PREBOUNDARY LENGTHENING AND ITS KINEMATIC CHARACTERISTICS IN MANDARIN CHINESE IN INTERACTION WITH FOCUS AND LEXICAL TONES

Hongmei Li[1,3], Sahyang Kim[2], Taehong Cho[3]

[1]Department of English, Yanbian University; [2]Hongik University;
[3]Hanyang Institute for Phonetics and Cognitive Sciences of Language (HIPCS), Hanyang University
angela.hongmeili@gmail.com, sahyang@hongik.ac.kr, tcho@hanyang.ac.kr

ABSTRACT

This study investigates preboundary lengthening (PBL) in Mandarin Chinese by exploring kinematic variation of Lip Aperture conditioned by boundary, focus, and lexical tones. Results showed that the lip opening and the following hold duration of CV words were longer phrase-finally, but in interaction with prominence. Under focus, PBL came with slower velocity with no spatial expansion, but under no focus, PBL came with spatial expansion but no slowing-down. This suggests that in the absence of focus-induced hyperarticulation, PBL comes with both temporal and spatial expansion, possibly counteracting a slowing-down effect. PBL also interacted with lexical tones. Compared to simplex Tone1, PBL was much more robust for Tone3 (low-dipping) and Tone4 (falling) with further augmented PBL for Tone3 under focus, presumably to make sufficient room for realizing their tonal complexity. These results imply that PBL is modulated by the phonetically-driven phonological needs for maximizing tonal contrast when licensed by prosodic structure.

Keywords: preboundary lengthening, kinematics, Mandarin Chinese, focus, lexical tones

1. INTRODUCTION

Preboundary lengthening (henceforth PBL) refers to a temporal expansion of phonological units at the right edge of a prosodic domain [4,12,19]. As a near-universal phonetic phenomenon, PBL is unequivocally realized across many languages as temporal lengthening in the acoustic domain, but its detailed articulatory manifestation and its scope are known to be language-specifically modulated by other higher-order linguistic factors [17,18,22,25].

The phonetic realization of PBL is known to be gradient in nature, with its effect strongest when closest to the boundary and more robust before a larger boundary than a smaller one [12,16,17,18,22]. Within the framework of Articulatory Phonology [2,14], these characteristics of PBL have been understood with the π-gesture model [3,4,7]. The π-gesture is a non-tract variable prosodic gesture that does not have a specified constriction degree and location of its own but overlaps with segmental constriction gestures. The π-gesture is assumed to modulate the rate of clock and control the realization of articulatory movements. As a consequence, articulatory gestures are realized with a longer, slower, and in some languages, spatially larger movement under a stronger influence of the π-gesture (i.e., at a stronger boundary) than when not [3,4,15]. In addition, its effect is the strongest at the juncture and gradually wanes across gestures even within a syllable in proportion to its distance from the boundary. The present study investigates the kinematic characteristics and the scope of PBL of monosyllabic CV words in Mandarin Chinese by examining the lip closing and opening gestures at the phrase-final and phrase-medial positions.

The investigation of PBL in Mandarin is particularly interesting, given the accumulating crosslinguistic evidence that the detailed manifestation of PBL is modulated by the prominence system of language [16,18,22,24]. In English, for example, PBL was not only realized on the final syllable regardless of stress, but also on a non-final stressed syllable [24]. PBL was further affected by phrase-level prominence in English, being modulated by the degree of prominence [18]. The interaction between PBL and the language-specific prominence system was also found with Japanese, a lexical pitch-accent language, and with Korean, an edge-prominence language without lexical level prominence, although no further interaction between PBL and phrase-level prominence was found in either language [16,22]. In order to further our understanding of the fine-grained kinematic realization of PBL and its interaction with language-specific prominence system, the current study focuses on how lexical tone system and phrase-level prominence in Mandarin interact to affect articulatory manifestation of PBL.

Mandarin has four lexical tones which are specified with different tonal targets [5,11]: a High-level tone (T1), a Low-High rising tone (T2), a Low-dipping tone (T3), and a High-Low falling tone (T4). There is a close correlation between contour-bearing ability and intrinsic duration of tone [26]: a longer duration is required when a tone has more pitch targets, a greater excursion, or a rising pitch excursion. Considering that each tone has distinctive tonal targets and intrinsic temporal structure, a question arises as to how PBL may be modulated by lexical tones in Mandarin. On the one hand, complex tones with more tonal targets would require more time to manifest phonological contrasts compared to simplex tones. Thus, one might expect a greater PBL effect for complex tones than for simplex tones. On the other hand, since lexical contrasts should be maintained through distinctive lexical tones both phrase-finally and phrase-medially, both complex and simplex tones may still be produced with a similar degree of PBL when phrase-final relative to when phrase-medial.

The present study also examines the effects of focus-induced prominence on PBL and lexical tones to explore whether and how the assumed interaction between lexical tones and PBL may be further conditioned by prominence. In particular, given that a boundary-related strengthening effect on tonal realization could be masked by the presence of tonal hyperarticulation under focus, testing the focus factor may further inform us whether and how the assumed contrast maximization of tones associated with PBL becomes more or less evident in the presence or absence of focus [7,8,10].

2. METHOD

2.1. Participants and procedures

Twelve native Mandarin speakers (6 females, 6 males, M_{age} = 25.2) participated in this study. They were born and raised in Northern China and had resided in Korea for fewer than 3 years at the time of recording. Two CV sequences (/pa/, /ma/) were recorded across four lexical tones. Each target word was embedded in a carrier sentence that was an answer to a question in a mini dialogue where Boundary (IP-medial vs. IP-final) and Focus (UnFoc vs. Foc) conditions varied, as in Table 1.

Table 1: Examples of target words in carrier sentences. Target words are underlined and italicized, and focused words are in bold. "#" represents an IP boundary. A number after each word indicates the lexical tone of the word (e.g., 1 = Tone1).

Boundary	Focus	Example sentences
IP-medial	UnFoc	A: [**mɑʊ1** mi1 pa1 pi4 ʂɤŋ4 ma?] Does **Cat** EIGHT win? B: [pu4] **#** [**ma1** mi1 <u>*pa1*</u> pi4 ʂɤŋ4.] No. **Mommy** <u>*EIGHT*</u> wins.
	Foc	A: [ma1 mi1 **ta1** pi4 ʂɤŋ4 ma?] Does Mommy **BUILD** win? B: [pu4] **#** [ma1 mi1 <u>***pa1***</u> pi4 ʂɤŋ4.] No. Mommy <u>***EIGHT***</u> wins.
IP-final	UnFoc	A: [ni3 tʂʰu1 **mɑʊ1** mi1 pa1 ma?] Do you play **Cat** EIGHT? B: [pu4] **#** [uɔ3 tʂʰu1 **ma1** mi1 <u>*pa1*</u>] **#** [pi4 ʂɤŋ4 pa?] No. I play **Mommy** <u>*EIGHT*</u>. Must win, right?
	Foc	A: [ni3 tʂʰu1 ma1 mi1 **ta1** ma?] Do you play Mommy **BUILD**? B: [pu4] **#** [uɔ3 tʂʰu1 ma1 mi1 <u>***pa1***</u>] **#** [pi4 ʂɤŋ4 pa?] No. I play Mommy <u>***EIGHT***</u>. Must win, right?

During the recording, the speakers were presented with a mini dialogue written in Chinese on the top of a computer screen, along with two picture cards (Fig.1). Next to the picture on each card, a monosyllabic word was displayed in Chinese. The target word (/ma/ or /pa/) was always on a card with a picture of a mommy, marked by a red seal. The combination of a picture and the target word yielded critical sequences such as /mɑʊ1mi1 pa1/ ('cat eight'), /ma1mi1 pa1/ ('mommy eight'), and /ma1mi1 ta1/ ('mommy build'). A target word (e.g., pa1) was followed by either an IP boundary (IP-final condition) or an IP-medial word boundary (IP-medial condition). In the focused condition, the target word on a mommy card was contrasted with a word on another mommy card, given in the prompt question (e.g., 'mommy build /ta1/' vs. 'mommy eight' /pa1/, Fig 1b). In the unfocused condition, the picture words were contrasted in the question-answer pair (e.g., 'cat eight' /pa1/ vs. 'mommy eight' /pa1/, Fig.1a), such that the target word was unfocused.

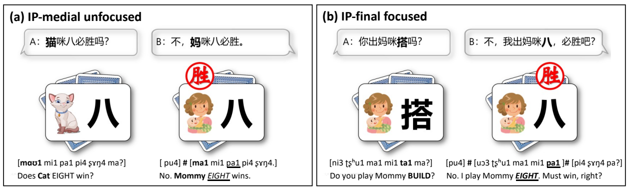

Figure 1: An example of the visual prompts used in the designed card game. Displayed in the right mommy card is a target word (*pa1*, 'eight').

The experiment was conducted at Hanyang Institute for Phonetics and Cognitive Sciences in Seoul. The articulatory data were obtained by using EMA (Electromagnetic Articulograph, Carstens AG501) with a sampling rate of 250 Hz. Acoustic data were recorded simultaneously with the articulatory data, with a Tascam US-4x4 audio interface and a SHURE KSM44A microphone at a sampling rate of 44 kHz. In total, 1920 tokens were collected: 2 CV sequences x 4 lexical tones x 2 boundaries x 2 focus conditions x 5 repetitions x 12 speakers. Thirty-five tokens with unintended prosodic renditions were excluded, leaving 1885 tokens for further data analysis.

2.2. Measurements and Statistical analysis

Articulatory data of lip movements were analyzed using a MATLAB-based software MVIEW [23]. Lip aperture was calculated using the Euclidean distance between the two sensors attached to the upper and lower lips. The gestural landmarks were identified based on the velocity profile, and seven kinematic measures for lip closing and opening gestures of CV were taken as follows (schematized in Fig. 2).

(1) **movement duration:** duration from onset to target (ms)
(2) **hold duration:** duration from target to release (ms)
(3) **formation duration**: duration from onset to release (ms)
(4) **time-to-pkvel**: duration from onset to peak velocity (ms)
(5) **movement displacement**: displacement from onset to target (mm)
(6) **maximum displacement**: displacement from onset to max constriction (mm)
(7) **peak velocity**: the highest velocity during lip movements (cm/sec)

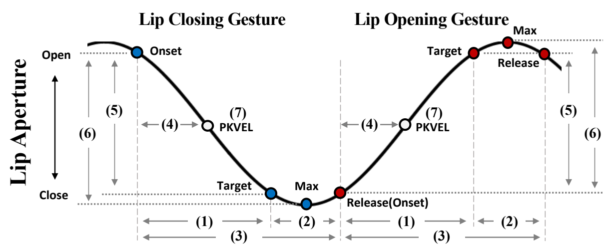

Figure 2: Schematized trajectories of lip closing and opening movements with kinematic landmarks and temporal-spatial measures.

A series of linear mixed-effects models were conducted by using lme4 [1] and lmerTest [20] packages. The seven kinematic measures were taken as dependent variables. Boundary, Focus, Tone, and their interactions were fixed factors. Boundary (IP-medial, IP-final) and Focus (unfocused, focused) were contrast-coded, while Tone (T1, T2, T3, T4) was treatment-coded. (The reference levels are underlined). The random structure included a random intercept by Speaker with random slopes by Boundary, Tone, and Focus and a random intercept by Item. The maximal random effect structure was employed if the model converged. When there were interactions between factors, pairwise comparisons of the posthoc tests were conducted. For the purpose of the present study, only the main effect of Boundary and its interactions with Focus and Tone are reported.

3. RESULTS

3.1. Articulatory characteristics of Mandarin PBL

The PBL effects were significant in almost all measures of the lip opening gesture. The lip opening gesture was longer in duration (in all temporal measures but *time-to-pkvel*), larger in displacement, and slower in peak velocity in the IP-final than in the IP-medial positions.

In contrast, none of the temporal measures showed any significant PBL effects for the lip closing gesture (although there was a trend effect in *total duration*). The lip aperture was significantly smaller IP-finally than IP-medially for both *movement* and *maximum displacement*. *Peak velocity*, however, was slower IP-finally than IP-medially for the lip closing gesture, as has been found with the lip opening gesture.

3.2. Interaction between PBL and prominence

As for the lip opening gesture, significant Boundary x Focus interactions were found in *movement duration* (Fig.3a) and *formation duration* measures. These interactions were due to the fact that the PBL effect was greater in the unfocused condition (β=-14.93, $p<.001$) than in the focused condition (β=-9.60, $p<.001$). No interaction was found in *time-to-pkvel* (Fig.3b).

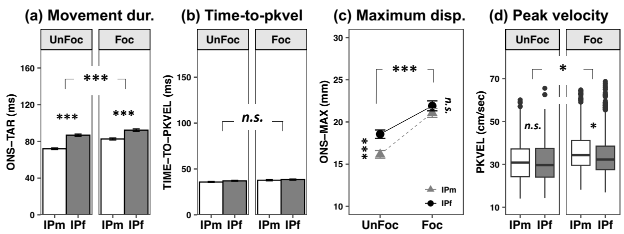

Figure 3: PBL x Prominence interactions for the lip opening gesture. Error bars show standard errors. (*** $p<.001$; * $p<.05$)

In the spatial dimension, a significant Boundary x Focus interaction was found in *maximum displacement* (Fig.3c) and in *movement displacement*, which was attributable to the fact that the larger displacement due to Boundary was only observed in the unfocused condition (β=-2.45, $p<.001$).

Peak velocity also showed a robust Boundary x Focus interaction, which may arise from the fact that the significantly slower movement in the IP-final than IP-medial position was only observed under focus (β=1.68, $p<.05$; Fig.3d).

As for the lip closing gesture, a trend interaction effect was found in *formation duration* only, which may be due to the fact that the significant PBL effect was only detected in the unfocused condition (β=-4.58, $p<.05$). No other significant interaction was found.

3.3. Interaction between PBL and lexical tone

Each of the three (potentially) complex tones, that is, T2 (a Low-High rising tone), T3 (a Low-dipping tone), and T4 (a High-Low falling tone) was compared with a simplex, High-level tone (T1) to examine the interaction between PBL and the lexical tone.

As for the lip opening gesture, there was no significant interaction between Boundary and Tone for T2, either in the temporal or in the spatial dimension. It means that T1 and T2 did not differ from each other in terms of their Boundary effects.

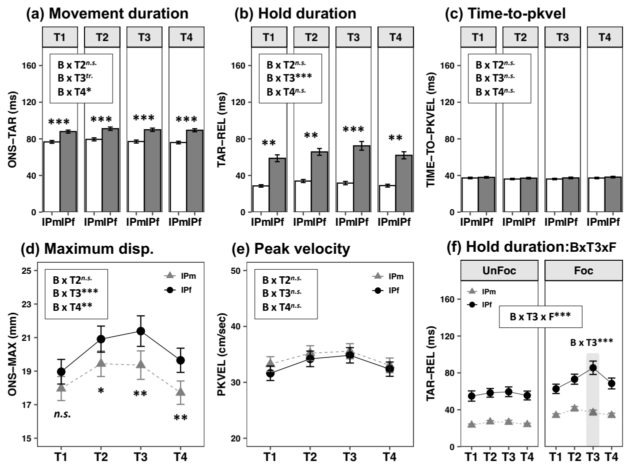

Figure 4: PBL x Tone interactions for the lip opening gesture. Error bars show standard errors. (*** $p<.001$; ** $p<.01$; * $p<.05$; *tr.* $.05<p<.08$)

T3 showed a significant Boundary x Tone interaction in *hold duration* (Fig. 4b), which was mainly due to the fact that the PBL effect (i.e., longer *hold duration* in the IP-final than in the IP-medial position) was significantly greater for T3 (β=-40.90, $p<.001$) than for T1 (β=-30.16, $p<.01$). This effect further interacted with Focus (β=19.27, $p<.001$; Fig.4f), showing that the Boundary x Tone interaction found with T3 was significant only in the focused condition. Presumably due to the extreme PBL effect of T3 in *hold duration*, similar Boundary x Tone interaction was also found in *formation duration*. However, no interaction was detected in *time-to-pkvel* as shown in Fig.4c.

T4 showed a significant interaction between Boundary and Tone in *movement duration* (Fig. 4a). The posthoc tests revealed that the interaction may arise from the fact that the PBL effect was greater in T4 (β=-13.32, p<.001) than in T1 (β=-11.15, p<.001). Significant two-way interaction was also found in *formation duration* but not in *hold duration* and *time-to-pkvel*.

In the spatial dimension for the lip opening gesture, there were significant Boundary x Tone interactions for T3 and T4 in both *movement displacement* and *maximum displacement*. As shown in Fig.4d, the interactions stemmed from the fact that lip opening was larger in the IP-final position than in the IP-medial position for T3 and T4, but that this boundary effect was not observed for T1.

As for *peak velocity* (Fig.4e), no interaction between Boundary and Tone was found in any tone comparisons.

For the lip closing gesture, significant Boundary x Tone interactions were found in *hold duration* and *formation duration* for T2, which may be contributable to the fact that the PBL effect was more reduced in T2 than in T1. There was a Boundary x Tone interaction found in *time-to-pkvel* for T4. No other interaction was found.

4. SUMMARY AND DISCUSSION

A basic finding of our study is that preboundary lengthening (PBL) in Mandarin Chinese was found for the (C-to-V) lip opening that was proximal to the boundary but not for the lip closing gesture for the onset C which was distal. This proximity effect could be accounted for the π-gesture model [3,4,7,15], in that its influence was stronger when the gesture was closer to the boundary. The lip opening gesture showed a *longer*, *larger*, and *slower* movement in the IP-final than in the IP-medial position. The slower movement could also be attributed to a clock-slowing effect of the π-gesture. Interestingly, however, *time-to-pkvel* was not necessarily longer associated with PBL, although it is often assumed to be influenced by the π-gesture. Instead, *hold duration* (the duration of the plateau-like portion after the target attainment of the lip opening gesture) contributed significantly to the PBL. As for the lip closing gesture preceding the opening gesture, there was no sign of PBL, though its movement itself was still *slower* in velocity but *smaller* in displacement which could counteract the lengthening effect (the smaller the displacement, the shorter the duration).

The PBL effect in Mandarin further interacted with focus-induced prominence. The boundary-induced temporal enhancement was greater in the unfocused than in the focused condition, and the spatial enhancement was only found when the target word was not focused. Results are in line with the findings in English [18], which showed the most robust PBL in the least prominent condition. These findings seem to suggest that when gestures are locally hyperarticulated by focus [7,10,21], and hence already spatio-temporally expanded enough, there is no sufficient room left to expand the gesture further to mark the end of a phrase boundary (a kind of ceiling effect).

What is also interesting is that the phrase-final slowing down was observed only in the focused condition, which yielded longer and slower (but not larger) movement under focus, unlike longer and larger phrase-final movement found in the absence of focus. This may indicate that speakers employ multiple strategies to achieve the prosodic goal of boundary marking, instead of utilizing one single kinematic parameter to encode linguistic structures [3,6,24].

Most crucially for the purpose of the present study, results showed that the PBL-related articulatory manifestation was differentially modulated by the lexical tone of the phrase-final word. The High-level tone (T1) and the rising tone (T2) were comparable in their PBL effects in both the temporal and spatial dimensions. In contrast, the Low-dipping tone (T3) and the falling tone (T4) showed Boundary x Tone interactions when compared with T1. In the temporal dimension, T3 and T4 showed greater PBL effects than T1. Recall that both T3 (low-dipping tone) and T4 (falling tone) have a low tonal target, and it is physiologically demanding to fully attain the underlying low pitch. Due to restricted time in the phrase-medial position, T3 and T4 must have been temporally suppressed and thus implemented with a short duration on the surface, which resulted in no durational difference between T1 and T3/T4 in the phrase-medial position. In the phrase-final position, however, both T3 and T4 are more robustly lengthened compared to T1, presumably to realize their tonal targets fully in the position. This pattern is notable particularly for T3. As it is physiologically challenging to continuously maintain a low target with vocal fold vibration, the pitch goes up to some extent after reaching the low target, such that vocal folds may reach a more comfortable equilibrium state. As a result, a hidden rising component of the dipping tone is manifested when more time is allowed in the phrase-final position, which makes T3 manifest like a complex contour tone. It is also notable that the lengthening of *hold duration* for T3 was observed under focus. This seems to indicate that the underlying tonal features of T3 became visible and fully implemented phrase-finally, at least partially due to local hyperarticulation to achieve tonal contrast maximization [8,9,10,21].

In the spatial dimension (in the tonal space), T3 and T4 showed larger displacement in the phrase-final than in the phrase-medial position, which was not observed with T1. This is also consistent with the full realization of L tone target in the phrase-final position. That is, with sufficient time, the jaw is likely to be lowered more, which further aided the larynx to reach the Low target in both T3 and T4. Taken together, the tonal targets for T3 and T4 were fully realized with longer duration and accompanied by larger spatial magnitude phrase-finally than phrase-medially.

In conclusion, our study shows that although PBL in Mandarin follows the cross-linguistically applicable patterns, it is modulated by higher-level prominence and lexical tones. From the point of view of the Articulatory Phonology, tone can also be deemed as laryngeal gesture [13]. In this context, our results suggest that segmental constriction gestures are coordinated with tone gestures and prosodic gestures (e.g., the π-gesture), and they are all activated to regulate the spatio-temporal kinematics of speech production. Results further support the view that low-level fine phonetic details can be fine-tuned by higher linguistic structures of a given language in a language-specific way.

5. ACKNOWLEDGEMENT

We thank the Mandarin Chinese speakers for their participation and the reviewers for the helpful comments. This work was supported by the Ministry of Education of the Republic of Korean and the National Research Foundation of Korea (NRF-2021S1A5C2A02086884 awarded to TC (PI) and SK (Co-PI)).

6. REFERENCES

[1] Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). Parsimonious mixed models. *arXiv preprint.arXiv:1506.04967*.

[2] Browman, C. P., & Goldstein, L. (1995). Dynamics and articulatory phonology. *Mind as motion: Explorations in the dynamics of cognition*, *175*, 194.

[3] Byrd, D., & Krivokapić, J. (2021). Cracking prosody in articulatory phonology. *Annual Review of Linguistics*, *7*(1), 31-53.

[4] Byrd, D., & Saltzman, E. (2003). The elastic phrase: Modeling the dynamics of boundary-adjacent lengthening. *Journal of Phonetics*, *31*(2), 149-180.

[5] Chao, Y. R. (1965). A grammar of spoken Chinese.

[6] Cho, T. (2006). Manifestation of prosodic structure in articulatory variation: Evidence from lip kinematics in English. *Laboratory phonology*, *8*, 519-548.

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

[8] Cho, T., & McQueen, J. M. (2005). Prosodic influences on consonant production in Dutch: Effects of prosodic boundaries, phrasal accent and lexical stress. *Journal of Phonetics*, *33*(2), 121-157.

[9] Choi, J., Kim, S., & Cho, T. (2016). Phonetic encoding of coda voicing contrast under different focus conditions in L1 vs. L2 English. *Frontiers in psychology*, *7*, 624.

[10] De Jong, K. (2004). Stress, lexical focus, and. segmental focus in English: patterns of variation in vowel duration. *Journal of Phonetics*, *32*(4), 493-516.

[11] Duanmu, S. (2007). *The phonology of standard Chinese*. OUP Oxford.

[12] Edwards, J., Beckman, M. E., & Fletcher, J. (1991). The articulatory kinematics of final lengthening. *the Journal of the Acoustical Society of America*, *89*(1), 369-382.

[13] Gao, M. (2008). Tonal alignment in Mandarin Chinese: An articulatory phonology account. *Unpublished Doctoral Dissertation (Linguistics), Yale University, CT*.

[14] 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*, 215-249.

[15] Iskarous, K., & Pouplier, M. (2022). Advancements of phonetics in the 21st century: A critical appraisal of time and space in Articulatory Phonology. *Journal of Phonetics*, *95*, 101195.

[16] Jang, J., & Katsika, A. (2020). The amount and scope of phrase-final lengthening in Seoul Korean. In *Proceedings of the 10th International Conference on Speech Prosody, Tokyo, Japan* (pp. 270-274).

[17] Kim, J. J., Baek, Y., Cho, T., & Kim, S. (2019). Preboundary lengthening and boundary-related spatial expansion in Korean. In *Proceedings of the 19th International Congress of Phonetic Sciences*.

[18] Kim, S., Jang, J., & Cho, T. (2017). Articulatory. characteristics of preboundary lengthening in interaction with prominence on tri-syllabic words in American English. *The Journal of the Acoustical Society of America*, *142*(4), EL362-EL368.

[19] Klatt, D. H. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *The Journal of the Acoustical Society of America*, *59*(5), 1208-1221.

[20] Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. (2017). lmerTest package: tests in linear mixed effects models. *Journal of statistical software*, *82*, 1-26.

[21] Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. In *Speech production and speech modelling* (pp. 403-439). Springer, Dordrecht.

[22] Seo, J., Kim, S., Kubozono, H., & Cho, T. (2019). Preboundary lengthening in Japanese: To what extent do lexical pitch accent and moraic structure matter?. *The Journal of the Acoustical Society of America*, *146*(3), 1817-1823.

[23] Tiede, M. (2010). MVIEW: Multi-channel. visualization application for displaying dynamic sensor movements. *development*.

[24] Turk, A. E., & Shattuck-Hufnagel, S. (2007). Multiple targets of phrase-final lengthening in American English words. *Journal of Phonetics*, *35*(4), 445-472.

[25] Xu, Y. (1999). Effects of tone and focus on the formation and alignment of f0contours. *Journal of phonetics*, *27*(1), 55-105.

[26] Zhang, J. (2014). Tones, tonal phonology, and tone sandhi. *The handbook of Chinese linguistics*, 443-464.