Differential effects of prosodic boundary on glottalization of word-initial vowels in Korean

Sungwok Hwang[1], Sahyang Kim[2], and Taehong Cho[1]

Hanyang Institute for Phonetics & Cognitive Sciences of Language (HIPCS), Hanyang University[1], Hongik University[2]
sungwokhwang@gmail.com, sahyang@hongik.ac.kr, tcho@hanyang.ac.kr

ABSTRACT

This study explores how the voice quality of word-initial vowels is modified by prosodic boundary in Korean. Prosodic boundary effects were examined where the boundaries were aligned with syntactic junctures, resolving structural ambiguity of coordinate structures (*Noun1-hako Noun2 animjʌn Noun3*, 'N1-and N2 or N3'). Results showed asymmetrical effects depending on whether the test word was a noun or a function word. Vowels tended to be glottalized *less* for N2 but *more* for *animjʌn* ('or') phrase-initially (at the Intonational-Phrase boundary) than phrase-medially. Notably, however, in the *less* glottalized case, there was much larger preboundary lengthening, compared to the *more* glottalized case. These results indicate that word-initial vowels are not necessarily glottalized to provide an invariant prosodic boundary marker. Rather, the modification of voice quality, we suggest, is modulated systematically by an optimization of the phonetics-prosody interplay that takes into account the relative contribution of available prosodic/segmental cues to signal prosodic structure.

Keywords: prosodic boundaries, glottalization, syntactic junctures, structural ambiguity

1. INTRODUCTION

Glottalization of word-initial vowel occurs in many languages. In general, it occurs either in a form of epenthetic glottal stop or with laryngealized (glottalized) voicing during the vowel [cf. 8]. Even in a language in which a glottal stop has a phonemic status, the vowel-initial allophonic glottalization still occurs often at a prosodic boundary, which may serve as a prosodic boundary marker [5, 12, 13] (See also [7] for further discussion on prominence as an important factor that induces glottalization.). For example, Mitterer et al. [12] suggested that in Maltese a glottal stop occurs either as a phoneme or as an epenthetic IP juncture correlate, such that Maltese listeners more likely perceive the glottalized sound as an epenthetic boundary marker when the preceding temporal structure as reflected in preboundary lengthening supports the percept of prosodic boundary. This study, though it focused on the perceptual aspects, implies that speakers produce various phonetic reflexes of prosodic juncture, whether suprasegmental or segmental, that marshal synergistically into signaling prosodic juncture, which are also reflected in the way the listener exploits the available cues to prosodic structure. Mitterer et al. [13] further revealed that word-initial glottalization could serve as a phonetic reflex of prosodic juncture that is often mapped onto a major syntactic juncture as a means of resolving syntactic ambiguity [e.g., 3, 9, 18].

The role of glottalization, however, is not limited to marking a large prosodic juncture such as an Intonational Phrase. But it may also be used to avoid vowel hiatus either within a word for marking a syllable boundary [e.g., 3] or across a word boundary thus delimiting a lexical boundary within a phrase [e.g., 4, 8]. Moreover, Garellek [7] further suggests that glottalization comes more consistently with prominence rather than with prosodic boundary especially in English, a head-prominence language [cf. 10]. Thus, a question remains open as to how glottalization is used to mark prosodic structure in a given language. In addition, given that the way that the syntax-prosody interface is actually realized may also vary depending on the language's syntactic and prosodic structure [cf. 6], it also remains open how phonetic reflexes of prosody including glottalization may be used in relation to syntactic disambiguation.

The present study uses Korean to explore how prosodic boundaries are realized in terms of temporal expansion and glottalization at a prosodic boundary and how these temporal and segmental reflexes of prosodic juncture are related to signalling syntactic junctures. To examine the syntax-prosody alignment, structurally ambiguous coordinate structures ([A] and [B or C] vs. [A and B] or [C]) will be used, which will provide different prosodic phrasings (with different prosodic boundaries) in relation to syntactic structure. Preboundary lengthening and the pause duration will be examined for the boundary-related temporal expansion, and word-initial vowels will be examined for the boundary-related glottalization that will be estimated by a spectral tilt measure (H1*-H2*) and HNR (Harmonic-Noise Ratio).

2. METHOD

2.1. Participants and speech materials

Seven males and seven females in their 20's who had no residential experience in the area other than Seoul Korea were recruited.

There were five test words: three proper nouns (person's names; /ali/, /aʧi/, /ami/) and two function morphemes (/-hako/ '-and' as a suffix, /animjʌn/ 'or' as a free morpheme). The nouns had the same vowel sequence (/aCi/). Each word appeared in six types of sentences which were answers to a question in a mini dialogue created for the purpose of the present study (see Table 1).

The three nouns (N1, N2, N3 crossed with a factorial combination) were used to create different coordinate structures that are structurally ambiguous:

(1) a. Early Closure: *(N1-hako) (N2 animjʌn N3)*
b. Late Closure: *(N1-hako N2) animjʌn (N3)*
glosses: *-hako* ('-and'), *animjʌn* ('or')

In (1a), a major syntactic juncture comes earlier after *-hako* ('-and'), which is suffixed to N1 (Early Closure), and in (1b), a major syntactic juncture comes later after N2 and before *animjʌn* ('or') (Late Closure). Moreover, as exemplified in Table 1, we used three different question types, which were meant to generate various target (answer) sentences with different information structures. The first question type was *what's happening?*, so that the answer would be produced with 'broad' focus (BF). The second question type was *who's coming to the party*, so that all three nouns (as person's names) together would receive 'narrow' focus. And the third question type was *Did you say (N1 and N2) or N3 are coming to the party?* In this case, the question was meant to indicate the Late Closure structure (pre-recorded), but the speaker was instructed to correct the sentence to have the Early Closure structure to be produced with 'contrastive' focus (see below for more details). Note that the contrast in this case is not the one between specific lexical items, but the one between two different syntactic parsings, so that, for example, a Late Closure parsing was corrected to be an Early Closure parsing by the speaker.

Table 1. An illustration of test sentences. Each content word (/ali/, /aʧi/, /ami/) could be placed in all three locations: Noun1 (N1), Noun2 (N2), Noun3 (N3).

(a) Info= BF	Q: musɨn ilija "What's happening?" A: *a N1-hako N2 animjʌn N3-ka ontɛ* "Well, N1, and N2 or N3 are coming."
(b) Info= NF	Q: ipʌn ʧʰukʧɛ: nuka ontɛ "Who's coming to the party?" A: a N1-hako N2 animjʌn N3-ka ontɛ "Well, N1, and N2 or N3 are coming."
(c) Info= CF	Q: mwʌ $_{IP}$ N1-hako N2 $_{IP}$ animjʌn N3-ka ʧʰukʧɛ: ontɛ "Did they say N1 and N2, or N3 are coming to the party?" A: ani N1-hako N2 animjʌn N3-ka ontɛ "No, they said N1, and N2 or N3 will."

2.2. Procedure

In each trial, the participants were presented with a visual display on the computer screen indicating an 'intended' coordinate structure to be used in an answer to the question which was pre-recorded by a female native speaker of Seoul Korean.

Participants heard the question first, and read the corresponding target sentence, guided by the visual cue to the underlying coordinate structure. Fig. 1 showed a case where the question type was *What's happning?*. Given that N1 alone was inside a circle, while N2 and N3 were grouped together by a circle, the answer was meant to be produced with an Early Closure coordinate structure.

The data were recorded in a soundproof booth with a Tascam HC-P2 digital recorder and a SHURE KSN44 microphone at a sampling rate of 44kHz. After the recording, the authors labelled target words and checked their prosodic boundary conditions. In total, 1512 tokens were collected (2 syntactic structures x 3 focus conditions x 6 word orders x 3 repetitions x 14 speakers), but 1428 tokens were used for analysis, with an exclusion of 84 tokens whose renditions did not sound natural, as confirmed by all three authors.

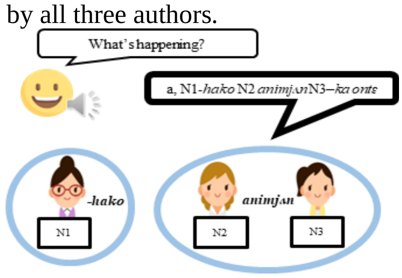

Fig. 1. An illustration of the visual display along with an auditory stimulus as a prompt question which was pre-recorded. This example was used to induce a broad focus with an Early Closure target sentence.

2.3. Measurements and statistical analyses

For N2, /animjʌn/, and N3, word-initial vowel's glottalization was assessed by H1*-H2* and HNR. In H1*-H2*, H1 and H2 refers to the first and second harmonics in the spectral domain, respectively. H1*-H2* is gained by subtracting the amplitude of H2 from the one of H1. The asterisk (*) means that the value was corrected for formants. Lower H1*-H2* values indicate greater glottalization. HNR refers to the harmonic-to-noise ratio, with lower HNR values indicating non-modal voice quality. Mean values of H1*-H2* and HNR were taken from three timepoints corresponding to three equally-divided portions of the whole vowel length. H1*-H2* and HNR values were extracted by using Voice Sauce [16, 17]. For /-hako/ ('-and'), N2, and /animjʌn/ ('or'), the acoustic duration of the last syllable and the temporal gap (pause) between the offset of the last syllable and the onset of the following word were measured using Praat [2].

The statistical analyses were carried out by using the *lme4* [1] packages in R [14] to investigate the influences of prosodic boundary on the four dependent variables: the duration of the last syllable and pause, H1*-H2*, and HNR. The fixed effects were Boundary (<u>Wd</u> vs. IP), Focus (<u>BF</u> vs. NF vs. CF), and Timepoint (<u>1st</u> vs. 2nd vs. 3rd). The underlined categories above were the reference level, and all factors were contrast-coded. The random-effect structure by speaker included random intercepts and random slopes for Boundary, Focus, and their interaction. Random slopes for Timepoint were not involved because the models failed to converge with those slopes. As Timepoint was incorporated as a factor primarily to examine how Boundary would interact with Timepoint, random slopes for Timepoint were excluded for the sake of convergence. The random effect by word order was also included in the model because a single set of the three content words (N1, N2, N3) was considered as one item. The random structure by word order took in only random intercepts because of the convergence issue.

3. RESULTS

3.1. Prosodic phrasing and syntax-prosody mapping

All the sentence tokens reported in the present study could be divided into three prosodic phrasing groups as shown in Fig. 2. Each group showed a consistent syntax-prosody alignment. The Early Closure construct was produced always with Phrasing Type 1 (100%), with a major syntactic juncture aligned with an IP boundary (#) between /-hako/ and N2. For the Late Closure construct, it was produced mostly with Phrasing Type 2 (81%), but often with Phrasing Type 3 (19%). An invariant phrasing aspect in line with the syntactic juncture was that for the Early Closure, an IP boundary was placed always after an function morpheme (*-hako* ‘and’) and for the Late Closure, it was placed always before a function morpheme (*animjʌn*, ‘or’) with an optional boundary after it.

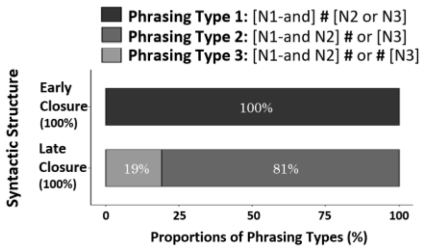

Fig. 2. Distribution of phrasing types for Early and Late Closure.

3.2. The first IP juncture between /-hako/ (‘-and’) and N2 (*N1-and # N2 or N3*)

Boundary had significant effects on the acoustic duration of the final syllable of /-ha<u>ko</u>/ (β=84.15, t=17.40, p<0.001) and pause (β=135.97, t=65.85, p<0.001). As can be seen in Fig. 3a-b, both the duration of /ko/ and the pause were significantly longer in IP-final than in IP-medial condition (by ca. 169ms and ca. 260ms, respectively). When these two measures were combined, there was a temporal expansion of approximately 420ms from IP-medial to IP-final condition. For the initial /a/ of N2, there was no clear sign of the difference in glottalization between the IP-initial and the IP-medial position, as can be inferred from Fig. 3c-d. For H1*-H2*, there was no Boundary effect (β=0.30, t=1.33, p=0.207), but Boundary interacted with Timepoint (β=0.52, t=8.29, p<0.001). As shown in Fig. 3(c), the value of the 2nd and 3rd timepoints tended to be higher in IP-higher than in IP-medial position (2nd: β=0.40, t=1.84, p=0.090; 3rd: β=0.80, t=2.15, p=0.051), indicating that, if there is any, the initial vowel of N2 is on a less glottalized (breathier) side. On the other hand, HNR showed a Boundary effect (β=-2.70, t=-7.99, p<0.001), indicating that the word-initial vowel was noisier IP-initially than IP-medially.

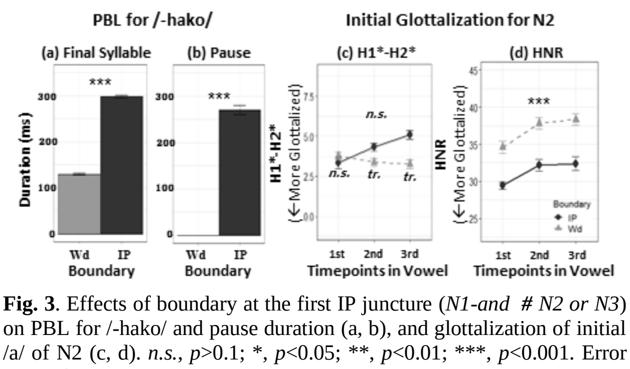

Fig. 3. Effects of boundary at the first IP juncture (*N1-and # N2 or N3*) on PBL for /-hako/ and pause duration (a, b), and glottalization of initial /a/ of N2 (c, d). *n.s.*, p>0.1; *, p<0.05; **, p<0.01; ***, p<0.001. Error bars refer to standard errors.

3.3. The second IP juncture between N2 and /animjʌn/ (*N1-and N2 # or N3*)

There was a significant effect of Boundary on the acoustic duration of the second (final) syllable of N2 (β=67.96, t=15.09, p<0.001) and pause (β=104.47, t=57.32, p<0.001). As shown in Fig. 4a-b, both the duration of the last syllable of N2 before the boundary and the pause were significantly longer at an IP boundary than at an Wd boundary (by ca. 120ms and ca. 210ms, respectively). When the two temporal measures were combined, there was a temporal expansion of approximately 330ms from the Wd boundary to the IP boundary. Compared with the boundary-related temporal expansion (420ms) near /-hako/ (‘-and’) (preboundary lengthening + pause), this temporal expansion (preboundary lengthening of N2 + pause at an IP) was relatively smaller.

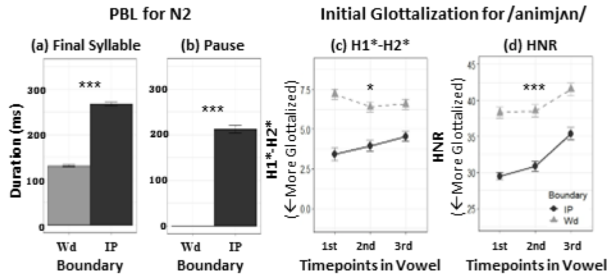

Fig. 4. Effects of boundary at the second IP juncture (*N1-and N2 # or N3*) on preboundary lengthening (PBL) of N2 and pause duration (a, b), and glottalization of initial /a/ of /animjʌn/ (‘or’) (c, d). *n.s.*, p>0.1; *, p<0.05; **, p<0.01; ***, p<0.001. Error bars refer to standard errors.

However, this time, as can be seen in Fig. 4c-d, there was a clearer indication of variation of glottalization as a function of boundary for the initial /a/ of /animjʌn/ (‘or else’). There was a significant effect of Boundary on H1*-H2* (β=-1.40, t=-2.86, p=0.013)—i.e., H1*-H2* was smaller IP-initially than IP-medially (Fig. 4c), indicating that the initial vowel of /animjʌn/ was more glottalized IP-initially. There was also a significant Boundary effect on HNR (β=-3.55, t=-5.92, p<0.001)—i.e., HNR was lower IP-initially than IP-medially (Fig. 4d), indicating that the spectral harmonics were less clear (noisier) IP-initially, in line with the (non-modal) glottalization pattern.

3.4. The third IP juncture between /animjʌn/ and N3 (*N1-and N2 or # N3*)

Boundary had significant effects on the acoustic duration of /mjʌn/ (the final syllable of /ani<u>mjʌn</u>/) (β=31.99, t=3.38, p=0.008) and pause (β=13.97, t=21.86, p<0.001). As can be seen in Fig. 5a-b, both the duration of /mjʌn/ and the pause were significantly longer at an IP boundary than at an Wd boundary (by ca. 70ms and 30ms, respectively). When these measures were combined, there was a total temporal expansion of approximately 100ms from the Wd to the IP boundary condition. Compared to the other IP junctures (which showed a total temporal expansion of 420~330ms), this IP juncture was marked by a far less temporal expansion. Nevertheless, for the post-boundary initial /a/ of N3, there was a clear variation in glottalization as a function of boundary. For H1*-H2*, there was no significant effect of Boundary but there was a significant interaction between Boundary and Timepoint (with 1st/2nd timepoint: β=0.66, t=3.049, p=0.002; with 1st/3rd timepoint; β=-1.07, t=-5.00, p<0.001). As can be seen in Fig. 5c-d, the interaction was due to the fact that there was a significant Boundary effect at the first timepoint (the beginning of the vowel) with a lower H1*-H2 (more glottalized) at the IP-initial position (β=-1.29, t=-2.96, p=0.016), while the effect disappeared into the vowel at the following timepoints. For HNR, there was a significant Boundary effect (β=-1.51, t=-3.42, p=0.011), indicating a substantially lower HNR of the vowel in the IP-initial than in the IP-medial position.

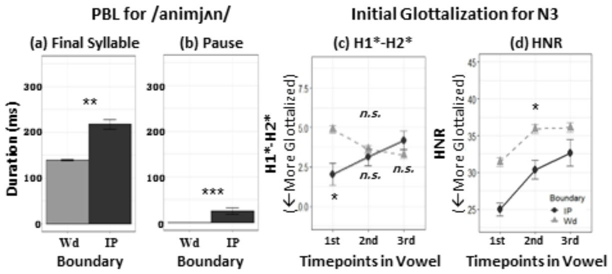

Fig. 5: Effects of boundary at the ‘optional’ third IP juncture (*N1-and N2 or # N3*) on preboundary lengthening (PBL) of /animjʌn/ and pause duration (a, b), and glottalization of initial /a/ of N3 (c, d). *n.s.*, p>0.1; *, p<0.05; **, p<0.01; ***, p<0.001. Error bars refer to standard errors.

4. SUMMARY AND DISCUSSION

The present study has examined how major syntactic junctures of coordinate structures are aligned with major prosodic boundaries to be used for resolving syntactic ambiguity in Korean and how the major prosodic boundary (IP) is phonetically expressed not only in the suprasegmental dimension of preboundary lengthening and pause duration, but also in the glottalization of post-boundary vowel that is definable in the segmental dimension.

One of the basic findings was that the major syntactic juncture was unequivocally aligned with the major prosodic boundary—i.e., the Intonational Phrase boundary. That is, for an Early Closure construct ([N1-and] # [N2 or N3]), the syntactic juncture (‘#’) was consistently aligned with an IP boundary; for a late closure construct ([N1-and N2] # or (#) [N3]), the first syntactic juncture ‘#’ before ‘or’ was consistently aligned with an IP, the second juncture ‘(#)’ was only occasionally (thus *optionally*) aligned with an IP. Such a syntax-prosody mapping means that the major syntactic juncture was expressed by phonetic reflexes of prosodic juncture including a temporal expansion reflected in preboundary lengthening and the pause duration.

But our results further indicated that although the Intonational Phrase may be phonologically defined, governed by the intonational phonology of the language [11], the phonetic reflexes of the same IP boundary could be realized in a gradient fashion. Recall that the magnitude of the temporal expansion associated with an IP boundary was substantially greater for the first IP juncture before N2 of Phrasing Type 1 than for the second IP juncture before ‘or’ of Phrasing Types 2 and 3. One possible reason may be related to a phonetic declination of phrase-final lengthening. Since the IP of Phrase Type 1 occurs earlier in the utterance, its temporal expansion could possibly be larger compared to when an IP occurs later in the utterance as in Phrasing Types 2 and 3. But it is also possible that preboundary lengthening can be controlled by the speaker as discussed in the preboundary lengthening literature [15, 19]. If so, the differential effects of the phonologically defined IP may well be interpreted to have come about as a consequence of the speaker control. Recall that the occurrence of an IP in [N1-and] # [N2 or N3] is invariantly aligned with a major syntactic juncture. From the listener’s point of view, the syntactic parsing is constructed as the speech unfolds over time [e.g., 13, 18]. Thus, the presence or absence of an early IP should play a pivotal role in determining whether the utterance is meant to be an early versus a late closure construct. The speakers may therefore produce the prosodic juncture with a more heightened phonetic clarity through an augmented temporal expansion to clearly signal the critical syntactic juncture. There is yet a third possibility that may not be mutually exclusive from the speaker control account. That is, because the excessive preboundary lengthening before the first IP was observed with the function morpheme *–hako* (‘-and’), one might assume that it has to do with the functionality of the morpheme. But the fact that it was not only the preboundary lengthening but also the pause duration that showed a substantial difference takes the support away from the third possibility.

The same account may apply to the differential effects between the second IP and the third IP. Recall that the temporal expansion was much greater for the second IP before ‘or’ of Phrasing Types 2 and 3 than for the third IP before N3 of Phrasing Type 3. For a late closure construct, the syntactic juncture was consistently aligned with the second IP, and only optionally with the third IP. In this case as well, the speaker could deliberately augment the strength of the prosodic juncture in reference to syntactic structure.

Finally, our results showed that an IP boundary does not always induce a substantial glottalization for the word-initial vowel as compared to a smaller prosodic boundary (i.e., in the phrase-medial position). Interestingly, however, the boundary-related glottalization appears to be correlated with the magnitude of temporal expansion. For the first critical IP for which the most robust temporal expansion was found, no additional boundary-related effect was observed with glottalization. But for the second IP, the following vowel of /animjʌn/ (‘or’) was substantially more glottalized IP-initially. A similar pattern was observed for the (optional) third IP. These results imply that the glottalization is not invariantly employed in marking a prosodic boundary, but rather, the voice quality (including glottalization) is likely to be modulated by the motor system that takes into account relative contributions of available suprasegmental and segmental (glottalization) cues to signal prosody structure.

In conclusion, our results indicate that prosodic structure that may be constructed based on phonologically determined prosodic categories is fleshed out with the gradient phonetic content in both segmental and suprasegmental dimensions. The phonetic content appears to be the outcome of a fine-tuning of phonetic encoding of prosodic structure in reference to syntactic factors as well as system-driven factors.

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] Bates, D., Mächler, M., Bolker, B., and Walker, S. 2015. “Fitting linear mixed-effects models using lme4,” J. Stat. Soft. 67(1), 1-48.

[2] Boersma, P., and Weenink, D. 2020. “Praat: Doing phonetics by computer” [computer program], version 6.0.40, http://www.praat.org/ (Last viewed 29 April 2020).

[3] Davidson, L., & Erker, D. 2014. Hiatus resolution in American English: The case against glide insertion. *Language* 90(2), 482-514

[4] Davidson, L. 2021. Effects of word position and flanking vowel on the implementation of glottal stop: Evidence from Hawaiian. *Journal of phonetics* 88, 101075.

[5] Dilley, L., Shattuck-Hufnagel, S., & Ostendorf, M. 1996. Glottalization of word-initial vowels as a function of prosodic structure. *Journal of phonetics* 24(4), 423-444.

[6] Elfner, E. 2018. The syntax-prosody interface: Current theoretical approaches and outstanding questions. *Linguistics Vanguard* 4(1), 20160081.

[7] Garellek, M. 2014. Voice quality strengthening and glottalization. *Journal of Phonetics* 45(1): 106-113

[8] Garellek, M. 2022. Theoretical achievements of phonetics in the 21st century: Phonetics of voice quality. *Journal of Phonetics* 94: 101155.

[9] Jun, S.-A. 2009. Prosody in sentence processing. In Li, P., Lee, C., Simpson, G., & Kim, Y. (Eds.), *Handbook of East Asian Psycholinguistics, Part III: Korean Psycholinguistics* (pp.423-432). Cambridge: Cambridge University Press.

[10] Jun, S.-A., editor 2014. *Prosodic Typology II: The phonology of intonation and phrasing*. Oxford University Press.

[11] Ladd, D. R. 2008. *Intonational phonology* (second edition). Cambridge: Cambridge University Press.

[12] Mitterer, H., Kim, S., & Cho, T. 2019. The glottal stop between segmental and suprasegmental processing: The case of Maltese. *Journal of Memory and Language* 108, 1-19.

[13] Mitterer, H., Kim, S., & Cho, T. 2020. The role of segmental information in syntactic processsing through the syntax-prosody interface. *Language and Speech* 64(4), 962–979

[14] R Core Team. 2022. “R: A language and environment for statistical computing,” R Foundation for Statistical Computing, Vienna, Austria, https://www. R-project.org/ (Last viewed 30 November 2022).

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

[16] Shue, Yen-Liang 2010. The voice source in speech production: Data, analysis and models. Los Angeles: University of California. http://www.phonetics.ucla.edu/voiceproject/ Publications/shue_dissertation.pdf

[17] Shue, Y.-L., Keating, P., Vicenik, C. & Yu, K. 2011. Voicesauce: A program for voice analysis. In Wai-Sum Lee & Eric Zee (eds.), Proceedings of the 17th International Congress of Phonetic Sciences, 1946–1849. Hong Kong.

[18] Snedeker, J., & Trueswell, J. 2003. Using prosody to avoid ambiguity: Effects of speaker awareness and referential context. *Journal of Memory and Language* 48, 103–130.

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