Journal of Phonetics 107 (2024) 101353

Contents lists available at ScienceDirect

Journal of Phonetics

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

Focus-induced tonal distribution in Seoul Korean as an edge-prominence language

Richard Hatcher [a], Hyunjung Joo [a,b,c], Sahyang Kim [d], Taehong Cho [a,b,*]

[a] Hanyang Institute for Phonetics and Cognitive Science, Hanyang University, Seoul, Republic of Korea
[b] Department of English Language and Literature, Hanyang University, Seoul, Republic of Korea
[c] Rutgers University, The State University of New Jersey, USA
[d] Hongik University, Seoul, Republic of Korea

ARTICLE INFO

Article history:
Received 21 November 2023
Received in revised form 18 June 2024
Accepted 6 August 2024
Available online 6 September 2024

Keywords:
Focus realization
Edge-prominence language
Seoul Korean
Phrasing
Tone–segment alignment
Hyperarticulation

ABSTRACT

This study investigates the phonetic realization of contrastive focus in short utterances in Seoul Korean, a so-called 'edge-prominence' language, which is assumed to express focus-induced prominence primarily through phrasing. The study explores how the distribution of phrase-level tones and their realization is influenced by focus in different positions of target words with different coda segmental makeups (/pam, pap/). Phrase-initially, focus displays a typical phrase-initial f0 rise for the L and H tones, with the L tone anchored to the focused monosyllabic word and the H tone to the following syllable, accompanied by a tonal expansion. This expansion results from an elevated f0 peak for the H while the L remains unchanged, showing tonal hyperarticulation only in the H tone. Phrase-medially, a similar f0 rise occurs under focus, but without robust tonal expansion. Crucially, the f0 rise is not accompanied by clear temporal or tonal evidence for the creation of a new phrase, demonstrating focus realization without phrasing. Phrase-finally, focus also shows no phrasing evidence. It results in an f0 fall, possibly due to tonal crowding of the L and H tones with the upcoming low boundary tone. However, this fall is distinct from a similar fall under no focus, suggesting a phonetic trace of the focal rise. Both initially and medially, the tonal realization of the f0 rise is affected by the segmental makeup (/pap/ vs. /pam/) only at the microprosodic level while maintaining the tonal targets, even in the face of physically adverse conditions for an f0 rise through the voiceless gap. The findings of the present study illuminate the intricate phonetic details of focus realization with a f0 rise in a language other than the well-studied West Germanic and Romance languages which employ word-level stress. The findings also shed new light on the relationship between focus and prosodic phrasing, implying that focus, previously argued to drive prosodic phrasing in Seoul Korean, is just one of several potentially competing structures that determine a sentence's phrasing, thereby underscoring the multidimensional nature of prosodic structure.

1. Introduction

Prosody fulfills two primary functions within speech: the organization of speech elements into hierarchically structured units and the highlighting of certain elements of an utterance as prominent (Beckman, 1996; T. Cho, 2022; Shattuck-Hufnagel & Turk, 1996). Prominence, when understood within this broad construal, can be understood as the accentuation of specific elements to assist listeners in identifying the key components of an utterance (cf. Cangemi & Baumann, 2020). While a speaker conveys prominence by modulating the phonetic realization of both the segmental and suprasegmental features of the prominent element in a language-specific manner, prominence-related prosodic typologies across languages have been discussed primarily in terms of how prominence is signaled by f0 patterns in conjunction with the intonational phonology of a given language (Jun, 2005, 2014b). For example, in English and German, prominence has been observed to influence the f0 contour in such a way as to raise the f0 of the prominent elements and/or reduce the f0 of non-prominent elements (Féry & Kügler, 2008; Liberman & Pierrehumbert, 1984; Xu & Xu, 2005).

* Corresponding author.
E-mail address: tcho@hanyang.ac.kr (T. Cho).

https://doi.org/10.1016/j.wocn.2024.101353

Even within the group of languages such as English and German classified under a similar prosodic type (e.g., Jun, 2005; Ladd, 2008), there exist idiosyncratic characteristics in f0 realization (e.g., how tones align with segmental materials), which vary not only across languages but also across dialects of the same language (refer to Chapter 5 of Ladd, 2008 for a related discussion). For example, prenuclear accentual rises occur later in German than in English, and they occur later in Southern German than in Northern German (e.g., Atterer & Ladd, 2004). These fine-grained yet systematic differences in f0 realizations across languages imply that although languages may be grouped together in phonological terms, the detailed tonal distribution in relation to signaling prominence is controlled by the speaker and may be incorporated into the phonetic grammar of the language (see Ladd, 2008; Cho, 2015 for a related discussion). Thus, studies that investigate the phonetic intricacies of how a language signals prominence enhance our understanding of linguistic prominence systems, revealing both language-specific features and typological similarities in how prominence is marked.

However, such studies predominantly arise from research on languages classified as so-called 'head-prominence' languages, characterized by robust pitch accent systems. In these languages, a lexically stressed syllable serves as the head of prominence, upon which a phrase-level pitch accent falls. Many other languages, in contrast, do not employ (post-lexical) pitch accent systems, but instead utilize various features of intonational phonology such as the phonetic expansion of a lexical pitch accent or prosodic phrasing to convey prominence (see Jun, 2014b, for a review of related prosodic typologies). Consequently, numerous questions remain unanswered about how languages, particularly those less studied than well-known head-prominence languages, express prominence in aspects such as the exact timing of tone–segment alignment and f0 scaling. This gap underscores the need for additional studies to broaden the scope of research into languages with diverse prosodic features, further exploring the relationship between the fine-grained phonetic details and the broader, macroscopic level of a language's intonational phonology.

In the present study, we examine the detailed f0 patterns and their alignment with segmental materials involving prominence marking in Seoul Korean to deepen our understanding of the phonetic intricacies of prominence marking in languages other than 'head-prominence' languages. While a comprehensive understanding of a language's prominence system extends beyond the scope of a single study, our aim is to take a significant step forward by concentrating on particular instances where prominence is associated with focus—a key feature typically underscored by elevated prosodic prominence.

1.1. Background on Seoul Korean intonation

In Seoul Korean, intonational patterns play a pivotal role in marking the language's prosodic structure (Jun, 1996, 1998). According to Jun's (2005) model of intonation, the prosodic structure of Seoul Korean encompasses at least two primary prosodic levels above the word. The level immediately above the word in Seoul Korean's prosodic structure is known as the Accentual Phrase (AP). This phrase is canonically characterized by two bitonal pitch sequences that indicate its boundaries, represented as [#TH. . .LH#].[1] However, the distribution of these tones can vary depending on the AP's phonological length. For APs that are shorter than four syllables, it is common for one or both of the internal tones to be omitted (Jun, 2005), indicating that there is some correlation between the number of syllables in an AP and the manifestation of AP tones. In contrast to phrases cross-linguistically, Korean APs are generally assumed to be marked primarily by an LH (rising) tone on the right edge without substantial lengthening (Jun, 1996), or if any, with a relatively small lengthening effect (Cho & Keating, 2001). Furthermore, as illustrated in Fig. 1, APs are nested within a larger prosodic structure called the Intonational Phrase (IP). The intonational pattern of an IP is primarily shaped by the combined contours of its constituent APs. However, the model posits that an IP-final boundary tone, indicated by the symbol '%', overrides the terminal edge tone of the rightmost AP, which is aligned with the right edge of the Intonational Phrase (IP).

Thus, the intonational system of Seoul Korean is characterized by intonational tones that are closely tied to prosodic boundaries. In Jun's phonological model, unlike head-prominence languages where a pitch accent marks a lexically stressed syllable, Seoul Korean has no specific intonational marking for prominent elements within its prosodic structure. Instead, the model posits that in Seoul Korean, prominence is conveyed through phrasing, so that the prosodic unit that is aligned with the left edge of a prosodic phrase, an Accentual Phrase, is typically assumed to convey prominence. (For this reason, Jun (2014a) calls Seoul Korean a "head-less AP language.") These characteristics appear to provide a basis for classifying Seoul Korean as an 'edge-prominence' language (Jun, 2005, 2014a). In theory, Seoul Korean, as an 'edge-prominence' language, is defined as a language where the edges of a prosodic phrase are marked by intonational tones, without additional refinement of prominence within the phrase.

In line with this prosodic characterization, existing research on Korean intonation has indeed shown that focus, particularly contrastive focus which is generally assumed to carry more prominence than non-contrastive focus (e.g., Bishop, 2013; Gussenhoven, 2007; Mücke & Grice, 2014), can be effectively conveyed through prosodic phrasing (Jeon & Nolan, 2017; Jun, 1996; Kim et al., 2006). Largely in line with these observations, Jun (1996) proposes that focus, arising from the information structure of an utterance, triggers the insertion of a prosodic boundary immediately before the focused element, generally followed by the dephrasing of post-focal words.[2] More recently, Jun (2007, 2011) has proposed a theory that maps focus directly onto prosody. According to this theory, the boundary inserted due to focus, arising from the information structure, forms an Intermediate Phrase (iP), which is presumed to be larger than an AP boundary, given its heightened phonetic salience compared to that of a typical AP boundary. However,

[1] The 'T' here represents either an L (Low) or H (High) tone, with the choice determined by the identity of the phrase-initial segment. By default, the tone is Low, but it becomes High when the onset consonant is a fortis or an aspirated stop/affricate, or a fricative (Jun 1996, 1998).

[2] Post-focal dephrasing in Seoul Korean is an optional phenomenon, the incidence of which is assumed to be influenced by several factors, including focus type, speech rate, and phonological length of the post-focal material (Jun et al., 2006; Jun & Lee, 1998).

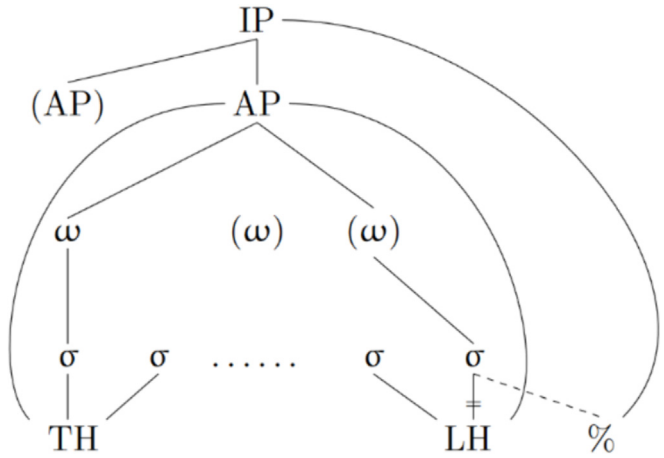

Fig. 1. Intonational structure of Seoul Korean following Jun (2005). 'T' refers to either a default L tone or a H tone, depending on the type of the segment (see footnote 1). % refers to a slot for a boundary tone, which is assumed to override the right-edge H tone of the Accentual Phrase.

empirical evidence suggests that the actual size of the prosodic boundary induced by focus varies. It may range from an Accentual Phrase (AP) boundary (Jun, 1996, 2005) to an Intonational Phrase (IP) boundary (Jeon & Nolan, 2017), beyond the Intermediate Phrase boundary recently proposed by Jun (2007, 2011). Indeed, focus often occurs on the initial word of an utterance, aligning its left edge with that of the Intonational Phrase that commences the utterance. Furthermore, independent phonetic studies that explore the interplay between focus and prosodic boundaries in terms of spatial and temporal segment realizations suggest that focused words can appear both IP-initially, aligned with an IP boundary, and IP-medially, aligned with smaller prosodic boundaries, even those below the level of an AP (e.g., Cho, 2021; Cho et al., 2014; Choi et al., 2020). The literature therefore indicates that focus derived from information structure is likely to trigger a prosodic boundary, the exact size of which may vary depending on other influencing factors. Nevertheless, it is generally assumed that the subsequent post-focus elements tend to undergo dephrasing (Jun, 1996, 2007, 2011), although not obligatorily (Jun & Lee, 1998), leading to post-focus words losing their typical phrasing and often resulting in an elongated low f0 span following the focused word.

In light of this context, it is clear that intonational patterns throughout an utterance in Seoul Korean can be influenced by focus-induced prominence. This is in principle similar to that in Western Germanic languages such as English and German, where the tonal distribution associated with one or more pitch accents is modified by the placement of focus. However, in Seoul Korean, as the term 'edge-prominence' suggests, the modification of intonational patterns delineating the edges of prosodic units theoretically goes hand in hand with phrasing (Jun, 1993, 1996, 2005, 2011). Yet, our understanding of the phonetic nuances that characterize the f0 contour associated with prominence, particularly under focus, remains incomplete. For example, the alignment of typical initial LH patterns with segmental materials and the modification of f0 in terms of scaling and timing have not been previously investigated in detail in relation to focus in Seoul Korean. In fact, H. Cho (2010, 2011) has provided some evidence indicating that the L and H tones of the AP-initial rise are systematically aligned with the rime midpoints of the first and second syllables, respectively. However, these studies examined alignment in contexts devoid of focus, leaving questions unanswered regarding the effects of focus on the phonetic details of f0 realization. It is also worth clarifying that our study does not pertain to general phrasing-related prominence effects which may arise with phrasings independent of focus realizations. Instead, our study concentrates on the influence of focus-induced prominence, stemming from information structure, which is generally assumed to intersect with phrasing.

In summary, the present study aims to investigate how Seoul Korean demonstrates intonational modifications attributed to focus-induced prominence occurring on monosyllable target words in various positions within an Intonational Phrase—i.e., IP-initial, IP-medial, and IP-final. This approach allows us to isolate and highlight the effects related to focus, while also examining how focus-related tonal realization varies based on the position of the focused word within the Intonational Phrase. It is also worth reiterating that 'contrastive' focus is chosen for investigation in the present study due to its well-documented significant impact on phonetic realization, resulting in robust focus-related intonational modifications. More broadly, the present study investigates how the intonational phonology of Seoul Korean, typologically considered an 'edge-prominence' language, is enriched with the fine phonetic nuances of f0 realization in association with focus-induced prominence. This investigation will also facilitate a discussion on how these patterns compared to prominence-related patterns found in languages classified as 'head-prominence' languages. In the following section, we will elaborate on specific research questions that stem from this overarching question.

1.2. Specific research questions

Our initial question addresses a rather basic issue yet to be explored: How does prominence in a focused context affect the direction and extent of pitch range expansion? Intonation is characterized by significant variations in the specific f0 values of tonal targets, such as L and H tones, which often establish the pitch range with the upper (f0 peak) and lower (f0 trough) values of the f0 contour in a given context. Observations of f0 modulation across languages suggest that pitch range variation is often characterized by compression or expansion, largely due to changes in peak values while changes in trough values remains rather minimal (Féry & Kügler, 2008; Liberman & Pierrehumbert, 1984). This asymmetry between H and L tones may be attributed to the proximity of the f0 trough to physiological constraints, while the upper bound for the f0 peak is less constrained by physical limitations (Bishop & Keating, 2012). On the other hand, a different picture emerges when examining a tonal language such as Mandarin Chinese (Lee et al., 2016; Chen & Gussenhoven, 2008) or a lexical pitch accent language such as South Kyungsang Korean (Cho et al., 2019). For these languages where tones participate in lexical contrast, focus (contrastive/narrow) has been shown to induce a distinctive bidirectional polarizing effect on the pitch range, characterized by the elevation of H tones paired with the depression of L tones. These studies, taken together, imply a divergence in pitch expansion in the focused context depending on whether tones are employed at a

phrasal level versus a lexical level. When used at the phrase (post-lexical) level, pitch expansion is constrained by physiological limitations. In this case, it is the f0 peak that primarily contributes to pitch expansion, signaling focus-induced prominence. On the other hand, when used at the lexical level, each tonal target appears to be 'hyperarticulated' in the context of focus-induced prominence to enhance the underlying tonal target. In this context, an L tone is realized with a lowered f0, and an H tone with a raised f0, showing a bidirectional f0 polarization.[3]

In light of this literature, we investigate the directionality of focus-induced pitch expansion in Seoul Korean. Some studies have already reported that there is a greater f0 span for the AP-initial rise in response to narrow focus for Seoul Korean (Chung & Kenstowicz, 1997; Jun, 2007; Jun & Lee, 1998). However, there remains a gap in our understanding the exact phonetic realization of this expansion and its directionality. The specific question here is whether Seoul Korean follows English and other non-tonal languages, where the expansion under focus predominantly affects the upper bound of pitch range, or displays a similar type of bidirectional polarizing expansion that has been described for South Kyungsang Korean. To determine if similar patterns occur in Seoul Korean, readers should closely examine the predicted contours to determine how the f0 level of tonal turning points is realized under focus in comparison to their non-focal alternatives. One might assume that Seoul Korean should follow similar patterns to non-tonal languages since it utilizes tones at the phrasal level rather than the lexical level. However, Seoul Korean presents a different case compared to the aforementioned two types. In non-tonal languages with a phrase-level pitch accent associated with lexical stress, tonal realization is primarily localized to the stressed syllable that receives a pitch accent. Similarly, in languages with lexically-specified tones, tonal realization is also primarily localized to the particular syllable that bears a specific tone. However, in Seoul Korean, edge-tones like LH at the beginning of a phrase (whether it is an Accentual Phrase or an Intonational Phrase) are not associated with specified syllables per se because they are typically realized over the first two or three syllables, depending on the length of the phrase. Despite the fact that the two tones (LH) spread over two or more syllables, due to the use of tones at the phrasal level, the realization of LH tones may still follow the general physiologically constrained pitch expansion found in non-tonal languages—namely, it is the elevation of the pitch peak associated with an H that contributes to the pitch expansion, rather than the lowering of the pitch trough associated with an L (cf. Barnes et al., 2020 and citations therein). Alternatively, it is possible that, because the two tones extend over two or more syllables, each tonal target is realized independently within its own domain of realization. This could result in the L tone being even lower and the H tone being even higher under focus, with each one enhancing the phonetic clarity of the phonologically defined tonal targets of L and H, respectively. In the present study, we aim to test these possibilities regarding the focus-induced expansion of pitch range in Seoul Korean.

Our second question addresses a more challenging issue regarding tonal alignment with the segmental string in different contexts in Seoul Korean. Tonal alignment pertains to the temporal positioning of tonal events to specific segmental landmarks (see Ladd 2008 for a review). Numerous languages have been observed to exhibit a relatively consistent pattern of anchoring their tonal inflection points to acoustic or articulatory landmarks within the segmental string, exemplified by cases in Greek (Arvaniti et al., 1998), English (Ladd et al., 1999), German (Grice et al., 2017; Mücke & Grice, 2014) and Dutch (Caspers & Heuven, 1993; Ladd et al., 2000). These patterns, found across languages, are largely consistent with the view often referred to as the 'segmental anchoring hypothesis.' This view suggests that tones are temporally aligned with the segmental string in ways that characterize the phonetic grammar of a given language, governing the temporal relationship between tonal and segmental realizations in a language-specific way. The patterns of consistent tonal alignment have primarily been explored in the context of pitch accents associated with lexical stress. Notably, however, studies on tonal alignment in French (Welby, 2006; Welby & Loevenbruck, 2005) have called into question the universality of segmental anchoring. These studies have found greater variability in the alignment of high (H) tones in French, suggesting that languages without lexical stress may exhibit less tonal stability. But recall that H. Cho (2011) showed some consistent alignment of the AP-initial LH tones with the rime of the first and the second syllable, respectively in Seoul Korean, indicating that some type of segmental anchoring occurs in Seoul Korean.

Considering the existing literature, therefore, it is essential to understand how tonal realizations in Seoul Korean align with the segmental string, regardless of whether Korean may show some variable tonal alignment patterns just like in French, compared to languages utilizing phrase-level pitch accent. This is because different languages, whether typologically similar or dissimilar, may have language-specific determinants that illuminate the timing relationship between laryngeal and supralaryngeal articulatory events, as discussed in Ladd (2006, 2008), which may reflect the ‘phonetic grammar’ of a given language (cf. Cho & Ladefoged, 1999; see Cho, 2015 for a related discussion). In the present study, we therefore explore the temporal realization of each of the assumed edge tones (L and H) in relation to the segmental string within a focused context. To fully grasp the impact of focus on tonal dynamics, readers should pay particular attention to how the timing of tones shift in focused conditions within the temporal domain. This will allow us to observe a general tendency for the tone–segment alignment conditioned by focus-induced prominence, following the general assumption that the temporal realization of tones relative to the segmental string is systematically controlled in a given language (Ladd, 2006, 2008). Particular attention will be given to how the second tone of LH aligns with the segments under focus. This is because, while the phonological association of the L tone is stipulated

[3] An apparent exception to this pattern from South Kyungsang Korean is found in Lee, Y.-C. et al. (2019). In that study, the authors found that corrective focus on a phrase-initial monosyllabic word bearing the lexical L pitch accent resulted in pitch range expansion on the subsequent post-focal word (cf. J. Kim & Jun, 2009). This suggests speakers of South Kyungsang Korean may have more than one strategy to indicate prominence on small prosodic units. The divergent results observed in the two studies may be attributable to differences in the speech material utilized. Cho et al. (2019) examined disyllabic words with LH or HL pitch accent melodies, whereas Lee et al. (2019) investigated monosyllabic numerals.

to be with the first syllable, the anchoring site for the subsequent H tone is not as clearly defined in the intonational grammar of Seoul Korean; its positioning may depend on other factors, such as phrase length as we briefly discussed above (e.g., Jun, 1996; cf. Shattuck-Hufnagel & Turk, 1996).

While investigating the aforementioned questions related to tonal expansion and alignment in focused contexts, we also examine how their relations may vary depending on the position of monosyllabic focused words within an Intonational Phrase—namely, IP-initial, IP-medial, or IP-final. It is well documented in the intonational literature that realizations of prominence-related tones may interact with boundary tones when, for example, a phrase-level accent such as a pitch accent in English and other Germanic languages, occurs on the final syllable of an Intonational Phrase to which a boundary tone is assigned as well. Such a context creates a tonal crowding of the pitch accent tones and boundary tones, as well as a phrasal tone, resulting in some modifications to accommodate all the tones within the final syllables (see Ladd, 2008 for a review; Arvaniti et al., 1998; Arvaniti & Ladd, 1995; Atterer & Ladd, 2004). One possible resolution is the compression of tones within the limited temporal space, which may be accompanied by a reduction in tonal scaling due to potential truncation effects, although the specifics of this resolution may vary across languages (e.g., Arvaniti et al., 2006). On the other hand, to the best of our knowledge, tonal crowding effects in Seoul Korean have not been extensively discussed in the literature, particularly those arising from interactions between phrasal accents and boundary tones. This oversight may stem from Jun's (1996, 2005) phonological model of Seoul Korean intonation, which suggests that the right-edge tone of an AP is overridden by the boundary tone at the level of an Intonational Phrase, thus avoiding tonal crowding, rather than creating it. This is particularly plausible if a focused monosyllabic word at the end of an Intonational Phrase does not form its own Accentual Phrase due to its short length, although such an AP is not categorically impossible. Our study probes this scenario, particularly when focus is placed on a phrase-final monosyllabic word, where phonologically, it is presumed that the right-edge H tone is replaced by a boundary tone. In particular, we test a possibility that the realization of focus, when imposed on a monosyllabic IP-final word, may still manifest itself with a tonal signature at the microprosodic level of fine phonetic detail, maintaining the level of prominence under focus. Should such fine-grained f0 nuances that signal focus persist in the final syllable, they may interact with the realization of the boundary tone, potentially unveiling a tonal crowding effect within this context.

Yet another question that will be addressed in the present study concerns the influence of segmental characteristics on tonal realization. This examination specifically focuses on the effects of segmental makeup and sonority by comparing voiceless obstruents with voiced sonorants in the coda of the monosyllabic focused words. Voiceless intervals positioned in the coda cause some discontinuity of f0 realization of intonational contours which may obscure pitch trajectory or possibly constrain the physical realization of f0, whereas voiced sonorants do not impose such constraints. Thus, the realization of tonal targets in the presence of voiceless intervals presents an adverse condition for understanding the underlying tonal realization in terms of tonal expansion and alignments with segments. Readers should closely observe how the presence of different segmental characteristics impacts the temporal and tonal adjustments within the focused conditions described. Roettger (2017) discusses possible adjustments that languages might employ to accommodate text-tune alignment under adverse conditions. These adjustments include truncation, where the tonal target is shortened, potentially resulting in less extreme f0 values; temporal compression, which condenses the pitch movement into a shorter time frame; and shifting of tones, where tones are positioned either earlier or later depending on the context.

In contrast, Nooteboom (1997) offers a perceptual perspective, suggesting that listeners can interpolate through short missing sections during voiceless intervals. This phenomenon, often referred to as 'filling in,' implies an unconscious cognitive process wherein the brain reconstructs the missing part of the pitch contour, ensuring the preservation of the intended contour shape. Similarly, in speech production, it is possible that speakers consider the segmental makeup and control the speech production to overcome constraints that may arise within different segmental contexts (cf. House (1990) and Barnes et al. (2014)), allowing them to consistently reach the assumed target. This possibility is theoretically grounded in light of the AM (Autosegmental-Metrical) Theories of Intonation, where tonal targets are defined independently of the segmental string (e.g., Ladd, 2008; Pierrehumbert, 1980). Taking these possibilities into account, we will investigate the effects of coda sonorancy on tonal realizations. Fig. 2 presents schematic representations of the LH rise in contexts with a sonorant (N = nasal) coda versus an obstruent (C = voiceless stop) coda. Fig. 2b and 2c depict two hypothetically schematized patterns, Fig. 2b shows a possible reduction of the f0 rise due to the voiceless gap—i.e., given that the f0 rise is suppressed during the voiceless gap, the expected global f0 rise may also be delayed, aside from some microprosodic f0 perturbation expected due to the voiceless onset. On the other hand, Fig. 2c demonstrates a hypothetically complete realization of the f0 rise as in the sonorant context of Fig. 2a, suggesting that the potential physical suppression on f0 interpolation (or f0 rise) through the voiceless gap is overcome to achieve the target f0 comparable to that in the sonorant context.

Finally, the results of this study will contribute to the important theoretical consideration on the relationship between focus realization and phrasing in Seoul Korean, specifically within the context under investigation. As mentioned earlier, we will examine the effects of focus on tonal realization in various positions within an Intonational Phrase, namely, IP-initial, IP-medial, and IP-final. In the latter two positions, from a theoretically neutral standpoint, the focus may align with either an AP or no phrasing at all. Recall, however, that Jun (2007, 2011) proposed that focus marking involves prosodic restructuring, initially suggesting the placement of the focused element at the beginning of an Accentual Phrase (Jun, 1996), and later refined to be an Intermediate Phrase. However, we leave the possibility open that such focus-prosody mapping is not obligatory for the following reasons. First, prominence especially under focus may also manifest itself through hyperarticulation of segmental realizations (see Cho, 2021 for a review) as well as intonational elements, independent of

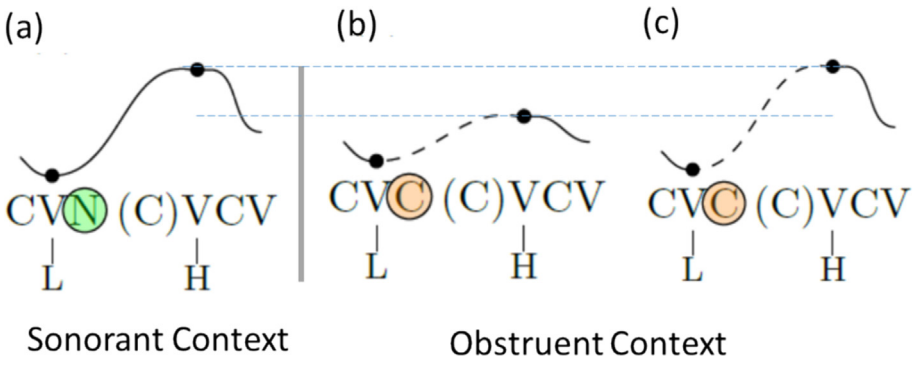

Fig. 2. Possible schematic representations of a focal rise with (a) a voiced sonorant coda and (b–c) a voiceless obstruent coda. Panel (b) illustrates a potential reduction of the f0 rise due to the discontinuity during the voiceless gap, while panel (c) shows a full realization of the f0 rise, despite the absence of physical facilitation for f0 interpolation, potentially akin to that in the sonorant context. Note that in (a), the optional obstruent consonant after 'N' is expected to undergo voicing when it is a lenis stop, indicating an interconsonantal voicing process.

phrasing. Second, prosodic phrasing is influenced by numerous other factors. These may extend beyond information structure to encompass aspects like speech rate, sentence length, and semantic coherence (Jun, 1996; Shattuck-Hufnagel & Turk, 1996). In other words, while Seoul Korean's intonational phonology typically involves phrasing when marking focus, such preference may be overridden by other factors that may, for example, demand grouping of focused words with preceding words, thus the focused word is not positioned at the beginning of a new phrase. Previous phonetic studies have indeed suggested that focused words can occur phrase-medially in certain cases (e.g., Cho, 2021; Cho et al., 2014; Choi et al., 2020; Lee, 2015). Additionally, the temporal effects of focus-induced prosodic strengthening further complicate the interpretation of prosodic boundaries, as the extended duration of phonetic units under focus may occur even in the absence of prosodic boundaries, blurring the distinctions traditionally used to identify prosodic structuring. Thus, we will explore the relationship between focus-induced tonal realization in the context of this study and prosodic phrasing, potentially challenging the focus-prosody mapping rule proposed by Jun (2007, 2011), and propose a revised model that accommodates focus realization independent of phrasing.

2. Methods

2.1. Participants

A total of fourteen native speakers of Seoul Korean (seven female and seven male) participated in an articulatory and acoustic recording session.[4] All participants were born and raised either in Seoul or the surrounding province, Gyeonggi-do. Participants were born between 1991 and 1998 (mean age 23). Participants were unaware of the purpose of the study, reported no hearing or speech problems, and were paid for their participation. The acoustic data were recorded at a sampling rate of 48 kHz, using a Tascam US-4x4 digital recorder, a SHURE KSN44 microphone and ART PRO MPA preamplifier in a sound-attenuated room at the Hanyang Institute for Phonetics and Cognitive Science of Language (HIPCS) in Seoul. This study was approved by the internal review board of HICPS and all data was anonymized before analysis.

2.2. Speech materials and task

Participants read short sentences within a mini discourse context (see below for more details). These sentences contained a target phrase consisting of five syllables, as, for example, underlined in Speaker B's response in (1), which was given in reply to Speaker A's question. (See Table 1 for a complete example set of speech materials).

(1)	**A**: /ipʌn tanʌnɨn ʌnni kuk twieta nwa/?	'Do (I) put the word behind the sister's soup this time?'
	B: /ani. ʌnni pam twie/	'No. Behind the sister's chestnut.'

This particular length was chosen as it was previously reported that sequences of five syllables or shorter are typically grouped within a single AP in Seoul Korean (Jun, 1996, 2003). This structure was thus meant to facilitate possible cases in which a test word, when positioned in the middle of this relatively short frame sentence, receives focus without forming a new AP phrase. A set of test words within the target phrase contained a pair of monosyllabic words /pap/ and /pam/, which systematically varied based on coda sonorancy, featuring either the voiceless stop /p/ or the voiced sonorant nasal /m/. (Note that /pam/ is homophonous, denoting both 'night' and 'chestnut,' and the data includes both cases, resulting in double the number of tokens compared to non-homophonous /pap/, meaning 'cooked rice').

These test words were incorporated into brief sentences within a mini-discourse framework, facilitating the manipulation of prosodic context and focus conditions. Target words were placed in either phrase-initial, -medial, or -final position. The mini-discourse structure encouraged participants to respond to a question under a designated prosodic condition. All sentences contained a single word under contrastive focus, either the target word or some adjacent word. Table 1 presents the set of sentences encompassing the test word /pam/ ('chestnut'). Overall, a total of fourteen experimental conditions were tested in the experiment, with two word-forms (*pam, pap)* and three phrase positions (IP-initial, IP-medial, IP-final) and two or

[4] The data under consideration were initially amassed by the second author in the course of her Master's research. Corresponding recordings pertaining to South Kyungsang Korean (SKK) speakers were examined in the study by Joo (2021).

Table 1
Example sentences with the test word /pam/ as a function of phrase position and prominence (corrective contrastive focus). The target words are underlined and focused words are in bold.

Initial	(a)	Focal	A:	*ipʌn tanʌnin* ***k^{h}o***ŋ *twieta nwa?*	'Should I put this word behind the **bean**?
			B:	*ani. #* ***pam*** *twieta nwa.*	'No. Put it behind the **chestnut**.'
	(b)	Pre-focal	A:	*ipʌn tanʌnin k^{h}o*ŋ ***ap^h****eta nwa?*	'Should I put this word **on top of** the chestnut?'
			B:	*ani. # pam* ***twi****eta nwa.*	'No. Put it **behind** the chestnut.'
Medial	(c)	Focal	A:	*ipʌn tanʌnin ʌnni* ***k^{h}o***ŋ *twieta nwa*	'Should I put this word behind Sister's **bean**?'
			B:	*ani. # ʌnni* ***pam*** *twie.*	'No. Behind Sister's **chestnut**.'
	(d)	Post-focal	A:	*ipʌn tanʌnin* ***op*a*** *pam twieta nwa?*	'Should I put this word behind **Brother's** chestnut?'
			B:	*ani. #* ***ʌnni*** *pam twie.*	'No. Behind **Sister's** chestnut.'
	(e)	Pre-focal	A:	*ipʌn tanʌnin ʌnni pam* ***ap^h****eta nwa?*	'Should I put this word **on top** of Sister's chestnut?
			B:	*ani. # ʌnni pam* ***twi****e.*	'No. **Behind** Sister's chestnut.'
Final	(f)	Focal	A:	*ipʌn tanʌnin nʌhii ʌnni* ***k^{h}o***ŋ*ini?*	'Is this your sister's **bean**?'
			B:	*ani. # uri ʌnni* ***pam****. # twieta nwa.*	'No. My sister's **chestnut**. Put it behind it.'
	(g)	Post-focal	A:	*ipʌn tanʌnin nʌhii* ***op*a*** *pamini?*	'Is this your brother's chestnut?'
			B:	*ani. # uri* ***ʌnni*** *pam. # twieta nwa.*	'No. My **sister's** chestnut. Put it behind it.'

three focus conditions (focused, pre-focal, post-focal). The phrase-medial condition differed from the other two phrase positions in having three possible focus contexts: focal, pre-focal, and post-focal. The other phrase conditions' position at the edge mandated a single unfocused context. This led to a total of seven possible phrase-focus conditions. It is also important to note that we examined not only the tonal realizations of these target words but also those of adjacent words (e.g., /ʌnni/ ('sister') or /twieta/ ('behind') preceding and following the target word) to compare them within the overall tonal context of the entire five-syllable phrase.

To obtain as natural a rendition of the intended response as possible, a set of card game scenarios was created. Participants were instructed to answer questions that indicated where to place the next card on the board. These scenarios, along with corresponding text and visual cues (as illustrated in Fig. 3), were presented visually on a computer screen. Each scenario included a question designed to elicit corrective contrastive focus, displayed on the left half of the screen, with the incorrect information marked by a red cross. The answer to the question was displayed on the right-hand side of the screen within a circumscribing circle around the corrected constituent. For instance, in Fig. 3a, the first scenario (matching the text in Table 1c) asks the participant the Korean equivalent of 'Should I put this word behind Sister's bean?' Here, the bean card is indicated to be incorrect via the red cross, and the participant is instead required to respond by saying the Korean equivalent of 'No. Behind Sister's chestnut.' By answering in this manner, the participant was supposed to produce a rendition in which the target word occupied the phrase-medial position in Korean and was under corrective focus (i.e., the IP-medial/focal condition). In another scenario provided in Fig. 3b (corresponding to the text in Table 1g), the participant is asked, 'Is this your brother's chestnut?' with a red cross indicating that the interlocutor's brother was not the intended recipient. The participant responds with, 'No, my sister's chestnut. Put it behind it.' In this case, the target word appears phrase-finally and is unfocused (i.e., IP-final/post-focal). The position of the target word and focus condition were manipulated in this way to elicit speech patterns in prosodic contexts.

Before the recording session, participants had a practice period during which they familiarized themselves with the experimental setting, scenarios, and speech materials to be elicited. During the recording session, the participant was presented with both the visual material on the computer screen as well as a prepared recording of the question to be answered. The prompt questions were pre-recorded by a native Seoul Korean speaker and prominence on the word to be corrected was indicated in red on the slide. Participants were asked to answer the question as naturally and casually as possible without pausing between the words. Each scenario was repeated twenty times in a randomized order resulting in a total of 5,880 tokens (14 speakers × 3 target words × (3 phrasal positions × 2 focus conditions + 1 additional unfocused context) × 20 repetitions).[5] After understanding the task during the practice period, speakers typically produced the test sentences without difficulty, aligning with the expected experimental conditions. Nevertheless, the prosodic renditions of the utterances were checked by three of the four authors, who were trained Korean ToBI transcribers following the conventions of Jun (2005) to verify that the utterances were natural and that the intended word was perceived as receiving prominence, as well as the size of the prosodic boundary between words within the five-syllable test phrase, including the boundary before and after the target words. Trials containing disfluencies like unnaturally added pauses or repetitions were omitted since they do not yield the f0 trajectories indicative of fluent utterance production. Moreover, some tokens were also removed because their focus marking wasn't appropriate, as determined by the three Korean ToBI transcribers, to the intended discourse context. Based on these standards, a total of 690 tokens were considered unsuitable, leaving 5,190 tokens for analysis (88.3%).

2.3. Data processing and measurements

The Montreal Forced Aligner (McAuliffe et al., 2017) was used to achieve word and phone-level segmentation via a pretrained Korean model (McAuliffe & Sonderegger, 2022). Subsequent to this automated segmentation, a manual inspection and necessary hand corrections were undertaken

[5] As a reviewer rightly pointed out, multiple repetitions with limited scenarios in the current study could lead to some reduction in stress or emphasis on the experimental speech materials. We agree that laboratory-controlled speech production experiments, like the current study, have various limitations, including the one mentioned by the reviewer. Acknowledging this potential limitation, we used visual aids to help the speaker focus on the context, which we believe would mitigate experimental artifacts from multiple repetitions. While multiple repetitions may have influenced the degree of pitch range scaling, as evidenced by the empirical contours detailed in Appendix B, there was negligible variation in the global contour shape across different prosodic contexts or focal targets for any given speaker. Consequently, any potential destressing due to multiple repetitions did not seem to alter the phonological structures, which is critical to the purpose of the current study.

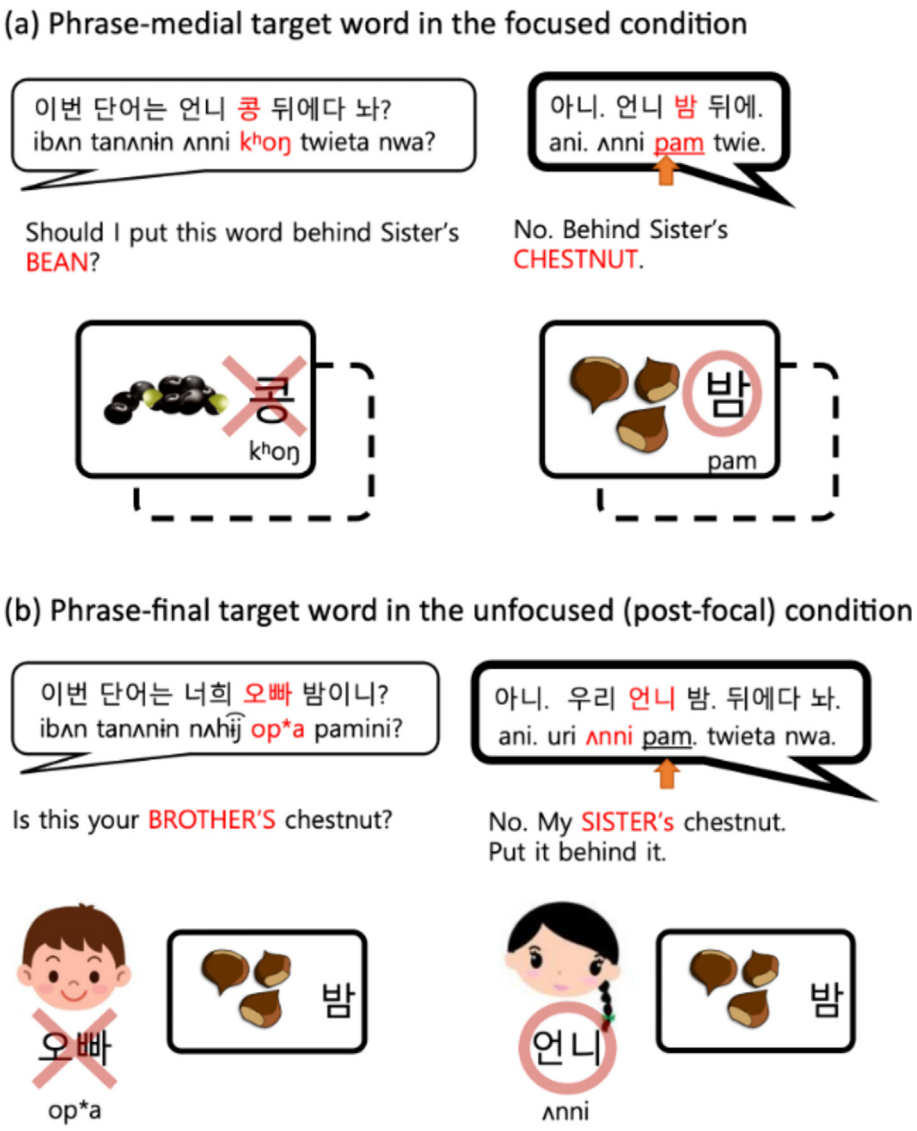

Fig. 3. Example illustrations of two simulations in a card game devised for the experiment eliciting a dialogue context for target word /pam/ ('chestnut') in which it is a) phrase-medial and focused and b) phrase-final and unfocused.

to ensure accuracy. As word-initial stops in Korean often contain a voiceless interval during closure which does not bear f0, these portions of the sentence were excluded from the Generalized Additive Modeling described below in Section 2.4. To foster a detailed and comparative analysis across phrasal positions and focus contexts, two additional tiers were created in Praat (Boersma & Weenick, 2021). The first tier strategically excludes all word-initial stops, providing a unit enabling fair comparison of word rimes across differing focus contexts. In contrast, the second tier, purposefully omitting only word-initial stops in sentence-initial positions or under focus where f0 is expecting to be absent, facilitates the generation of models encapsulating the maximal f0 contour.

F0 trajectories were computed across the sonorant portion of words via VoiceSauce (Shue et al., 2011), using the STRAIGHT algorithm (Kawahara et al., 2005) (40–500 Hz; 1 ms time step) and manually inspected to avoid the measurement error of pitch doubling or halving. The default 1000 Hz sampling rate used for fundamental frequency (f0) measurements in VoiceSauce was deemed excessively high. To address this, the recorded data was downsampled to reduce the sampling rate to 100 Hz. The downsampled data was generated by applying a filter eliminating all but those data points occurring at 10 ms intervals. Prior to statistical analysis, both time and f0 values were normalized. F0 was z-score standardized by speaker to account for cross-speaker differences in pitch level and range. The time dimension was linearly normalized for each word token (scale: 0–1).

Duration measures were taken for both sonorant and non-sonorant intervals. The average duration measures were later used to scale the output of the statistical model to reveal any temporal effects of prominence on words in sentences and to reconstruct the global f0 contour of the phrase.

2.4. Data analysis

To capture cross-speaker generalities in the global intonational contour, a generalized additive mixed model (henceforth GAMM) was constructed for each of the phrase positions in R (R Core Team, 2022) using the bam function[6] of the *mgcv* package (Wood, 2019). To account for any autocorrelation in the model residuals for time, the bam function was supplied with a rho (ρ) value corresponding to the autocorrelation function value at lag = 1 as recommended by Wieling (2018). To account for the interaction of focus on individual words in the phrase, an interaction factor was created combining the factors Word and Focus and Item, i.e., whether the target wordform was *pam* or *pap*. The model incorporated a fixed effect for condition (the interaction factor), along with cubic regression spline smooths for time and an interaction between time and condition. It also included factor smooths for the interaction between time and participant. The R code for the GAMM for phrase-initial targets is shown in (1). The output of the GAMMs were then visually compared to the f0 tracks of actual renditions obtained with *mausmooth* (Cangemi, 2015) to validate the GAMM models.

(1) Structure of generalized additive mixed model used for phrase-initial targets

```
# creation of interaction term
ini <- mutate(labFocItem = interaction(Word, Focus, Item))
bam(scaleF0 ~ labFocItem
    + s(Time, by = labFocItem, bs="cr", k = 5)
    + s(Time, Speaker, by = labFocItem,
      bs="fs", m = 1),
  data = ini, method="ML",
  rho = r1, AR.start = ini$start.event)
```

The data were subset to include renditions to be analyzed separately for effects of focus and coda condition at each of three positions:

- ***Phrase-initial target words*:** The phrases in this condition are of the form: **/ani._____twieta nwa/** ('No. Put it behind the TARGET'). This subset yielded a total of 1,485 observations (995 /pam/; 490 /pap/) for a total of 92,443 data points (66,201 for sentences with /pam/, 26,242 for /pap/).[7] The focused condition refers to the phrase-initial words (/pam/, /pap/) being focused, and the unfocused condition to the following word (/twieta/ ('behind it') being focused, so that the target word becomes pre-focal.
- ***Phrase-medial target words*:** The phrases in this condition are of the form: **/ani. ʌnni_____twie/** ('No. Behind Sister's TARGET'). In this context, target words were preceded by /ʌnni/ ('older sister') and followed by /twie/ ('behind'). This means that the target words can be 'pre-focal' when the following word /twie/ is focused, or 'post-focal' when the preceding word /ʌnni/ is focal, resulting in three focus-related levels: focused, unfocused (pre-focal), unfocused (post-focal). This subset yielded a total of 2,505 observations (1,598 for sentences with /pam/; 907 for sentences with /pap/) for a total of 157,874 data points (111,737 for sentences with /pam/, 46,137 for /pap/).
- ***Phrase-final target words*:** The phrases in this condition are of the form: **/ani. uri ʌnni _____/** ('No. Our sister's TARGET'). This subset yielded a total of 1,581 observations (1,054 for sentences with /pam/; 527 for sentences with /pap/) for a total of 72,088 data points (48,182 for sentences with /pam/, 23,906 for /pap/).

Alongside the global contour comparisons of GAMM models, linear mixed-effects models (lmer) were constructed in R with the *lme4* package (Bates et al., 2021) to examine differences in the level and timing of f0 extrema in relation to Focus and/or Phrase conditions. In the specified models, fixed effects were contingent upon the type of comparison being performed and included the factors Target Item (/pam/ or /pap/) and either the Focus (focused, pre-focal, post-focal) or the Phrase Position (IP-initial, IP-medial, IP-final). The model's random effects structure comprised random intercepts for each speaker, along with slopes for the speaker's Focus or Phrase Position, adhering to the principles of maximal model specification recommended by Barr et al. (2013). However, when faced with issues of nonconvergence or a singular fit error, the model was simplified to include only the random intercepts and the relevant slopes for Focus or Phrase Position. The assessment of all fixed effects was conducted using the *lmerTest* package (Kuznetsova et al., 2017). The specifications for all linear models and their corresponding summary tables are provided in Appendix A.

3. Results

3.1. Tonal realizations for utterances with phrase-initial targets

To understand the results based on the GAMM, we need to clarify how the GAMM smooths are presented in this study. We will therefore begin by explaining the figures used to report the results. Fig. 4a illustrates the aggregate f0 contours of the entire five syllable phrase, emphasizing the monosyllabic target words that appear in the phrase-initial position. These contours are plotted separately for targets in the focal context (shown in the upper panel) and the pre-focal context (shown in the lower panel), which allow for the qualitative observation of the overall distribution of tones and their alignment patterns with the segmental string around the target words. These GAMM smooths are displayed on an estimated real-time scale, which is determined by the average duration of each word across their respective focus contexts. Notice that the f0 discontinuity in the smooths corresponds to a voiceless obstruent portion, such as found in the voiceless coda context of /pap/, followed by the voiceless lenis stop as the onset of the following word /twieta/, where f0 cannot be measured. Fig. 4b and 4c, on the other hand, show GAMM smooths accompanied by difference plots. These plots evaluate any significantly different portions in the GAMM smooths stemming from the focal versus pre-focal contexts of the target words. The analysis compares each test word pair for focus-related differences, focusing only on the sonorant portions where f0 is measurable. Therefore, the data presented in Fig. 4b and c exclusively pertains to voiced sonorant speech segments, with the voiceless gap omitted. To facilitate these comparisons, time is linearly

[6] Within the *mgcv* package in R, BAM (Big Additive Models) and GAM functions apply splines for modeling nonlinear predictor-response relationships in generalized additive models (GAMs). BAM is tailored for large datasets, enhancing computational efficiency and memory management, making it preferable for handling extensive data compared to GAM.

[7] The dataset contains two homophones for /pam/, 'night' and 'chestnut,' and only one for pap 'cooked rice,' resulting in twice as many tokens for the former.

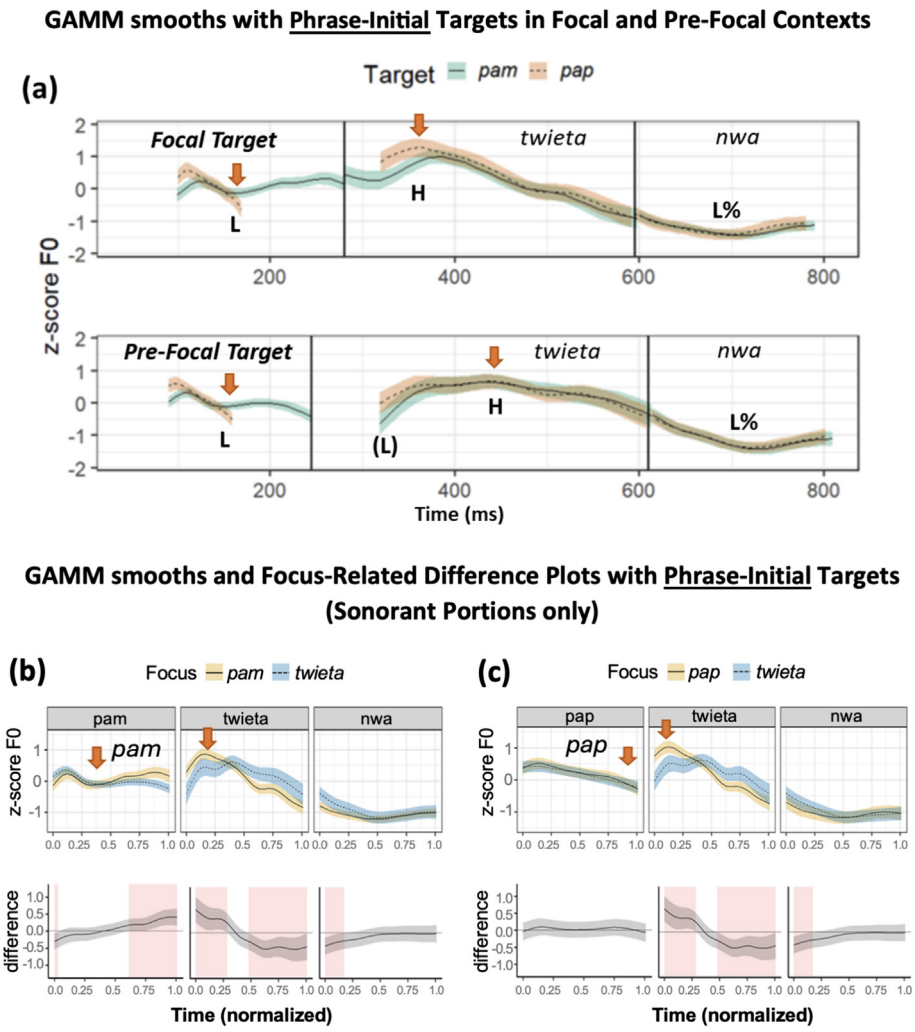

Fig. 4. GAMM smooths of target words in phrase-initial positions. (a) Aggregate f0 contours for targets in focal versus pre-focal contexts, represented on the estimated real-time scale; (b) and (c) GAMM smooths and focus-related difference plots in /pam/ and /pap/ contexts, respectively, based solely on sonorant portions with time normalized on a scale ranging from 0 to 1 for each compared word rime.

normalized based on the sonorant portions, creating a scale from 0 to 1 for each word comparison. Finally, although individual speaker variation falls outside the scope of the present study, for the sake of completeness, we have included f0 plots for each speaker in the Appendix B, which reflect all repeated individual utterance tokens in focal versus non-focal contexts. As we report the results, it is important to keep in mind the foundational research questions that guide this study: (1) What role does focus play in determining the pitch range and the alignment of intonational tones in Seoul Korean? (2) How does the position of the focused word within a sentence interact with focus to influence these tonal realizations? (3) What impact does the consonant of the target word have on the realization of these tones? (4) Do the observed contours suggest that focus is predominantly realized through prosodic phrasing?

Let us now examine the f0 contours especially in relation to the expected phrase-initial LH tones across focal and coda contexts. As Fig. 4a illustrates, an f0 trough (f0 minimum) appears for the initial L tone during the monosyllabic target word, occurring well before the midpoint of the rime, at approximately one-third of its way. This is explicitly evident in the sonorant context of /pam/ (marked by a solid line and green ribbon), while it is implicitly estimated for /pap/, considering the period of the voiceless coda (marked by a dotted line and orange ribbon). As for the focus effect, notice that the anchoring point of the L tone remains largely consistent in both the focal and pre-focal contexts, as shown in the upper and lower panels of Fig. 4a, respectively. What diverges from the initial f0 pattern due to focus, however, is the location of the f0 peak for the subsequent H tone. When the phrase-initial target word is focused (shown in the upper panel of Fig. 4a), the f0 peak occurs near the onset of the next word /twieta/,[8] whereas in the pre-focal context (that is, when the next word /twieta/ is focused), it is further delayed towards the second syllable of /twieta/. Specifically, when /pam/ is in focus (indicated by

[8] A reviewer noted the similarity between this alignment pattern and 'peak delay' of pitch accents observed in languages such as English and Japanese. It is conceivable that the LH rise associated with narrow corrective focus in these languages might be phonologically linked to a single syllable, with the H tone phonetically aligning with the onset of the following vowel. However, the intonational phonology of Seoul Korean assigns an L tone to the first syllable and an H tone to the second syllable when an AP consists of more than one syllable, as is the case with the current speech materials. The early peak of the H on the post-focal word could stem from the phonetic realization of focus, where the earlier alignment of the H tone leads to a more rapid—and thus more prominent—rise in F0 (cf. Xu & Xu; 2005).

a green ribbon in the upper panel of Fig. 4a), f0 rises continuously up to the beginning of /twieta/, forming the f0 peak for the H tone (marked by an arrow). Conversely, when /pam/ is pre-focal (indicated by a green ribbon in the lower panel of Fig. 4a), the f0 minimum behaves more like an elbow rather than a f0 trough. The level of this f0 minimum remains relatively consistent for a period, exhibiting a slight rise, and begins to decrease again well before the acoustic offset of the target word towards the next word. This leads to a realization of a deeper f0 trough for the L tone at the onset of the next word that is focused. It appears that the focused word draws the L tone towards its initial syllable.

As for the coda effect, there is a noticeable difference between the sonorant and voiceless obstruent conditions. Considering when the initial target word with the sonorant coda (/pam/) is in the focal context (indicated by a green ribbon in the upper panel of Fig. 4a), the f0 continues to rise after its trough during the remaining sonorant portion of the rime. This demonstrates an f0 interpolation through the coda /m/ between an L and an H target, the latter of which occurs at the beginning of the next word. Conversely, in the voiceless obstruent context of /pap/ (indicated by an orange ribbon in the upper panel of Fig. 4a), f0 tends to fall by the end of the sonorant portion (the vowel) of the first syllable. There follows a distinct discontinuity in f0, after which f0 sharply rises into the following f0 peak for the H tone (marked by an arrow) around the onset of the subsequent word /twieta/. It appears that, regardless of whether explicit f0 interpolation occurs with the sonorant coda or not with the obstruent coda, the realizations of the H tone remain largely consistent in terms of both the alignment with the onset of the subsequent syllable and the f0 height. In other words, the intended targets for f0 height and timing appear to be achieved even when there is an f0 discontinuity caused by the obstruent coda. This suggests a form of virtual f0 interpolation across the voiceless gap to reach the f0 target for the H tone, similar to the physical facilitation provided by the sonorant coda. Notice that this consistency across the coda conditions is also evident when the focus shifts to the subsequent word, as shown in the lower panel of Fig. 4a.

However, there remains a small but significant difference at the microprosodic level due to the coda's segmental makeup. The f0 height is significantly higher (β = 0.380, SE = 0.046, p < 0.001) and aligns earlier during the vowel (β = 25.7, SE = 5.39, p < 0.001) after /pap/ than after /pam/, which can also be inferred from Fig. 4a. This difference appears to result from the segmental perturbation effect on f0 in the /pap/ context. Specifically, in the context of /pap/, when the obstruent /p/ combines with the following /t/ in /twieta/, it creates a voiceless fortis stop for /t/ (due to post-obstruent tensification in Korean within an Accentual Phrase). In contrast, in the context of /pam/, the following /t/ in /twieta/ becomes voiced. Given that the voiceless stop has a known effect of elevating f0 at the beginning of the vowel, the segmental differences appear to account for the microprosodic change with an earlier and elevated f0 peak after /pap/.

Another observation from Fig. 4a highlights that there is also a focus-induced variation in the scaling (height) of the two peaks of the H tone. That is, the f0 range of the LH, estimated based on the sonorant context, is significantly larger when the initial target is focused than when the next word is (β = 0.505, SE = 0.068, p < 0.001). Specifically, the upper bound of the f0 range (i.e., f0 peak for the H tone) is elevated when the focus is on the target word (as seen in the upper panel of Fig. 4a) compared to when the focus shifts to the subsequent word (illustrated in the lower panel of Fig. 4a) (β = 0.263, SE = 0.029, p < 0.001). On the other hand, as can be seen in Fig. 4b and 4c, the level of the f0 trough for L does not differ at all when the target is focal versus pre-focal. This suggests that the observed expansion of scaling under focus is due to an increased f0 for H, but not to a simultaneous decrease in f0 for L.

Fig. 4b and c further confirm these focus-induced variations in the f0 contours, accompanied by statistical difference tests. Both figures confirm that the initial f0 trough for the L tone anchors largely at a consistent time point in relative terms on a normalized time scale—i.e., roughly at one-third of the way through the sonorant rime in the context of /pam/ and at the acoustic offset of the vowel in the context of /pap/, as marked by an arrow in the leftmost panel of Fig. 4b and c). Furthermore, these figures confirm that differences due to the location of focus stem primarily from the later portion of the rime for /pam/ and from the beginning portion of the subsequent word /twieta/, as highlighted by the vertical pink blocks. A similar f0 contour is generally observed for the phrase with /pap/. As can be seen in Fig. 4c, the significant difference due to the location of focus is reflected during the beginning portion of the second word /twieta/. The difference is due to the f0 peak for the H tone occurring earlier at the beginning of the subsequent word /twieta/ in the focal condition (as marked by an arrow in the middle panel of Fig. 4c) than in the pre-focal condition.

3.2. Tonal realizations for utterances with phrase-medial targets

Recall that in the phrase-medial context, the target words occur in three focus levels: focused, pre-focal (when the following word /twie/ is focused), and post-focal (when the preceding word /ʌnni/ is focused). Fig. 5a illustrates the aggregate f0 contours of the entire phrase, emphasizing the monosyllabic target words that appear in the phrase-medial position. These contours are plotted separately for targets in the focal context (shown in the top panel), the pre-focal context (shown in the middle panel) and the post-focal context (shown in the bottom panel). As in the preceding section, this separation allows for the qualitative observation of the overall distribution of tones and their alignment patterns with the segmental string around the target words. For the purpose of GAMM analyses, we compare focal and pre-focal target words and focal with post-focal target words, separately. GAMM smooth and the focus-related difference plots for the f0 contour of phrase-medial target words are presented in Fig. 5b–e. Fig. 5b and c compare the focal and pre-focal contexts, and Fig. 5d and 5 compare the focal and post-focal context with the target words positioned in the middle of the phrase.

Let us first examine the distribution of the phrase-initial LH tones associated with /ʌnni/ when the target words occur phrase-medially. As seen in Fig. 5a, all three focus-related contexts start with an L tone followed by an H tone. Both tones occur within the disyllabic first word /ʌnni/ (‘older sister’), with the L tone in the first syllable and the H tone in the second syl-

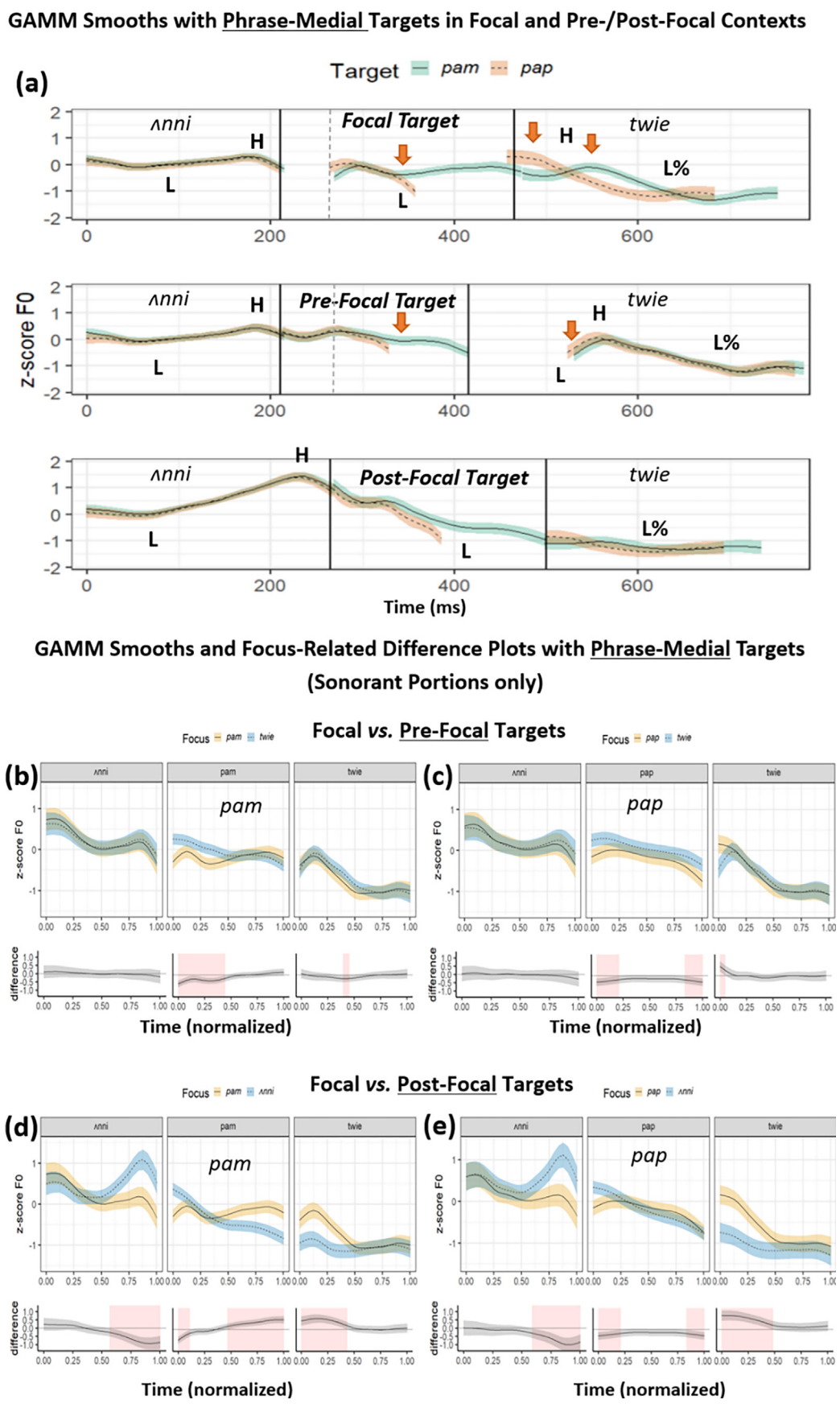

Fig. 5. GAMM smooths of target words in phrase-medial positions. (a) Aggregate f0 contours for targets in focal, pre-focal, and post-focal contexts, represented on the estimated real-time scale; (b, c) GAMM smooths and focus-related difference plots between focal and pre-focal contexts; and (d, e) between focal and post-focal contexts, based solely on sonorant portions with time normalized on a scale ranging from 0 to 1 for each compared word.

lable. This distribution is similar to the LH tone distribution when the monosyllabic target words occur phrase-initially, as reported in the previous section, where the H tone appears in the subsequent syllable belonging to the next word. It is also worth noting that when the initial contextual word /ʌnni/ is focused, it is produced with an increased f0 height associated with the H tone on the second syllable compared to when it is not focused. On the other hand, there is no further lowering of f0 for the L tone. This modification of f0 scaling in terms of f0 height can be observed in the leftmost panels of Fig. 5d and e. In these panels, the GAMMs reveal no differences around the f0 trough, but they do show clear differences around the f0 peak, which are highlighted by the vertical pink blocks.

Now, moving to the tonal realization of the monosyllable target words, when they are in focus, the target words are produced with a clear f0 trough during the rime. This is evident in the sonorant context of /pam/, as marked by an arrow in the upper center panel of Fig. 5a, where the f0 trough occurs roughly one-third of the way through the sonorant rime. In contrast, in the obstruent context of /pap/, the f0 minimum is aligned with the end of the sonorant portion (the vowel). In both contexts, an f0 peak for the H tone follows in the subsequent syllable, which corresponds to the first syllable of the next (phrase-final) word /twie/. However, the exact timing of this peak varies depending on the preceding coda, as indicated by arrows in the top rightmost panel of Fig. 5a (further discussion below). Overall the alignment of LH tones with the segmental string around the phrase-medial target words under focus is similar to that found in phrase-initial target words in the same focal context.

When the focus shifts to the next word (making the target words pre-focal), as shown in the middle panel of Fig. 5a, the LH tonal alignment also appears to shift rightward, particularly the anchoring point of the L tone. In this pre-focal context, similar to the pre-focal context with the initial target words, a local f0 minimum arises around the midpoint of the rime, marked by an arrow. However, f0 continues to decrease toward the syllable's end, resulting in a noticeable f0 trough for the L tone at the beginning of the subsequent focused word. Here again, the focused word appears to draw the L tone towards its initial syllable. Conversely, when the focus shifts to the preceding word (this time, making the target words post-focal), a different pattern emerges. As shown in the bottom panel of Fig. 5a, the f0 steadily decreases from the f0 peak established during the initial word /ʌnni/, extending through the target words. This ultimately leads to an f0 minimum for the L% tone associated with the phrase-final word /twie/. Once again, this is consistent with post-focal dephrasing, as there are no apparent realizations of post-focal LH within the interpolation from the H to L% tones, which would otherwise indicate the presence of a possible new phrase.

In terms of coda effects, we also observe similar patterns to those in the phrase-initial cases. As can be seen in the top panel of Fig. 5a, the tonal target for the H tone appears to be fully realized in both the sonorant and obstruent contexts of /pam/ and /pap/ under focus, even though the latter lacks the physical facilitation of f0 provided by the sonorant coda. This once again indicates a form of virtual f0 interpolation, where the assumed target for the H tone is reached. There is also a difference, however, at the microprosodic level in the timing (β = −31.858, SE = 2.652, p < 0.001) and height of the f0 peak (β = 0.37, SE = 0.086, p < 0.001) after /pap/ compared to /pam/. This shows the segmental perturbation effect, such that the f0 peak is higher and comes earlier near the onset of the vowel of the next word /twie/ after /pap/ than /pam/, as marked by arrows in the top rightmost panels of Fig. 5a. (Regarding the focus effects on scaling, making direct comparisons around the target words between focal and non-focal contexts seems unfeasible because tonal distributions in the phrase-medial contexts are not fixed like the phrase-initial cases. Recall that in the initial context, the target words display an initial rise of LH tones in both focal and pre-focal contexts, enabling direct comparisons in f0 scaling).

Fig. 5b–e further illustrate the results of GAMMs on focus-induced differences in the f0 contours, which have been normalized based on the sonorant portions only. In general, significant differences due to focus during the target words stem from the portions that are related to forming an f0 trough for the L tone. Let us first consider differences between focal versus *pre-focal* target words. As can be seen in Fig. 5b and c, significantly different portions (marked by the vertical pink blocks) due to focus center around the f0 trough in both /pam/ and /pap/ when in focus, while such an f0 event associated with the L tone is not observed with the target words in the pre-focal context. (Note that in both /pam/ and /pap/ under focus, the syllable starts with a relatively lower f0, forming a small initial rise at the very beginning before transitioning into an f0 trough. We do not have an explanation for this phenomenon in the context of intonational phonology, except that it appears to arise with focus at the microprosodic level.) Now, let us consider differences between focal versus *post-focal* target words (Fig. 5d and e). Significant differences in the f0 contours during the target word emerge due to the target words being focal and their role in post-focal dephrasing. In the sonorant context of /pam/ (Fig. 5d), significantly different portions of the f0 smooth in /pam/ are associated with the initial small rise under focus and an additional rise after the f0 peak, resulting in the formation of an H tone in the next syllable. But in the post-focal context, no such f0 events are observed due to dephrasing effects as discussed above. In the obstruent context of /pap/ (Fig. 5e), significant differences in the f0 contour are also attributed to the initial small rise and the end of the vowel, which result in the formation of an f0 trough for the L tone under focus. On the other hand, when under no focus, the post-focal target word shows a further decrease in f0 as part of the dephrasing process, contributing to the significant differences.

3.3. Tonal realizations for utterances with phrase-final targets

Fig. 6a illustrates the overall f0 contours with target words in the phrase-final position when the targets are in focus (focal) versus when they are not (post-focal). Fig. 6b and c depict the GAMM smooth and focus-induced difference plots on a normalized time scale based on the sonorant portions in the sonorant context of /pam/ and in the obstruent context of /pap/, respectively. See the introduction of Section 3.1 for further details of these visual representations.

Let us begin by examining the overall tonal distribution when the target words are in focus, shown in the upper panel

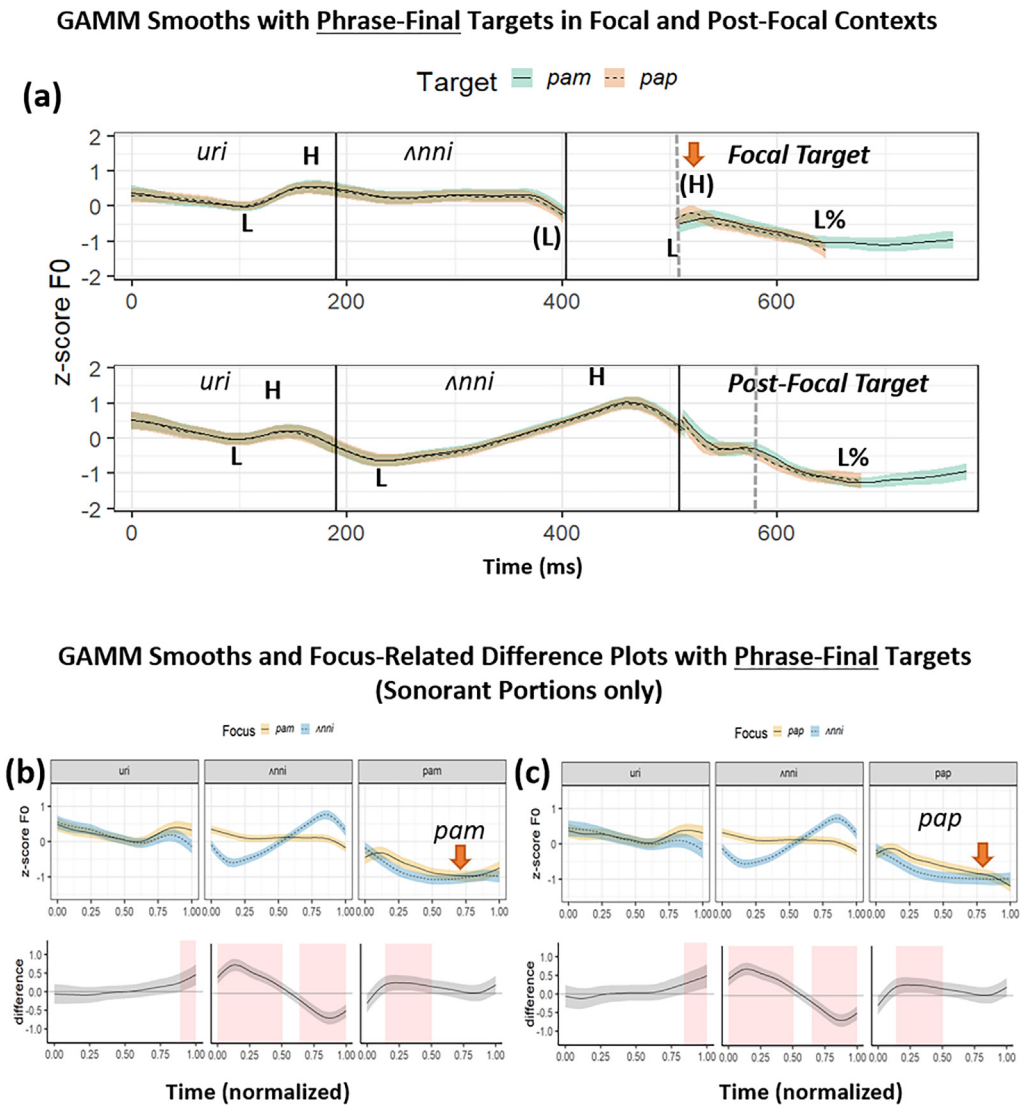

Fig. 6. GAMM smooths of target words in phrase-final positions. (a) Aggregate f0 contours for targets in focal versus post-focal contexts, represented on the estimated real-time scale; (b) and (c) GAMM smooths and focus-related difference plots, in /pam/ and /pap/ contexts, respectively, based solely on sonorant portions with time normalized on a scale ranging from 0 to 1 for each compared word.

of Fig. 6a. As previously observed, the initial LH tones are realized during the first two syllables of the initial contextual word /uri/ ('my'). However, the tonal distribution on the second contextual word /ʌnni/ does not appear to be clearly determined. Nevertheless, there is a clear trend of f0 lowering towards the end of the word /ʌnni/. This f0 lowering seems to continue, possibly resulting in an f0 trough at the beginning of the subsequent target words that are focused. This phenomenon resembles what we observed with the phrase-medial target words in the pre-focal context, as discussed in the preceding section, indicating a possible attraction of the L tone towards the onset of the following focused word.

When the monosyllable target word is focused and positioned at the end of a phrase, it presents a unique context that encompasses both focus realization and phrase-finality with a boundary tone on a single syllable. This context potentially causes tonal crowding effects if both the focus and the boundary each operate with their own tone. As can be estimated from the upper panel of Fig. 6a, the focused target words indeed do not exhibit the typical f0 rise of LH tones that are observed in the non-final focused contexts, previously discussed in Sections 3.1 and 3.2. Upon closer examination of the f0 contours, however, there is a small f0 rise that is evident at the onset of the final target word, as indicated by an arrow in the upper right panel of Fig. 6a. While such a local f0 rise may be attributed to a segmental perturbation effect, the local lowering of f0 for L, continuing from the preceding pre-focal syllable, could potentially indicate the phonetic residue of the H tone, albeit significantly reduced or nearly indiscernible in magnitude. This is because segmental f0 perturbation typically manifests with the f0 peak often aligned with the very beginning of the following vowel, showing a typical fall into the mid portion of the vowel. However, in the current case, there was a consistent rising pattern before falling towards the L%. This pattern becomes evident when examining individual tokens, as exemplified in Appendix C. Moreover, all authors, two of whom are experienced Korean ToBI experts, independently identified the rising f0 pattern perceptually, supporting this argument.

Thus, it is conceivable that the reduction of the f0 rise in magnitude is attributable to a possible downward pull from the subsequent L% boundary tone. If this were the case, the reduction of f0 rise in magnitude could be attributable to a possible downward pull from the subsequent L% boundary tone. In other words, an upward rising tonal movement towards the f0 target for the assumed H tone appears to be truncated by the initiation of the downward tonal movement towards the f0 target for the L% boundary tone, due to the time pressure imposed on the final syllable during which both tonal targets must be realized. In a similar vein, the extremely early rise of f0 can be viewed as a result of tonal crowding between the focus-related LH and the boundary tone L%–i.e., the earlier realization of the assumed focus-related LH tones provides a pathway for the realization of the L% boundary tone.

Further supportive evidence for the tonal crowding interpretation can be obtained when comparing GAMM smooths and focus-induced difference plots for the target words between the focused and the unfocused (post-focal) contexts, as shown in the rightmost panels of Fig. 6b (/pam/) and 6c (/pap/). Crucially, with these GAMM smooths on a normalized time scale, the f0 minimum associated with the L% comes later when the final target word is focused than when it is not focused, as marked by an arrow in the figure, in both the sonorant and obstruent contexts of /pam/ and /pap/. The GAMMs indeed reveal significant differences, as highlighted by the vertical pink blocks in the figures, corresponding to these observations. The focus-induced differences are evident from the vowel onset, continuing for approximately half of the sonorant portion of the target word's rime. This pattern arises because the descent of f0 towards the L% target is delayed to the right under focus across both sonorant and obstruent coda contexts (Fig. 6b and c). Furthermore, the level of the f0 minimum for the L% tone is elevated in the focused context, compared to that in the post-focal context. The elevation of the L% tone is possibly due to the tonal coarticulatory impact of the assumed H tone on the L%, particularly evident in the sonorant context of /pam/ (β = 0.072, SE = 0.015, p < 0.001). Overall, these observations suggest that the tonal crowding effect is responsible for the changes in pitch: The initial rise in f0 for the assumed LH tones due to focus pushes the descending f0 for the final L% to the right, causing it to shift rightward and resulting in a higher f0 minimum. Conversely, in the absence of the focus-induced LH tone within the post-focal context of the target word, the L% boundary tone occurs earlier and reaches a lower f0 minimum, indicating the absence of a tonal crowding effect.

Finally, before moving on to the next section, it is worth noting another type of f0 scaling modification due to focus during the contextual word /ʌnni/. As can be seen from the lower center panel of Fig. 6a, there is an expansion of the f0 range for the LH tones. Similar to the case when /ʌnni/ is focused phrase-initially, in this phrase-medial context, the f0 peak for the H tone in /ʌnni/ is clearly heightened due to focus. Crucially, however, unlike the phrase-initial case, /ʌnni/ in this medial context shows a more pronounced lowering of f0 for the L tone under focus. Note, however, that direct comparisons of the levels of L and H tones between the focused and unfocused cases of /ʌnni/ are not possible, because in the unfocused case, there is no clear realization of LH tones to compare. Nevertheless, the level of the f0 trough for this L tone is significantly lower than that for the L tone of the preceding (phrase-initial) word /uri/ (β = −0.326, SE = 0.071, p = 0.005), as can be inferred from the lower leftmost panel of Fig. 6a. In addition, the level of the f0 trough for the focused /ʌnni/ is significantly lower than that for the same word in the phrase-initial context (β = −0.180, SE = 0.065, p = 0.016). In a similar fashion, the f0 peak for focused /ʌnni/ is significantly lower than the same word in phrase-initial context (β = −0.457, SE = 0.082, p < 0.001). The magnitude of the rise for phrase-medial focused /ʌnni/ is significantly smaller than the phrase-initial counterpart (β = 0.277, SE = 0.086, p = 0.007). (Note that, compared to the phrase-initial context, the monosyllable target words in the phrase-medial context also exhibit a small but significant lowering effect on the f0 trough under focus, although the extent of this effect is smaller (β = −0.284, SE = 0.049, p < 0.001).) At present, we do not have a theoretically informed explanation for the position-related difference in the extent of f0 expansion. However, the additional lowering of f0 for the L tone, which is particularly evident in the phrase-medial context, may be due to more preparatory time being available to enhance the L tone under focus, as opposed to when focus occurs phrase-initially. This appears to be particularly noticeable in contexts where both the preceding initial word (/uri/) and the actual word under focus (/ʌnni/) consist entirely of sonorants, which may facilitate such expansion more than when the word under focus is a monosyllable with an obstruent onset (/pam/ and /pap/).

3.4. Effects of focus on phrasings

In this section, we examine whether the distribution of focus-induced prominence results in any noticeable changes to phrasing, particularly when the focus falls on the target words in non-initial positions—namely, phrase-medial and phrase-final contexts. (Recall that the target words in the present study are produced always with a new phrase in the phrase-initial contexts, regardless of their focus status).

Let us first examine the case with the target words in the phrase-medial context. As shown in the top panel of Fig. 5a, we have already observed that when the focus falls on the phrase-medial target word, it starts not only with the f0 trough for the L tone, but the preceding (pre-focal) phrase-initial word /ʌnni/ is produced with an LH pattern, showing a slight but noticeable f0 rise towards the end of the pre-focal context before the phrase-medial focused target word. This indicates that the overall distribution of tones still forms a typical LHLH pattern, with the first LH, though reduced in rising, falling on the pre-focal word, the focused word starting with the second LH, and an L% at the end. Here, the question arises as to whether the second LH, aligned with the focused target words (/pam/, /pap/), can be considered as evidence for forming a new Accentual Phrase (as proposed by Jun, 1996), or a new Intermediate Phrase (as proposed later by Jun, 2011). One could argue that the deeper f0 trough observed for the L tone in this case may signal the start of a new AP (or an Intermediate Phrase). This depth in the f0 trough can be attributed to the focus, which tends to lower the f0 more than what is typically seen in a phrase-medial L tone. Normally, this phrase-medial L tone would be less pronounced — not as deep as an initial L tone — unless influenced by such a focus. Indeed, such low-

ering of the L is often regarded as an indicator of a new AP, as proposed by Jun (1996, 2011). Moreover, as can be seen in the top panel of Fig. 5a, a clear voiceless gap is evident, associated with the onset /p/ of the target words /pap/ and /pam/. This suggests that it does not undergo the lenis stop voicing between sonorants, which would otherwise be the case if the stop were to occur in the middle of an AP since the lenis stop voicing process is considered to be blocked by the AP boundary (Jun, 1996, p. 95, Jun, 1998). Based on these observations, one could conclude that the second LH, aligned with the focused target words, indeed signals a formation of a new AP due to focus.

However, there are reasons to consider an alternative perspective against the focus-induced initiation of a new AP. First, the preceding phrase-initial word /ʌnni/ (the top leftmost panel of Fig. 5a) does not end with an f0 rise that is typical for the H as an edge tone marking the end of an AP. Instead, it is notably undershot, a pattern frequently observed when the H tone appears in the middle of a phrase (e.g., Jun, 1996). Second, the reduced f0 rise for /ʌnni/ bears a striking resemblance to that in the middle leftmost panel of Fig. 5a. In the latter, both the initial word /ʌnni/ and the target word are unfocused, and they are grouped together for a single AP with no discernible tonal and segmental signatures that indicate an Accentual Phrase boundary between the two. (Note that all the utterances are assumed to start with an AP whose left edge is aligned with the left edge of an immediately higher level of phrase, in line with the Strict Layering Hypothesis (Selkirk, 1984, as is also adopted by Jun, 1996). Thus, the level of the f0 rise for /ʌnni/ in the middle of an AP when the following target word is unfocused (in the middle panel of Fig. 5a) is comparable to when the following target word is focused (in the top panel of Fig. 5a). This undermines the possibility that the focused target word forms a new Accentual Phrase or the like. Similarly, the duration of /ʌnni/ preceding the focused target word is relatively short, akin to the duration of /ʌnni/ when the following target word is not focused. There is indeed no statistical difference in the duration of /ʌnni/ between when the following word is focused and when it is not ($\beta = -5.342$, SE = 6.078, $p = 0.395$). Although an AP may not be subject to phrase final lengthening as is usually observed with a larger phrase (Jun, 1995, 2005), it is often reported that it may be accompanied by a small but significant phrase-final lengthening effect (e.g., Cho & Keating, 2001). Moreover, even if an AP does not induce any phrase-final lengthening effect, one would still expect a relatively longer duration of /ʌnni/ when it alone forms a single Accentual Phrase than when it is produced as part of an AP with the following target word. This is because the syllable duration is, all else being equal, expected to be relatively shorter in the latter case, due to a polysyllabic shortening effect that is usually observed within an AP. In other words, had /ʌnni/ been produced as an independent Accentual Phrase followed by another AP initiated by the focused target word, its duration would have been relatively longer than when it is grouped together with the following unfocused target word.

Given these observations, it appears that focus can manifest itself within an Accentual Phrase. From this perspective, the increased duration and voiceless gap observed at the onset of /p/ in the target word, often considered a signature of a new AP, can instead be attributed to the segmental hyperarticulation occurring in the phrase-medial word under focus. This point is discussed further in the Discussion section. In other words, when focus falls on a word in the middle of a phrase, it can inhibit the lenis stop voicing without necessarily creating an AP, simply heightening the phonetic clarity of the underlying voiceless characteristic of the lenis stop, as a phonetic feature enhancement through localized hyperarticulation (Cho, 2016; de Jong, 1995). In this account, the focused word can also influence LH tone alignment phrase-internally rather than creating a new phrase, causing the L tone to anchor in the first syllable of the focused word, followed by a subsequent rise, which serves to further enhance the prominence associated with the focus. In fact, as discussed before, we observed cases where the L tone appears to be attracted to the initial syllable of the focused word in the phrase-medial and final contexts.

This alternative account that the focus can be realized in the middle of a phrase without creating a new phrase is further buttressed when we consider the cases with the target words positioned at the end of the phrase. Recall that even though the phrase-final target word under focus is produced with a small f0 rise that may correspond to the LH tones, its magnitude is not substantial enough to be considered as forming a new AP (see the upper rightmost panel of Fig. 6a). Moreover, the preceding contextual word /ʌnni/ was not produced with a rising tone that would have otherwise signified an upcoming AP boundary. Instead, the opposite was true—i.e., it was produced with a falling f0 that continued to the beginning of the following syllable of the target word that is focused. As such, there is no tonal evidence for the formation of a new Accentual Phrase when focus falls on the phrase-final target word. Here again, the observed voiceless gap, as shown in the upper rightmost panel of Fig. 6a, can be interpreted as a hyperarticulation of the voiceless lenis stop in the onset of the target word under focus. Specifically, the blocking of the lenis stop voicing can be attributed to a focus-induced enhancement of its voiceless phonetic feature, localized to the phrase-medial word under focus, rather than to a boundary-related effect.

Yet another supportive case can be observed in the tonal realization around the contextual word /twieta/ when it receives focus, positioned after the target words in the phrase-initial context. Once again, as shown in the lower center panel of Fig. 4a, when the second word /twieta/ is in focus, the preceding word does not end with a rising tone, which would otherwise serve as intonational evidence for an upcoming AP boundary. Instead, the f0 falls as a continuation of the initial L tone. Recall that in this context, the L tone does not appear to be completed within the first monosyllabic word. Instead, the f0 continues to descend until the onset of the subsequent word /twieta/, where it forms a distinct f0 trough for the L tone at the beginning of the focused /twieta/. This continuous f0 decline across the word boundary suggests that the two words are phonologically grouped together to form an AP, rather than being separated by an AP boundary. It is also noticeable that in this case, the subsequent f0 peak for the H tone is further shifted to the right. These observations suggest that focus modifies the distribution of tonal (LH) patterns without generating a new AP. This case demonstrates yet another instance

where the L tone aligns with the first syllable of the focused word, and the H tone shifts rightward onto the subsequent syllable, all without altering the phrasing.

4. Summary and discussion

The present study investigated the distribution of tones in Seoul Korean within relatively short Intonational Phrases when contrastive focus falls on monosyllabic target words (/pam/ and /pap/) in different positions—namely, IP-initial, IP-medial, and IP-final. In the following discussion, we will summarize and discuss findings aligned with the specific questions upon which the present study embarked.

4.1. Focus effects on the direction and extent of pitch range for the L and H tones

Our first question sought to investigate the direction and extent of pitch range modification associated with the phrase-level realization of L and H tones (the assumed edge tones at the Accentual Phrase level), especially when contrastive focus falls on monosyllabic target words in IP-initial position. (Hereafter, 'contrastive focus' will simply be referred to as 'focus' unless specified otherwise.) The results indicated that when the target words are focused, the pitch range of the f0 rise for both L and H tones is indeed expanded, compared to their pre-focal (unfocused) counterparts. However, the directionality of the tonal expansion did not indicate a bidirectional polarization, which would involve 'hyperarticulation' of both L and H tones under focus—a phenomenon observed in tone languages (e.g., Lee et al., 2016; Chen & Gussenhoven, 2008) and lexical pitch accent languages (Cho et al., 2019). Instead, the focus-induced expansion was attributed to the heightened f0 peak of the H tone on the second syllable, while the f0 trough for the L tone on the first syllable remained unchanged, showing an asymmetrical contribution of each tone to the expansion. (The same pattern was also observed with the contextual word /ʌnni/ ('older sister') when it occurred phrase-initially with and without focus.)[9] This asymmetry is in line with the typical directionality of focus-induced pitch range expansion observed with bitonal elements in other non-tonal West-Germanic languages (e.g., Féry & Kügler, 2008; Liberman & Pierrehumbert, 1984). This is presumably driven by physiological constraints—namely, the lower bound for the f0 trough is closer to physiological limits, while the upper bound for the f0 peak is less restricted, allowing for a greater degree of freedom for expansion in the upper range (Pierrehumbert, 1980). Seoul Korean appears to follow this general phonetically-driven cross-linguistic trend, even though its L and H tones may function as independent tonal targets within their respective domains, rather than as bitonal elements like L + H* or L*+H, where the two tones seem more tightly associated with each other. It appears that speakers of non-tonal languages including Seoul Korean opt for the low-cost option by primarily modifying the upper bound of the pitch range for expressing focus, in spirit of the principle of effort minimization (e.g., Lindblom, 1990; Flemming, 1995), when the prominence gained does not directly enhance lexical distinction.[10]

Before moving to the next section to address our second question, it is worth noting that this interpretation also has further implications for the phonological status of the L tone in the lenis stops of our target words /pam/ and /pap/. Related to the use of the L tone in association with the lenis stop, recent studies have suggested that the distinction between lenis and aspirated stops in Seoul Korean is no longer marked by VOT differences—historically shorter for lenis than for aspirated stops—but rather by L and H tones, which have become the primary phonetic features for phonological and lexical distinctions. This modification of the cue primacy has led some researchers to posit that Seoul Korean has undergone a tonogenetic sound change (e.g., Kang, 2014; Bang et al., 2018). However, the fact that the L tone in Seoul Korean is not hyperarticulated, showing no enhancement of f0 lowering, contrasts with languages where tones are used for lexical distinction. This includes Kyungsang Korean, a lexical pitch accent language in which the L tone has been shown to undergo further lowering under focus (e.g., Cho et al., 2019). The absence of such L tone enhancement in Seoul Korean challenges the 'Tonogenetic Account,' which posits that tones are primary features that yield phonological contrasts and, consequently, lexical distinctions. Were the L tone used as the primary phonological feature for lexical distinction in Seoul Korean, one might predict that it would exhibit tonal hyperarticulation, similar to what has been observed in other tone or lexical pitch accent languages. Instead, it appears that the L tone serves as a phrase-level tone assigned post-lexically by the intonational phonology of the language. This is consistent with the 'Prosodic Account' proposed by Choi et al. (2020). This account posits that the use of the L tone is conditioned by prosodic structure, which allows speakers to utilize L versus H tones when available phrase-initially at the post-lexical level. The rationale is that since the post-lexical L and H tones invariably occur in the phrase-initial position, corresponding to the lenis and the aspirated stop, respectively, the existing VOT differences have become redundant. Thus, here again, speakers seem to opt for a low-cost option, but this time by eschewing redundancy—i.e., by forgoing redundant VOT features in that prosodic context, namely, when it is licensed by the prosodic structure.

4.2. Tone-segment alignment under focus on the IP-initial, IP-medial, and IP-final target words

Our second question pertained to an understanding of the general alignment patterns of the LH phrasal tones with the segmental string in the focal and non-focal contexts. As for the phrase-initial context, a basic pattern found is that when the target word receives contrastive focus, the left-edge L tone is realized during the first syllable, which also forms the mono-

[9] In a phrase-medial position, the contextual word /ʌnni/ exhibited a more pronounced lowering of the L tone under focus, compared to its phrase-initial occurrence. However, this lowering effect was not attributed solely to focus, as both initial and medial words were compared while under focus. Instead, this effect was partly interpreted as phonetic facilitation of lowering in the middle of a phrase, preceded by the vowel-final word /uri/. In addition, this lowering effect could also be partly attributed to the general tendency of f0 declination over the course of speech.

[10] Another comparable case driven by the physiological constraints on f0 range variation may be the pattern of catathesis (downstep)—i.e., the f0 declination over the course of an utterance is more pronounced on the f0 peaks than on the f0 troughs (e.g., Liberman & Pierrehumbert, 1984; Beckman & Pierrehumbert, 1986).

syllabic target word (/pam/, /pap/) and the subsequent H tone in the second syllable (the first syllable of the following contextual word /twieta/). This pattern is largely consistent with what has been generally observed for the IP-initial LH tones in Seoul Korean (Jun, 1996, 2005; Jeon & Nolan, 2017), including the cases with no focus on the initial words (e.g., Cho, 2010, 2011). However, the specific alignment patterns identified in this study warrant further discussion about the segmental anchoring of L versus H, as elaborated below.

Recall that the L tone was found to form an f0 trough approximately one-third of the way through the rime in the sonorant context of /pam/, but at the acoustic offset of the vowel in the obstruent context of /pap/. These alignment points show a slight deviation from the estimated alignment points reported in H. Cho (2011) for targets with no focus (be it narrow or contrastive)—i.e., the f0 trough tends to occur around the midpoint of the rime in the first syllable when in sonorant contexts, with no focus on the initial target words. Note, however, that our intention is not to establish a specific time point, like the one-third point of the rime, as the invariable anchoring point dictated by the language. This is because its alignment may vary depending on various factors which are imposed on the target words or the utterance as a whole. Specifically, in an acoustic study and its accompanying model, which focuses on a three-syllable target word followed by a monosyllabic case marker, H. Cho (2010, 2011) suggests that such variation may reflect an interaction between the language's preferred anchoring point and the level of time pressure imposed, for instance, on the tonal movement by the speaking rate of the utterance. Thus, the alignment timing of the L tone observed in the present study (roughly a one-third of the rime in the sonorant context) simply indicates a particular alignment pattern in the specific focus context with monosyllabic target words.

Nevertheless, H. Cho (2011) further demonstrates an important aspect of the timing difference between the L and H tones. Specifically, the L trough for the left-edge L tone exhibits relatively stable temporal alignment compared to the subsequent f0 peak of the H tone. This stability is evident in the fact that the alignment time for the L tone is less influenced by changes in speaking rate than that of the H tone. Such relative temporal stability for the initial L tone despite durational differences due to global speaking rate changes is interpreted as indicating that the language's preferred anchoring point takes precedence over any potential adjustments due to time pressure. In fact, consistent with this finding, our GAMM analyses of sonorant portions on a normalized (relative) time scale also indicate that the timing of the local f0 trough for the left-edge L tone during the monosyllabic target word remains constant. In our case, the constancy occurs regardless of the obvious duration change when the target word is focused versus unfocused (pre-focal), being longer versus shorter, respectively, as also observed in Jeon & Nolan (2017), rather than due to changes in the global speaking rate as tested in H. Cho (2011). Our finding thus builds on H. Cho's finding, suggesting that in Seoul Korean, the presumed anchor point is given more weight than any possible alterations caused by time pressure, in this case, the temporal variation arising from localized focal versus non-focal effects. This also dovetails with the assumption that the left-edge L tone is phonologically associated with the left-edge syllable, as discussed in Jun (1996, 2005), which in turn implies a relatively stable segmental anchoring point determining the fine phonetic detail of tone–segment alignment in conjunction with the language's intonational phonology (as discussed in Ladd, 2008).

However, a caveat is necessary regarding the interpretation of the relative timing invariance associated with the left-edge L tone's alignment. Recall that when the initial target words were not focal (with the focus on the subsequent contextual word /twieta/), the initial f0 trough during the first syllable was not fully formed, resembling more of an elbow without a clear rise. This was followed by the f0 falling more deeply at the beginning of the next word that was focused (see the lower panel of Fig. 4a). One might interpret this as suggesting that the anchorage of the initial L tone could be drawn towards the focused word, namely, to the second syllable, which is the first syllable of the next contextual word receiving focus. However, we also observed the striking similarity in the timing of the local f0 trough for the monosyllabic target words in both focal and non-focal contexts, implying a phonetic reflex of the L tone even in the non-focal context. We therefore propose that the left-edge L tone is consistently assigned to the first syllable as stipulated by the intonational phonology of Seoul Korean, and that an extra L tone should be assigned to the first syllable of the subsequent word receiving focus, although the extra L tone may be seen as a left-edge tone of a new phrase (see below for further discussion).

In the case of the H tone, its f0 peak was found to form at the beginning of the first syllable of the subsequent contextual word /twieta/ when the initial target words were focused.[11] Some small timing adjustments were also noted due to the segmental makeup of the preceding coda conditions. Specifically, the f0 peak occurred slightly earlier at the beginning of the vowel in /twieta/ following the obstruent context of /pap/, which was attributed to the segmental perturbation effect of the voiceless /t/ on f0 realization. However, a different picture emerged when the focus was on the contextual word /twieta/. In this case, the f0 peak shifted significantly to the right, approximately during the second syllable of /twieta/, though the exact boundary between the /i/ of /twi/ and /e/ in /twieta/ was not clearly determined. Recall that in this case, there was no clear phonetic evidence for any noticeable f0 peak before this delayed f0 peak when /twieta/ was focused. Put together, when the focus shifted to the contextual word /twieta/ following the initial target word, there was not only a deeper f0 trough at the beginning of the focused /twieta/, resulting from a continued f0 fall from the preceding local f0 minimum, but also a noticeable shift of the f0 peak to the right.

As for the phrase-medial target words, we also observed a clear f0 trough during the rime of the focused target word. Sim-

[11] This phenomenon, where focal expansion can extend to a subsequent post-focal element if the focused constituent is too brief to accommodate the focal rise, diverges from patterns observed in tonal languages such as Mandarin. In Mandarin, pitch range expansion is typically confined to the focused element, regardless of its monosyllabic nature (Y.-C. Lee et al., 2016; Xu, 1999). This variance could be attributed to the differing origins of tone: phrasal specification in Seoul Korean versus lexical specification in Mandarin Chinese. Alternatively, the distinction may arise from limitations on the association of tones with the segmental string. In Seoul Korean, the allocation of Accentual Phrase (AP) tones is generally restricted to one per syllable unless an AP consists of a monosyllabic word (Jun 1996), whereas Mandarin allows syllables to carry contour tones. Consequently, in Seoul Korean, the focal rise must span at least two syllables, including one that is post-focal.

ilar to the phrase-initial contexts under focus, this f0 trough was positioned roughly one-third of the way through the rime in the sonorant context of /pam/, but at the acoustic offset of the vowel in the obstruent context of /pap/. Again, there followed a f0 peak indicating a subsequent H tone being aligned on the first syllable of the following contextual word /twie/ which was phrase-final. These L and H alignment patterns are remarkably similar to those observed with the phrase-initial target words receiving focus, further substantiating a consistent focus-induced temporal realization of the L and H tones, operating in both phrase-initial and phrase-medial contexts.

Finally, when the monosyllabic target word occurred IP-finally, there disappeared clear phonetic evidence at the macroscopic level of tones for focus-induced LH patterns. Instead, there was a unmistakable f0 falling for the boundary tone L%, in line with the phonological assumption that the right-edge tone is overridden by the boundary tone (Jun, 1996, 2005). However, we observed a small f0 rise at the microprosodic level at the beginning of the focused target word's vowel. This was interpreted as a phonetic remnant of the focus-induced LH tones that may be suppressed, although not completely, in the final syllable, rather than categorically replaced by the boundary tone. Upon examination of the tonal realization during the final syllable, we identified two key features as evidence of a tonal crowding effect: the extremely early (albeit reduced) rise of f0, and the relatively delayed realization of the f0 trough for the L% tone under focus. In this effect, the f0 rise was pushed leftward, away from the L% (tonal compression), and the f0 rise for the H tone is substantially truncated by this shift (tonal truncation). To the extent that this interpretation holds, it suggests that the assumed focus-induced LH tones, regardless of their association with focus or an Accentual Phrase (a topic we will discuss later), which may otherwise appear absent at the macroscopic level of intonational phonology, are still expressed in fine phonetic detail.

4.3. Effects of segmental makeup of the coda on the f0 realization

Another question the present study aimed to answer concerned the extent to which the segmental makeup of the coda as a sonorant versus a voiceless obstruent (/pam/ vs. /pap/) would physically constrain the f0 rise from the L to the H tone. The primary distinction between the coda conditions lied in the presence or absence of f0 discontinuity in /pap/ and /pam/, respectively. As noted earlier, after the obstruent context of /pap/, there was a segmental perturbation effect at the micro-prosodic level on the onset of the following contextual word /twieta/, resulting in a slightly higher and earlier f0 peak compared to the case after /pam/. Apart from this micro-prosodic effect, the results showed that the f0 rise from the L to the H tone over a voiceless obstruent coda is largely similar to that over a sonorant coda, both in achieving the target f0 height and in terms of alignment. This implies that in the presence of a voiceless gap, a form of 'virtual' f0 interpolation takes place, aiming to meet the assumed f0 target for the H tone. Such a virtual interpolation, bridging the gap, effectively ensures that the f0 peak is not undershot, countering the challenging physical context of voicelessness. Thus, it appears that just as listeners can unconsciously interpolate through short missing sections during voiceless intervals to preserve the intended contour shape (as observed by Nooteboom, 1997), speakers also make similar adjustments to achieve the intended target. In such a production process, speakers may need to exert additional articulatory effort to overcome contextually imposed adverse conditions. The effort can be interpreted as the speaker's deliberate action to avoid the undershoot of the target specified by the underlying H tone, rather than merely maintaining the contour shape. This interpretation is consistent with the fundamental assumptions of AM Theories of Intonation (refer to Ladd, 2008 for a review), where phonologically assigned tonal targets are assumed to be achieved independently of the segmental string.

4.4. Relationship between focus-induced tonal distribution and phrasing

Finally, the results of the present study provide a foreground for further discussion on the relationship between focus-induced tonal distribution and phrasing. As introduced at the beginning of this paper, Jun (2007, 2011) proposed that focus marking in Seoul Korean (especially narrow or contrastive focus) is generally achieved by means of phrasing: the focused element starts a new phrase, such as an Accentual Phrase or an Intermediate Phrase. However, our study departed from this assumption, considering the possibility that, although focus is often produced with the focused element positioned at the left edge of a phrase in probabilistic terms, the left-edge alignment condition is not mandatory. This allows for instances where focus can occur within the phrase under certain circumstances. We discussed this possibility in detail in Section 3.4, titled 'Effects of focus on phrasings.' In this section, we will provide a brief summary.

In Section 3.4, we highlighted the cases of 'phrase-medial' and 'phrase-final' target words (/pam/ or /pap/) receiving focus, as they have the potential to form a new phrase when focused. As for the 'phrase-medial' case, the pre-focal contextual word /ʌnni/ that preceded the focused target word did not end with the typical f0 rise characteristic of the H as an edge tone marking the end of an AP. Instead, we interpreted this pre-focal rise as being undershot, as commonly observed phrase-medially (e.g., Jun, 1996), with a noticeable reduction of the f0 peak and the f0 falling well before the right edge, hence not clearly marking the end of an Accentual Phrase or an Intermediate Phrase. The observation that the following focused target word began with a relatively deep L tone could thus imply that focus prompts the assignment of an extra L tone to the non-initial target word independently of phrasing. Furthermore, the reduced f0 rise in /ʌnni/ preceding the focused target word closely resembled its pattern before an unfocused target word. In this latter case, the two consecutive words /ʌnni/ and the target word were both unfocused, and were clearly grouped together as part of a single AP. The similarity of the f0 reduction in both cases reinforces our interpretation that focusing the target word in the examined IP-medial context does not necessarily lead to the formation of a new phrase at the level of an Accentual Phrase or an Intermediate Phrase. The possibility for the phrase-independent focus realization was crystallized by the evidence from the target words positioned phrase-finally. In this case, the monosyllabic focused word was aligned with the right edge of the Intonational Phrase. Crucially, the

absence of tonal evidence, such as a canonical phrase-initial f0 rise for the L and H tones during the phrase-final target word under focus, indicated that no new phrase formation occurred phrase-finally. Furthermore, there was no noticeable pre-focal f0 rise that could otherwise have signaled the end of a phrase as an indication of a potential prosodic juncture at the phrase level.

Comparatively, our findings also resonate with quite a few previous studies with an assumption that focus occurred without introducing a new phrase (e.g., Cho & Keating, 2001; Y. Lee, 2015; Cho et al., 2014; Choi et al., 2020). Among these studies, a particular attention can be paid to the research conducted by Y.-C. Lee (2015). This study investigated the phonetic realization of corrective (contrastive) focus within phone numbers that form a whole subsequence, produced by speakers of various languages including Seoul Korean. In the case of Seoul Korean speakers, when the corrective emphasis was placed on numbers positioned either in the middle or at the end of a subsequence of the digits of a phone number, the pitch expansion that had been typically observed with focus, was either significantly diminished or entirely absent. Such observations also hint at the capacity of Seoul Korean speakers to effectively convey corrective focus even in the absence of prosodic breaks. It suggests that when there's a compelling communicative reason, in this specific case, to maintain a prosodic group, such as the subgrouping of a phone number, speakers are adept at conveying the necessary corrective focus without resorting to prosodic rephrasing. This is again in line with our view of the autonomy of prosody, underscoring that while prosody can operate independently, it still remains flexible adapting to structural constraints that arise in a particular communicative context (in this case, forming a subsequence of the digits of a phone numbers), other things being equal.

Recall, however, that in both IP-medial and IP-final contexts, the onset /p/ of the target words /pam/ and /pap/ was still produced with increased duration and a clear voiceless gap, demonstrating the blocking of the lenis stop voicing process in the intervocalic contexts. (Remember that the target words were preceded by the vowel-final word /ʌnni/, creating an intervocalic context in which /p/ was subject to the lenis stop voicing process.). It might be argued that such a blocking of the lenis stop voice rule in Seoul Korean could be considered as evidence for the presence of an AP or a higher phrase, given that its domain of application is generally assumed to be within an Accentual Phrase (e.g., Jun, 1998, 2005). However, given the lack of evidence supporting the creation of an AP or a higher-level prosodic structure with the focused target word, the prevention of lenis stop voicing in this context could be interpreted as localized hyperarticulation (enhancing the underlying voiceless feature) linked to focus (e.g., de Jong, 1995; Cho, 2016), rather than being related to phrasing.[12] In fact, a visual inspection of the entire dataset indicated that the lenis stop in the phrase-medial context was completely devoiced in the focused context, and the closure duration was twice as long as the often voiced tokens in the unfocused context (ranging from 107 to 111 ms when focused to 55–65 ms when unfocused, phrase-medially). Both of these findings indicate hyperarticulation, which typically involves both spatial and temporal expansion of articulation.

These observations lead us to propose a revised account of the relationship between focus and phrasing in Seoul Korean. Following Jun (2005, 2007, 2011), we acknowledge that as an 'edge-prominence' language, Seoul Korean's prosodic system typically maps focus (especially narrow or contrastive focus) onto phrasing. More broadly, Jun's proposal represents a focus-prosody mapping theory, implying that the information structure directly influences the prosodic structure of an utterance, aligning the focused element with the left edge of a phrase. But we suggest that Jun's proposed focus-prosody mapping is only one of multiple constraints that could play a role in determining the final shape of the prosodic structure. This is in line with the theoretical tenet that prosodic structure is an independent grammatical entity, parsed in its own right (Beckman, 1996; Shattuck-Hufnagel & Turk, 1996), allowing for the possibility of negating or mitigating the impact of a particular constraint when necessary in forming a prosodic structure. Thus, the weight of any constraint imposed on determining prosodic structure can, in principle, be adjusted based on its interaction with other constraints in specific contexts that require them. These constraints encompass various factors, including utterance length, speaking rate, semantic weight, syntactic structure, and information structure, as discussed in Shattuck-Hufnagel and Turk (1996) and Jun (1996). It is evident that focus is just one among many factors at play. With respect to current findings, our test sentences were produced in a specific context: they are short, comprising only five syllables from three words in total, and the target word that receives contrastive focus is monosyllabic. In such short utterances with monosyllabic target words, a length-driven factor appears to come into effect, favoring focus realization on the monosyllabic target word without the creation of a new phrase.

Before concluding, we present three potential representations of focus based on our findings and discussions thus far. These representations are specific to our cases with monosyllabic target words in short utterances, allowing for phrase-internal focus realization independently of phrasing. Nonetheless, our proposed representations, encompassing both phrase-internal cases and the more conventional phrasing-initial one as discussed below, aim to establish a foundational framework for the development of more comprehensive and unified theories regarding representations of focus realization in Seoul Korean.

Fig. 7a shows a representation when contrastive focus falls on the phrase-initial target word. In this case, the realization of the remaining post-focal AP tones after the initial L and H tones is suppressed due to a post-focal dephrasing (or reduction) process, showing a relatively straightforward interpolation from the f0 peak of the H to the phrase-final boundary tone L%. In the representation, the post-focal dephrasing is depicted as deletion of these tones, but one might assume that these tones may not be assigned to begin with generating a post-focal dephrasing contour. Notice that we have also specified the

[12] An alternative view posits that focus may simply induce a lengthening effect in the absence of hyperarticulation, a possibility our figures might suggest. However, research on languages such as English (de Jong 1995) and Arabic (de Jong and Zawaydeh 2002) indicates that prominence-induced prosodic strengthening often involves hyperarticulation, not merely an increase in duration. In our data, we observe both the prolonged durations and significantly higher Voice Onset Time (VOT) in initial lenis consonants of focal target words. This combination of features supports our analysis that hyperarticulation, incorporating both temporal extension and articulatory enhancement, more comprehensively signals focus than lengthening alone.

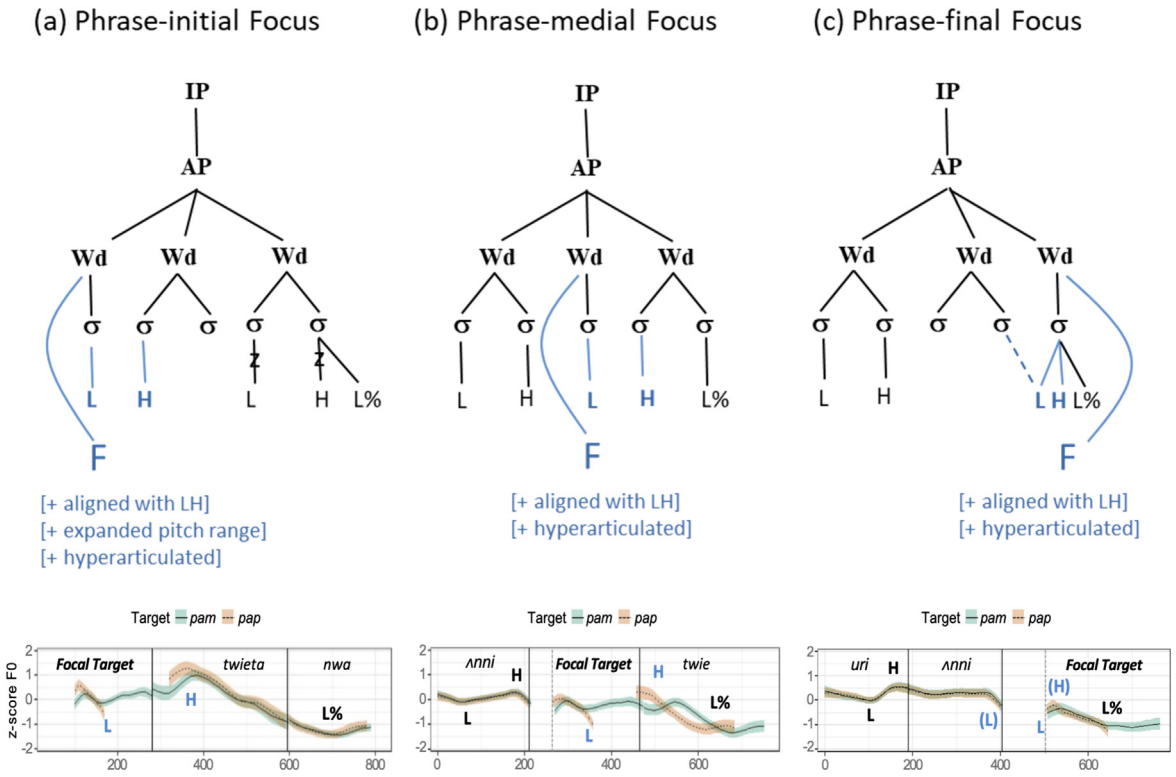

Fig. 7. Three possible representations of focus realization when focus falls on the monosyllabic target word at phrase-initial (a), phrase-medial (b), and phrase-final (c) positions. The solid blue lines indicate the association of tones aligned with contrastive focus. The dotted blue line represents a potential phonetic effect of L anchored at the final syllable, which may spread to the preceding syllable due to possible tonal crowding. Each of the lower panels shows the aggregate tonal contours taken from Fig. 4a, Fig. 5a, and Fig. 6a, respectively. (It is worth noting that within Jun's (1996, 2005) intonational phonology framework for Seoul Korean, although the association of the AP-internal (non-peripheral edge) tones tends to be with the syllable next to the edge, the choice of the anchoring syllable exhibits flexibility in contrast to the edge tone.). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

phrase-initial focus with three features: [+aligned with LH], [+expanded pitch range], and [+hyperarticulated]. These features characterize the monosyllabic focused word as being aligned with the L tone followed by the H tone, accompanied by an expanded pitch range and the expected segmental hyperarticulation. The [+aligned with LH] feature is accompanied by the f0 rise that creates perceptual saliency of focused elements, as a way of expressing prominence. The addition of [+expanded pitch range] enhances the level of the perceptual salience, a hallmark of focus realized phrase-initially.[13] Fig. 7b shows a representation when focus falls on the phrase-medial target word. Here again, the focus is marked by [+aligned with LH], as the phrase-medial focused word is also aligned with the L tone realized on the focused word followed by the H tone, creating perceptual saliency of focused elements, being followed by the L% boundary tone. It is also marked by [+hyperarticulated] as evident in the blocking of the lenis stop voicing, which is presumed to enhance the voicelessness feature for maximizing lexical distinction. Finally, Fig. 7c illustrates a scenario where the focus is placed on the phrase-final target word. As previously proposed, focus is still specified by [+aligned with LH], despite a significant reduction in the f0 rise due to tonal crowding. This specification is justified by the fine-grained phonetic evidence at the micro-prosodic level for the underlying LH tones. Additionally, the [+hyperarticulated] feature is evident in the segmental realization, with lenis stop voicing being blocked under focus.

5. Conclusion

The present study aimed to understand the realization characteristics of focus-related phrase-level tones, often referred to as edge tones that are post-lexically assigned at the level of the Accentual Phrase in Seoul Korean. As discussed thus far, we believe this study has indeed taken a step toward this goal by examining specific instances—i.e., the tonal realization in relation to focus on monosyllabic target words (/pam/ and /pap/) in three positions (IP-initial, IP-medial, IP-final) within short utterances. The findings and their consequential implications can be summarized as follows.

The focus is primarily marked by the L tone aligning with the monosyllabic word in focus, followed by the H tone in the next syllable, which creates an f0 rise that carries perceptual saliency to mark prominence. Notably, focus is most prominently expressed at the IP-initial position, where it aligns with the edge-related default prominence, possibly accompanied by domain-initial strengthening, characteristic of Seoul Korean as an 'edge-prominence' language. This combined effect results in an expanded pitch range that enhances the perceptual saliency of the focused elements at the phrase-initial position.[14]

[13] Since substantial pitch range expansion was observed only phrase-initially, it seems to generate a combined effect of focus and domain-initial strengthening, maximizing prominence, representing the strongest prosodic strengthening effect (see Cho, 2016 for a related discussion on prosodic strengthening).

[14] According to Terken (1991), listeners do not judge prominence based solely on the absolute f0 value of intonational peaks. Instead, they compute an f0 floor, or baseline, against which these values are gauged, facilitating a clearer distinction of the focused item from its surrounding context (cf. Gussenhoven's (2004) Effort Code).

Irrespective of focus conditions, however, the default edge-prominence appears to be upheld through the rather invariant alignment timing of the left-edge tone, L. This underscores that the language appears to assign greater weight to the presumed segmental anchoring point over other potential constraints, highlighting the primacy of aligning the tone with the phrase's left edge. The invariant nature of this phenomenon is further evidenced by how tones are consistently realized across both voiceless and voiced codas, overcoming adverse physical conditions to achieve the target H tone. Crucially, however, our results imply that focus may not necessarily align with the edge of a phrase in this particular context when the target word occurs phrase-medially or phrase-finally. This contradicts the general theoretical assumption that focus must be marked by means of phrasing in Seoul Korean. Nevertheless, there seems to be an invariant pattern in tonal expression for focus. Specifically, the AP-internal tones of L and H are realigned with the non-phrase-initial focused word, as if they were attracted to the focused element. This pattern holds true even though their realizations are greatly reduced phrase-finally due to tonal crowding. We also proposed three possible representations tailored to the short utterances we examined, where a monosyllabic target word in focus can be situated in one of three positions (IP-initial, IP-medial, IP-final). These representations show a unified way of expressing focus-induced prominence across the positions, such that the focused element exhibits the characteristics of [+aligned with LH] and [+hyperarticulated]. Furthermore, an additional feature of [+expanded pitch range] enhances saliency when focus is combined with left-edge prominence (or domain-initial strengthening).

These representations and the specific claims made in our present study must be considered provisional, as they pertain to specific cases involving monosyllabic target words in short utterances. These cases may not impose the length-related constraints seen in longer utterances, which often lead to the generation of multiple Accentual Phrases. More broadly, we acknowledge the fundamental limitations of a single study such as ours in comprehensively capturing all the intricate features of focus realization within a language's intonational phonology. It is therefore essential to recognize that future studies may unveil additional constraints, which will in turn refine our understanding of the weighting assigned to focus representation and phrasing by the linguistic system. As such, further studies are warranted to gain a comprehensive understanding of the nuanced intricacies of the focus-prosody mapping in Seoul Korean, particularly in its capacity as an 'edge-prominence' language. It will also bridge the existing gap in the literature, where the focus-prosody mapping is more extensively explored in West Germanic and Romance languages with post-lexical pitch accents in their intonational phonology.

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.

CRediT authorship contribution statement

Richard Hatcher: Writing – original draft, Visualization, Methodology, Formal analysis. **Hyunjung Joo:** Methodology, Data curation. **Sahyang Kim:** Writing – review & editing, Validation, Methodology, Funding acquisition, Data curation, Conceptualization. **Taehong Cho:** Writing – original draft, Visualization, Validation, Supervision, Methodology, Investigation, Funding acquisition, Data curation, Conceptualization.

Acknowledgements

We would like to thank all the Seoul Korean speakers for their participation of the experiment. We also thank the members of HIPCS as well as the audiences of HISPhonCog 2023 in Seoul and ICPhS 2023 in Prague for their valuable comments. Part of this work with preliminary analyses was published in the Proceedings of ICPhS 2023 as Hatcher et al. (2023): https://guarant.cz/icphs2023/677.pdf. This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2021S1A5C2A02086884) awarded to T. Cho (corresponding author).

Appendix A

Table A1
Model output of fixed effects for peak f0 of phrase-initial rise in sentences with phrase-initial targets. Contrasts are sum coded. There is a significant effect of target word, with higher peaks following /pap/. There is also a significant effect of focus, with focused targets (coded as "focused") evidencing higher peaks than focus on the phrase-medial word, /twieta/ (coded as "unfocused").

Fixed effects				
Term	$\hat{\beta}$	**SE($\hat{\beta}$)**	**t**	***p***
(Intercept)	1.07	0.08	14.02	0<.001
Itempam − pap	−0.24	0.03	−8.60	0<.001
Focusfocused − unfocused	0.25	0.07	3.43	0.004
Itempam − pap × Focusfocused − unfocused	−0.32	0.05	−5.80	0<.001

Model formula: `maxF0 ~ Item * Focus + (1 + Focus | Speaker)`.
Marginal R^2 = 0.07.

Table A2
Model output of fixed effects for peak alignment of phrase-initial rise as measured as lag from vowel onset of /twieta/ in sentences with phrase-initial targets. Contrasts are sum coded. There is a significant effect of target word, with earlier peaks following /pap/.

Fixed effects				
Term	$\hat{\beta}$	**SE($\hat{\beta}$)**	**t**	***p***
(Intercept)	67.69	8.17	8.28	<0.001
Itempam − pap	21.45	3.60	5.95	<0.001
Focusfocused − unfocused	−58.3	10.20	−5.72	<0.001
Itempam − pap × Focusfocused − unfocused	8.41	7.21	1.17	0.243

Model formula: `timeAtMax ~ Item * Focus + (1 + Focus | Speaker)`.
Marginal R^2 = 0.16.

Table A3
Model output of fixed effects for the magnitude of the phrase-initial rise in f0 in sentences with phrase-initial targets. Contrasts are sum coded. There is a significant effect of focus, with focused targets (coded as "focused") evidencing greater rises than focus on the phrase-medial word, /twieta/ (coded as "unfocused").

Fixed effects				
Term	$\hat{\beta}$	**SE($\hat{\beta}$)**	**t**	***p***
(Intercept)	2.02	0.11	13.17	<0.001
Itempam − pap	−0.27	0.04	−6.89	<0.001
Focusfocused − unfocused	0.51	0.07	7.37	<0.001
Itempam − pap × Focusfocused − unfocused	−0.39	0.08	−4.99	<0.001

Model formula: `range ~ Item * Focus + (1 + Focus | Speaker)`.
Marginal R^2 = 0.10.

Table A4
Model output of fixed effects for peak f0 of phrase-medial rise in sentences with phrase-medial targets. Contrasts are sum coded. There is a significant effect of target word, with higher peaks following /pap/.

Fixed effects				
Term	$\hat{\beta}$	**SE($\hat{\beta}$)**	**t**	***p***
(Intercept)	0.33	0.05	6.45	<0.001
Itempam − pap	−0.19	0.03	−5.65	<0.001
Focusfocused − unfocused	−0.01	0.10	−0.19	0.855
Itempam − pap × Focusfocused − unfocused	0.25	0.07	3.84	0.001

Model formula: `maxF0 ~ Item * Focus + (1 + Focus | Speaker)`.
Marginal R^2 = 0.16.

Table A5
Model output of fixed effects for peak alignment of phrase-medial rise as measured as lag from vowel onset of /twie/ in sentences with phrase-medial targets. Contrasts are sum coded. There is a significant effect of target word, with earlier peaks following /pap/.

Fixed effects				
Term	$\hat{\beta}$	**SE($\hat{\beta}$)**	**t**	***p***
(Intercept)	18.37	0.98	18.76	<0.001
Itempam − pap	1.79	0.01	2.94	0.003
Focusfocused − unfocused	−3.06	2.03	−1.51	0.15
Itempam − pap × Focusfocused − unfocused	−1.04	1.22	−0.86	0.39

Model formula: `timeAtMax ~ Item * Focus + (1 + Focus | Speaker)`.
Marginal R^2 = 0.03.

Table A6
Model output of fixed effects for minimum f0 of final boundary tone in sentences with phrase-final targets. Contrasts are sum coded. There is a significant difference for focus, with sentences with a phrase-medial focus (coded as "unfocused") having lower troughs than phrase-final focus (coded as "focused").

Fixed effects				
Term	$\hat{\beta}$	**SE($\hat{\beta}$)**	**t**	***p***
(Intercept)	−1.32	0.07	−19.48	<0.001
Itempam − pap	−0.07	0.02	−4.47	<0.001
Focusfocused − unfocused	−0.07	0.03	−2.36	0.033
Itempam − pap × Focusfocused − unfocused	0.07	0.03	2.38	0.017

Model formula: `minF0 ~ Item * Focus + (1 + Focus | Speaker)`.
Marginal R^2 = 0.01.

Table A7
Model output of fixed effects for minimum f0 of phrase-initial and −medial rises in sentences with phrase-final targets. There is a significant effect for Word, with phrase-medial /ʌnni/ evidencing lower troughs than phrase-initial /uri/.

Fixed effects				
Term	$\hat{\beta}$	**SE($\hat{\beta}$)**	**t**	***p***
(Intercept)	−0.59	0.07	−8.33	<0.001
Word [ʌnni]	−0.33	0.10	−3.36	0.005

Model formula: `minF0 ~ Word * Focus + (1 + Word | Speaker)`.
Marginal R^2 = 0.06.

Table A8
Model output of fixed effects for minimum f0 of phrase-initial and −medial rises on /ʌnni/ sentences with phrase-medial and phrase-final targets respectively. Contrasts are sum coded. There is a significant effect for phrasal position with phrase-medial /ʌnni/ showing lower troughs.

Fixed effects				
Term	$\hat{\beta}$	**SE($\hat{\beta}$)**	**t**	***p***
(Intercept)	−0.71	0.05	−13.05	<0.001
PhrasePositionIP medial − IP final	0.17	0.07	2.52	0.026

Model formula: `minF0 ~ PhrasePosition * Focus + (1 + PhrasePosition | Speaker)`.
Marginal R^2 = 0.03.

Table A9
Model output of fixed effects for magnitude of phrase-initial and −medial rises on /ʌnni/ sentences with phrase-medial and phrase-final targets respectively. Contrasts are sum coded. There is a significant effect for phrasal position, with phrase-initial /ʌnni/ (coded as "IP medial") showing greater rises.

Fixed effects				
Term	$\widehat{\beta}$	**SE($\widehat{\beta}$)**	**t**	***p***
(Intercept)	1.76	0.05	22.84	<0.001
PhrasePositionIP medial − IP final	0.29	0.07	3.29	0.006

Model formula: `range ~ PhrasePosition * Focus + (1 + PhrasePosition | Speaker)`.
Marginal R^2 = 0.07.

Table A10
Model output of fixed effects for minimum f0 of phrase-medial rises on target words /pam/ and /pap/ in sentences with phrase-medial targets. Contrasts are sum coded. There is a significant effect for focus, with phrase-medial focused targets evidencing lower troughs.

Fixed effects				
Term	$\widehat{\beta}$	**SE($\widehat{\beta}$)**	**t**	***p***
(Intercept)	−0.57	0.06	−9.93	<0.001
Itempam − pap	0.08	0.02	4.52	<0.001
Focusfocused − unfocused	0.30	0.10	3.15	0.008
Itempam − pap: Focusfocused − unfocused	−0.28	0.35	−8.01	<0.001

Model formula: `minF0 ~ Item * Focus + (1 + Focus | Speaker)`.
Marginal R^2 = 0.06.

Table A11
Model output of fixed effects for word duration of phrase-initial (pre-prefocal) and −medial (prefocal) /ʌnni/ in sentences with phrase-medial or phrase-final targets respectively. Contrasts are sum coded. There is no significant effect of focus location between pre-prefocal contexts (coded as "unfocused") and prefocal contexts (coded as "focused").

Fixed effects				
Term	$\widehat{\beta}$	**SE($\widehat{\beta}$)**	**t**	***p***
(Intercept)	302.46	6.88	43.98	<0.001
Itempam − pap	8.83	2.30	3.84	<0.001
Focusfocused – unfocused	6.29	6.15	1.02	0.323
Itempam – pap: Focusfocused – unfocused	−5.826	4.60	−1.266	0.21

Model formula: `duration ~ Item * Focus + (1 + Focus | Speaker)`.
Marginal R^2 = 0.005.

Appendix B

The superimposed f0 contours of the two focus conditions for sentences with phrase-initial targets are provided in Fig. B1. These smoothed contours were created in Praat (Boersma & Weenick, 2021) via the script, MAUSMOOTH (Cangemi, 2015). Different scales are used according to the sex of the particiants and the two left columns, representing the productions by female participants, are plotted from 100 to 300 Hz, while the right columns, productions by male participants, are scaled from 100 to 200 Hz. For each speaker, the lefthand plot shows the contour for sentences with phrase-initial focus, whereas phrase-medial focus is illustrated on the right. While the actual pitch range varies by speaker, the effect of focus position is evident in the location of the peak. When focus shifts from phrase-initial to phrase-medial position, the peak of the rising contour shifts rightword. For many of the speakers, the magnitude of the rise differs by focus position, with phrase-initial focus showing greater rises. Both of these patterns were also evident in the GAMMs discussed in section 3.1.

Fig. B2 presents overlaid f0 contours for three focal conditions in sentences with phrase-medial targets, generated using the MAUSMOOTH (Cangemi, 2015) in Praat (Boersma and Weenick, 2021). Contours for female participants (left columns) range from 100 to 300 Hz, whereas male participants' contours (right columns) span 100–200 Hz. Each speaker's data includes a left plot for phrase-initial focus, a central plot for phrase-medial focus, and a right plot for phrase-final focus. Notably, the peak location shifts rightward as focus moves from phrase-initial to phrase-medial, with phrase-initial focus exhibiting larger rises. Phrase-medial and phrase-final focus show similar contour shapes, with the latter's lack of distinct rise potentially attributed to the boundary tone's association with the phrase-final syllable. These trends align with the GAMMs outlined in section 3.2.

Fig. B3 displays superimposed f0 contours for two focus conditions in sentences with phrase-final targets, generated using MAUSMOOTH (Cangemi, 2015) script in Praat (Boersma & Weenink, 2021). Female participants' contours (left columns) range from 100 to 300 Hz, while male participants' contours (right columns) span 100–200 Hz. Each speaker's data features a left plot for phrase-medial focus and a right plot for phrase-final focus. Phrase-medial focus typically exhibits a mid-sentence rise on /ʌnni/, whereas phrase-final focus shows minimal or no rise. The absence of a pronounced rise in phrase-final focus might be attributed to the boundary tone associated with the phrase-final syllable. Notably, the boundary tone's fall, represented at the contour's end, varies by focus, with a later trough occurring when focus is on the final word. These observations are consistent with the GAMMs discussed in section 3.3.

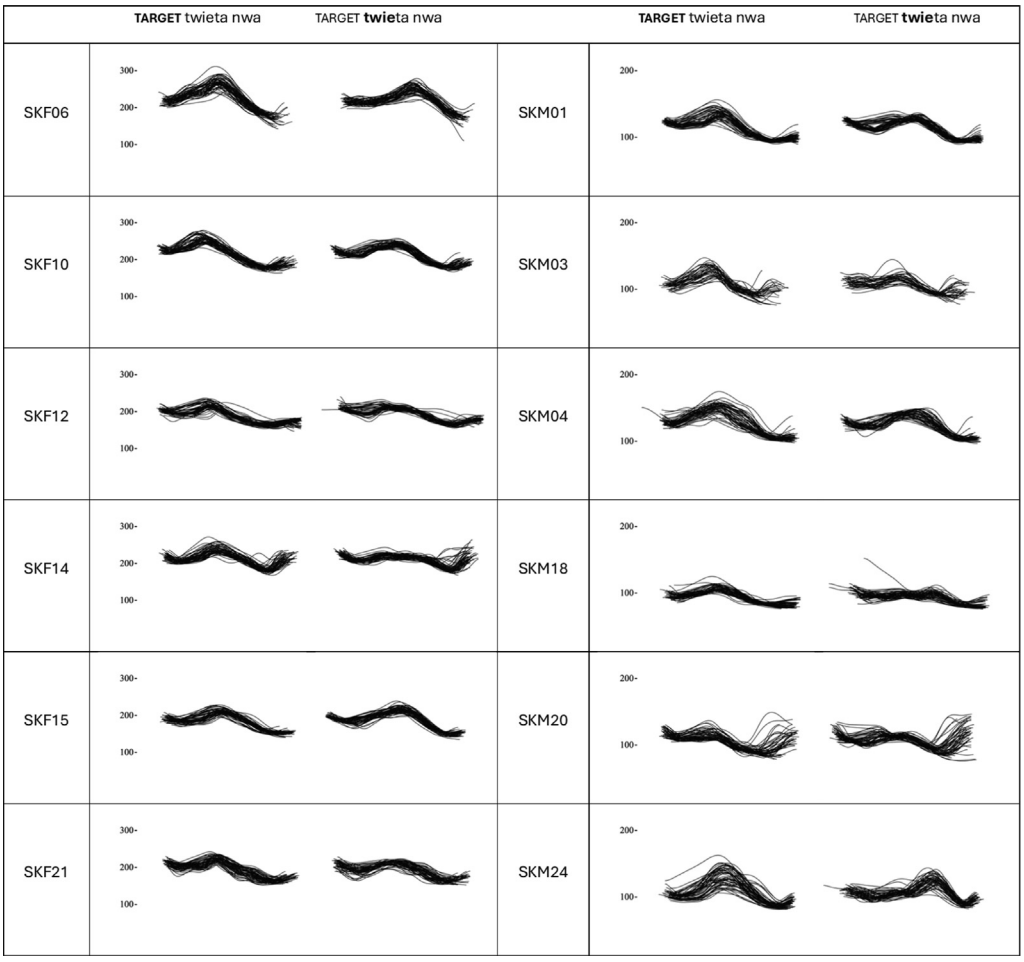

Fig. B1. Time-normalized, superimposed f0 contours are depicted for phrase-initial, and −medial focus in sentences with phrase-initial target words plotted collectively and each speaker's data presented individually.

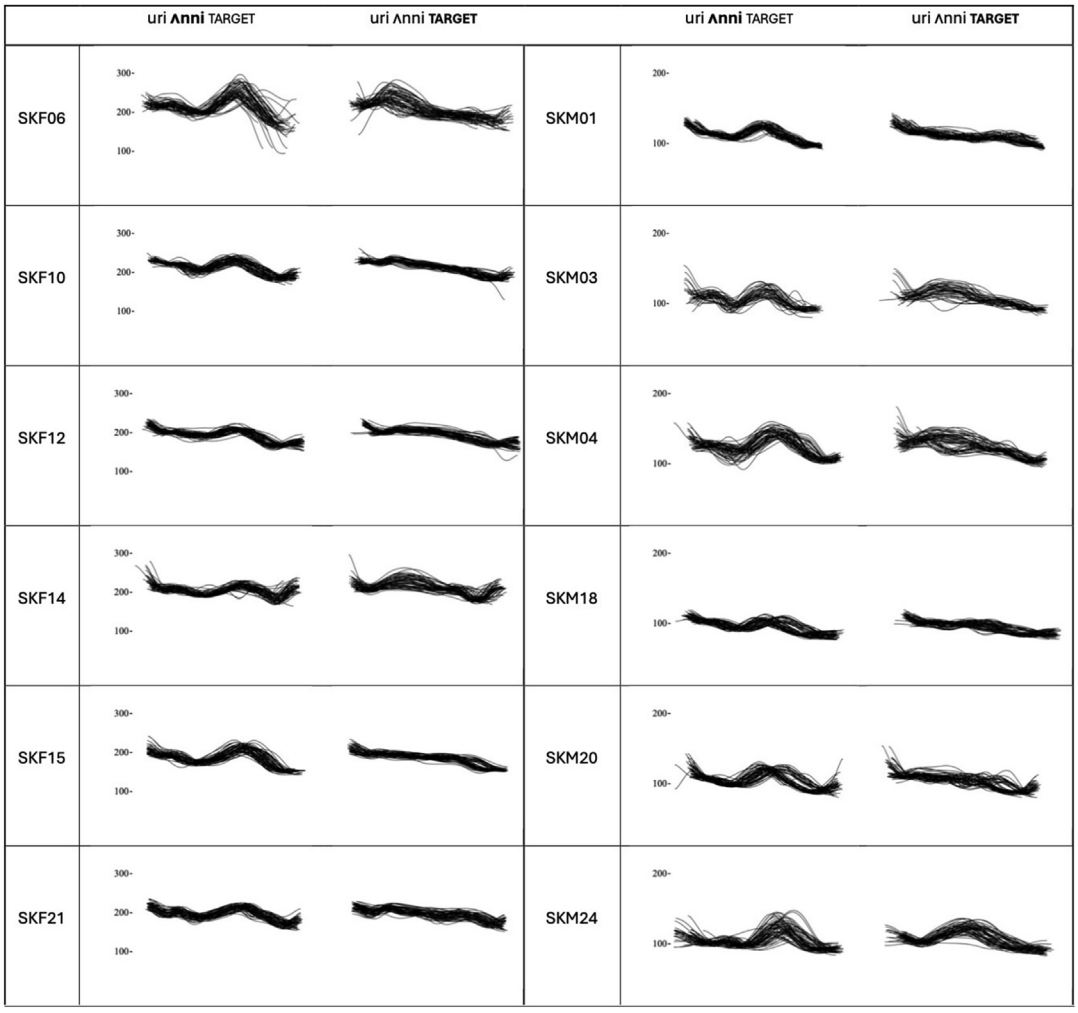

Fig. B2. Time-normalized f0 contours, superimposed for phrase-initial, -medial, and -final focus conditions, are presented for sentences containing phrase-medial target words plotted collectively and each speaker's data displayed individually.

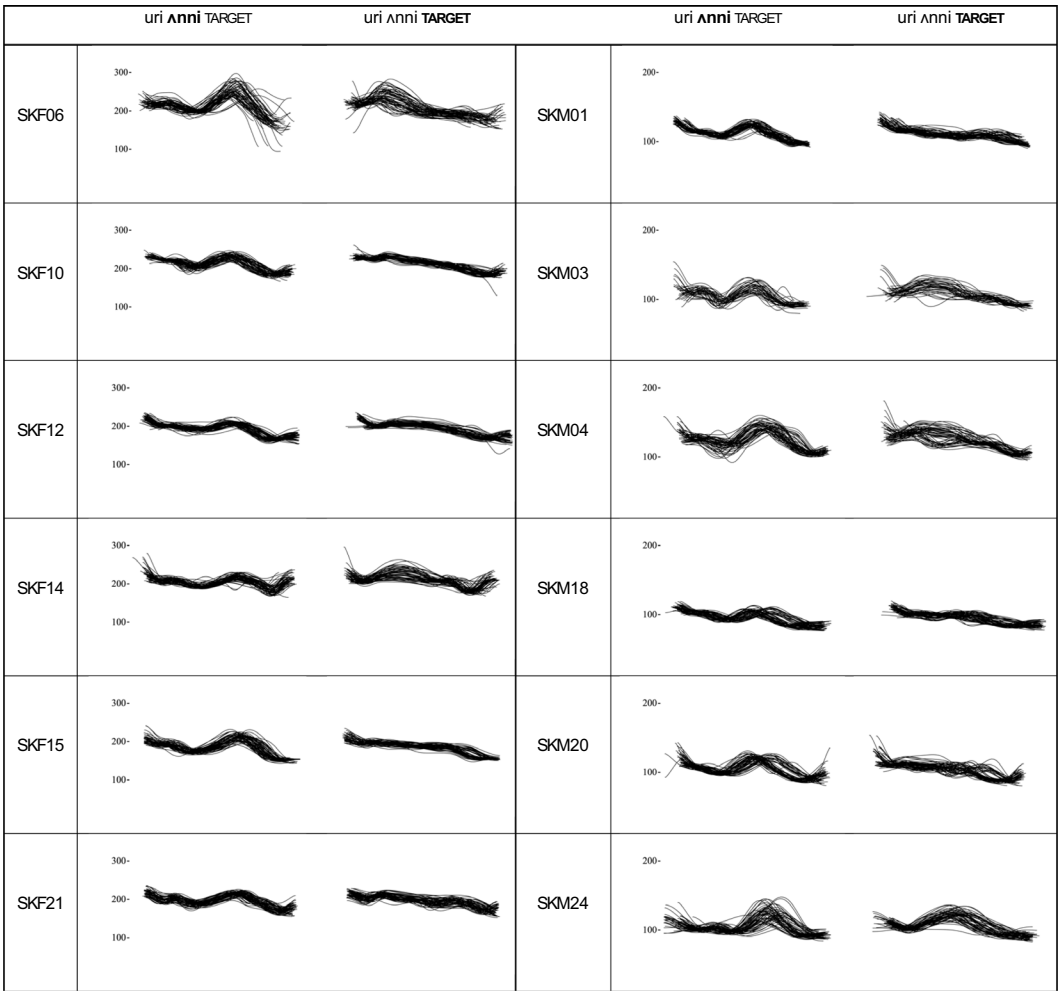

Fig. B3. Time-normalized, superimposed f0 contours are depicted for phrase-initial, -medial, and -final focus in sentences with phrase-medial target words plotted collectively and each speaker's data presented individually.

Appendix C

This appendix provides four tokens demonstrating phrase-final focus on target words /pam/ and /pap/. These words exhibit a slight fall on the offset of the pre-target word /ʌnni/ and a slight rise at the onset of the target word's rime. These tokens from two different speakers exemplify individual instances displaying the pattern observed in the GAMM smooths of Fig. 6a. Contrary to what is predicted from consonant-induced f0 perturbation, we find a rise at the onset following aspiration instead of the expected fall.

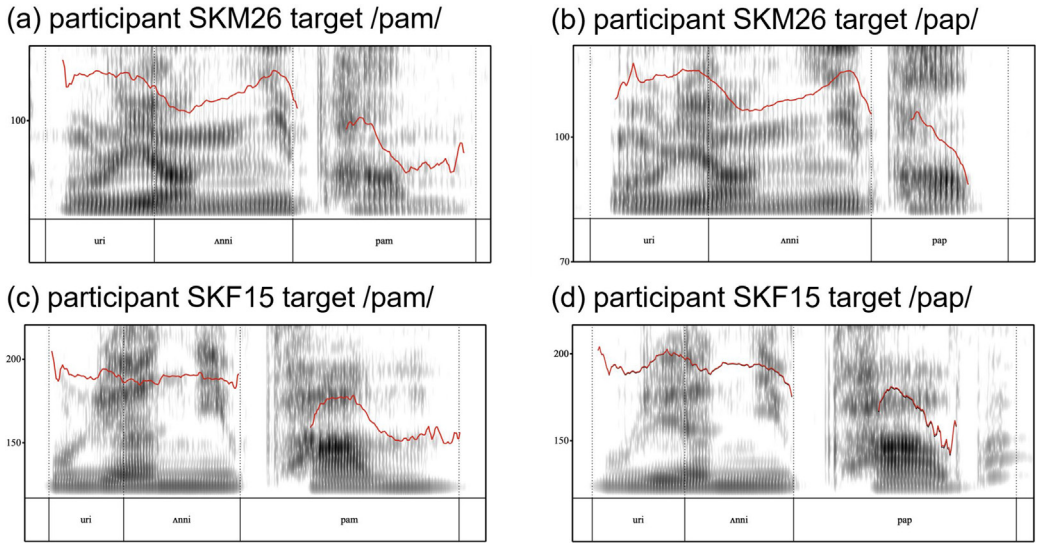

Fig. C1. Tokens of phrase-final focus for speakers SKM26 (a–b) and SKF15 (c–d) with slight falling-rising contours near the onset of the focused word suggesting the possible residue of the LH rise.

References

Arvaniti, A., & Ladd, D. R. (1995). Tonal alignment and the representation of accentual targets. In *Proceedings of the XIIIth International Congress of Phonetic Sciences* (Vol. 4). https://kar.kent.ac.uk/id/eprint/46541.

Arvaniti, A., Ladd, D. R., & Mennen, I. (1998). Stability of tonal alignment: The case of Greek prenuclear accents. *Journal of Phonetics, 26*(1), 3–25. https://doi.org/10.1006/jpho.1997.0063.

Arvaniti, A., Ladd, D. R., & Mennen, I. (2006). Tonal association and tonal alignment: Evidence from Greek polar questions and contrastive statements. *Language and Speech, 49*(4), 421–450. https://doi.org/10.1177/00238309060490040101.

Atterer, M., & Ladd, D. R. (2004). On the phonetics and phonology of "segmental anchoring" of F0: Evidence from German. *Journal of Phonetics, 32*(2), 177–197. https://doi.org/10.1016/S0095-4470(03)00039-1.

Barnes, J., Brugos, A., Veilleux, N., & Hufnagel, S. S. (2014). Segmental influences on the perception of pitch accent scaling in English. In *7th International Conference on Speech Prosody 2014* (pp. 1125–1129). https://doi.org/10.21437/SpeechProsody.2014-214.

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. https://doi.org/10.1016/j.wocn.2017.09.005.

Barnes, J., Mixdorff, H., & Niebuhr, O. (2020). Phonetic variation in tone and intonation systems. In C. Gussenhhoven & A. Chen (Eds.), *The Oxford Handbook of Language Prosody.* https://doi.org/10.1093/oxfordhb/9780198832232.013.9.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001.

Bates, D., Maechler, M., Bolker [aut, B., cre, Walker, S., Christensen, R. H. B., Singmann, H., Dai, B., Scheipl, F., Grothendieck, G., Green, P., Fox, J., Bauer, A., & simulate.formula), P. N. K. (shared copyright on. (2021). *lme4: Linear Mixed-Effects Models using "Eigen" and S4* (1.1-27.1) [Computer software]. https://CRAN.R-project.org/package=lme4.

Beckman, M. E. (1996). The parsing of prosody. *Language and Cognitive Processes, 11*(1–2), 17–68. https://doi.org/10.1080/016909696387213.

Bishop, J. (2013). Prenuclear Accentuation in English—Phonetics, Phonology and Information Structure [PhD Thesis].

Beckman, M. E., & Pierrehumbert, J. B. (1986). Intonational structure in Japanese and English. *Phonology, 3*, 255–309. https://doi.org/10.1017/S095267570000066X.

Bishop, J., & Keating, P. (2012). Perception of pitch location within a speaker's range: Fundamental frequency, voice quality and speaker sex. *The Journal of the Acoustical Society of America, 132*(2), 1100–1112. https://doi.org/10.1121/1.4714351.

Boersma, P., & Weenick, D. (2021). *Praat: Doing phonetics by computer* (6.1.40) [Computer software].

Cangemi, F. (2015). *mausmooth. Praat script* [Computer software].

Cangemi, F., & Baumann, S. (2020). Integrating phonetics and phonology in the study of linguistic prominence. *Journal of Phonetics, 81*. https://doi.org/10.1016/j.wocn.2020.100993 100993.

Caspers, J., & van Heuven, V. J. (1993). Perception of low-anchoring versus hĺgh-anchoring of Dutch accent-lending pitch rises. *Working Papers, 41*, 4.

Cho, H. (2010). *A Weighted-Constraint Model of F0 Movements* [PhD Thesis]. Massachusetts Institute of Technology.

Chen, Y., & Gussenhoven, C. (2008). Emphasis and Tonal Implementation in Standard Chinese. *Journal of Phonetics, 36*(4), 724–746. https://doi.org/10.1016/j.wocn.2008.06.003.

Cho, H. (2011). The timing of phrase-initial tones in Seoul Korean: A weighted-constraint model. *Phonology, 28*(3), 293–330.

Cho, T. (2015). Language effects on timing at the segmental and suprasegmental levels. In *The Handbook of Speech Production* (pp. 505–529). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118584156.ch22.

Cho, T. (2016). Prosodic boundary strengthening in the phonetics-prosody interface. *Language and Linguistics Compass, 10*(3), 120–141. https://doi.org/10.1111/lnc3.12178.

Cho, T. (2021). The Phonetics-Prosody Interface and Prosodic Strengthening in Korean. In *The Cambridge Handbook of Korean Linguistics* (pp. 248–293).

Cho, T. (2022). Linguistic Functions of Prosody and Its Phonetic Encoding with Special Reference to Korean. In *Japanese/Korean Linguistics* (Vol. 29, pp. 1–24).

Cho, T., & Keating, P. A. (2001). Articulatory and acoustic studies on domain-initial strengthening in Korean. *Journal of Phonetics, 29*(2), 155–190. https://doi.org/10.1006/jpho.2001.0131.

Cho, T., Kim, D. J., & Kim, S. (2019). Prosodic strengthening in reference to the lexical pitch accent system in South Kyungsang Korean. *The Linguistic Review, 36*(1), 85–115. https://doi.org/10.1515/tlr-2018-2008.

Cho, T., & Ladefoged, P. (1999). Variation and universals in VOT: Evidence from 18 languages. *Journal of Phonetics, 27*(2), 207–229. https://doi.org/10.1006/jpho.1999.0094.

Cho, T., Yoon, Y., & Kim, S. (2014). Effects of prosodic boundary and syllable structure on the temporal realization of CV gestures. *Journal of Phonetics, 44*, 96–109. https://doi.org/10.1016/j.wocn.2014.02.007.

Choi, J., Kim, S., & Cho, T. (2020). An apparent-time study of an ongoing sound change in Seoul Korean: A prosodic account. *PLOS ONE, 15*(10). https://doi.org/10.1371/journal.pone.0240682 e0240682.

Chung, S.-J., & Kenstowicz, M. (1997). Focus Expression in Seoul Korean. *Harvard Studies in Korean Linguistics*, *VII*, 93–105.

de Jong, K. (1995). The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation. *Journal of the Acoustical Society of America, 97*(1), 491–504.

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. https://doi.org/10.1006/jpho.2001.0151.

Féry, C., & Kügler, F. (2008). Pitch accent scaling on given, new and focused constituents in German. *Journal of Phonetics, 36*(4), 680–703. https://doi.org/10.1016/j.wocn.2008.05.001.

Flemming, E. S. "Auditory Representations in Phonology." PhD Thesis, UCLA, 1995.

Grice, M., Ritter, S., Niemann, H., & Roettger, T. B. (2017). Integrating the discreteness and continuity of intonational categories. *Journal of Phonetics, 64*, 90–107. https://doi.org/10.1016/j.wocn.2017.03.003.

Gussenhoven, C. (2004). *The phonology of tone and intonation*. Cambridge University Press.

Gussenhoven, C. (2007). Types of Focus in English. In C. Lee, M. Gordon, & D. Büring (Eds.), *Topic and Focus: Cross-Linguistic Perspectives on Meaning and Intonation* (pp. 83–100). Netherlands: Springer. https://doi.org/10.1007/978-1-4020-4796-1_5.

House, D. (1990). *Tonal perception in speech*. Lund Univ. Press ; Chartwell-Bratt.

Jeon, H.-S., & Nolan, F. (2017). Prosodic Marking of Narrow Focus in Seoul Korean. *Laboratory Phonology, 8*(1), 2. https://doi.org/10.5334/labphon.48.

Joo, H. (2021). *Tonal alignment with articulatory gestures in South Kyungsang Korean* [M.A., Hanyang University]. https://repository.hanyang.ac.kr/handle/20.500.11754/159687.

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

Jun, S.-A. (1995). Asymmetrical Prosodic Effects on the Laryngeal Gesture in Korean. In B. Connell & A. Arvaniti (Eds.), *Phonology and phonetic evidence* (pp. 235–253). Cambridge University Press.

Jun, S.-A. (1996). *The phonetics and phonology of Korean prosody: Intonational phonology and prosodic structure*. Garland Pub.

Jun, S.-A. (1998). The Accentual Phrase in the Korean prosodic hierarchy. *Phonology, 15*(2), 189–226. https://doi.org/10.1017/S0952675798003571.

Jun, S. -A. (2003). The effect of phrase length and speech rate on prosodic phrasing. In *Proceedings of the XVth International Congress of Phonetic Sciences* (pp. 483–486).

Jun, S. -A. (2005). Korean Intonational Phonology and Prosodic Transcription. In *Prosodic Typology: The Phonology of Intonation and Phrasing*. Oxford University Press.

Jun, S.-A. (2007). The intermediate phrase in Korean: Evidence from Sentence Processing. In T. Riad & C. Gussenhoven (Eds.), *Tones and Tunes: Experimental Studies in Word and Sentence Prosody* (Vol. 2). De Gruyter Mouton. https://www.degruyter.com/document/doi/10.1515/9783110207576/html.

Jun, S.-A. (2011). Prosodic markings of complex NP focus, syntax, and the pre-/post-focus string. In *West Coast Conference on Formal Linguistics (WCCFL)* (pp. 214–230).

Jun, S. -A. (2014a). Prosodic typology: By prominence type, word prosody, and macro-rhythm. In S.-A. Jun (Ed.), *Prosodic Typology II: The Phonology of Intonation and Phrasing* (pp. 520–539). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199567300.001.0001.

Jun, S.-A. (Ed.). (2014b). *Prosodic Typology II: The Phonology of Intonation and Phrasing*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199567300.001.0001.

Jun, S.-A., Kim, H.-S., Lee, H.-J., & Kim, J.-B. (2006). An experimental study on the effect of argument structure on VP focus. *Korean Linguistics, 13*(1), 89–113.

Jun, S. -A., & Lee, H. -J. (1998). Phonetic and phonological markers of contrastive focus in Korean. In *5th International Conference on Spoken Language Processing (ICSLP 1998)*, paper 1087-0. https://doi.org/10.21437/ICSLP.1998-151.

Kawahara, H., Cheveigné, A. de, Banno, H., Takahashi, T., & Irino, T. (2005). Nearly defect-free F0 trajectory extraction for expressive speech modifications based on STRAIGHT. In *Ninth European Conference on Speech Communication and Technology*.

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. https://doi.org/10.1016/j.wocn.2014.03.005.

Kim, H.-S., Jun, S.-A., Lee, H.-J., & Kim, J.-B. (2006). *Argument Structure and Focus Projection in Korean*. Dresden, Germany: Speech Prosody.

Kim, J., & Jun, S. -A. (2009). Prosodic structure and focus prosody of South Kyungsang Korean. *Eohak Yeongu (Language Research)*.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*, 1–26. https://doi.org/10.18637/jss.v082.i13.

Ladd, D. R. (2006). Segmental anchoring of pitch movements: Autosegmental association or gestural coordination? *Italian Journal of Linguistics, 18*(1), 19–38.

Ladd, D. R. (2008). *Intonational phonology* ((2nd ed.)). Cambridge University Press.

Ladd, D. R., Faulkner, D., Faulkner, H., & Schepman, A. (1999). Constant "segmental anchoring" of F0 movements under changes in speech rate. *The Journal of the Acoustical Society of America, 106*(3), 1543–1554. https://doi.org/10.1121/1.427151.

Ladd, D. R., Mennen, I., & Schepman, A. (2000). Phonological conditioning of peak alignment in rising pitch accents in Dutch. *The Journal of the Acoustical Society of America, 107*(5), 2685–2696. https://doi.org/10.1121/1.428654.

Lee, Y.-C. (2015). *Prosodic Focus Within and Across Languages [PhD Thesis]*. University of Pennsylvania.

Lee, Y.-C., Kim, D., & Cho, S. (2019). The Effect of Prosodic Focus Varies by Phrasal Tones: The Case of South Kyungsang Korean. *Linguistics Vanguard, 5*(1). https://doi.org/10.1515/lingvan-2019-0010.

Lee, Y.-C., Wang, T., & Liberman, M. (2016). Production and perception of tone 3 focus in Mandarin Chinese. *Frontiers in Psychology, 7*. https://doi.org/10.3389/fpsyg.2016.01058.

Liberman, M., & Pierrehumbert, J. B. (1984). Intonational Invariance under Changes in Pitch Range and Length. In M. Halle, M. Aronoff, & R. T. Oehrle (Eds.), *Language sound structure: Studies in phonology* (pp. 157–233). MIT Press.

Lindblom, B. (Ed.). (1990). *Explaining Phonetic Variation: A Sketch of the H&H Theory. In Hardcastle, W. & Marchal (Eds.), A. Speech Production and Speech Modelling*, (pp. 403–439). Dordrecht: Kluwer.

McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., & Sonderegger, M. (2017). Montreal forced aligner: Trainable text-speech alignment using kaldi. *Interspeech, 2017*, 498–502 10.21437/Interspeech.2017-1386.

McAuliffe, M., & Sonderegger, M. (2022). *Korean MFA acoustic model v2.0.0a*.

Mücke, D., & Grice, M. (2014). The effect of focus marking on supralaryngeal articulation – Is it mediated by accentuation? *Journal of Phonetics, 44*, 47–61. https://doi.org/10.1016/j.wocn.2014.02.003.

Nooteboom, S. (1997). The prosody of speech: Melody and rhythm. In W. J. Hardcastle & J. Laver (Eds.), *The Handbook of Phonetic Sciences* (pp. 640–673). Blackwell Publishing Ltd.. https://doi.org/10.1111/b.9780631214786.1999.x.

Pierrehumbert, J. B. (1980). *The phonology and phonetics of English intonation* [PhD Thesis]. Massachusetts Institute of Technology.

R Core Team. (2022). *R: A Language and Environment for Statistical Computing* (4.1.2) [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/.

Roettger, T. B. (2017). Tonal placement in Tashlhiyt: How an intonation system accommodates to adverse phonological environments. *Zenodo*. https://doi.org/10.5281/ZENODO.814472.

Selkirk, E. O. (1984). *Phonology and syntax: The relation between sound and structure*. MIT Press.

Shattuck-Hufnagel, S., & Turk, A. E. (1996). A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research, 25*(2), 193–247.

Shue, Y.-L., Keating, P. A., Vicenik, C., & Yu, K. (2011). Voicesauce: A Program for Voice Analysis. *Proceedings of International Congress of Phonetic Sciences*, 1846–1849.

Terken, J. (1991). Fundamental frequency and perceived prominence of accented syllables. *The Journal of the Acoustical Society of America, 89*(4), 1768–1776. https://doi.org/10.1121/1.401019.

Welby, P. (2006). French intonational structure: Evidence from tonal alignment. *Journal of Phonetics, 34*(3), 343–371. https://doi.org/10.1016/j.wocn.2005.09.001.

Welby, P., & Loevenbruck, H. (2005). *Segmental "Anchorage" and the French late rise* (pp. 2369–2372).

Wieling, M. (2018). Analyzing dynamic phonetic data using generalized additive mixed modeling: A tutorial focusing on articulatory differences between L1 and L2 speakers of English. *Journal of Phonetics, 70*, 86–116. https://doi.org/10.1016/j.wocn.2018.03.002.

Wood, S. (2019). *mgcv: Mixed GAM Computation Vehicle with Automatic Smoothness Estimation version 1.8-38 from CRAN* (1.8-38) [R]. https://rdrr.io/cran/mgcv/.

Xu, Y. (1999). Effects of tone and focus on the formation and alignment of f0 contours. *Journal of Phonetics, 27*(1), 55–105. https://doi.org/10.1006/jpho.1999.0086.

Xu, Y., & Xu, C. X. (2005). Phonetic realization of focus in English declarative intonation. *Journal of Phonetics, 33*(2), 159–197. https://doi.org/10.1016/j.wocn.2004.11.001.