Chapter 9

The Phonetics-Prosody Interface and Prosodic Strengthening in Korean

Taehong Cho

9.1 Introduction

In speech communication, in order for the speaker to deliver a linguistic message to the listener successfully, the speaker must be able to encode the linguistic message in phonetic forms (phonetic encoding) in a way that allows the listener to decode it as intended by the speaker (phonetic decoding). One of the essential linguistic structural components that underlie the phonetic encoding-decoding process is prosodic structure (e.g., Keating 2006). It serves as a frame for articulation, by specifying (1) how phonological constituents (e.g., phonemes, syllables, and words) are to be grouped to form phrases that constitute a spoken utterance (boundary marking), and (2) which of the phonological constituents are to be produced with heightened phonetic saliency or prominence relative to the other constituents (prominence marking) (e.g., Beckman 1996; Shattuck-Hufnagel and Turk 1996). To build up a particular prosodic structure of a spoken utterance, the speaker must be able to control phonetic implementation of both segmental and suprasegmental features to signal landmarks important to prosodic structure, such as prosodic junctures and prominent (stressed or accented) units. Thus, the fine phonetic detail of the spoken utterance is fine-tuned to reflect the prosodic structure, which is often described as the phonetics-prosody interface. The phonetics-prosody interface has been examined in particular reference to strengthening of phonetic features which contributes to marking boundaries and prominence, the two important functions of prosodic structure.

In this chapter, I will first discuss some general theoretical considerations regarding prosodic structure as an autonomous grammatical entity (Section 9.2) and the phonetics-prosody interface in conjunction with prosodic strengthening (Section 9.3). I will then review various aspects of the phonetics-prosody interface in speech production in Korean with
special reference to prosodic strengthening (Section 9.4) and phonological processes (Section 9.5). Finally, I will briefly touch on the role of the phonetics-prosody interface in speech perception (Section 9.6).

9.2 Plurality and Autonomy of Prosodic Structure

Prosodic structure is assumed to be created online in a speech planning process (e.g., Keating and Shattuck-Hufnagel 2002; cf. Krivokapić 2012) that takes into account various aspects of linguistic components such as phonology, syntax, semantics, information structure, pragmatics, and discourse (e.g., Jun 1993). The speaker then produces an utterance according to a planned prosodic structure, by modifying not only suprasegmental features such as pitch, duration, and amplitude, but also segmental realization, since the prosodic structure is signaled by an array of phonetic cues at both the suprasegmental and segmental levels. This process is known as prosodic encoding or phonetic encoding of prosodic structure (e.g., Keating 2006; see Cho 2016 for a review). The examples in (1) highlight the function of prosodic encoding involving prosodic structuring or phrasing in Korean.

(1) a. 공사가다방하다
   koy.sa.ka.ta.ma.ya ha.ta
b. (koy.sa.ka) # (ta.ma.ya ha.ta)
   (“public and private matters”-NOM) (“to be busy”)
   “(someone) is very busy with various public and private matters”
c. (koy.sa.ka) # (ta) # (ma.ya ha.ta)
   (“construction-NOM”) (“all”) (“to mess things up”)
   “the construction is all messed up”
d. (koy.sa) # (ka.ta) # (ma.ya ha.ta)
   (“construction site”) (“to go”) (“to mess things up”)
   “(things) are messed up while going to a construction site”
e. (koy) # (sa.ka.ta) # (ma.ya ha.ta)
   (“ball”) (“to buy and go”) (“to mess things up”)
   “(things) are messed up while (someone) is going somewhere after buying a ball”

(Note: “( )” refers to a prosodic phrase, “#” a prosodic boundary. For example, if “( )” refers to an accentual phrase, “#” becomes an accentual phrase boundary.)

Imagine that the string of syllables in (1a) is produced monotonously without any noticeable segmental and suprasegmental change (e.g., in the absence of a pause and durational modification) between the syllables. In this case, prosodic structure is not properly encoded in the speech signal, and, therefore, there is no phonetic clue that allows the listener to understand the linguistic message embedded in the prosodic structure as
intended by the speaker. The utterance with no prosody renders different interpretations of prosodic structure, hence creating multifaceted ambiguity as exemplified in (1b–e). For example, the string of syllables can be parsed with a particular phrasing pattern as in (1b) (koŋ.sa.ka) (ta.maŋ.ha.ta) which is consistent with an idiomatic expression (“being very busy with various public and private matters”). With this particular prosodic phrasing, the last syllable ka of the first phrase (koŋ.sa.ka) functions as a nominative marker, so that koŋ.sa is likely parsed as part of the subject NP (a lexicalized compound of koŋ “public” and sa “private”). Furthermore, to be consistent with the intended idiomatic expression, the speaker also has to form a phrase of (ta.maŋ.ha.ta) – i.e., by grouping the following syllable ta with the remaining syllables to mean ta.maŋ (“to be busy”). In contrast, if the syllable ta forms a single phrase as in (1c) (koŋ.sa.ka) (ta) (maŋ.ha.ta), it is likely interpreted as a monosyllabic adverb (“all”), so that the sentence becomes no longer consistent with the idiomatic expression.

Further possible phrasings are exemplified in (1d–e). The prosodic grouping of (koŋ.sa) (ka.ta) (maŋ.ha.ta) in (1d) indicates that the syllable ka serves as a verb root in (ka.ta) “to go and”, rather than a nominative marker. With this prosodic phrasing, the preceding string (koŋ.sa) is likely parsed as an NP, rendering the sentence to mean “things were screwed up while (someone) was going to the construction site”. Finally, with the prosodic grouping of (koŋ) (sa.ka.ta) (maŋ.ha.ta) in (1e), (koŋ) forming a single phrase is likely be parsed as an object NP “ball”, and (sa.ka.ta) as a verb (“to buy and go”), rendering an entirely different meaning.

Figure 9.1 illustrates two possible prosodic structures (or phrasings), which may correspond to (1b) and (1e) respectively. Following Jun (2000), it is assumed that an Accentual Phrase (AP) is by default assigned tones that
form a pattern of T[H ... L]H, where T refers to either an L tone or an H tone, depending on the laryngeal feature of the AP-initial segment (i.e., the initial syllable receives an H tone with a fortis or aspirated consonant, and an L tone elsewhere). The boundary tone (e.g., H%, L%) may override the final H tone of the AP. The prosodic structure depicted in Figure 9.1a is consistent with (1b) (koŋ,sa,ka) # (ta,maŋ,ha,ta), consisting of two APs in one Intonational Phrase (IP), so that the prosodic boundary “#” refers to an AP boundary. Figure 9.1b depicts a possible prosodic structure to be consistent with (1e) where one IP consists of three APs.

These phrasing possibilities demonstrate how the same string of segments can be parsed into different prosodic structures in conjunction with different lexical and syntactic structural possibilities, indicating that the construction of prosodic structure is conditioned by the morphosyntactic construction (e.g., Nespor and Vogel 1986; Selkirk 1984, 1995). However, the construction of prosodic structure is conceived to be autonomous from the syntactic construction (e.g., Beckman 1996; Shattuck-Hufnagel and Turk 1996; Jun 1998; Keating and Shattuck-Hufnagel 2002; Cho 2016). For example, all of the assumed AP boundaries as denoted by “#” in Figure 9.1a and 9.1b may be replaced by IP boundaries without changing the assumed morphosyntactic structures, leaving the core linguistic meanings intact. More specific examples are given in (2).

(2) 영만이가 음악을 만들었어
   [juŋmanika]NP [[imakil]NP mantilas*s]VP
   “Youngman-NOM” “music-ACC” “to write-PAST”
   “Youngman majors in music”

(3) a. (juŋmanika) # (imakil mantilas*s)
    b. (juŋmanika) # (imakil) # (mantilas*s)
    c. (juŋmanika imakil) # (mantilas*s)
   (where “#” = an AP or an IP boundary)

The sentence in (2) has an NP-VP construction, and the VP has an object NP embedded in it. Given that a major syntactic boundary such as the one between the NP and the VP in (2) is likely to align with a major prosodic boundary (Selkirk 1984), (3a) may represent an optimal, possibly default, prosodic phrasing that reflects the syntactic structure. Crucially, however, the other two phrasings shown in (3b–c) are equally possible, overriding the syntactically conditioned “default” phrasing. As discussed in Shattuck-Hufnagel and Turk (1996), such multiple phrasings of the same syntactic structure indicate that although the syntactic structure may influence prosodic phrasing, the prosodic structure of a given utterance can be independent from the syntactic construction. The plurality of prosodic phrasings, therefore, leads to a theoretical assumption that the prosodic structure is a grammatical component of linguistic structure and is parsed in its own right (Beckman 1996).
In this structural view of prosody, the term prosody does not merely refer to low-level prosodic features such as pitch, duration, and amplitude, but to the abstract grammatical structure which governs the phonological organization of segments into the higher-order prosodic constituents and the distribution of prominence within each constituent. Thus, prosody is definable as embracing “phonetic and phonological aspects of prosody, including both the higher level organization, with its constituent boundaries and prominence markers, and the phonetic reflexes of this organization in the pattern of F0, duration, amplitude and segment quality/reduction within an utterance” (Shattuck-Hufnagel and Turk 1996: 196).

### 9.3 The Phonetics-Prosody Interface

As has been discussed thus far, the premise underlying the structural view of prosody is that prosody is an integral part of speech with linguistic functions essential to speech production and perception – i.e., delimitative and culminating functions (boundary marking and prominence marking, respectively). As a corollary, phonetic realization of speech segments is fine-tuned, engendering systematic phonetic variation at both the segmental and the suprasegmental levels in accordance with the prosodic structure in which they occur. A broad consensus among researchers who have adopted this view of prosody (see Fletcher 2010 or Cho 2016 for a review) is that a description of any particular sound pattern or the sound system of a language can never be completed without understanding its interaction with the higher-order prosodic structure of the language. This view has led to a proliferation of research under the rubric of the phonetics-prosody interface which may be defined as “the interaction of sounds and sound systems with prosodic structure in the grammatical system of the language” (Cho 2016: 123). (Note, however, that the phonetics-prosody interface as used in this context does not imply that there is a clear-cut boundary between phonetics and prosody.)

The overarching question that has underlain the research on the phonetics-prosody interface is how the speaker signals or encodes prosodic structure in the fine-grained phonetic detail. As for the delimitative function of prosody, prosodic constituents are generally demarcated by intonational and durational features. For example, in Figure 9.1 the final syllable of an IP is assigned a boundary tone as marked by L% and undergoes substantial lengthening (known as phrase-final lengthening or preboundary lengthening). It is important to clarify that the phrasings depicted in Figure 9.1 display phonological representations, so that it specifies only a (categorical) boundary tone as stipulated by the intonational phonology of the language. The preboundary lengthening effect, which is a gradient phonetic process by nature, is therefore not specified. Whether phonetic or phonological, the modification of both the temporal and tonal
realization at the end of the IP is an important attribute of the phonetics-prosody interface, which helps organize smaller prosodic constituents into larger ones, marking a major prosodic boundary in the prosodic structure. Unlike the right edge of the IP which is marked by robust temporal and tonal features, however, the right edge of the AP (a prosodic phrase smaller than an IP) in Korean is marked primarily by the tonal feature – i.e., the final H tone of the TH...LH tonal sequence. Jun (1995a) suggested that the right edge of the AP does not undergo final lengthening, nor does it involve an increase in amplitude (Jun 1995b), although it may be subject to a small but significant temporal expansion (Cho and Keating 2001). (See below for further discussion on tones and preboundary lengthening in Korean.)

The boundary-related tonal and/or temporal modification of segmental realization discussed above pertains primarily to the right edge of a prosodic constituent. The past few decades, however, have witnessed a further development of the phonetics-prosody interface which points to particular patterns of phonetic realization on the other side of the prosodic constituent – i.e., the left edge, which is also assumed to participate in demarcating the prosodic constituents. For example, Keating et al. (2003) examined articulatory data, obtained with an electropalatography (EPG) system, of initial consonants in four languages (English, Korean, French, and Taiwanese). The cross-linguistic data showed that initial lingual consonants (e.g., /n, t/) were produced with a more extended linguopalatal contact in the initial position of a larger prosodic domain (constituent) (e.g., IP-initial position) than of a smaller one (e.g., word-initial position), although the languages differed in terms of the detailed phonetic mapping between the degree of the linguopalatal contact and the level of the prosodic domain. They referred to this articulatory phenomenon as the domain-initial strengthening (DIS) which contributes to the demarcation of prosodic constituents, hence participating in the delimitative function of prosody (e.g., Fougeron and Keating 1997; Fougeron 2001; Cho and Keating 2001, 2009). Note that prosodic constituents are often referred to as prosodic domains as they may serve as domains of tonal distribution and of the application of phonological rules (cf. Selkirk 1984, 1995; Jun 1998). I will return to this point regarding the prosodic structural conditioning of phonological rule application later in Section 9.4.2.

The phonetics-prosody interface is also concerned with prominence marking which serves the culminative function of prosody. Let us first consider the relative prominence of phonological constituents in prosodic structure in English. English employs lexical stress which interacts with a phrase-level prominence in conjunction with a pitch accent within an Intermediate Phrase. The lexically stressed syllable that receives the postlexical pitch accent becomes the head of the Intermediate Phrase in English (Beckman 1996). English is, therefore, often classified as a head-prominence language (Jun 2014). The term pitch accent used for English
highlights the fact that the phrase-level accentuation is associated with a salient pitch movement as stipulated by the intonational phonology of English which assigns a starred tone (e.g., H* or L+H*) to the syllable that receives a pitch accent (Pierrehumbert and Beckman 1988; Ladd 2008). (The pitch accent in English differs from the pitch accent system in Japanese which is largely lexically determined – see Kubozono 2018). But the tonal feature is not the only phonetic reflex of pitch accent in English. The pitch-accented syllable is also accompanied by a spatio-temporal expansion of segmental realization, so that, for example, vowels under accent become more peripheral in the vowel space (both acoustically and articulatorily) and longer than those that do not receive the phrase-level accent (de Jong 1995; Cho 2005, 2008). (See also Baumann and Winter 2018, for the effects of prosodic vs. non-prosodic factors on prominence perception in German.)

The prominence-marking system in Korean is fundamentally different from that in English. Korean does not employ word-level prominence (i.e., lexical stress), and thus relative prominence in Korean does not interact with the word-level prosody. Instead, the distribution of prominence in Korean is generally assumed to be determined primarily by the post-lexically assigned tonal pattern on an AP, as shown in Figure 9.1. The AP-initial T and the AP-final H are in principle associated with the phrase edges – the initial and the final syllables, respectively. Jun (2014) suggests that the edge-based tonal distribution within an AP characterizes prominence marking in Korean, which classifies Korean as an edge-prominence language. Edge-prominence languages are taken to employ a so-called macro-rhythm because the rhythm of the language is determined at the phrase level. (Jun refers to the rhythm in English as micro-rhythm as the rhythm is created by an alternation of smaller units such as syllables or feet within and across words.)

The edge-based prominence system in Korean may have further ramifications for the phonetics-prosody interface in Korean, especially with respect to the degree of domain-initial strengthening. Keating et al. (2003), for example, observed more robust domain-initial strengthening effects in Korean than in English, and attributed the cross-linguistic difference to the fact that Korean marks prosodic structure primarily by phrasing, while English employs both phrasing and prominence marking. It is certain that the authors did not mean to suggest that there is no prominence marking function in prosodic structuring in Korean since a language is assumed to utilize prominence in reference to prosodic structure. But this observation alludes to a possible inseparability between phrasing and prominence marking in Korean. This is precisely because the delimitative function of prosodic boundary marking goes hand in hand with the culminative function of prominence marking in edge-prominence languages like Korean.
At this point, it is noteworthy that Korean is often considered to be similar to French as both languages employ an AP whose right edge is marked by a rising tone (H). But the right-edge marking of AP in Korean differs from that in French. The AP-final syllable in French is generally associated with the final (stressed) syllable of the AP-final word (when the final vowel is not reduced), so that it involves an increase in both duration and amplitude (Jun and Fougeron 2002; cf. Vaissière 1991). This characterizes the AP-final (right-edge) syllable in French as accented, on which basis Jun (2014) classifies French as a head/edge-prominence language. The AP-final syllable in Korean, which is marked primarily by the tonal feature (not accompanied by either a substantial final lengthening or an increase in amplitude), must therefore not be taken to be accented in the same sense as is described in French.

### 9.3.1 Syntagmatic and Paradigmatic Contrast Enhancement

In the literature on the phonetics-prosody interface, the linguistic functions of boundary vs. prominence marking have been discussed in terms of linguistic contrast enhancement – i.e., *syntagmatic and paradigmatic contrast enhancement* (see Fougeron 1999 and Cho 2011, 2016 for a review).

As illustrated in Figure 9.2, the term *syntagmatic* describes the structural relationships between segments that form a sequence in speech. Given that boundary marking is syntagmatically, or structurally, motivated, it is expected to bring about an enhancement of the syntagmatic contrast between neighboring segments at the prosodic junctures. The syntagmatic contrast enhancement at the prosodic juncture often involves enhancements of #CV contrast and V#C contrast. For example, the consonant at the prosodic juncture becomes more consonant-like in the spatio-temporal dimension through an increase in both the constriction degree and the constriction duration. This reduces the consonant’s sonority, resulting in an enhancement of its consonantality. On the other hand, the vowel at the prosodic juncture becomes more vowel-like through an increase in the degree of mouth opening, which enhances its sonority feature (sonority expansion).

![Figure 9.2](image_url)
Researchers have often interpreted the linguistic function of the DIS effect on the consonant as enhancing the syntagmatic contrast in #CV and V#C (see Cho 2016 for a review). Based on DIS effects in English, Fougeron and Keating (1997) suggested that the boundary-induced enhancement of #CV and/or V#C contrast would facilitate the parsing of the continuous incoming speech signal into words and, therefore, the lexical segmentation. Cho, McQueen, and Cox (2007) indeed tested the perceptual role of DIS in lexical segmentation of a two-word sequence (e.g., bus tickets) which involved temporary lexical ambiguity—i.e., bust is a competitor of bus in the bus#tickets sequence. They showed that the acoustic consequence of the DIS effect on the onset of the postboundary word (e.g., tickets) facilitated lexical segmentation via resolving lexical ambiguity. The result suggests that fine-grained phonetic detail of domain-initial strengthening is indeed exploited by listeners in speech comprehension.

The term paradigmatic, as marked by the vertical arrows in Figure 9.2, describes the relationship among linguistic units such as phonemes that can substitute for each other in a given context. The enhancement of paradigmatic contrast often refers to the maximization of phonemic distinction of contrastive sounds that may arise with prominence marking. For example, prominence involves a so-called localized hyperarticulation (de Jong 1995) which may be defined as a spatio-temporal expansion of segmental realization that results in an enhancement of phonemic contrast of segments under prominence. The term localized is used to indicate that the effect is localized to a stressed syllable (in English) that receives a phrase-level accent, so that it is differentiated from the communicatively driven hyperarticulation which is extended globally to the entire utterance in the theory of Hyper- and Hypo-articulation (H&H theory, Lindblom 1963).

The prominence-induced localized hyperarticulation involves an enhancement of distinctive features. For example, English /ʊ/ under accent is produced with a more retracted tongue, enhancing [+back] (de Jong 1995); and English /i/ under accent is produced with a more advanced tongue body along with a higher F2, enhancing [−back] (Cho 2006). Studies on the consonants in English also suggest that prominence may induce enhancement of language-specific phonetic features. For example, S. Kim, J. Kim, and Cho (2018) showed that the phonological voicing contrast between the voiceless and voiced stops in English is polarized under prominence along the VOT dimension, but the phonological contrast enhancement is primarily due to the voiceless stops being produced with longer VOT values, i.e., in a direction that enhances their voicelessness. The voiced stops, on the other hand, were not found to be produced with more negative VOT values, which, therefore, does not necessarily heighten their phonetic voicing. The results suggest that it is not the phonological features ([+−voice]) but the phonetic features ([+spread glottis] and [−spread glottis]) that undergo the prominence-related strengthening. (Following Keating 1984, the curly brackets refer
to phonetic features.) It is also worth mentioning that a phonologically identical voiceless stop in Dutch is produced with shorter VOT when the stop is accented than when it is unaccented, consistent with enhancement of the language-specific phonetic feature \{−spread glottis\} rather than the coarsely defined phonological feature \{−voice\} (Cho and McQueen 2005).

9.3.2 Prosodic Strengthening
From the phonetic point of view, both types of contrast enhancement patterns, whether syntagmatic or paradigmatic, involve some kind of phonetic strengthening effects that increase phonetic clarity in the spatio-temporal dimension in prosodically strong contexts. The term prosodic strengthening has been used in the literature as a cover term to refer to the strengthening of segmental realization in association with the delimitative and culminative functions. As discussed above, there is a relatively clear dichotomy in linguistic function between the two kinds of prosodic strengthening in English. The dichotomy may be seen as coming from the fact that English is a stressed-time, head-prominence language in which lexical stress is integrated into a higher-order prominence system, which dissociates prominence marking from boundary marking (e.g., Keating et al. 2003; Cho and Keating 2009; Cho 2011; Cho, Lee, and Kim 2011; Cho, Kim, and Kim 2017; Jun 2014). Again, as discussed above, the dichotomy becomes fuzzy when we consider an edge-prominence language like Korean whose prominence system is rather loosely defined: Korean does not specify lexically defined stressed syllables so that higher-order prominence is not superimposed on a lexically specified location.

The special status of phrasing in Korean has, therefore, drawn the particular attention of many researchers to edge phenomena in Korean, and sparked numerous phonetic studies on the phonetics-prosody interface with special reference to phonetic events that occur at phrase edges. In the remainder of this chapter, I will review previous studies as well as some ongoing studies on the phonetics-prosody interface in Korean primarily from the perspective of prosodic encoding in speech production. I will then touch on some issues of prosodic decoding in speech perception with implications for linking prosodic encoding to decoding.

9.4 Prosodic Strengthening in Korean

9.4.1 The Left-Edge Effects: Domain-initial Strengthening
9.4.1.1 DIS Effects on Nasal Consonants
Figure 9.3 illustrates a DIS effect on Korean /n/ whose linguopalatal contact, reflected in the EPG data in Cho and Keating (2001), varies
systematically as a function of boundary strength. As marked by an arrow in the figure, the area of the palate (the roof of the mouth) contacted by the tongue becomes progressively larger in proportion to the boundary strength. It is also noticeable that the denti-alveolar contact for /n/, the nominal point of articulation, is fully attained in the IP-initial position, and the articulatory target is undershot in the initial position of smaller prosodic domains. The increase in the linguopalatal contact was found to be accompanied by an increase in the hold (seal) duration of the oral constriction, showing a strong correlation between the spatial and temporal variation of the segmental realization. Cho and Keating (2001) interpreted this DIS effect in terms of the time-dependent articulatory undershoot hypothesis (Lindblom 1963; Moon and Lindblom 1994). It was hypothesized that the consonant is assigned sufficient time in the initial position of a larger prosodic domain for executing the articulatory action, which allows for a full attainment of the assumed articulatory target. On the other hand, the time assigned to the initial consonant of a smaller prosodic domain was assumed to be insufficient to attain the target, resulting in target undershoot.

The target undershoot hypothesis is consistent with fundamental assumptions of the theory of $\pi$-gesture (prosodic gesture) (e.g., Byrd 2000, 2006; Byrd et al. 2000; Byrd and Saltzman 1998, 2003; Byrd, Krivokapić, and Lee 2006). The tenet of this theory is that boundary-related temporal variation comes about as a result of modulation of a so-called $\pi$-gesture which is governed by prosodic constituency. The $\pi$-gesture modulates the rate of the clock that controls the temporal activation of constriction gestures in the vicinity of the prosodic juncture, so that any segments (or articulatory gestures) that are under the influence of the $\pi$-gesture may lengthen (or shorten) largely in proportion to boundary strength. Under the time-dependent articulatory undershoot hypothesis, the domain-initial articulatory strengthening effects on /n/ can be interpreted as stemming from temporal expansion under the influence of the $\pi$-gesture which

![Figure 9.3 Domain-initial articulatory strengthening of Korean /n/ as reflected in the linguopalatal contact (re-drawn based on Figure 3 in Cho and Keating 2001). The filled ovals refer to electrodes in an electropalatography (EPG) that are contacted by the tongue during the occlusion of the nasal stop.](image-url)
allows for a full attainment of the assumed articulatory target. (See Section 9.4.1.2 for further discussion on the π-gesture.)

Another important point to be made regarding the DIS effect on /n/ is that despite the fact that /n/ is a sonorant consonant, the oral constriction of /n/ is strengthened, which increases the degree of its consonantality (i.e., the more constricted, the more consonant-like) and, therefore, reduces the degree of its sonority. The sonority reduction (or the strengthening of the consonantality) was further evident in the reduced duration of the nasal murmur and the weakening of the nasality amplitude, which has been observed not only in Korean as reported in Cho and Keating (2001) (see also Jang, Kim, and Cho 2018), but also in other languages such as English and French (Fourgeron and Keating 1997; Fougeron 2001; Cho and Keating 2009; Cho, D. Kim, and S. Kim 2017). The resulting reduced sonority of /n/ across languages is interpretable as increasing the /n/’s consonantality (i.e., the less sonorant, the more consonant-like), which has the effect of enhancing syntagmatic contrast at the prosodic juncture through CV contrast enhancement.

The reduced nasality of the nasal consonant in domain-initial position may appear to have a further consequence for the sound system of Korean. Korean has been thought to undergo a denasalization process in which a word-initial nasal consonant is produced without any discernible nasality during the oral constriction, even to the extent that non-native listeners would misperceive it as oral (Yoshida 2008; Y. Kim 2011; Jang et al. 2018). The acoustic data reported in Jang et al. (2018) indicate that the degree of denasalization – reflected in the duration of the nasal murmur and its nasal energy (obtained from A1−P0, the difference between the peak amplitude of the first formant and the nasal amplitude peak at a lower spectral frequency) – varies depending on the prosodic position in which the nasal-initial word occurs. Consonantal nasality was found to be substantially more reduced in the IP-initial than in the word-initial position. Furthermore, the degree of vowel nasalization after the nasal consonant (#NV) was found to be more reduced at the beginning of the following vowel compared to that observed in the later part of the vowel. This is the exact opposite of the expected coarticulatory nasalization effect as found in English: more nasality is observed when it is measured at an earlier point in the vowel in #NV than at a later point (Cho, Kim, and Kim 2017).

The prosodic structural conditioning of denasalization in Korean implies that the denasalization has originated from the domain-initial strengthening effect, which enhances consonantality (the opposite of sonority). Moreover, further analyses of the data obtained for Jang et al. (2018) revealed that the process is gradient. As shown in Figure 9.4, although the nasality of /n/ is greatly reduced in IP-initial positions, speakers showed different degrees of denasalization ranging from complete denasalization, to partial denasalization, to no denasalization. Given that Jang et al.’s speakers were relatively young (age range:
twenty-one to twenty-eight; mean 24.4), the denasalization process does not appear to have been phonologized, but is better characterized as a gradient phonetic effect driven by the DIS. A cross-generation study is warranted for understanding how the possible sound change has progressed in the language. I will return to this point in Section 9.4.1 in discussing the prosodic-structural modulation of a coarticulatory process.

9.4.1.2 DIS Effects on Oral Consonants

Cho and Keating (2001) also examined DIS effects on the oral stops /t, tʰ, t*/ (lenis, aspirated, and fortis, respectively; cf. Cho, Jun, and Ladefoged 2002), and found that the oral constriction degree and the closure (seal) duration are also augmented in the initial position of a larger prosodic domain, effects which appear to be more robust than those observed in other languages, as discussed in Keating et al. (2003).

In an acoustic-aerodynamic study, Cho and Jun (2000) further examined how the phonetic realization of the three-way (laryngeal) contrast of stops in Korean is conditioned by the DIS. The results are recast in Figure 9.5. As can be seen in the figure, for the aspirated stop /pʰ/, VOT was longer and the amount of airflow was greater in the initial position of a larger prosodic domain than in that of a smaller one. The fortis stop /p*/ on the other hand, showed the opposite pattern, being produced with a reduction in both VOT and airflow. The lenis stop /p/ was consistently produced with a relatively shorter VOT and a reduced airflow, maintaining its distinction from the aspirated stop across different prosodic positions. But the lenis stop too showed an increase in the magnitude of VOT and airflow in the initial position of a larger prosodic domain, showing a similar DIS effect as the aspirated stop.

Cho and Jun interpreted these DIS effects as a combined effect of both syntagmatic and paradigmatic enhancement induced by boundary marking in Korean. The laryngeal contrast between the aspirated and the fortis

![Figure 9.4](image-url) Some gradient aspects of denasalization in Korean. Tokens were selected from the data collected for the study of Jang et al. (2018): a monosyllabic word /nat/ ("sickle") from IP-initial and IP-medial positions (preceded by a /a/-final word). The downward arrows indicate some acoustic evidence for the degree of denasalization.
stops was maximized at the prosodic juncture, due to an enhancement of \{spread glottis\} for the aspirated stop (as reflected in an increase in VOT and airflow) and an enhancement of \{constricted glottis\} for the fortis (as reflected in a decrease or no change in VOT and airflow). At the same time, the increased VOT/airflow for the aspirated stop was also taken to contribute to enhancing the syntagmatic CV contrast. The DIS effects on the lenis stop were taken to be syntagmatically, rather than paradigmatically, driven under the assumption that the lenis stop is unspecified for any laryngeal feature (Lombardi 1991). (It is noteworthy that the clear-cut distinction between the aspirated and the lenis stop in VOT and airflow indicates that the Seoul speakers who participated in the experiment back in 1999 did not show a merger in VOT, which has been observed among young Seoul Korean speakers – Silva 2006; Kang 2014; Bang et al. 2018).

The assumed change in the glottal opening area (as reflected in VOT and airflow measures) as a function of domain-initial strengthening was more directly observed in a fiberscopic study by Jun, Beckman, and Lee (1998). Results of this study revealed that the aspirated stop is produced with a larger glottal opening area during the closure (and at the time of release) in the AP-initial position than in the AP-medial position, while the reverse is true for the fortis stop. This illuminates the articulatory underpinnings of enhancements of \{spread glottis\} vs. \{constricted glottis\}, in line with the view that prosodic strengthening operates based on language-specific phonetic features (e.g., Cho and McQueen 2005; Cho 2016). Jun et al. (1998) also observed an increase in the glottal opening area for the lenis stop in the AP-initial position. In the case of the lenis stop, which is assumed to be unspecified for any laryngeal feature, one might expect that the default laryngeal opening gesture for a voiceless stop would be strengthened domain-initially. These findings, setting
aside the theoretical issue of the laryngeal feature underspecification for the lenis stop, demonstrate that the three-way phonological contrast is maintained more clearly in the domain-initial position, showing a kind of paradigmatic enhancement.

Seen from a different angle, however, the observed DIS effects on laryngeal features can be seen as increasing a segment’s consonantality in a direction opposite to sonority expansion. The domain-initial increase in VOT for both the aspirated and the lenis stops may be seen as enhancing the consonantality as the longer aspiration makes the consonant more consonant-like (Pierrehumbert and Talkin 1992; Cho 2016). The shortening of VOT for the domain-initial fortis stop is accompanied by a lengthening of the closure duration (Cho and Keating 2001), which also increases the consonantality. These results then imply that DIS also brings about a syntagmatic (CV) contrast enhancement in parallel with paradigmatic enhancement.

In a kinematic study, Cho, Son, and Kim (2016) have provided evidence for the assumed dual function of boundary marking in Korean. They observed kinematic variation of the lip closing (V-to-#C) and lip opening (#C-to-V) gestures for /p, pʰ, p*/ in different prosodic positions. They reported that the consonantal lip closing (V-to-#C) gesture was characterized by temporal variation with a larger, longer, and slower articulation. The temporal expansion of the V-to-#C movement pattern was interpreted to be comparable to boundary effects found in English, which may be best accounted for by a change in a rate of the clock (possibly modulated by the π-gesture) as schematized in Figure 9.6a. Following Byrd and Saltzman (2003), the spatial expansion (the larger articulation) that accompanied the temporal expansion was assumed to be attributable to the possibility that the reduced gestural overlap due to the time-stretching of sequenced gestures would minimize the likelihood of spatial truncation. The vocalic lip opening (#C-to-V) gesture (after the release), however, showed a somewhat different pattern: It was larger (with spatial expansion in displacement) and faster, but not longer. This #C-to-V movement pattern was broadly comparable to the prominence-induced effects reported in English, which were taken to be better accounted for by a change in target which caused a spatial expansion of the #C-to-V gesture as schematized in Figure 9.6b.

Taken together these results regarding both laryngeal and supralaryngeal phonetic reflexes of boundary marking indicate that boundary-related prosodic strengthening in Korean involves both paradigmatic and syntagmatic contrast enhancements, serving a dual function – i.e., delimitative and culminative. They also provide converging evidence for the typological classification of Korean as belonging to the edge-prominence language group (Jun 2014), in which edge-marking and prominence marking may not be separable, hence forming a combined effect.
In the previous section, I have discussed boundary-related prosodic strengthening as a combined prosodic effect in which boundary marking is inseparable from prominence marking in Korean. A question then follows as to whether and how prominence may be independently marked by the linguistic system. One way of understanding the pure prominence effect in Korean may be to examine the possible prominence effect of focus that comes from the information structure of the utterance (Kucerova and Neeleman 2012), and its interaction with boundary-related strengthening effects. It has been suggested that focus realization induces prominence on the word or syllable under focus and, as briefly mentioned above, it may serve as a diagnostic for what specific phonetic content (among various possible phonetic correlates) is used to mark phonological contrast in a given language (de Jong 2004; de Jong and Zawaydeh 2002).

9.4.2 Prominence Effects in Conjunction with Focus in Korean

In the previous section, I have discussed boundary-related prosodic strengthening as a combined prosodic effect in which boundary marking is inseparable from prominence marking in Korean. A question then follows as to whether and how prominence may be independently marked by the linguistic system. One way of understanding the pure prominence effect in Korean may be to examine the possible prominence effect of focus that comes from the information structure of the utterance (Kucerova and Neeleman 2012), and its interaction with boundary-related strengthening effects. It has been suggested that focus realization induces prominence on the word or syllable under focus and, as briefly mentioned above, it may serve as a diagnostic for what specific phonetic content (among various possible phonetic correlates) is used to mark phonological contrast in a given language (de Jong 2004; de Jong and Zawaydeh 2002).

9.4.2.1 Focus-Induced Prominence versus Boundary Effects on Segmental Realization in Korean

Cho, Lee, and Kim (2011) indeed explored the effects of boundary and focus on the acoustic realization (i.e., the VOT of the aspirated stop, and F1/F2 of the following vowel) of a word-initial syllable and the interaction between boundary and focus. (Note that the focus condition in Cho et al. 2011 was called the prominence condition.) As for the effects on the vowel, their results, as recast in Figure 9.7, indicate that both the boundary and the focus factors caused an expansion of the vowel space. This suggests that boundary marking involves strengthening effects on the vowel realization largely comparable to the focus-induced prominence effect, although the exact direction of strengthening and the vowel space expansion may differ for different vowels and depending on the source of strengthening.
These effects, however, differ from those found in English (Cho and Keating 2009), which indicate only a prominence-induced modification of the vowel realization. This again implies that boundary marking in Korean involves strengthening that may engender an enhancement of paradigmatic contrast which is observed only in conjunction with prominence marking in English, although the focus-related effect appears to be more robust than the boundary-related effect in Korean.

As for the effects on the consonant, Figure 9.8 (adapted from Cho et al. 2011) shows the effects of boundary and focus on the VOT of the aspirated stop. As can be seen in Figure 9.8a, VOT was found to be longer in the IP-initial position than in the IP-medial position, and longer in the focused than in the unfocused condition, but the effect size was larger for the focus factor than for the boundary factor. Crucially, a significant interaction was reported between boundary and focus. As can be seen in Figure 9.8b, the boundary-induced VOT difference did not come about in the focused (most prominent) condition and turned out to be significant only when the test word was unfocused. Cho et al. (2011) attributed this focus-dependent boundary effect to a ceiling effect: because the focus-induced strengthening effect was stronger than the boundary-induced strengthening effect, there was no room for further strengthening due to boundary when focus marks prominence. As far as VOT is concerned, it appears that the focus-induced prominence effect enhances the (spread glottis) feature across the board, whereas the boundary marking comes into effect in a less prominent context in which both the syntagmatic and paradigmatic contrasts may otherwise blur.
These results together indicate that both the boundary-related strengthening effect and the focus-induced effect give rise to a kind of hyperarticulation of segments which may maximize the phonological contrast of the segments (i.e., paradigmatic enhancement). The focus factor engenders the more robust strengthening effect of prominence, which may differ from a possible prominence effect that arises from boundary marking. This suggests that focus realization regulates prominence quite independently from boundary marking, even to the extent that it masks the boundary-related strengthening effect. This further implies that focus marking in Korean may not be directly mediated by the boundary marking system. This interpretation is congruous with what Mücke and Grice (2014) suggested with respect to focus marking in German: focus marking in German is not mediated by a phonological accent system, but it may directly regulate the phonetic realization.

9.4.2.2 The Rise of Fundamental Frequency and the Fall of Voice Onset Time: A Prosodic Structural Account of Sound Change

The phonological contrast between the lenis and aspirated stops has traditionally been considered to be differentiated primarily by VOT in tandem with F0, along with some additional voice quality difference (e.g., the vowel after a lenis stop is breathier than that after an aspirated stop) (Cho, Jun, and Ladefoged 2002). In the past several decades, however, the cue primacy relationship between VOT and F0 has begun to change, so that the VOT difference has been substantially diminishing while the F0 difference is considered to have become a primary cue to the lenis-aspirated contrast, especially among young speakers. Given the primacy of the F0 cue, the change has been discussed in terms of an on-going tonogenetic sound change (Silva 2006; Kang 2014; Bang et al. 2018), which is assumed
to have been led by young female speakers (Oh 2011). (For general issues of the relationship between VOT and F0, see Kirby 2018; Ladd and Schmid 2018; Coetzee et al. 2018). The rise of F0 as a primary cue to the laryngeal contrast, however, is conditioned by prosodic structure, being limited to the onset of a word that is aligned with the left edge of a larger prosodic constituent – i.e., in domain-initial (either AP-initial or IP-initial) positions.

Importantly, the use of F0 in the domain-initial position has its origin in the intonational phonology of Seoul Korean (Jun 1993, 2000). As mentioned above, the intonational structure of an AP is TH...LH, and an H tone is assigned to the AP-initial syllable when its onset is aspirated, but an L tone when it is lenis. The segmentally conditioned tonal assignment indicates that the low-level segmental pitch perturbation (a higher F0, localized to the beginning of the vowel, when the onset consonant is fortis/aspirated than when it is lenis) was phonologized as a tonal contrast between L and H in the intonational grammar of the language. Jun (1996) describes this as the influence of microprosody (i.e., a segmental pitch perturbation) on the macroprosody (i.e., the higher-level intonational structure), which may be seen as a case of domain-initial strengthening. The macroprosody here refers to the tonal contrast as a post-lexical contrast, as it does not directly pertain to the segmental (phonological) contrast that gives rise to a lexical contrast. On the contrary, the purported tonogenetic sound change as described above implies that the segmental pitch perturbation has been phonologized at the level of segmental phonology, which engenders the use of tones for the segmental contrast between the lenis and aspirated stops that gives rise to a lexical contrast.

But given that the segmentally conditioned H vs. L tones are already used post-lexically in the intonational grammar of the language, the notion of tonogenesis, if one prefers to use the term, applies only to the level of macroprosody. In other words, it may not be a case of so-called transphonologization at the level of segmental phonology (Hyman 2008), as no tones have been newly introduced into the sound system in Korean – at least in its current state. This differs from a typical case of tonogenesis in which the pitch perturbation that may occur redundantly at the phonetic level becomes phonologized and, therefore, is used contrastively at the phonemic level as in Kammu and Cham (see Kingston 2011 for a review of various cases of tonogenesis).

Given the use of tones already available in Korean, I propose that the ongoing sound change is not tonogenetic, in a strict sense, but is better characterized as a change in the way that the existing tones are used, through an interaction between the segmental phonology and the intonational phonology in Korean. The tonal specifications in AP-initial (and IP-initial) positions are conditioned by the laryngeal contrast between the lenis and aspirated stops and, therefore, the tones assigned post-lexically are indexed with the phonological contrast. As a result, the phonological contrast between the lenis and aspirated stops is cued by fixed
intonational properties so that VOT, which once played a primary role, has become redundant in signaling the phonological contrast. The propensity to minimize production effort (or the principle of effort minimization) (cf. Flemming 1995; Boersma 1998) could then have led to diminishment of the use of redundant VOT, giving primacy to the invariently available post-lexical tones, which provide sufficient information flow for the phonological contrast. This illustrates a unique case in Korean. The segmentally conditioned tonal contrast at the level of intonational phonology, which originated from the segments’ low-level pitch perturbation, is now re-used as a primary cue to the segmental contrast. Crucially, again, the use of the tonal cues is position-specific, being limited to domain-initial positions in reference to the prosodic structure of a given utterance.

In sum, under the prosodic-structural account, the use of tones is not a tonogenetic process, but a cue reduction process driven by the principle of effort minimization. In this view, L and H tones are not taken to be the primary phonetic features of the phonological contrast, but rather they are post-lexical features from the intonational phonology. In contrast, the tonogenetic sound change account assumes that the L and H tones have become primary features for the lenis and aspirated stops, respectively. One way of testing these possibilities may be to examine how tonal realization is modified by focus which, as discussed above, may serve as a metric for assessing what phonetic content may underlie the phonological contrast. If the L and H tones are primary distinctive features, focus may enhance the contrast between L and H in such a way that they are polarized along the phonetic F0 dimension: under focus, the L tone becomes lower and the H tone becomes higher. Standard Mandarin Chinese which employs lexical tones shows such a polarization effect. Chen and Gussenhoven (2008) reported that in a focused (emphatic) context, a high tone becomes higher and a low tone lower, in comparison to when the tones are realized in unfocused (non-emphatic) contexts. For the remainder of this section, I will discuss the nature of the tonal realization under focus (i.e., tonal polarization effects) in Seoul Korean in comparison with that in Kyungsang Korean, which employs a lexical pitch accent system, with a view to understanding the phonological status of the tones in these languages.

9.4.2.2.1 The Case of South Kyungsang Korean
In an acoustic study, Cho, D. Kim, and S. Kim (2019) tested the tonal polarization effect under focus in South Kyungsang Korean, a lexical pitch accent language in which tones are assumed to be lexically specified, regulating the phonological contrast of otherwise segmentally identical strings (e.g., /are/ LH “the day before yesterday” vs. /are/ HL “the bottom”). The authors specifically explored a theoretical debate regarding the nature of the Low tone: whether the Low tone is an underlying tone specified in the lexicon or is assigned later (post-lexically) by the initial boundary Low
tone of AP at the level of intonational phonology (Kim and Jun 2009). (Note that Kim and Jun 2009 suggested that the initial H tone is a lexical property but the L tone is a post-lexical one.) They hypothesized that if the Low tone is part of the underlying tonal representation, it will be phonetically enhanced to get even lower under focus, showing a bidirectional polarization (the H tone becomes higher, and the L tone becomes lower). They further assumed that there would be no a priori reason to predict such a focus-induced lowering effect, if the tone is a post-lexical property. Their results indeed showed that contrastive focus on the tonal difference induced a bidirectional phonetic polarization of the Low-High tonal contrast of the word-initial vowel (\texttt{ar} LH vs. \texttt{ar} HL). The H tone became even higher and the L tone became even lower when the tone-bearing word was focused rather than unfocused, which supports the view that the L tone is part of the underlying phonological representation of the word. The L tone appears to be actively controlled by the speaker to augment the lowering of the L tone to maximize the phonological tone contrast in the phonetic dimension.

9.4.2.2.2 The Case of Seoul Korean

Choi, Kim, and Cho (in preparation) (see also Choi et al. 2018) further hypothesized that if the L and H tones in Seoul Korean had become the primary phonetic features rather than post-lexically assigned intonational properties (i.e., the tonogenetic sound change account), they would show a similar bidirectional polarization effect under focus as has been observed in Mandarin Chinese and South Kyungsang Korean. As shown in Figure 9.9, however, the results revealed no lowering effect on the L tone under focus, in either the younger or older groups, even though focus was on the assumed H-L tonal contrast in minimal pairs such as /\texttt{th}anto/ (“trajectory”) vs. /\texttt{t}anto/ (“short sword”). In fact, the L tone showed a small but significant increase in F0 in the focused condition, showing the opposite of the focus-induced phonetic enhancement of the L tone found in Mandarin Chinese (a tone language) and South Kyungsang Korean (a lexical pitch accent language), in both of which a tone directly affects phonological/lexical

![Figure 9.9](image-url)
contrast. Thus, the focus-induced raising of F0 for the L tone in Seoul Korean may be interpreted as being more consistent with the use of post-lexical tonal properties conditioned by prosodic structuring, rather than as the emergence of tonal features for phonological contrast at the level of segmental phonology. It appears that unlike the case of the lexically specified L tone in South Kyungsang Korean, the post-lexical L tone under focus is not actively controlled by the speaker. Instead, the raising of the L tone under focus may be seen as a result of heightened articulatory force applied to articulation (including the laryngeal articulation) which may contribute tension to the laryngeal muscles, effectively raising the F0.

It is, however, worth noting that the focus-induced increase in F0 for the lenis stop was smaller in the younger speaker group than in the older speaker group. This may be a reflex of sound change in progress, in the direction of using the tones as part of the phonological representation. If so, at some point in the future, the phonological contrast may eventually change to resemble those that utilize tones as distinctive phonological features, showing a bidirectional polarization pattern.

9.4.3 Domain-Final (Right Edge) Effects in Korean
Thus far I have reviewed studies on the phonetic reflexes of boundary marking at the left edge of prosodic constituents, some of which may relate to prominence marking. Boundary marking, of course, also appears on the other side of the prosodic juncture, the right edge of a prosodic constituent. In fact, substantial phonetic reflexes of prosodic boundaries have been found before the prosodic juncture across languages, since phonetic reflexes should be able to signal the end of the prosodic constituent (usually a word) that is being parsed into a larger prosodic constituent such as an AP or IP.

9.4.3.1 Preboundary Tones in Korean
As introduced earlier in this chapter, theories of the phonetics-prosody interface assume that prosodic structure supplements information about boundary strength with tonal marking, especially for higher prosodic boundaries – IP and AP boundaries – such that the AP-final tones and boundary tones are phonologically associated with the right edge of the AP and the IP, respectively, signaling upcoming prosodic junctures of a different boundary strength.

As explained above, an AP in Seoul Korean is by default assigned an underlying tonal structure of #TH...LH# (where # = AP boundary). The tonal distribution within an AP may be further constrained by the structure of the AP-initial syllable (open vs. closed) and the vowel quantity (long vs. short) (Lim and de Jong 1999; Park 2004) (but note that vowel quantity is no longer maintained in contemporary Seoul Korean). S. Kim (2004) transcribed Korean intonation patterns in read speech and radio drama, and
reported that about 88% of AP-initial syllables of multisyllabic content words start with an initial rising (#LH) tone (when the initial consonant is neither an aspirated nor a fortis (tense) consonant), and AP-final syllables end with a rising (LH#) tone for about 85% of multisyllabic content words, consistent with the assumed underlying tonal distribution within an AP. This frequency of tonal distribution was observed regardless of the structure of the AP-initial syllable. As for the boundary tones at the end of an IP, various intonation patterns have been identified such as L%, H%, LH%, and HL%, which may override the AP-final tone (Jun 2000).

An important issue to be further explored with respect to the tonal realization at the right edge is exactly how the tonal target is aligned with the segmental string. Theories of the phonetics-prosody interface are interwoven with theories of intonational phonology (Bruce 1977; Pierrehumbert 1980; Pierrehumbert and Beckman 1988; Ladd 2008). One of the important contributions of intonational phonology to theories of phonetics and phonology is to bridge gradient/continuous F0 events in the physical dimension and categorical phonological patterns that are part of the prosodic structure of a given utterance. An important notion is that, once tonal targets are assigned categorically to a syllable or a word, the structure of the tune (the overall shape of F0) that is superimposed over the segmental string of a given utterance is generated by phonetic interpolation between tonal targets (Pierrehumbert 1980). This assumption was further developed into the Autosegmental Metrical theory of intonation (henceforth the AM theory) (Bruce 1977; Pierrehumbert 1980; Pierrehumbert and Beckman 1988; Ladd 2008), based on the concept in Autosegmental Phonology (Goldsmith 1976, 1990) that tonal components are realized in a tier that is independent of the segmental tier. The tones in one tier, however, are not arbitrarily determined over the segmental string, but are associated with the segmental string in another tier by some kind of association rule. For example, a pitch accent of L+H* is associated with a lexically stressed syllable of the word that receives the pitch accent. Similarly, the AP-final H tone in Korean is associated with an AP-final syllable (see the dotted association lines in Figure 9.1 that associate the tonal target with the phonological unit, the syllable).

The phonological representations of tone-segment association are also exemplified in Figure 9.1. They are certainly useful in mapping tones and segments at the phonological level, but the phonological association fails to predict the phonetic timing detail of how the tonal target is actually aligned with the segmental string. Quite a few studies have explored exactly how the tone is aligned or anchored with the segmental string under the rubric of the Segmental Anchoring Hypothesis (SAH) (Arvaniti, Ladd, and Mennen 2000; Atterer and Ladd 2004; Prieto and Torreira 2007; Ladd et al. 2009). The tenet of the SAH is that tonal targets such as F0 troughs and crests for L and H tones are aligned in principled ways with specific locations or landmarks in the segmental structure (such as the midpoint...
or the end of the vowel or syllable) in a given language. The same phonological tone-segment association may then entail differential tone-segment anchoring patterns. For example, both nuclear and prenuclear peaks are likely aligned later in Scottish Standard English (SSE) than in Southern British English (RP) (Ladd et al. 2009), and certain accentual F0 peaks tend to come later in an American English variety spoken by Southern Californians than in a variety spoken by Minesotans (Arvaniti and Garding 2007), although it is often difficult to pinpoint the exact segmental anchors that may interact with various other factors (such as syllable structure and phonetic vowel length) (see Chapter 5 of Ladd 2008 for a thorough review, and Cho 2015 for a related discussion).

In accounting for the language-specific segmental anchoring effects, Ladd (2004, 2006, 2008) suggested that segmental anchoring involves relative timing between laryngeal (F0) and supralaryngeal (segmental) gestures, to be specified in the phonetic grammar of the language in a similar way as for VOT, giving rise to cross-linguistic variation (Cho and Ladefoged 1999)—i.e., stops with the same phonological voicing specification (voiceless unaspirated or aspirated stops) surface with VOTs that differ systematically across languages (see also Cho, Whalen, and Docherty 2019 for related discussion). Languages may take on any of a continuum of segmental anchoring possibilities, and they may also choose a target VOT value along the VOT continuum, both of which can be expressed uniformly in terms of intergestural timing between laryngeal and supralaryngeal gestures.

Given the cross-linguistic evidence for language-specific fine-tuning of segmental anchoring, it remains to be explored how tonal targets in Korean are aligned with the segmental string. A good starting point is to examine the tone-segment alignment for the AP-final H tone in Korean in comparison with that in French, as both Korean and French are considered to have a similar AP-final rising tonal pattern (Jun and Fougeron 2002). Tremblay et al. (2018), based on the literature as well as their own data of one French speaker and one Korean speaker, noted a possible difference in the tone-segment alignment between Korean and French within an AP-final syllable. As seen in Figure 9.10, the AP-final H peak was shown to be realized earlier in Korean than in French. (The difference shown in Figure 9.10 was used for an artificial language learning study to test the perceptual consequences of the potential alignment difference, as will be discussed later in Section 9.5.3.) More systematic quantitative studies are certainly called for to corroborate the generalizability of the cross-linguistic difference in fine-grained tone-segment anchoring, which is again assumed to be stipulated in the phonetic grammar of each language.

### 9.4.3.2 Preboundary Lengthening in Korean

Along with preboundary tones, preboundary lengthening (or phrase-final lengthening) is one of the most robust cues to the upcoming boundary. It is
a cross-linguistically recurrent phenomenon whose effect may be attributable to universally applicable low-level phonetic constraints. It is often assumed to stem from a relaxation of articulatory gestures following a natural physical tendency that causes articulatory movement to slow down towards the end of the utterance before the cessation of the movement (Lindblom 1963). It can also be seen as a kind of supralaryngeal declination over the course of an utterance (Fowler 1988; Vayra and Fowler 1992; Berkovits 1993, 1994; Krakow, Bell-Berti, and Wang 1995; Tabain 2003), showing a gradual temporal declension. As discussed by Cho (2015), however, the seemingly physiologically driven effect may be further controlled by the speaker in a language-specific way, thereby interacting with other aspects of the language such as lexical stress (e.g., English; Turk and Shattuck-Hufnagel 2007; S. Kim, Jang, and Cho 2017), mora (e.g., Japanese; Shepherd 2008) and vowel quantity (e.g., Finnish; Nakai et al. 2008). The fundamental assumption, as suggested by Cho (2015, 2016), is that the degree of preboundary lengthening must be specified in a linguistic description of the phonetics-prosody interface as part of the phonetic grammar of the language (Keating 1984, 1990; Cho and Ladefoged 1999) just as the realization of boundary tones is.

Important theoretical considerations concern the scope of preboundary lengthening – i.e., to what extent preboundary lengthening spreads beyond the final syllable to the preceding syllables, and how the scope interacts with other phonological factors. In an acoustic study, for example, Turk and Shattuck-Hufnagel (2007) demonstrated multiple targets of preboundary lengthening in English, such that preboundary lengthening may be extended to a long-distant stressed non-final (antepenultimate) syllable in a trisyllabic word (Michigan), possibly skipping the unstressed penultimate syllable. A more recent study on English (S. Kim, Jang, and Cho 2017) further indicated that the degree of preboundary lengthening as a function of boundary strength interacts with prominence, showing a lesser degree of
preboundary lengthening when final syllables receive a focus-induced nuclear pitch accent. Kim, Jang, and Cho (2017), however, did not observe an attraction of preboundary lengthening to the antepenultimate syllable regardless of whether it receives a pitch accent or not. Cho, J. Kim, and S. Kim (2013), on the other hand, observed that the antepenultimate (initial) syllable of the English word *banana* is subject to preboundary lengthening, despite the fact that the syllable is unstressed. These studies imply that the exact scope of preboundary lengthening in a given language cannot be categorically determined, as it may interact with numerous factors, engendering the surface variability, although it does interact with the prominence system in some ways.

To the best of my knowledge, however, our understanding of preboundary lengthening in Korean is even more limited. An AP is known to be produced with no discernible preboundary lengthening at the end (Jun 1993, 1996) or, if it does exist, the effect is very small (Cho and Keating 2001). This implies that an AP boundary is marked primarily by the phonologically determined tonal distribution (i.e., #TH...LH#) rather than by preboundary lengthening, so that the final rising tone plays an important role as a preboundary phonetic event to mark the AP boundary (S. Kim and Cho 2009; S. Kim, Broersma, and Cho 2012). On the other hand, the final syllable of an IP undergoes substantial final lengthening (Jun 1993; Cho and Keating 2001; Jang and Katsika 2020) as in many other languages. We do not, however, know much about the exact nature of such preboundary lengthening in Korean, either, and are left with many questions open as to, for example, what the scope of preboundary lengthening in Korean is, and how it is determined in interaction with other factors such as syllable structure (open vs. closed), segmental content (high vs. low vowel; sonorant vs. obstruent) and prominence (focused vs. unforced), to name a few.

9.5 Prosodic-Structural Conditioning of Phonetic and Phonological Processes in Korean

In the previous sections, I have discussed how the fine phonetic detail of segmental and suprasegmental realizations may serve as a phonetic hallmark of higher-order prosodic structure with special reference to prosodic strengthening that may arise with boundary and prominence marking. In this section, I will discuss how specific phonetic and phonological processes may also be conditioned by prosodic strengthening, which again illuminates the prosodic structure as an integral part of speech production that permeates into various levels of both phonetic and phonological processes.
9.5.1 Phonetic Granularity of Coarticulatory Process and Prosodic Strengthening in Korean

A coarticulatory process is often conceived of as an inevitable and automatic phonetic process which has its origin in physiological and biomechanical constraints imposed on the human speech production system (see Kühnert and Nolan 1999 for a general review). For example, when a vowel and a nasal consonant are produced in a row as in *ban*, the velum lowering gesture necessary for /n/ must initiate during the vowel in order to make a smooth transition from the oral to the nasal articulation. As a result, the later part of the vowel is coarticulated (or co-produced) with the velum lowering gesture, introducing some acoustic nasality into the vowel, a phenomenon known as coarticulatory vowel nasaslization. Since such a coarticulatory process is gradient in nature, being continuous and gradual in temporal and spatial dimensions without modifying the phonological category of segments involved, traditional views such as the one crystallized in the Sound Patterns of English (SPE) tradition (Chomsky and Halle 1968) postulate that it is not part of the grammar and, therefore, is something that should be studied outside the realm of linguistics. The sense of dichotomy between gradient and categorical processes, which are roughly equivalent to phonetic and phonological processes, however, has been steadily eroding since the 1980s. An increasing number of studies have adopted advanced scientific methodologies in investigating fine-grained phonetic detail of speech sounds and provided ample evidence for the importance of scalar and gradient aspects of speech in understanding the linguistic sound system. Subphonemic and, therefore, non-contrastive phonetic events, traditionally deemed to be beyond the speaker’s control (i.e., the result of automatic low-level physiological phenomena), have been reinterpreted as part of the grammar – i.e., something that is controllable by the speaker in reference to various higher-order linguistic factors, and that may fine-tune phonological contrasts or shape phonetic forms in accordance with language-specific phonetic rules (see Cho 2011, 2015 for a review).

In an acoustic study, Cho, D. Kim, and S. Kim (2017) explored how prosodic-structural or strengthening factors might modulate a coarticulatory vowel nasalization process in reference to phonological contrasts in English (see Cho 2004 for an articulatory study of vowel-to-vowel coarticulation that explores a similar question). Their results indicated that different sources of prosodic strengthening (boundary vs. prominence marking) bring about different types of linguistic contrast enhancement in coarticulatory vowel nasalization. As for the prominence-induced strengthening effect, in both carryover (#NVC) and anticipatory (CVN#) coarticulatory contexts (e.g., *mob* and *bomb*), the consonant’s [nasality] was found to be enhanced under focus, whereas the vowel showed coarticulatory resistance to the nasal influence even when the nasal
consonant received a phonological contrastive focus (e.g., *mob* vs. *bob*; *bomb* vs. *bob*). On the other hand, boundary strength was found to regulate different types of enhancement patterns, depending on domain-initial vs. final prosodic positions. In the domain-initial position (e.g., *mob*), boundary strength decreased the consonant’s [nasality] and the degree of V-nasalization, thus enhancing CV contrast (i.e., the consonant becomes more consonant-like by a reduction of its sonority, and the vowel becomes more vowel-like by an enhancement of its orality). Interestingly, however, the opposite was found for the domain-final nasal (e.g., *bomb*). The nasal consonant itself was lengthened, and encroached upon the preceding vowel, so that the vowel was more nasalized, showing coarticulatory vulnerability. The results of this study, therefore, illustrate the systematic phonetic granularity of coarticulatory variation as a function of prosodic factors and provide a concrete case in English, demonstrating that coarticulatory V-nasalization is under speaker control and is fine-tuned in a linguistically significant way.

Following up on Cho, D. Kim, and S. Kim’s (2017) study, Jang, S. Kim, and Cho (2018) examined the effects of boundary and focus-induced prominence on coarticulatory vowel nasalization in Korean and discussed their results in terms of implications for cross-linguistic similarities and differences. As was found in English, Korean showed a focus-induced reduction of V-nasalization in both CVN# and #NVC, interpretable as coarticulatory resistance in reference to prosodic strengthening, which enhances the vowel’s [oral] feature. This was the case despite the following cross-linguistic difference: English is a head-prominence language displaying interaction with the word-level stress system, whereas Korean is an edge-prominence language with no such interaction with word-level prosody. This implies that focus-induced prominence operates on the vowel regardless of the language’s prosodic system, heightening the phonetic clarity of the vowel’s orality by modulating a low-level phonetic process.

The phonetic details of the coarticulatory process, however, elucidate cross-linguistic differences between Korean and English. First, the nasality of the initial N in #NVC in Korean was reduced under focus, the exact opposite of what was observed with the initial N in English. This apparent cross-linguistic difference reflects the domain-initial denasalization process in Korean, discussed above. As the authors suggested, given the denasalization, it is not the N’s underlying [nasal] feature, but rather the derived [oral] feature, that is enhanced under focus. It is also interesting to note that V-nasalization in Korean was found to be extremely reduced at the point in the vowel most proximal to N, the source of V-nasalization in #NVC. The lack of a phonetic proximity effect on #NVC is also consistent with the denasalization process of N in Korean. This, however, does not mean that the [nasal] feature is completely delinked. The authors provided some evidence for the presence of nasal murmur during N and for the maintenance of the vowel nasality difference between NVC and CVC in the early part of the vowel. The
denasalization process in Korean thus may not be a completely phonological process, but rather – as I discussed above – a domain-initial phonetic process in which the velum lowering gesture in NVC is aligned later than in English, rendering the ostensible denasalization.

Second, the final nasal (in CVN#) in Korean did not show preboundary lengthening. Instead, it was shortened, and its nasality was weakened IP-finally. This was again the opposite of what was observed in English. Korean, however, still showed an increase in V-nasalization IP-finally, just as in English. The boundary-related coarticulatory effect in CVN# (i.e., the increased V-nasalization) in both English and Korean may be understood as stemming from a loosening of articulatory linkage between the oral constriction gesture and the velum lowering gesture, leading to more nasal leaking during the vowel (hence more nasalization). The authors further explained that the final N shortening may be consistent with what Beddor (2009) suggested: variation in coarticulatory V-nasalization reflects a relatively constant activation time of the velum lowering gesture which is variably aligned to the oral gestures. When the velum lowering gesture overlaps more with the vowel’s oral gesture, the N-duration is compensatorily shortened, showing an inverse relationship between vowel nasalization and N-duration. But based on the fact that English showed no such inverse relationship (i.e., it showed an increase in both N-duration and vowel nasalization in CVN#), the authors suggested that the magnitude of the velum lowering gesture may be greater in English than in Korean, and it is expanded IP-finally, stretching over both the vowel and the following consonant, as schematized in Figure 9.11.

Alternatively, as Jang, S. Kim, and Cho (2018) discussed, languages may employ differential intergestural timings between the oral and the velic gestures (as is also schematized in Figure 9.11) resulting in cross-linguistic variation, just as languages employ differential intergestural timings between the oral release gesture and the laryngeal gesture, which accounts for cross-linguistic variation in VOT (Cho & Ladefoged 1999) or tone-segment anchoring (Ladd 2008). If so, the cross-linguistic difference

![Figure 9.11](image_url)
in the vowel nasalization process in both CVN# and #NVC may be accounted for by the language-specific timing difference: the velum lowering gesture in CVN# is aligned earlier with respect to the oral gesture in Korean than it is in English, but in #NVC, it is aligned later in Korean than in English.

In sum, the cross-linguistic comparison of coarticulatory V-nasalization provides insights into the phonetics-prosody interface from a cross-linguistic perspective, illuminating how the detailed phonetic manifestation of a seemingly low-level coarticulatory process may indeed be controllable by the speaker, and how it brings about different phonetic forms in reference to prosodic-structurally driven strengthening in a language-specific way.

9.5.2 Prosodic Strengthening Delimiting Phonological Processes

As briefly mentioned above, the phonological relevance of prosodic phrasing has been discussed in terms of how the domain of application of phonological rules is delimited by prosodic boundary. In relation to Korean, Jun (1993, 1998) explained that the notion of AP, which is intonationally defined as a phonologically meaningful prosodic unit, is further supported by segmental phonology—i.e., AP serves as a domain of application for some phonological rules in Korean. For example, the Lenis Stop Voicing Rule where a lenis stop becomes voiced between vowels applies when the triggering vowels and the stop occur within the same AP (e.g., /namu/ # /toma/ → [namu] # [dome], where “#” = an AP-medial word boundary, but /namu/ # /toma/ → [namu] # [toma] where # = an AP boundary). Another case is found with the Post-Obstruent Tensing Rule where a lenis stop becomes tense after an obstruent when the triggering obstruent and the stop occur within the same AP (e.g., /hankuk/ # /toma/ → [hankuk] # [t*oma], where # = an AP-medial word boundary). (See S.-J. Kim 2000 and Jun 1998 for other possible rules whose domain of application is assumed to be an AP).

These examples highlight the role of prosodic structure in phonological processes, demonstrating an interaction between prosodic structuring and phonology. It is also worth noting that the delimitation of rule application may also be seen as a kind of domain-initial strengthening, in that the domain-initial segment is strengthened in such a way as to resist the contextual influence that would otherwise cause phonological modification, weakening its phonemic identity on the surface (and therefore violating phonological faithfulness constraints).

An important issue that needs further elucidation is the nature of such prosodic-structural conditioning of rule application whose detailed effect may vary as a function of boundary strength. The rules discussed above are generally assumed to be completely blocked by a stronger prosodic boundary, or an IP boundary (i.e., when an AP boundary is also aligned with an IP boundary). But it is unclear whether such a complete blocking also applies
when an AP boundary occurs in an IP-medial context, and is thus not aligned with an IP boundary. Cho and Keating (2001) in fact provided some evidence that phonetically voiced variants of lenis stops do occur occasionally in AP-initial positions, suggesting that the delimitation of the lenis stop voicing rule may be gradient as a function of boundary strength (see Choi et al., in preparation, for similar results.) More studies are called for to understand the nature of the interaction between prosodic strengthening and phonological processes. Such studies should also include the effects of prominence, as it is also possible that phonological modification may be further constrained by strengthening due to prominence similarly to the way that prominence-induced strengthening causes coarticulatory resistance, maintaining the segment’s canonical phonetic form.

### 9.6 Linking Prosodic Encoding in Production to Prosodic Decoding in Perception

Thus far I have discussed various aspects of prosodic strengthening from the production point of view, illuminating the phonetic encoding of prosodic structure. A question that naturally follows concerns to what extent the structural information contained in the speech signal may be exploited by the listener to decode the prosodic structure as intended by the speaker. Unlike the large body of research on the phonetics-prosody interface from the speaker’s point of view, however, fewer studies have been available in the literature, especially on Korean, regarding the decoding of prosodic structure from the listener’s point of view. In this section, I will briefly review some of the studies that have explored the linking of prosodic encoding to decoding.

#### 9.6.1 Prosodic-Structural Modulation of Phonetic Categorization

Some recent studies on speech perception have provided evidence of prosodically conditioned modulation of speech perception in phonetic categorization (e.g., S. Kim and Cho 2013; Mitterer, Cho, and Kim 2016). Let’s consider the phonetic form of a stop whose VOT value falls between voiced (with short-lag VOT) and voiceless (with long-lag VOT) categories along the VOT continuum in English. S. Kim and Cho (2013) showed that the listener is likely to perceive this form as a voiceless stop when it is presented in a prosodic context in which a voiceless stop is produced with a relatively shorter VOT (e.g., phrase-medial position). The listener perceives the same phonetic form as a voiced stop when it is presented in a prosodic context (e.g., phrase-initial position) in which a voiceless stop is produced with a relatively longer VOT. This implies that listeners compute prosodic structure and modulate perception of segments by making reference to the granularity of segmental variation that arises from prosodic
structure being computed online in speech processing. This prosodic-structural modulation of phonetic categorization in English was evident not only in the perception of native English listeners, but also of Korean learners of L2 English, indicating the robustness of perceptual modulation due to prosodic structure that may apply to both L1 and L2 speech. (See Mitterer et al. 2016 for a further discussion on how the prosodic modulation effect may be disentangled from perceptual modulation due to variation in speech rate.)

9.6.2 The Use of Language-Specific Phrase-Level Tones in Lexical Segmentation

Kim and Cho (2009) investigated the role of phrase-level tones in word segmentation in Korean using word-spotting experiments in which listeners had to detect a possible word out of a stream of nonsensical multiple syllables. They particularly asked to what extent the language-specific edge tones of an AP would help listeners with lexical segmentation. Given that the AP boundary is always aligned with a word boundary, and given that it is accompanied by systematic tonal cues (the AP-final H tone and the AP-initial L tone, except when the initial segment is fortis or aspirated), they hypothesized that the post-lexical (phrase-level) intonational cues would help listeners with lexical segmentation, which would in turn facilitate spotting the word out of an otherwise nonsensical string of syllables. The results indeed showed that listeners performed far better at spotting a word in the H#L context – i.e., when the word was preceded by a H tone, and the initial of the word itself started with a L tone – than in non-canonical tonal contexts (e.g., L#L, L#H). They further observed that listeners did not necessarily benefit from the presence of preboundary lengthening in the optimal tonal context (H#L) consistent with the AP boundary percept, and that preboundary lengthening came into effect only when the tonal context was not canonical, providing weak support to the AP boundary percept. This study therefore confirmed that the speaker’s post-lexical tonal marking of a prosodic boundary is indeed utilized by the listener in lexical processing, presumably through a prosodic decoding process – i.e., computing prosodic structure based on the intonational cues available in the speech signal.

S. Kim, Boersma, and Cho (2012) extended the study to further test the role of language-specific intonational cues in processing an unfamiliar language. They employed an artificial language (AL) learning paradigm in which listeners had to listen to a stream of several unknown words in a row in an unfamiliar (artificial) language repeatedly for a certain period of time (about 20 minutes). Listeners were then asked to identify the words that were used in the language, given two forced choices (a word in the AL vs. a non-word). In general, listeners are able to identify target words above chance level, based on transitional probabilities available in the stream of
the words (i.e., the probability for a particular syllable to come after another particular syllable is higher within a word than across a word boundary) (Saffran, Newport, and Aslin 1996; Saffran 2001). S. Kim et al. (2012) then investigated the possible roles of available prosodic cues in listeners’ learning of the AL. If a prosodic cue points to a prosodic juncture, word boundaries are likely to be detected, which will in turn help learning the language. Specifically, the authors asked to what extent the use of prosodic features would be similar across listener groups with different native language backgrounds (Korean vs. Dutch) or whether it is language-specifically driven in learning an unfamiliar language. The presence of preboundary lengthening in the target words indeed improved the listeners’ learning performance regardless of their native language background (Korean or Dutch). But Korean listeners performed far better than Dutch listeners in learning the AL when the final syllable of the target words was associated with an H tone (in the final F0 rising condition). These results imply, as was also suggested by the authors, that while tonal and durational cues may contain universally applicable auditory-perceptual salience, the way that they are used is language specific, constraining the use of available prosodic cues in processing an unfamiliar language.

9.6.3 The Use of Fine-Grained Phonetic Detail in Segmental Anchoring and Tonal Scaling

As briefly mentioned above, Korean and French have an AP whose right edge is associated with an H tone, but they differ in the details of tone-segment alignment (segmental anchoring), with the F0 peak coming relatively earlier in Korean than in French (see Figure 9.10). Tremblay et al. (2018) also observed that the following L tone across an AP boundary tended to be lower in Korean than in French. (This may be due to a lower L tone target specified in Korean, or the target may be realized earlier because the onset starts earlier, as reflected in the earlier F0 peak for the preceding H tone.) In an AL study, the authors tested how Korean listeners would process fine-grained phonetic detail with regard to the segmental anchoring of the AP-final H tone and the tonal scaling of the following L tone. Their experimental conditions may be schematized as in Figure 9.12.

The results can be summarized as follows. First, when the tonal target (L) is fully attained as in Figure 9.12a (Low L Scaling), the fine-grained phonetic detail in H-alignment (segmental anchoring) did not influence the Korean listeners’ performance in learning an unfamiliar AL. This indicates that the Korean listeners process the preceding H tone phonologically (or categorically) as long as the (following) L tone matches with Korean intonational phonology.

Second, when the L-tone is undershot as in Figure 9.12a (Mid L Scaling), the Korean listeners showed a poorer performance in the early H peak condition,
despite the Korean-like H peak alignment pattern. In Korean (Jun 1993, 2000), the L tone is often undershot in AP-medial positions. It appears then that when the L tone realization deviates from the listener’s L1 experience (i.e., the intonational phonology of the listener’s native language), the listeners become more attentive to the fine-grained phonetic detail in the pre-boundary H-tone alignment. But the question is why the Korean-like early H peak alignment hindered the learning. A phonetic account can be offered. On the one hand, the later H peak (the French-like pattern) imposes a steep slope on the H-to-L pitch movement into the following vowel, giving a salient boundary percept. The perceptual (low-level auditory) salience of the steep slope may have counteracted the adverse effect of the mismatched L-tone (mismatched in terms of its tonal scaling, i.e., the L undershoot), showing a learning effect comparable to the no prosody (control) condition. On the other hand, the earlier H peak (the Korean-like pattern) which results in a less steep slope during the H-to-L pitch movement shows no such boundary detecting effect, hence the poorer learning performance.

Finally, the Korean listeners performed worse when the L tone was more raised as in Figure 9.12c (High L Scaling) than in Figure 9.12b (Mid L Scaling). In fact, their performance was even worse than in the control (no prosody) condition. This was when the L tone realization deviated most from the canonical L tone as stipulated by Korean intonational phonology. As I noted parenthetically in Figure 9.12c, the most raised L tone is no longer indicative of the L target associated with the initial syllable. Instead, it appears to fall roughly halfway between the preceding H tone and the L tone (for the second syllable of the following word). It is as if the tone for the initial syllable were unspecified, thus showing a kind of interpolation from H to L which would be observable only across syllables that are AP-medial in Korean. The severely mismatched tone that is not aligned with

---

**Figure 9.12** Schematics for the conditions of three experiments in Tremblay et al. (2018). (a) Low L Scaling with early/late H peaks; (b) Mid L Scaling with early/late H peaks; (c) High L Scaling with early/late H peaks. The black lines refer to the early H peak condition, and the gray line to the late H peak condition. These figures were drawn based on the methodological descriptions in Tremblay et al. (2018). (The parenthetical remarks “target fully attained”, “target undershot” and “target unspecified/interpolating” are not used in the original paper.)
the edge appears to override any potential learning effect that arises from transitional probabilities in the control (no prosody) condition. As was the case with the L undershoot (Figure 9.12b), however, the Korean listeners’ performance was slightly but significantly better in the later H peak condition, which again is interpretable as a phonetic effect coming from the perceptual salience of a steeper pitch fall.

These results imply that the phonetic difference in tone-segment alignment for an H may be perceived phonologically (or assimilated to the L1 intonational pattern) in a prosodically optimal context (i.e., when the target of the following L tone is fully attained). This, however, does not mean that fine-grained phonetic difference plays no role. When the prosodic boundary is not fully supported by the cues available in the signal, listeners appear to be more attentive to the fine phonetic detail, presumably driven by a universally applicable acoustic-phonetic saliency (see S. Kim and Cho 2009 for a related discussion).

9.6.4 Prosodic Modulation in Phonological Inferencing: The Case of Korean Post-Obstruent Tensing

As discussed above, the domain of phonological rule application is often delimited by prosodic structure. As described in Section 9.4.2, the Post-Obstruent Tensing Rule (POT) in Korean is one of the phonological processes assumed to be blocked by an AP boundary. S. Kim, Mitterer, and Cho (2018) employed an eye-tracking paradigm to test how Korean listeners process segmental variation that arises from the POT. This kind of segmental variation is called phonological variation since it changes the segmental identity categorically as a consequence of a legitimate phonological rule application. Because of the categorical change, such a rule often creates ambiguity on the surface. As shown in Figure 9.13, for example, the phonetic contrast of the minimal pair /puli/ ("beak") and /p*uli/ ("root") is neutralized within an AP when the word is preceded by /[jʌndusɛk]/ ("light green") which ends with an obstruent that triggers tensing of the following

![Figure 9.13](image-url)  
**Figure 9.13** Ambiguity created by an application of Post-obstruent Tensing. This process is assumed to occur within an AP, but to be blocked when the triggering obstruent occurs across an AP boundary.
lenis stop. As a result, the surface form of [p*uli] in [jʌndusɛk]+[p*uli] can underlingly be either /puli/ (“beak”) or /p*uli/ (“root”). S. Kim et al. (2018) first tested whether the presence of the context (/k/) that licenses POT would help listeners infer the underlying representation, a phenomenon known as phonological inferencing (Gaskell and Marslen-Wilson 1996; Gaskell and Snoeren 2008). The authors provided some evidence for such phonological inferencing: the listeners’ eyes indeed fixated on the target word /puli/ on the computer screen more often upon hearing the word in the presence of the triggering /k/ (e.g., [jʌndusɛk]+[p*uli]) than in its absence (e.g., preceded by [jʌnnolɑŋ] “light yellow”, as in [jʌnnolɑŋ]+[p*uli]) within an AP. Figure 9.14a illustrates more fixations on the lenis target upon hearing the phonetic tense form [p*] in the presence of the triggering /k/ than in the control condition (in the absence of /k/). This suggests that Korean listeners activate the lexical representation of a phonologically variable surface form (in this case, the tensified form of the underlying lenis stop) by reversing the phonological rule.

Crucially, the authors also tested whether the presence of another prosodic cue in the preceding (triggering) context would further modulate such phonological inferencing. The prosodic cue was the F0 rise of the preceding context without any durational change. This experimental factor makes it possible to test the effect of F0 alone on the way that the underlying representation is recovered by the listener upon hearing the phonetic tense form on the surface. Given that the F0 rise was consistent with an AP-final H, it was hypothesized that Korean listeners would not resort to phonological inference because the F0 rise would provide prosodic support for an AP boundary, which blocks the POT rule. As shown in Figure 9.14b, the phonological inferencing effect as observed in the absence of F0 (Figure 9.14a) disappeared. The fact that the F0 cue alone (without a temporal cue) blocked the phonological inferencing effect supports the view that phonological inferencing is indeed modulated

![Figure 9.14](image-url)  
Figure 9.14 Eye fixations to the target word with the lenis stop (/p/) upon hearing the phonetic tense form [p*] in the /k/ context that licenses the Post-obstruent Tensing rule in the absence of F0 rise (a) and the presence of F0 rise (b) in the preceding context. (** refers to significance at p<0.01.)
by listeners’ computation of prosodic structure rather than by a perceptual adjustment to temporal variation. More broadly, this study again implies that lexical processing (i.e., spoken word recognition) involves segmental analysis in parallel with computation of prosodic structure, calling for refinement of any existing lexical processing theories to take into account the role of prosodic decoding in lexical processing (see Cho et al. 2007 for a further discussion on the role of the interaction between segmental analysis and prosodic analysis in lexical processing).

9.7 Conclusion

In this chapter, I have reviewed some of the studies on the phonetics-prosody interface in reference to prosodic strengthening in Korean with a view to illuminating multi-dimensional aspects of the Korean sound system. As many of the low-level phonetic phenomena have their origins in physiological and biomechanical constraints imposed on the human speech production and perception system, similar patterns may occur across languages. For example, languages manipulate suprasegmental features such as intensity, pitch, and duration to yield a synergistic effect increasing the relative salience of the prosodic unit under prominence. They signal a prosodic juncture by slowing down articulation approaching it. But the fine phonetic details of cross-linguistically recurrent patterns, which are presumably driven by common low-level phonetic underpinnings, may never be the same, as their final phonetic shape is determined by the phonetic grammar of the particular language (Keating 1984, 1985; Kingston and Diehl 1994). The language-specific phonetic grammar (or phonetic knowledge) gives rise to phonetic arbitrariness within phonetic universals (Cho and Ladefoged 1999; cf. Cho, Whalen, and Docherty 2019). I propose that, central to this phonetic arbitrariness, is the prosody of the language as an integral part of speech production. The studies reviewed in this chapter indeed provide converging evidence that a large part of the phonetic arbitrariness that makes Korean unique among the world’s languages is in fact determined by the way that phonetics interacts with prosodic structure in a Korean-specific way.

Korean, which may be classified as an edge-prominence language, has its unique prosodic system. We have seen that the intonational system of Korean stipulates the tonal distribution within an AP whose edges are the loci of phonological association of tonal targets. The rule-driven edge tones come with segmental strengthening (domain-initial strengthening) especially at the left edge of a prosodic constituent, which appears to serve both delimitative and culminative functions, thus making it hard to disentangle boundary and prominence marking in Korean. As in many
other languages, phonetic reflexes of prosodic structure in Korean reflect a kind of prosodic strengthening that heightens the phonetic clarity of segments that occur at prosodically important landmarks. Prosodic strengthening also leads to enhancement of paradigmatic and syntagmatic contrasts that arise from the dual function of prosodic structure. Again, the exact way in which prosodic strengthening is phonetically realized, and in which linguistic contrasts are enhanced, is governed by the prosodic component of the grammar of the language in interaction with other higher-order linguistic components of the language. We have in fact seen some evidence that prosodic structure (and its encoding) does not simply regulate the grouping of phonological units and the signaling of their relative salience, but does so specifically by making reference to phonetics, phonology, information structure, and other linguistic components of the language.

The view of prosody as an integral part of speech production entails further that the prosodic structure of an utterance is phonetically encoded into the speech signal, carrying information about various components of the linguistic structure of the language. An important question here is how listeners then decode the linguistic message to be conveyed by prosodic structure. I touched on this question and discussed how prosodically conditioned fine-grained phonetic detail that arises from a particular tonal distribution, tone-segment alignment, and rule-driven segmental variation may influence Korean listeners’ ways of perceiving speech sounds. The observed perceptual modulation as a function of prosodically conditioned phonetic variation implies that Korean listeners indeed compute or decode the intended prosodic structure and adjust their perception according to this computed prosodic structure. More broadly, this suggests that any auditory and perceptual aspects of speech should be elucidated by considering the prosodic-structural modulation of speech perception.

Research on the phonetics-prosody interface in Korean has clearly advanced our understanding of the nature of the Korean sound system in which phonetics and prosody are intertwined in a linguistically systematic way. We are, however, still at an embryonic stage, left with new challenges. How can our knowledge on the phonetics-prosody interface in Korean be reinforced to build a unified model of speech production and perception that captures the complex interplay between phonetics and prosody in both descriptively and explanatorily adequate ways? And how can the influence of various other higher-order linguistic structures (phonology, morphology, syntax, semantics, information structure, discourse) be integrated into such a model? It is hoped that the Korean linguistic community will work together to develop a unified model of speech production and perception that crystallizes the linguistic functions of prosody in the general architecture of the grammar of Korean.
References


