HOW DOES FOCUS-INDUCED PROMINENCE INFLUENCE REALIZATION OF EDGE TONES AND SEGMENTAL ANCHORING IN SEOUL KOREAN?

Richard Hatcher[1], Hyunjung Joo[1,2], Sahyang Kim[1,3], & Taehong Cho[1]

[1]Hanyang Institute for Phonetics & Cognitive Sciences of Language (HIPCS), Hanyang University, [2]Rutgers University, [3]Hongik University

richard.j.hatcher.jr@gmail.com, hyunjung.joo@rutgers.edu, sahyang@hongik.ac.kr, tcho@hanyang.ac.kr

ABSTRACT

This study investigated how focus-induced prominence influences the global F0 contour within an Accentual Phrase (AP) embedded in an Intonational Phrase in Seoul Korean. Since an AP typically contains edge tones (#LH…LH#), it was tested how realization of underlying LH is conditioned by focus. Monosyllabic target words occurred phrase-initially/finally, and the coda sonorancy was varied (*pam* vs. *pap*) to examine whether/how the tone bearing ability would influence tonal realization. Results showed that focus on phrase-initial words induced pitch range expansion (LH), evident in H-peak anchored early in the post-focal word. But phrase-finally focus induced no F0 expansion but slightly delayed L%, showing an interaction with boundary tones. The coda sonorancy yielded local F0 perturbation, but not global F0 modification. Our results suggest that the segmental anchoring and range of phonologically-defined tones are systematically modulated by prominence, illuminating an interplay between tonal and segmental realizations in reference to information structure.

Keywords: edge tones, focus-induced prominence, segmental anchoring, intonation, Korean

1. INTRODUCTION

Linguistic prominence is often realized by the intonational contour. Languages like English and German are classified as head-prominence languages, as linguistic prominence relies on pitch-accented syllables serving as head of the prosodic phrase [15, 14, 9] In these languages, when prominence is licensed by focus, it is realized in the speech signal by expanded pitch range on the focused (and pitch accented) elements and subsequent deaccenting or pitch compression on the post-focal elements [13]. Other languages, such as Korean, lack the accent system and are said to employ prosodic phrasing in order to mark prominence. In such an edge prominence system, focus is often assumed to trigger insertion of a prosodic boundary, with focused words aligned to the left edge of a prosodic phrase [10, 6].

Setting aside the theoretical issue of whether there should be a one-to-one mapping between focus and phrasing in Korean, it is clear that edges of a prosodic phrase are marked by edge tones of some sort in such an edge-prominence language. What is less clear, however, concerns whether and how prominence may influence the tonal realization of both edge tones and non-edge tones in shaping the global tune of an intonational structure in a language. The present study attempts to fill this gap by investigating the tonal manifestation of focus-induced prominence and its influence on the intonational landscape of the Accentual Phrase (AP) embedded in an Intonational Phrase (IP) in Seoul Korean.

The Korean AP is intonationally defined by two rising contours (/#LH…LH#) at its edges, although they may or may not be fully realized depending on the number of syllables within an AP as well as other (non-)linguistic factors. (While the identity of the initial tone is conditioned by the laryngeal status of the AP-initial segment, we will only focus on the initial L tone in this study assigned for sonorants and lenis obstruents.) As shown in Figure 1, an AP is embedded in a larger prosodic phrase, the Intonational Phrase. While the tonal composition of an AP directly contributes to the form of an IP, especially at its left edge, the IP-final boundary tone (marked by %) overrides the right-aligned AP edge tone.

Previous studies of prominence in Korean intonation broadly agree that focus (especially the narrow or contrastive focus) is expressed by phrasing, followed by optional dephrasing of any post-focal materials [5, 12, 4]. But analyses may differ on the precise level of the phrase boundary initiated by focus. There are accounts for the Accentual Phrase [5, 6], the Intermediate Phrase [7, 8], and the Intonational Phrase [4]. In addition,

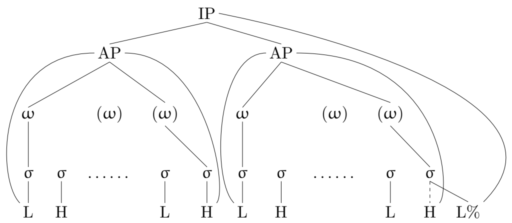


Figure 1: Intonational structure of Seoul Korean following Jun [6]

focus is also described as influencing the phonetic realization of phrasal tones, whereby focused H tones were realized with higher F0 [4].

The purpose of the present study is not to attempt to contribute to these theoretical accounts, but to provide more empirical data to illuminate the nature of the relationship between prominence and tonal realization. We therefore investigate in more detail how focus effects on the F0 contour are constrained by the intonational grammar of Korean. We specifically focus on how focus-induced F0 range expansion is further conditioned by the prosodic position of the focused elements, and how edge tones are anchored with segments and further interact with boundary tones under prominence.

2. METHODS

Data were collected from 14 native speakers of Seoul Korean (7 female, 7 male) in their twenties. Participants were required to have been born and reside in Seoul.

There were two monosyllabic target words varying in the sonority of their coda (*pam* "chestnut; night", *pap* "cooked rice"). Table 1 provides an example set of question-answer pairs for target word *pam*. In order to induce corrective focus, participants were asked to make a contrast between the words in bold in Sentences A (a question) and B (a corrective answer to A). To control for the possible effects of phrasing, target sentences (B in Table 1) were constructed to consist of three words[1] (5 syllables total) with the target word occupying either the phrase-initial or -final position. The phrase position of the test words were varied to investigate how the effects of F0 contour change in different prosodic position.

In the experiment, participants were presented with a mini dialogue via a visual aid provided on the computer screen. The participants were asked to read the target sentence (as Speaker B) in response to the prime sentence presented auditorily as well as visually. Participants were asked to produce the sentence naturally without pausing between the

Initial	Focus	A:	*ipʌn tanʌnɨn* ***kuk*** *twiɛta nwa?*
			Should I put this word behind the soup?
		B:	*ani.#**pam** twiɛta nwa.*
			No. Put it behind the **chestnut**.
	No Focus	A:	*ipʌn tanʌnɨn pam* ***apʰɛta*** *nwa?*
			Should I put this word on the chestnut?
		B:	*ani.#pam* ***twiɛta*** *nwa.*
			No. Put it **behind** the chestnut?
Final	Focus	A:	*ipʌn tanʌnɨn nʌhɨ͡i ʌnni* ***kʰoŋini?***
			Is this one (for) your sister's bean?
		B:	*ani.#uri ʌnni* ***pam.****#twiɛta nwa.*
			No. Our sister's **chestnut**. Put it under.
	No Focus	A:	*ipʌn tanʌnɨn nʌhɨ͡i* ***op*a*** *kʰoŋini?*
			Is this (for) your brother's chestnut?
		B:	*ani.#uri* ***ʌnni*** *pam.#twiɛta nwa.*
			No. Our **sister's** chestnut. Put it under.

Table 1: An example set of test words with varying focus and boundary conditions.

words.

The recording was made in a sound-attenuated booth, using a SHURE KSM44 condenser microphone and a Tascam US-4x4 digital recorder at a sampling rate of 48 kHz. In total, 3360 tokens were collected (3 items x 2 boundaries x 2 focus types x 20 repetitions x 14 speakers). From these, 90 tokens were discarded due to incorrect phrasing and an additional 79 tokens were excluded for incorrect or ambiguous focusing.

The acoustic recordings were force-aligned via the Montreal Forced Aligner [16], using a pre-trained Korean model [17]. An additional tier was created in Praat [1] segmenting only the sonorant portions of words capable of bearing F0.

Acoustic measurement of F0 was computed at nine equidistant time-points across the sonorant portion of the word in VoiceSauce [19], using the STRAIGHT algorithm [11]. Prior to statistical analysis, F0 was z-score normalized to address interspeaker differences in pitch range and height. Generalized additive mixed models (GAMM) [21] were constructed using the `bam` function of the *mgcv* package [20] in R [18] adjusting for autocorrelation in the time dimension by specifying a ρ value corresponding to the autocorrelation function at lag $= 1$. To account for the interaction of focus on separate words of the phrase, an interaction factor was created combining factors Word and Focus. This interaction term was then used as fixed effect and as smoothing term. Full random smooths were added for Speaker.[2]

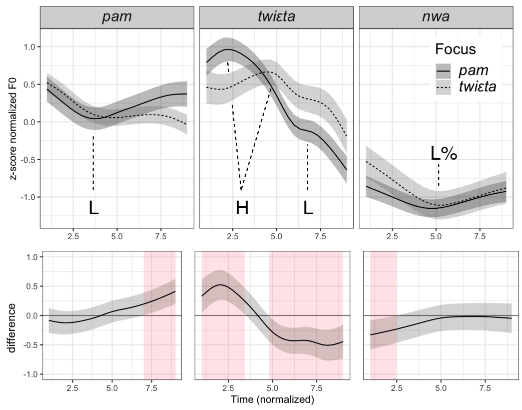
3. RESULTS

3.1. Phrase-initial Target words

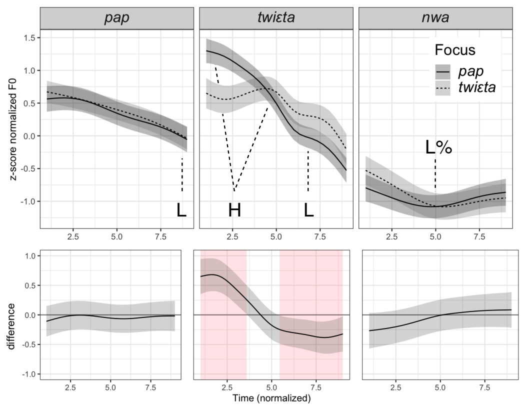
In order to investigate differences between phrase-initial vs -medial focus, a GAMM was fit on the phrase-initial subset of the corpus. The subset yielded a total of 1600 observations and 42,740 data points (28,944 for *pam*, 13,796 for *pap*).

GAMM smooth and difference plots of the F0 contour for phrase-initial target words are shown in Figure 2. These smooth plots were created with the `plot_smooths` function of the *tidymv* R package [3]. The figures are organized such that beneath each GAMM's smooth plot is provided the difference plot indicating which portion of the two curves are significantly different.

The plots for both phrases containing *pam* and *pap* evidence substantial differences in the F0 curve of the phrase-medial word *twiɛta* "behind." For both, phrase-initial focus results in an earlier-anchored, higher-scaled peak in comparison to the corresponding phrase-medial focus. The F0 contour of the initial word evidences only a small difference by focus conditions for *pam*. Phrase-initial *pap*, however, bears the same contour regardless of its prominence status.



(a) *pam*



(b) *pap*

Figure 2: Visualization of the non-linear smooths and difference plots for phrase-initial target word contexts. The pointwise 94%-confidence intervals are shown as grey ribbons. The pink vertical bars represent portions of the curve where the two focus conditions significantly differ.

The difference in the predicted contours of the two target words is perhaps exaggerated by the effects of time-normalization. The sonorant portion of *pam* was approximately twice as long as that of *pap* on average. If we restrict our attention to the first half of the contour for *pam*, the shape of the contour is comparable to that of *pap*, i.e. a fall to a local F0 minimum. The shortened contour of *pap*, however, does not result in a lowering or a shift in the anchoring of the following peak. Instead, the local F0 extrema occur with no explicit rising contour as illustrated in the upper left two panels of Figure 2b.

The effect of the target word's coda on the F0 of the following word is shown in Figure 3. Focused *pap* induces a higher initial F0 value in the following word. This may due to local perturbation resulting from the *post-obstruent tensing* rule in Korean whereby the following lenis obstruent is tensified due to the presence of the preceding obstruent (i.e., /pap twi/ > [pap t*wi]) within the same AP. Tensified consonants yield higher F0 due to their laryngeal status.

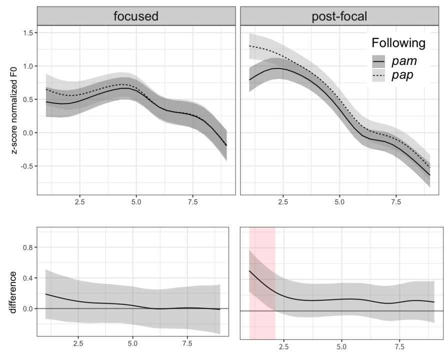


Figure 3: Visualization of non-linear smooths an difference plots of phrase-medial *twiɛta* under focal (left) and post-focal (right) contexts. The pointwise 95%-confidence intervals are shown as grey ribbons. The pink vertical bars represent portions of the curve where the two focus conditions significantly differ.

3.2. Phrase-final Target words

Another GAMM was fit to the subset of the corpus where the test words occupy phrase-final position. In these sentences, focus occurs either on the target word phrase-finally or on the immediately preceding word *ʌnni* "older sister." This subset provided a total of 1,591 observations and 42,948 data points (28,674 for *pam*, 14,274 for *pap*).

GAMM smooth and difference plots of the F0 contour for phrase-final target words are provided in Figure 4. These plot the model's non-linear predictions for the F0 contours for phrase-final (dark line) and phrase-medial (light dotted line) focus. Beneath the smooth plots are situated difference plots illustrating which portion of a word's F0 contour differs significantly by focus.

Phrase-medial focus on *ʌnni* results a substantial F0 rise over the disyllabic word. An effect of prominence-induced scaling is also manifest on the initial word *uri* "our," which displays a lowering effect at its right edge pre-focally perhaps suggesting tonal coarticulation with the following tone.

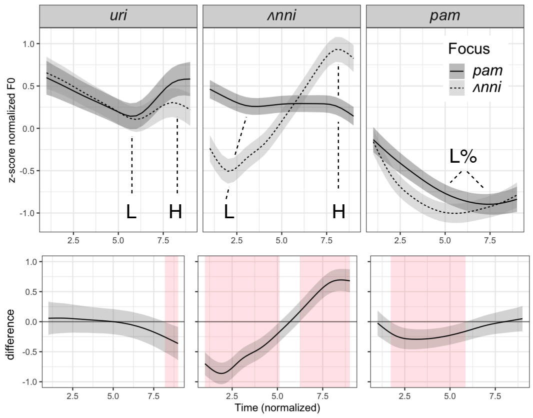
Phrase-final focus differs from initial and medial focus in the absence of the typical rise associated with AP-initial position. Instead, both focal and post-focal phrase-final words exhibit a fall, whose exact shape differs according to focus. For both *pam* and *pap*, the F0 minimum occurs later under focus. This is especially apparent for *pap*, which consists solely of a fall; the trough occurring at the phrase edge.

In contrast to the raising of the peak under initial focus, there appears to be no scaling effect of focus on the phrase-initial trough. For both words regardless of focus, the GAMM predicts an F0 minimum of zero, i.e. a speaker's mean F0 value. The trough of post-focal words, assumed to represent the boundary tone occurs earlier, thereby permitting the possibility of a small rise at the right edge.

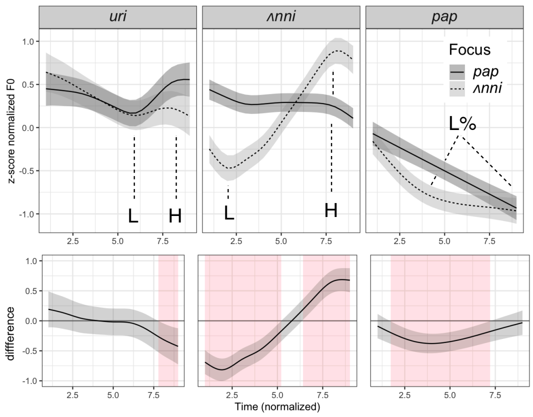
4. DISCUSSION

These results clearly show that the F0 manifestation of focus-induced prominence is sensitive to phrase position. Monosyllabic phrase-initial words incapable of bearing both tones necessary for the AP-initial rise show little effect of prominence on F0. Notably, however, the effect occurs on the post-focal word. The domain for prominence-induced scaling of F0, therefore, is not to the focused word *per se*, but may be the entire AP including any dephrased material. The scaling of the AP-initial L tone was not effected by focus in our data, contrasting with the results from South Kyungsang Korean [2], which was shown to exhibit a "polarizing" of initial pitch accents under focus. Note, however, the prosodic system of South Kyungsang differs crucially in bearing lexically-specified pitch accents as opposed to the phrasally specified tones of Seoul Korean.

Phrase-final focus differs from other positions in lacking an F0 rise. Instead final words evidence a fall anchored late in the word. The late anchoring of the L may perhaps represent the dual association of AP initial L tone and IP-final boundary tone. This behavior suggests a more complex relationship between the boundary tone and AP tones than that of simply overriding.



(a) *pam*



(b) *pap*

Figure 4: GAMM for phrase-final target words (a) *pam* and (b) *pap*. The pointwise 95%-confidence intervals are shown as grey ribbons. The pink vertical bars represent portions of the curve where the two focus conditions significantly differ.

Finally, it is worth noting that coda sonorancy plays a role in determining the shape of the contours investigated above. In both contexts investigated, the F0 contour of *pap* consisted solely of a fall as shown in the top left panels of Figs 2b & 4b. The phrase-initial trough representing the L tone was anchored with the vowel offset, with the sonorant coda of *pam* interpolating to the peak associated with the following syllable. While these effects are attributed to the tone bearing ability of the coda, what remains invariant is the timing of L and H targets regardless of whether there is an obstruent gap or not. Phrase finally, the difference seems to bear on the phonological system, restricting the choice of boundary tone (e.g., better accommodating a complex tone LH% with the sonorant coda.)

In conclusion, the present study has shown that tonal distribution in Seoul Korean is influenced by both its intonational grammar and low-level phonetic differences in consonant sonorancy. Prominence may interact with boundary-induced strengthening and the existence or absence of phrase-final boundary tones.

5. ACKNOWLEDGEMENTS

This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2021S1A5C2A02086884) awarded to TC (PI) and SK (Co-PI).

6. REFERENCES

[1] P. Boersma and D. Weenick, "Praat: Doing phonetics by computer," Feb. 2021.

[2] T. Cho, D. J. Kim, and S. Kim, "Prosodic strengthening in reference to the lexical pitch accent system in South Kyungsang Korean," *The Linguistic Review*, vol. 36, no. 1, pp. 85–115, Feb. 2019.

[3] S. Coretta, J. v. R. I. functions from itsadug), and M. W. I. functions from itsadug), "Tidymv: Tidy Model Visualisation for Generalised Additive Models," Sep. 2022.

[4] H.-S. Jeon and F. Nolan, "Prosodic Marking of Narrow Focus in Seoul Korean," *Laboratory Phonology*, vol. 8, no. 1, p. 2, Jan. 2017.

[5] S.-A. Jun, *The Phonetics and Phonology of Korean Prosody: Intonational Phonology and Prosodic Structure*, ser. Outstanding Dissertations in Linguistics. New York ; London: Garland Pub, 1996.

[6] ——, "Korean Intonational Phonology and Prosodic Transcription," in *Prosodic Typology: The Phonology of Intonation and Phrasing*. Oxford University Press, 2005.

[7] ——, "The intermediate phrase in Korean: Evidence from Sentence Processing," in *Tones and Tunes: Experimental Studies in Word and Sentence Prosody*, T. Riad and C. Gussenhoven, Eds. De Gruyter Mouton, Feb. 2009, vol. 2.

[8] ——, "Prosodic Markings of Complex NP Focus, Syntax, and the Pre-/Post-focus String," in *West Coast Conference on Formal Linguistics (WCCFL)*. Somerville, MA,: Cascadilla Proceedings Project, 2011, pp. 214–230.

[9] ——, "Prosodic typology: By prominence type, word prosody, and macro-rhythm," in *Prosodic Typology II: The Phonology of Intonation and Phrasing*, S.-A. Jun, Ed. Oxford: Oxford University Press, 2014, pp. 520–539.

[10] S.-A. Jun and H.-J. Lee, "Phonetic and phonological markers of contrastive focus in Korean," in *5th International Conference on Spoken Language Processing (ICSLP 1998)*. ISCA, Nov. 1998, pp. paper 1087–0.

[11] H. Kawahara, A. de Cheveigné, H. Banno, T. Takahashi, and T. Irino, "Nearly defect-free F0 trajectory extraction for expressive speech modifications based on STRAIGHT," in *Ninth European Conference on Speech Communication and Technology*, 2005.

[12] H.-S. Kim, S.-A. Jun, H.-J. Lee, and J.-B. Kim, "Argument Structure and Focus Projection in Korean," in *Speech Prosody*, Dresden, Germany, 2006.

[13] F. Kügler and C. Féry, "Postfocal Downstep in German," *Language and Speech*, vol. 60, no. 2, pp. 260–288, Jun. 2017.

[14] D. R. Ladd, *Intonational Phonology*, 2nd ed., ser. Cambridge Studies in Linguistics. Cambridge ; New York: Cambridge University Press, 2008.

[15] M. Liberman and J. B. Pierrehumbert, "Intonational Invariance under Changes in Pitch Range and Length," in *Language Sound Structure: Studies in Phonology*, M. Halle, M. Aronoff, and R. T. Oehrle, Eds. Cambridge, Mass: MIT Press, 1984, pp. 157–233.

[16] M. McAuliffe, M. Socolof, S. Mihuc, M. Wagner, and M. Sonderegger, "Montreal Forced Aligner: Trainable Text-Speech Alignment Using Kaldi," in *Interspeech 2017*. ISCA, Aug. 2017, pp. 498–502.

[17] M. McAuliffe and M. Sonderegger, "Korean MFA acoustic model v2.0.0a," https://mfa-models.readthedocs.io/acoustic/Korean/Korean MFA acoustic model v2_0_0a.html, Tech. Rep., May 2022.

[18] R Core Team, "R: A Language and Environment for Statistical Computing," R Foundation for Statistical Computing, Vienna, Austria, 2022.

[19] Y.-L. Shue, P. A. Keating, C. Vicenik, and K. Yu, "Voicesauce: A Program for Voice Analysis," in *Proceedings of International Congress of Phonetic Sciences*, Hong Kong, 2011, pp. 1846–1849.

[20] S. Wood, "Mgcv: Mixed GAM Computation Vehicle with Automatic Smoothness Estimation version 1.8-38 from CRAN."

[21] S. N. Wood, *Generalized Additive Models: An Introduction with R*, second edition ed., ser. Chapman & Hall/CRC Texts in Statistical Science. Boca Raton: CRC Press/Taylor & Francis Group, 2017.

[1] This is excluding the initial negation /ani/ 'no,' which is phrased as an independent IP.

[2] The R code for the GAMM construction of the phrase-initial target words is shown below in Listing 1.

Listing 1: R code for phrase-initial GAMM construction

```
# creation of interaction term
ini <- mutate(
  labFoc = interaction(word, focus))
# main effect of focus
bam(scaleF0 ~ labFoc
# smooths by focus
+ s(Time, by=labFoc, bs="cr", k=9)
# random smooths for Speaker
+ s(Time, Speaker, by=labFoc,
    bs="fs", m=1),
data=ini, method="fREML",
rho=r1, AR.start=ini$start.event)
```