focus condition (focus, post-focus) and focus position (onset, nucleus, coda) of the target syllable. Stop onset closure duration, VOT, vowel duration, and f0max are reported individually below.

3.1.1. Stop onset closure duration

Onset closure duration of the lenis stop onset [t] in target syllable /tat/ with varying syllable-internal focus positions is investigated. (The measurements for aspirated and fortis stop onsets are not included as these have different intrinsic closure durations compared to that of lenis onset stops. Focus effects on closure duration for each distinct onset consonant are reported in Section 3.2.1). The onset lenis stop closure duration data are shown in Fig. 4. Note that *post-focus* refers to *unfocused* (control) counterparts of the focused syllable.

Focus condition affects onset closure ($\chi^2(1) = 143.75$, $p < .001$), lengthening the duration by 8.11 ms ± 0.64 (std. error) under focus compared to post-focus conditions. The effect of position on closure duration measures, however, is not found to be significant ($\chi^2(2) = 4.11$, $p = .128$), nor is the interaction of focus and focus position ($\chi^2(2) = 2.93$, $p = .231$).

Focus lengthens the closure duration of the onset stop regardless of where the focused segment is located within a syllable. That is, onset, nucleus, and coda focus all had the same basic effect on onset closure lengthening, suggesting that the minimal size or scope of focus is bigger than a segment-only sized interval. This argues against the syllable substructure focus hypothesis.

3.1.2. Voice onset time (VOT)

In this section, VOT values as a function of focus and syllable position are analyzed specifically for voiceless lenis [t] that occurs in onset in target syllables /tat/ and /tan/ (Fig. 5).[5]

The effect of focus condition and focus position on onset [t] VOT of /tat/ examined by linear mixed effects regression model comparisons shows that there exists no interaction effect of focus and syllable position ($\chi^2(2) = 1.35$, $p = .510$). In addition, there is no main effect of focused syllable position (onset, nucleus, vs. coda) ($\chi^2(2) = 4.05$, $p = .132$). However, the main effect of focus is significant ($\chi^2(1) = 141.13$, $p < .001$). This indicates that the onset consonant's VOT is significantly length-

[5] Just as for the report on onset closure duration, measures for other stop consonants (aspirated and fortis) not included in this the analysis as they have different intrinsic VOT values from the lenis stops (in focus conditions: 27.3 ms ± 18.4 [std. dev] for /t/, 42.8 ms ± 17 [sd] for /t^h/, and 10.2 ms ± 4.1 [sd] for /t*/; in post-focus conditions: 14.8 ms ± 4.6 [sd] for /t/, 29.2 ms ± 10 [sd] for /t^h/, and 13.3 ms ± 4.1 [sd] for /t*/ in the current data). Focus effects on VOT for each individual onset consonant are reported in section 3.2.1.

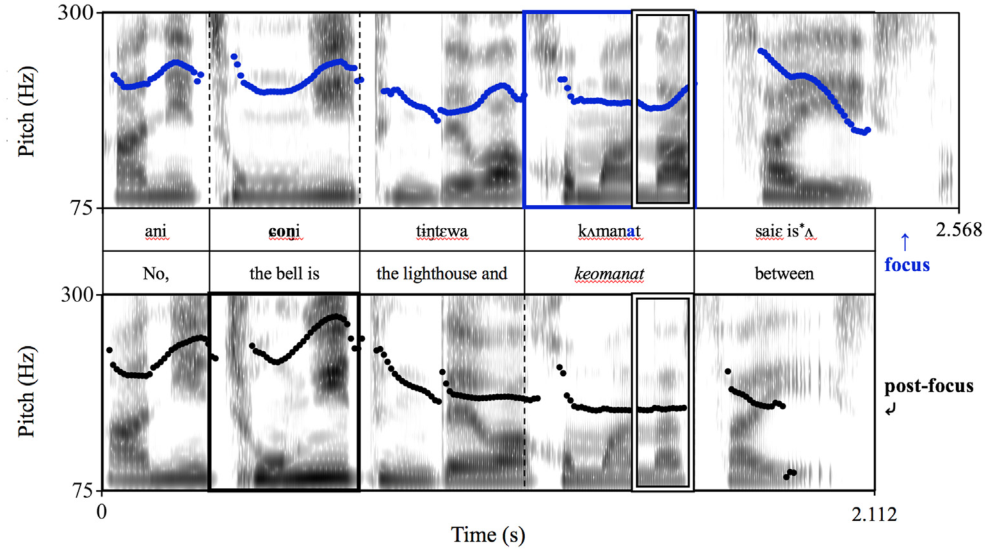


Fig. 3. Spectrogram and pitch tracks for two utterances: (i) focus material with focus on the target nonce-word [kʌmana**t**] (in blue) and (ii) post-focus material with focus on the preceding word *the bell* (in black). The measurement was always taken from the white outlined box region. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

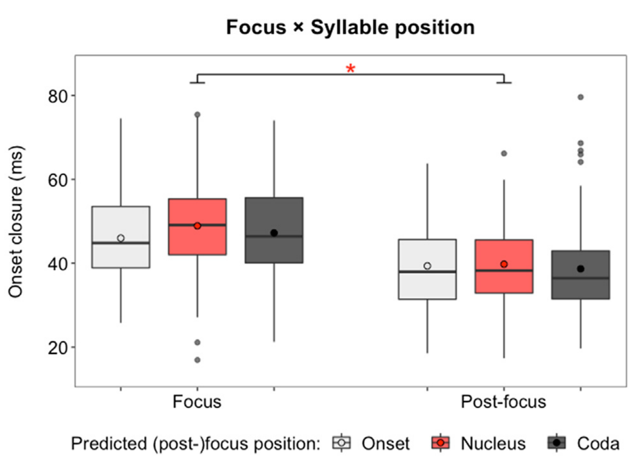


Fig. 4. Focus and focus position effects on onset closure for lenis [t] in /tat/ (In this figure and throughout, bars indicate the median with a horizontal line; the mean with a color-coded dot inside each box; interquartile range [IQR] with the box, and intervals between minimum and maximum values within 1.5 × IQR with the vertical bars.)

ened, by 13.2 ms ± 1.04 (std. error), under focus regardless of which sub-syllable position (onset, nucleus, or coda) is correctively focused.

Likewise, the comparison of nucleus and coda focus position for onset /t/ VOT in /tan/ shows that there is a significant main effect of focus condition ($\chi^2(1) = 101.36$, $p < .001$), lengthening VOT by 19.8 ms ± 1.83 (std. error) under focus. However, no effect of focus position ($\chi^2(1) = 1.57$, $p = .210$) nor any interaction between the two factors ($\chi^2(1) = 1.32$, $p = .251$) is found (Two male speakers' results [M5 & M7] were unavailable in the current report as they did not have any instance of measurable positive VOT in one of the post-focus conditions.)

It is worth noting that in *post-focus* conditions, voiceless lenis stop onset was frequently reduced and became voiced so that VOT values were not available. This is consistent with the possibility that *under focus* there is lesser overlap between the [t] and its tautosyllabic vowel, allowing the more robust glottal opening gesture for /t/ to yield a longer VOT interval (see Byrd & Saltzman, 2003, regarding such a mechanism at prosodic phrase edges).

Overall, similarly to the results for onset closure, the results for VOT of the target syllable's onset consonant imply that the scope of focus effect is larger than a single segment that is focused, given that both corrective onset and coda focus lengthen VOT of the syllable's onset consonant.

3.1.3. Vowel duration

Fig. 6 presents the effects of focus condition (focus, post-focus) and focus position (onset, nucleus, coda) of the target syllable /tat/.[6]

The model comparisons reveal that the interaction effect of focus condition and focus position is not significant ($\chi^2(2) = 2.98$, $p = .226$); nor is the simple main effect of focus position ($\chi^2(2) = 3.54$, $p = .170$). The main effect of focus condition is significant, lengthening vowel duration by 10.62 ms ± 1.03 (std. error) under segmental focus (mean vowel duration in focus: 88.9 ms ± 25.7 [sd]; in post-focus: 78.3 ms ± 15.5 [sd]; $\chi^2(1) = 99.37$, $p < .001$).

Again, the findings argue against a syllable structure-sensitive focus effect in that the scope of the focus effect does not appear to be specific to a segment-size interval. In sum, correctively focusing a segment within a CVC syllable indeed lengthens the duration of that syllable's onset closure, VOT, and vowel, but the effect is the same regardless of where within a syllable the focus is induced (i.e., on onset, nucleus, or coda position).

[6] It is worthwhile to note that comparing vowel duration with onset /t/ to that with coda /t/ in the target /tat/ syllable is an imperfect comparison, as tense consonants are neutralized to a lenis stop in the coda position phonotactically; moreover, the coda /t/ was frequently deleted in the post-focus condition. A direct comparison between onset and coda focus for an *identical* consonant is possible in this dataset in the case of focusing /n/ in the onset of /nat/ and the /n/ in the coda of /tan/; of course this has other imperfections. That said, separate analyses of /nat/ and of /tan/ shows that focusing the onset [n] lengthens vowel duration (onset /n/: $t = 8.46$, $p < .001$), coda /n/: $t = 5.77$, $p < .001$).

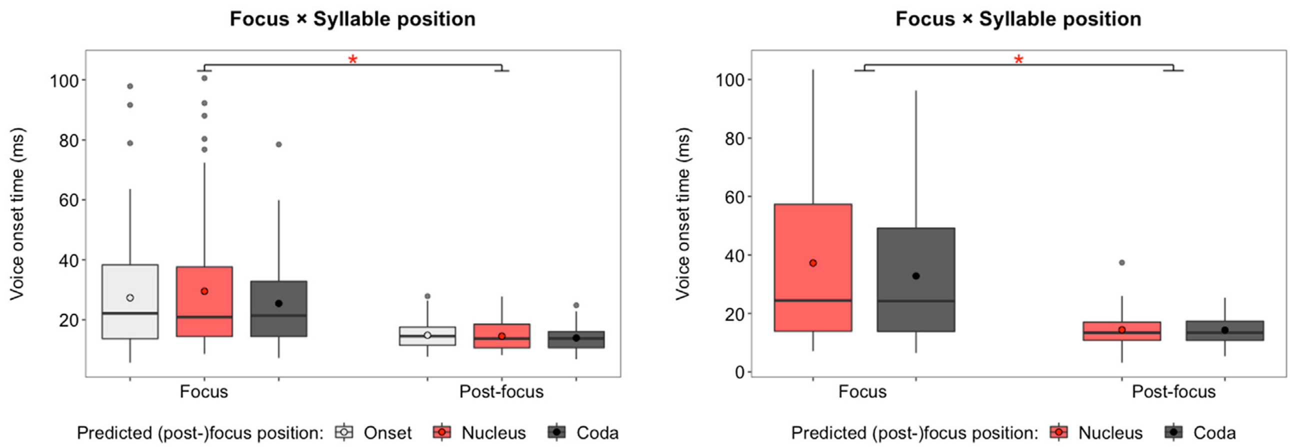


Fig. 5. Focus condition and position effects on VOT for lenis onset [t] (left: /tat/, right: /tan/).

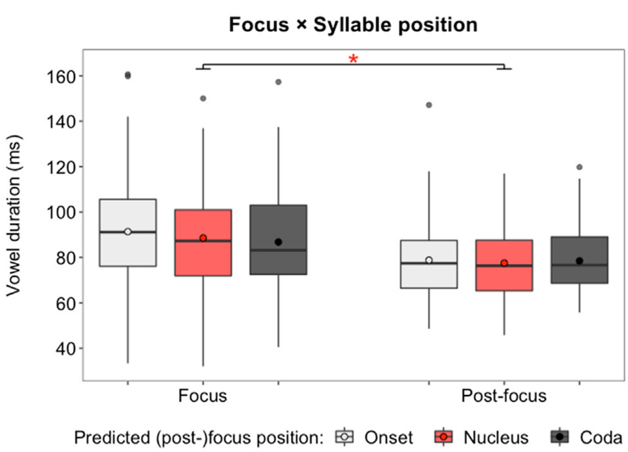


Fig. 6. Focus condition and focus location effects on vowel duration (ms).

3.1.4. F0 maximum

In a consideration of f0 maximum (f0max), female and male speakers are analyzed separately. Linear mixed effects model comparisons on f0max values of female speakers during the vowel (for /nat, tat, t^h^at, t*at/) reveal a significant main effect of focus condition, raising f0 by 41.9 Hz ± 1.6 (std. error) under focus ($\chi^2(1) = 500.37$, $p < .001$). There is no interaction effect of focus condition and focus position ($\chi^2(1) = 0.10$, $p = .747$) and no main effect of focus position ($\chi^2(1) < 0.01$, $p = .993$) (Fig. 7: left).

Likewise, f0max for male speakers similarly show a significant main effect of focus condition, lengthening f0 by 23.2 Hz ± 0.9 (std. error) under focus ($\chi^2(1) = 478.44$, $p < .001$), and like the female speakers, there is no main effect of focus position ($\chi^2(1) = 0.02$, $p = .898$) nor interaction of focus condition and position ($\chi^2(1) = 0.03$, $p = .867$) (Fig. 7: right).

Maximum fundamental frequencies in both female and male speakers were higher in the focused condition, indicating that correctively focusing any segment within a syllable, such as its onset or nucleus, enlarges the fundamental frequency excursion in the syllable's nuclear vowel.

3.2. Focus and segment compositional effects

The second main hypothesis of the study predicts that the prosodic focus gesture will modulate all concurrent gestures in its domain. That is, the observed acoustic reflections of a prosodic focus gesture will be a function of the specific articulatory gestures co-active with the focus event because the focus gesture affects all and only the gestural activations within its domain, i.e. acts "transgesturally" (Byrd & Saltzman, 2003). In this section, the measurements of stop onset closure duration, VOT, vowel duration, and f0 maximum are analyzed so as to investigate the segment-specific details of focus realization. In this portion of the analysis, the various onset consonants included in the stimuli's target syllable are examined— i.e., the three contrasting oral stops of Korean (/t, t^h, t*/), which are often grouped into lax (lenis /n, t/) and tense (aspirated /t^h/ and fortis /t*/); because of the available distribution in the stimuli, they are examined under onset focus (and post-focus) conditions only.

3.2.1. Stop onset closure duration and VOT

The interaction of focus condition and segment identity on stop closure duration was significant ($\chi^2(2) = 79.17$, $p < .001$), as was main effects of focus condition ($\chi^2(1) = 287.28$, $p < .001$) and segment identity ($\chi^2(2) = 20.32$, $p < .001$) (Fig. 8). Closure durations for all onset segment were lengthened under focus, as revealed in regression models for each onset consonant (Table 2). Regarding the interaction effect, the amount (ms) of lengthening under focus increased as the stop closure duration itself increased—i.e. the longer the closure interval of the onset stop consonant the more that interval lengthened when it was focused (/t/ < /t^h/ < /t*/).

Overall, this finding is consistent with the expectation that all the stop onsets exhibit the lengthening effect of focus on their stop closures, presumably due to local slowing of the consonant articulator, in this case, the tongue tip.

In examining the VOT of the onset stop consonant, the interaction between focus condition and segment identity is significant ($\chi^2(2) = 85.31$, $p < .001$), as well as of course the main effects of focus ($\chi^2(1) = 65.84$, $p < .001$) and segment identity ($\chi^2(2) = 16.12$, $p < .001$). Linear mixed models fit by maximum likelihood using Satterthwaite's method on VOT for lenis and aspirated stop onsets exhibit significant VOT lengthening under focus, lengthening VOT by 12.55 ms ± 1.91 (std. error) for /t/ and by 13.52 ms ± 1.36 (std. error) for /t^h/ (/t/: $t = 6.58$, /t^h/: $t = 9.93$, all $p < .001$). VOT of fortis stop onsets, however,

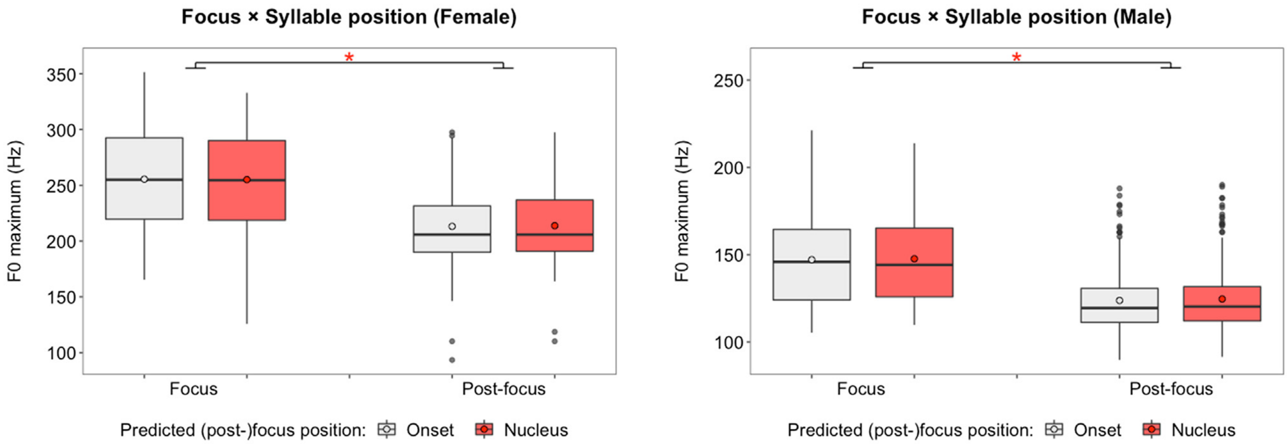


Fig. 7. Focus and focus position effects on f0 maximum (left: female, right: male).

is shorter in focus conditions, shortening VOT by 3.13 ms ± 0.41 (std. error) (/t*/: $t = -7.62$, $p < .001$) (Fig. 9).

It is worth noting that while a significant VOT shortening is found for the focused fortis onsets, the magnitude of this mean difference is quite tiny (3 ms). Furthermore, we observe larger variabilities of VOT values in the focused compared to the unfocused conditions for the lenis and the aspirated stops, while VOTs of the fortis stop show less variability than those of other two stop categories. This may be arising from inter-speaker variability; for onset /t/, by-speaker coefficients of variance, or intercepts, of VOT values range from 10.2 to 49.5 ms when focused and from 12.6 to 16.7 when unfocused. Alternatively, this pattern may be due to the unbalanced number of data points between focused and unfocused conditions, as VOT was frequently not observable in the latter condition.[7] The difference in variance among stops is compatible with contemporary Seoul Korean sound change patterns. Fortis stops clearly have had and continue to have a shorter VOT than lenis and aspirated stops, but the lenis versus aspirated distinction in VOT that used to exist prominently is diminishing, and this may yield larger inter-speaker variability.

The VOT results support an account of focus realization in which the glottal opening gestures of the lenis and aspirated stops undergo slowing under focus and perhaps lesser overlap with the upcoming vowel, while the absence of such a glottal opening gesture for fortis stops provides no such opportunity for lengthening its VOT values.

3.2.2. Vowel duration

The linear mixed effects model comparisons on vowel duration reveal a main effect for focus ($\chi^2(1) = 194.96$, $p < .001$), a main effect for onset segment identity ($\chi^2(3) = 17.94$, $p < .001$), and an interaction effect of the two ($\chi^2(3) = 23.75$, $p < .001$) (Fig. 10). A regression model for each segment (i.e., /n, t, t^h, t*/) shows that nuclear vowel duration is significantly lengthened under onset focus for all stop consonants (Table 3). Vowel lengthening under onset focus is predicted by the compositional focus hypothesis whereby all focused onset stops are accompanied by a slower activation of the tautosyllabic vocalic gesture that is concurrent to the focus gesture and being coproduced with the syllable's focus onset consonant production.

The lengthening of the vowel duration in the neighborhood of the aspirated stop has the smallest focus effect compared to other segments (7 ms). This could be arising from the fact that the glottal abduction interval producing the long positive VOT 'obscures' (via overlap) the initial portion of the vowel delaying the onset of the measured acoustic vowel duration. Notably, the fortis stop, which does not have a glottal abduction gesture, is accompanied by the longest lengthening of vowel duration under focus (19 ms).

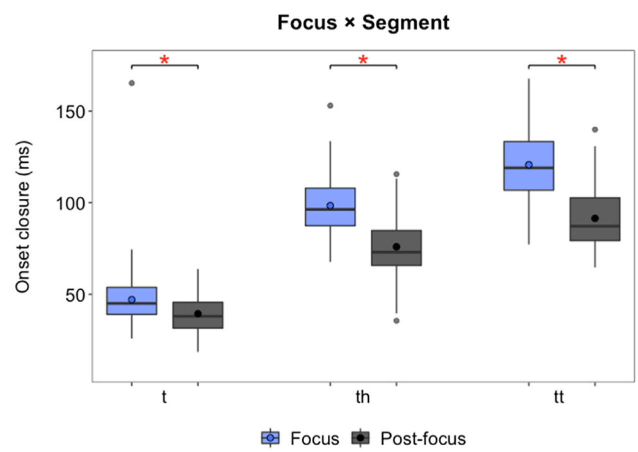


Fig. 8. Focus and individual segment effects on stop onset closure (aspirated stop /t^h/ and fortis stop /t*/ denoted as *th* and *tt*, respectively).

Table 2
Mean onset closure duration (ms) for each onset segment.

	Duration under focus (ms)	Lengthening under focus (std. err) (ms)	*t*-value
/t/	47.04	7.69 (1.52)	5.05**
/t^h/	98.36	22.45 (1.51)	14.86**
/t*/	120.69	29.29 (1.84)	15.93**

***p* < .001.

[7] Levene's tests for homogeneity of variance show that both the lenis and the aspirated stops have significantly different variances between the focused and unfocused conditions (/t/: $F(1, 185) = 44.5$, $p < .001$, /t^h/: $F(1, 222) = 31.3$, $p < .001$).

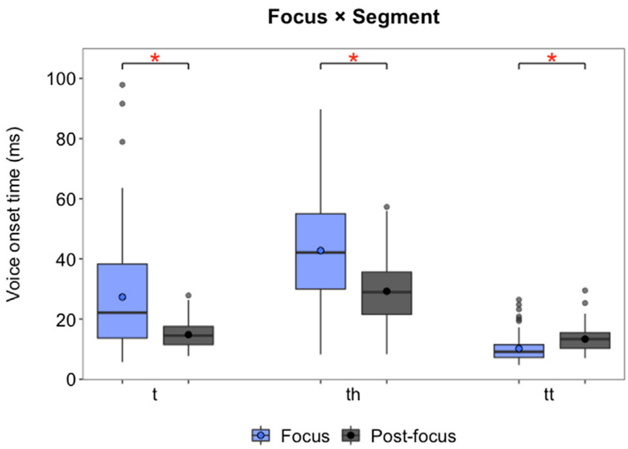


Fig. 9. Focus and individual segment effects on VOT (/t, t^h, t*/) (aspirated stop /t^h/ and fortis stop /t*/ denoted as *th* and *tt*, respectively).

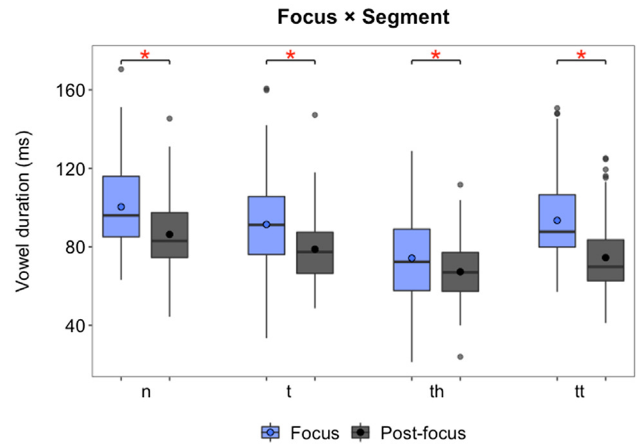


Fig. 10. Focus and individual segment (/n, t, t^h, t*/) effects on vowel duration (aspirated stop /t^h/ and fortis stop /t*/ denoted as *th* and *tt*, respectively).

3.2.3. F0 maximum

Under focal accent, f0 values have been seen cross-linguistically to be perturbed, and that perturbation may be a raising or a lowering. In this Korean dataset the general result for f0 maximum indicates that there is overall pitch raising in the focus condition (Fig. 10). For female speakers, f0max exhibits an interaction effect between focus condition and segment (discussed further below) ($\chi^2(3) = 79.76$, $p < .001$), and there are significant main effects of both focus condition ($\chi^2(1) = 237.68$, $p < .001$) and segment identity ($\chi^2(3) = 15.16$, $p = .002$) (Fig. 11: left). Individually, all female speakers but one raises their pitch under focus. Interestingly, one speaker has higher or lower pitch depending on which onset consonant is focused—tense onsets raise f0 and lax onsets lower f0. This individual result is considered further in the discussion section below. Linear mixed models using Satterthwaite approximations for each segment (/n, t, t^h, t*/) show that the tautosyllabic vowel for all stop segments undergo pitch raising in the onset focus condition (Table 4).

The results for the male speakers also show that the focus condition raises f0 maximum values compared to the f0 values in post-focus condition. There is a main effect of focus condition ($\chi^2(1) = 248.28$, $p < .001$) and a main effect of segment identity ($\chi^2(3) = 14.61$, $p = .002$), as well as the interaction of condition and segment (discussed further below) ($\chi^2(3) = 35.29$, $p < .001$) (Fig. 11: right). All of the individual speakers' results indicate that f0max is significantly raised under focus. Linear mixed effects regression models for each segment indicate that f0 values on vowels after each of the four segments are significantly higher under focus (Table 5).

The mean f0 maximum values and the amount of raising (i.e., the difference between focus and post-focus conditions) for each segment are summarized in Tables 4 (female) and 5 (male). Although (mean) f0 maximum is higher under focus in all segments, the degree of pitch raising varies as a function of the tenseness of the segment. For female speakers' lax stop onsets (/n/ and /t/), f0 is 20–25 Hz higher in the focus condition (similarly it is 15–20 Hz higher for male speakers). On the other hand, there is a > 60 Hz difference between focus and post-focus conditions for tense stops (/t^h/ and /t*/) produced by female speakers (similarly it is over a 30 Hz increase of pitch for male speakers).

Table 3
Mean vowel duration (ms) for each onset segment.

	Duration under focus (ms)	Lengthening under focus (std. err) (ms)	*t*-value
/n/	100.33	13.0 (1.65)	8.46**
/t/	91.36	12.55 (1.84)	6.82**
/t^h/	74.27	6.98 (1.58)	4.42**
/t*/	93.43	19.07 (1.48)	12.92**

***p* < .001.

The difference between lax/tense stops in their reflection of pitch raising under focus is also noticeable in the individual speaker's results in Table 6. All of the speakers show significant pitch raising when focusing tense (aspirated or fortis) stops (although 2 of the 16 speakers [F8 & M4] exhibit pitch raising under focus on just one or the other of the tense stops). However, only 6 speakers extend this to both the lenis /t/ and /n/, with another 6 speakers having no f0 raising under focus for either lenis stop (/n, t/) and 4 showing raising only for one of the two lenis stops.

The results on f0 maximum show that focusing a segment induces significant pitch raising regardless of which segment is focused. Additionally, we observe different degrees of pitch modulation effect between tense and lax onsets—tense onset consonants exhibiting more pitch raising and more consistent raising across speakers. This is in line with our assumption that tense onsets bear a H_{tenseC} gesture in their segmental composition.

4. Discussion

Focus is an instance of linguistic prominence in which speakers highlight new or important information. The goal of this study is to examine corrective focus effects on a syllable for varying focus positions and segments within that syllable. The results show that correctively focusing a single segment in a CVC syllable lengthens acoustic durations at a variety of locations within that syllable—lengthening onset closure, VOT, and vowel duration, and also raising f0 maximum values of the syllable's nuclear vowel. The nature of the observed focus effect in terms of lengthening and pitch perturbation is

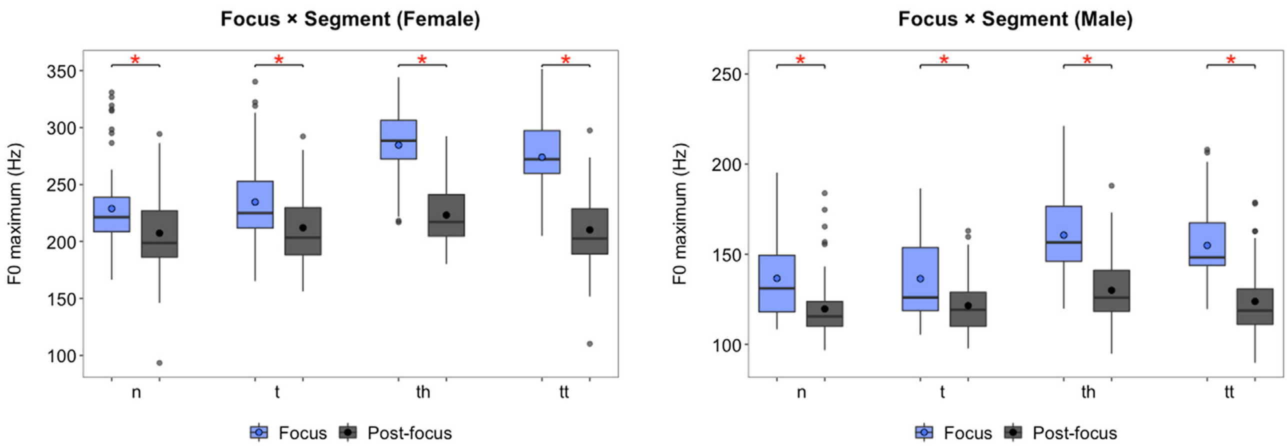


Fig. 11. Focus and individual segment (/n, t, t^h, t*/) effects on f0 maximum (left: female, right: male) (aspirated stop /t^h/ and fortis stop /t*/ denoted as *th* and *tt*, respectively).

Table 4
F0 maximum (Hz) of individual segments—female.

	F0max under focus (Hz)	Raising under focus (std. err) (Hz)	*t*-value
/n/	228.88	21.58 (4.06)	5.32**
/t/	234.65	22.59 (3.39)	6.66**
/t^h/	284.66	61.51 (3.69)	16.68**
/t*/	274.07	63.79 (4.68)	13.63**

***p* < .001.

Table 5
F0 maximum (Hz) of individual segments—male.

	F0max under focus (Hz)	Raising under focus (std. err) (Hz)	*t*-value
/n/	136.67	16.98 (2.21)	7.68**
/t/	136.46	14.93 (2.24)	6.67**
/t^h/	160.55	30.56 (2.64)	11.57**
/t*/	154.87	31.01 (2.60)	11.92**

***p* < .001.

consistent with the action of a focus gesture that engenders the slowing of gestural activations and its concomitant decrease in overlap, and increase in gestural magnitude. Additionally, the scope of the focus effect—which can be interpreted as the activation interval of the focus gesture—extends beyond the duration or activation interval of a single focused segment; the acoustic properties of surrounding gestures within the syllable are modulated as well as those of the focused segment itself. Focus effects span throughout the syllable even with corrective focus nominally or pragmatically placed on an individual segmental unit, suggesting that the scope of focus is syllable-sized, rather than segment-sized (as shown in Fig. 1 right, as compared to left). This supports the hypothesis that the smallest domain of prosodic prominence is a syllable (Van Heuven, 1994).

In future work the question of whether sub-syllable structure has effects on the details of focus production can also be addressed (albeit not in Korean) by examining focus in syllables having complex onset and coda consonant sequences. For example, one could examine the interaction between the focus effect and the 'c-center' effect (Browman & Goldstein, 1988; Byrd, 1995; Goldstein, Chitoran, & Selkirk, 2007; Nam et al., 2009). Given that multiple consonant onsets overlap only *in part* with the vowel production, focus effects—assuming a ramped-activation μ-gesture with its strongest activation being achieved in the middle of the syllable's vowel—will be stronger at the final (right) edge of an onset cluster than at its initial (left) edge given that the co-production or overlap of the cluster with the vowel is greater later in the onset sequence, within reach of the strong portion of the prosodic gesture. Conversely, in a lengthy coda cluster, the right edge is least likely to exhibit these focus effects as the prosodic gesture activation is waning (although its leftmost edge may show some small effect).

The results of the current study support the postulation of a transgestural influence of focus gestures. Due to different gestural composition for each of the three (oral) stop onsets in Korean, the focus effects are manifested in a systematically variable fashion (as represented in Fig. 2). Among lenis, aspirated, and fortis stops, the first two require a glottal opening gesture yielding a positive VOT, and the last two have a H_{tenseC} gesture that is related to raised pitch. As these gestures undergo activation-slowing through modulation by the focus gesture, the observed surface manifestations of the stops' component gestures are predicted to yield systematic acoustic variabilities. This prediction was supported in the current results that show that only lenis and aspirated stops have significant lengthening of VOT under focus. Lenis and aspirated stops have a glottal opening (GLO) gesture co-active with the nuclear tongue dorsum (TD) gesture, whereas fortis stops lack the presence of a GLO gesture. Accordingly, we observe VOT lengthening only on the former stops, though tautosyllabic vowel lengthening is seen in the vicinity of all stop consonants. These acoustic duration results are consistent with a gestural score in which the focus gesture affects all concurrently active gestures (whether GLO and/or TD gestures), as predicted by the transgestural prosodic modulation account hypothesized in Byrd and Saltzman (2003).

Next, let us turn to the prediction that only tense stops baring the H_{tenseC} gesture would show pitch raising under focus. All onset consonants exhibited focus effects of pitch raising on their nuclear vowel but to a different degree. This may be due to an interplay of the focus gesture and phrasal accent gesture(s) in Korean within its prosodic phrases (AP & IP). If there are interactions and competitions among multiple proso-

Table 6
Focus effects for individual segments on f0 maximum (Hz) for each speaker (Linear mixed model *t*-tests using Satterthwaite's method; gray shading: lack of pitch raising).

	Lax onsets		Tense onsets	
	/n/	/t/	/tʰ/	/t*/
	F0 raise (std. err)	F0 raise (std. err)	F0 raise (std. err)	F0 raise (std. err)
F1	−20.2 (2.2)**	−15.7 (6.1)	86.3 (4.0)**	83.2 (4.9)**
F2	54.5 (2.7)**	44.3 (5.2)**	93.6 (4.0)**	116.2 (7.4)**
F3	24.2 (11.2)	22.7 (4.1)**	92.8 (4.9)**	82.4 (8.1)**
F4	42.9 (4.9)**	40.2 (9.2)**	83.2 (5.0)**	80.1 (3.5)**
F5	10.5 (2.3)*	18.1 (2.2)**	15.5 (2.6)**	23.3 (4.4)**
F6	38.7 (15.6)	12.5 (8.5)	45.2 (7.7)**	26.1 (8.5)*
F7	34.2 (4.0)**	43.9 (7.0)**	36.0 (6.0)**	43.3 (5.2)**
F8	−11.1 (5.1)	14.7 (8.0)	39.6 (5.3)**	56.6 (21.2)
M1	11.4 (3.7)	3.1 (4.2)	18.4 (4.3)*	19.3 (4.7)**
M2	19.8 (8.6)	25.7 (4.5)**	23.8 (6.1)*	26.8 (7.6)*
M3	25.8 (3.1)**	2.2 (1.9)	29.5 (6.0)*	24.8 (4.5)**
M4	15.6 (2.8)**	13.1 (2.6)**	7.2 (2.7)	14.2 (1.8)**
M5	50.1 (3.3)**	51.8 (4.2)**	78.3 (5.8)**	83.7 (5.1)**
M6	1.4 (2.0)	6.7 (4.3)	29.2 (4.3)**	27.1 (3.1)**
M7	10.5 (2.7)*	12.0 (6.0)	30.0 (3.1)**	22.6 (3.9)**
M8	1.3 (1.9)	4.8 (1.9)	28.1 (4.7)**	29.6 (3.2)**

**$p < .001$, *$p < .01$.

dic and/or modulation gestures that are concurrently active, the μ-gestural approach (Saltzman et al., 2008) offers a plausible elaboration of the π-gesture framework. In this approach, the f0 pattern of Accentual and Intonational Phrases in Korean may be modeled by a syllable oscillator that alternates between Low and the High tonal gestures; these accentual tone gestures could potentially co-occur with focus gestures, represented with a μ-gesture that acts transgesturally in its domain or scope of activation.

As a side note, let us return to the one speaker in the current study who showed patterns different from the others in modulating the pitch-related gesture under focus. While 15 out of 16 speakers consistently raised their pitch under focus, Speaker F1 in contrast raised or lowered her pitch depending on the focused segment's tenseness—for lax onsets, f0 values of the tautosyllabic vowel were lower in focus conditions, whereas for tense onsets (aspirated and fortis stops), f0 was higher (Fig. 12).

Although we have postulated here that there is no tone gesture for lax onsets underlyingly, the possibility that this speaker plausibly has a $L_{\mathbf{laxC}}$ tone gesture in addition to a H_{tenseC} gesture must be entertained. The tonal patterns distinguishing lax and tense stops are participating in an on-going sound change phenomenon in Seoul Korean (Kang, 2014; Lee, 2018; Silva, 2006), and it is possible that two steps in this tonogenesis may arise. Namely the H_{tenseC} tone gesture identified in the intrinsic spatiotemporal representation of tense stops may also perhaps be subsequently (diachronically) followed up by a L_{laxC} tone gesture added into the system. In the current data, all but one of sixteen speakers have focal pitch raising on tense stops but no pitch *lowering* on lax ones under focal prominence, as would be expected for a H_{tenseC} gesture in the representation of tense stops but no corresponding tonal gestures associated with the lax stops. That said, the single speaker who does show bi-directional pitch modulation might be postulated to have both a H_{tenseC} gesture and a L_{laxC} gesture for tense and lax stops, respectively, causing higher or lower pitch values depending on the identity of the focused segment undergoing spatiotemporal expansion under focus.

If, as we and others suggest, multiple prosodic/modulation gestures can be co-produced, interesting questions arise as to how the phrasal gestures interact with modulations engendered by focus (see for example related discussion in Katsika et al. (2014), Katsika (2016), and Saltzman et al. (2008). Can focus gestures in Korean reveal ('de-neutralize') the tense/lax tonal contrast exhibited in phrase-initial position? Or will focus perhaps engender a separate AP (or IP) via the emergent creation of a phrasal boundary?

In the current study, the target focused syllables were located at the final syllable of the tri-syllabic pseudo-word with an LLH or an LHH Accentual/Intonational Phrase pattern, so the focused syllable always had an underlying High tone in its prosodic phrase. If the gesture associated with a phrasal prominence is dominant over the prosodic gesture associated with focus (perhaps via blending strength in a Task Dynamics implementation), target syllables in focus will uniformly present higher pitch values than preceding syllables due to the underlying High tone of the Accentual Phrase being weighed more heavily. Alternatively, if the prosodic focus gesture dominates the effects of the Accentual Phrase tonal patterns, there will be an expansion in the spatiotemporal properties of any concurrent pitch-related gestures.

Such a LH phrasal accent pattern can be modeled as an instance of a rhythmic oscillator, as proposed by Lee (2018). It is possible that when a focus gesture is present, it modulates not only tonal (and constriction gestures) associated with the stop consonants but also the phrasal accentual tonal gestures that are concurrently active as well. For example, in our fortis/aspirated initial CVC stimuli, a H_{tenseC} gesture is present in the intrinsic gestural organization along with an additional oscillating High tone co-active due to the target syllable's position within the Accentual/Intonational Phrase. On the other hand, a lax stop initial target has only one co-active High tone, the one that is due to the alternating oscillator associated with the Accentual Phrase f0 prosody. Therefore, the strength of output pitch raising due to focus effects in this phrasal position would be greater for tense stops (/tʰ, t*/) compared to lax stops (/n, t/), because the prosodic focus gesture—or μ-gesture—

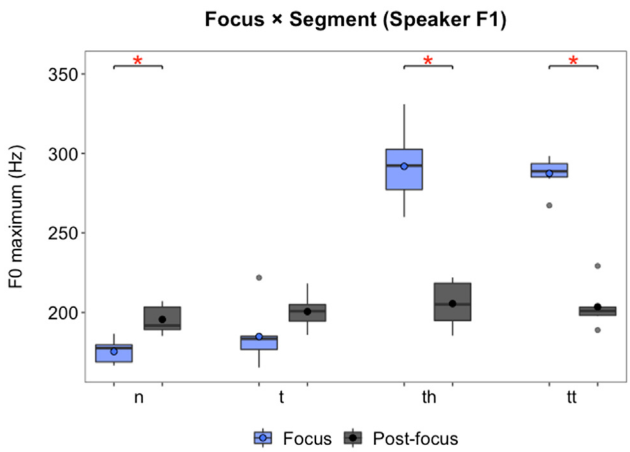


Fig. 12. Focus and individual segment effects on f0 maximum for Speaker F1.

can target both the intrinsic pitch raising gesture of the tense stop (i.e., H_{tenseC} gesture) as well as the accentual High tone gesture that occurs at this position in the Accentual Phrase.

An alternative interpretation of the results is that the informational structure (i.e., focus event) changes the prosodic phrase boundary structure. A focus gesture may give rise to an additional phrase boundary, perhaps due to lessened overlap among active gestures in its domain and/or perhaps due to the informational load it carries. This is in line with the Jun and Lee's (1998) claim that contrastive focus may initiate a separate Accentual or Intonational Phrase. In this scenario, there will be an interaction between the focus (μ_s) gesture and the π (μ_t)-gesture (that instantiates the 'spun off' phrasal boundary). Most importantly though, in both of these scenarios there are critically multiple prosodic gestures that are active (at least in languages with rhythmic stress/accentual patterns). There is good reason to pursue a theoretical approach that allows for the possibility of such overlapping, i.e. co-active, prosodic gestures, much as Katsika et al. (2014) and Katsika (2016) describe in their investigations of the interaction between lexical stress and boundary tone/lengthening for Greek.[8]

5. Conclusions

The current study investigates focus modulation effects in varying sub-syllabic domains. The acoustic data obtained in the current experiment suggests that prominence effects are manifested throughout a focused syllable with systematic variations depending on the specific gestural structures, i.e. consonants, that are active during the domain of a focal prominence. These results are consistent with an account in which a focus gesture modulates co-active articulatory gestures within its domain, slowing the activations of those gestures. The patterns of focus modulation did not differ with respect to varying elicited sub-syllabic focus positions within a syllable, consistent with the hypothesis that the smallest domain of prosodic prominence is a syllable. Lastly, there was also some evidence suggesting that multiple prosodic gestures (prominence gestures and accentual gestures) may be interacting with regard to the observed focus effects on f0 patterns.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by NIH Grant DC03172 (Byrd, PI) and the USC Department of Linguistics. The authors thank the reviewers, editor, and guest editors for their guidance, and thank Louis Goldstein and Elsi Kaiser for their comments on an earlier version of this study using a smaller dataset.

Appendix A. Statistical models

A.1. Focus condition × syllable position effects

Dependent variables: dv = {onset_closure_tat, vot_tat, vot_tan, vowel_tat, f0_female, f0_male}

Independent variables: Focus condition (condition) & focused syllable position (position)

Linear regression model specifications:

full.model = lmer(dv ~ condition * position +1(1|subject) + (1|item), data, REML = FALSE)

reduced.model = lmer(dv ~ condition + position + (1|subject) + (1|item), data, REML = FALSE)

condition.model = lmer(dv ~ condition + (1|subject) + (1|item), data, REML = FALSE)

position.model = lmer(dv ~ position + (1|subject) + (1|item), data, REML = FALSE)

Model comparisons:

Interaction effect: anova(full.model, reduced.model)

Main effect of focus condition: anova(reduced.model, position.model)

Main effect of syllable position: anova(reduced.model, condition.model)

A.2. Focus condition × segment effects

Dependent variables: dv = {onset_closure, vot, vowel, f0_f, f0_m}

Independent variables: Focus condition (condition) & focused segment (segment)

Linear regression model specifications:

full.model = lmer(dv ~ condition * segment + (1|subject) + (1|item), data, REML = FALSE)

reduced.model = lmer(dv ~ condition + segment + (1|subject) + (1|item), data, REML = FALSE)

condition.model = lmer(dv ~ condition + (1|subject) + (1|item), data, REML = FALSE)

segment.model = lmer(dv ~ segment + (1|subject) + (1|item), data, REML = FALSE)

Model comparisons:

Interaction effect: anova(full.model, reduced.model)

Main effect of focus condition: anova(reduced.model, segment.model)

[8] In the current study of Korean, it is worth noting that for most subjects no high tone activity was observed for the tense consonants in the post-focus condition. Silva (2006) has observed that phrase-internally the tonal difference between tense and lax Korean stops is neutralized. Our data is consistent with that observation.

Main effect of segment quality: anova(reduced.model, condition.model)

Appendix B. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.wocn.2019.100933.

References

Audacity Team. (2017). Audacity: Free audio editor and recorder (Computer software). Version 2.1.3. Retrieved from http://www.audacityteam.org/ (4 August 2017).

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.

Bates, D., Maechler, M., Bolker, B., & Walker, D. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48.

Baumann, S., & Winter, B. (2018). What makes a word prominent? Predicting untrained German listeners' perceptual judgments. *Journal of Phonetics, 70*, 20–38.

Bion, R. A. H. (2009). SegmentFiles.praat (Praat script). Version 3.0, Retrieved from https://sites.google.com/site/centerforinfantstudies/home/files/SegmentFiles.praat (13 October 2017).

Boersma, P. & Weenink, D. (2017). Praat: Doing phonetics by computer (Computer software). Version 6.0.28, Retrieved from http://www.praat.org/ (16 May 2017).

Bolinger, D. L. (1961). Contrastive accent and contrastive stress. *Language, 37*, 83–96.

Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H., & White, J. S. S. (2009). Generalized linear mixed models: A practical guide for ecology and evolution. *Trends in Ecology & Evolution, 24*(3), 127–135.

Botinis, A., Fourakis, M., & Prinou, I. (1999). Prosodic effects on segmental durations in Greek. *Sixth European conference on speech communication and technology.* Budapest, Hungary.

Browman, C. P., & Goldstein, L. (1988). Some notes on syllable structure in articulatory phonology. *Phonetica, 45*, 140–155.

Browman, C. P., & Goldstein, L. (1992). Articulatory phonology: An overview. *Phonetica, 49*, 155–180.

Byrd, D. (1995). C-centers revisited. *Phonetica, 52*(4), 285–306.

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

Byrd, D., Krivokapić, J., & Lee, S. (2006). How far, how long: On the temporal scope of phrase boundary effects. *Journal of the Acoustical Society of America, 120*, 1589–1599.

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

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

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

Cho, T., & Jun, S. A. (2000). Domain initial strengthening as enhancement of laryngeal features: Aerodynamic evidence from Korean. *Chicago Linguistics Society, 36*, 31–44.

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). Articulatory and acoustic studies on domain-initial strengthening in Korean. *Journal of Phonetics, 29*(2), 155–190.

Cho, T., Kim, D., & Kim, S. (2017). Prosodically-conditioned fine-tuning of coarticulatory vowel nasalization in English. *Journal of Phonetics, 64*, 71–89.

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

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

Cole, J., Hualde, J. I., Smith, C. L., Eager, C., Mahrt, T., & de Souza, R. N. (2019). Sound, structure and meaning: The bases of prominence ratings in English, French and Spanish. *Journal of Phonetics, 75*, 113–147.

R Core Team. (2017). R: A language and environment for statistical computing. In: R Foundation for statistical computing, Vienna, Austria. Version 3.5.1, Retrieved from http://www.R-project.org/ (16 May 2017).

Crosswhite, K. (2013). TextGridMaker.praat (praat script). Retrieved from http://www.ling.upenn.edu/courses/Fall_2013/ling520/TextGridMaker.Praat (13 October 2017).

Dart, S. N. (1987). An aerodynamic study of Korean stop consonants: Measurements and modeling. *Journal of the Acoustical Society of America, 81*(1), 138–147.

De Jong, K. (2004). Stress, lexical focus, and segmental focus in English: Patterns of variation in vowel duration. *Journal of Phonetics, 32*(4), 493–516.

De Jong, K., & Zawaydeh, B. (2002). Comparing stress, lexical focus, and segmental focus: Patterns of variation in Arabic vowel duration. *Journal of Phonetics, 30*(1), 53–75.

Fletcher, J. (1991). Rhythm and final lengthening in French. *Journal of Phonetics, 19*(2), 193–212.

Gao, M. (2008). Tonal alignment in Mandarin Chinese: An articulatory phonology account. *Ph.D. dissertation.* University of Southern California.

Goldstein, L., Byrd, D., & Saltzman, E. (2006). The role of vocal tract gestural action units in understanding the evolution of phonology. In Michael Arbib (Ed.), *Action to language via the mirror neuron system, 215-249. From action to language: The mirror neuron system*. Cambridge: Cambridge University Press.

Goldstein, L., Chitoran, I., & Selkirk, E. (2007). Syllable structure as coupled oscillator modes: Evidence from Georgian vs. Tashlhiyt Berber. In *Proceedings of the XVIth international congress of phonetic sciences* (pp. 241–244).

Goldstein, L., Nam, H., Saltzman, E., & Chitoran, I. (2009). Coupled oscillator planning model of speech timing and syllable structure. In G. Fant, H. Fujisaki, & J. Shen (Eds.), *Frontiers in phonetics and speech science* (pp. 239–250). Beijing: The Commercial Press.

Han, J.-I. (1996). *The phonetics and phonology of tense and plain consonants in Korean.* Ph.D. dissertation. Cornell University.

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

Hirose, H., Park, H. S., & Sawashima, M. (1983). Activity of the thyroarytenoid muscle in the production of Korean stops and fricatives. *Annales Bulletin RILP, 17*, 73–81.

Hong, K., Niimi, S., & Hirose, H. (1991). Laryngeal adjustments for the Korean stops, affricates and fricatives: An electromyographic study. *Annual Bulletin, Research Institute of Logopedics and Phoniatrics, 25*, 17–31.

Jun, S. A. (1993). *The phonetics and phonology of Korean prosody.* Ph.D. dissertation. The Ohio State University.

Jun, S. A. (1995). Asymmetrical prosodic effects on the laryngeal gesture in Korean. In B. Connell & A. Arvaniti (Eds.), *Papers in laboratory phonology IV: Phonology and phonetic evidence* (pp. 235–253). Cambridge: Cambridge University Press.

Jun, S. A. (1998). The accentual phrase in the Korean prosodic hierarchy. *Phonology, 15*(2), 189–226.

Jun, S. A. (014). Prosodic typology: By prominence type, word prosody, and macro-rhythm. In S.-A. Jun (Ed.), *Prosodic typology II: The phonology of intonation and phrasing*. Oxford University Press.

Jun, S. A., & Kim, H. S. (2007). VP focus and narrow focus in Korean. In *Proceedings of the XVIth international congress of phonetic sciences* (pp. 6–10).

Jun, S. A., & Lee, H. J. (1998). Phonetic and phonological markers of contrastive focus in Korean. *Fifth International Conference on Spoken Language Processing (ICSLP).* Sydney, Australia.

Kagaya, R. (1971). Laryngeal gestures in Korean stop consonants. *Annual Bulletin Research Institute of Logopedics and Phoniatrics, University of Tokyo, 5*, 15–23.

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

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

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), 3909–3917.

Katsika, A. (2016). The role of prominence in determining the scope of boundary-related lengthening in Greek. *Journal of Phonetics, 55*, 149–181.

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*, 62–82.

Keating, P. A., Cho, T., Fougeron, C., & Hsu, C.-S. (2003). Domain-initial strengthening in four languages. In *Papers in laboratory phonology VI: Phonetic interpretations* (pp. 145–163). Cambridge: Cambridge University Press.

Kim, C. W. (1965). On the autonomy of the tensity feature in stop classification (with special reference to Korean stops). *Word, 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.

Kim, H., Maeda, S., Honda, K., & Crevier-Buchman, L. (2018). The mechanism and representation of Korean three-way phonation contrast: External photoglottography, intra-oral air pressure, airflow, and acoustic data. *Phonetica, 75*(1), 57–84.

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

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2013). lmertest: Tests for random and fixed effects for linear mixed effect models (lmer objects of lme4 package). R Package Version 2.0-3.

Lee, Y. (2018). *The prosodic substrate of consonant and tone dynamics.* Ph.D. dissertation. University of Southern California.

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.

Lee, Y. C., & Xu, Y. (2010). Phonetic realization of contrastive focus in Korean. In *Speech Prosody 2010* (pp. 1–4). Chicago, USA.

Lennes, M. (2004). Collect_data_from_two_tiers_in_files.praat (Praat script). GNU General Public License. Retrieved from http://www.helsinki.fi/~lennes/praat-scripts/#calculate (13 October 2017).

Löfqvist, A., & Gracco, V. L. (1999). Interarticulator programming in VCV sequences: Lip and tongue movements. *The Journal of the Acoustical Society ofAmerica, 105*(3), 1864–1876.

Maran, L. R. (1973). On becoming a tone language: A Tibeto-Burman model of tonogenesis. In L. M. Hyman (Ed.), *Consonant types and tone, Southern California occasional papers in linguistics* (pp. 98–114).

Mücke, D., & Grice, M. (2014). The effect of focus marking on supralaryngeal articulation–Is it mediated by accentuation? *Journal of Phonetics, 44*, 47–61.

Mücke, D., Nam, H., Hermes, A., & Goldstein, L. (2012). Coupling of tone and constriction gestures in pitch accents. *Consonant clusters and structural complexity.* Munich, Germany: de Gruyter.

Nam, H., Goldstein, L., & Saltzman, E. (2009). Self-organization of syllable structure: A coupled oscillator model. *Approaches to Phonological Complexity*, 299–328.

Nam, H., Saltzman, E., Krivokapić, J., & Goldstein, L. (2008). Modeling the durational difference of stressed vs. unstressed syllables. *Proceedings of the 8th phonetic conference of China (PCC 2008)*. Beijing, China.

Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effects models in S and SPLUS*. New York: Springer.

Saltzman, E. L. (1995). Dynamics and coordinate systems in skilled sensorimotor activity. In *Mind as Motion: Explorations in the dynamics of cognition* (pp. 149–173).

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

Saltzman, E., Nam, H., Krivokapić, J., & Goldstein, L. (2008). A task-dynamic toolkit for modeling the effects of prosodic structure on articulation. In *Speech Prosody 2008* (pp. 175–184). Campinas, Brazil.

Shipp, T., & McGlone, R. E. (1971). Laryngeal dynamics associated with voice frequency change. *Journal of Speech and Hearing Research, 14*, 761–767.

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

Smith, C.L., Erickson, D., & Savariaux, C. (2019). Articulatory and acoustic correlates of prominence in French: Comparing L1 and L2 speakers. *Journal of Phonetics, 77*, 100938.

Van Heuven, V. J. (1994). What is the smallest prosodic domain? *Phonological Structure and Phonetic Form: Papers in Laboratory Phonology, 3*, 76–97.

Zahner, K., Kutscheid, S., & Braun, B. (2019). Alignment of f0 peak in different pitch accent types affects perception of metrical stress. *Journal of Phonetics, 74*, 75–95.