ALIGNMENT OF PROSODIC AND SYNTACTIC JUNCTURES AND VOWEL-INTIAL GLOTTALIZATION IN SYNTACTIC DISAMBIGUATION IN ENGLISH

Jae-Eun Jennifer Shin[1], Sahyang Kim[2] & Taehong Cho[1]

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

ABSTRACT

This study examines how ambiguity of coordinate structures is resolved by prosodic junctures in American English, and how vowel-glottalization serves as a prosodic boundary marker. Results indicated that syntactic junctures were unequivocally aligned with Intonational-Phrase boundaries (IP-boundary=#), critically placed before conjunctions ([***Anna***] # *and (#)* [***Annie**-or-**Angie***]; [***Anna**-and-**Annie***] # *or (#)* [***Angie***]), and optionally after them. Qualitatively, vowels were substantially glottalized both IP-initially and IP-medially, but spectral tilt measures revealed some boundary effects at the critical juncture, interacting with prominence: Vowels of conjunctions were glottalized more IP-initially (critical IP-boundary) only when pitch-accented, whereas no boundary effect was observed on vowels of nouns (optional IP-boundary). Moreover, both pre- and post-boundary lengthening were greater at the critical than optional IP-boundary. Finally, no prominence-induced glottalization was observed, and the opposite was true in some cases. These results suggest that the phonologically-defined IP category is realized not with the same phonetic content, but with phonetic modulation in reference to syntax and information structure.

Keywords: prosodic boundary, syntactic ambiguity, vowel glottalization, syntax-prosody interface, prominence

1. INTRODUCTION

In spoken language, speakers must present speech through a phonetic output to be understood clearly by the listener. For the phonetic form to arise, speakers go through phonetic encoding which is later decoded by the listener. Prosodic structure is considered to be an essential component of this process which regulates boundary and prominence marking to provide a detailed framework for articulation (see [3] for review). Such prosodic structuring can be further modulated in reference to other linguistic structures, possibly through the syntax-prosody interface that deals with prosodic phrasing and syntactic parsing [e.g., 7,10,12,13]. This interface may become more conspicuous when clearer manifestation of syntactic structure is demanded for resolving syntactic ambiguity that might arise, for example with coordinate structures (e.g., [A] and [B or C] vs. [A and B] or [C]). To resolve syntactic ambiguity, speakers often signify a specific syntactic structure by aligning a major syntactic juncture with a major prosodic juncture such as the one arising with the intonational phrase (IP) boundary.

Evidence of syntactic disambiguation through syntax-prosody mapping has indeed been cumulated in the literature [e.g., 7,10,13,14,18]. The notion of syntax-prosody interface may be related to how syntactic junctures are aligned with prosodic junctures [7,14]. Our view, however, goes beyond a simple mapping between the two junctures of prosody and syntax. Although prosodic boundaries may be phonologically defined, and thus treated in categorical terms, we hypothesize that the syntax-prosody interface can be defined in a much more granular way [3,12]—i.e., both suprasegmental and segmental features associated with a particular prosodic structure may be realized with fine-phonetic details that are tuned to reflect the underlying syntactic structure.

In the present study, we explore this possibility by examining how syntactically ambiguous coordinate structures in American English are resolved through syntax-prosody mapping, and how prosodic-structurally driven phonetic features may be realized in reference to syntactic structure. To this end, we will examine durational measures as well as a voice quality-related segmental feature as phonetic reflexes of prosodic juncture. For the durational features, preboundary lengthening, pause duration, post-boundary (initial) lengthening will be measured; and for the segmental feature, glottalization of vowels in vowel-initial words will be measured.

Glottalization refers to irregular creaky phonation that arise from low open quotient, low airflow and ventricular incursion [4]. Glottalization can be found in various contexts and be facilitated by several factors. [6] explains that glottalization in American English occurs amongst several segments or syllables. In American English, glottalization is most likely found for initial vowels [4,6,8,9]. As [4] suggests, the salient nature of creaky phonation may further enhance contrast amongst other acoustic cues. Therefore, glottalization employment helps with segmental contrasts, and a vowel hiatus occurring across a word boundary is often resolved with glottalization or an epenthetic glottal stop [4,5,9]. Glottalization is also found to occur with a vowel at a larger prosodic boundary [6,8,14] as well as under prominence [6, 8]. Thus, glottalization can serve as markers of both boundary and prominence presumably due to an increase in accompanied phonetic salience, relative to modal voice (cf. [8]). In the present study, we will examine how the degree of glottalization, as measured by spectral tilt measure (H1-H2) and noise-related measures HNR and CPP, may vary with the boundary strength that is mapped onto a major syntactic juncture, and how it may also manifest itself under prominence, especially when it occurs in a pitch-accented context compared to an unaccented context.

2. METHOD

2.1. Participants, materials and procedure

Fourteen native speakers of American English (7 male, 7 female; aged 19-35) participated in the acoustic recording experiment. They had been living in Seoul Korea for a short-term stay.

Speech materials were constructed with coordinate structures containing [Name1] and [Name2] or [Name3] as in *'Anna and Annie or Angie'* which created different (structurally ambiguous) syntactic structures: *[N1] and [N2 or N3]* (the Early Closure construct) *versus [N1 and N2] or [N3]* (the Late Closure construct). The early closure construct has a major syntactic juncture after 'N1', and the later closure construct after 'N2.' A mini dialogue context was used to induce these two types of syntactic parsing to be produced by the participants in different contexts of information structure. As exemplified in Table 1, the mini dialogue contained a question (pre-recorded) and an answer (as the target sentence). When the question was *What is going on?,* the answer sentence to be provided by the participant had 'broad' focus; and when the question was *WHO will come to the party?,* the answer sentence had 'narrow' focus on all the noun phrases; and when the question was *Did they say (Anna and Annie) or (Angie) will come?*, the answer sentence had 'contrastive' focus, in the sense that the late closure construct meant by the question was corrected to be the early closure construct. Thus, the focus used in the present study did not fall on a particular lexical item, but either on the whole utterance or on a particular coordinate structure. We employed these contexts to induce various prosodic phrasing patterns that could be aligned with an intended syntactic structure. The three names were crossed over the three locations (*N1 and N2 or N3*), yielding 6 combinations, each of which was produced with two different syntactic parsings (Early vs. Late).

Table 1. Speech Materials according to Focus Type. Only the Answer category has been recorded. Early Closure has a syntactic juncture before 'and' whereas Late Closure (LC) has one before 'or'. Both cases have been recorded for all focus types with randomization of names and closure.

	Question	Answer	
Broad Focus	What is going on?	Well, (Anna) and (Annie or Angie) are coming.	**EC**
		Well, (Anna and Annie) or (Angie) are coming.	**LC**
Narrow Focus	WHO will come to the party?	Well, (Anna) and (Annie or Angie) will.	**EC**
		Well, (Anna and Annie) or (Angie) will.	**LC**
Contras-tive Focus	Did they say (Anna and Annie) or (Angie) will come?	No. They said, (Anna) and (Annie or Angie) will.	**EC**
	Did they say (Anna) and (Annie or Angie) will come?	No. They said, (Anna and Annie) or (Angie) will.	**LC**

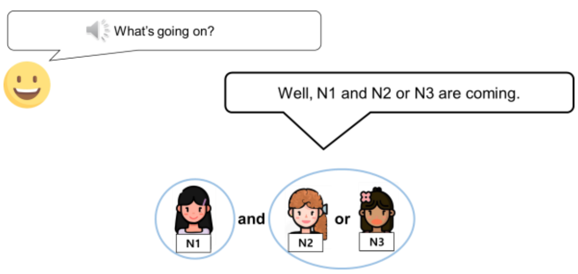

Fig. 1. Speakers were presented the target sentences as shown in this figure. Speakers were to produce the syntactic structure shown in the illustration

In the recording session, two coordinate structures to be produced by the participant were cued by a visual aid as illustrated in Fig. 1. Upon hearing an auditory prominent sentence (pre-recorded by a female speaker), the participant was instructed to produce the written sentence as an answer to a question prompt with a grouping of people as marked by circles displayed on a computer screen.

2.2. Measurements and analyses

Prosodic boundary types and pitch accents of all the obtained tokens were checked by all three authors (two of them were experienced ToBI transcribers), and the data were coded in terms of Boundary (IP, ip, Wd) and pitch accent (pitch accented vs. unaccented. The occurrence of an intermediate phrase (ip) boundary was relatively less frequent, and we excluded the ip tokens from analysis, so that we could observe a clear difference between the IP and the Wd boundaries. Names were generally produced with a pitch accent, so we did not analyse the pitch effects associated with N1, N2, and N3.

To examine glottalization of initial vowels, we used Voice Sauce [16, 17] to measure H1*-H2* (corrected, H1H2c, the difference between amplitudes of the first and the second harmonic) and two noise-related measures CPP (Cepstral Peak Prominence) and HNR (Harmonics to Noise Ratio). For HNR, we used HNR35 in Voice Sause to account for HNR up to 3,500 Hz. According to [9], H1-H2 (difference between amplitude of first and second harmonic) reveals vocal fold constriction with low values indicating creaky voicing. [9] further mentions phonation types can be sufficiently discriminated with additional measurements of CPP. H1H2 and CPP are considered crucial in evaluation. HNR is also a noise-related measurement that is utilized in non-modal voicing [4]. These parameters have been accounted for with Voice Sauce from three time points. The vowel was divided into three equi-interval portions, and these measures were taken from these three portions (Time points 1~3). To examine temporal variation along with the degree of glottalization, we also measured the duration of the phrase-final syllable (preboundary lengthening) and of the phrase-initial syllable (domain-initial/post-boundary lengthening) as well as the pause duration at the prosodic juncture. In the present study, we combined the preboundary lengthening and pause duration to assess the boundary-related lengthening as these two measures are often closely related to each other to signal a prosodic boundary.

A series of linear mixed-effect models were run using lme4 [2] package in R [15] with fixed factors of Boundary (IP, Wd), Closure (EC, LC), and Pitch Accent (accented, unaccented). Dependent variables were H1H2c, HNR35, CPP as well as preboundary lengthening + pause duration and initial duration (i.e., duration of the phrase-initial vowel). Note that we added the focus difference (broad, narrow, contrastive) as a control factor in the models, and do not report their effects in interaction with Boundary and Pitch Accent. Speaker differences have been accounted for with a random intercept for each model.

3. RESULTS

3.1. Syntax-prosody mapping

The results of prosodic transcription indicate that the major syntactic junctures in both the Early closure and Late closure contexts were consistently aligned with an Intonational Phrase (IP). As shown in Fig. 2, for the Early Closure construct ([N1] # and [N2 or N3]), an IP boundary occurred after N1 without exception; and for the Late Closure construct ([N1 and N2] # or [N3]), an IP boundary occurred after N2 without exception. Thus, these are *critical* IP boundary conditions matched with the syntactic junctures. Along with this obligatory IP juncture, speakers occasionally placed an additional IP boundary elsewhere (12%, 14%) as shown in Fig. 2—e.g., for the Early Closure construct, in addition to the obligatory IP after 'N1', they placed an optional '(#)' IP boundary for about 12% of the time ([N1] # and [N2 (#) or N3]).

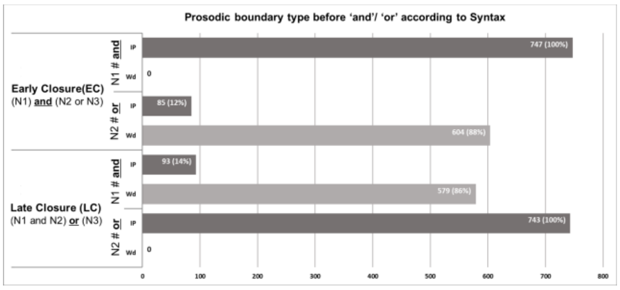

Fig. 2. Distribution of IP boundaries versus Wd boundaries in line with syntactic junctures of two different coordinate structures.

3.2. N2 and N3 (Second, Third Name)

Our qualitative phonetic transcription of the vowels of the names at N2 and N3 (as confirmed by all three authors) indicated that they were all produced with a clear percept of glottalization (creaky phonation) regardless of the boundary conditions. The results of spectral analyses indeed confirmed our auditory impression. There were no significant effects of Boundary at any measured point on any spectral measure: H1H2c, HNR35, and CPP, as can be also inferred from Fig. 3a-c. On the other hand, as shown in Fig. 3d, there was a significant effect of Boundary on preboundary lengthening (N2: β=-53.64, t=-6.12, p<0.001, N3: β=-33.51, t=-4.47, p<0.001). (Note that the pause duration was not included here because these nouns were not followed by any detectable pause even at an IP boundary). There was no boundary effect on the post-boundary syllable duration.

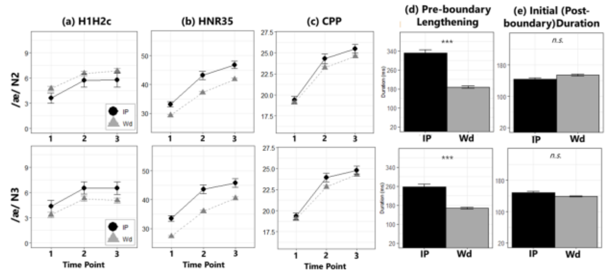

Fig. 3. Boundary effects on H1H2c, HNR35, CPP, Pre-boundary lengthening and Initial (post-boundary) duration for the vowel /æ/ of the second and third names (N2, N3). Error bars represent standard errors. The lower the spectral tilt values, the more glottalized (creakier).

3.3. Boundary effect on 'and' and 'or'

Our initial qualitative inspection of the voice quality of the vowels of 'and' and 'or' also indicated that they were all glottalized regardless of the boundary types. But our initial impression clearly varied as a function of whether these conjunctions were pitch accented or not.

For /æ/ of 'and', there was no Boundary effect on any of the measures, but there were significant interactions between Boundary and Accent for H1H2c, HNR and CPP (H1H2c: β=-2.74, t=-3.42, p<0.001, HNR35: β=10.47, t= 6.68, p<0.001, CPP: β<-0.01, t= 4.62, p<0.001). As shown in Fig. 4a-c (upper panels), 'and' was produced with more glottalization at the IP boundary than at the Wd boundary, especially in the accented condition, as reflected in a clear interaction between Boundary and Accent on H1H2. The other two noise-related measures showed some interaction effects, but not as clearly as H1H2c, indicating that there was no clear-cut boundary effect difference as a function of Pitch Accent. On the other hand, as shown in Fig. 4d, preboundary lengthening (+pause) showed a clear boundary effect (accented: β=308.47, t=5.75, p<0.001, unaccented: β=400.618, t=39.55, p<0.001) with no interaction with Pitch Accent. Interestingly, 'and' also showed a longer vowel duration of itself (initial duration), showing an initial strengthening effect (accented: β=60.55, t=11.03, p<0.001, unaccented: β=31.62, t=2.51, p<0.05),

The vowel of 'or' has also shown significant interaction between Boundary and Pitch Accent for HNR and CPP (HNR35: β=5.4, t=4.43, p<0.001, CPP: β=1.57, t=2.42, p<0.05), and H1H2c at Timepoint 2 (β<-0.01, t= 2.75, p<0.01). As shown in Fig. 4a-c (lower panels), there was a significant Boundary effect on H1H2c in the accented condition, but not in the unaccented conditions. However, results of HNR35 and CPP were not consistent with those of H1H2c (see Fig. 4b-c, lower panels), indicating that the noise-related measures did not provide interpretable results for the boundary effect on glottalization. As for the temporal measure, there was a clear boundary effect on preboundary lengthening (+pause) (accented: β=301.72, t=10.77, p<0.001, unaccented: β=280.47, t=29.23, p<0.001) (Fig. 4d, lower panel). There was also a boundary effect on initial (post-boundary) lengthening but only in the accented condition (β=38.87, t=3.62, p<0.001) (Fig. 4e, lower panel).

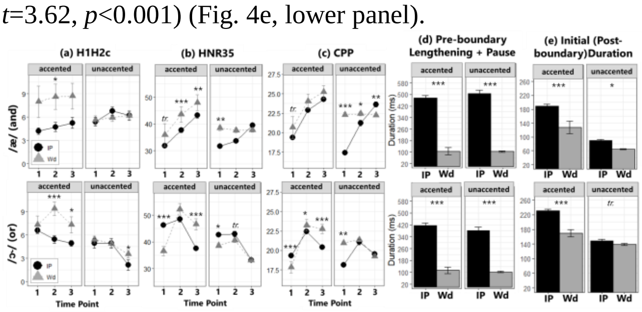

Fig. 4. Boundary effects on H1H2c, HNR35, CPP, Pre-boundary lengthening and Initial (post-boundary) duration for the vowel /æ/ of 'and' and /ɔ˞/ of 'or' regarding pitch accent. Error bars represent standard errors. The lower the spectral tilt values, the more glottalized (creakier).

3.4. Prominence Effect on 'and' and 'or'

In the previous section, interactions between Boundary and Pitch Accent were focused on boundary effects. This section focuses on prominence (Pitch Accent) effects that may vary across different boundary conditions. The pitch accent effects are summarized in Fig. 5. For 'and', H1H2c indicated an increased glottalization at the IP boundary, but the opposite was true at the Wd boundary (Fig. 5a, upper panel). HNR35 and CPP did not show any consistent focus effects, either. For ‘or’, all the measures indicated that if there was anything, vowels tended to be *less* glottalized in the accented than in the unaccented condition, as can be inferred from Fig. 5a-c (lower panels).

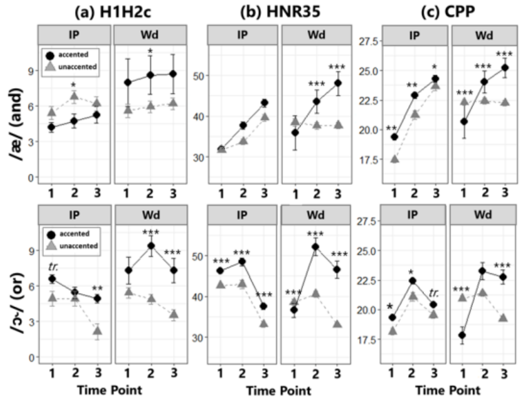

Fig. 5. Prominence effects on H1H2c, HNR35, CPP, Pre-boundary lengthening and Initial (post-boundary) duration for the vowel /æ/ of ‘and’ and /ɔ˞/ of ‘or’ regarding boundary types. Error bars represent standard errors. The lower the spectral tilt values, the more glottalized (creakier).

4. DISCUSSION AND CONCLUSION

In the present study, we examined how different coordinate structures (*N1 and N2 or N3*) in English are expressed by prosodic boundary marking (or phrasing) to disambiguate otherwise structurally ambiguous surface forms, and how preboundary lengthening and glottalization may serve as phonetic reflexes of prosodic junctures that are mapped onto syntactic junctures.

Results indicated that syntactic junctures were invariantly aligned with IP boundaries. With the experimental task of grouping nouns of the phrase *Anna and Annie or Angie* as guided by a visual aid, speakers consistently produced an IP boundary that was aligned with a major syntactic boundary according to an intended syntactic parsing. For the Early Closure construct, a critical IP boundary was placed consistently (100%) at the major syntactic juncture between *N1* and the conjunction ‘and’ ([N1] # and [N2 (#) or N3]), and optionally (12%) an additional IP after N2. Similarly, for the Late Closure construct, a critical IP boundary was placed consistently (100%) at the major syntactic juncture between *N2* and the conjunction ‘or’ ([N1 (#) and N2] # or [N3]), and optionally (14%) an additional IP after N1. The alignment of major syntactic junctures with prosodic junctures are generally in line with the assumptions made in the literature on the Syntax-Prosody interface [10, 13, 14, 18].

Most remarkably, we observed some differential effects of the IP boundary depending on whether the IP boundary was used *critically* at the major syntactic juncture or optionally at a non-critical juncture (with prosody and syntax misalignment). At the critical juncture (with syntax-prosody alignment), a substantial temporal expansion was observed, which came from both preboundary and post-boundary lengthening as well as from the pause duration. Moreover, at this critical juncture, there was also a substantial increase in vowel-initial glottalization (as reflected in H1H2c, HNR and CPP) especially when in the pitch accented condition. On the other hand, at the non-critical prosodic boundary (when it was not aligned with a major syntactic juncture), the boundary effect was quite reduced. Not only did we fail to find the boundary-induced glottalization, but we also observed a reduced magnitude of preboundary lengthening with no detectable pause or post-boundary lengthening.

The phonetic expression of syntactic junctures may be mediated by prosodic boundaries (see [7] for a related discussion), and an IP boundary may be phonologically defined [11]. But, the clear differences in the way that phonetic reflexes of prosodic structure were realized depending on syntactic structures imply that phonetic encoding of prosodic structure makes direct reference to syntax at least to some extent.

Another related point to be made concerns the boundary effect on glottalization in interaction with prominence. Recall that spectral tilt measures revealed some boundary effects at the critical junctures of ‘and’ and ‘or’ only when pitch accented. In fact, this goes hand in hand with observations of [6] where phrase-initial syllables were more likely to be found with glottalization when they were also pitch accented. But we found no clear prominence-induced glottalization, and the opposite was true in some cases. This appears to run counter to what has previously been observed in English [8, 9] in which prominence may be an important trigger of glottalization. We do not have a clear account to offer, but our qualitative observations of those word-initial vowels indicated that they were all creaky regardless of the boundary and pitch accent types. In fact, glottalization may also be employed to resolve the vowel hiatus (in addition to the function of marking a larger prosodic boundary) [5]. Since detecting a word boundary in the vowel hiatus context may become harder in the absence of prominence, the salience of glottalization may come into play, counteracting a possible prominence effect.

Finally, our results leave a question unsolved as to whether the observed differential effects in relation to syntax-prosody alignment may have to do with different lexical properties of the target words. At the *critical* juncture, the target words were always conjunctions ‘and’ and ‘or.’ These were the ones produced with boundary-induced glottalization and post-boundary lengthening. On the other hand, at the non-critical juncture of syntax-prosody misalignment, the target words were nouns (*Annie*, *Angie*, *Anna*), which showed none of these effects. These observed differences may then be due at least in part to initial-strengthening neutralization for content words. [1] states that the lengthening of initial vowels is neutralized since it may otherwise hinder the information of stress within a word. Such prominence may be essential for content words especially for trochaic words as used in the present study, but not for function words which are often susceptible to reduction in various contexts. If so, vowels such as the initial /æ/ for the target word like *Anna* may not be subject to initial strengthening effects that are readily observable for the function words. But, this stress-related account does not explain why there was still a robust preboundary lengthening and a substantial pause at the critical juncture. Thus, our interim conclusion to be made is that our results still support the view that the phonetic encoding of prosodic structure is modulated in reference to syntax.

5. ACKNOWLEDGMENTS

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] Barnes, J. A. (2001). Domain-initial strengthening and the phonetics and phonology of positional neutralization. *Proceedings of the Northeast Linguistic Society (NELS)*, 32, vol. 1, pp.1-20.

[2] Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1-48.

[3] Cho, T. 2022. The phonetics-prosody interface and prosodic strengthening in Korean. In S. Cho and J. Whitman (Eds.), *The Cambridge Handbook of Korean Linguistics* (pp. 248-293). Cambridge University Press.

[4] Davidson, L. 2021. The versatility of creaky phonation: Segmental, prosodic, and sociolinguistic uses in the world's languages. *Wiley Interdisciplinary Reviews: Cognitive Science*, 12(3), e1547.

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

[6] 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.

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

[8] Garellek, M. 2014. Voice quality strengthening and glottalization. *Journal of Phonetics*, 45, 106-113.

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

[10] Jun, S. A. 2009. Prosody in sentence processing. *Ping Li (General ed.), Handbook of East Asian Psycholinguistics, Part III: Korean Psycholinguistics (C. Lee, G. Simpson, & Y. Kim, Eds.)*. London: Cambridge University Press. 423-43.

[11] Ladd, D. R. 2008. *Intonational phonology*. Cambridge University Press.

[12] Mitterer, H., Kim, S., & Cho, T. 2021. Glottal stops do not constrain lexical access as do oral stops. *PloS one*, 16(11), e0259573.

[13] Mitterer, H., Kim, S., & Cho, T. 2021. The Role of Segmental Information in Syntactic Processing Through the Syntax–Prosody Interface. *Language and Speech*, 64(4), 962-979.

[14] Nespor, M., & Vogel, I. 2012. *Prosodic phonology*. Dordrecht: Foris

[15] R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

[16] Shue, Y. L. 2010. *The voice source in speech production: data, analysis and models.* Los Angeles: University of Calinfornia

[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, 1846-1849. 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.