SOME ASYMMETRICAL PRE- *VERSUS* POST-FOCAL EFFECTS ON ARTICULATORY REALIZATION OF PROMINENCE DISTRIBUTION IN KOREAN

Suyeon Im[1], Sahyang Kim[2] & Taehong Cho[3]

[1]Soongsil University, [2]Hongik University, [3]Hanyang Institute for Phonetics & Cognitive Sciences of Language, Hanyang University
sim@ssu.ac.kr, sahyang@hongik.ac.kr, tcho@hanyang.ac.kr

ABSTRACT

This study examines how CVC words in non-prominent contexts are kinematically realized in Korean, compared to when focused. Results on lip aperture showed some asymmetric effects between pre-focal and post-focal contexts. While both pre-/post-focal gestures were much more reduced than focal counterparts, the difference tended to be clearer for the pre-focal/focal than the post-focal/focal comparison. There was also some evidence for post-focal gestures being relatively less reduced than pre-focal gestures. We attribute this small asymmetry to differential auditory impacts of non-focal gestures on focal gestures, assuming that gestures are auditory-perceptually more salient with a drastic energy increase from the preceding gesture than the opposite direction. Relatively less reduced post-focal gestures compared to pre-focal gestures suggest that focus-induced accentuation spreads to the following, rather than preceding, gesture. These results imply that although prominence may be phonologically determined, the articulatory realization is fine-tuned, driven by perceptual and articulatory optimization of prominence distribution.

Keywords: prosody, focus, prominence distribution. articulatory kinematics, Korean

1. INTRODUCTION

Prominence is one way of making speech signal perceptually salient to the listener, so that the intended linguistic message is delivered clearly to the listener as intended by the speaker (cf. [1]). In a language like English, prominence is often localized on a lexically stressed syllable, especially when it receives a pitch accent that is post-lexically assigned by the prominence system of the language. Such a prominence effect is most clearly evident when the pitch accent is assigned in conjunction with focus associated with information structure [2, 3, 4, 5]. A language like Korean does not employ lexical stress, so that it may not be localized to a specific stressed syllable, but the notion of prominence itself may still apply to Korean, so that it makes a linguistic unit more salient than its neighbouring units to enhance the perceptual salience of the linguistic message under prominence.

As such, prominence may be defined in a relative term not only in the hierarchy of stress in a given language, but also in terms of relative perceptual salience (e.g., [6, 7]). This suggests that a speech unit (a syllable or a prosodic word) can be perceived salient when the unit itself is hyperaticulated more than neighboring segments. From another angle, this also means that the unit is perceived salient when neighboring units are reduced. Thus, one could infer that such relative salience associated with prominence is employed cross-linguistically by the prominence system of any language, regardless of whether the language employs lexical stress as in English or not as in Korean.

In the present study, we investigate how CVC words are reduced in non-prominent (non-focal) contexts adjacent to the focal context in Korean, relative to when they are focused. That is, we explore how a language which does not employ lexical stress in the prominence system expresses prominence distribution in relative terms and how the results may compare to existing prominence-related data available in other languages such as English (e.g., [3, 4, 6, 8, 9, 10]) that uses a typologically different prominence system. Moreover, we use an Electromagnetic Articulograph, so that we can examine how reduction in non-focal contexts versus hyperarticulation in focal contexts are kinematically expressed in both spatial and temporal dimensions.

A specific question to be addressed is how articulatory reduction of pre-focal versus post-focal words may be similar to or different from each other, as compared to when the same words are focused. Both the pre- and post-focal conditions are expected to induce articulatory reduction, but given the directional asymmetry of pre- versus post-focal conditions, we expect that the degree of reduction will also be asymmetrical. It is hypothesized that the degree of articulatory reduction will be greater in the pre-focal than in the post-focal context, assuming that the auditory-perceptual impacts are likely to be greater when there is a drastic and rapid increase in articulatory force from a non-focal gesture on the following focal gesture rather than the other way around (e.g., [11]). This asymmetrical effect can also be predicted independently by how the domain of accentual impact is often defined. For example, it has been suggested that accentual impact in English may spread to the right rather than to the left, showing a kind of carry-over effect, although the domain may be defined phonologically (e.g., a foot; [12, 13]).

2. METHOD

2.1. Participants & Speech Materials

As part of a larger project building an articulatory database, twelve speakers of Seoul Korean (6 male and 6 female college students, M_{age} = 23.3 years) participated in an articulatory experiment using an Electromagnetic Articulograph (Carstens AG501). They were instructed to imagine playing a card game in which the participant would play a role to answer the interlocutor's question regarding where to put a card (with a target word written on it) on the board. The participants read provided written sentences in response to question sentences with the pictures of the cards visually presented along on a computer screen. The question sentences were designed to elicit corrective focus in the answer sentences on (a) the target word (focal condition), (b) the preceding word (pre-focal condition) and (c) the following word (post-focal condition). The target words were /pap/ ('rice') and /pam/ ('night'/'chestnut') in a phrase-medial position (in the middle of an Intonational Phrase) with the same preceding and following words. An example of test sentences is shown in Table 1.

Table 1: Example of test sentences. Target words are underlined. Focused words are in bold.

(a) focal
Q: *[ipʌn tanʌnɨn ʌnni kuk twiɛta nwa]?*
'Do (I) put the word behind the sister's soup this time?'
A: *[ani. ʌnni **pap** twiɛ].*
'No. Behind the sister's **rice**. '
(b) pre-focal
Q: *[ipʌn tanʌnɨn ʌnni pap ap^{h}ɛta nwa]?*
'Do (I) put the word in front of the sister's soup this time?'
A: *[ani. ʌnni pap **twiɛ**].*
'No. **Behind** the sister's rice.'
(c) post-focal
Q: *[ipʌn tanʌnɨn op*a pap twiɛta nwa]?*
'Do (I) put the word behind the brother's rice this time?'
A: *[ani. **ʌnni** pap twiɛ].*
'No. Behind the **sister's** rice.'

2.2. Measurement

Eight sensor coils were attached to the nose (R1), the two lips (L1, L2), the two gums (R2, J), the tongue tip (T1), tongue body (T2), and the tongue dorsum (T3) as illustrated in Figure 1. The lip aperture (i.e., the Euclidian distance between the upper lip and the lower lip) was used to examine kinematic characteristics of lip closing and opening gestures of CVC words with bilabial consonants (/p/ and /m/). Figure 2 illustrates gestural landmarks used for kinematic measurements. The onset (ONS) and the target (TARG) of each (closing or opening) gesture were determined as the time points when the velocity reached 20% of the peak velocity (PKVEL) as shown in Figure 2. Movement duration (ms; DUR), displacement (mm; DISP), and peak velocity (cm/sec; PKVEL) were measured for each movement from the 2043 tokens in the corpus (3 words * 3 focus conditions * 20 repetitions * 12 speakers - 117 excluded tokens whose renditions deviated from intended phrasings and focus distribution), using MVIEW [14].

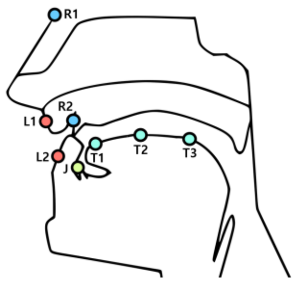

Figure 1: Location of eight sensor coils for articulatory experiment using an Electromagnetic Articulograph (Carstens AG501).

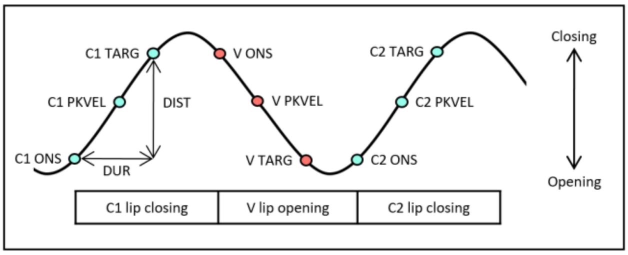

Figure 2: Gestural landmarks for C1 lip closing gesture, V lip opening gesture and C2 closing gesture.

2.3. Statistical Analysis

A linear mixed-effects model was run for each movement (C1 closing, C-to-V opening, C2 closing) with Focus (focal, pre-focal, post-focal) and Word (/pap/, /pam/) as fixed factors including their interaction. (Note that Word was included as a control factor, and given that there was no Focus x Word interaction, we did not report the statistical details of Word effects.) The random structure included by-subject intercept and slope for all the fixed factors. The models were run in R [15] using the lme4 package [16].

3. RESULTS

3.1. C1 (Onset) Lip Closing Gesture

For C1-closing gesture, duration showed no significant difference among focal, pre-focal, and post-focal conditions (pre-focal/focal: β=-1.44, *t*=-1.25, *n.s*; post-focal/focal: β=.48, *t*=.34, *n.s.*). On the other hand, as shown in Figure 3, both displacement and peak velocity measures showed a significant difference between the focal and the pre-focal conditions, such that compared to the focal condition, the pre-focal C1 closing movement was *smaller* in displacement and *slower* in peak velocity (disp., β=-.69, *t*=-3.58, $p<.01$; pkvel., β=-1.35, *t*=-2.86, $p<.05$). But there was no difference between the focal and the post-focal conditions (disp., β=.05, *t*=.20, *n.s.*; pkvel., β=-.26, *t*=-.57, *n.s.*). (Note that the overall duration of C1-closing movement was longer for /pam/ than /pap/, but it did not influence the focus-related effects.) There were no other significant effects or interactions among factors including.

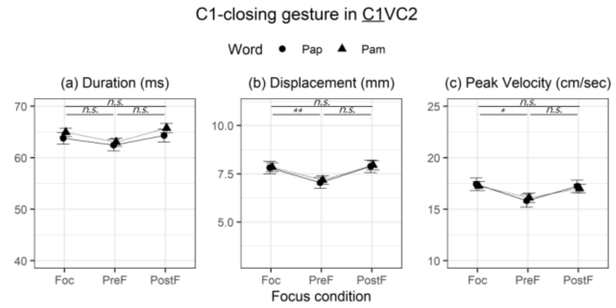

Figure 3: C1 lip-closing duration (left), displacement (middle), and peak velocity (right) plotted by Word (/pap/, /pam/) and Focus condition (focal, pre-focal, post-focal).

3.2. C-to-V Lip Opening Gesture

For C-to-V lip opening gesture, the focal condition significantly differed from both pre- and post-focal conditions in all the measures. As shown in Figure 4, C-to-V opening was longer in duration, larger in displacement and faster in peak velocity in the focal than in the non-focal conditions (focal/pre-focal: duration, β=-18.67, *t*=-8.50; disp., β=-5.23, *t*=-6.22; pkvel., β=-8.36, *t*=-5.93, all at $p<0.01$; focal/post-focal: duration, β=-16.36, *t*=-7.75; disp., β=-4.84, *t*=-6.05, pkvel., β=-7.66, *t*=-6.04, all at $p < .001$). (Note that C-to-V opening gesture was generally longer, larger and faster for /pam/ than /pap/, but it did not influence the focus-related effects.) No other significant effects between conditions were observed.

3.3. C2 (Coda) Lip Closing Gesture

There was a significant difference among the focus

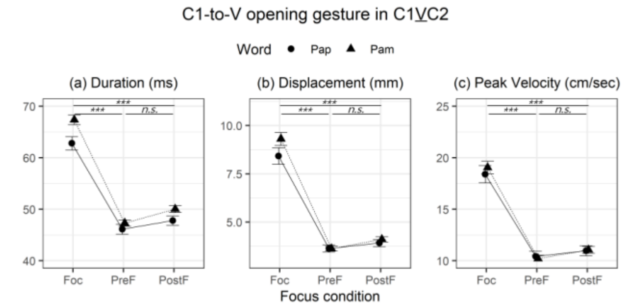

Figure 4: The V lip opening duration (left), displacement (middle), and peak velocity (right) plotted by Word (/pap/, /pam/) and Focus condition (focal, pre-focal, post-focal).

conditions in the C2 closing gesture. C2-closing movement was longer, larger and faster in the focused than in the unfocused condition (focal/pre-focal: duration, β=-9.90, *t*=-4.54; disp., β=-5.21, *t*=-7.20; pkvel, β=-9.44, *t*=-8.43, all at $p<.001$; focal/post-focal: duration, β=-8.06, *t*=-4.84; disp., β=-4.66, *t*=-6.51; pvkel, β=-8.34, *t*=-7.34, all at $p < .001$). Crucially, there was also a significant difference between pre- and post-focal conditions, so that post-focal C2-closing gesture was larger in displacement and faster in peak velocity (disp., β=-8.34, *t*=-7.34, $p < .001$; pkvel, β=-8.34, *t*=-7.34, $p < .001$). (Note that C2-closing gesture was generally longer for /pam/ than /pap/ especially in the focal context, but it did not influence the focus-related effects.) There were no other significant effects or interactions among factors.

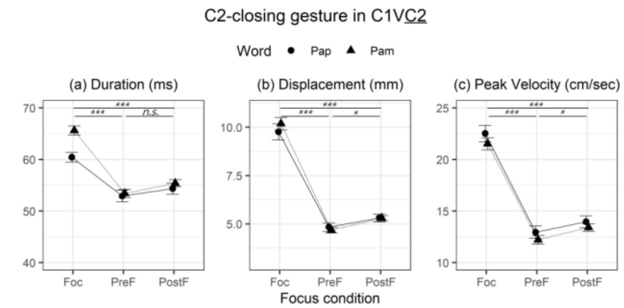

Figure 5: The C2 lip closing duration (left), displacement (middle), and peak velocity (right) plotted by Word (/pap/, /pam/) and Focus condition (focal, pre-focal, post-focal).

4. DISCUSSION

This study examined how articulatory Lip Aperture movements (C1-closing, C-to-V opening, and C2-closing) of C1VC2 words with bilabial consonantal contexts (/pam, pap/) would be realized in non-prominent contexts (pre-focal and post-focal) compared to when they are in the prominent (focal) context in Korean.

A basic finding was that non-focal gestures were much more reduced in both spatial and temporal dimensions than focal gestures that received a corrective contrastive focus. In other words, focal gestures were hyperarticulated (being larger, longer and faster than non-focal gestures), being 'prominent' above the surrounding non-focal words. This hyperarticulation pattern in Korean is largely consistent with the hyperarticulation pattern generally reported in English [3, 6, 17]. The present study taken together with previous studies demonstrates cross-linguistic similarities—i.e., gestures under focus are hyperarticulated in much the same way across languages, regardless of whether a language is a head-prominence language like English (where prominence is localized to a stressed syllable) or an edge-prominence language like Korean (where prominence is primarily expressed by phrasing) (cf. [18]).

As for the specific research question of how pre-focal versus post-focal effects may differ from each other relative to focal effects, our results indicated that pre-focal gestures tended to be reduced more than post-focal gestures. This asymmetry was evident in two cases. For one thing, C1-closing gesture when in the pre-focal condition was reduced (in displacement and peak velocity) compared to when in the focal condition, whereas the same C1-closing gesture of the post-focal word showed no such reduction. For another, while C2-closing gesture was substantially reduced when in both pre- and post-focal conditions, it was the pre-focal C2-closing gesture that was reduced more (as evident in displacement and peak velocity), relative to the post-focal C2-closing gesture. (Recall that C1-to-V opening gesture was reduced to a similar extent in both pre- and post-focal conditions relative to the focal condition.)

These results indicate that the nature of reduction differs depending on the directionality of prominence distribution—i.e., whether it occurs in the pre-focal or post-focal context. On the one hand, the pre-focal word was reduced as a whole from the beginning C1-onset gesture to the final C2-closing gesture which was immediately adjacent to the focal word. On the other hand, the post-focal word was reduced in a rather progressively gradient way. The beginning (C1-closing gesture) of the post-focal word which was immediately adjacent to the focal element was not reduced at all in any kinematic measure, and the reduction became evident on the following C1-to-V opening and C2-closing gestures. And even the final C2-closing gesture of the post-focal word was reduced relatively less than the final C2-closiong gesture of the pre-focal word. These asymmetrical results can be interpreted in terms of differential auditory-perceptual effects and articulatory propensity of post-focal reduction.

The substantial across-the-board reduction of the pre-focal word is indeed consistent with what has previously been reported in other languages such as English [9, 10, 19] and French [20]. This cross-linguistically similar effect can be taken as suggesting that the speaker plans to reduce the preceding gesture deliberately to enhance the auditory-perceptual impact on the focused word that follows as discussed in [19], which would in turn bring about contrast maximization (e.g., [10]). This possibility is consistent with the general assumption that speech production is auditory-perceptually more salient and impactful with a drastic energy increase from the immediately preceding acoustic event than the opposite direction (see [11] for related discussion on asymmetrical auditory-perceptual effect.).

On the other hand, the post-focal production appears to be subject to the propensity to carry over the accentuation to the right rather than the other way around. The progressively increasing reduction to the post-focal word can be considered as the evidence of 'spill-over' (carry-over) effect as discussed in [12, 13, 21]. Such a rightward spread of articulatory force associated with prominence is consistent with a view that accentuation operates within a phonologically-defined domain of accentuation in English (e.g., a foot structure [12, 13]). But we suggest that the gradual rightward spread of accentuation also reflects the efficiency of the production system—i.e., more effort is required to drastically reduce the articulatory force immediately after the focal articulation than do so gradually. Thus, the observed gradual attenuation of the focus effect on the following non-focal unit can be interpreted as suggesting that the speaker avoids a drastic reduction unless otherwise required by the system, in line with a principle of effort minimization that underlies speech production [22].

5. CONCLUSION

The present study was the first to explore how prominence distribution would be reflected in kinematic terms in Korean, especially with respect to reduction patterns of pre-focal versus post-focal gestures as compared to focal gestures. The general differences in the focal versus the non-focal contexts illuminate that relative prominence is kinematically realized in a form of 'hyperarticulation' in much the same way across languages. Furthermore, the directional asymmetry of pre-focal versus post-focal effects further implies that although prominence may be defined differently in the phonology of a given language, articulation of prominence is fine-tuned by the production system of the language that optimizes prominence distribution taking into account both the listener-oriented auditory-perceptual saliency and the speaker-oriented motor efficiency.

6. REFERENCES

[1] Cangemi, F., Baumann, S. 2020. Integrating phonetics and phonology in the study of linguistic prominence. *J. Phonetics.* 81, 100993.
[2] Cho, T. 2011. Laboratory phonology. In: Kula, N. C., Botma, B., Nasukawa, K. (eds), *The Continuum Companion to Phonology*. Continuum, 343–368.
[3] de Jong, K. 1995. The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation. *J. Acoust. Soc. Am.* 97, 491–504.
[4] de Jong, K. 2004. Stress, lexical focus, and segmental focus in English: Patterns of variation in vowel duration. *J. Phonetics*. 32, 493–516.
[5] Fletcher, J. 2010. The prosody of speech: Timing and rhythm. In: Hardcastle, W. J., Laver, J., Gibbon, F. E. (eds). *The Handbook of Phonetic Sciences (2nd ed.).* Blackwell, 523–602.
[6] Beckman, M. E., Edwards, J. 1994. Articulatory evidence for differentiating stress categories. In: Keating, P. A. (ed), *Phonological Structure and Phonetic Form: Papers in Laboratory Phonology III*. Cambridge University Press, 7–33.
[7] Beckman, M. E., Pierrehumbert, J. 1986. Intonational structure in Japanese and English. *Phonology Yearbook*, 3, 255–309.
[8] Cho, T. 2006. Manifestation of prosodic structure in articulatory variation: Evidence from lip kinematics in English. In: Goldstein, L., Whalen D. H., Best, C. T. (eds), *Papers in Laboratory Phonology VIII: Varieties of Phonological Competence (Phonology and Phonetics)*. Mouton de Gruyter, 519–548.
[9] Erickson, D., Lehiste, I. 1995. Contrastive emphasis in elicited dialogue: Durational compensation. *Proc. 13th ICPhS* Stockholm, 4, 352–355.
[10] Jensen, C. 2004. *Stress and accent: Prominence relations in Southern Standard British English.* Ph.D. dissertation, University of Copenhagen, 1–192.
[11] Wright, R. 2004. A review of perceptual cues and cue robustness. *Phonetically Based Phonology*, 34, 57.
[12] Turk, A. E., Sawusch, J. R. 1997. The domain of accentual lengthening in American English. *J. Phonetics*, 25, 25–41.
[13] Turk, A. E., White, L. 1999. Structural influences on accentual lengthening in English. *J. Phonetics*, 27, 171–206.
[14] Tiede, M. 2005. MVIEW: Software for visualization and analysis of concurrently recorded movement data. Haskins Laboratories.
[15] R Core Team. 2022. R: A language and environment for statistical computing. R Foundation for Statistical Computing. https://www.R-project.org/.
[16] Bates, D., Mächler, M., Bolker, B., Walker, S. 2015. Fitting linear mixed-effects models using lme4. *J. Statistical Software*, 67(1), 1–48.
[17] de Jong, K. 1991. An articulatory study of consonant-induced vowel duration changes in English. *Phonetica*, 48, 1–17.
[18] Jun, S. A. 2006. Intonational phonology of Seoul Korean revisited. *Japanese-Korean Linguistics*, 14, 15–26.
[19] Cho, T., Kim, J., Kim, S. 2013. Preboundary lengthening and preaccentual shortening across syllables in a trisyllabic word in English. *J. Acoust. Soc. Am.* 133(5), 384–390.
[20] Fougeron, C., Jun, S.-A. 1998. Rate effects on French intonation: Prosodic organization and phonetic realization. *J. Phonetics*. 26, 45–69.
[21] Sluijter, A. M. C., van Heuven, V. J. 1995. Effects of focus distribution, pitch accent and lexical stress on the temporal organization of syllables in Dutch. *Phonetica* 52, 71–89.
[22] Flemming, E. 1995. *Auditory representations in phonology.* Ph.D. dissertation, UCLA.

7. ACKNOWLEDGEMETNS

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).