THE DOMAIN OF BOUNDARY-RELATED LENGTHENING IN TOKYO JAPANESE, A MORA-TIMED LANGUAGE

Sherry Chien, Karen Tsai, Argyro Katsika

University of California, Santa Barbara
sherrychien@ucsb.edu, karentsai@ucsb.edu, akatsika@linguistics.ucsb.edu

ABSTRACT

Although boundary-related lengthening is well-established, it is unclear whether the stretch of speech affected is determined by a grammatical domain (e.g., mora, syllable, word) or whether it corresponds to a fixed interval at the boundary. Here, we use Electromagnetic Articulography (EMA) to address this issue in Tokyo Japanese, a mora-timed language. Disyllabic initial-accented words with all possible combinations of moraic structure in each syllable were tested in phrase-final and phrase-medial positions. Analyses of consonant gestures in each syllable's onset show that boundary-related lengthening extends beyond the phrase-final mora or syllabic rhyme to the onset of the phrase-final syllable when the latter is light (CV), but not when it is heavy (CVC or CVV). No evidence for boundary-related lengthening was found in the onset of the penultimate syllable, suggesting that the phrase-final word is not the domain of the effect. Results are discussed within the framework of Articulatory Phonology.

Keywords: boundary lengthening, moraic structure, Tokyo Japanese, articulation, Articulatory Phonology

1. INTRODUCTION

Phrase-final lengthening is the phenomenon in which acoustic and articulatory durations are longer in phrase-final positions as compared to their phrase-medial counterparts (e.g., acoustics: [1]–[5]; articulation: [6]–[11]). The effect is so widely attested in numerous languages and language varieties that it is considered a language universal [12, 13]. Yet, the scope of the effect, i.e., the stretch of speech affected, remains an understudied and ambiguous issue. An important parameter that stays unknown is whether the scope of the effect is determined by a grammatical domain. For instance, there is evidence suggesting that the effect can extend over several pre-boundary segments, with the exact domain of the effect varying, possibly on a language-specific manner, from the phrase-final syllable rhyme [1, 14, 15] to the phrase-final foot [4] or even word [16]. Alternatively, the effect may not scope over a grammatically defined domain, but instead over a fixed interval at the boundary [17]. Furthermore, it seems that the scope and degree of the effect is further fine-tuned by several factors that also affect phonetic duration, such as segment type, vowel length, syllable structure as well as language-specific features. For instance, in English, phrase-final lengthening has been found to apply to the rhyme of the phrase-final syllable [5], while other studies have suggested that the effect begins earlier, extending to the onset of the final syllable in cases when its vowel is reduced or is not a diphthong [1]. Another factor that emerges from the literature as playing an important role in determining the scope of the effect is position of lexical stress and/or phrase-level pitch accent [10] (See also [4], [5]; but see [18]). Recent work in Japanese shows interactions of the effect with lexical pitch accent as well [19].

Here, we also turn to Japanese. While the majority of studies have looked into boundary-related effects in stress languages like English and Dutch, where the most relevant phonological unit is mainly the syllable, mora-timed languages like Japanese can offer an interesting perspective on the domain of phrase-final lengthening, since the candidate domains of mora, syllable and word can be disentangled from each other. Previous findings on this issue in Japanese has been limited and inconclusive. An acoustic study by Shepherd [20] has found that phrase-final positions showed proportionally greater lengthening in short vowels than long vowels, and took this to mean that the domain of final lengthening in Japanese is one mora. On the other hand, studies from Seo et al [21, 22] suggest that the scope of boundary lengthening is better explained by syllabic rather than moraic structure, as lengthening on final CV syllables was comparable to that of the combined effect of lengthening in the two moras in the rhyme of a final CVN syllable. Moreover, Seo et al's [22] study showed that the degree of phrase-final lengthening was greater when the penultimate syllable had a moraic nasal. Finally, the number of segments in the rhyme was found to affect the degree of lengthening, as the scope of boundary was found to be more limited in words that have more segments (e.g., CVN.CVN) compared to those that have fewer segments (e.g., CV.CV).

In this study, we examine the interaction between moraic structure and phrase-final lengthening in Tokyo Japanese via Electromagnetic Articulography

(EMA). Disyllabic initial-accented words with all possible combinations of moraic structure in each syllable (see Table 1) are tested in phrase-final and phrase-medial positions. If the domain of phrase-final lengthening is the syllable, we expect to see the final syllable to lengthen in test disyllabic words with various moraic structures [21, 22]. If the domain of phrase-final lengthening is the mora, we expect to see lengthening limited to the final mora across the different syllable weights [20]. If the scope is domain-agnostic, effects should be seen on a specific interval at the end of the phrase (e.g., [17]). If lengthening is attracted to heavy non-final syllables, long vowels and geminate consonants should have the same effect as moraic nasals in Seo et al's [21, 22] research.

2. METHOD

2.1. Participants and recording apparatus

Participants were 6 native speakers of Tokyo Japanese (1 male and 5 females, M_{age} = 22 years, range 20-25 years). All participants could read Japanese orthography and reported no speech, hearing, or vision problems. They were not aware of the purpose of the experiment and were compensated for their participation.

Data were collected via the AG501 three-dimensional electromagnetic transduction device (Carstens Medizinelektronik). Five sensors were attached to tongue dorsum, tongue body center, tongue tip, upper lip, and lower lip. An additional five sensors were attached as reference points: upper incisor, lower incisor, nose, and left and right ears. Audio recordings were performed simultaneously with the kinematic recordings with a Sennheiser shotgun microphone set at a sampling rate of 16 kHz.

2.2. Experimental design and procedure

Test words were 11 disyllabic initial-accented words with every possible combination of moraic structure in the first and second syllable, as shown in Table 1.

All test words were embedded in frame sentences in a phrase-medial (PhM) control position or in a phrase-final (PhF) test condition, as illustrated in Table 2. Frame sentences were controlled for length (number of moras) and pitch accent of the following verb, which was varied to create sensical utterances. Short dialogues were used to create contexts that prompted frame sentences embedded with the target words as responses.

All sentences and stimuli were presented in Japanese orthography on a computer screen placed about 3 feet away from the participant. Context sentences were read silently and target sentences aloud. Altogether, there were 2 phrasal positions, 10 test words (manipulating moraic structure), 9 repetitions, and 6 speakers for a total of 1188 utterances collected.

Mora Count	Moraic Structure	Syllabic Weight	Test Word	Gloss
2	CV.CV	LL	ma*mi	'name'
2	CV.CV	LL	na*mi	'name'
3	CV.CVV	LH	ma*ni:	'money'
3	CV.CVN	LH	bi*naɴ	'handsome man'
3	CVV.CV	HL	bi:*mu	'beam'
3	CVN.CV	HL	baɴ*bi	'southern area'
3	CVC.CV	HL	bag*gu	'bag'
4	CVN.CVV	HH	meɴ*bo:	'cotton swab'
4	CVV.CVV	HH	ba:*bi:	'Barbie'
4	CVV.CVN	HH	ma:*biɴ	'name'
4	CVC.CVN	HH	mak*kuɴ	'name'

Table 1: Test words by moraic structure and resulting syllable weight. L = light syllable, H = heavy syllable

Phrasal Position	**Test Sentence**
Phrase-medial (PhM)	[hoɴto: ni **na*mi** makasi̥ta?]$_{IP}$ really Nami defeated? 'Really (you) defeated Nami?'
Phrase-final (PhF)	[hoɴto: ni **na*mi** ?]$_{IP}$ [makasi̥ta?]$_{IP}$ really Nami defeated? 'Really Nami? (You) defeated (her)?'

Table 2: Sample frame sentences for test word na*mi.

2.3. Data analysis

Kinematic labels for all consonants (C) were made using a semiautomatic procedure that identifies constriction gestures (Tiede, Haskins Laboratories). Labial C gestures (/m/) were labelled on the lip aperture tract, coronal C gestures (/t, d, n/) on the tongue tip vertical displacement tract and dorsal C gestures (/k, g/) on the tongue dorsum vertical displacement tract.

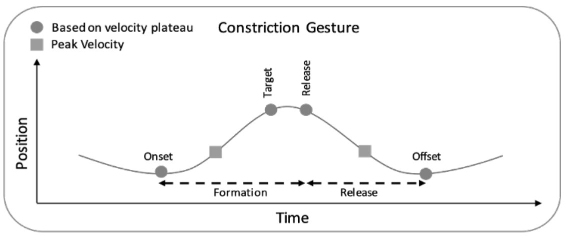

Figure 1: Schematic representation of a gestures and labelled timepoints.

The labelling procedure detected the following timepoints in each test C gesture using velocity criteria: onset, target, release, and offset of the gesture. The timepoints at peak velocity of both the formation and the release phase were also labelled. Labels were used here to calculate the duration of the formation and the release phases of each test constriction gesture. Formation duration was defined as the interval between the onset and release of the gesture, and release duration as the interval between the release and the offset of the gesture, as shown in Figure 1.

The moraic nasals /ɴ/ that precede a bilabial stop /b/ in the words such as /baɴ*bi/ and /meɴ*boː/ are segmented as one single gesture (i.e., /ɴb/), as moraic nasal /ɴ/ in Japanese assimilates to the following stop in terms of place of articulation [23]. Consonants that are realized as geminates, such as the velar stops /g/ and /k/ in /bag*gu/ and /mak*kuɴ/, respectively, are also segmented as one single gesture. This is because both /ɴb/ and geminates appear as a single constriction.

The constriction gestures for each onset consonant of the disyllabic test words were analyzed. We refer to the onset consonants of the first and second syllable as C1 and C2 respectively. Separate linear mixed effects models with formation duration and release duration as response variables were fitted for each test C gesture using the lmerTest package [24] in R [25]. Random effects included random intercepts by speaker and word. The three fixed factors included were Boundary (phrase-medial vs. phrase-final, referred to as PhM and PhF respectively), First Syllable's Moraic Structure and Second Syllable's Moraic Structure (levels for two last factors: light [CV], heavy with a long vowel [CVV], heavy with a moraic coda [CVC]).

The drop1 function in R [25] was used to determine the optimal model. Random Effects Principle Components Analysis (rePCA) was used to determine the best random effect structure, and emmeans with Holm correction [26] were used to derive pairwise comparisons (α =0.05).

3. RESULTS

Results are reported for each test C gesture separately. C2 gesture is presented first since it is closer to the IP boundary, and thus more likely to be affected by it.

3.1. C2 formation duration

Statistical analyses detected an interaction effect between boundary and second syllable's moraic structure on the formation duration of C2 gesture ($F(2)$= 20.279, p<0.0001). Post-hoc pairwise comparisons clarified that boundary had a lengthening effect on the C gesture in the onset of the phrase-final syllable (C2 gesture) only when that syllable was light (CV) (p<0.001), and not when it was heavy either with a long vowel (CVV) or a moraic coda (CVC) (see Figure 2). This indicated that phrase-final lengthening can extend beyond the phrase-final mora or syllabic rhyme to the onset of the phrase-final syllable, but only when the syllable is light. As Figure 2 illustrates, this pattern might arise from a ceiling effect on C gestures in onsets of heavy syllables.

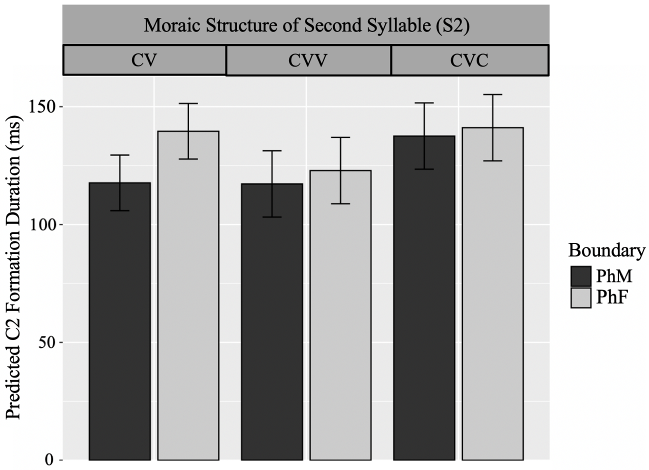

Figure 2: Predicted formation duration (ms; with standard error) of C2 gesture as a function of Boundary (PhM, PhF) by moraic structure (CV, CVV, CVC) of the second syllable' (S2).

3.2. C2 release duration

The release duration of C2 gesture showed an interesting three-way interaction of Boundary with the moraic structure of both the first and the second syllable ($F(4)$= 6.2144, p<0.0001).

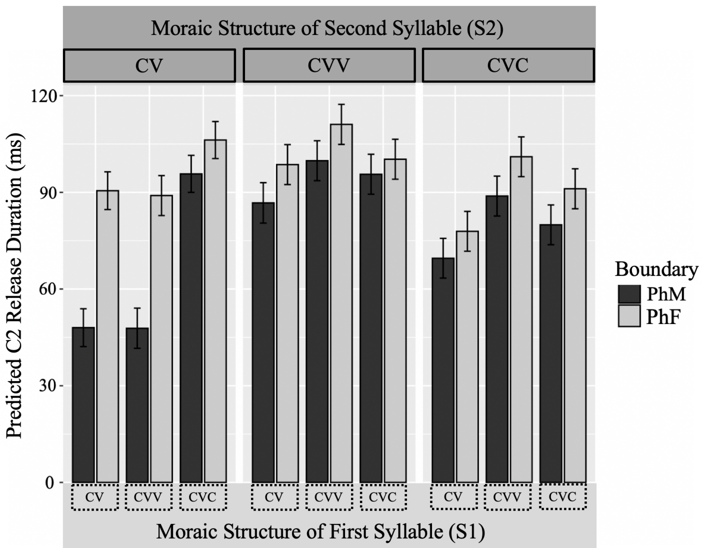

Figure 3: Predicted release duration (ms; with standard error) of C2 gesture as a function of boundary (PhM, PhF) by moraic structure (CV, CVV, CVC) of first (S1) and second (S2) syllable.

According to the pairwise comparisons, C2 release underwent phrase-final lengthening only in words with CV.CV or CVV.CV structures, in which the second syllable is light and the first syllable is not closed with a moraic coda (p<0.0001 for both comparisons). Thus, similarly to C2 formation, phrase-final lengthening extends from the right edge of the IP to C2 release only when the final syllable is light, with the following exception: this pattern does not hold in cases of CVC penultimate syllables. Once again, as Figure 3 indicates, these patterns seem to arise from a ceiling effect, since phrase-medial C gestures in CV.CV and CVV.CV are much shorter than their counterparts in the other moraic structures (p<0.0001 for all comparisons).

3.3. C1 formation duration

No effect of Boundary was detected on the formation duration of C1 gesture.

3.4. C1 release duration

A three-way interaction effect among the fixed factors of Boundary, First Syllable's Moraic Structure and Second Syllable's Moraic Structure was detected for C1 release (F(4)=2.4901, p=0.04). Pairwise comparisons showed that the source of this interaction was not the phrase-final lengthening effect. As illustrated in Figure 4, C1 release was not lengthened phrase-finally in any of the combinations of moraic structure for the two syllables of the phrase-final words.

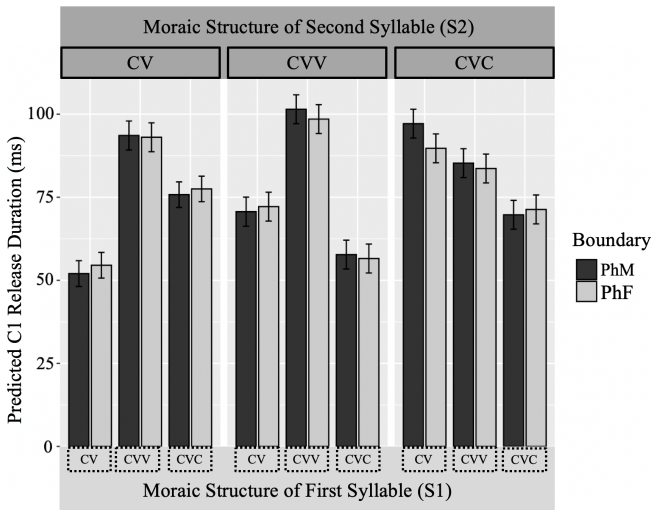

Figure 4: Predicted release duration (ms; with standard error) of C1 gesture as a function of boundary (PhM, PhF) by moraic structure (CV, CVV, CVC) of first (S1) and second (S2) syllable.

4. DISCUSSION

In order to examine the scope of boundary lengthening in Tokyo Japanese, this study investigated the durational profile of C gestures in the two syllabic onsets of initial-accented disyllabic words. No systematic boundary-related lengthening was found in the onset consonant of the penultimate syllable, suggesting that the phrase-final word is not the domain of the effect in this language, as was proposed for German in [16].

For the word-final syllable, boundary effects were found up to the onset of the phrase-final syllable, but only when that final syllable was light. Absence of phrase-final lengthening on C onsets of heavy syllables (CVV and CVC) is presumably due to a ceiling effect. These results in combination suggest that the domain of phrase-final lengthening in Tokyo Japanese is either the phrase-final mora, being in agreement with [20], or the phrase-final rhyme, suggested in e.g., [5]. Next steps of our analysis will assess acoustic durations of all C and V components of phrase-final syllables in an attempt to differentiate between these two alternative accounts.

We further propose that in order to better understand the presence of phrase-final lengthening on the onset of the word-final syllable, one should consider the fact that the test words used here are initially-accented. Lexical pitch accent has been found to affect the scope of phrase-final lengthening, such that words with non-final pitch accent present a larger scope of the effect than words with final pitch accent (e.g., [19], see also [21]). An account of these patterns could be provided from within the framework of Articulatory Phonology (e.g., [27]), in which prosodic effects at boundaries are instantiated by prosodic modulation gestures, called π-gestures, that control the spatio-temporal profile of the C and V constriction gestures that overlap with them [17]. These π-gestures have been proposed to have a dual coordination in stress languages [10]: one with the phrase-final vowel (V) gesture and one with the prosodic modulation gesture, called μ-gesture, that gives rise to stress [28]. Based on [19], Tokyo Japanese presents a similar coordination system for their π-gestures: these are coordinated both with the lexical pitch accent (when present) and the phrase-final mora/vowel gesture. When accent is initial, phrase-final lengthening is initiated earlier, extending towards the onset of the final syllable. In light syllables, lengthening actually reaches the onset C gesture, but in heavy syllables, which are longer due to their moraic structure, π-gesture does not get to overlap with the onset C gesture, and thus the latter does not get affected by it.

5. ACKNOWLEDGEMENTS

This work was supported by NSF Grant 1551428 and funds from the UCSB Academic Senate.

6. REFERENCES

[1] D. K. Oller, “The effect of position in utterance on speech segment duration in English,” *J. Acoust. Soc. Am.*, vol. 54, no. 5, pp. 1235–1247, Nov. 1973, doi: 10.1121/1.1914393.

[2] R. Berkovits, “Durational Effects in Final Lengthening, Gapping, and Contrastive Stress,” *Lang. Speech*, vol. 37, no. 3, pp. 237–250, Jul. 1994, doi: 10.1177/002383099403700302.

[3] T. Cambier-Langeveld, “The Domain of Final Lengthening in the Production of Dutch,” *Linguist. Neth.*, vol. 14, no. 1, pp. 13–24, Jan. 1997, doi: 10.1075/avt.14.04cam.

[4] L. White, “English speech timing : a domain and locus approach,” Doctoral dissertation, University of Edinburgh, 2002. Accessed: Jan. 06, 2023. [Online]. Available: http://hdl.handle.net/1842/23256

[5] A. E. Turk and S. Shattuck-Hufnagel, “Multiple targets of phrase-final lengthening in American English words,” *J. Phon.*, vol. 35, no. 4, pp. 445–472, Oct. 2007, doi: 10.1016/j.wocn.2006.12.001.

[6] J. Edwards, M. E. Beckman, and J. Fletcher, “The articulatory kinematics of final lengthening,” *J. Acoust. Soc. Am.*, vol. 89, no. 1, pp. 369–382, Jan. 1991, doi: 10.1121/1.400674.

[7] D. Byrd and E. Saltzman, “Intragestural dynamics of multiple prosodic boundaries,” *J. Phon.*, vol. 26, no. 2, pp. 173–199, Apr. 1998, doi: 10.1006/jpho.1998.0071.

[8] D. Byrd, S. Lee, D. Riggs, and J. Adams, “Interacting effects of syllable and phrase position on consonant articulation,” *J. Acoust. Soc. Am.*, vol. 118, no. 6, pp. 3860–3873, Dec. 2005, doi: 10.1121/1.2130950.

[9] D. Byrd, J. Krivokapić, and S. Lee, “How far, how long: On the temporal scope of prosodic boundary effects,” *J. Acoust. Soc. Am.*, vol. 120, no. 3, pp. 1589–1599, Sep. 2006, doi: 10.1121/1.2217135.

[10] A. Katsika, “The role of prominence in determining the scope of boundary-related lengthening in Greek,” *J. Phon.*, vol. 55, pp. 149–181, Mar. 2016, doi: 10.1016/j.wocn.2015.12.003.

[11] S. Kim, J. Jang, and T. Cho, “Articulatory characteristics of preboundary lengthening in interaction with prominence on tri-syllabic words in American English,” *J. Acoust. Soc. Am.*, vol. 142, no. 4, pp. EL362–EL368, Oct. 2017, doi: 10.1121/1.5005132.

[12] J. Vaissière, “Language-Independent Prosodic Features,” in *Prosody: Models and Measurements*, A. Cutler and D. R. Ladd, Eds. Berlin, Heidelberg: Springer, 1983, pp. 53–66. doi: 10.1007/978-3-642-69103-4_5.

[13] M. D. Tyler and A. Cutler, “Cross-language differences in cue use for speech segmentation,” *J. Acoust. Soc. Am.*, vol. 126, no. 1, pp. 367–376, Jul. 2009, doi: 10.1121/1.3129127.

[14] L. H. Nakatani, K. D. O’Connor, and C. H. Aston, “Prosodic Aspects of American English Speech Rhythm,” *Phonetica*, vol. 38, no. 1–3, pp. 84–105, Jan. 1981, doi: 10.1159/000260016.

[15] C. W. Wightman, S. Shattuck-Hufnagel, M. Ostendorf, and P. J. Price, “Segmental durations in the vicinity of prosodic phrase boundaries,” *J. Acoust. Soc. Am.*, vol. 91, no. 3, pp. 1707–1717, Mar. 1992, doi: 10.1121/1.402450.

[16] K. J. Kohler, “Prosodic Boundary Signals in German,” *Phonetica*, vol. 40, no. 2, pp. 89–134, Mar. 1983, doi: 10.1159/000261685.

[17] D. Byrd and E. Saltzman, “The elastic phrase: modeling the dynamics of boundary-adjacent lengthening,” *J. Phon.*, vol. 31, no. 2, pp. 149–180, Apr. 2003, doi: 10.1016/S0095-4470(02)00085-2.

[18] D. Byrd and D. Riggs, “Locality interactions with prominence in determining the scope of phrasal lengthening,” *J. Int. Phon. Assoc.*, vol. 38, no. 2, pp. 187–202, Aug. 2008, doi: 10.1017/S0025100308003460.

[19] K. Tsai and A. Katsika, “Pitch accent and phrase boundaries: Kinematic evidence from Japanese,” in *Speech Prosody 2020*, May 2020, pp. 215–219. doi: 10.21437/SpeechProsody.2020-44.

[20] M. A. Shepherd, “The scope and effects of preboundary prosodic lengthening in Japanese,” *USC Work. Pap. Linguist.*, vol. 4, pp. 1–14, 2008.

[21] J. Seo, S. Kim, H. Kubozono, and T. Cho, “Preboundary lengthening in Japanese: To what extent do lexical pitch accent and moraic structure matter?,” *J. Acoust. Soc. Am.*, vol. 146, no. 3, pp. 1817–1823, Sep. 2019, doi: 10.1121/1.5122191.

[22] J. Seo, S. Kim, H. Kubozono, and T. Cho, “Interaction between rhythmic structure and preboundary lengthening in Japanese,” in *PProceedings of the 19th International Congress of Phonetic Sciences*, Melbourne, Australia, 2019, pp. 637–641.

[23] T. J. Vance, *The Sounds of Japanese with Audio CD*. Cambridge University Press, 2008.

[24] A. Kuznetsova, P. B. Brockhoff, and R. H. B. Christensen, “lmerTest Package: Tests in Linear Mixed Effects Models,” *J. Stat. Softw.*, vol. 82, no. 13, pp. 1–26, Dec. 2017, doi: 10.18637/jss.v082.i13.

[25] R Core Team, “R: A language and environment for statistical computing.” R Foundation for Statistical Computing, Vienna, Austria., 2022. [Online]. Available: https://www.R-project.org/

[26] R. V. Lenth *et al.*, “emmeans: Estimated Marginal Means, aka Least-Squares Means.” Dec. 06, 2022. Accessed: Jan. 07, 2023. [Online]. Available: https://CRAN.R-project.org/package=emmeans

[27] Browman, C.P., Goldstein, L.M. 1992. Articulatory phonology: An overview. *Phonetica* 45, 155-180

[28] Saltzman, E., Nam, H., Krivokapić, J., Goldstein, L. 2008. A task-dynamic toolkit for modeling the effects of prosodic structure on articulation. *Proc. of the Speech Prosody Conference 2008*, 175-184.