Linguistic Functions of Prosody and Its Phonetic Encoding with Special Reference to Korean

TAEHONG CHO

Hanyang Institute for Phonetics and Cognitive Sciences of Lan-guage, Hanyang University, Seoul

1 Introduction

This review article discusses some fundamental issues regarding linguistic functions of prosody that underlies speech variation on the surface. An important premise of the current discussion is that some significant portion of speech variation that appears to come about beyond the speaker control as a consequence of low-level phonetic processes is in fact conditioned systemat-ically by multiple factors that stem from higher-order linguistic and non-lin-guistic structures. Central to such speech variation, as I will discuss, is pro-sodic structure that plays a pivotal role in modulating phonetic realization in reference to other structural information that may be available in the planning process of speech production. In the next sections, I will first discuss how speech variation may be related to higher-order linguistic structures such as information structure and syntax in conjunction with prosodic structure (Sec-tion 2). I will then outline how prosodic structure may be created in the speech planning process (Section 3), followed by discussing some intractable issues

Japanese/Korean Linguistics 29.
Edited by Kaoru Horie, Kimi Akita, Yusuke Kubota, David Y. Oshima, and Akira Utsugi.

regarding the phonetic granularity that cannot be easily captured by a phonologically defined prosodic structure but must be reflected in the speech planning process (Section 4).

2 Speech Variation and Prosodic Structure

Speech is variable by nature as the speaker's motor execution is affected by various factors that stem from both speech-internal and external sources. One obvious speech-internal factor that contributes to speech variation is a physiological one. It is biomechanically difficult, if not impossible, for a speaker to assume an exact same articulatory posture (or to coordinate speech organs) to produce the exact same speech utterance even when it is repeated immediately one after another in the same communicative context. We, however, do not notice there exists such speech variation because it occurs at a phonetic level that does not participate in modulating the speaker's motor execution in any meaningful ways. But this source of speech variation magnifies when it comes to variation across speakers of the same variety or dialect of the language. This is because speakers inevitably differ in their anatomical dimensions of the articulatory apparatus. The resulting speaker variation thus provides speaker-specific idiosyncratic information that may serve as indexical information about the speaker, allowing us, for example, to identify who is talking over the phone (see Dellwo, Huckvale & Ashby, 2019, for related discussion). This type of speech variation, however, does not pertain directly to a delivery of linguistic message intended by an interlocutor, although such indexical information contributes to it to some extent (cf. Levi & Pisoni, 2007).

Another source of speech variation that is more pertinent to exchanging linguistic message is a paralinguistic one. We can easily picture ourselves producing the Korean greeting word *annjʌŋhasejo?* (안녕하세요? 'How are you?') differently depending on our mood or emotion on the day we say it. That is, by changing the way we produce the same utterance (e.g., *annjʌŋhasejo?*), we deliver a different paralinguistic message (e.g., mood) alongside the linguistic message to the listener. Thus, the paralinguistically driven source of variation may serve as indexical information about the speaker's mood or emotion.

Finally, more directly pertinent to a delivery of linguistic message is speech variation that makes reference to linguistic structure. For a better illustration of this point, let's imagine situations in (1a-d) in which a child may ask various questions in English.

(1) a. Mommy, what did you say?
    b. Mommy, who did you say dislikes hamburgers?
    c. Mommy, what did you say Daddy dislikes?

d. Mommy, can we go to McDonalds?

(2) I said, Daddy dislikes hamburgers.

Here in response to each question in (1), a morpho-syntactically identical sentence with the same phonological (segmental) composition may be used as an answer in (2)—i.e., *I said, Daddy dislikes hamburgers.* But each question in (1), especially (1a-c), requires a different level of information that entails different types of focus realization on the surface (see Gussenhoven, 2008 for a review). (1a) requires new information embedded in the entire utterance, so that the whole sentence is expected to become a locus of required information—that is, the information structure of the interlocution entails *broad focus* falling on the whole utterance. On the other hand, the required information for (1b) and (1c) is much narrower in scope, specific to "who" and "what", respectively. This type of focus is called *narrow focus* falling on the specific location "Daddy" or "hamburgers' in correspondence to "who" and "what." Finally, in response to (1d), (2) can still be an answer, flouting the maxim of relation, but in an emphatic way (see below).

2.1 Prosodic Structure in Reference to Information Structure

While information structure determines the locus of information to be focused, it does not mean that it determines exactly how focus information is phonetically realized. There must be some kind of interaction between information structure and phonetic component of the grammar which governs phonetic implementation (motor execution). Whether information structure directly informs the phonetic component of the grammar is beyond the scope of the present review, but it is reasonable to assume for now that this is done via prosodic structure—i.e., information structure influences prosodic structure first which in turn provides an overall production 'skeleton' or articulatory frame according to which speech units are organized and articulated. So the basic premise, as mentioned at the outset of this review article, is that prosodic structure is a central component which interacts with various other linguistic structures, so that an interaction between a higher-order linguistic structure and the phonetic component is mediated by prosodic structure (see Cho, 2016, 2022 for further discussion). In this view, the information structure given in each case of (3a-c) is assumed to inform the speech planning process, so that a particular prosodic structure for a planned utterance is constructed in accordance with the information required by the interlocutor.

(3) a. [Daddy] [dislikes hamburgers] (*broad focus*)

H* L- (H*) H* L-L%

b. [Daddy dislikes hamburgers] (*narrow focus on 'Daddy'*)

(L+)H* L-L%

c. [Daddy dislikes hamburgers] (*narrow focus on 'hamburgers'*)

(L+)H* L-L%

d. [Daddy] [dislikes] [hamburgers] (*emphatic rendition*)

(L+)H* L- (L+)H* L- (L+)H* L-L%

In (3), prosodic structure for each utterance is expressed roughly in line with prosodic labelling conventions of English Tones and Break Indices (ToBI) (Beckman & Ayers, 1994; Beckman, Hirschberg & Shattuck-Hufnagel, 2005). The labelling conventions were developed based on the Autosegmental-Metrical (AM) Theory of Intonational Phonology (e.g., Pierrehumbert & Beckman, 1988; Beckman, 1996; Shattuck-Hufnagel & Turk, 1996; Ladd, 2008). Prosodic structure as in (3) assumes association between a nuclear pitch accent (H* or L+H*) and a stressed syllable of a word to which the pitch accent is assigned. An 'L-' refers to a phrasal tone that follows a last nuclear pitch accent within a phrase, so that it configures the tune of the phrase filling the gap between a nuclear pitch accent and the end of a phrase, called an Intermediate phrase. Note that when there is more than one pitch accent in an Intermediate Phrase as in the second phrase ('dislikes hamburgers') of (3a), the last one becomes a nuclear pitch accent, and the preceding pitch accent is called a prenuclear pitch accent. An 'L%' refers to a boundary tone that is generally aligned with the last syllable of a larger phrase, called an Intonational Phrase, which is assumed to be the largest phrase in the hierarchy of prosodic structure.

The ToBI labelling as given in each utterance in (3) illustrates two important linguistic functions of prosodic structure—i.e., prominence marking (prominence distribution) and boundary marking (prosodic phrasing). Prominence is marked primarily by assigning a pitch accent to a word that is meant to be relatively more salient than any other prosodic constituents within an Intermediate Phrase. This is pertinent to our discussion on focus realization. In a broad focus condition, a pitch accent may be assigned to multiple words across the utterance as the whole utterance is meant to be the locus of information required by the interlocutor. Thus, a typical prominence distribution under broad focus is that a pitch accent is assigned to all three content words as in (3a). Prominence distribution is intricately related to prosodic boundary marking (or phrasing). In this particular case, the utterance that forms an Intonational Phrase is divided into two Intermediate Phrases as marked by L- in the middle. Note also that the prosodic structure assumed for (3a) may vary in the same broad focus condition, so that the verb *dislikes* may not receive a pitch accent, as indicated by a parenthetical H* and the entire utterance may be produced with one Intermediate Phrase under one Intonational Phrase.

(3b-c) illustrates two possible prosodic structures with the locus of information being 'Daddy' in response to 'Who' and 'hamburgers' in response to

'What'. In these cases, it is most likely that one pitch accent is assigned to each utterance, and prominence is distributed in such a way that a pitch accent falls on the word that should be most salient in accordance with the information structure. Finally, (3d) illustrates a prosodic structure for the utterance where each content word forms a separate Intonational Phrase. One can imagine a situation where this rendition may occur—e.g., when a child tenaciously asks the same question again and again, the mom may say the utterance with a prosodic structure similar to (3d) preceded by 'How many times do I have to tell you!' This kind of rendition is much more emphatic compared to (3a) that may occur under broad focus.

Information structure, as outlined above, is assumed to inform the speech planning process where a particular prosodic structure for a planned utterance is constructed. It is proposed that a selection of prosodic structure occurs at least in part in reference to the information required by the interlocutor as in (3a-d). In such a speech planning model that I currently envisage, once an abstract prosodic structure (abstract in the sense that only categorically defined information is included such as phrasings and tonal assignments) is constructed in reference to information structure in the speech planning process, it is fed into the phonetic component of the grammar. Note that in a well received speech production model as proposed by Levelt and colleagues (Levelt 1989; Levelt et al., 1999), this kind of selection process is done by a devise called "Prosody Generator," which influences "phonetic spell-out procedures" that determines a final "phonetic plan." Such phonetic plan should then be passed on to the phonetic component for motor execution. I propose that it is in this phonetic component where abstract phonological and prosodic units that comprise a planned utterance is fleshed out with actual phonetic content governed by a language-specific phonetic grammar (see Cho, 2015 for a related review). The phonetic grammar, as discussed in Keating (1984, 1985, 1990) and Cho & Ladefoged (1999), characterizes language arbitrariness in phonetic implementation, so that, for example, the same abstract phonological labelling for a stop consonant such as [voiced] or [aspirated] does not translate into the same phonetic values across languages (see Cho, Whalen & Docherty, 2019, for a recent survey). Rather, the actual phonetic implementation occurs in a language-specific way (or governed by a language-specific phonetic grammar), engendering, for example, phonetic variation in VOT for the same phonological label of [voiced] or [aspirated]. Similarly, association between tones and segmental units is specified in a symbolic representation at an abstract level, and it does not tell us about the fine phonetic detail of how the tonal target, for example, of L+H* is phonetically aligned with the segmental string. In fact, the tone-segment alignment for the same L+H tonal composition is best characterized as a gradient process that varies from language to language or from variety to variety of the same language (e.g., Arvaniti, Ladd & Mennen, 2000; Atterer & Ladd, 2004; Ladd,

Schepman, White, Quarmby & Stackhouse, 2009). The reader is referred to Chapter 5 of Ladd, 2008, where this issue is discussed under the rubric of the *Segmental Anchoring Hypothesis*. Here again, this subtle fine phonetic detail of the tone-segment alignment that differs across languages can be attributed to the language-specific phonetic rule which operates at the phonetic component of the grammar. It fine-tunes the actual phonetic implementation in a language-specific way that renders phonetic variation across languages (see Cho, 2015 for more discussion).

Now let us consider similar cases in Korean as in (4) and (5):

(4) a. *ʌmma mwʌlako hasjʌs*ʌjo?*
엄마, 뭐라고 하셨어요?
'Mommy, what did you say?'

b. ʌmma, nuka hɛmpʌkʌlɨl silʌhantako hasjʌs*ʌjo?
엄마, 누가 햄버거를 싫어한다고 하셨어요?
'Mommy, who did you say dislikes hamburgers?'

c. ʌmma, ap*aka mwʌl silʌhantako hasjʌs*ʌjo?
엄마, 아빠가 뭘 싫어하신다고 하셨어요?
'Mommy, what did you say Daddy dislikes?'

d. ʌmma, mɛktonaldi kato twɛjo?
엄마, 맥도날드 가도 돼요?
'Mommy, can we go to McDonalds?'

(5) a. *broad focus*
((ap*aka**) (**hɛmpʌkʌlɨl**) (**silʌhasjʌ**))**
아빠가 햄버거를 싫어하셔.
'Daddy' 'hamburgers' 'dislikes'

b. narrow focus on **ap*aka** ('Daddy')
((ap*aka hɛmpʌkʌlɨl silʌhasjʌ**))**

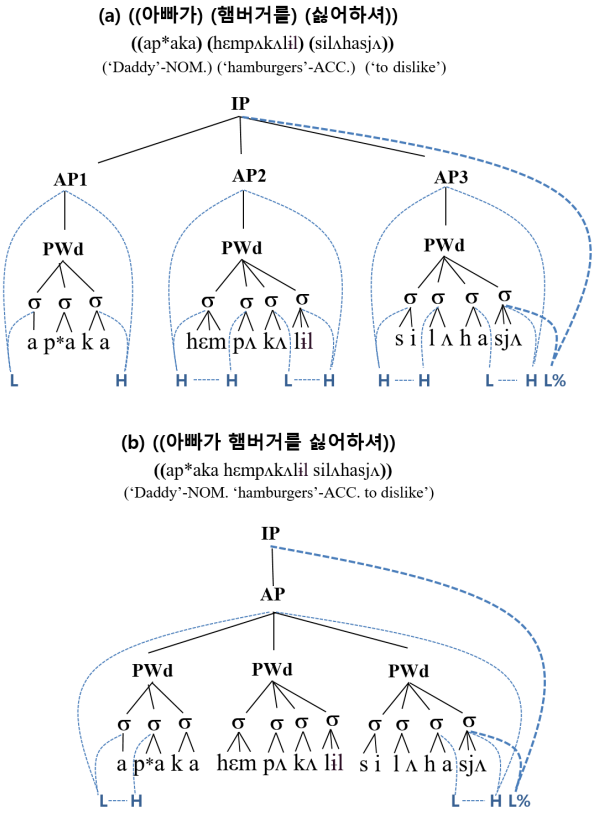
c. narrow focus on **hɛmpʌkʌlɨl** ('hamburgers')
((ap*aka) **(hɛmpʌkʌlɨl** silʌhasjʌ**))**

d. an extremely emphatic case
((ap*aka)) ((hɛmpʌkʌlɨl)) ((silʌhasjʌ))

Here, Korean employs a different prosodic system for focus realization, which stems from the difference in the typology of prosody between the two languages (Jun, 2014)—i.e., English as a head-prominence language versus Korean as an edge-prominence language. In English, a pitch accent is assigned to the head of an Intermediate Phrase, which is usually a lexically stressed syllable of a content word that is meant to receive prominence. On the other hand, Korean does not employ lexical stress and pitch accent in its prominence marking system, but rather it distributes prominence in terms of phrasing. For example, in a neutral or broad focus context, the utterance in

(5a) forms three small phrases, known as Accentual Phrases (AP), as marked by single parentheses. One or more APs can be grouped to form a larger prosodic constituent, an Intonational Phrase, as marked by double parentheses in (5).

(6)



The information about prosodic phrasing and tonal distribution provided here is largely in line with K-ToBI that has been developed to provide prosodic labelling conventions for Seoul Korean (Jun, 2000). The figure in (6a) illustrates a schematized prosodic structure with each word forming a separate AP. An AP is assigned with phrasal tones that form a canonical pattern of T(H…L)H. The phrasal tones at both ends are essential edge tones that demarcate the beginning and the end of an AP, while the parenthetical tones can be deleted (or completely undershot) specially when an AP is relatively shorter (bisyllabic), and can be realized on a word that contains more than two syllables. The initial 'T' can be an L tone or a H tone, conditioned by the laryngeal feature of the AP-initial segment, so that the initial syllable receives

a H tone with fortis or aspirated consonants (i.e., fortis/aspirated stops and affricates and fricatives /s, s*, h/) and an L tone elsewhere. So in (6a), AP2 and AP3 begin with a H tone since their initial segments are /h/ and /s/, respectively, whereas AP1 whose initial segment is a vowel begins with an L tone. As these tones are phrasal tones assigned to an AP, their association lines originate from an AP node. The boundary tone (e.g., H%, L%) that demarcates the end of an IP can in theory override the final phrasal tone (H) of the AP. But it should be in principle possible that an AP-final H and a boundary tone are combined to be realized as a complex tone HL%. But in Jun (2000), such a bitonal realization is taken to be a boundary tone which overrides the AP final H. Note that in the figures in (6), PWd refers to a Prosodic Word that is a prosodic constituent smaller than an AP, but it is often the case as in (6a) that one PWd forms one AP.

Now compare (5a) with (5b) whose prosodic structure is schematized in (6b). When *ap*aka* ('Daddy'-Nom) receives narrow focus in (5b), the utterance forms one phrase, so that the focused word is positioned at the left edge of the phrase to be prominent over the remainder of the phrase, hence an edge-prominence language. Since the three APs in a typical phrasing pattern in (5a) merges into one AP headed by the focused word at the left edge, this rephrasing is often called a 'dephrasing'—i.e., possible AP boundaries are deleted after a focused element. Similarly, in (5c), the focused word ('hamburgers') becomes the left edge of an AP. This AP, as a dephrasing process of focus realization, encompasses the following word that would otherwise form a separate AP in a neutral context. English shows a similar process but it is not phrasing but a placement of nuclear pitch accent on post-focal prosodic words that is suppressed. In the English case, the post-focal string is said to be 'de-accented' rather than 'de-phrased.' Finally, in (5d), as in English, all three words may form separate IPs, so that each word occurs at the edge of a largest phrase. This type of phrasing may occur in a context which requires extremely empathic speech. (Note that an intermediate level of phrase may in theory play a role in constructing prosodic structure in Korean, but how it is defined has not been fully articulated (see Jun, 2007 and 2011 for related discussion).) Here again, as discussed with English cases, prosodic structure that is constructed for a given sentence is assumed to provide an articulatory frame based on which abstract phonological units are organized and eventually fleshed out with phonetic content in a language-specific way governed by the phonetic grammar of the language, Seoul Korean.

2.2 Prosodic Structure in Reference to Syntax

In the previous section, I discussed how a morpho-syntactically same sentence structure in both English and Korean may be produced differently as a function of information structure, thus engendering considerable speech variation. We have now compelling evidence that this kind of speech variation

is not a random noise which impedes communicative processes, but it signals linguistic information. Crucially, such high-order linguistic information does not influence phonetic implementation directly, but through its interaction with prosodic structure, in such a way that a particular prosodic structure for a given utterance is determined in reference to information structure. But information structure is only one of many factors that influences a formation of prosodic structure. Another important contributing factor is syntactic structure. Examples in (7), which were adopted from Cho (2022), highlight the linguistic function of prosodic structure in interaction with syntactic structure.

(7) a. 공사가다망하다
koŋ.sa.ka.ta.maŋ.ha.ta
b. 공사가 다망하다
(*koŋ.sa.ka*) # (*ta.maŋ.ha.ta*)
('public and private matters'-NOM) ('to be busy')
'(someone) is busy with various public and private matters'
c. (*koŋ.sa.ka*) # (*ta*) # (*maŋ.ha.ta*)
('construction-NOM') ('all') ('to mess things up')
'the construction is all messed up'
d. (*koŋ.sa*) # (*ka.ta*) # (*maŋ.ha.ta*)
('construction site') ('to go') ('to mess things up')
'(things) are messed up while going to a construction site'
e. (*koŋ*) # (*sa.ka.ta*) # (*maŋ.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: These examples are adopted from (1) in Cho (2022))

The string of syllables in (7a) written without any space between syntactic constituents does not tell us about its underlying morpho-syntactic structure. In (7b), the same string of syllables is now written with a space between indented words. This space is in fact aligned with syntactic juncture between two major syntactic constituents, a subject NP (*koŋ.sa.ka*) and a VP (*ta.maŋ.ha.ta*), thus indicating the sentence's underlying syntactic structure. The orthographic convention is useful in a case like this to disambiguate a possible syntactic (structural) ambiguity. The same, however, does not hold in spoken language based on which a child acquires his/her mother tongue. As space is to written language, so is prosody to spoken language. Imagine that the sentence in (7a) is produced with no prosody—that is, with completely flat intonation and fixed duration and loudness across the syllables that comprise the sentence. As discussed in Cho (2022), the utterance with no prosody (i.e., no change in suprasegmental features such as pitch, duration and amplitude) does not provide any predictions about the speaker's planned

prosodic structure. It therefore renders multiple interpretations of prosodic structure, creating syntactic-structural ambiguity as exemplified in (7b-e).

For example, the prosodic juncture (boundary) of (7b) between two APs of (*koŋ.sa.ka*) and (*ta.maŋ.ha.ta*) is aligned with a major syntactic boundary between an NP and a VP, to mean '(someone) is very busy with various public and private matters.' With this prosodic phrasing, the last syllable *ka* of the first AP (*koŋ.sa.ka*) functions as a nominative marker, indicating that the preceding two syllables *koŋ.sa* form a subject NP (a lexicalized compound of *koŋ* 'public' and *sa* 'private'.) Another AP that groups the following four syllables (*ta.maŋ.ha.ta*) together indicates that it is most likely interpreted as one word meaning 'to be busy.' But the same first syllable *ta* in *ta.maŋ.ha.ta* may form a single AP as in (7c) *(koŋ.sa.ka) (ta) (maŋ.ha.ta)* in which case *ta* is likely to be interpreted as a monosyllabic adverb ('all'). (7d) shows another prosodic phrasing pattern that signals a different syntactic parsing. In (7d), the phrasing with three separate APs (*koŋ.sa*), (*ka.ta*) and (*maŋ.ha.ta*) is likely to mean that 'things are messed up while (someone) is going to the construction site.' There is yet another possible prosodic grouping with a different set of three APs—i.e., (*koŋ*) (*sa.ka.ta*) (*maŋ.ha.ta*) in (7e), indicating a different morpho-syntactic parsing. Here, the first AP (*koŋ*) is likely to be parsed as an object NP 'ball', and the second AP (*sa.ka.ta*) as a verb ('to buy and go'), meaning that '(things) are messed up while (someone) is going somewhere after buying a ball.'

The different phrasing patterns in (7) thus demonstrate that prosodic structure for a given utterance is built up in reference to morpho-syntactic structure (e.g., Nespor & Vogel, 1986; Selkirk, 1984, 1995). Crucially, however, this does not mean that syntactic structure governs prosodic structure nor does information structure (see Keating & Shattuck-Hufnagel, 2002, for related discussion). In fact, current theories of prosody reiterate its autonomy in the architecture of linguistic structure with a view that prosodic structure is a grammatical entity parsed in its own right (Beckman, 1996, Shattuck-Hufnagel & Turk (1996), Keating & Shattuck-Hufnagel, 2002; and see Cho, 2016 and 2022 for related discussion). The examples in English and Korean given in (3) and (5) are indeed in support of the autonomy of prosody. Recall that the same syntactic structure *Daddy dislikes hamburgers* both in English and Korean may be produced with different prosodic structures, leaving its morpho-syntactic structure (and its core linguistic meaning) intact. Similar evidence is found in Shattuck-Hufnagel and Turk (1996) who uses a classic example as shown below in (8). This illustrates discrepancy between prosodic structure and syntactic structure. Here a well-formed prosodic structure in (8a) is mismatched with a well-formed syntactic parsing in (8b) under the assumption that the head of PP (propositional phrase) must not be separately parsed from its complement NP.

(8) a. Well-formed prosodic structure
(Sesame Street is brought to you ***by), (****the Children's Television Workshop)*

b. Well-formed syntactic structure
*(Sesame Street is brought to you), (****by*** *the Children's Television Workshop)*

(from Shattuck-Hufnagel & Turk, 1996)

All these pieces of evidence taken together suggest that although syntax may influence the speaker's choice of prosody, a final construction of prosodic structure for a given utterance can be independent from syntactic structure. In the next sections, I will elaborate on how a prosodic structure may be selected, and how it may influence phonetic implementation. Before moving on, however, it is important to clarify that the term 'prosody' used in this paper does not simply refer to suprasegmental phonetic features such as pitch, duration and amplitude or intonational properties of the language whose variation signals prosodic boundaries and prominence distribution. As discussed in Shattuck-Hufgnael & Turk (1996) and Cho (2022), it must also refer to abstract prosodic structure itself that determines the phonological organization of speech units into higher-level prosodic constituents and their relative prominence within these constituents. In other words, prosody is defined as embracing these two phonetic and phonological (structural) aspects both of which must be simultaneously taken into account when investigating prosody within or across languages.

3 Prosodic Encoding

Thus far, I have discussed some core functions of prosodic structure or 'prosody.' It gives rise to speech variation on the surface. Such variation, however, is not a random noise that might arise with speech-internal factors such as the one that is inevitably created due to speaker-specific anatomical conditions. Rather, it reflects prosodic structure in relation to various speech-external factors that stem from either paralinguistic reasons or higher-order linguistic structures such as information structure and syntax. Other factors that also influence prosodic structure may include phonology, morphology, pragmatics and discourse (see Jun, 1993, Shattuck-Hufnagel & Turk, 1996, Cho, 2022 for related discussion).

The interplay between prosodic structure and other higher-order structural factors has implications for speech planning. From a perspective of speech planning, a selection of a particular prosodic structure for a given utterance can be considered as an end product after all these factors and their interactions as influencers on prosodic structure having been taken into account. Such a prosodic structure selection (building) process is related to *Prosodic Encoding*. The term is often used to indicate what particular higher-

order information is 'encoded' or reflected along some suprasegmental phonetic dimensions (e.g., pitch, duration, amplitude) of prosody as can be inferred in its use in 'prosodic encoding of information structure' (e.g., Kügler & Calhoun, 2021; Cole & Chodroff, 2020) or 'prosodic encoding of topic and focus' (e.g., Wang & Xu, 2011). But one can define Prosodic Encoding in concert with the structural view of prosody as discussed as the end of the previous section. That is, Prosodic Encoding may be defined as a process in speech planning that builds a final prosodic structure that must 'encode' multifaceted information stemming from a number of factors that marshal to influence prosodic structure.

This structural view of Prosodic Encoding also concerns realization of suprasegmental phonetic features as the prosodic structure building process entails specification of suprasegmental phonetic features that are needed to produce a planned prosodic structure. One such suprasegmental (prosodic) feature that may immediately come to mind is pitch as it plays an important role in determining two essential elements of prosodic structure—i.e., prominence distribution and boundary marking. This is in consistent with Autosegmental-Metrical Theory of Intonational Phonology (see Ladd, 2008 for a review) that assumes that prosodic structure is defined primarily by specification of tones or tone targets. The specification of tones is also an important part of ToBI conventions both in English and Korean. Thus, information about pitch at some level of detail must be contained in prosodic encoding process at least in the form of tonal targets.

What about duration? In ToBI conventions, the break indices imply that temporal information may be included in a prosodic structure in correlation with the strength of prosodic juncture. But the AM theory of Intonational Phonology does not specify the temporal information in phonological terms. In this framework, the temporal realization on the surface must be driven by some kind of phonetic implantation rule that translates the tonally-marked boundary and prominence of prosodic structure along the temporal dimensions. It is not, however, theoretically impossible to include durational information in Prosodic Encoding. Given the importance of temporal information in the phonetic realization of prosodic structure, I envision a successful model of Prosodic Encoding must specify durational features in some form.

One promising place to look for the inclusion of temporal information in Prosodic Encoding is Articulatory Phonology (Browman & Goldstein, 1992; Goldstein, Byrd & Saltzman, 2006; Byrd & Krivokapić, 2021). Articulatory Phonology assumes that phonological primitives are articulatory gestures that are defined in terms of both spatial and temporal dimensions along which articulatory movements occur. Byrd & Krivokapić (2021) provides an insightful discussion of how encoding of boundary and prominence information may be done by so-called 'modulation' gestures such as π-gesture and μ-gesture. The π-gesture (or 'prosody' gesture) is assumed to govern the temporal realization of gestures at prosodic junctures (Byrd & Saltzman, 2003)

whereas the μ-gesture modulates both the temporal and spatial realization of gestures under prominence (Saltzman, Nam, Krivokapić & Goldstein, 2008). For example, an assumed strength of prosodic juncture determines the activation level of π-gesture which modulates the rate of clock that controls the temporal realization of articulatory gestures—the stronger the prosodic juncture, the slower the articulatory movement. Note that the π-gesture itself is a non-track variable gesture that does not have a task to be realized in terms of vocal track constriction degree and location. It is a mere modulation gesture that overlaps with actual constriction gestures. On the other hand, the μ-gesture modulates both the spatial and temporal realization of articulatory gestures (or to be precise, there are two kinds of modulation gestures, temporal and spatial) in reference to the degree of prominence that is determined by lexical stress and phrasal pitch accent. The activation level of these modulation gestures, which determines the temporal realization of articulatory gestures, may be determined later in the production process after a planned prosodic structure is created or is specified at a stage where prosodic structure is created. This is a question beyond the scope of the current discussion, but it is an important one that is hoped to be answered by gesture-based theories of speech production.

Let us now assume that Prosodic Encoding returns a particular prosodic structure along with proper specification of suprasegmental features that are needed to signal boundary and prominence distribution. This entails another important question. Will the articulatory motor execution system for a planned utterance generate the identical speech output if the same prosodic structural information comes about as a result of prosodic encoding? Of course, this question is valid only when physically-conditioned speech variation that I discussed at the outset of this paper is effectively factored out. Let's consider (9a) in which a prosodic structure of the Korean utterance *ap*aka hɛmpʌkʌlil silʌhasjʌ* ('Daddy dislikes hamburgers') is specified with one IP with three APs. Given that an AP in Korean is quite narrowly defined primarily in terms of tones in Jun (2000), its suprasegmental phonetic implementation of the prosodic structural information illustrated in (9a) may appear to be straightforward, thus not causing too much variation on the surface.

But things are not as simple as they appear. Let's compare two actual phonetic outputs of the same sentence produced based on the same prosodic structure as specified in (9a). Their acoustic waveforms and spectrograms are given in (10). Even a quick eyeballing of (10) indicates that the two seemingly identical utterances that may be transcribed as having the same prosodic structure are notably different along some phonetic dimensions such as pitch range, amplitude and duration. Some of the noticeable differences are marked by numbered squares in the figures. Compared to (10b), (10a) is produced with longer duration and greater amplitude for the first syllable /a/ (①); shorter duration and lower amplitude for /s/ (②), longer duration and greater

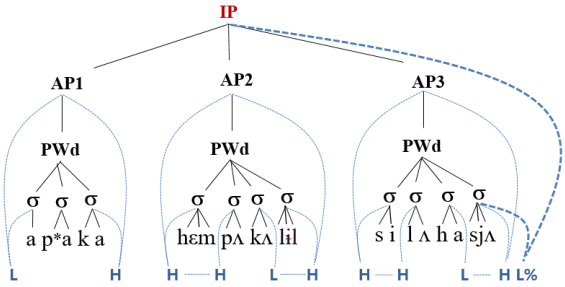
amplitude for the last vowel /jʌ/ (③), and a smaller pitch range for the second AP (④).

(9)

(a) ((아빠가) (햄버거를) (싫어하셔)) (One IP, Three APs)

((ap*aka) (hɛmpʌkʌlil) (silʌhasjʌ))

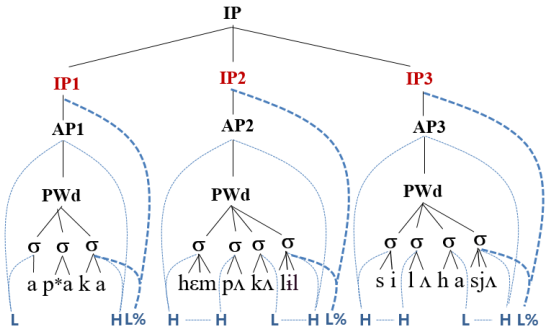
('Daddy'-NOM.) ('hamburgers'-ACC.) ('to dislike')



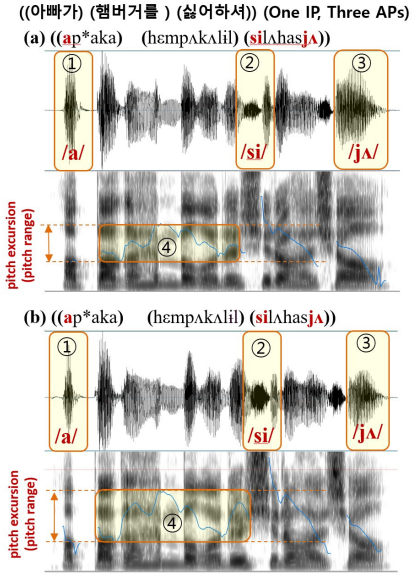
(b) ((아빠가)) ((햄버거를)) ((싫어하셔)) (Three IPs)

((ap*aka)) ((hɛmpʌkʌlil)) ((silʌhasjʌ))

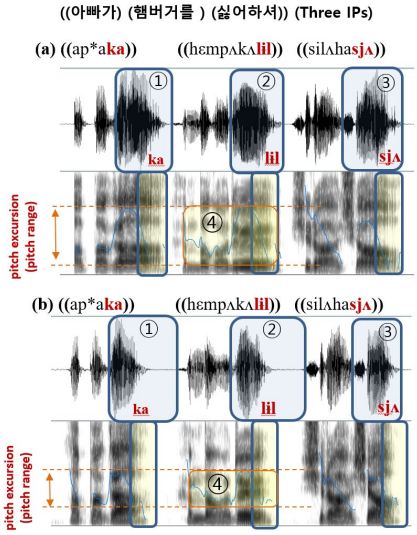
('Daddy'-NOM.) ('hamburgers'-ACC.) ('to dislike')



(10)



(11)



Discrepancy between utterances having the same prosodic specifications becomes more conspicuous with (9b) where the utterance has three IPs. Note that following Krivokapić and Byrd (2012), an IP is assumed to be recursive, so that an IP may govern one or more IPs (see also Ladd, 2008 for related

discussion). The acoustic waveforms and spectrograms of the two possible utterances of the same prosodic structure as specified in (9b) are given in (11). Here again, several substantial differences between the two utterances become immediately apparent. First, the final syllable of each IP (①-③), is longer in (11a) than in (11b). Second, the pause duration after the first and the second IP (①, ②) is shorter in (11a) than in (11b) (see Krivokapić, 2007, for further discussion on the relationship between phrasal length and pause duration). Third, the vowel (or the rhyme) of the IP-final syllable is more glottalized or creakier in (11a) than in (11b), as indicated by extended irregular periods evident in the spectrograms. Fourth, the pitch range for the second IP (④) is smaller in (11a) than in (11b).

4 Prosodic Strengthening and Fine Phonetic Detail

The acoustic-phonetic differences between the utterances illustrated in (10) and (11) in the previous section suggest that prosodic structure, as output of prosodic encoding (defined as such in the present discussion) does not suffice to account for fine-grained phonetic detail that differs between utterances with the same prosodic labelling. Moreover, there is a growing body of literature that demonstrates that the difference in fine-phonetic detail is not confined to suprasegmental phonetic features but it is reflected to a notable degree in the realization of segmental features (see Cho, 2016 for a general review and Cho, 2022 for a review on this issue in Korean). Variation in segmental realization that stems from prosodic structure has been discussed under the rubric of *prosodic strengthening.* Prosodic strengthening is used as a cover term to refer to the strengthening of realization of segmental features in marking two important elements of prosodic structure, boundary and prominence. While the reader is referred to Cho (2022) for a review on prosodic strengthening in Korean, I will briefly recapitulate some of the previous findings for an illustration of this point in Korean. (12) summarizes segmental variation due to prosodic structure at the left edge of a prosodic constituent known as *domain-initial strengthening*, and (13) summarizes focus-induced prominence effects.

(12) *Domain-initial strengthening in Korean* (boundary-related prosodic strengthening at the left edge of prosodic constituent)
   a. Alveolar stop and nasal consonants are produced with an increased constriction between the tongue tip/blade and the palate (as measured with an electropalatography (EPG) system) in domain-initial position (at the left edge of a larger prosodic constituent such as an Intonational Phrase) compared to the same segments that occur in domain-medial position (at the left edge of a smaller prosodic constituent such as a Prosodic Word) (Cho & Keating, 2001).

b. Both lenis and aspirated stops /p, p^h/ are produced with a longer VOT and a larger amount of airflow domain-initially than domain-medially (Cho & Jun, 2000).
c. The consonantal lip closing movement from a preceding vowel to a word-initial bilabial consonant, as measured with an Electromagnetic Articulograph (EMA), is larger in displacement, and longer in movement duration and slower in movement velocity in domain-initial position, compared to the same movement that occurs in domain-medial position (Cho, Son & Kim, 2016).
d. The F1-F2 acoustic vowel space as measured by an /i/-/a/-/u/ Euclidean area is expanded domain-initially (Cho, Lee & Kim, 2011).
e. Nasal consonants in NV are produced with reduced nasality (or tend to be denasalized) domain-initially, which is interpreted as suggesting a decrease of sonority but an increase of consonantality for an initial consonant due to domain-initial strengthening (Cho & Keating, 2001; Jang, Kim & Cho, 2018).
f. Vowels in NV are nasalized less domain-initially than domain-medially in line with the reduced nasality of the domain-initial nasal consonant (Jang, et al., 2018).

(13) Prominence-related prosodic strengthening induced by focus
a. The F1-F2 acoustic vowel space as measured by an /i/-/a/-/u/ Euclidean area is expanded under narrow focus. This focus-induced vowel space expansion is largely similar to the boundary-induced (domain-initial) expansion of vowel space described in (12d), although the exact directionality of the expansion is not the same (Cho, et al., 2011).
b. The aspirated stop /p^h/ is produced with a long VOT in the focused condition than in the unfocused one (Cho, et al., 2011), similar to domain-initial strengthening effect on VOT described in (12b).
c. Nasal consonants are produced with reduced nasality in the focused condition than in the unfocused one, similar to the boundary-related domain-initial strengthening effect. Again, in line with the reduced nasality of the nasal consonant, focus induces a reduction of coarticulatory vowel nasalization in both CVN and NVC context (Jang et al., 2018). The authors interpreted the results as suggesting a prominence-induced coarticulatory resistance that enhances the vowel's [oral] feature as was also found in English (Cho, Kim & Kim, 2017).
d. Both lenis and aspirated stops are produced with higher F0 on the following vowel in the focused condition than in the unfocused one,

although the effect size is larger for aspirated stops (Choi, Kim & Cho, 2020).[1]

5 Phonetic encoding of prosodic structure

In the previous two sections, two important points were made. First, utterances produced with a particular prosodic structure that is built up during a prosodic encoding process may still differ in fine phonetic detail. Second, prosodic structure is signaled by an array of both suprasegmental and segmental phonetic features, so that the speaker produces an utterance according to a planned prosodic structure, by adjusting not only suprasegmental phonetic features, but also segmental realization as discussed above. Let's now return to the question that I raised earlier. Does the same prosodic structure that comes about as a result of Prosodic Encoding lead the articulatory motor execution system to generate the same phonetic output? We now have compelling evidence that the answer is no. Given the phonetic granularity that arises along both the suprasegmental and segmental dimensions with a particular prosodic structure, the phonologically-informed symbolic representation of prosody specified with tones and possibly with temporal features does not suffice to capture the fine-grained phonetic detail on the surface. This means that before a particular prosodic structure as output of prosodic encoding in a planning process is fed into the articulatory motor execution system, it must go through another process that mediates between prosodic encoding and phonetic implementation. The process can be called *Phonetic Encoding*. This term has been used in psycholinguistic models of speech production most heavily by Levelt and colleagues (Levelt, 1989; Levelt, Roelofts & Meyer, 1999). In traditional terms, Phonetic Encoding refers to a process in which phonological forms (as output of Phonological Encoding or Phonological Spell-out) are fleshed out with specific phonetic content to give rise to the specific phonetic form on the surface.

As discussed in depth in Keating & Shattuck-Hufnagel (2002) and Keating (2006), Phonetic Encoding in Levelt's model is rather limited as it does not take into account prosodic-structurally conditioned speech variation that we have discussed so far. Keating & Shattuck-Hufgnael (2002) propose that

[1] The increase of F0 for the lenis stop, which is often hypothesized to be phonologically specified with a low tone in the present-day Seoul Korean, stands in sharp contrast with the focus-induced phonetic enhancement of the L tone found in Mandarin Chinese (a tone language; cf. Chen & Gussenhoven, 2008) and South Kyungsang Korean (a lexical pitch accent language; cf. Cho, Kim & Kim, 2019), in both of which a tone concerns directly the phonological/lexical contrast. Based on this observation, Cho (2022) proposes that the L and H tones that may serve as acoustic correlates of the lenis-aspirated stop contrast are post-lexical tonal properties conditioned by prosodic structure, rather than what have emerged as phonological features due to a tonogentic sound change.

Phonetic Encoding must refer to prosodic structure, provided that prosodic structure is available earlier in the planning process. Keating (2006) provides further insights into this issue under the rubric of *Phonetic Encoding of Prosodic Structure*. This term, perhaps first theoretically elaborated by Keating (2006), can be broadly used to reiterate the fact that the detailed phonetic plan as output of Phonetic Encoding must contain sufficient information about the fine-grained phonetic variation as a function of prosodic structure. The phonetic plan will thus contain enough information to guide the articulatory motor execution system in producing an utterance as planned by the speaker. Keating concludes:

> "In summary, when a speaker plans for the phonetic aspects of speech production, prosodic structure organizes the treatment of possibly every feature in every segment, and the interactions of segments. One aspect of this dependence is the relation between the strength of a prosodic position, and the phonetic strength of a segment in that position. A theory of phonetic encoding that incorporates this basic fact is a major challenge, but an important one." (Keating, 2006: 183)

Is this the end of the story? The answer is only partly yes. We are still left with a question unsolved. Does Phonetic Encoding, even if it refers to prosodic structure, account for fine phonetic detail that may vary between utterances that are specified with the same prosodic structure generated in the planning process? More specifically, questions related to phonetic differences of the sort shown in (10) and (11) must be answered adequately. For example, where and how in the planning process is the durational difference of the pause determined for two utterances that are assumed to have the same prosodic structure? The same question applies to other differences such as in pitch range, voice quality (e.g., degree of creakiness or glottalization), the degree of phrase-final lengthening, and so on and so forth. As discussed above, in current AM theories of prosodic structure and intonational phonology, prosodic structure is rather coarsely defined, in such a way that it provides no principled predictions on the phonetic granularity beyond what may be predicted based on coarse prosodic specifications available in a prosodic structure.

One way to address this issue is by devising a way that Phonetic Encoding of Prosodic Structure includes the subtle phonetic granularity as intended by the speaker. In such a scheme, one can still maintain the basic tenet of AM theory—i.e., phonetic encoding of prosodic structure must refer to the abstract prosodic structure whose prosodic specifications are coarsely defined, but it should also have some kind of a built-in device or mechanism in the process of Phonetic Encoding that fine-tunes phonetic realization in reference to any other factors that contributes to building the granularity of the phonetic realization. Such a model is in fact reminiscent of a proposal made by Mücke & Grice (2014) that the effect of focus driven by information structure in

German is not mediated by accentuation that refers to prominence distribution of prosodic structure, but it directly influences phonetic realization in reference to information structure. More generally, in a speech planning model that I envisage, as I discussed in Section 2.1, such a fine-tuning of phonetic realization should occur at the phonetic component of the grammar which modulates the Phonetic Encoding process at the level of fine phonetic detail and in a language-specific way (cf. Keating, 1984, 1990; Cho & Ladefoged, 1999; Cho, et al., 2019).

Another way to address the question is by assuming that prosodic structure itself is defined in a gradient way, so that once a prosodic structure is initially generated at some point in the planning process possibly in reference to syntax (see Keating & Shattuck-Hufnagel, 2002 for related discussion), it may continue to be refined in a gradient fashion in reference to other factors that might influence prosodic structuring. After all the factors having been taken into account, a final prosodic structure is fed into the phonetic plan with all the necessary information to guide the motor execution system to generate the final surface phonetic form of the utterance as planned by the speaker. A gradient view of prosodic structure is adopted by the π-gesture model where boundaries are represented gradiently rather than categorically (Byrd & Saltzman, 2003; see Byrd & Krivokapić, 2021, for related discussion), but it remains to be seen how prosodic structure as a whole can be represented in gradient terms.

6 Conclusion

In the present review, I have discussed how low-level speech variation along both segmental and suprasegmental dimensions may be related to higher-order linguistic structures such as information structure and syntax. The surface phonetic form of an utterance must therefore contain information that comes down from higher-order linguistic structures. It is underscored that prosodic structure plays a central role in shaping the surface phonetic form of an utterance in reference to such higher-order structural information. The phonetics-prosody interface (a fine-tuning of phonetic realization in reference to prosodic structure) is not as simple as it may look as it must take into account phonetic variation that cannot be captured by a coarsely-defined prosodic structure in currently prevalent linguistic theories of prosody. From a perspective of speech planning, prosodic structure, determined by Prosodic Encoding, must be encoded into the speech signal (Phonetic Encoding of Prosodic Structure) in such a way to reflect the phonetic granularity which is beyond what phonological descriptions of prosody can account for but is critical in decoding a linguistic message from the variable speech signal. Admittedly, the discussion made here leaves more questions than answers regarding the intricate interaction between phonetics and prosodic structure. But I am

certain that it will spark further research on detailed mechanisms of the phonetics-prosody interface in relation to various levels of linguistic structure, which will help to crystalize the role of prosodic component in the general architecture of the grammar of Korean and of other languages.

Acknowledgements

This review article was written in honor of Pat Keating whose work on phonetic grammar and phonetic encoding of prosodic structure had inspired a great deal of discussion made here, but all errors in conceptualizing her work remain my own. I would like to thank Sahyang Kim for her insightful comments on an earlier version of this ms. This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2021S1A5C2A02086884)

References

Arvaniti, A., Ladd, D. R. and Mennen, I. 2000. What is a starred tone? Evidence from Greek. *Papers in Laboratory Phonology V: Acquisition and the Lexicon*, eds. M. Broe and J. Pierrehumbert, 119-131. Cambridge: Cambridge University Press.

Atterer, M. and Ladd, D. R. 2004. On the phonetics and phonology of 'segmental anchoring' of F0: evidence from German. *Journal of Phonetics* 32: 177-97.

Beckman, M. E. 1996. The parsing of prosody. *Language and Cognitive Processes* 11: 17-67.

Beckman, M. E. and Ayers, G. M. 1994. Guidelines for ToBI Labelling. Online manuscript and files Available at http://www.ling.ohio-state.edu/phonetics/E_ToBI.

Beckman, M. E., Hirschberg, J. and Shattuck-Hufnagel, S. 2005. The Original ToBi System and the Evolution of the ToBi Framework. *Prosodic Typology*, ed. S.-A. Jun, 9-54. Oxford: Oxford University Press.

Browman, C. and Goldstein, L. 1992. Articulatory phonology: An overview. *Phonetica* 49: 155-180.

Byrd, D. and Krivokapić, J. 2021. Cracking Prosody in Articulatory Phonology. *Annual Review of Linguistics* 7: 31–53.

Byrd, D. and Saltzman, E. 2003. The elastic phrase: modeling the dynamics of boundary-adjacent lengthening. *Journal of Phonetics* 31: 149–180.

Chen, Y and Gussenhoven, C. 2008. Emphasis and tonal implementation in Standard Chinese. *Journal of Phonetics* 36: 724–746.

Cho, T. 2015. Language effects on timing at the segmental and suprasegmental levels. *The Handbook of Speech Production*, ed. M. Redford, 505–529. Hoboken, NJ: Wiley-Blackwell.

Cho, T. 2016. Prosodic boundary strengthening in the phonetics-prosody interface. *Language and Linguistics Compass* 10(3): 120-141.

Cho, T. 2022. The phonetics-prosody interface and prosodic strengthening in Korean. *The Cambridge Handbook of Korean Linguistics*, eds. S. Cho and J. Whitman, 248-293. Cambridge: Cambridge University Press.

Cho, T. and Jun, S. 2000. Domain-initial strengthening as featural enhancement: Aerodynamic evidence from Korean. *Chicago Linguistics Society* 36: 31-44.

Cho, T. and Keating, P. 2001. Articulatory and acoustic studies of domain-initial strengthening in Korean. *Journal of Phonetics* 29: 155-190.

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

Cho, T., Kim, D. and Kim, S. 2019. Prosodic strengthening in reference to the lexical pitch accent system in South Kyungsang Korean. *The Linguistic Review* 36(1): 85-116.

Cho, T. and Ladefoged, L. 1999. Variation and universals in VOT: evidence from 18 languages. *Journal of Phonetics* 27: 207-229.

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

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

Cho, T., Whalen, D. H. and Docherty, G. 2019. Voice onset time and beyond: Exploring laryngeal contrast in 19 languages. *Journal of Phonetics* 72: 52-65.

Choi, J., Kim, S. and Cho, T. (2020). An apparent-time study of an ongoing sound change in Seoul Korean: A prosodic account. *PloS ONE* 15(10): e0240682, 1-29. https://doi.org/10.1371/journal.pone.0240682

Cole, J. and Chodroff, E. (2020). Prosodic encoding of information structure in nuclear and prenuclear positions in American English. *The Journal of the Acoustical Society of America* 148: 2725. (Abstract).

Dellwo, V., Huckvale, M. and Ashby, M. 2007. How is individuality expressed in voice? An introduction to speech production and description for speaker classification. *Speaker Classification I: Fundamentals, Features, and Methods*, ed. Ch. Mueller, 1-20. Berlin, Heidelberg: Springer.

Goldstein, L., Byrd, D., & Saltzman, E. 2006. The role of vocal tract gestural action units in understanding the evolution of phonology. *Action to Language via the Mirror Neuron System*, ed. M. Arbib, 215-249. New York: Cambridge University Press.

Gussenhoven, C. 2008. Types of focus in English. *Topic and Focus. Studies in Linguistics and Philosophy, vol. 82,* eds. C. Lee, M. Gordon and D. Büring, 83-100. Dordrecht, The Netherlands: Springer.

Jang, J., Kim, S. and Cho, T. 2018. Focus and boundary effects on coarticulatory vowel nasalization in Korean with implications for cross-linguistic similarities and differences. *The Journal of the Acoustical Society of America* 144: EL33-39.

Jun, S.-A. 1993. *The Phonetics and Phonology of Korean Prosody.* Ph.D. dissertation, Ohio State University.

Jun, S.-A. 2000. K-ToBI (Korean ToBI) Labelling Conventions (Version 3.1). Online manuscript and files available at http://www.linguistics.ucla.edu/people/jun/ktobi/K-tobi.html.

Jun, S.-A. 2007. The intermediate phrase in Korean Intonation: Evidence from sentence processing. *Tones and tunes: Studies in word and sentence prosody*, eds. C. Gussenhoven & T. Riad, 143–167. Berlin: Mouton de Gruyter.

Jun, S.-A. 2011. Prosodic markings of complex np focus, syntax, and the pre-/post-focus string. *Proceedings of the 28th West Coast Conference on Formal Linguistics,* eds. M. B. Washburn, K. McKinney-Bock, E. Varis, A. Sawyer, and B. Tomaszewicz, 214–230. Somerville, MA USA: Cascadilla Proceedings Project.

Jun, S.-A. 2014. Prosodic typology: by prominence type, word prosody, and macrorhythm. *Prosodic Typology II: The Phonology of Intonation and Phrasing*, ed. S.-A. Jun, 520–539. Oxford: Oxford University Press.

Keating, P. A. 1984. Phonetic and phonological representation of stop consonant voicing. *Language* 60: 286–319.

Keating, P. A. 1985. Universal phonetics and the organization of grammars. *Phonetic Linguistics: Essays in Honor of Peter Ladefoged*, ed. V.A. Fromkin, 115–132. Orlando FL: Academic Press.

Keating, P.A. 1990. Phonetic representations in a generative grammar. *Journal of Phonetics* 18: 321-334.

Keating, P. A. 2006. Phonetic encoding of prosodic structure. *Speech Production: Models, Phonetic Processes, and Techniques*, eds. J. Harrington and M. Tabain, 167-186. New York and Hove: Psychology Press.

Keating, P. A. and Shattuck-Hufnagel, S. 2002. A prosodic view of word form encoding for speech production. *UCLA Working Papers in Phonetics* 101: 112-156.

Krivokapić. J. 2007. Prosodic planning: Effects of phrasal length and complexity on pause duration. *Journal of Phonetics* 35: 162-179.

Krivokapić, J. and Byrd, D. 2012. Prosodic boundary strength: An articulatory and perceptual study. *Journal of Phonetic*s 40: 430–442.

Kügler, F. and Calhoun, S. 2021. Prosodic encoding of information structure: A typological perspective. *The Oxford Handbook of Language Prosody*, eds. C. Gussenhoven and A. Chen, 454-467. Oxford: Oxford University Press.

Ladd, D. R. 2008. *Intonational Phonology* (2nd edition). Cambridge, UK: Cambridge University Press.

Ladd, D. R., Schepman, A., White, L., Quarmby, L. M. and Stackhouse, R. 2009. Structural and dialectal effects on pitch peak alignment in two varieties of British English. *Journal of Phonetics* 37: 145-161.

Levelt, W. J. M. 1989. *Speaking: From Intention to Articulation.* Cambridge MA: MIT Press.

Levelt, W. J. M., Roelofs, A. and Meyer, A. S. 1999. A theory of lexical access in speech production. *Brain and Behavioral Sciences* 22(1): 1-38.

Levi, S. V. and Pisoni, D. B. 2007. Indexical and linguistic channels in speech perception: Some effects of voiceovers on advertising outcomes. *Psycholinguistics Phenomena in Marketing Communications,* ed. T. M. Lowrey, 203–219. Mahwah: Lawrence Erlbaum.

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

Nespor, M., and Vogel, I. 1986. *Prosodic phonology.* Dordrecht: Foris Publications.

Pierrehumbert, J. and Beckman, M. 1988. *Japanese Tone Structure.* Cambridge, MA: MIT Press.

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

Saltzman, E., H. Nam, J. Krivokapić, and L. Goldstein. 2008. A task-dynamic toolkit for modeling the effects of prosodic structure on articulation. *Proceedings of the Speech Prosody 2008 Conference*, eds. P. A. Barbosa, S. Madureira, and C. Reis. Campinas, Brazil.

Selkirk, E. 1984. *Phonology and Syntax: The Relation between Sound and Structure.* Cambridge: MIT Press

Selkirk, E. 1995. Sentence prosody: intonation, stress, and phrasing. *The Handbook of Phonological Theory*, ed. J. A. Goldsmith, 550–569. Oxford: Blackwell.

Wang, B. and Xu, Y. 2011. Differential prosodic encoding of topic and focus in sentence-initial position in Mandarin Chinese. *Journal of Phonetics* 39: 595–611.