Phonetic variations driven by prosodic structure and their function in speech production

By

Seulgi Shin
M.A., University of Kansas, 2017
M.A., Hanyang University, 2015
B.A., Hanyang University, 2011

Submitted to the graduate degree program in Linguistics and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Chair: Dr. Annie Tremblay

Dr. Allard Jongman

Dr. Joan Sereno

Dr. Jie Zhang

Dr. Michael S. Vitevitch

Date Defended: 27 August 2021

The dissertation committee for Seulgi Shin certifies that this is the approved version of the following dissertation:

Phonetic variations driven by prosodic structure and their function in speech production

Chair: Annie Tremblay

Date Approved: 27 August 2021

Abstract

The present study investigates how prosodic strengthening driven by prosodic boundaries is manifested in relation to its linguistic function in the production of different speech styles, focusing on English plosives /p, t, k, b, d, g/ and nasals /m, n/ in IP-initial position compared to IP-medial position in interactive and read speech. Boundary-induced prosodic strengthening has been observed in domain-initial positions compared to domain-medial positions in articulatory and acoustic dimensions. This phenomenon has been accounted for by the *syntagmatic contrast enhancement* account, which stipulates that the contrast between neighboring consonants and vowels is enhanced, and by the *paradigmatic contrast enhancement* account, which posits that phonological contrasts such as voicing and manner of articulation contrasts are enhanced. By testing these two accounts in read and interactive speech, the present study aimed to provide a more comprehensive understanding of the mechanisms behind boundary-induced prosodic strengthening and investigate potential factors that modulate boundary-induced prosodic strengthening.

In an interactive speech task, disyllabic English words that contained word-initial plosives and nasals were elicited in IP-initial and IP-medial positions in a task where a pair of participants interacted to find out what was going on in Animal Village. In each trial, one participant was given pictorial information, and the information needed to be verbally delivered to the other participant whose task was to ask a question and choose the scene among the given pictorial options based on the verbal description. In a read speech task that was administered with a few days later, the same target words were elicited in IP-initial and IP-medial positions in a task where participants were asked to read written sentences. The present study analyzed the data of 18 participants who participated in both the interactive and read speech tasks. Prosodic

strengthening was evaluated by comparing durational, amplitudinal, and spectral measurements of the initial consonant in IP-initial and IP-medial positions.

The analysis revealed that the patterns of prosodic strengthening were contingent on speech sounds and its acoustic correlates. For plosives, the VOT results showed evidence of syntagmatic contrast enhancement in read speech whereas they showed evidence of paradigmatic contrast enhancement in interactive speech. Spectral acoustic correlates showed paradigmatic contrast enhancement in both interactive and read speech when there was prosodic strengthening. Based on the fact that syntagmatic contrast enhancement was only observed in read speech, speech style influences prosodic strengthening and its linguistic function such that speakers may simply mark a prosodic juncture without the need for directly delivering messages to listeners in read speech, rather than having to help listeners identify the speech sound better by making it phonologically more distinct from other speech sounds as in interactive speech. Unlike plosives, except for durational acoustic correlates, prosodic strengthening on nasals showed the pattern of syntagmatic contrast enhancement in both read and interactive speech possibly due to fact that English nasals are phonetically and phonologically distinct enough from non-nasal consonants so speakers may not need to enhance the distinction between them when delivering messages to listeners in interactive speech.

Overall, the present study suggests that prosodic strengthening driven by prosodic boundary and its linguistic function can be modulated by whether or not speech style is interactive (i.e., interactive vs. read speech) and by whether speech sounds are phonetically and phonologically distinct enough when compared to other speech sounds. Prosodic strengthening driven by prosodic boundary and its linguistic function should be understood in the interaction with different factors that modulate its patterns.

Acknowledgments

Thanks to my professors, family, and friends, I have grown as a researcher and as a person throughout this long and difficult journey. Thank you all for being there for me. I cannot express more how grateful I am for the support and encouragement that I have received, and I would not have completed this dissertation on my own.

I would like to first express my sincere appreciation to my advisor, Annie Tremblay, for guiding me with patience and enthusiasm throughout my time in the graduate program. You always pushed me further to produce quality research, and your passion for research showed me how I would go on with my research. Working with you expanded my areas of research interest, and it has now become a priceless asset to me. I would like to thank my committee members, Allard Jongman, Joan Sereno, Jie Zhang, and Michael Vitevitch for your encouragement and invaluable comments that help improve my dissertation. I am thankful to KU Linguistics for providing me with the opportunity and resources to pursue my research.

I would like to thank my family for their patience and love. Special thanks to my husband, Charlie, for being a lovely husband, my best friend, and an inspiring collaborator. In moments of difficulty, you supported me to gain a new perspective on life and research and move forward with more confidence. I learned from you how to be kind to others and especially to myself. Mom and dad, thank you. Your unconditional love and belief in me fueled my passion toward my research. Thanks to my siblings, Sori and Taeji, for giving me great comfort and joy. I always felt connected to you two and never felt that I was alone. I will miss the times when I stayed up all night working with you two. I am also grateful that I have a wonderful brother-in-law, Jihoon, and niece, Minseol, who provided unlimited emotional support for me. I would like

to thank my family in the US, Ellen, Kathleen, Dany, and Peter for giving me warmth and stability in this challenging time of my life.

Lastly, thank you, Nick and Lena, who went through exciting and difficult times in Lawrence together. Your friendship meant a lot to me, and I will remember our silly conversations. I also want to thank my friends in Sound group and fellow students in the department for their feedback and assistance for my projects. I would like to say thank you to all the friends whom I have met along the way.

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction

Prosody plays an important role in the speech planning stages that speakers go through when producing language. The model of speech production proposed by Levelt and colleagues (Levelt, Roelofs, & Meyer, 1993, 1999) includes five speech planning stages: (i) activation of lexical concepts based on speakers' intention (semantic encoding); (ii) retrieval of lemmas with associated syntactic information from the mental lexicon (lexical encoding); (iii) specification of morphological structure based on the syntactic and metrical characteristics of the lemmas (morphological encoding); (iv) activation of phonological forms taking into account the metrical frame of the lemma and its syllabification (phonological encoding); and (v) retrieval, concatenation, and contextual adjustment of stored syllabic gestural scores for a word (phonetic encoding). Importantly, in connected speech where words are concatenated, a prosody generator sets prosodic frames for syllabic templates and for larger prosodic contexts in terms of how to group words into chunks (boundary marking) and how to deliver information (prominence marking), along with associated prosodic parameters, including duration, loudness, f0 and pauses (see Keating & Shattuck-Hufnagel, 2002; Keating, 2006). Researchers have worked on specifying how this prosodic structure is implemented—that is, how the prosodic frames stored in speakers' linguistic system are planned and realized at a phonetic level.

The proposed models of prosodic structure stipulate that prosodic structure is hierarchically organized based on prosodic grouping and prominence marking (e.g., Beckman & Edwards, 1990, 1994; Beckman & Pierrehumbert, 1986; Bolinger, 1958, 1965; Ladefoged, 1975; Nespor & Vogel, 1986; Pierrehumbert & Beckman, 1988; Selkirk, 1978, 1986; Vanderslice & Ladefoged, 1972). The organization of prosodic structure and its acoustic correlates differ from language to language; yet, researchers have found similarities across many languages in the

phonetic realization of segments as a function of the size of prosodic boundaries and prosodic prominence (e.g., Byrd, 2000; Byrd & Saltzman, 1998; Cho, 2006; de Jong, Beckman, & Edwards, 1993; Edwards, Beckman, & Fletcher, 1991; Fowler, 1995; Mücke & Grice, 2014; Shattuck-Hufnagel & Turk, 1998; Turk & Sawusch, 1997; Turk & White, 1999; Wightman, Shattuck-Hufnagel, Ostendorf, & Price, 1992). More specifically, prosodically salient positions such as words at edges of prosodic phrases and words under prosodic prominence yield lengthening and some sort of (possibly cumulative) strengthening effect on the phonetic realization of segments. This has been referred to as *prosodic strengthening.*

Prosodic strengthening appears to be present in those prosodically salient positions across different languages (cf. Kuzla & Ernestus, 2011) but the strengthening patterns driven by prosodic boundaries are found to be quite inconsistent compared to those driven by prominence. Based on previous studies where prosodic strengthening driven by prosodic boundary is investigated with limited sets of segments in lab-setting reading experiments, it is difficult to see how prosodic strengthening operates in a larger set of segments that differ in many features including manner, voicing, and place of articulation, and what other potential factors can influence prosodic strengthening. For a better understanding of prosodic strengthening driven by prosodic boundaries, its generalizability and idiosyncrasy across different types of segments need to be tested. Moreover, given that the essential purpose of speech production is to communicate with interlocutors, it is important to investigate how prosodic structure is manifested in interactive speech where speakers have interactive communication. In fact, previous research has found that information about the prosodic structure of a language modulates listeners' recognition of words (e.g., Christophe, Peperkamp, Pallier, Block, & Mehler, 2004; Salverda, Dahan, & McQueen, 2003; Salverda, Dahan, Tanenhaus, Crosswhite, Masharov, & McDonough,

2007; Shin & Tremblay, 2018; Tremblay, Broersma, Coughlin, & Choi, 2016; Tremblay, Cho, Kim, & Shin, 2019;). Thus, prosodic strengthening might manifest itself differently in read speech and interactive speech when given the absence versus presence of interlocutors (listeners), respectively. For example, speakers may be more likely to produce prosodic strengthening as a way of helping listeners' perception of words (e.g., helping listeners reduce lexical competition) in interactive speech than in read speech. The present study aims to address these issues by investigating how prosodic strengthening is manifested through English stops (i.e., [p, t, k, b, d, g]) and nasals (i.e., [m, n]) that differ in voicing and/or place of articulation in read (non-interactive) and interactive speech, and how its phonetic implementation can be explained by two different accounts: syntagmatic vs. paradigmatic contrast enhancement accounts.

Phonetic Encoding of Prosodic Structure and Possible Linguistic Accounts

Previous studies have shown that segments in domain-initial prosodic positions are realized with lengthening and more exaggerated articulation (reflected in relevant articulatory and acoustic parameters) than those in domain-medial prosodic positions (e.g., Bombien, Mooshammer, & Hoole, 2013; Byrd et al., 2000; Byrd et al., 2006; Byrd & Saltzman, 2003; Cho & Keating, 2001; Cho & Jun, 2000; Cho & Keating, 2009; Dilley, Shattuck-Hufnagel, & Ostendorf, 1994, 1996; Fougeron, 2001; Fougeron & Keating, 1997; Georgeton & Fougeron, 2014; Hsu & Jun, 1998; Keating, Cho, Fougeron, & Hsu, 2003; Keating, Wright & Zhang, 1999; Kim, Kim, & Cho, 2018; Pierrehumbert & Talkin, 1992). Articulatorily, the observed patterns at prosodic boundary junctures have been accounted for by the π-gesture (prosodic gesture) theory proposed by Byrd and Saltzman (e.g., Byrd, 2000; Byrd, Kaun, Narayanan, & Saltzman, 2000;

Byrd, Krivokapić, & Lee, 2006; Byrd & Saltzman, 1998; Byrd & Saltzman, 2003; Saltzman, 1995). A π-gesture is an abstract gesture that influences the temporal realization of articulatory gestures at prosodic boundaries. Under the influence of a π-gesture, articulatory gestures slow down their movements, yielding lengthening, and become less overlapped with each other. The influence of a π-gesture on articulatory gestures increases towards a prosodic juncture and wanes gradually away from it. In line with this account, Fougeron and Keating (1997) have suggested, as one of the accounts, that the mechanisms for lengthening and decreased overlap between articulatory gestures might be responsible for the larger gestural magnitude in terms of linguopalatal contact (articulatory strengthening) at prosodic boundaries partly under the observation of a correlation between the degree of lengthening in domain-initial prosodic positions and the degree of gestural magnitude in some language such as French and Korean (e.g., Cho & Keating, 2001; Fougeron, 2001). The lengthening of articulatory gestures and decreased overlap between them allows more time for them to reach the target without undershoot. However, this correlation is weaker in English (Fougeron & Keating, 1997), which means that lengthening does not necessarily accompany a larger gestural magnitude. This might imply that while π-gesture controls the lengthening of phonological units, articulatory strengthening can be operated by different mechanisms in English. Thus, more mechanisms that can possibly explain the patterns of strengthening have been explored in different languages.

Two different accounts have been proposed to further explain why prosodic strengthening takes place in domain-initial prosodic positions across languages. One of the accounts is referred to as *syntagmatic contrast enhancement* or as CV enhancement (e.g., Cho & Keating, 2001; Cho & Keating, 2009; Kim, Kim & Cho, 2018). This view suggests that a consonant increases its consonantality compared to a neighboring vowel and a vowel increases

its vocalicity compared to a neighboring consonant such that prosodic strengthening enhances the sonority contrast (i.e., difference in articulatory openness) between consonants and vowels (Fougeron & Keating, 1977; see also Straka, 1963)[1]. For example, domain-initial prosodic positions, compared to domain-medial prosodic positions, have shown decreased nasality for /n/ (e.g., Cho & Keating, 2009; Fougeron, 1999, 2001), increased linguopalatal contact for voiceless stops and /n/ (e.g., Cho & Keating, 2001; Fougeron, 2001; Fougeron & Keating, 1997; Keating et al., 2003; Keating, Wright & Zhang, 1999), longer Voice Onset Time (VOT) for voiceless stops (e.g., Cho & Keating, 2009; Jun, 1993; Lisker & Abramson, 1967; Pierrehumbert & Talkin, 1992), and a greater spatial expansion of articulatory gestures such as the tongue closing gesture for consonants (Byrd et al., 2006), all of which contribute to increasing the consonantality of the consonant in relation to the subsequent vowel.

Prosodic strengthening in domain-initial prosodic positions has also been interpreted as one type of local hyperarticulation (de Jong, 1995) that enhances phonological contrasts, also known as *paradigmatic contrast enhancement* or as phonological contrast enhancement (e.g., Cho & Jun, 2000; Georgeton & Fougeron, 2014; Hsu & Jun, 1998). This particular pattern of enhancement is language-specific and depends on the phonological system of the language. In Taiwanese, for example, a longer VOT for the aspirated stop /k^h/, a shorter VOT for the voiced stop /b/, and no change in VOT for the unaspirated stop /t/ were found in domain-initial positions compared to domain-medial positions (Hsu & Jun, 1998), suggesting that the aspirated and voiced stops became more phonologically distinctive from each other in domain-initial positions.

[1] The sonority contrast here is primarily defined by articulatory openness. The prosodic strengthening that ultimately supports syntagmatic contrast enhancement was early on observed from linguopalatal contact at domain-initial positions where consonants showed greater linguopalatal contact and vowels showed less linguopalatal contact, resulting in the increased difference in openness between neighboring consonants and vowels (e.g., Fougeron & Keating, 1997).

Similarly, the Korean aspirated stop /p^h/ and lenis stop /p/ showed a longer VOT but the fortis stop /p*/ (unaspirated) showed a shorter VOT in domain-initial positions such as IP-initial and AP-initial positions than in domain-medial positions (Cho & Jun, 2000). Unlike Taiwanese, the phonological distinction in terms of VOT in Korean was enhanced by differentiating aspirated and lenis stops from fortis stops, presumably because Korean aspirated and lenis stops in domain-initial positions are going through a diachronic change such that the VOT range is merged for aspirated and lenis stops and instead F0 primarily distinguishes them (e.g., Kang & Guion, 2008; Kim, Beddor, & Horrocks, 2002; Lee & Jongman, 2012; Lee, Politzer-Ahles, & Jongman, 2013). Paradigmatic contrast enhancement was also found for vowels in Georgeton and Fougeron (2014), where French vowels (i.e., /i, e, ɛ, a, y, ø, œ, u, o, ɔ/) enhanced their features to be more different from one another in IP-initial position: lip opening was larger for all vowels in IP-initial position, with unrounded vowels showing more lip opening than rounded vowels; F2 and F2-F1[2] were lower for back vowels than front vowels in IP-initial position compared to IP-medial position; and low and mid vowels were realized with a higher F1 than high vowels in IP-initial position compared to IP-medial position. Studies that have looked at vocalic gestures across a prosodic boundary are suggestive of paradigmatic contrast enhancement (e.g., Byrd et al., 2006; Shin, Kim, & Cho, 2015; Tabain 2003; Tabain & Perrier 2005). For example, Shin at al. (2015), who examined /a/-to-/i/ tongue body movement in Korean, observed a greater spatial expansion of the closing gesture for the high vowel /i/ across an IP boundary compared to within an IP (i.e., across a Wd boundary). These studies suggest that

[2] French back vowels /u, o, ɔ/ are known to be focal vowels that have a sharp concentration of energy due to the merging of F1 and F2 (Schwartz, Boë, Vallée, & Abry, 1997; Vaissière, 2011). F2-F1, thus, can be used as a cue to back vowels. Note that /i/ and /y/ are also defined as focal vowels.

boundary-induced strengthening can have the linguistic function of enhancing phonological contrasts in the language.

English Prosodic Structure

Researchers have proposed different terms and systems for the hierarchical organization of prosodic constituents (Beckman & Pierrehumbert, 1986; Hayes, 1989; Nespor & Vogel, 1986; Pierrehumbert & Beckman, 1988; Selkirk, 1978, 1980). The present study will adopt Beckman and Pierrehumbert's (1986) proposal for the analysis of prosodic structure. From the largest to the smallest prosodic unit above the foot, English prosodic constituents have been analyzed as being comprised of the Utterance (Ut), the Intonational Phrase (IP), the Intermediate Phrase (iP), which is more or less equivalent to the Phonological Phrase (PP) in Nespor and Vogel (1986) and Hayes (1995), and the Prosodic word (Wd) (see also Wightman et al., 1992 and Shattuck-Hugnagel & Turk, 1996). The Utterance, which is the largest constituent in which phonological rules can apply, have one or more IPs that are phonetically marked by substantial lengthening of the final syllable at the end of the phrase and is often followed by a pause (e.g., Ladd & Campbell, 1991; Lehiste, Olive, & Streeter., 1976; Selkirk, 1984; Wightman et al., 1992). Prosodic phrasing is determined by many factors such as syntactic and prosodic length, complexity of the sentence, speech rate, balancing of prosodic constituent sizes, contrastive focus, semantic coherence, and, punctuation in written language (e.g., D'Imperio et al., 2005; Elordieta, Frota, & Vigário, 2005; Ferreira, 1991, 1993; Frazier, Clifton & Carlson, 2004; Frota, 2014; Gee & Grosjean, 1983; Hellmuth, 2004; Kalbertodt, Primus, & Schumacher, 2015; Krivokapić, 2007; Nespor & Vogel, 1986; Selkirk, 2000, 2005; Watson & Gibson, 2004). The hierarchical structure of prosodic constituents is illustrated in Figure 1.

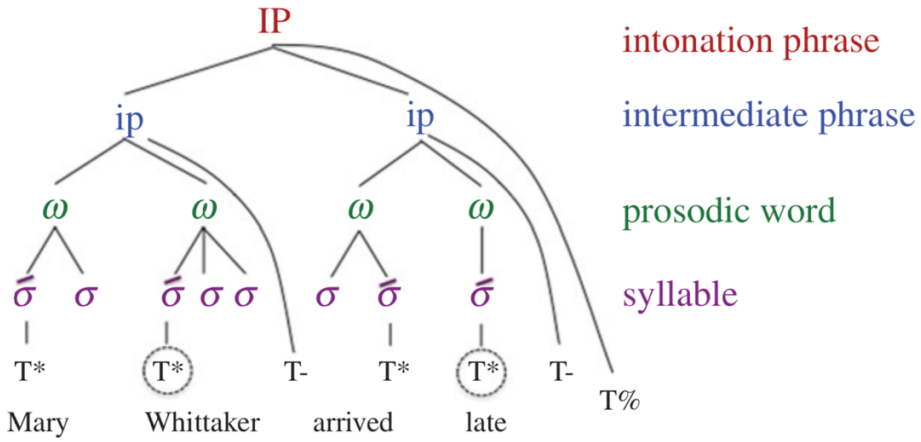


Figure 1: A schematic prosodic structure in English. This figure is borrowed directly from Krivokapić (2014, p. 2). *T* represents tone that can either be low tone (L) or high tone (H). *T** represents a pitch accented tone. The circled *T** represents the nuclear pitch accent. *T-* represents a phrase tone. *T%* represents a boundary tone.

Languages can have different levels of prominence. English has at least two levels of prominence: lexical stress at the word level and intonational pitch accent at the phrase level (see Beckman & Pierrehumbert, 1986; Keating, 2006). At the word level, syllables can have primary or secondary stress with fully realized vowels. At the phrase level, English has been analyzed as having the following pitch accents: H*, L* L*+H, L+H*, and !H, which are determined by the type of information that speakers want to convey with a word in a sentence. The pitch accent that occurs toward the end of an IP is called the nuclear pitch accent, and earlier ones are called prenuclear pitch accents (see Beckman & Ayers, 1997). These types of phrasal-level prominence closely associate with different levels of prosodic constituents (Beckman & Edwards, 1990, 1994). IPs contain at least one nuclear pitch accent (L* or H*) followed by a phrase tone (L- or H-), followed by a boundary tone (L% or H%). IPs can have one or more iPs that contain at least one nuclear pitch accent followed by a phrase tone. An iP is the domain of F0 range lowering after an accented syllable (catathesis or downstep) (Beckman & Pierrehumbert, 1986). An iP can have one or more Wds. Wds have at least one stressed syllable that potentially receives a nuclear

or pre-nuclear pitch accent at the phrase level. As shown in Figure 1, for example, the word *Mary* receives a pre-nuclear pitch accent, and the word *Whittaker* receives a nuclear pitch accent. The two words consist of an iP [*Mary Whittaker*] of which the nuclear pitch accent is the head. This iP [*Mary Whittaker*] ending with a phrase tone (T-) and the other iP [*arrived late*] also ending with a phrase tone (T-) together form of an IP that ends with a phrase tone (T-) and a boundary tone (T%). This hierarchical prosodic structure will be employed in the analysis of the effect of prosodic structure on speech production in previous studies on English and in the present study. However, although there are (at least) two levels of prominence in English, prominence will generally refer to phrasal pitch accent in the present study since word-level stress will not be investigated.

Phonetic Variations as a Function of Prosodic Boundaries (and Prominence) in English

The phonetic realization of segments has been examined for different prosodic boundaries in English (e.g., Byrd et al., 2006; Cho & Keating, 2009; Cole et al., 2007; Dilley et al., 1996; Fougeron & Keating, 1997; Lisker & Abramson, 1967; Pierrehumbert & Talkin, 1992). As with many other languages, the existing data in English generally suggest that segments in domain-initial prosodic positions and segments in prominent (i.e., pitch accented) words tend to be produced with lengthening and exaggerated articulations in acoustic and articulatory dimensions compared to segments in domain-medial positions and segments in unaccented words (respectively). Many studies investigating prosodic strengthening in English interpreted the patterns driven by the different levels of prosodic boundaries and prominence in relation to the syntagmatic and paradigmatic contrast enhancement accounts; yet, on closer

examination, these prosodic strengthening patterns in relation to linguistic functions vary greatly among the different studies.

Pierrehumbert and Talkin (1992) reported how prosodic boundary (phrase-initial vs. phrase-medial)[3] and accentuation (accented vs. unaccented) influenced the phonetic realization of /h/ (e.g., hogfarmer), /ʔ/ (e.g., August) (glottalization at the beginning of vowels), and /t/[4] (i.e., tomahawk) in the initial syllable in English. Two speakers were instructed to read sentences that were blocked by prosodic condition, where the target syllable was: (i) pitch accented in phrase-medial condition; (ii) accented with contrastive focus in phrase-medial condition; (iii) accented in phrase-initial condition; and (iv) unaccented in phrase-medial condition. The results showed that prosodic boundary and accentuation influenced segments durationally and gesturally: /h/ was produced with more lengthening and less voicing (as reflected in lower Root Mean Square (RMS) energy) after a phrase boundary (i.e., phrase-initial condition) than after a word boundary (i.e., phrase-medial conditions). Gestural magnitude (magnitude of CV gestures) was inferred for /h/ from plotting the RMS energy for /h/ against the RMS energy for the following vowel, with this measure being found to increase in phrase-initial condition compared to the phrase-medial conditions. The results also showed that accented syllables (i.e., accented conditions) increased CV gestural magnitude for /h/, making /h/ more consonant-like and the following vowels more vowel-like, compared to unaccented syllables (i.e., unaccented condition). Based on the production of /h/, both phrase-initial positions (compared to phrase-medial positions) and

[3] Pierrehumber and Talkin (1992) referred to the prosodic boundary of their interest as a phrase boundary and compared a phrase boundary (phrase-initial) and a word boundary (phrase-medial). Based on their description, the phrase boundary was likely the iP boundary rather than the IP boundary.

[4] Pierrehumbert and Talkin (1992) initially did not aim to investigate /t/ in their stimuli, but they included /t/ with a small subset of data in their analysis.

accented syllables (compared to unaccented syllables) increased the distinction between /h/ and the following vowel.

Similarly, with a small subset of the data, the authors observed that speakers produced /t/ with greater aspiration duration (VOT) in phrase-initial position than in phrase-medial position (including all three phrase-medial conditions). Even though the effect was smaller than the effect of prosodic boundary, the comparison between accented and unaccented conditions revealed that /t/ in an accented syllable was also produced with greater aspiration duration compared to that in an unaccented syllable. Like in the production of /h/, thus, the production of /t/ was interpreted to increase the distinction between /t/ and the following vowel in phrase-initial positions (compared to phrase-medial positions) and accented syllables (compared to unaccented syllables), but alternatively it can also be interpreted as the increase in the distinction between /t/ and its voicing counterpart /d/, possibly supporting paradigmatic contrast enhancement. In addition, as found in Dilley et al. (1996), greater glottalization for /ʔ/ (represented by a more reliable and noticeable occurrence of glottalization) was found at a phrase-initial position than at a phrase-medial position, which can be interpreted as prosodic strengthening without syntagmatic or paradigmatic contrast enhancement. Overall, these results generally show prosodic strengthening in phrase-initial positions (compared to phrase-medial positions). More specifically, the patterns for /h/ induced by prominence showed syntagmatic contrast enhancement, and those for /t/ can be interpreted as syntagmatic contrast or paradigmatic contrast enhancement. Thus, the results from Pierrehumbert and Talkin (1992) raised the possibility that syntagmatic contrast or paradigmatic contrast enhancement can be induced by both prosodic boundary and prominence in English.

Studies that employed linguopalatal contact as an index of the degree of oral constriction to investigate phonetic variations in domain-initial versus domain-medial positions in English

also reported prosodic strengthening (e.g., Cho & Keating, 2009; Fougeron & Keating, 1997). For example, Fougeron and Keating (1997) examined linguopalatal contact for /n/ and the following vowel /o/ in the initial syllable of the iteration of /no/ sequences to test the effect of levels of prosodic boundaries on segments and to determine if the resulting articulatory variations were cumulative as a function of the prosodic boundary hierarchy. The authors compared domain-initial positions with domain-medial positions that differed in their prosodic strengths. In order to control for lexical stress, they varied the positions of lexical stress in the stimuli, but pitch accenting was not controlled. They found for /n/ that the linguopalatal contact was greater and the acoustic duration was longer in domain-initial positions than in domain-medial positions. This effect was generally cumulative as a function of the prosodic boundary hierarchy such that the greatest linguopalatal contact and the longest consonantal duration were found in Ut-initial or IP-initial position and decreased along the prosodic boundary hierarchy from Ut-initial or IP-initial position to syllable-initial position. The linguopalatal contact for the following vowel /o/ was more reduced in domain-initial positions than in domain-medial positions, showing that speakers made the vowel more vowel-like. Thus, the results provide further evidence of syntagmatic contrast enhancement by showing that a consonant increases its consonantality and the following vowel increases its vocalicity. Nevertheless, given that the speakers were three phoneticians (including one of the authors), the hypotheses need to be tested with naïve participants. More importantly, since pitch accenting was not controlled in their stimuli, one cannot conclude that syntagmatic contrast enhancement stemmed solely from boundary-induced strengthening.

Cho and Keating (2009) further examined the linguopalatal articulation and acoustic correlates of prosodic strengthening in English /t/ and /n/ in the initial syllable of non-words (i.e.,

Tababet /tɛbəbɛt/, Nebaben (/nɛbəbɛn/) to investigate the phonetic variations of these two segments in Ut-initial and Ut-medial positions. Prominence (accented vs. unaccented) was also manipulated on target words that began with /t/ and /n/ in the study. Four native American English speakers (trained phoneticians) were asked to read the presented sentences and repeat each sentence five-times in a block. Linguopalatal contact as an index of the degree of oral constriction revealed that /t/ and /n/ were produced with larger linguopalatal contact in Ut-initial position than in Ut-medial position, suggesting that /t/ and /n/ became more consonant-like in Ut-initial position than in Ut-medial position. Linguopalatal contact for /t/ and /n/ did not show prominence-induced strengthening. RMS burst energy for /t/ was lowered in Ut-initial position compared to Ut-medial position.[5] Contrary to the effect of prosodic boundaries, RMS burst energy for /t/ was increased in accented syllables compared to unaccented syllables, making /t/ more consonant-like. In terms of VOT for /t/, although no influence of prominence was found, there was an interaction between prosodic boundaries and prominence such that a longer VOT was reliably found in Ut-initial position than in Ut-medial position only when the target words were unaccented. The VOT pattern was interpreted by the authors as suggesting an enhancement of the consonantality of /t/ in Ut-initial position compared to Ut-medial position. For /n/, nasal energy[6] was increased in accented syllables compared to unaccented syllables. The authors interpreted this increased nasal energy as enhancing the [+nasal] feature, suggesting paradigmatic contrast enhancement in accented syllables compared to unaccented syllables. In

[5] Cho and Keating (2009) explained that the lowered RMS burst energy was due to speed of the CV opening that is positively correlated with the release burst energy. Based on the observation in Cho (2006), the CV lip opening movement in English was faster in accented syllables compared to unaccented syllables, but the faster lip opening movement was not found in domain-initial position compared to domain-medial position. If the CV lip opening is somewhat slower in Ut-initial position than in Ut-medial position due to greater linguopalatal contact, RMS burst energy can be reduced.

[6] In Cho and Keating (2009), nasal energy was measured by taking the means over the RMS acoustic energy of the entire nasal duration.

contrast, nasal energy was reduced in Ut-initial position compared to Ut-medial position when the target words were unaccented, making the nasals more consonant-like with less sonority. Overall, Cho and Keating (2009) concluded that prosodic boundaries and prominence influenced /t/ and /n/ differently: /t/ and /n/ in Ut-initial position enhanced their consonantality compared to those in Ut-medial position, suggesting syntagmatic contrast enhancement; while /t/ in accented syllables enhanced its consonantality (at least in terms of RMS burst energy) compared to that in unaccented syllables (suggesting syntagmatic contrast enhancement), /n/ in accented syllables made itself more distinctive from other non-nasal consonants by enhancing its nasal feature [+nasal] (suggesting paradigmatic contrast enhancement). Again, the study raises the question of whether the results with phonetically trained speakers and a small set of segments (i.e., /t/ and /n/) are generalizable to a larger sample of naïve speakers. Moreover, the interpretation of the effect of prosodic boundary is problematic because syntagmatic contrast enhancement and paradigmatic contrast enhancement cannot be teased apart with the voiceless stop /t/ being tested without its counterpart, the voiced stop /d/. For example, a longer VOT observed in Ut-initial position compared to Ut-medial position can suggest syntagmatic contrast enhancement as the authors concluded, but it can also suggest paradigmatic contrast enhancement in that /t/ might have enhanced its voicelessness by increasing its VOT in contrast to /d/. Thus, it is important to include segments such as voiced stops in order to tease apart these two different accounts.

More recently, Kim, Kim, and Cho (2018) investigated the effect of prosodic boundary (IP-initial vs. IP-medial positions) and prominence (accented with contrastive focus vs. unaccented) on English stops by including both word-initial voiceless and voiced stops (i.e., /p, t, b, d/). They used disyllabic words that were stressed either on the first syllable (i.e., trochaic: *panel, tanner, banner, Daniel*) or on the final syllable (i.e., iambic: *panache, Tenise, banal,*

Denise) in order to further manipulate lexical stress. Participants were asked to read the sentences presented orthographically. The study found that voiceless stops in trochaic words stayed constant in terms of VOT in IP-initial and IP-medial positions, whereas voiced stops increased their VOT in IP-initial position compared to IP-medial position. As a result, the difference in VOT between voiceless and voiced stops was smaller in IP-initial position than in IP-medial position. On the other hand, the effect of prominence on voiceless and voiced stops in trochaic words showed different patterns such that both voiceless and voiced stops increased their VOT in accented syllables compared to unaccented ones, yet the difference in VOT between voiceless and voiced stops was greater in accented syllables than in unaccented ones. Similarly, voiceless stops in iambic words were also constant in VOT across IP-initial and IP-medial positions whereas voiced stops increased in VOT in IP-initial position compared to IP-medial position, resulting in a smaller difference in VOT between voiceless and voiced stops in IP-initial position than in IP-medial position. Neither voiceless nor voiced stops in iambic words changed in VOT between in accented and unaccented syllables.

The authors concluded that for both trochaic and iambic words, although voiceless stops did not show boundary-induced prosodic strengthening, the results suggest syntagmatic contrast enhancement for the following reasons. First, the difference between voiceless and voiced stops was smaller in IP-initial position than IP-medial position, and thus there was no enhancement of the distinction between voiceless and voiced stops in IP-initial position. Second, at least for trochaic words, when the interaction between prosodic boundary and prominence was separately analyzed for voiceless and voiced stops, the authors found an effect of boundary on voiceless stops in unaccented syllables that did not receive contrastive focus. On the other hand, at least for trochaic words, the results suggest that prominence-induced prosodic strengthening created

paradigmatic contrast enhancement, with the difference between voiceless and voiced stops being greater in an accented syllable than in an unaccented syllable. Based on these results, the authors proposed that prosodic strengthening for voiceless and voiced stops in English is realized in different ways depending on its source (i.e., prosodic boundary or prominence), with boundary-induced prosodic strengthening yielding syntagmatic contrast enhancement but with prominence-induced prosodic strengthening yielding paradigmatic contrast enhancement. However, several questions remain to be answered, such as whether prosodic strengthening can enhance other types of phonological contrast such as place of articulation, and whether these patterns of prosodic strengthening extend to other speech styles.

These questions were addressed to some degree in Cole et al. (2007), who investigated prosodic strengthening in more natural speech by analyzing a corpus based on FM Radio news speech (American English) (Boston University Radio News corpus, Ostendorf, Price & Shattuck-Hufnagel, 1995). The corpus consists of four news stories read by four professional announcers in radio speech style in a lab setting. The study examined the phonetic implementation of word-initial and pre-vocalic stop consonants /p, b, t, d, k, g/ in IP-initial position compared to IP-medial position and in accented syllables compared to unaccented syllables in order to test syntagmatic contrast enhancement (i.e., CV contrast enhancement) and paradigmatic contrast enhancement (i.e., phonological contrast enhancement), including both voicing and place of articulation contrasts. For the investigation of boundary-induced strengthening, the authors could only examine the realization of /t/ and /d/ in IP-initial and IP-medial positions in unaccented syllables[7] due to the rarity of instances of accented words in IP-

[7] Since the analysis included both content and function words, most of the unaccented words in Cole et al. (2007) were function words. A preliminary analysis was conducted to see whether the acoustic measurements patterned differently when function words were included in or removed from the data set. This analysis revealed that there was no substantial difference between the two data sets.

initial position for other segments within the corpus. VOT, f0 at the onset of voicing, closure duration, and burst amplitude were measured.

The results revealed that /t/ and /d/ in IP-initial position in unaccented syllables did not show prosodic strengthening compared to the same segments in IP-medial position. Instead, Cole et al. (2007) found less variability in the phonetic realization of /t, d/ in IP-initial position than in IP-medial position, suggesting a more precise control of articulation in IP-initial position than in IP-medial position. The word-initial and pre-vocalic stop consonants /p, b, t, d, k, g/ in IP-medial position were examined to investigate the influence of prominence (accented vs. unaccented). f0 and closure duration were generally found to increase for both voiceless and voiced stops in accented syllables compared to unaccented syllables. These results suggest that even though prominence tends to increase f0 and closure duration for both voiceless and voiced stops, the directionality of this effect does not enhance the distinction between voiceless and voiced stops. In terms of VOT, two speakers (announcers) showed an increased VOT for voiceless stops and a decreased VOT for voiced stops in accented syllables compared to unaccented syllables, while the other two speakers showed an increased VOT for both voiceless and voiced stops. At a first glance, the phonological contrast seemed to have been enhanced by the two speakers and the CV contrast seemed to have been enhanced by the other two speakers. Even when VOT was increased for both voiceless and voiced stops, however, the difference in VOT between voiceless and voiced stops in accented syllables was greater than that in unaccented syllables. Thus, despite this discrepancy in the results for VOT between the four speakers, one common finding was that the difference in VOT between voiceless and voiced stops was greatly increased in accented syllables compared to unaccented syllables, suggesting greater separation between voiceless and voiced stops in accented syllables than in unaccented syllables. Moreover, the

place distinction between labials and velars or alveolars was greater in terms of VOT and closure duration as cues to place of articulation in accented syllables compared to unaccented syllables, such that labials were more separated from velars or alveolars by a shorter VOT and a longer closure duration in accented syllables than in unaccented syllables; however, no enhancement of the distinction between velars and alveolars was found, although there were inconsistencies across speakers. The authors concluded that, although boundary-induced prosodic strengthening was not observed, prominence-induced prosodic strengthening showed the enhancement of phonological contrasts, including voicing and place of articulation contrasts. One issue that arises is that even though the study aimed to investigate more *natural* speech rather than lab-setting speech by analyzing audio-recorded radio news scripts, the corpus comprised recordings of read speech, which might have influenced speakers' use of prosodic structure. Thus, one of the open questions that remain to be answered is how prosodic structure operates in terms of prosodic strengthening and its linguistic function in more natural (i.e., non-read) speech. More discussion of this question follows in the "Considerations of Speech Styles: Read Speech vs. Interactive Speech" section.

Overall, prosodic strengthening driven by prosodic boundary is not always observed in English, whereas prosodic strengthening driven by prominence is consistently observed. This also appears to be true in relation to linguistic functions such that, in those contexts where boundary-induced prosodic strengthening is examined, the patterns of prosodic strengthening are not systematically found but in fact vary depending on the segment of interest and how those segments are compared, whereas prominence-induced prosodic strengthening enhances phonological contrasts in most cases. Thus, it is difficult to ascertain the nature of prosodic strengthening driven by prosodic boundary. Some important issues need to be addressed.

One of these issues is the language specificity of how prosodic strengthening is realized depending on the source of strengthening (i.e., boundary marking vs. prominence marking). Given that boundary marking and prominence marking have different functions (i.e., the former signals where phrases begin and end, and the latter is the locus of discourse information, e.g., new vs. given vs. focused), some researchers suggest that boundary marking and prominence marking are differently encoded in speech production (e.g., Beckman & Edwards, 1994; Edwards et al., 1991; Kim, Kim, & Cho, 2018). In particular, unlike other languages such as Korean, French, and Taiwanese, in which boundary-induced prosodic strengthening can enhance phonological contrasts, English has been suggested to be a language in which boundary-induced prosodic strengthening enhances CV contrast whereas prominence-induced prosodic strengthening enhances phonological contrast (e.g., Cho & Keating, 2009; Kim, Kim, & Cho, 2018) despite the inconsistent findings of previous literature in English. Even in cases where prosodic strengthening driven by prosodic boundary can be accounted for by both syntagmatic contrast and paradigmatic contrast enhancement, the patterns of prosodic strengthening in domain-initial position compared to domain-medial position were interpreted as having a syntagmatic contrast enhancement function (e.g., Cho & Keating, 2009; Pierrehumbert & Talkin, 1992). For example, recall that Cho and Keating (2009) interpreted the increased VOT for voiceless stop /t/ in Ut-initial position compared to Ut-medial position as an indication of syntagmatic contrast enhancement when, in fact, both syntagmatic contrast enhancement and paradigmatic contrast enhancement can predict an increased VOT for voiceless stops, such that VOT can be lengthened to make voiceless stops more consonant-like compared to the following vowels or to make them more voiceless stop-like compared to voiced stops. Cho and colleagues suggest that the enhancement patterns of prosodic strengthening differ depending on its source in

English because, as a stress-timed language, English integrates lexical stress into the higher-order prominence system, thus with boundary-induced prosodic strengthening deviating from prominence-induced prosodic strengthening (for details, see Cho 2016). If this is the case, other languages that employ lexically defined stress in their prominence marking system similarly to English may also show different enhancement patterns of boundary-induced and prominence-induced prosodic strengthening, with boundary-induced prosodic strengthening inducing syntagmatic contrast enhancement and prominence-induced prosodic strengthening inducing paradigmatic contrast enhancement.

However, languages such as German, in which phonological and phonetic categories of plosives are similar to English and in which the prominence system employs lexically defined stress, do not appear to show similar prosodic strengthening patterns as English in relation to linguistic functions (e.g., Kuzla & Ernestus, 2011). For example, Kuzla and Ernestus (2011) compared word-initial /p, t, k, b, d, g/ (i.e., *packen* 'to pack', *Tank* 'tank', *Karten* 'cards', *backen* 'to bake', *Dank* 'thanks', *Garten* 'gerden') in three different prosodic positions: (i) major boundary accompanied by a pause and a boundary tone, (ii) minor boundary accompanied by no pause but a boundary tone, and (iii) word boundary accompanied by no pause and boundary tone in an unaccented syllable. They found that the VOT for the fortis plosives /p, t, k/[8] decreased from a smaller prosodic boundary to a larger prosodic boundary (i.e., from a word boundary to a major boundary) whereas the VOT for the lenis plosives /b, d, g/ was not affected by the different sizes of prosodic boundary. The results did not support syntagmatic contrast enhancement in that fortis and lenis plosives did not increase VOT at a larger prosodic boundary

[8] Kuzla & Ernestus (2011) categorized German plosives into fortis /p, t, k/ and lenis /b, d, g/ because they acknowledged the fact that German /b, d, g/ are not truly voiced with glottal vibration word-initially. For this reason, they applied the same categorization to English by referring to *voiceless* stops /p, t, k/ as fortis and *voiced* stops /b, d, g/ as lenis.

compared to a smaller prosodic boundary, nor did they support paradigmatic contrast enhancement in that the difference in VOT between fortis and lenis plosives was smaller at a larger prosodic boundary than at a smaller one. Despite the similarities between the prosodic systems of German and English, as well as in the phonological/phonetic categories of plosives, the patterns of prosodic strengthening in relation to linguistic functions were inconsistent with those found in English. Although we do not expect the patterns to be exactly the same between languages that employ lexical stress, it still raises the question of whether the linguistic function of prosodic strengthening (i.e., syntagmatic vs. paradigmatic contrast enhancement) indeed differs depending on its source (i.e., prosodic boundary vs. prominence) also in English. Thus, when the inconsistent findings of boundary-induced prosodic strengthening in English are considered together, it may be premature to conclude that boundary-induced strengthening only has the linguistic function of enhancing CV contrasts in English based on a limited set of data. The present study tries to provide a more comprehensive investigation of boundary-induced prosodic strengthening and seeks to shed further light on another potential factor that can influence the enhancement patterns of prosodic strengthening in relation to linguistic function: speech style (i.e., read vs. interactive speech).

In addition, the evidence that prosodic strengthening can take place in domain-initial prosodic positions without enhancing differences between neighboring segments (consonants vs. vowels) or without enhancing a particular phonological contrast further complicates the understanding of prosodic strengthening in English (e.g., Dilley, Shattuck-Hufnagel, & Ostendorf, 1994, 1996; Garellek, 2012, 2014). Dilley and her colleagues examined the glottalization of vowels at the edges of prosodic domains (i.e., IP-initial boundary vs. iP-initial boundary vs. Wd-initial boundary) in English FM radio news speech that includes four news

stories read by professional announcers in a lab setting (see Ostendorf, Price & Shattuck-Hufnagel, 1995). They found that the rate of vowel glottalization increased as the prosodic constituent became larger.[9] Glottalization does not enhance the sonority of vowels, nor does it enhance a phonological contrast with other vowels. Given that glottalization in domain-initial positions can be a physiological consequence driven by low subglottal pressure preceded by a pause (see Slifka, 2006), it might be that the phonetic variations observed in domain-initial positions, termed as prosodic strengthening, are not linguistically driven.

Another issue is that it is often difficult to tease apart syntagmatic and paradigmatic contrast enhancement accounts. As indicated above, many previous studies observed increased VOT for voiceless plosives in domain-initial positions compared to domain-medial positions, and interpreted these results as suggesting that plosives enhance their consonantality against the neighboring vowel when in fact the increased of VOT for voiceless plosives can also be interpreted as an enhancement of their voicelessness compared to voiced plosives. Thus, prosodic strengthening for both voiceless and voiced plosives should be tested to tease those accounts apart. There are three possible enhancement patterns of prosodic strengthening. Figure 2 shows the schematic of three possible patterns of VOT realization for voiceless and voiced plosives in the case of syntagmatic contrast enhancement, paradigmatic contrast enhancement, and the concurrence of syntagmatic and paradigmatic contrast enhancement when there is prosodic strengthening.

[9] Dilley et al. (1996) examined the effect of prosodic prominence (accented vs. unaccented) along with that of prosodic boundaries. The rate of glottalization was higher in domain-initial positions than in domain-medial positions when words were not accented, and it was highest in domain-initial positions than in domain-medial positions when words were accented. It should be also noted that the glottalization in domain-initial positions for accented words was not caused by pitch accents with a low tone (i.e., L* and L*+H), which often correlates with glottalization (e.g., Pierrehumbert & Frisch, 1994).

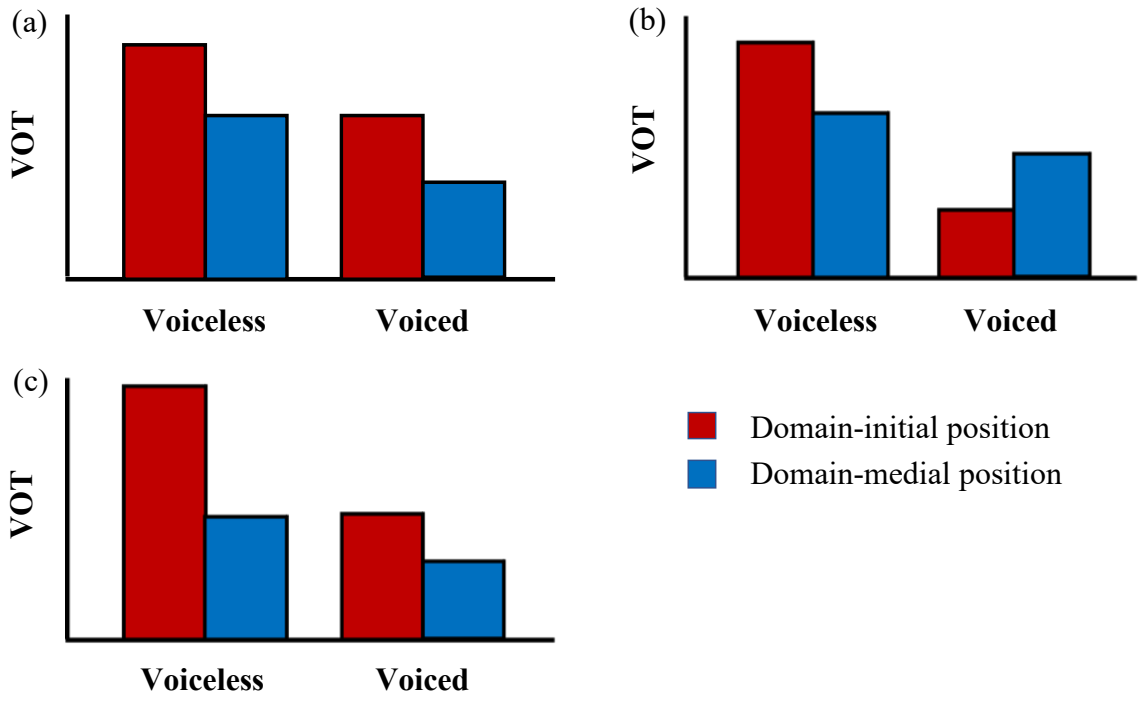


Figure 2: A schematic of possible patterns of VOT realization for voiceless and voiced plosives in relation to the linguistic function of prosodic strengthening: (a) syntagmatic contrast enhancement, (b) paradigmatic contrast enhancement, and (c) syntagmatic + paradigmatic contrast enhancement.

Based on the VOT results of previous studies, recall that syntagmatic contrast enhancement results in increased VOT for both voiceless and voiced plosives in domain-initial positions compared to domain-initial positions (Figure 2 (a)). On the other hand, as shown in Figure 2 (b), paradigmatic contrast enhancement results in increased VOT for voiceless plosives but decreased VOT for voiced plosives in domain-initial positions compared to domain-medial positions so that the difference in VOT between voiceless and voiced plosives is enhanced in domain-initial positions compared to domain-medial positions. In theory, it is also possible to find evidence for both syntagmatic contrast enhancement and paradigmatic contrast enhancement. For example, as depicted in Figure 2 (c), both voiceless and voiced stops can be produced with a longer VOT that enhances their consonantality but the contrast between

voiceless and voiced stops can also be enhanced in domain-initial positions compared to domain-medial positions if the increase in VOT for voiceless stops is greater than that for voiced stops, resulting in a greater difference in VOT between voiceless and voiced stops in domain-initial positions than domain-medial positions. Similarly, nasals can tease apart syntagmatic and paradigmatic contrast enhancement. Syntagmatic contrast enhancement predicts that nasals will show less nasality and be more consonant-like in domain-initial positions compared to domain-medial positions, whereas paradigmatic contrast enhancement predicts that they will show more nasality and be more distinct from other manners of articulation such as stops and fricatives in domain-initial positions compared to domain-medial positions. All these possibilities were explored and discussed in the present study by examining the production of English voiceless and voiced plosives and nasals in different prosodic positions.

Last but not least, the understanding of paradigmatic contrast enhancement driven by boundary-induced prosodic strengthening has been limited to voicing contrast. Cole et al. (2007) attempted to examine the enhancement of the contrast between places of articulation, in addition to the enhancement of the voicing contrast, in IP-initial position compared to IP-medial position by looking at English plosives /p, t, k, b, d, g/ that differ in voicing and place of articulation. Due to the rare occurrence of an IP boundary in the corpus (Boston University Radio News), however, only /t, d/ in IP-initial position could be compared with those in IP-medial position, so ultimately the enhancement of place of articulation could not be tested in the study. The present study broadens the understanding of paradigmatic contrast enhancement of the contrast between places of articulation by exhaustively examining English plosives /p, t, k, b, d, g/ and nasals /n, m/, which can occur word-initially, in IP-initial position compared to IP-medial position. Acoustic parameters that are known to distinguish places of articulation such as spectral peak of

the burst (e.g., Dorman, Studdert-Kennedy, & Raphael, 1977; Edwards, 1981; Fant, 1973; Fischer-Jørgensen, 1954; Halle, Hughes, & Radley, 1957; Keating, Byrd, Flemming, & Todaka, 1994; Keating & Lahiri, 1993; Repp & Lin, 1989; Winitz, Scheib, & Reeds, 1972; Zue, 1976) and F2 (e.g., Fowler, 1995; Jongman, Wayland, & Wong, 2000; Sussman, Bessell, Dalston, & Majors, 1997; Sussman, McCaffrey, & Matthews, 1991; Sussman & Shore, 1996; Zhao, 2010) are measured to directly test the enhancement of the contrast between places of articulation.

All in all, the present study seeks to fill these gaps by providing a more comprehensive understanding of prosodic strengthening under the influence of prosodic boundaries. The focus is on the realization of English voiceless and voiced plosives and nasals in IP-initial vs. IP-medial (Wd-initial) positions. In addition, we investigated the above issues in read speech and interactive speech. Through interactive speech, the present study tries to provide insight into how speakers make use of prosodic strengthening in a more natural setting that involves interaction with an interlocutor. If we assume that the goal of speech is to be understood, it is important to comprehend the use of prosodic strengthening in everyday interactive communication. The following section provides more details on the matter.

Considerations of Speech Styles: Read Speech vs. Interactive Speech

One of the issues that researchers have overlooked in the investigation of prosodic strengthening is speech style. Speech style in the present study specifically includes read speech that is not directed to listeners vs. interactive speech that occurs in the course of interactions between speakers. Although our everyday use of spoken language mostly occurs in the form of conversational speech, most previous studies that have investigated the phonetic realization of prosodic structure have focused on read speech in a lab setting (e.g., Bombien et al., 2013; Byrd

et al., 2000; Byrd et al., 2006; Byrd & Saltzman, 2003; Cho & Keating, 2001; Cho & Keating, 2009; Fougeron, 2001; Fougeron & Keating, 1997; Keating et al., 2003; Keating et al., 1999; Kim, Kim, & Cho, 2018; Pierrehumbert & Talkin, 1992). A few studies have investigated FM Radio news speech corpus (Ostendorf et al., 1995) that supposedly contains more natural speech (e.g., Cole et al., 2007; Dilley et al., 1994, 1996), but this corpus consists of four news stories that were read by professional radio news announcers. Thus, these studies still investigated the influence of prosodic structure on the phonetic realization of segments in read speech, but outside of a lab setting. The investigation of read speech in a lab setting has its advantages in that it is easier to control experimental conditions and unexpected discourse situations; yet, it is not clear whether the conclusions that are drawn from the prosodic data collected under these controlled conditions can be generalized to more natural speaking styles such as interactive speech.

While the effect of speech style on prosodic strengthening remains unclear, it is well established that speech style (read vs. interactive (spontaneous) speech) has an important influence on prosodic phrasing and prominence in relation to intonational contour, pauses, the presence or absence of speech disfluencies, the reduction or full realization of segments, and pitch range, among others (e.g., Ayers, 1992; Barry, 1995; Blaauw, 1992; Bruce, 1995; Hirschberg, 2000; Kohler, 1995; Silverman, Blaauw, Spitz, & Pitrelli, 1992). Researchers have found that different speech styles elicit variations in how prosodic phrases are grouped and where and how prominence falls in a sentence. For example, Hirschberg (2000) showed that the intonational contour at the right edge of the Intonational Phrase often shows a rise (due to a high boundary tone) in interactive speech but a fall (due to a low boundary tone) in read speech. She also showed that the distribution of pitch accents can be also different such that read speech

carries substantially more H* and fewer L+H* or L* compared to interactive speech. Because pauses often mark grammatical junctures such as larger prosodic boundaries and are indicative of speech planning processes in speech production (e.g., Krivokapic, 2007; see also Krivokapic, 2012), the occurrence and duration of pauses also differs between read and interactive speech (e.g., Deese, 1980; Goldman-Eisler, 1972). When there is a given script as in read speech, pauses occur mainly at syntactic junctures, but they can also occur elsewhere in interactive speech (Goldman-Eisler, 1968), and thereby the occurrence of pauses is more predictable in read speech than in interactive speech. Read speech requires less planning since there is a given script in which speech plan is already provided, and, as a result, yields shorter pause duration than interactive speech that requires more planning. Thus, speech style influences the implementation of prosodic structure. However, the previous studies that compared interactive speech to read speech sought to investigate how speaking styles influenced general patterns of prosodic structure rather than investigating the effect of speaking styles on prosodic strengthening. It is unclear how prosodic structure is phonetically manifested especially in interactive speech compared to read speech.

Therefore, another goal of the present study is to investigate prosodic strengthening in more natural interactive speech, unlike most of the previous studies on prosodic strengthening, which elicited read speech in a lab setting to test whether and how speakers make use of prosodic strengthening. Read speech differs from interactive speech in terms of (i) how much speech planning is involved and (ii) whether or not the speaker's speech is directed to a listener. These differences in read and interactive speech might yield different patterns of prosodic strengthening. In read speech, prosodic structure is already provided through linguistic notations such as punctuation. For example, in previous studies, an IP boundary that is smaller than an

utterance boundary and larger than an iP boundary and a word boundary may be marked by a comma (e.g., Bombien et al., 2013; Byrd, 2000; Byrd et al., 2006; Cho, 2006; Fougeron, 2001; Georgeton & Fougeron, 2014; Keating et al., 2003; Kim et al., 2018; Pierrehumbert & Talkin, 1992). Thus, there is less speech planning involved, which is evident with the fewer occurrences and shorter duration of pauses. Recall that a pause is one of the important markers of an IP boundary. In order to make up for the lack of pause in read speech, speakers might focus more on signaling the left edge of an IP boundary by enhancing the contrast between neighboring consonants and vowels. In fact, pause duration has been found to be inversely proportional to final lengthening at the right edge of an IP boundary such that more final lengthening is followed by a shorter pause (e.g., see also Byrd & Saltzman, 2003; Ferreira, 1993). This may also be true for the left edge of an IP boundary. In addition, without a listener, the speaker does not need to enhance the production of lexical entries in the process of phonetic encoding. Thus, as a means of purely marking a prosodic juncture, the speaker might enhance the contrast between neighboring consonants and vowels rather than enhancing phonological contrasts. On the other hand, interactive speech involves more speech planning, as is evident from the greater occurrences of pauses and longer pause duration. Thus, pauses already provide abundant cues to IP boundaries, so speakers do not need to focus on signaling the left edge of the boundary. Instead, since interactive speech is directed to a listener, speakers may focus more on helping the listener better access lexical entries by enhancing phonological contrasts such that an initial segment becomes more distinctive from other segments that can potentially confuse the listener. Therefore, given the different processes that underlie the production of read speech and interactive speech, investigating the effect of speech style on prosodic strengthening could shed light on the nature of the linguistic functions of boundary-induced prosodic strengthening.

In addition, the investigation of prosodic strengthening in interactive speech helps ascertain whether prosodic strengthening is a true characteristic of (more) naturally occurring speech rather than an artifact of highly controlled lab speech. Most of the previous studies that found an effect boundary-induced prosodic strengthening elicited sentences by asking participants to read sentences in a highly controlled setting where only a few types of sentences were repeated throughout an experiment and where participants were asked to repeat the sentences when their production did not meet the rendition of the intended prosodic structure (e.g., Byrd et al., 2006; Cho & Keating, 2009; Cho et al., 2007; Fougeron & Keating, 1997; Kim, Kim, & Cho, 2018; Pierrehumbert & Talkin, 1992). This controlled lab setting could result in an artificially amplified effect of prosodic boundary because prosodic structure may prime itself, compared to more naturally elicited interactive speech. Therefore, it is important to investigate prosodic strengthening in both read and interactive speech and compare them for a better understanding of prosodic strengthening.

Research Questions and Hypotheses

The research questions investigated in this dissertation are as follows: (i) How does prosodic boundary (IP-initial vs. IP-medial positions) influence the phonetic realization of English plosives /p, t, k, b, d, g/ and nasals /n, m/ in relation to their linguistic functions (syntagmatic vs. paradigmatic contrast enhancement)? And (ii) How does speech style (read vs. interactive speech) influence the patterns of boundary-induced prosodic strengthening in relation to their linguistic functions?

We hypothesize that prosodic strengthening is observed and reflected in the acoustics of English plosives and nasals in IP-initial position compared to IP-medial position in both read and

interactive speech. However, read speech and interactive speech are predicted to show different patterns of prosodic strengthening in relation to linguistic functions. In read speech, speakers might focus more on marking the left edge of an IP boundary (see the "Considerations of Speech Styles: Read Speech vs. Interactive Speech" section). Thus, prosodic strengthening in read speech is predicted to enhance syntagmatic contrast by making English plosives and nasals more consonant-like (less sonorous). In interactive speech, speakers might focus more on helping listeners' perception of an initial sound correctly (see the "Considerations of Speech Styles: Read Speech vs. Interactive Speech" section). Thus, prosodic strengthening in interactive speech is predicted to enhance paradigmatic contrast by enhancing voicing contrast and/or the contrast between places of articulation for plosives and by enhancing nasality (compared to oral sounds) and the contrast between places of articulation.

In order to answer the first research question (i), the present study investigated prosodic strengthening on English plosives and nasals that enabled us to examine various acoustic correlates that can tease apart the syntagmatic and paradigmatic contrast enhancement accounts and those that can test the enhancement of place of articulation contrast as well as voicing contrast for the evaluation of paradigmatic contrast enhancement. For plosives, VOT and RMS amplitude were examined to tease apart syntagmatic vs. paradigmatic contrast enhancement: syntagmatic contrast enhancement would yield longer positive VOT and higher RMS amplitude for both voiceless and voiced plosives in IP-initial position than IP-medial position; paradigmatic contrast enhancement would yield longer positive VOT and higher RMS amplitude in IP-initial position than in IP-medial position for voiceless plosives whereas it would yield shorter positive VOT (or more prevoicing) and lower RMS amplitude in IP-initial position than in IP-medial position for voiced plosives. Spectral peak of the burst and F2 were examined specifically to

investigate the enhancement of the contrast between place of articulation as one of the means of testing paradigmatic contrast enhancement. For nasals, the amount of nasality was inferred from nasal duration, F1 bandwidth, and the maximum and mean value of A1 that can tease apart syntagmatic and paradigmatic contrast enhancement. Syntagmatic contrast enhancement would yield less nasality, which can be indicative of more consonantality: shorter nasal duration, narrower F1 bandwidth, and higher maximum and mean value of A1 (see the "Prosodic Strengthening Driven by Prosodic Boundary and Its Function: Syntagmatic vs. Paradigmatic Contrast Enhancement" section for more details). In contrast, paradigmatic contrast enhancement would yield more nasality indirectly compared to non-nasal (oral) consonants: longer nasal duration, wider F1 bandwidth, and lower maximum and mean value of A1. As for plosives, F2 was examined to investigate the contrast enhancement between place of articulation in order to test paradigmatic contrast enhancement.

The second research question (ii) was answered by creating two tasks each of which was designed to elicit target consonants in the intended prosodic boundaries (i.e., IP-initial & IP-medial positions) in different speech styles: read vs. interactive speech so that prosodic strengthening in read and interactive speech could be directly compared and evaluated. The interactive speech task requires a pair of speakers who need to interact with each other to finish the task whereas the read speech task requires a speaker to read written sentences without a listener. As predicted above, these tasks that vary in whether there is a listener or not would inform us whether speech style can modulate boundary-induced prosodic strengthening and its linguistic function.

Chapter 2: Methods

Participants

Thirty-seven native American English speakers aged from 18 to 39 (23 females; mean: 23.1; SD: 5.5) were recruited online. They first participated in the interactive speech experiment and next in the read speech experiment. The participants were compensated with $20 after they finished the read speech-experiment. The recording quality was checked for each participant by visually inspecting the Signal to Noise Ratio (SNR), and the participants who produced recordings of poor quality were excluded from the analysis. When the recording quality from a participant was good in one experiment but not in the other, the participant was not included in the analysis. Ultimately, eighteen native American English speakers were included in the analyses.

Materials

The consonants of interest are stops /p, t, k, b, d, g/ and nasals /m, n/ in American English. Disyllabic English adjectives that begin with a voiceless stop (e.g., *picky, tipsy, kissy*), a voiced stop (e.g., *busy, ditsy, giddy*), or a nasal (e.g., *mini, nippy*) were used as target words. Two sets of target words were selected that had their initial consonant followed by either [ɪ] or [ʌ]. All target words were lexically stressed in the initial syllables. Each target word was followed by a mono- or disyllabic noun whose initial segment did not share manner of articulation with the target word onset (e.g., *kissy* in *Kissy Snail*; *giddy* in *Giddy Lamb*). The target word and following noun together formed a proper noun so that the target word consistently received pitch accent (e.g., Liberman & Prince, 1977). The experiment targeted the elicitation of an H* pitch

accent on the target word and only those target words produced with an H* pitch accent were included in the analysis (see the "Acoustic Analyses" section).

As shown in Table 1, the target word was positioned in IP-initial or IP-medial position in sentences. IP-initial boundaries were elicited by placing a prepositional phrase (separated from the main clause by a comma) at the beginning of the sentence. Verbs that often accompany a prepositional phrase (e.g., *put, kiss, hit, ask* etc.) were used in the sentences so that the target word would not be located at the end of a sentence, the latter being more prone to influences from F0 declination, boundary tones, and phrase-final lengthening. In addition to prosodic boundary, linear position (i.e., early vs. late) was manipulated in order to take simple locational differences in a sentence (i.e., located earlier vs. located later) into account as a potential confounding factor when evaluating the effect of prosodic boundary (i.e., IP-initial vs. IP-medial positions).

Table 1: Design and example stimuli. The target word is underlined. The IP boundary is marked with #. In this example, the segment of interest is the voiceless bilabial stop [p].

Prosodic conditions	Linear position	Vowel	Example sentences
IP-initial	Early	[ɪ]	In the morning, # Picky Mole is calling Forky on the phone.
		[ʌ]	In the morning, # Puffy Horse is putting Goosy in the box.
	Late	[ɪ]	At 7:20 in the morning, # Picky Mole is calling Forky on the phone.
		[ʌ]	At 7:20 in the morning, # Puffy Horse is putting Goosy in the box.
IP-medial	Early	[ɪ]	In the morning, # Forky is calling Picky Mole on the phone.
		[ʌ]	In the morning, # Goosy is putting Puffy Horse in the box.
	Late	[ɪ]	At 7:20 in the morning, # Forky is calling Picky Mole on the phone.
		[ʌ]	At 7:20 in the morning, # Goosy is putting Puffy Horse in the box.

The manipulation of linear position was achieved by removing or including an utterance-initial phrase containing a specific time when an event was happening (e.g., *at 7:20*) in the sentences. For the early position, the phrase containing a specific time was removed so that the target word would be located earlier in the sentence; for the late position, it was included so that the target word would be located later in the sentence. As a result, the target word in the early position of the IP-initial condition was located the earliest in the sentence, and the target word in the late position of the IP-medial condition was located the latest in the sentence. The target word in the early position of the IP-medial condition and the target word in the late position of the IP-initial condition had the same absolute position in the sentence (in terms of number of syllables that preceded them). Any difference between them can therefore not be attributed to their linear position in the sentence. Finally, vowel context (i.e., [ɪ] or [ʌ]) was manipulated such that two participants who interacted with each other in the interactive speech experiment were given different sets of target words where the target segment was followed by either [ɪ] or [ʌ] in order to reduce the likelihood of phonetic convergence between the two participants during their interaction (e.g., Pardo, 2006; Delvaux & Soquet, 2007).

Each participant produced 32 sentences per condition (128 sentences throughout the entire experiment), but 256 different sentences were produced by the two participants throughout the entire experiment (128 for each [ɪ] and [ʌ]), yielding the following design: 8 consonants ([p, t, k, b, d, g, m, n]) x 2 prosodic positions (IP-initial vs. IP-medial positions) x 2 linear positions (early vs. late positions) x 2 speech styles (interactive vs. read speech) x 2 repetitions x 2 vowel contexts ([ɪ] or [ʌ]). The experiment also included 32 filler sentences in which the target word was located in utterance-initial position without the time information (e.g., without *at 7:20 in the morning*). These fillers were included to further vary the location of the target words throughout

the experiment. Six lists were created, with each list having different orders of trials. Each list was assigned to a pair of participants. In a list, there were two blocks, each of which contained all experimental trials interspersed with the filler trials. These blocks ultimately worked as the two repetitions included in the design. Trials in each block were pseudo-randomized so that participants would not have sentences that have the same target word in different prosodic positions (i.e., IP-initial vs. IP-medial positions) in a row, which might elicit contrastive focus on the target word.

Sixteen pictures were created for the familiarization phase, the interactive speech task, and the read speech task. As shown in Figure 3, the pictures for the familiarization phase included two animal characters with their names (e.g., *Goosy, Puffy Horse*), a moving image in which the characters interact (presented as GIFs), and a written description of their interaction as a partial phrase.

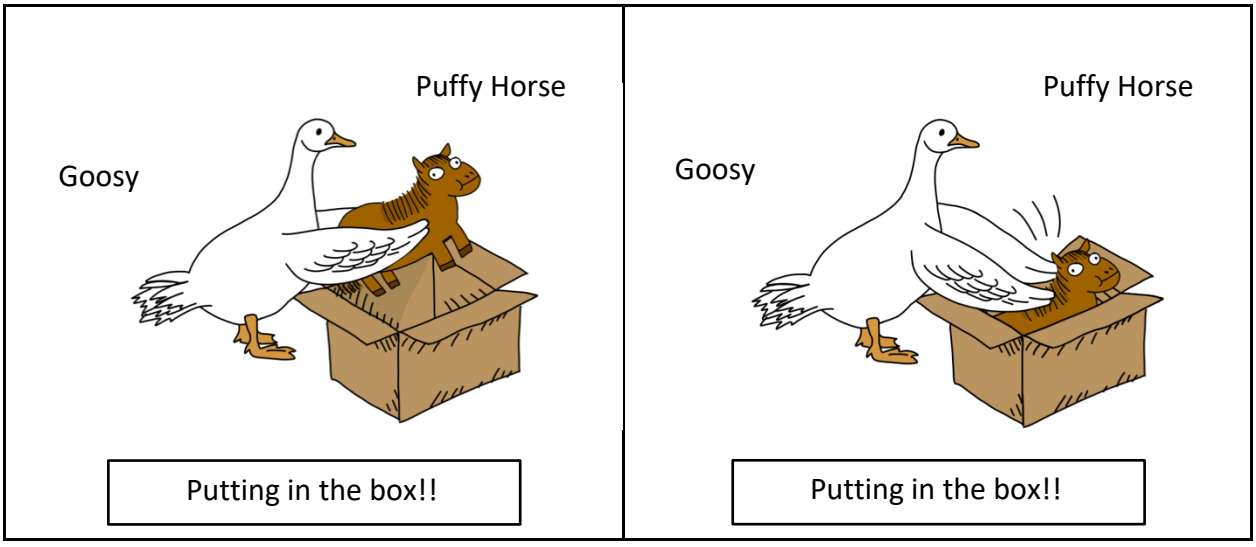
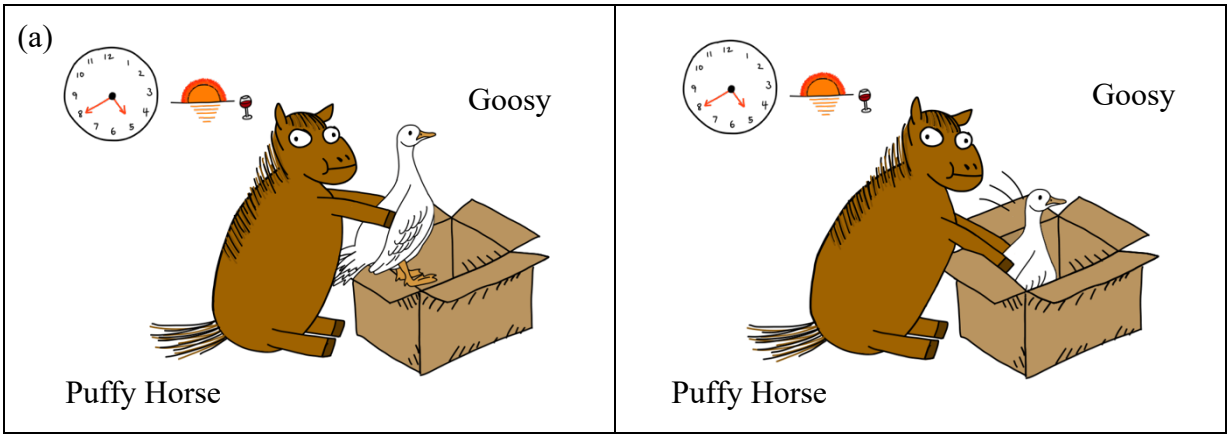


Figure 3: An example of pictures in the familiarization phase. These pictures were presented as a connected action (as a GIF).

For the interactive speech task, as shown in Figure 4 (a) and Figure 5 (a), the pictures described the scenes in which one animal character (i.e., *Puffy Horse*) is doing something to the other character (e.g., *Goosy*). The names of the animal characters were written in the scenes so that participants did not have to memorize/remember the names of the characters. The elements of information that made up the scenes, such as the time of events, the agent of an action, and the theme of the action, were conveyed from left to right in consideration of where those elements were located in the intended sentences. The time information was represented as a clock and the sun was located in the upper left corner. The agent of the action, an animal character, was located in the middle of the scene, and the theme of the action, the other animal character, was located on the right side of the target. These scenes were paired with a picture that included a question (e.g., *At 5:40 in the evening, what is happening?*) and two options (i.e., A and B) together as in Figure 4 (b) and Figure 5 (b).



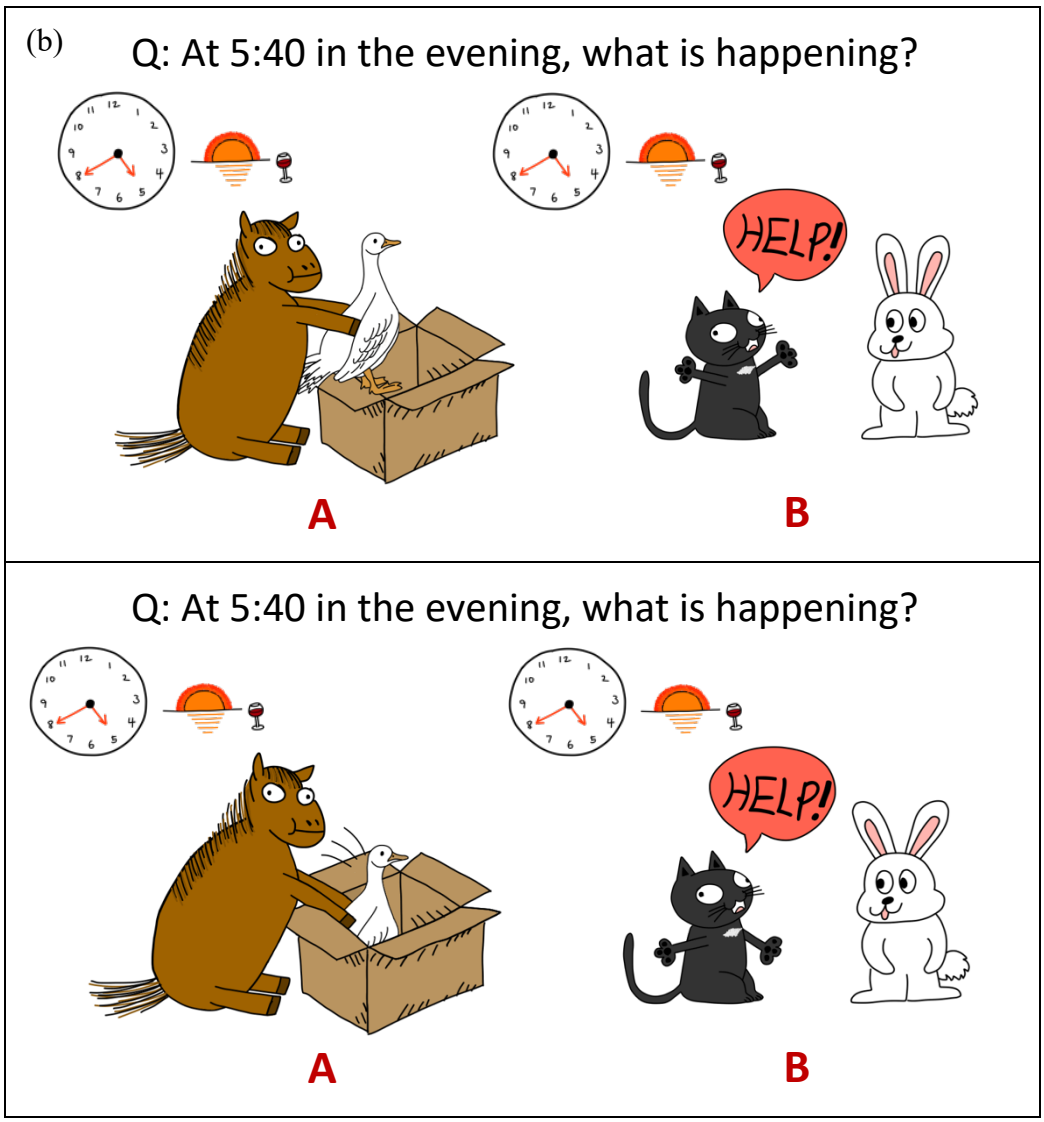


Figure 4: An example of scenes in an experimental trial for the interactive speech task. The target word is Puffy and located in IP-initial position. A participant sees (a), and the other participant sees (b) with the question. Two pictures for each (a) or (b) were presented as a connected action (as a GIF).

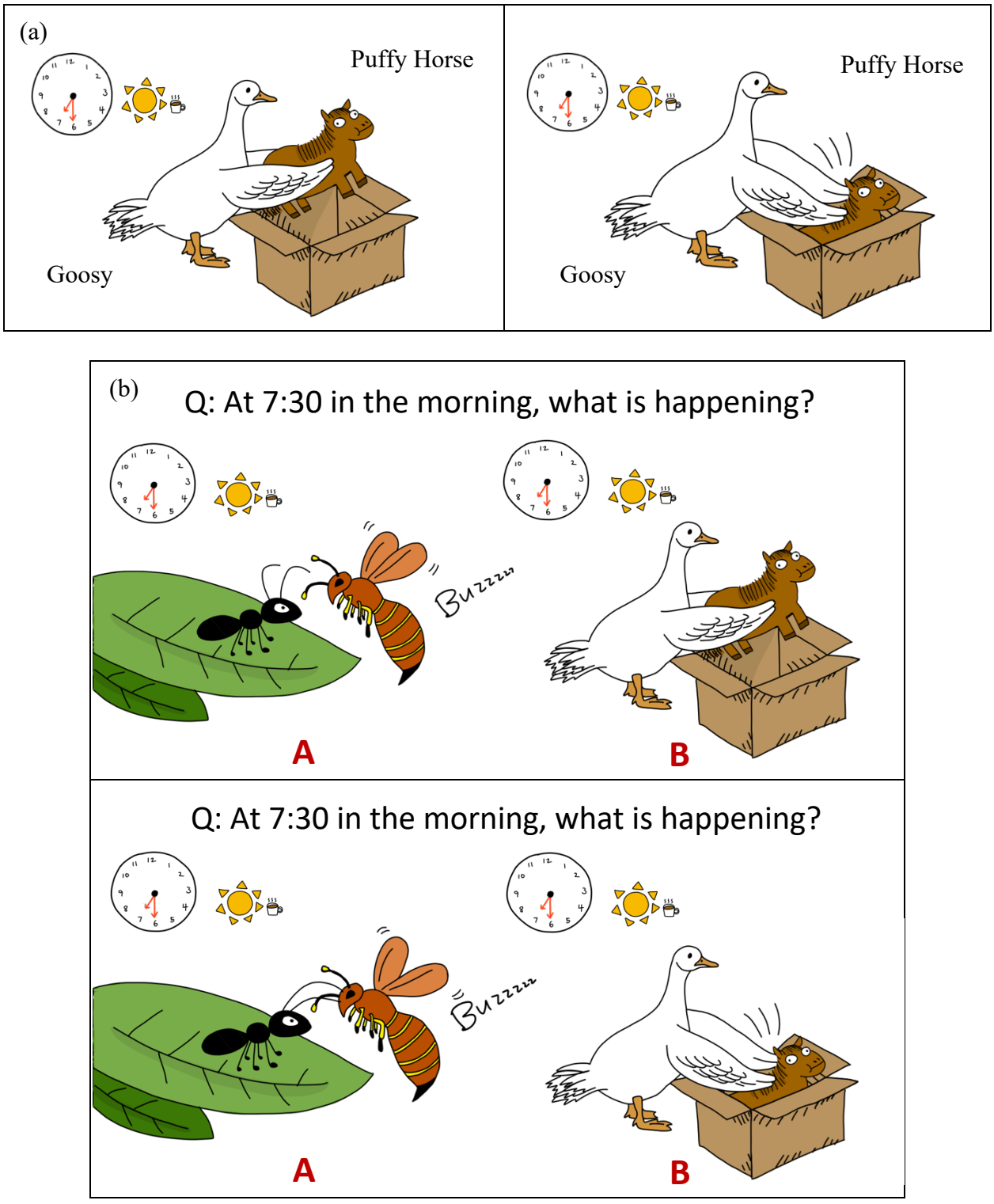


Figure 5: An example of scenes in an experimental trial for the interactive speech task. The target word is Puffy and located in IP-medial position. A participant sees (a), and the other participant sees (b) with the question. Two pictures for each (a) or (b) were presented as a connected action (as a GIF).

Finally, the pictures for the read speech task were created to have the same scenes as for the interactive speech task in order to make the two tasks similar, but, as an important manipulation, a full description of the scenes was presented orthographically, as shown in Figure 6.

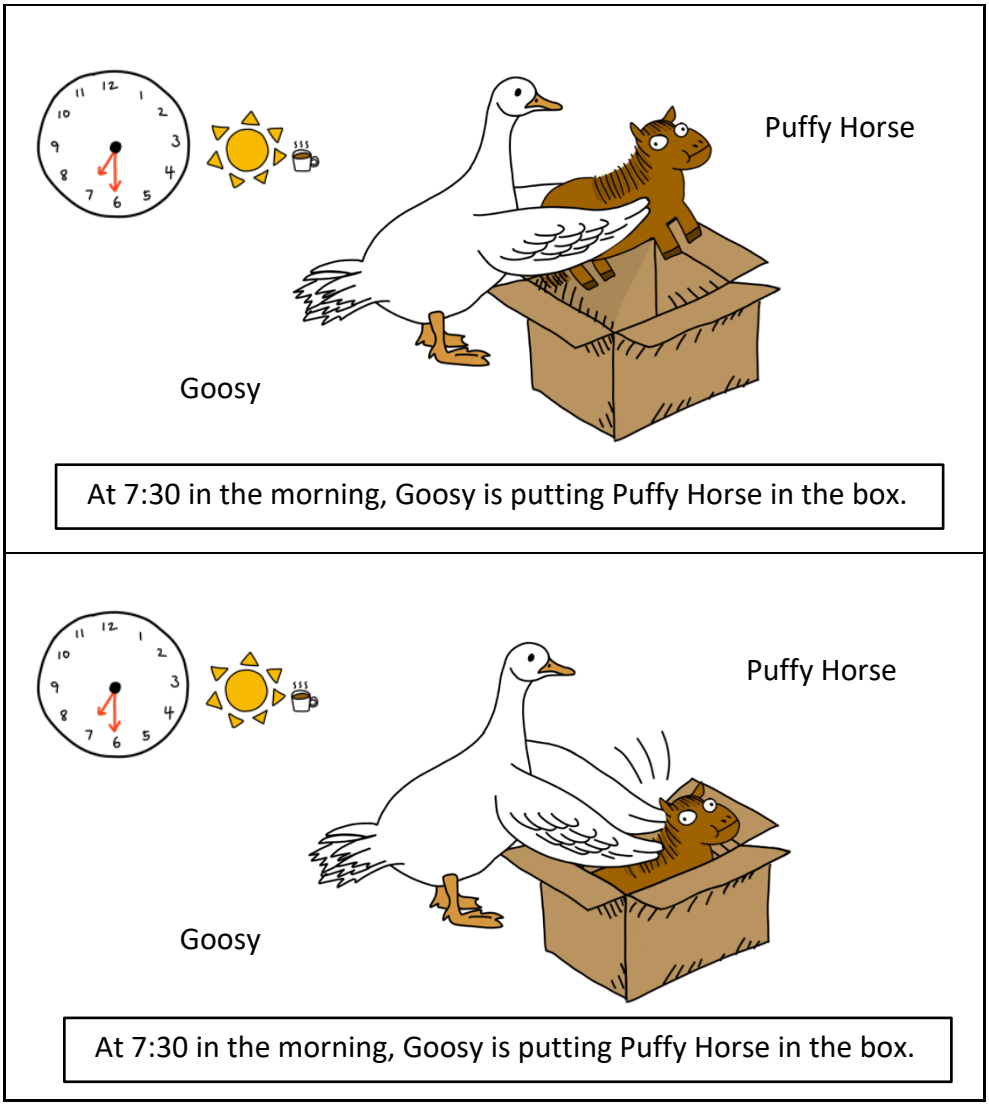


Figure 6: An example of scenes in an experimental trial for the read speech task. The target word is Puffy and located in IP-medial position. These pictures were presented as a connected action (as a GIF).

Procedures

Participants completed (i) the language background questionnaire, (ii) the familiarization phase, (iii) the interactive speech task, and (iv) the read speech task (in this order) on two

separate days online via Zoom (Zoom Video Communications Inc). They did so in two testing sessions over two different days. During the first testing session, participants signed an electronic consent form and completed a language background questionnaire. Subsequently, they completed the familiarization phase and the interactive speech task with another participant. During the second testing session, participants completed the read speech task without another participant. There were at least two days of gap between the interactive speech task and the read speech task.

In the familiarization phase, participants were asked to familiarize themselves with what animals were doing in Animal Village, what their names were, and how their actions were described in the presented scenes on their monitor (Figure 3). The familiarization phase was designed to avoid too many sentential variants and obtain consistency in how the words were accented and where the targeted prosodic boundaries were realized. As shown in Figure 3, full sentences were never presented in the familiarization phase. The familiarization phase took approximately 3-5 minutes for each participant to go through all the scenes.

Next, for the interactive speech task, participants were randomly paired and engaged in a story telling game together. The purpose of the game was to figure out what animals in Animal Village are doing at certain times of the day. Participant A first asked a question such as “at 7:30 in the morning, what is happening?” and Participant B provided a detailed description of the scene seen on their monitor to Participant A. Participants were asked to describe the details of the scenes, including time when it was included in the scene. All participants were able to construct the targeted sentence by describing the scene from left to right without any instruction. After Participant A listened to the description that Participant B provided, they selected the scene that corresponded to that described by Participant B among two scenes given as options on their

monitor, as shown in Figure 4 (b) and 5 (b). Participants alternated their role throughout the experiment such that, in the next trial, Participant B asked a given question to Participant A, Participant A described the scene they were looking at to Participant B, and Participant A chose the scene that Participant B had just described among two options of scenes. There were two practice trials that contained scenes that were not presented in the experimental or filler trials. They were designed to show the participants how they would alternate their roles and what they should expect to see in the trials. In the practice trials, the participants were told once more that the *names* of characters would be presented on each scene. The experimenter was monitoring the experiment in the same zoom meeting but did not intervene in the interaction between participants unless there were technical problems (e.g., disconnection from the internet, inaccessibility of the materials, misalignment of a trial between participants).

At last, in the read speech task, participants were asked to *read* sentences that were orthographically provided on the computer screen (Figure 6). The read speech task included the scenes that they described to the other participant in the interactive speech task. Participants were instructed to go through the trials in the read speech task alone. As in the interactive speech task, the experimenter was monitoring the experiment in the same zoom meeting but did not intervene once participants started to read the given sentences unless there were technical problems.

At each meeting, participants made recordings (WAV audio format) on their mobile phone using Awesome Voice Recorder (Ver. 8.0.7) while they went through experimental trials and sent them to the experimenter at the end of the meeting (method adapted from Zhang, Jepson, Lohfink, & Arvaniti, 2020). In the recorder, the sampling rate was set at 22,050 Hz. Before the experiment, the participants were instructed to be in a quiet environment without any

distractions or interruptions and place their phone as high as their lips but approximately 11 inches (30cm) away from them.

Acoustic Analyses

Following the exclusion of data with poor audio quality (as specified in the "Participants" section), the data of 18 participants were included in the analyses. Before extracting acoustic measurements, the elicited sentences were checked in terms of prosodic boundaries. The prosodic transcription of prosodic boundaries followed the English ToBI labelling conventions (version 3.0) by Beckman and Elam (Beckman & Elam, 1997). Using Praat (Boersma & Weenink, 2018), prosodic boundaries (IP-initial vs. IP-medial boundaries) were marked. The prosodic boundaries were initially checked by a trained English ToBI transcriber. Twenty percent of the sentences were randomly selected and cross-checked by a second English ToBI transcriber who did not have any information about the experimental design or where the target boundary was supposed to be. IP boundary was defined as a juncture that entailed lengthening and significant tone lowering with an L% boundary tone or with a rising H% boundary tone on the word at the right edge of the phrase, with this IP boundary often being followed by a pause. Boundary tones (L% or H%) are preceded by a gradual declination or ascent of a phrasal tone (L- or H-) and reach the lowest or highest pitch in the individual speakers' pitch range over a phrase. When the prosody preceding the target word did not meet these criteria and thus did not provide clear evidence of an IP boundary, the target word was considered as being in IP-medial position (i.e., Prosodic Word-initial position in the present study). Of the sentences that were cross-checked by a second transcriber, half were judged by the first transcriber as having the prosodic boundary intended by the experimental design and half were judged as not having a

prosodic boundary that meets one or more of the above criteria. The first and second transcribers showed a 98% agreement rate on the transcription.

Whether the target words received the targeted pitch accent was also checked. Recall that the target word was the first word in the character name so that the target word would receive the pitch accent. The presence of a H* pitch accent was determined impressionistically and by examining the local pitch excursion on the target word. Only the target words that received a H* pitch accent were included in the analyses, for three reasons. First, our materials were designed to avoid the pitch accent L+H* signaling 'corrective focus' (also known as contrastive focus) that can potentially override the effect of prosodic boundary in English (e.g., Cho & Keating, 2009; Cho, Lee, & Kim, 2014). Second, the types of pitch accent that participants produce are expected to influence the acoustic parameters that are relevant for the present study. For example, whether words receive H* or L* can influence amplitude and f0 (e.g., Shue et al., 2010). The present study only focused on the pitch accent H* in order to minimize unnecessary influences from different types of pitch accent. The trials in which the target words did not receive an H* pitch accent were thus excluded from the analysis: 67 trials (3%) out of 2304 trials in total (21 for IP-initial position in interactive speech; 12 for IP-medial position in interactive speech; 7 for IP-initial position in read speech; 27 for IP-medial position in read speech). The trials in which the participants showed some disfluency and/or hesitation and that contained too much background noise were also excluded from the analysis. The analysis included 1687 trials (73%, 776 for the interactive speech task and 911 for the read speech task) out of 2304 trials in total.

Using Praat (Boersma & Weenink, 2018), the following acoustic parameters were measured for voiceless and voiced stops and nasals to test whether and, if so, how speakers

enhance segments in terms of CV contrast and/or phonological contrast in prosodically salient positions such as IP-initial position compared to prosodically less salient positions such as IP-medial position. The acoustic parameters for stops included VOT, RMS burst amplitude, spectral peak of burst, and F2, and those for nasals included nasal duration, f1 bandwidth, mean A1, max A1 and F2. The details are discussed in the following sections (the "Measurements for stops" and "Measurements for nasals" sections).

Measurements for Stops

VOT. Two types of VOT were measured: raw VOT and the ratio of VOT over word duration. The ratio values were considered to account for speech rate differences among the different recordings (e.g., Edward, 1981). VOT was defined as the amount of time (in milliseconds) from the burst release of the stop to the onset of voicing by inspecting waveforms and spectrograms. The onset of voicing was determined by the start of periodicity in the waveform. Prevoicing was located when there was voicing that preceded the burst of the stop release in IP-initial position. It was determined in a similar manner in IP-medial position, but more caution was required due to the absence of pauses in running speech. In some cases, because the preceding word ended with /ŋ/ as in *calling* or *pulling*, its voicing residue could continue, and it was difficult to tell it apart from the prevoicing of the stop in the closure duration. The voicing observed between the preceding word and the target word in IP-medial position was considered as prevoicing only when the voicing filled at least half of the closure duration. This strategy was adopted from Abramson and Whalen (2017), although their discussion of this issue did not consider different prosodic positions and was limited to stops in an intervocalic position within a single word.

RMS burst amplitude. Two types of RMS amplitude of the burst were measured for stops: the raw RMS burst amplitude and the RMS burst amplitude of stops relative to the RMS amplitude of the whole word (RMS burst amplitude for a stop – RMS amplitude of the word). The burst part of stops needed to be identified first to measure the RMS burst amplitude, but the precise onset and offset of the burst was difficult to identify since the burst was often not separable from the aspiration in the inspection of waveforms and spectrograms. For this reason, instead of identifying the exact burst in stops, the present study used a half-Hamming window (the latter half of the Hamming window) so that the initial part of noise, which is the most likely to include the burst, could be weighted more. Within this window, the RMS burst amplitude was then calculated by taking the square root of the mean of the sum of squares of the amplitude values for each sample point divided by the window size. The RMS amplitude of the burst was normalized relative to the RMS amplitude of the word in order to account for participants' overall loudness in each recording. The RMS amplitude for the word was calculated in the same way as stops but over the whole word. The relative RMS burst amplitude values were calculated by subtracting the RMS amplitude of the word from the RMS burst amplitude of the stop.

Spectral peak. A Fast Fourier Transform (FFT) was calculated over the half Hamming window (the latter half of the Hamming window), again assuming that the burst occurs somewhere at the beginning of stops. The peak was examined from the spectrum within the interval of 550 Hz to 10k Hz.

F2. F2 was measured at the onset of the vowel following the stop.

Measurements for Nasals

Nasal duration. Two types of nasal duration were measured: the raw nasal duration and the nasal duration relative to the duration of the whole word. The nasal duration was defined as the time interval from the onset of the nasal (which was determined by locating the onset of weak f1 [low in amplitude]) to the onset of the vowel (which was determined by locating the onset of darker formants [darker bands]).

F1 bandwidth. The bandwidth of F1 was obtained in the mid-point of nasals (e.g., Hawkins & Stevens, 1985; Styler, 2017).

Max A1 & mean A1. A1 represents the amplitude of F1. In order to calculate the max and mean values of F1 amplitude, the F1 bandwidth was first extracted. The peak and mean amplitudes were searched within the F1 bandwidth (i.e., from F1 – (0.5xF1bandwidth) to F1 + (0.5xF1bandwidth)).

F2. F2 was measured for nasals at the onset of the following vowel.

Statistical Analyses

Using the lme4 package in R (Bates, Mächler, Bolker, & Walker, 2015), separate linear mixed-effect models were built for each acoustic parameter as dependent variable: for stops, the raw and relative VOT, the RMS burst amplitude, the spectral peak of the burst and F2 at the onset of the following vowel; and for nasals, the raw and relative nasal duration, the F1 bandwidth, and the mean A1 and F2 at the onset of following vowels. For the durational and amplitudinal acoustic parameters for stops, the linear mixed-effects model included as fixed effects: linear position (which was defined by whether the target word was located early or late in a sentence regardless of prosodic boundary; baseline = early position), speech style (baseline =

interactive speech), prosodic boundary (baseline = IP-initial position), voicing (baseline = voiceless), place of articulation (baseline = bilabial), and/or their interactions. The model also included participant as a random intercept. Vowel context (baseline = [ɪ]) was also included in the model as an additional fixed effect (without its interaction with other predictor variables) in order to control for between-subjects variability (which was introduced by assigning each vowel to each participant in a pair) in the evaluation of the other within-subject variables (i.e., linear position, speech style, prosodic boundary, voicing and place of articulation). For the spectral acoustic parameters of stops, the linear mixed-effects models included linear position, speech style, prosodic boundary, place of articulation, and/or their interactions as fixed effects, gender (baseline = female) without its interaction with other predictor variables as an additional fixed effect, and participant as a random intercept for stops in different vowel contexts (i.e., [ɪ] and [ʌ]). Because spectral acoustic parameters cannot be interpreted without considering vowel contexts, separate models were built separately for each vowel context (i.e., [ɪ] and [ʌ]). As a consequence, each model included different participants since, as stated above, each vowel context was assigned to each participant of a pair.

For nasals, with the exception of F2, the acoustic parameters were analyzed by using the linear mixed-effects models that included linear position (baseline = early position), prosodic boundary (baseline = IP-initial position), speech style (baseline = interactive speech), place of articulation (baseline = bilabial), and/or their interactions, as well as vowel context (baseline = [ɪ]) without its interactions with other predictor variables as an additional fixed effect, and participant as a random intercept. For F2, the linear mixed-effects models included linear position (baseline = early position), prosodic boundary (baseline = IP-initial position), speech style (baseline = interactive speech), and place of articulation (baseline = bilabial), and their

interactions as fixed effects, as well as gender (baseline = female) without its interactions with other predictor variables as an additional fixed effect, and participant as a random intercept. The models for F2 were run separately for nasals in different vowel contexts (i.e., [ɪ] and [ʌ]).

Subsequently, the linear mixed-effects models built for each acoustic parameter were backward fit via the step() function in the lmerTest package in R (Kuznetsova, Brockhoff, & Christensen, 2017), which iteratively removes terms from the model, starting with the highest-order interactions, based on the absence of a significant difference in likelihood ratio tests between the larger and smaller model. After backward fitting the models, the linear mixed-effects model with the best fit was reported. The levels of the fixed-effects variable in the linear mixed-effects model with the best fit were compared using the emmeans () function (Lenth, 2021). In order to simplify the interpretation of complex models, subsequent analyses were conducted separately on interactive and read speech when there was an interaction between four or more predictor variables (including speech style) that contributed to the model fit. In this case, the alpha level was adjusted to .025 for each subsequent analysis.

Chapter 3: Results

Stops

Raw and Relative VOT

The analysis of raw VOT will be reported first. In the analysis of raw VOT, the backward fitting of the big model revealed that the linear mixed-effects model with the best fit included an interaction between speech style, prosodic boundary, voicing, and place of articulation ($F(2, 1255) = 3.09$, $p<.05$) and an interaction between linear position, speech style, and voicing ($F(1, 1254) = 7.94$, $p<.01$). Given the significant four-way interaction between speech style, prosodic boundary, voicing, and place of articulation, subsequent analyses were conducted separately on the interactive and read speech conditions in order to simplify the interpretation of the results. For interactive speech, the linear mixed-effects model with the best fit included an interaction between prosodic boundary, voicing, and place of articulation ($F(2, 534) = 4.57$, $p<.025$) and an interaction between linear position and voicing ($F(1, 530) = 8.00$, $p<.01$). Table 2 summarizes the results of the linear mixed-effects model with the best fit on raw VOT for *interactive speech,* with early position (linear position), IP-initial position (prosodic boundary), voiceless (voicing), bilabial (place of articulation), and [ɪ] (vowel context) as baseline.

Table 2: Summary of linear mixed- effects model with best fit on raw VOT for interactive speech (α = .025).

	Estimate	**Standard Error**	***t***	***p***
(Intercept)	32.32	6.94	4.66	<.001
Linear position (late)	12.04	4.69	2.57	<.025
Boundary (IP-medial)	–13.49	7.77	–1.74	.08
Voicing (voiced)	–68.53	8.49	–8.07	<.001
Place (alveolar)	17.89	7.67	2.33	<.025
Place (velar)	33.20	7.94	4.18	<.001
Vowel ([ʌ])	10.25	6.20	1.66	.12
Linear position (late) x Voicing (voiced)	–19.18	6.73	–2.85	<.01
Boundary (IP-medial) x Voicing (voiced)	29.93	11.42	2.62	<.01
Boundary (IP-medial) x Place (alveolar)	–6.82	11.08	–0.62	.054
Boundary (IP-medial) x Place (velar)	–31.43	11.78	–2.67	<.01
Voicing (voiced) x Place (alveolar)	6.40	11.12	0.58	.056
Voicing (voiced) x Place (velar)	–27.35	11.11	–2.46	<.025
Boundary (IP-medial) x Voicing (voiced) x Place (alveolar)	–11.33	16.15	–0.70	.048
Boundary (IP-medial) x Voicing (voiced) x Place (velar)	36.70	16.84	2.18	.030

First, we focus on the results that are related to our main interest, prosodic boundary. The linear mixed-effects model with the best fit revealed that the three-way interaction between prosodic boundary, voicing, and place of articulation stemmed in part from the fact that the interaction between prosodic boundary and voicing was not observed across places of articulation. As shown in Table 2, in interactive speech, there was a significant interaction between prosodic boundary and voicing for bilabial stops. The releveling of place of articulation in the model revealed that there was a significant interaction between prosodic boundary and

voicing for velar stops (β = 66.63, *SE* = 12.30, *t*[534] = 6.42, *p*<.001) but not for alveolar stops (β =18.60, *SE* = 11.43, *t*[533] = 1.63, *p*>.1). Figure 7 shows the raw VOT for voiceless and voiced stops at bilabial, alveolar, and velar places of articulation in IP-initial and IP-medial positions in interactive speech.

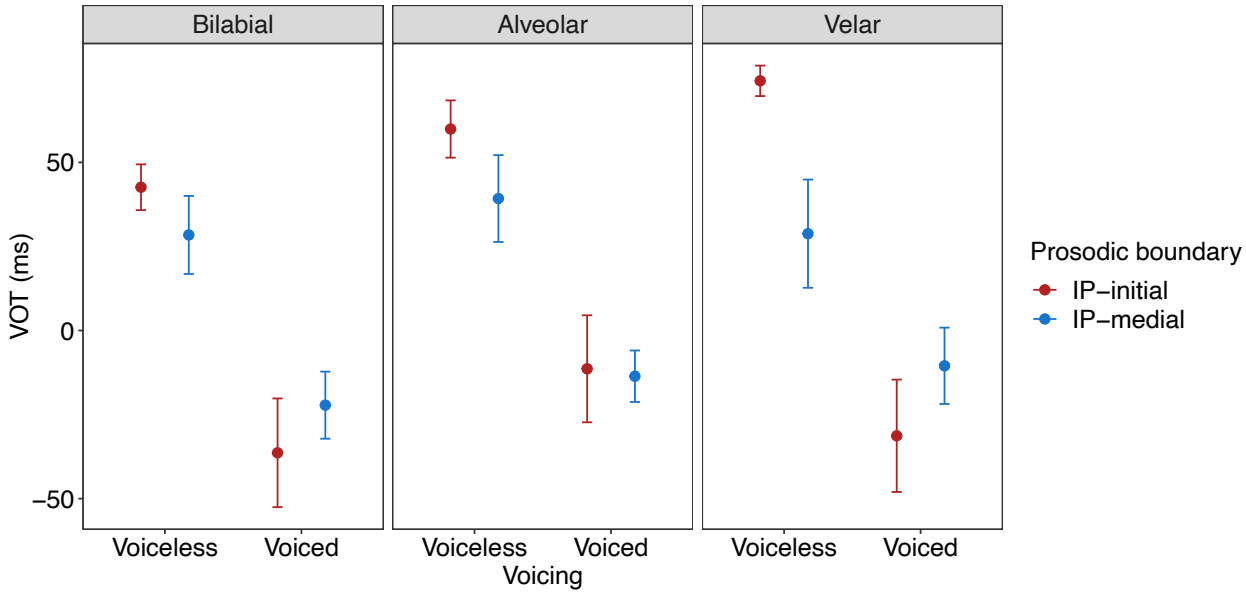


Figure 7: Raw VOT for voiceless and voiced stops at bilabial, alveolar, and velar places of articulation in IP-initial and IP-medial positions in interactive speech.

As shown in Figure 7, the significant interaction between prosodic boundary and voicing for bilabial place of articulation was driven by the raw VOT being higher (i.e., longer voicing lag) for voiceless stops in IP-initial position than for voiceless stops in IP-medial position (Table 2) but being lower (i.e., more prevoicing) for voiced stops in IP-initial position than for voiced stops in IP-medial position (β = –16.44, *SE* = 8.37, *t*[534] = –1.96, *p*=.05), ultimately resulting in a greater difference in raw VOT between voiceless and voiced stops in IP-initial position (β = 78.1, *SE* = 7.63, *t*[529] = 10.24, *p*<.001) compared to IP-medial position (β = 48.2, *SE* = 8.50, *t*[534] = 15.67, *p*<.001). Similarly, velar place of articulation showed an interaction between

prosodic boundary and voicing such that the raw VOT for voiceless stops was higher (i.e., longer voicing lag) in IP-initial position than in IP-medial position (β = 44.92, *SE* = 8.87, *t*[536] = 5.07, $p<.001$) whereas the raw VOT for voiced stops was lower (i.e., more prevoicing) in IP-initial position than in IP-medial position (β = –21.71, *SE* = 8.54, *t*[533] = –2.54, $p<.025$), resulting in a greater difference in raw VOT between voiceless and voiced stops in IP-initial position (β = 105.5, *SE* = 8.08, *t*[530] = 13.05, $p<.001$) compared to IP-medial position (β = 38.8, *SE* = 9.34, *t*[535] = 4.16, $p<.001$). For the alveolar place of articulation, the difference in raw VOT between voiceless and voiced stops appeared to be greater in IP-initial position (β = 71.7, *SE* = 8.09, *t*[530] = 8.87, $p<.001$) than in IP-medial position (β = 53.1, *SE* = 8.08, *t*[532] = 6.57, $p<.001$), but the interaction between prosodic boundary and voicing was not statistically significant. Thus, in interactive speech, the difference in raw VOT between voiceless and voiced stops was enhanced in IP-initial position compared to IP-medial position.

The significant interaction between prosodic boundary, voicing, and place of articulation in interactive speech was also partially due to the fact that an effect of place of articulation on the raw VOT was only found in IP-initial position, not in IP-medial position. Figure 8 shows the raw VOT for voiceless and voiced stops at bilabial, alveolar, and velar places of articulation in IP-initial and IP-medial positions in interactive speech (as Figure 7 does), but the conditions were reorganized to highlight the interaction between prosodic boundary and place of articulation.

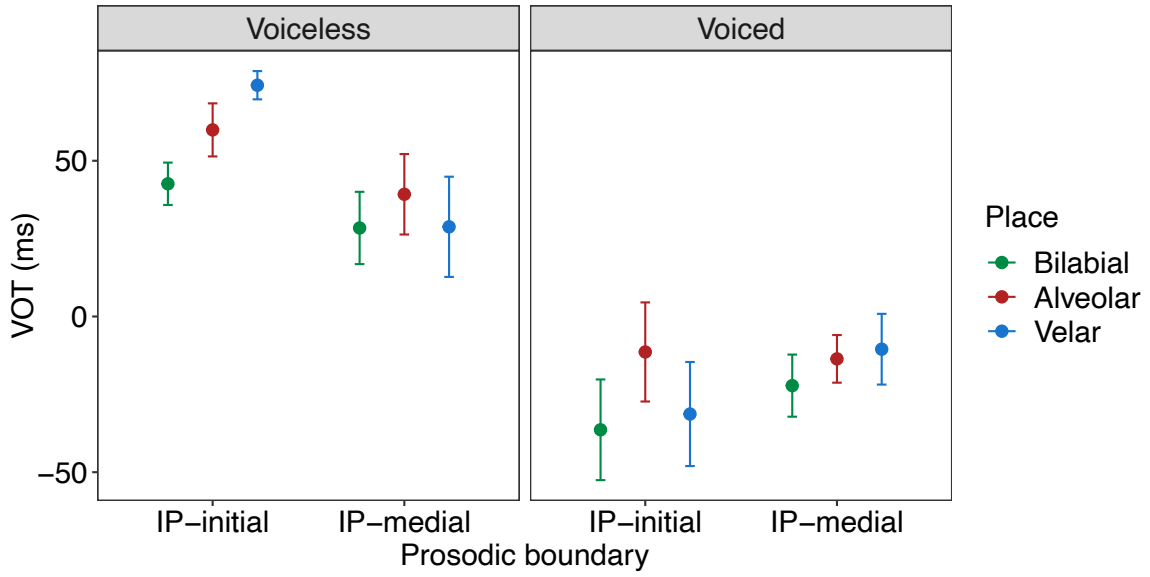


Figure 8: Raw VOT for voiceless and voiced stops at bilabial, alveolar, and velar places of articulation in IP-initial and IP-medial positions in interactive speech.

As shown in Figure 8, the raw VOT for both voiceless and voiced stops in interactive speech was distinguished better between bilabial, alveolar, and velar places of articulation in IP-initial position than in IP-medial position. More specifically, for voiceless stops, the difference in raw VOT between bilabial and other places of articulation was significant in IP-initial position, with the bilabial place of articulation being shorter in raw VOT than the alveolar or velar place of articulation (Table 2), but this effect disappeared in IP-medial position (bilabial – alveolar: β = –11.07, *SE* = 7.99, *t*[532] = –1.39, *p*>.1; bilabial – velar: β = –1.76, *SE* = 8.72, *t*[535] = –0.20, *p*>.1). For voiced stops, the raw VOT was significantly different between alveolar and other places of articulation in IP-initial position, with the alveolar place being longer in raw VOT than the bilabial or velar place of articulation (bilabial – alveolar: β = –24.30, *SE* = 8.05, *t*[530] = –3.02, *p*<.01; alveolar – velar: β = 18.45, *SE* = 8.16, *t*[530] = 2.26, *p*<.025), but not in IP-medial position (bilabial – alveolar: β = –6.15, *SE* = 8.55, *t*[535] = –0.72, *p*>.1; alveolar – velar: β = –

4.97, *SE* = 8.63, *t*[535] = –0.58, *p*>.1). In interactive speech, thus, the difference in raw VOT between places of articulation increased in IP-initial position compared to IP-medial position.

Next, we examine whether the effect of prosodic boundary on the raw VOT that was reported above could be explained by a difference in linear position (early vs. late). Figure 9 shows the raw VOT for voiceless and voiced stops in early and late positions in the interactive speech condition.

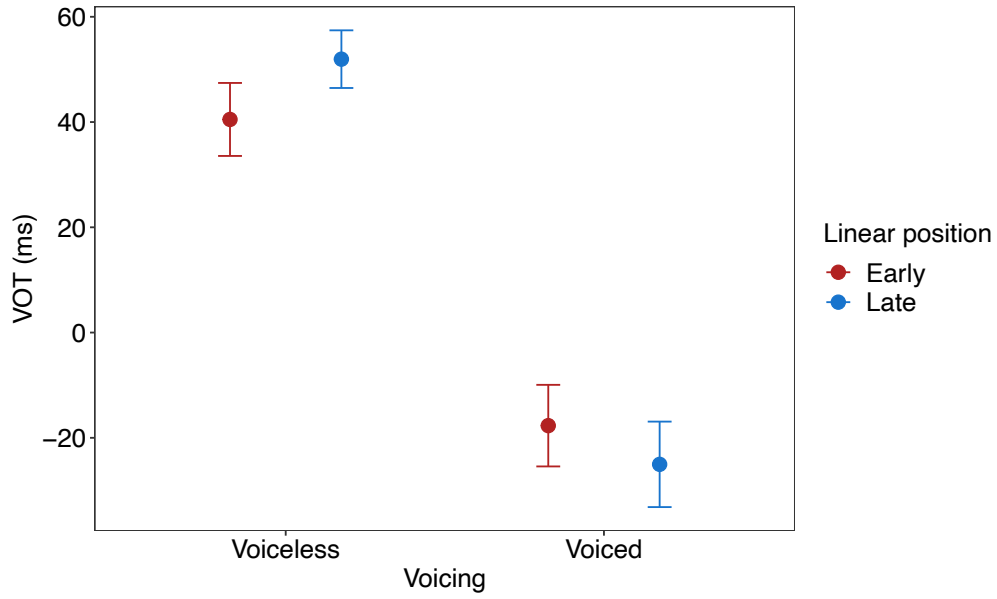


Figure 9: Raw VOT for voiceless and voiced stops in early and late positions in interactive speech.

As shown in Figure 9, in interactive speech, the raw VOT for voiceless stops was lower (i.e., shorter voicing lag) in the early position than in the late position (Table 2), whereas the raw VOT for voiced stops was higher (i.e., less prevoicing) in the early position than in the late position (β = 7.14, *SE* = 4.82, *t*[529] = 1.48, *p*>.1), resulting in a smaller difference in raw VOT between voiceless and voiced stops in the early position (β = 56.3, *SE* = 4.87, *t*[531] = 11.58, *p*<.001) compared to the late position (β = 75.5, *SE* = 4.70, *t*[532] = 16.07, *p*<.001). Recall that,

in the interaction between prosodic boundary and voicing, the raw VOT for voiceless stops was higher in IP-initial position than in IP-medial position whereas the raw VOT for voiced stops was lower in IP-initial position than in IP-medial position (Figure 9). These patterns show the opposite directionality when compared with the patterns observed for the interaction between linear position and voicing. Given that the IP-initial position (prosodic boundary) is comparable with the early position (linear position) and the IP-medial position (prosodic boundary) is comparable with the late position (linear position), if the effect of prosodic boundary is in fact confounded with the effect of linear position, the directionality of their effects should be the same. However, the results showed otherwise. Therefore, the effect of prosodic boundary in the raw VOT of interactive speech cannot be attributed to an effect of linear position.

Unlike interactive speech, the linear mixed-effects model with the best fit for read speech included a main effect of prosodic boundary ($F(1, 721) = 28.45$, $p<.001$), a main effect of voicing ($F(1, 719) = 620.72$, $p<.001$), and a main effect of place of articulation ($F(2, 718) = 20.04$, $p<.001$), without any interactions. Table 3 summarizes the results of the linear mixed-effects model with the best fit on the raw VOT for *read speech,* with IP-initial position (prosodic boundary), voiceless (voicing), bilabial (place of articulation), and [ɪ] (vowel context) as baseline. Figure 10 shows the raw VOT for voiceless and voiced stops at bilabial, alveolar, and velar places of articulation in IP-initial and IP-medial positions in read speech.

Table 3: Summary of linear mixed-effects model with best fit on raw VOT for read speech (α = .025).

	Estimate	Standard Error	*t*	*p*
(Intercept)	38.58	4.44	8.68	<.001
Boundary (IP-medial)	–12.19	2.28	–5.35	<.001
Voicing (voiced)	–56.52	2.27	–24.93	<.001
Place (alveolar)	10.66	2.75	3.88	<.001
Place (velar)	17.48	2.79	6.27	<.001
Vowel ([ʌ])	11.86	6.12	1.94	.07

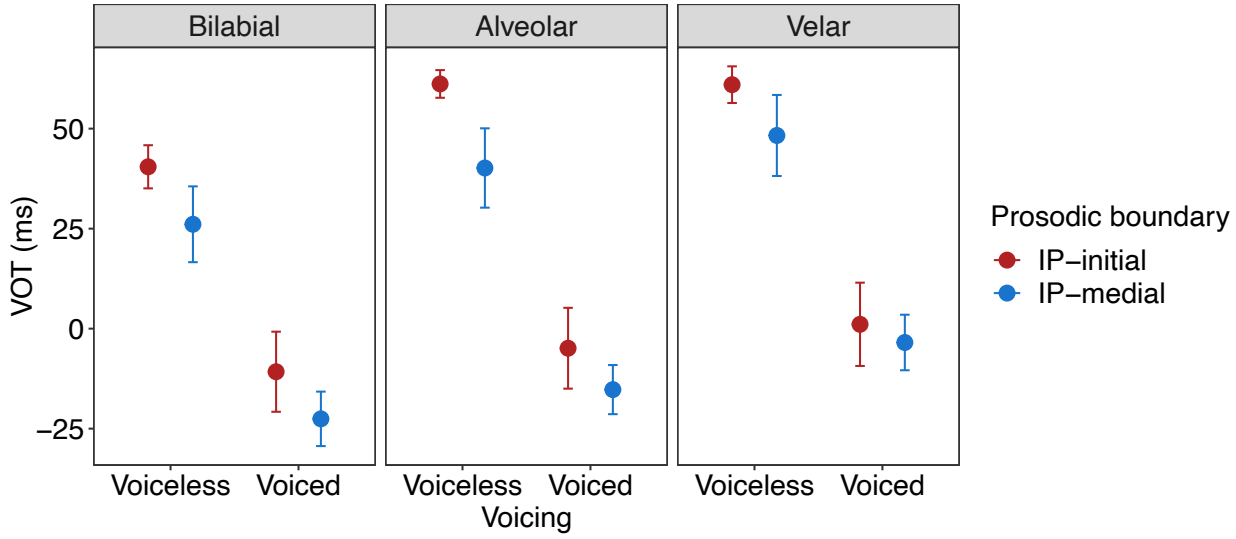


Figure 10: Raw VOT for voiceless and voices stops at bilabial, alveolar, and velar places of articulation in IP-initial and IP-medial positions in read speech.

As shown in Table 3, the linear mixed-effects model with the best fit revealed that the raw VOT was higher (i.e., longer voicing lag) in IP-initial position than in IP-medial position for both voiceless and voiced stops across places of articulation. Without any interactions, the effect of prosodic boundary was comparable across voiceless and voiced stops and across bilabial,

alveolar, and velar places of articulation in read speech. In addition, the linear mixed-effect model with the best fit also revealed that the effect of linear position did not contribute to the model fit without interacting with other predictor variables ($F(1, 714) = 0.79$, $p>.1$), whereas prosodic boundary did. Thus, the effect of linear position again did not appear to be confounded with the effect of prosodic boundary in read speech. Therefore, the results from interactive and read speech taken together suggest that the difference in raw VOT between voiceless and voiced stops or between bilabial, alveolar, and velar places of articulation increased in IP-initial position compared to IP-medial position in interactive speech, whereas it did not differ between the IP-initial and IP-medial positions in read speech.

We now move on to the analyses of relative VOT over word duration. The backward fitting of the big model yielded a linear mixed-effects model that included an interaction between prosodic boundary, speech style, and voicing ($F(1, 1253) = 9.90$, $p<.01$), an interaction between linear position, speech style, and voicing ($F(1, 1253) = 7.02$, $p<.01$), and finally an interaction between linear position, prosodic boundary, speech style, and place of articulation ($F(2, 1253) = 3.43$, $p<.05$). Since the four-way interaction between linear position, prosodic boundary, speech style, and place of articulation contributed the model fit, subsequent analyses were conducted separately on interactive and read speech to simplify the interpretation of the results. In the subsequent analyses for interactive speech, the linear mixed-effects model with the best fit revealed that there were an interaction between prosodic boundary, voicing, and place of articulation ($F(2, 534) = 3.74$, $p<.025$) and an interaction between linear position and voicing ($F(1, 530) = 7.39$, $p<.001$). Table 4 summarizes the results of the linear mixed-effects model with the best fit on relative VOT for *interactive speech* with early position (linear position), IP-initial position (prosodic boundary), voiceless (voicing), bilabial (place of articulation), and [ɪ]

(vowel context) as baseline. Figure 11 shows the relative VOT measurements for voiceless and voiced stops at bilabial, alveolar, and velar places of articulation in IP-initial and IP-medial positions in interactive speech.

Table 4: Summary of linear mixed-effects model with best fit on relative VOT (over word duration) for interactive speech (α = .025).

	Estimate	**Standard Error**	***t***	***p***
(Intercept)	0.11	0.02	6.04	<.001
Linear position (late)	0.03	0.01	2.60	<.01
Boundary (IP-medial)	–0.05	0.02	–2.52	<.025
Voicing (voiced)	0.20	0.02	–8.69	<.001
Place (alveolar)	0.04	0.02	1.90	.05
Place (velar)	0.09	0.02	3.99	<.001
Vowel ([ʌ])	0.03	0.02	1.90	.07
Linear position (late) x Voicing (voiced)	–0.05	0.02	–2.74	<.01
Boundary (IP-medial) x Voicing (voiced)	0.08	0.03	2.59	<.01
Boundary (IP-medial) x Place (alveolar)	–0.02	0.03	–0.54	>.1
Boundary (IP-medial) x Place (velar)	–0.08	0.03	–2.43	<.025
Voicing (voiced) x Place (alveolar)	0.03	0.03	0.81	>.1
Voicing (voiced) x Place (velar)	–0.07	0.03	–2.41	<.025
Boundary (IP-medial) x Voicing (voiced) x Place (alveolar)	–0.03	0.04	–0.68	>.1
Boundary (IP-medial) x Voicing (voiced) x Place (velar)	0.09	0.05	1.93	.05

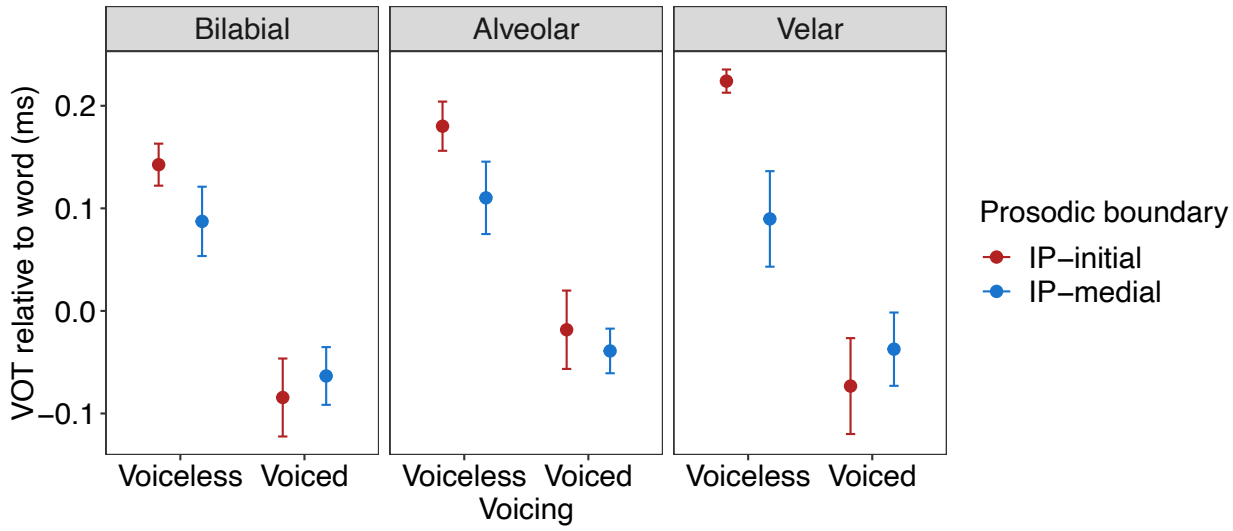


Figure 11: Relative VOT (over the word duration) for voiceless and voiced stops at bilabial, alveolar, and velar places of articulation in IP-initial and IP-medial positions in interactive speech.

We begin by examining the results for the effect of prosodic boundary, which is the main interest of the present study. In interactive speech, the three-way interaction between prosodic boundary, voicing, and place of articulation contributed to the model fit, partly reflecting the fact that the interaction between prosodic boundary and voicing was significant for bilabial (Table 4) and velar places of articulation (β = 0.17, *SE* = 0.03, *t*[534] = 5.05, *p*<.001), but not for the alveolar place of articulation (β = 0.05, *SE* = 0.03, *t*[533] = 1.62, *p*>.1). As shown in Figure 11, for bilabial and velar places of articulation, voiceless stops were higher in relative VOT in IP-initial position than in IP-medial position (bilabial: β = 0.05, *SE* = 0.02, *t*[530] = 2.52, *p*<.025; velar: β = 0.13, *SE* = 0.02, *t*[536] = 5.43, *p*<.001), whereas voiced stops were lower in relative VOT in IP-initial position than in IP-medial position; although this effect was not significant (bilabial: β = –0.03, *SE* = 0.02, *t*[534] = –1.20, *p*>.1; velar: β = –0.04, *SE* = 0.02, *t*[533] = –1.63, *p*>.1), it resulted in a greater difference in relative VOT between voiceless and voiced stops in IP-initial position compared to IP-medial position. For the alveolar place of articulation,

however, both voiceless and voiced stops were higher in relative VOT in IP-initial position than in IP-medial position, although the effect was statistically significant only for voiceless stops (voiceless: $\beta = 0.07$, $SE = 0.02$, $t[534] = 3.22$, $p<.01$; voiced: $\beta = 0.02$, $SE = 0.02$, $t[536] = 0.84$, $p>.1$) and did not interact with prosodic boundary and voicing. In interactive speech, therefore, the difference in relative VOT between voiceless and voiceless stops increased in IP-initial position compared to IP-medial position at least for bilabial and velar places of articulation.

The three-way interaction between prosodic boundary, voicing, and place of articulation in interactive speech also stemmed in part from the fact that the effect of place of articulation was found only in IP-initial position, not in IP-medial position, and the effect of place of articulation depended on voicing. Figure 12 shows relative VOT for voiceless and voiced stops at bilabial, alveolar, and velar places of articulation in IP-initial and IP-medial position in interactive speech (as Figure 11 does), but the conditions were reorganized to highlight the interaction between prosodic boundary and place of articulation.

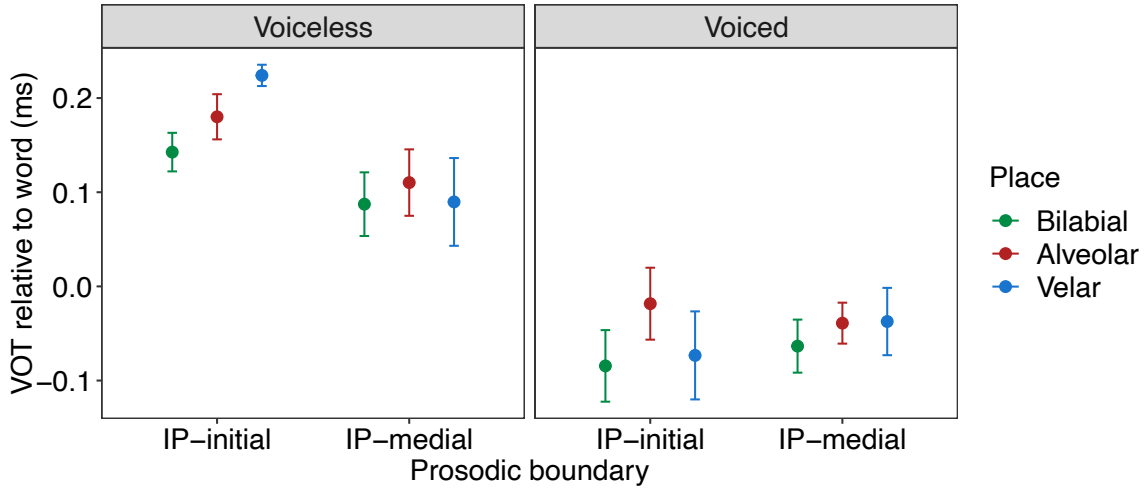


Figure 12: Relative VOT (over the word duration) for voiceless and voiced stops at bilabial, alveolar, and velar places of articulation in IP-initial and IP-medial positions in interactive speech.

As shown in Figure 12, for both voiceless and voiced stops, the difference between places of articulation generally increased in IP-initial position relative to IP-medial position. The linear mixed-effects model revealed that, for voiceless stops, relative VOT was significantly different between bilabial and velar places of articulation only in IP-initial position, with the bilabial place of articulation being lower in relative VOT than the velar place of articulation (Table 4), but with this significant difference disappearing in IP-medial position (β = –0.01, SE = 0.02, t[534] = –0.35, p>.1). In addition, voiceless stops tended to differ in relative VOT between alveolar and velar places of articulation in IP-initial position, with the alveolar place of articulation being lower in relative VOT than the velar place of articulation (β = –0.05, SE = 0.02, t[533] = –2.13, p=.03), but, again, with this effect disappearing in IP-medial position (β = 0.02, SE = 0.02, t[535] = 0.64, p>.1). For voiced stops, relative VOT for the bilabial place of articulation was significantly lower than that for the alveolar place of articulation in IP-initial position (Table 4), whereas no significant difference between bilabial and alveolar places of articulation was found in IP-medial position (β = –0.02, SE = 0.02, t[534] = –0.78, p>.1). Voiced stops also differed in relative VOT between the alveolar and velar places of articulation in IP-initial position (β = 0.05, SE = 0.02, t[530] = 2.30, p<.025) but not in IP-medial position (β = –0.01, SE = 0.02, t[535] = –0.24, p>.1). Thus, in interactive speech, the difference in relative VOT between places of articulation increased in IP-initial position compared to IP-medial position.

Now that we have shown that the effect of prosodic boundary on relative VOT in interactive speech differs depending on the voicing and place of articulation of the stop, we examine whether this effect could be explained by linear position. Recall that the linear mixed-effects model with the best fit included an interaction between linear position and voicing (F(1,

530) = 7.39, p<.001). Figure 13 shows relative VOT for voiceless and voiced stops in early and late positions in interactive speech.

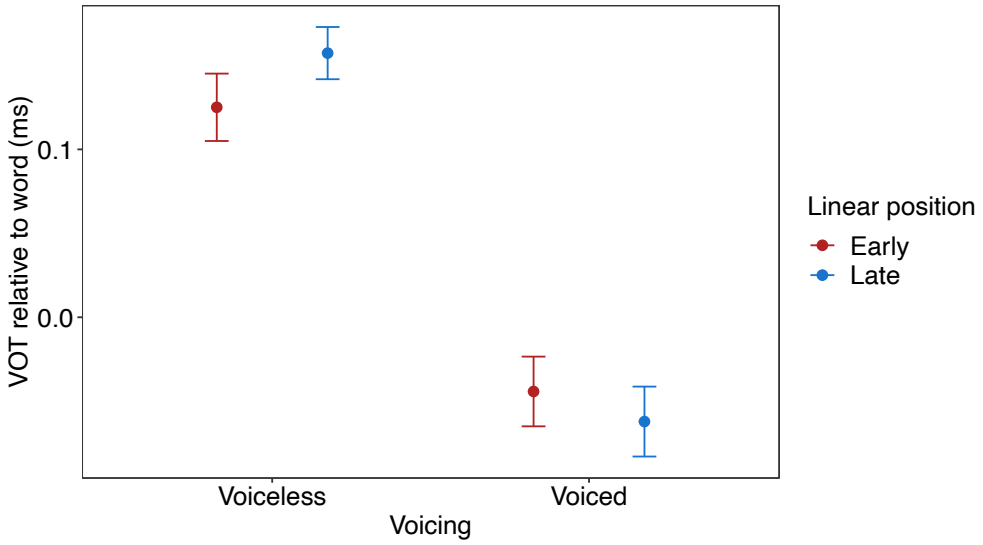


Figure 13: Relative VOT (over the word duration) for voiceless and voiced stops in early and late positions in interactive speech.

As shown in Figure 13, in interactive speech, voiceless stops showed a lower relative VOT in early position than in late position (β = –0.03, SE = 0.01, t[529] = –2.60, p<.001), whereas voiced stops showed a higher relative VOT in early position than in late position (β = 0.02, SE = 0.01, t[529] = 1.30, p>.1), resulting in a smaller difference between voiceless and voiced stops in early position (β = 0.17, SE = 0.01, t[531] = 12.61, p<.001) compared to the late position (β = 0.22, SE = 0.01, t[532] = 17.00, p<.001). If the effect of prosodic boundary found above was purely due to whether the target words were located earlier or later in a sentence (i.e., linear position), the interaction between linear position and voicing should show patterns that are similar to the interaction between prosodic boundary and voicing. However, prosodic boundary and linear position showed opposite patterns on voicing in terms of relative VOT. This means

that the effect of prosodic boundary cannot be accounted for by the effect of linear position in interactive speech.

In the analysis of relative VOT in read speech, the linear mixed-effects model with the best fit included a main effect of prosodic boundary ($F(1, 723) = 49.94$, $p<.001$), a main effect of voicing ($F(1, 721) = 688.85$, $p<.001$), and a main effect of place of articulation ($F(2, 719) = 17.92$, $p<.001$). Table 5 summarizes the results of the linear mixed-effects model with the best fit on relative VOT for voiceless and voiced stops at bilabial, alveolar, and velar places of articulation in IP-initial and IP-medial positions in read speech, with IP-initial position (prosodic boundary), voiceless (voicing), bilabial (place of articulation), and [ɪ] (vowel context) as baseline. Figure 14 shows relative VOT for voiceless and voiced stops at bilabial, alveolar, and velar places of articulation in IP-initial and IP-medial positions in read speech.

Table 5: Summary of linear mixed-effects model with best fit on relative VOT (over word duration) for read speech ($\alpha = .025$).

	Estimate	Standard Error	*t*	*p*
(Intercept)	0.140	0.013	10.59	<.001
Boundary (IP-medial)	–0.049	0.007	–7.08	<.001
Voicing (voiced)	–0.182	0.007	–26.27	<.001
Place (alveolar)	0.029	0.008	3.46	<.001
Place (velar)	0.050	0.008	5.95	<.001
Vowel ([ʌ])	0.031	0.018	1.75	.09

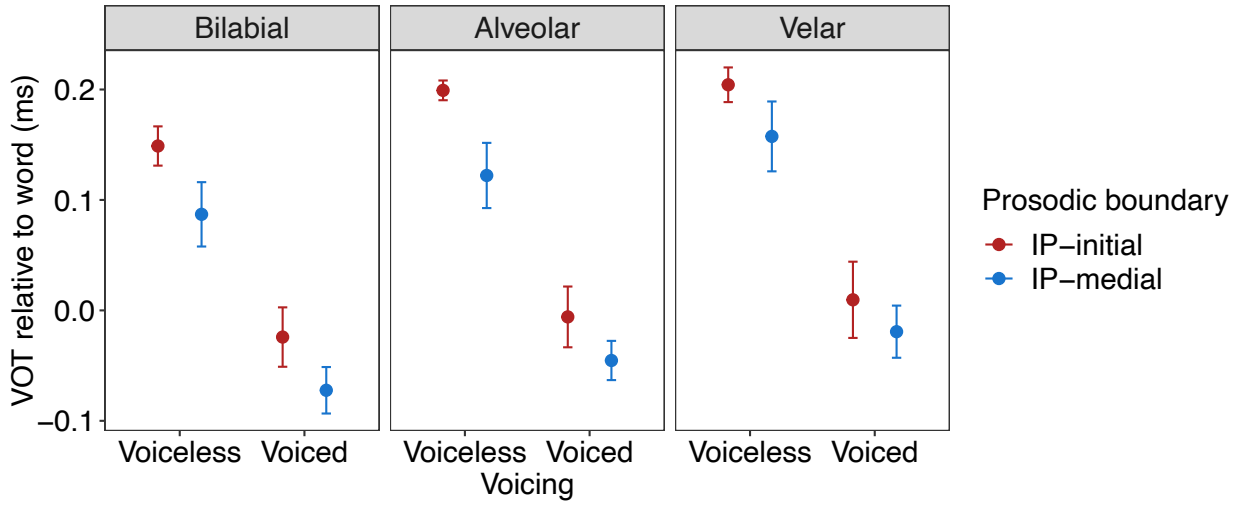


Figure 14: Relative VOT (over the word duration) for voiceless and voiced stops at bilabial, alveolar, and velar place of articulation in IP-initial and IP-medial positions in read speech.

The linear mixed-effects model with the best fit (Table 5) revealed a main effect of prosodic boundary without any interactions with voicing and/or place of articulation in read speech. As shown in Figure 14, the main effect of prosodic boundary indicates that relative VOT was significantly higher in IP-initial position than in IP-medial position (Table 5) regardless of voicing and/or place of articulation. The linear mixed-effects model with the best fit also revealed that linear position and its interaction with other predictor variables did not contribute to the model fit in read speech. Thus, the observed main effect of prosodic boundary on relative VOT in read speech cannot be explained by a difference in linear position. Taken together, the difference in relative VOT between voiceless and voiced stops or between bilabial, alveolar, and velar places of articulation increased in IP-initial position compared to IP-medial position in interactive speech, whereas it did not differ between IP-initial and IP-medial positions in read speech.

In sum, these results for both raw and relative VOT show that the difference between voiceless and voiced stops or between bilabial, alveolar, and velar places of articulation

increased in IP-initial position compared to IP-medial position only in interactive speech, supporting paradigmatic contrast enhancement. On the other hand, in read speech, raw and relative VOT values increased in IP-initial position compared to IP-medial position, becoming more consonantal regardless of voicing and/or place of articulation, thus supporting syntagmatic contrast enhancement.

RMS Burst Amplitude

For the RMS amplitude for the burst, the backward fitting of the big model showed that the linear mixed-effects model with the best fit included an interaction between prosodic boundary and voicing ($F(1, 1289) = 8.60$, $p<.01$), an interaction between prosodic boundary and speech style ($F(1, 1289) = 7.43$, $p<.001$), an interaction between voicing and place ($F(2, 1289) = 12.87$, $p<.001$), and a main effect of linear position ($F(1, 1289) = 5.08$, $p<.05$). Table 6 summarizes the results of the linear mixed-effects model with the best fit on RMS burst amplitude, with early position (linear position), IP-initial position (prosodic boundary), interactive speech (speech style), voiceless (voicing), bilabial (place of articulation), and [ɪ] (vowel context) as baseline. First, we report how the effect of prosodic boundary interact with voicing or speech style. Figure 15 presents the RMS amplitude of the burst for voiceless and voiced stops in IP-initial and IP-medial positions.

Table 6: Summary of linear mixed-effects model with best fit on the RMS amplitude for the burst part of stops.

	Estimate	Standard Error	*t*	*p*
(Intercept)	52.99	1.10	48.29	<.001
Linear position (late)	–0.52	0.23	–2.26	<.05
Speech style (read)	–1.91	0.33	–5.86	<.001
Voicing (voiced)	1.00	0.45	2.20	<.05
Boundary (IP-medial)	–0.90	0.42	–2.14	<.05
Place (alveolar)	5.40	0.39	13.80	<.001
Place (velar)	1.52	0.40	3.77	<.001
Vowel ([ʌ])	–1.34	1.66	–0.81	>.1
Speech style (read) x Boundary (IP-medial)	1.27	0.47	2.73	<.01
Voicing (voiced) x Place (alveolar)	–1.78	0.56	–3.21	<.01
Voicing (voiced) x Place (velar)	1.04	0.57	1.83	.067
Voicing (voiced) x Boundary (IP-medial)	1.35	0.46	2.93	<.01

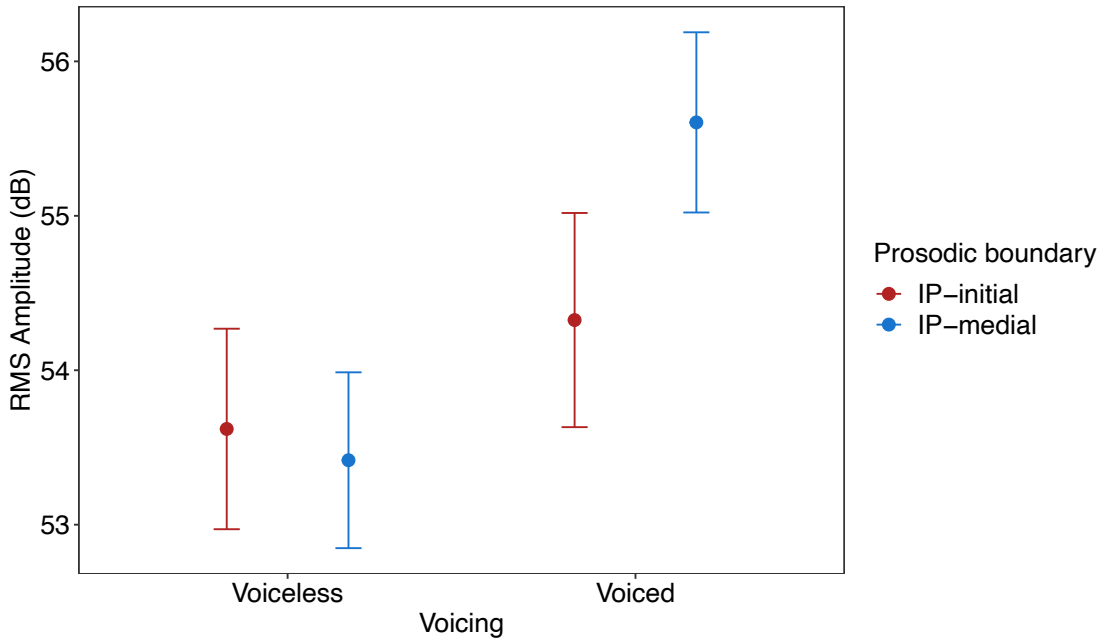


Figure 15: RMS amplitude of the burst for voiceless and voiced stops in IP-initial and IP-medial positions.

As shown in Figure 15, voiceless stops showed no significant difference in RMS burst amplitude between IP-initial and IP-medial positions (β = 0.26, SE = 0.33, t[1290] = 0.79, p>.1), whereas voiced stops showed a significant difference in the RMS amplitude of the burst between IP-initial and IP-medial positions such that the RMS amplitude was significantly lower in IP-initial position than in IP-medial position (β = –1.09, SE = 0.33, t[1290] = –3.31, p<.01). As a result, the difference in RMS amplitude between voiceless and voiced stops was smaller in IP-initial position (β= –0.75, SE = 0.33, t[1289] = –2.32, p<.05) than in IP-medial position (β= –2.10, SE = 0.33, t[1289] = –6.45, p<.001). Thus, the RMS amplitude did not increase, but rather decreased the difference between voiceless and voiced stops in IP-initial position compared to IP-medial position.

While prosodic boundary interacted with voicing, the linear mixed-effects model with the best fit revealed that prosodic boundary also interacted with speech style. Figure 16 shows the RMS amplitude of the burst of stops in IP-initial and IP-medial positions in interactive and read speech.

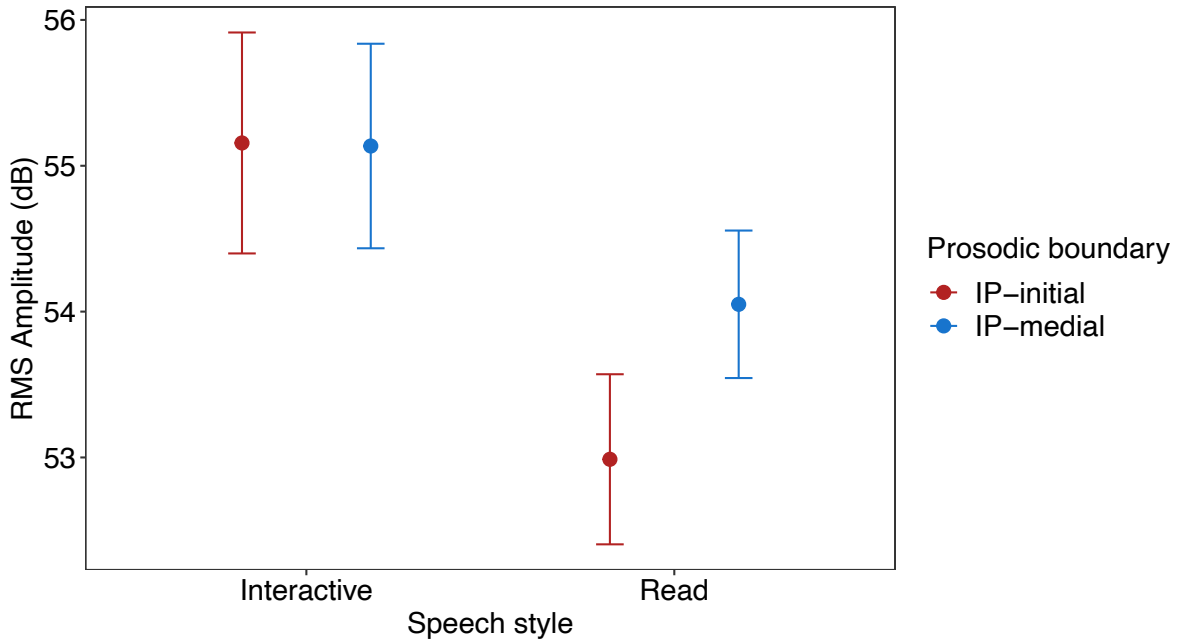


Figure 16: RMS amplitude in the burst of stops in IP-initial and IP-medial positions in interactive and read speech.

Figure 16 revealed that the RMS amplitude of the stop burst significantly differed between IP-initial and IP-medial positions only in read speech, with the RMS amplitude being significantly lower in IP-initial position than in IP-medial position (β= –1.05, *SE* = 0.31, *t*[1290] = –3.42, *p*<.001). This effect of prosodic boundary disappeared in interactive speech (β= 0.22, *SE* = 0.35, *t*[1290] = 0.63, *p*>.1). In other words, the effect of prosodic boundary that the lowered RMS amplitude of the stop burst in IP-initial position compared to IP-medial position was found only in read speech.

Now, we turn to the issue of whether the effect prosodic boundary found above can be explained by the effect of linear position. Figure 17 shows the RMS burst amplitude of the burst in IP-initial and IP-medial positions in early and late positions. Note that the effect of prosodic boundary and the effect of linear position are presented together in Figure 17 just so that the directionality of the effects can be compared.

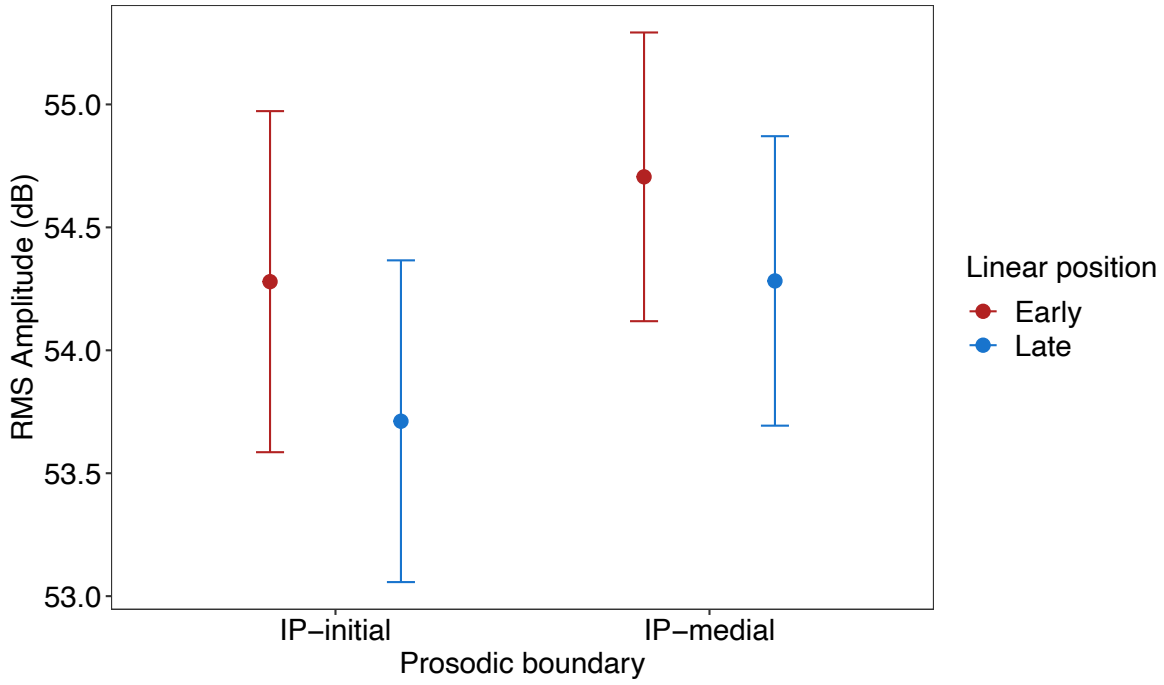


Figure 17: RMS amplitude of the burst in IP-initial and IP-medial positions in early and late positions.

Table 6 shows that the linear mixed-effects model with the best fit on RMS amplitude did not include interactions between linear position and other predictor variables, but there was a main effect of linear position such that the RMS amplitude of the stop burst was higher in early position than in late position (Figure 17). However, as shown in Figure 17, the overall mean RMS amplitude (i.e., across other predictor variables) was lower in IP-initial position (mean: 53.98) than in IP-medial position (mean: 54.49), showing the opposite directionality when compared with the patterns found for linear position. Thus, the patterns of RMS amplitude driven by prosodic boundary cannot be explained by linear position.

Taken together, only voiced stops, not voiceless stops, showed a lowering of the RMS amplitude in IP-initial position compared to IP-medial position, yielding to a decrease in difference between voiceless and voiced stops in IP-initial position compared to IP-medial position. In addition, the general lowering of the RMS amplitude of the stop burst in IP-initial

position compared to IP-medial position was observed only in read speech. The effect of prosodic boundary appears to be limited to a specific type of stop (i.e., voiced stop) and a specific speech style (i.e., read speech). In terms of linguistic function of the effect of prosodic boundary, the observed patterns seem to support neither syntagmatic nor paradigmatic contrast enhancement. These results will be discussed in detail in the "Discussion" section.

Spectral Peak of the Burst

As stated above, the analyses of the spectral peak of the burst were conducted separately for stops followed by different vowels [ɪ] and [ʌ]. We first focus on the acoustic analyses for stops followed by the vowel [ɪ]. The big model for stops followed by the vowel [ɪ] was backward fit on the spectral peak of the burst and reduced to a linear mixed-effects model that included an interaction between speech style, prosodic boundary, and place of articulation ($F(2, 791) = 2.65$, p=.07). Table 7 summarizes the results of the linear mixed-effects model with the best fit on the spectral peak of the burst at the onset of the following vowel [ɪ], with interactive speech (speech style), IP-initial position (prosodic boundary), and bilabial (place of articulation) as baseline. Figure 18 shows the spectral peak of the burst at the onset of the following vowel [ɪ] for stops at bilabial, alveolar, and velar places in IP-initial and IP-medial positions in interactive and read speech.

Table 7: Summary of linear mixed-effects model with best fit on spectral peak of burst at the onset of the following vowel [ɪ].

	Estimate	Standard Error	*t*	*p*
(Intercept)	1584.20	254.44	6.23	<.001
Speech style (read)	253.91	261.29	0.98	>.1
Boundary (IP-medial)	–194.83	281.87	–0.69	>.1
Place (alveolar)	4079.19	281.49	14.49	<.001
Place (velar)	2015.99	276.32	7.30	<.001
Gender (male)	546.19	408.69	1.34	>.1
Speech style (read) x Boundary (IP-medial)	–147.29	374.70	–0.39	>.1
Speech style (read) x Place (alveolar)	–1243.03	378.17	–3.29	<.01
Speech style (read) x Place (velar)	–575.97	375.43	–1.53	>.1
Boundary (IP-medial) x Place (alveolar)	–1125.59	402.14	–2.80	<.01
Boundary (IP-medial) x Place (velar)	–55.22	411.00	–0.13	>.1
Speech style (read) x Boundary (IP-medial) x Place (alveolar)	1198.71	532.91	2.25	<.05
Speech style (read) x Boundary (IP-medial) x Place (velar)	391.25	543.03	0.72	>.1

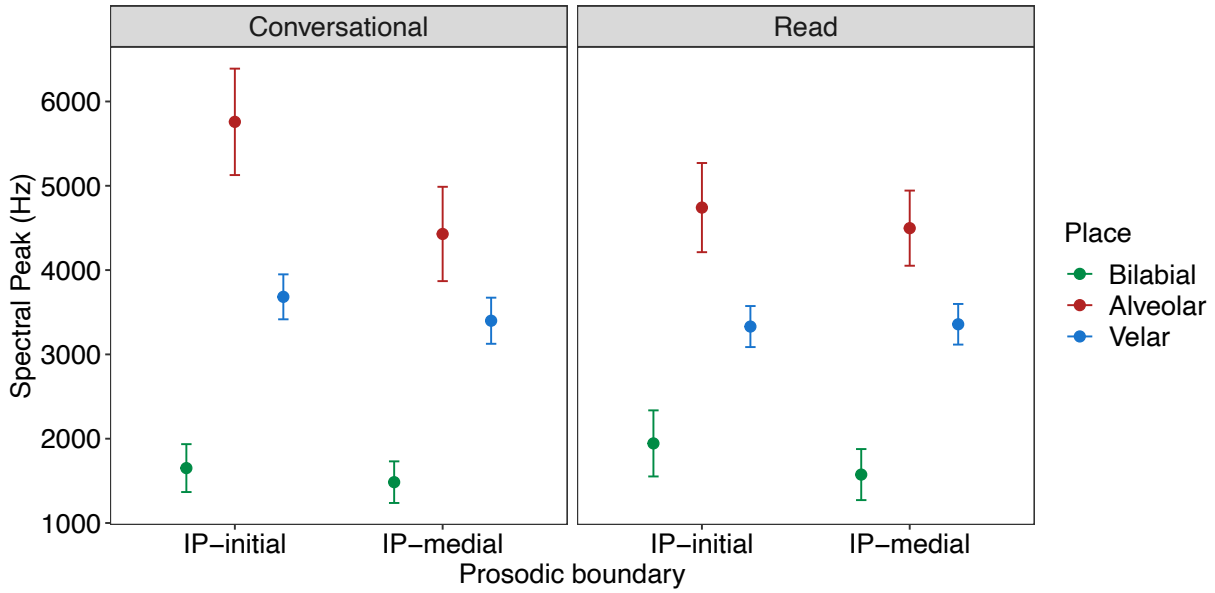


Figure 18: Spectral peak of the burst at the onset of the following vowel [ɪ] for stops at bilabial, alveolar, and velar places of articulation in IP-initial and IP-medial positions in interactive and read speech.

The three-way interaction between speech style, prosodic boundary, and place of articulation stemmed in part from the fact that an interaction between prosodic boundary and place of articulation was found in interactive speech but not in read speech. More specifically, as shown in Figure 18, for the interactive speech condition, only the alveolar place of articulation yielded a significantly higher spectral peak in IP-initial position than in IP-medial position (β =1320.43, SE = 289, t[793] = 4.57, p<.001), resulting in a greater difference between bilabial and alveolar stops in IP-initial position (Table 7) than in IP-medial position (β = –2954, SE = 287, t[791] = –10.29, p<.001). However, the releveling of speech style in the model revealed that the difference in burst spectral peak between bilabial and alveolar places of articulation was similar across IP-initial and IP-medial positions (β = 73.12, SE = 349.32, t[790] = 0.21, p>1). In addition, this heightened spectral peak of the burst for stops at the alveolar place of articulation in IP-initial position compared to IP-medial position in interactive speech also resulted in a greater difference between alveolar and velar places of articulation in IP-initial position (β = 2063, SE = 286, t[791] = 7.22, p<.001) than in IP-medial position (β = 993, SE = 302, t[792] = 3.29, p<.01). Again, however, the releveling of speech style in the model revealed that the difference in burst spectral peak between alveolar and velar places of articulation was similar across IP-initial and IP-medial positions (β = 262.92, SE = 354.77, t[790] = 0.74, p>.1). Moreover, the effect of prosodic boundary found above cannot be accounted for by an effect of linear position because linear position or its interaction with other predictor variables did not contribute to the model fit. Taken together, for the spectral peak of the burst at the onset of the vowel [ɪ], the increased difference between bilabial, alveolar, and velar place of articulation in IP-initial position compared to IP-medial position was only found in interactive speech.

Next, the results for stops followed by the vowel [ʌ] are examined. The backward fitting of the big model for stops followed by the vowel [ʌ] revealed a reduced linear mixed-effects model that included a main effect of place of articulation across other predictor variables ($F(2, 499) = 113.39$, $p<.001$). Table 8 summarizes the results of the linear mixed-effects model with the best fit on the spectral peak of the burst for stops followed by the vowel [ʌ], with bilabial (place of articulation) and female (gender) as baseline. Figure 19 shows the spectral peak of the burst for bilabial, alveolar, and velar stops followed by the vowel [ʌ] in IP-initial and IP-medial positions in interactive and read speech.

Table 8: Summary of linear mixed-effects model with best fit on the spectral peak of the burst at the onset of the following vowel [ʌ].

	Estimate	**Standard Error**	***t***	***p***
(Intercept)	1806.92	752.57	2.40	.059
Place (alveolar)	2819.31	195.97	14.39	<.001
Place (velar)	666.77	201.25	3.31	<.001
Gender (male)	–1073.11	1134.25	–0.95	>.1

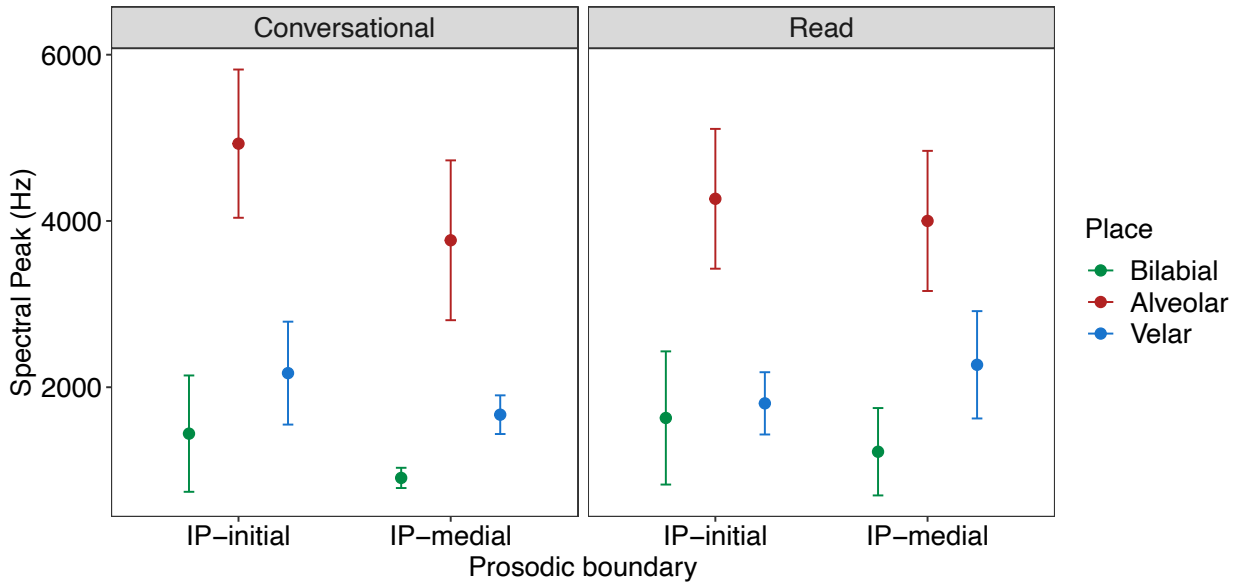


Figure 19: Spectral peak of the burst at the onset of the following vowel [ʌ] for stops at bilabial, alveolar, and velar places of articulation in IP-initial and IP-medial positions in interactive and read speech.

As shown in Table 8, a main effect of prosodic boundary or its interaction with speech style and/or place of articulation was not included in the linear mixed-effects model with the best fit on the spectral peak of the burst. However, it is noteworthy that, as Figure 19 shows, the differences in the spectral peak of the burst between bilabial and alveolar places of articulation and between alveolar and velar places of articulation were greater in IP-initial position than in IP-medial position in interactive speech, and these patterns were not observed in read speech. The interaction between speech style, prosodic boundary, and place of articulation did not reach significance potentially due to the greater variability in the realization of the spectral peak for stops followed by the vowel [ʌ] compared to those followed by the vowel [ɪ]. Moreover, the linear mixed-effects model with the best fit also showed that linear position or its interaction with the other predictor variables did not influence the spectral peak of the burst.

Therefore, the results, at least for stops followed by the vowel [ɪ], have shown that the difference in the spectral peak between stops at bilabial, alveolar, and velar places of articulation

increased in IP-initial position compared to IP-medial position only in interactive speech and not in read speech. In other words, the enhancement of phonological distinctions in terms of the spectral peak of the burst between bilabial, alveolar, and velar stops in IP-initial position compared to IP-medial position was only found in interactive speech, supporting paradigmatic contrast enhancement.

F2

For the spectral peak of the burst, all analyses of F2 were also conducted separately for stops followed by different vowels [ɪ] and [ʌ]. First, the results for stops followed by the vowel [ɪ] are reported. The backward fitting of the big model for stops followed by the vowel [ɪ] revealed that the linear mixed-effects model with the best fit included a main effect of speech style ($F(1, 797) = 5.21$, $p<.05$) and a main effect of place of articulation ($F(2, 796) = 168.49$, $p<.001$). Table 9 summarizes the results of the linear mixed-effects model with the best fit on F2 at the onset of vowel [ɪ], with interactive speech (speech style), bilabial (place of articulation), and female (gender) as baseline.

Table 9: Summary of linear mixed-effects model with best fit on F2 at the onset of the following vowel [ɪ].

	Estimate	Standard Error	*t*	*p*
(Intercept)	2241.85	39.71	56.46	<.001
Speech style (read)	–30.04	13.07	–2.30	<.05
Place (alveolar)	–15.45	15.58	–0.99	>.1
Place (velar)	246.28	15.88	15.51	<.001
Gender (male)	–164.59	88.92	–1.85	>.1

As the linear mixed-effects model with the best fit on F2 revealed (Table 9), there was no interaction between prosodic boundary and place of articulation, suggesting that prosodic boundary did not change the effect of place of articulation on F2 at the onset of the following vowel [ɪ]. The linear mixed-effects model with the best fit also showed that linear position or its interaction with other predictor variables also did not influence F2 at the onset of the following vowel [ɪ].

For F2 at the onset of the vowel [ʌ], on the other hand, the backward fitting of the big model yielded a linear mixed-effects model with the best fit that included an interaction between prosodic boundary and place of articulation ($F(2, 483) = 5.68$, $p<.01$) and an interaction between prosodic boundary and linear position ($F(1, 483) = 4.10$, $p<.05$). Table 10 summarizes the results of the linear mixed-effects model with the best fit on F2 at the onset of the vowel [ʌ], with IP-initial position (prosodic boundary), early (linear position), and bilabial (place of articulation) as baseline. First, we focus on the effect of boundary by examining the interaction between prosodic boundary and place of articulation. Figure 20 shows F2 at the onset of the vowel [ʌ] for stops at bilabial, alveolar, and velar places of articulation in IP-initial and IP-medial positions.

Table 10: Summary of linear mixed-effects model with best fit on F2 at the onset of the following vowel [ʌ].

	Estimate	Standard Error	*t*	*p*
(Intercept)	1321.71	47.26	27.97	<.001
Linear position (early)	22.65	24.34	0.93	>.1
Boundary (IP-medial)	115.01	34.94	3.29	<.01
Place (alveolar)	428.03	29.57	14.47	<.001
Place (velar)	494.16	30.11	16.41	<.001
Gender (male)	–135.38	63.11	–2.15	.08
Linear position (early) x Boundary (IP-medial)	–70.05	34.61	–2.02	<.05
Boundary (IP-medial) x Place (alveolar)	–19.37	41.97	–0.46	>.1
Boundary (IP-medial) x Place (velar)	–134.59	43.31	–3.11	<.01

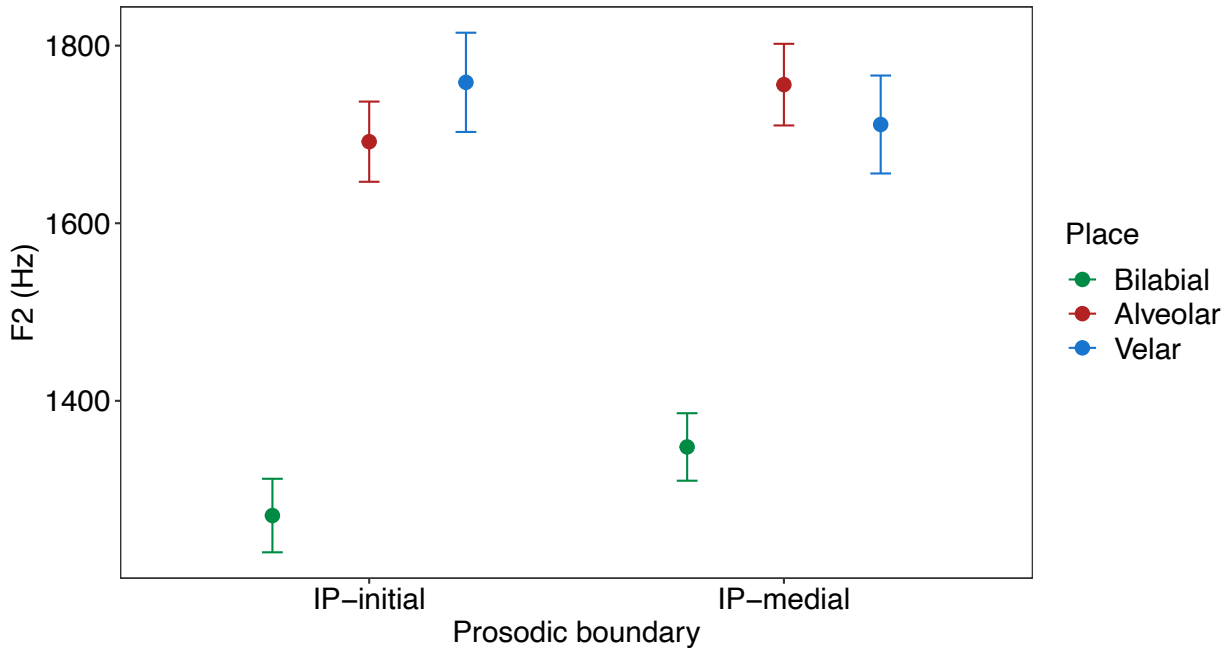


Figure 20: F2 at the onset of the following vowel [ʌ] for stops at bilabial, alveolar, and velar places of articulation in IP-initial and IP-medial positions.

As shown in Figure 20, F2 at the onset of the vowel [ʌ] for bilabial and alveolar stops was lower in IP-initial position than in IP-medial position (bilabial: β = –80.0, *SE* = 30.6, *t*[483] = –2.62, *p*<.01; alveolar: β = –60.6, *SE* = 28.9, *t*[483] = –2.10, *p*<.05), whereas F2 for velar stops

was higher in IP-initial position than in IP-medial position (β = 54.6, *SE* = 30.8, *t*[484] = 1.77, *p*=.07). As a result, as Table 10 shows, in the examination of the differences in F2 between places of articulation, a greater difference in F2 between bilabial and velar places of articulation was observed in IP-initial position (β = –494.2, *SE* = 30.1, *t*[483] = –16.41, *p*<.001) than in IP-medial position (β = –359.6, *SE* = 31.1, *t*[483] = –11.56, *p*<.001). The releveling of place of articulation also revealed that the difference in F2 between alveolar and velar places of articulation was greater in IP-initial position (β = –66.1, *SE* = 29.7, *t*[483] = –2.22, *p*=.06) than in IP-medial position (β = 49.1, *SE* = 29.8, *t*[483] = 1.65, *p*<.1) (Figure 20).

Now, we move on to the interaction between prosodic boundary and linear position to see whether the effect of prosodic boundary is confounded with the effect of linear position. Figure 21 shows F2 at the onset of the following vowel [ʌ] for stops in IP-initial and IP-medial positions in early and late positions.

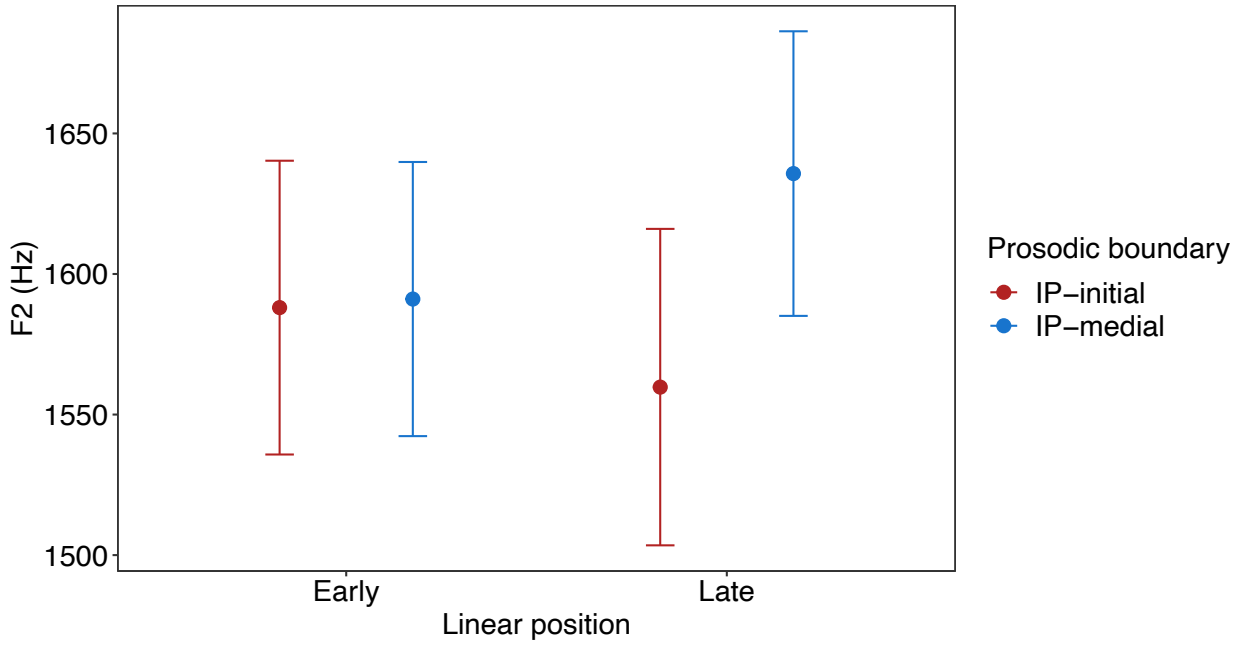


Figure 21: F2 at the onset of the following vowel [ʌ] in IP-initial and IP-medial positions in early and late positions.

As shown in Figure 21, the interaction between prosodic boundary and linear position was due in part to the greater difference in F2 at the onset of the vowel [ʌ] between IP-initial and IP-medial positions in the late position, with F2 being lower in IP-initial position than in IP-medial position (β = –63.69, SE = 24.5, t[484] = –2.60, p<.01), compared to the early position, where F2 was not different across prosodic boundary conditions (β = 6.36, SE = 24.7, t[483] = 0.26, p>.1). In other words, the effect of prosodic boundary was greater in the late position than in the early position, suggesting that the effect of prosodic boundary on F2 can be influenced by linear position. Nonetheless, this result does not suggest that the effect of prosodic boundary is in fact an effect of linear position. If the effect of prosodic boundary was purely due to the absolute location of the target words (i.e., located earlier vs. later in a sentence), its interaction with other predictor variables should also be explained by linear position. However, since place of articulation interacted only with prosodic boundary, the enhancement of differences in F2 between places of articulation was dependent on prosodic boundary but not on linear position.

Therefore, the difference between bilabial, alveolar, and velar places of articulation increased in IP-initial position compared to IP-medial position, at least for F2 at the onset of the following vowel [ʌ]. These results suggest that there was an enhancement of phonological distinction between places of articulation in IP-initial position compared to IP-medial position in both interactive and read speech, supporting paradigmatic contrast enhancement.

Nasals

Nasal Duration

The analysis of the raw nasal duration is reported first. The big model that was backward fit on the raw nasal duration measurements revealed that an interaction between prosodic

boundary and speech style ($F(1, 444) = 17.08$, $p<.001$), an interaction between prosodic boundary and linear position ($F(1, 443) = 4.92$, $p<.05$), and a main effect of place of articulation ($F(1, 443) = 8.30$, $p<.01$) contributed to the model. Table 11 summarizes the linear mixed-effects model with the best fit on the raw nasal duration, with early position (linear position), interactive speech (speech style), IP-initial position (prosodic boundary), and bilabial (place of articulation) as baseline.

Table 11: Summary of linear mixed-effects model with best fit on raw nasal duration.

	Estimate	**Standard Error**	***t***	***p***
(Intercept)	92.57	4.25	21.79	<.001
Linear position (late)	–11.22	2.84	–3.96	<.001
Speech style (read)	–23.51	2.86	–8.23	<.001
Boundary (IP-medial)	–26.29	3.68	–7.15	<.001
Place (alveolar)	–5.63	1.96	–2.87	<.01
Vowel ([ʌ])	3.21	5.30	0.61	>.1
Linear position (late) x Boundary (IP-medial)	8.69	3.92	2.22	<.05
Speech style (read) x Boundary (IP-medial)	16.29	3.95	4.12	<.001

We begin with the interaction between prosodic boundary and speech style, which is of one of our interests in the present study. Figure 22 shows the raw nasal duration in IP-initial and IP-medial positions in interactive and read speech.

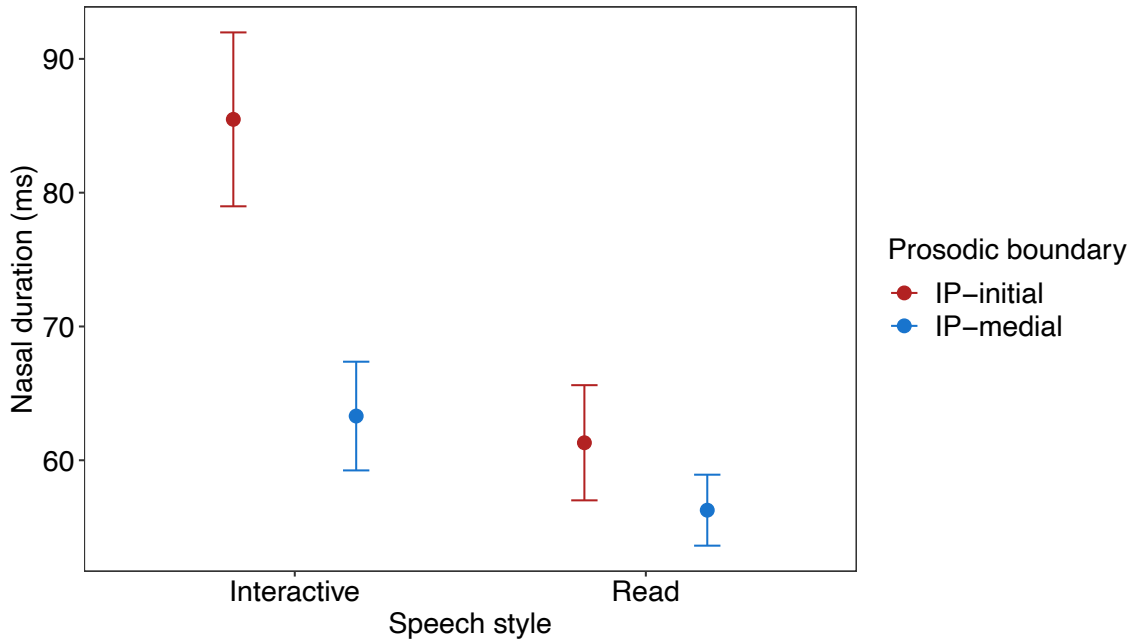


Figure 22: Raw nasal duration in IP-initial and IP-medial positions in interactive and read speech.

As shown in Figure 22, the raw nasal duration was longer in IP-initial position than in IP-medial position in both interactive speech (β = 21.95, SE = 3.00, t[447] = 7.32, p<.001) and read speech (β = 5.66, SE = 2.63, t[443] = 2.15, p<.05). However, the difference in the raw nasal duration between IP-initial and IP-medial positions was significantly greater in interactive speech than in read speech (Table 11). In sum, the raw nasal duration increased in IP-initial position compared to IP-medial position, and the increase was greater in interactive speech than in read speech.

In addition, the interaction between prosodic boundary and linear position that was found in the model is examined again to evaluate whether the effect of prosodic boundary can be accounted for by an effect of linear position. Figure 23 presents the raw nasal duration in IP-initial and IP-medial positions in early and late positions.

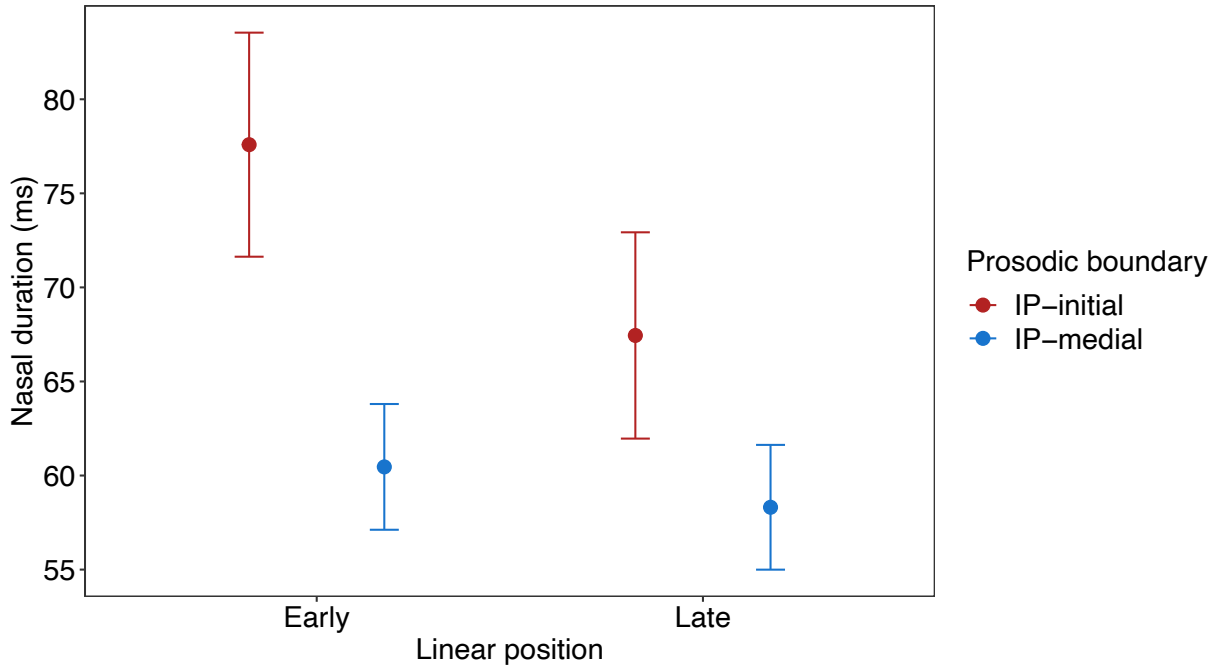


Figure 23: Raw nasal duration in IP-initial and IP-medial positions in interactive and read speech.

As Figure 23 shows, the raw nasal duration was higher in IP-initial position than in IP-medial position in both the early and late positions (β = 18.15, SE = 2.90, t[446] = 6.26, p<.001; β = 9.46, SE = 2.72, t[444] = 3.48, p<.001, respectively). Nevertheless, the difference in the raw nasal duration between IP-initial and IP-medial positions was greater in early position compared to late position (Table 11). This means that for the raw nasal duration, the effect of prosodic boundary gets weakened from early position to late position but is maintained in both positions.

Next, we report the analysis of the relative nasal duration over the word duration. As for the raw nasal duration, the backward fitting of the big model on the relative nasal duration revealed that the linear mixed-effects model with the best fit included an interaction between prosodic boundary and speech style (F(1, 443) = 9.19, p<.01), an interaction between prosodic boundary and linear position (F(1, 442) = 4.51, p<.05), and a main effect of place of articulation (F(1, 443) = 47.25, p<.001). Table 12 summarizes the linear mixed-effects model with the best

fit on the relative nasal duration, with early position (linear position), interactive speech (speech style), IP-initial position (prosodic boundary), and bilabial (place of articulation) as baseline. Firstly, the interaction between prosodic boundary and speech style is examined. Figure 24 shows the relative nasal duration over the word duration in IP-initial and IP-medial positions in interactive and read speech.

Table 12: Summary of linear mixed-effects model with best fit on relative nasal duration over the word duration.

	Estimate	**Standard Error**	***t***	***p***
(Intercept)	0.30	0.01	22.98	<.001
Linear position (late)	–0.03	0.01	–3.42	<.001
Speech style (read)	–0.04	0.01	–4.72	<.001
Boundary (IP-medial)	–0.05	0.01	–5.07	<.001
Place (alveolar)	–0.04	0.01	–6.86	<.001
Vowel ([ʌ])	0.02	0.02	0.91	>.1
Linear position (late) x Boundary (IP-medial)	0.02	0.01	2.12	<.05
Speech style (read) x Boundary (IP-medial)	0.03	0.01	3.01	<.01

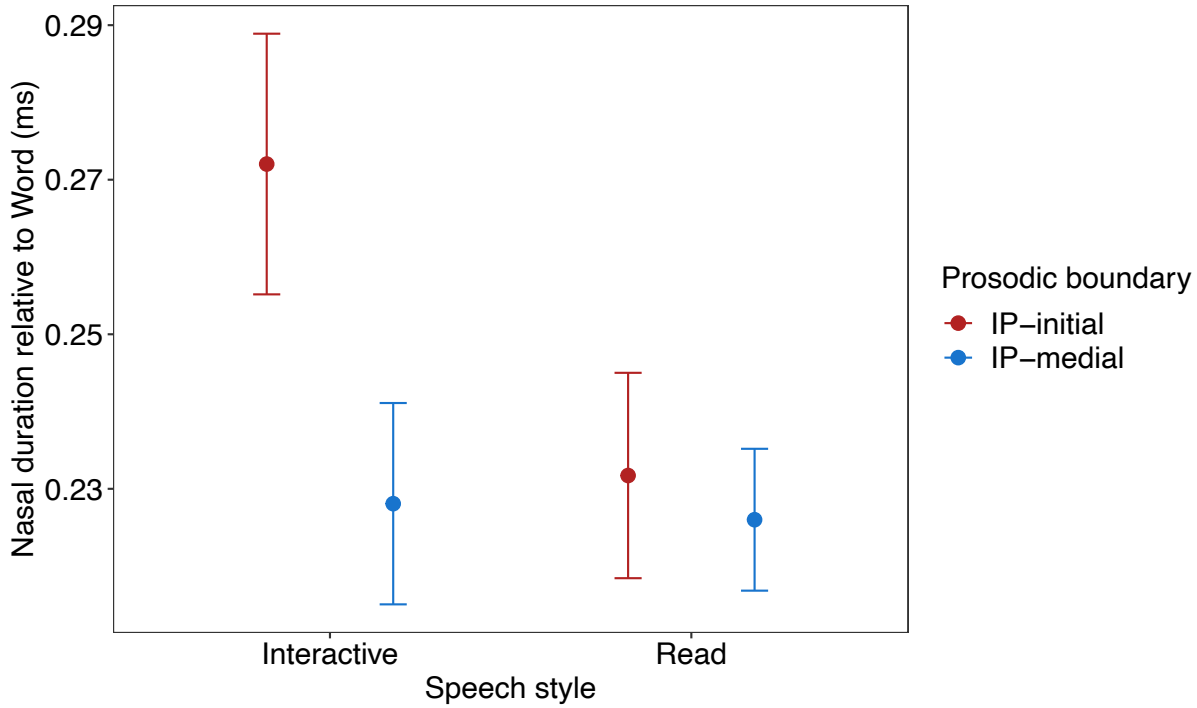


Figure 24: Relative nasal duration over the word duration in IP-initial and IP-medial positions in interactive and read speech.

As shown in Figure 24, the relative nasal duration was significantly higher in IP-initial position than in IP-medial position in interactive speech (β = 0.04, SE = 0.01, t[446] = 4.83, p<.001), whereas the relative nasal duration in read speech did not differ between IP-initial and IP-medial positions (β = 0.01, SE = 0.01, t[442] = 0.98, p>.1). As a result, as Table 12 shows, the difference in the relative nasal duration between IP-initial and IP-medial positions was significantly greater in interactive speech than in read speech. In sum, the relative nasal duration increased in IP-initial position compared to IP-medial position, but the increase was present only in interactive speech.

The relative nasal duration showed a similar interaction between prosodic boundary and linear position as the raw nasal duration did. Figure 25 shows the relative nasal duration in the IP-initial and IP-medial positions of the early and late positions.

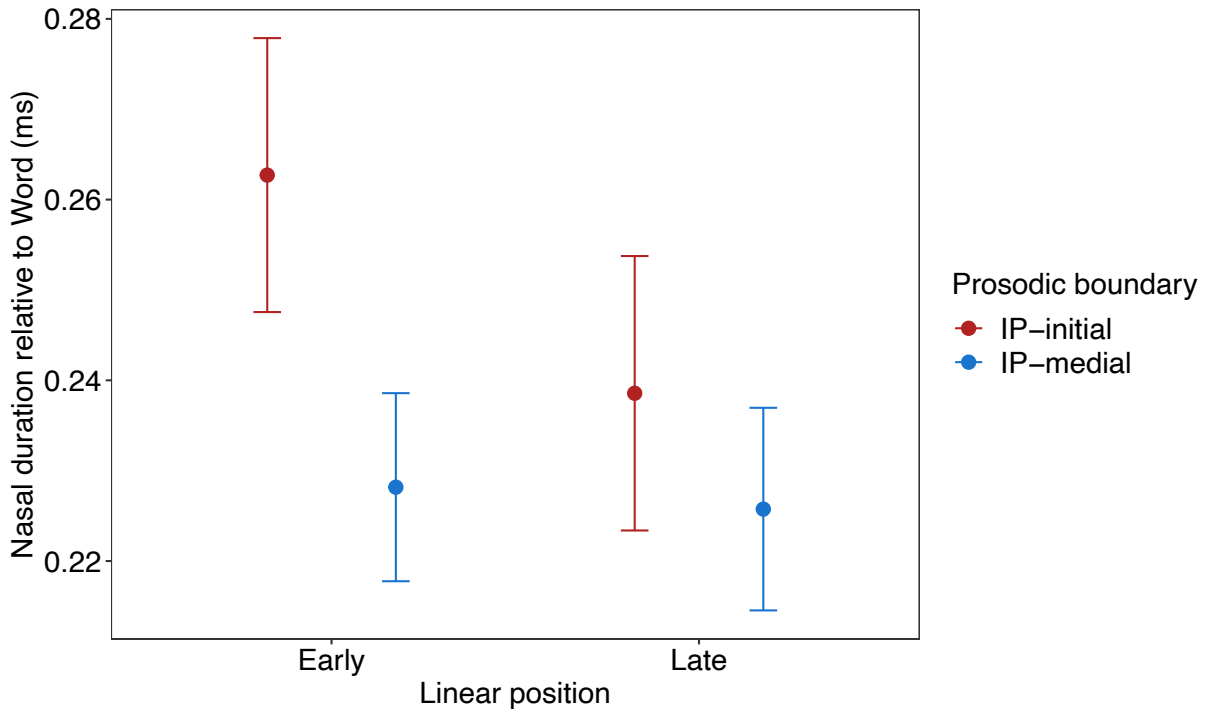


Figure 25: Relative nasal duration in IP-initial and IP-medial positions in early and late positions.

As illustrated in Figure 25, the relative nasal duration was significantly higher in IP-initial position than in IP-medial position only in the early position (β = 0.04, *SE* = 0.01, *t*[445] = 4.38, *p*<.001), not in the late position (β = 0.01, *SE* = 0.01, *t*[443] = 1.61, *p*>.1), resulting in a greater difference in the relative nasal duration between IP-initial and IP-medial positions in the early position compared to the late position (Table 12). Therefore, the relative nasal duration showed that linear position interacted with prosodic boundary such that the earlier in a sentence, the stronger the effect of prosodic boundary is.

Given that an increase in nasal duration results in an increase in the overall amount of nasality, the results showed that the nasality of nasals increased in IP-initial position compared to IP-medial position, and this effect of prosodic boundary was greater in interactive speech than in read speech. Thus, the results of the raw and relative nasal duration are in line with the paradigmatic contrast enhancement account.

F1 Bandwidth

The backward fitting of the big model on F1 bandwidth revealed that the linear mixed-effects model with the best fit included an interaction between prosodic boundary and speech style ($F(1, 444.47) = 5.073$, $p<.05$). Table 13 summarizes the linear mixed-effects model with the best fit on F1 bandwidth, with interactive speech (speech style), IP-initial position (prosodic boundary), and [ɪ] (vowel context) as baseline. Figure 26 shows the F1 bandwidth in IP-initial and IP-medial positions in interactive and read speech.

Table 13: Summary of linear mixed-effects model with best fit on F1 bandwidth for nasals.

	Estimate	**Standard Error**	***t***	***p***
(Intercept)	162.10	28.61	5.67	<.001
Speech style (read)	–25.59	17.49	–1.46	>.1
Boundary (IP-medial)	13.13	18.37	0.72	>.1
Vowel ([ʌ])	28.15	41.85	0.67	>.1
Speech style (read) x Boundary (IP-medial)	54.47	24.32	2.24	<.05

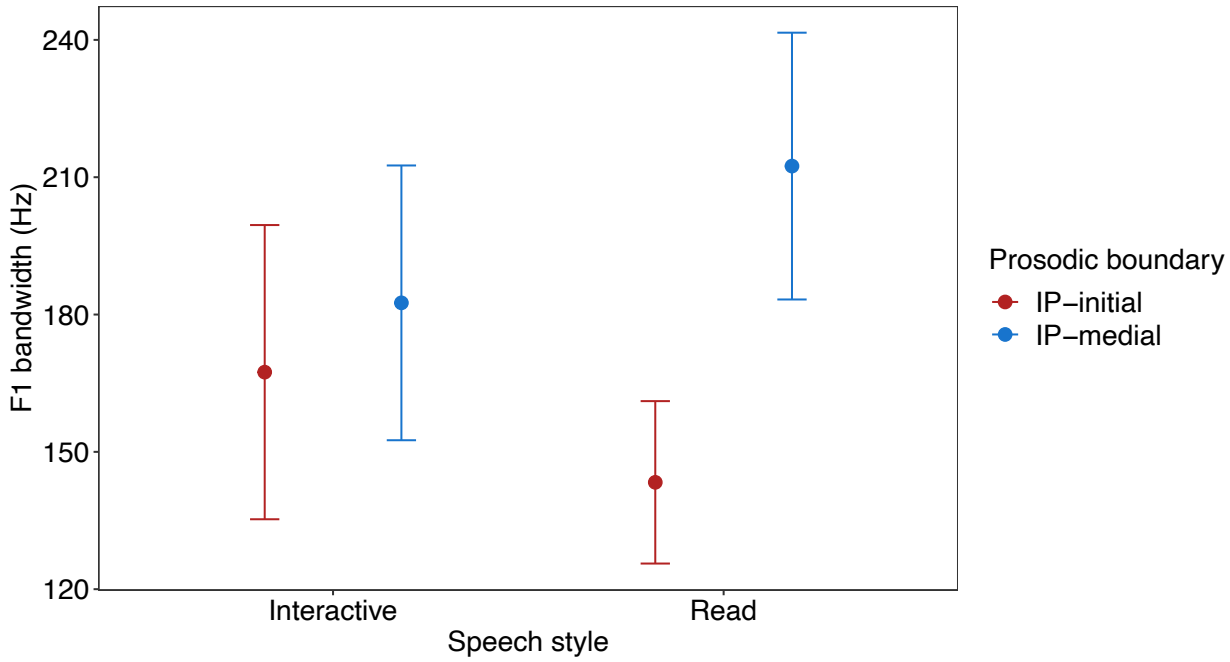


Figure 26: F1 bandwidth in IP-initial and IP-medial positions in interactive and read speech.

Figure 26 shows that the F1 bandwidth was lower in IP-initial position than in IP-medial position in both interactive and read speech. As shown in Table 13, the linear mixed-effects model with the best fit revealed that the lowering of F1 bandwidth in IP-initial position, compared to IP-medial position, was not statistically significant in interactive speech, whereas the releveling of the model revealed that this effect of prosodic boundary was statistically significant in read speech (β = 67.6, *SE* = 16.2, t[444] = 0.98, p<.001). In addition, neither the effect of linear position nor the interaction between linear position and the other predictor variables contributed to the model fit, indicating that linear position did not influence the realization of F1 bandwidth for nasals, and suggesting that the effect of prosodic boundary cannot be explained by linear position.

To summarize, F1 bandwidth was not influenced by prosodic boundary in interactive speech. On the other hand, the effect of prosodic boundary was observed in read speech such that

F1 bandwidth was lowered in IP-initial position compared to IP-medial position. Thus, nasality decreased in IP-initial position compared to IP-medial position only in read speech, suggesting that syntagmatic contrast enhancement can account for the effect of prosodic boundary on F1 bandwidth in read speech.

Max & Mean A1

We begin with the analysis of maximum A1. The backward fitting of the big model on the maximum value of A1 revealed a reduced linear mixed-effects model that included a main effect of prosodic boundary ($F(1, 443) = 5.45$, $p<.05$) and a main effect of place of articulation ($F(1, 442) = 16.12$, $p<.001$) with no interaction between them. Table 14 summarizes the results from the linear mixed-effects model with the best fit on the maximum value of A1 for nasals, with IP-initial position (prosodic boundary), bilabial (place of articulation), and [ɪ] (vowel context) as baseline. The main effect of prosodic boundary, which is our main interest, is reported. Figure 27 presents the maximum value of A1 in IP-initial and IP-medial positions.

Table 14: Summary of linear mixed-effects model with best fit on maximum value of A1 for nasals.

	Estimate	**Standard Error**	***t***	***p***
(Intercept)	52.05	1.38	37.61	<.001
Boundary (IP-medial)	0.80	0.34	2.33	<.05
Place (alveolar)	–1.35	0.34	–4.02	<.001
Vowel ([ʌ])	–2.25	2.19	–1.03	>.1

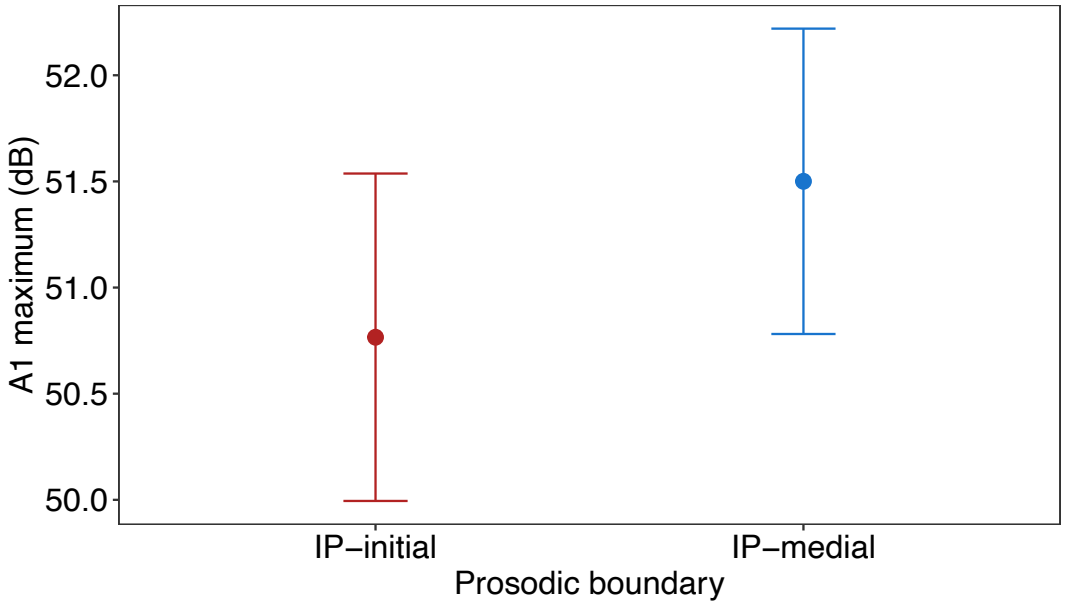


Figure 27: Maximum of A1 in IP-initial and IP-medial position.

Figure 27 shows that the maximum value of A1 was lower in IP-initial position than in IP-medial position across speech styles and places of articulation (Table 14). The lowered maximum value of A1 indicates less airflow in the nasal cavity in IP-initial position compared to IP-medial position (see the "Discussion" section for more details). Thus, in both interactive and read speech, nasality did not increase but decreased in IP-initial position compared to IP-medial position in terms of the maximum value of A1. In addition, given that there was an effect of prosodic boundary but no effect of linear position or interactions between linear position and other predictor variables for the maximum value of A1, the effect of prosodic boundary was not confounded with the effect of linear position.

Next, the influence of the predictor variables on the mean value of A1 is examined. The big model was backward fit, revealing a linear mixed-effects model with the best fit that included a main effect of speech style ($F(1, 445) = 3.85$, $p=.05$) and a main effect of place of articulation ($F(1, 446) = 4.84$, $p<.05$). Table 15 summarizes the results from the linear mixed-effects model

with the best fit on the mean value of A1 for nasals, with interactive speech (speech style), bilabial (place of articulation), and [ɪ] (vowel context) as baseline.

Table 15: Summary of linear mixed-effects model with best fit on mean value of A1 for nasals.

	Estimate	**Standard Error**	***t***	***p***
(Intercept)	52.40	1.38	38.08	<.001
Speech style (read)	–0.61	0.36	–1.70	.09
Place (alveolar)	–1.17	0.36	–3.27	<.01
Vowel ([ʌ])	–2.61	2.19	–1.19	>.1

As shown in Table 15, no influence of prosodic boundary on the mean value of A1 was found. The linear mixed-effects model with the best fit revealed that the mean value of A1 for nasals at the bilabial place of articulation was higher than that for nasals at the alveolar place of articulation. Moreover, the mean value of A1 was higher in interactive speech than in read speech. However, as the linear mixed-effects model showed, prosodic boundary did not influence the effect of speech style or place of articulation. Finally, there was no effect of linear position found in the model either.

In conclusion, the maximum value of A1 was lower in IP-initial position than in IP-medial position across speech styles and places of articulation, suggesting a decreased nasality and increased consonantality in IP-initial position compared to IP-medial position in both interactive and read speech. Both maximum and mean values of A1 differed between nasals at bilabial and alveolar places of articulation, but no enhancement of distinctions between bilabial and alveolar places of articulation was observed in IP-initial position compared to IP-medial

position. Thus, the results suggest that the effect of prosodic boundary on the maximum value of A1 can be accounted for by syntagmatic contrast enhancement.

F2

First, the results for the F2 of nasals followed by the vowel [ɪ] are reported. The backward fitting of the big model revealed that the linear mixed-effects model with the best fit included an interaction between prosodic boundary and place of articulation ($F(1, 268) = 4.99$, $p<.05$). Table 16 summarizes the results of the linear mixed-effects model with the best fit on F2 at the onset of the vowel [ɪ] for nasals, with IP-initial position (prosodic boundary), bilabial (place of articulation), and female (gender) as baseline. Figure 28 shows F2 at the onset of the vowel [ɪ] for bilabial and alveolar nasals in IP-initial and IP-medial positions.

Table 16: Summary of linear mixed-effects model with best fit on F2 at the onset of the following vowel [ɪ] for nasals.

	Estimate	**Standard Error**	***t***	***p***
(Intercept)	2024.84	49.12	41.22	<.001
Boundary (IP-medial)	–73.26	30.29	–2.42	<.05
Place (alveolar)	194.86	30.90	6.31	<.001
Gender (male)	–110.43	106.00	–1.04	>.1
Boundary (IP-medial) x Place (alveolar)	94.24	42.11	2.24	<.05

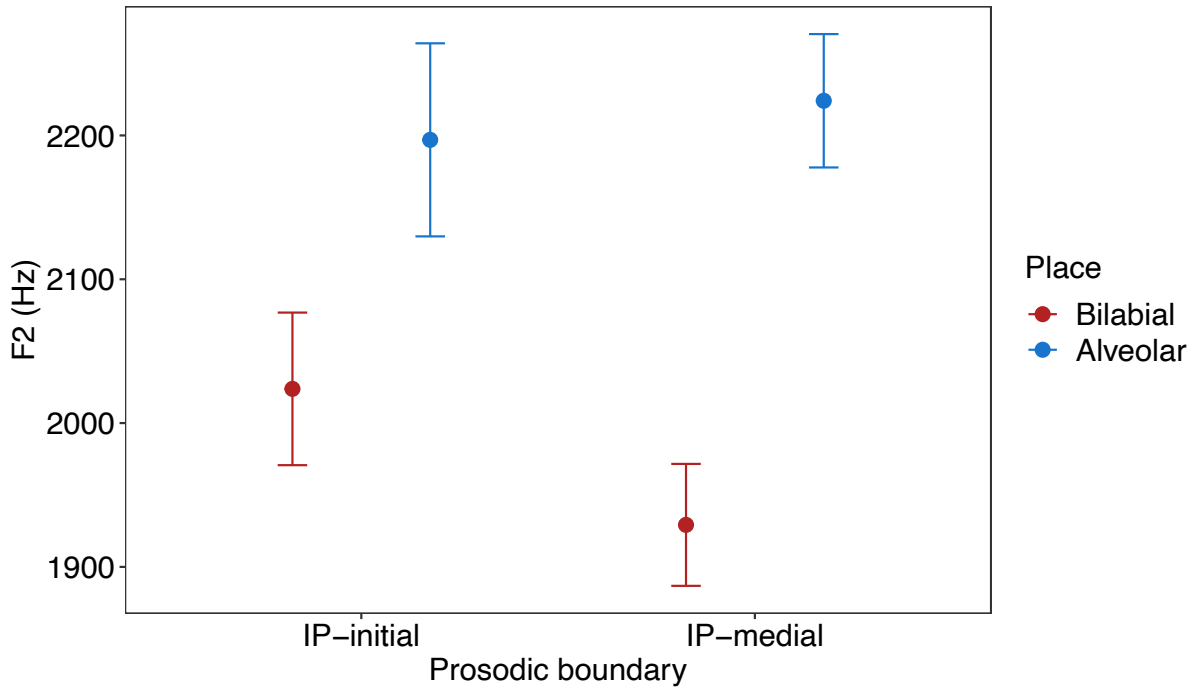


Figure 28: F2 at the onset of the following vowel [ɪ] for nasals at bilabial and alveolar places of articulation in IP-initial and IP-medial positions.

Figure 28 shows that F2 at the onset of the following vowel [ɪ] was lower for bilabial nasals than for alveolar nasals in both IP-initial and IP-medial positions (IP-initial position: β = –195, SE = 30.9, t[269] = –6.30, p<.001; IP-medial position: β = –289, SE = 28.7, t[268] = –10.06, p<.001), but this difference in F2 between bilabial and alveolar nasals was smaller in IP-initial position than in IP-medial position (Table 16). The lesser distinction between bilabial and alveolar nasals in IP-initial position compared to IP-medial position was driven by bilabial nasals, such that F2 at the onset of the vowel [ɪ] for bilabial nasals was significantly higher in IP-initial position than in IP-medial position (Table 16) whereas, with the releveling in the model, F2 at the onset of the following vowel [ɪ] for alveolar nasals was constant across IP-initial and IP-medial positions (β = 21.0, SE = 30.0, t[269] = 0.70, p>.1).

Now, we examine F2 following the nasal at the onset of the vowel [ʌ]. The backward fitting of the big model revealed that the linear mixed-effects model with the best fit included a main effect of place of articulation ($F(1, 168) = 453.16$, $p<.001$). Table 17 summarizes the results of F2 at the onset of the following vowel [ʌ], with bilabial (place of articulation) and female (gender) as baseline.

Table 17: Summary of linear mixed-effects model with best fit on F2 at the onset of vowel [ʌ] for nasals.

	Estimate	**Standard Error**	***t***	***p***
(Intercept)	1351.70	60.18	22.46	<.001
Place (alveolar)	475.71	22.37	21.26	<.001
Gender (male)	–202.72	90.00	–2.25	.074

As shown in Table 17, the linear mixed-effects model with the best fit on F2 at the onset of [ʌ] did not include any interactions between prosodic boundary and speech style or place of articulation. F2 at the onset of [ʌ] was lower for bilabial nasals than for alveolar nasals (Table 17). However, more importantly, this difference in F2 between bilabial and alveolar nasals did not change by prosodic boundary.

Taken together, the effect of prosodic boundary was found for F2 at the onset of the vowel [ɪ] only when preceded by bilabial nasals, resulting in less distinction between bilabial and alveolar nasals in IP-initial position compared to IP-medial position. The results did not show an enhancement of the difference between places of articulation in IP-initial position compared to IP-medial position. Thus, the results suggest that prosodic strengthening was not observed on F2 for nasals and thereby the patterns cannot be accounted for by paradigmatic contrast enhancement.

Summary of the Results

Table 18 summarizes the results for all acoustic parameters that were examined in the present study. The table indicates whether there was prosodic strengthening that was marked by the enhancement of an acoustic parameter in IP-initial position compared to IP-medial position. More importantly, the observed patterns for each acoustic parameter were assessed for whether the patterns could be accounted for by syntagmatic contrast enhancement (i.e., CV enhancement) or paradigmatic contrast enhancement (i.e., phonological contrast enhancement). Note that the enhancement of nasality was considered as paradigmatic contrast enhancement by comparing it indirectly to oral consonants.

Table 18: Summary of the results in relation to prosodic strengthening and linguistic accounts. N.S. indicates that the difference between IP-initial and IP-medial positions was not statistically significant, thus no prosodic strengthening was found.

Manner	**Acoustic parameter**	**Speech style**	**Prosodic strengthening (IP-initial compared to IP-medial)**	**Supporting accounts**
Stop	Raw VOT (raw & relative)	Interactive	Voiceless: longer Voiced: shorter	Paradigmatic (voicing & place)
		Read	Longer	Syntagmatic
	RMS burst amplitude	Interactive	N.S.	–
		Read	Lower	–
	Spectral peak of burst	Interactive	Higher (only for alveolar)	Paradigmatic (place)
		Read	N.S.	–
	F2	Interactive	Bilabial: lower Alveolar: lower Velar: higher	Paradigmatic (place)
		Read	Bilabial: lower Alveolar: lower Velar: higher	Paradigmatic (place)

Nasal	Nasal duration (raw & relative)	Interactive	Longer	Paradigmatic (nasality)
		Read	Longer	Paradigmatic (nasality)
	F1 bandwidth	Interactive	N.S.	–
		Read	Lower	Syntagmatic
	Max A1	Interactive	Lower	Syntagmatic
		Read	Lower	Syntagmatic
	Mean A1	Interactive	N.S.	–
		Read	N.S.	–
	F2	Interactive	Higher (only for bilabial)	–
		Read	Higher (only for bilabial)	–

As shown in Table 18, broadly speaking, plosives tend to show different patterns of prosodic strengthening in relation to linguistic function from nasals. Accordingly, the detailed directionality of prosodic strengthening for each acoustic correlate and how it can be explained by two accounts (i.e., paradigmatic vs. syntagmatic contrast enhancement) are discussed for plosives first in the following "Discussion" section, and next for nasals and how nasals pattern differently from plosives. We then discuss the patterns of prosodic strengthening in different speech styles and their implications.

Chapter 4: Discussion

Prosodic Strengthening Driven by Prosodic Boundary and Its Function: Syntagmatic vs. Paradigmatic Contrast Enhancement

The general pattern of how prosodic boundary influences the phonetic implementation of English plosives and nasals is discussed first. As expected, the phonetic implementation of English plosives and nasals was found to differ as a function of the level of prosodic boundary (i.e., higher/larger vs. lower/smaller prosodic boundaries in terms of prosodic hierarchy) such that English plosives and nasals were realized with more exaggerated articulation that is reflected in acoustic parameters in IP-initial position compared to IP-medial position. These findings are in general consistent with the previous studies in that some form of prosodic strengthening was observed in IP-initial position compared to IP-medial position, and it was not always observed. However, unlike most of the previous studies that suggested syntagmatic contrast enhancement to explain boundary-induced prosodic strengthening in English, the present study discovered that boundary-induced prosodic strengthening in English can be accounted for by paradigmatic contrast enhancement, as well as syntagmatic contrast enhancement, depending on the segment of interest and its acoustic correlates and depending on speech style. Durational acoustic correlates of plosives and nasals showed a more consistent effect of prosodic boundary than amplitudinal or spectral acoustic correlates. In addition, the present study showed that prosodic strengthening patterns differently in relation to its linguistic function depending on whether speakers read given sentences or produced sentences when interacting with a listener. Lastly, simple locational differences (located earlier vs. later in a sentence) can influence prosodic strengthening on segmental duration, but prosodic strengthening cannot be accounted for by simple locational differences because the segmental duration still varies as a function of the

strength of prosodic boundaries in different locations. A detailed discussion of the results for plosives and nasals is provided below.

Plosives

VOT shows the clearest differences in the patterns of prosodic strengthening between read and interactive speech. The results from read speech agree with those from previous studies where prosodic strengthening was investigated in read speech such that both voiceless and voiced plosives become less sonorous, suggesting syntagmatic contrast enhancement. The present study provides convincing evidence for syntagmatic contrast enhancement by showing that, in read speech, voiced plosives are produced with less prevoicing (i.e., shorter negative VOT) in IP-initial position than in IP-medial position, whereas voiceless plosives are produced with a longer voicing lag (i.e., longer positive VOT) in IP-initial position than in IP-medial position. Moreover, there was no enhancement of the distinction between voiceless and voiced stops in terms of VOT (Figure 2 (a)), refuting the possibility of co-occurring syntagmatic and paradigmatic contrast enhancements (Figure 2 (c)).

Unlike read speech, the patterns of prosodic strengthening in interactive speech did not replicate those in previous studies. Voiceless plosives lengthened voicing lag (i.e., longer positive VOT) whereas voiced plosives lengthened prevoicing (i.e., longer negative VOT). These patterns of prosodic strengthening found in interactive speech suggest paradigmatic contrast enhancement potentially via feature enhancement such that voiceless plosives enhanced the phonetic feature {voiceless aspirated} whereas voiced plosives enhanced the phonetic feature {voiced} in IP-initial position compared to IP-medial position, resulting in more distinction between voiceless and voiced plosives in IP-initial position than in IP-medial position. Feature

enhancement has been suggested to be closely related to paradigmatic contrast enhancement (e.g., Cho & Jun, 2000; de Jong, 1995, 2004; Georgeton & Fourgeron, 2014; Hsu & Jun, 1998). However, one might be puzzled by the fact that the phonetic feature {voiced} was enhanced for voiced plosives in American English where voiced plosives have shown a short-lag VOT word-initially (e.g., Flege, 1982; Lisker & Abramson, 1964; Smith, 1978; Westbury, 1979) and thereby they can be specified with the phonetic feature {voiceless unaspirated}, in contrast to voiceless plosives that can be specified with the phonetic feature {voiceless aspirated} (e.g., Keating, 1984, 1990). In the present study, speakers produced word-initial voiced plosives with prevoicing for 80% of the analyzed data.[10] In fact, recently, some studies have reported that the production of voiced stops in American English varies dialectally, with speakers from Southern United States being more likely to produce voiced stops with prevoicing (e.g., Herd, 2020; Jacewicz, Fox, & Lyle, 2009). Thus, it is reasonable to assume that the phonetic feature {voiced} was enhanced for voiced plosives in IP-initial position compared to IP-medial position to be more distinct from voiceless plosives in the present study. However, it is important to note that feature enhancement does not always work with paradigmatic contrast enhancement. For example, Cho and McQueen (2005) showed that Dutch /t/ was realized with a shorter VOT, enhancing the phonetic feature {voiceless unaspirated}, whereas /d/ was realized with longer prevoicing, enhancing the phonetic feature {voiced}, in domain-initial positions compared to domain-medial positions. As a result, they did not find the enhancement of the contrast between /t/ and /d/ in their study. The exact nature of the relation between feature enhancement and paradigmatic contrast enhancement should be further explored.

[10] Unfortunately, the regional variation was not well controlled in the present study. Information about where in the US the speakers were born and raised was not collected.

In interactive speech, the present study also found the enhancement of differences in VOT between places of articulation in IP-initial position, compared to IP-medial position in which no distinction between places of articulation was made, again suggesting paradigmatic contrast enhancement. For voiceless plosives in interactive speech, the VOT for the bilabial place of articulation was enhanced to be shorter than that for the alveolar or velar places of articulation, making voiceless plosives at bilabial, alveolar, and velar places articulation more distinct from each other in IP-initial position than in IP-medial position despite the overall increase of VOT across places of articulation in IP-initial position compared to IP-medial position. These patterns of distinction in IP-initial position accords with the patterns found in Nearey and Rochet (1994), where VOT for word-initial voiceless plosives (and voiced plosives with a short-lag VOT) was found to be longest for the velar place of articulation, intermediate for the alveolar place of articulation, and shortest for the bilabial place of articulation (e.g., Lisker & Abramson, 1964; Nearey & Rochet, 1994). For voiced plosives, the VOT for the alveolar place of articulation was enhanced to have less prevoicing, and thereby became more distinctive from the bilabial and velar places of articulation in IP-initial position compared to IP-medial position. Based on the VOT distribution patterns in previous studies where prevoicing was observed for word-initial voiced plosives, voiced plosives at bilabial and alveolar places of articulation show a similar range of prevoicing, all of which are longer in prevoicing than (word-initial) voiced stops at the velar place of articulation (e.g., Hunnicutt & Morris, 2016; Lisker & Abramson, 1964). In the present study, speakers might have tried to further differentiate voiced plosives at bilabial and alveolar places of articulation in IP-initial position than in IP-medial position in terms of VOT by reducing the prevoicing of voiced plosives at the alveolar place of articulation. Thus, based on the VOT data, the present study shows that prosodic strengthening driven by prosodic

boundary has the function of enhancing paradigmatic contrast of place of articulation, as well as voicing, in interactive speech.

Prosodic strengthening for the RMS amplitude of the burst was also observed in IP-initial position compared to IP-medial position, and varied depending on voicing or speech style. However, the patterns of prosodic strengthening for the RMS amplitude of the burst were difficult to comprehend in terms of linguistic function. Articulatorily, voiceless plosives are formed by a longer occlusion of the articulators that leads to more oral pressure built up behind the closure than voiced plosives; as a result, voiceless plosives are produced with a higher intensity burst (i.e., higher energy of the turbulence noise in the burst release) than voiced plosives (e.g., Halle, Hughes, & Radley, 1957; Stevens, 1998). Thus, if consonantality for both voiceless and voiced plosives is enhanced by strengthening/lengthening the occlusion of the articulators[11] in IP-initial position compared to IP-medial position, the RMS amplitude of the burst is expected to be increased in IP-initial position compared to IP-medial position. However, although there was a lowering of the RMS amplitude of the burst in IP-initial position, compared to IP-medial position, particularly in read speech, the present study shows that the RMS amplitude of the burst did not increase in IP-initial position compared to IP-medial position for both voiceless and voiced plosives, resulting in no increase of their consonantality in IP-initial position compared to IP-medial position. Hence, these prosodic strengthening patterns do not suggest syntagmatic contrast enhancement. In addition, in both read and interactive speech, the lowering of the RMS amplitude of the burst was observed in IP-initial position compared to IP-medial position only for voiced plosives, not for voiceless plosives. As a result, it did not lead to

[11] Closure duration can indicate the length of the occlusion of articulators before the release burst. However, closure duration was not examined because it was difficult to tease apart pause and closure duration at an IP boundary (i.e., IP-initial position).

a greater distinction between voiceless and voiced plosives, but rather to a smaller distinction between them in IP-initial position than in IP-medial position. These patterns do not support paradigmatic contrast enhancement, either. Thus, neither syntagmatic nor paradigmatic contrast enhancement can explain the limited patterns of prosodic strengthening for the RMS amplitude of the burst.

One of the issues that require some discussion is the fact that the realization of the RMS amplitude of the burst for voiceless and voiced plosives (regardless of prosodic boundary) in the present study showed the opposite directionality from the predicted patterns above. Figure 29 shows the first syllable of *ditsy* in IP-initial position in interactive speech. We suspect that the higher RMS amplitude of the burst for voiced plosives than voiceless plosives might be driven by the prevoicing that was often carried well into the burst and noise part of voiced plosives in the present study (Figure 29).

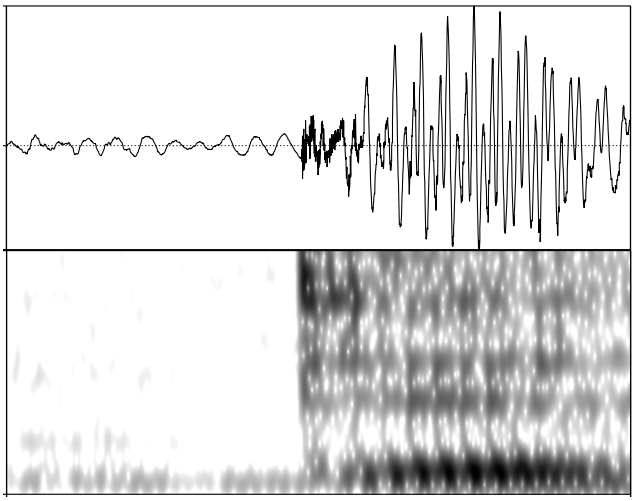


Figure 29: Waveform and spectrogram of the initial syllable in a target word *Ditsy* in IP-initial position in interactive speech.

Thus, as far as the patterns of prosodic strengthening for voiced plosives are concerned, the lowering of the RMS amplitude of the burst in IP-initial position compared to IP-medial

position might be the result of reducing prevoicing as reflected in the VOT data in read speech, possibly suggesting syntagmatic contrast enhancement. However, this explanation is still insufficient to argue for syntagmatic contrast enhancement in that the lowering of RMS amplitude of the burst for voiced plosives was found across speech styles but the reduction of prevoicing for voiced plosives was found only in read speech. Recall that, in interactive speech, the prevoicing for voiced plosives increased. In addition, the RMS amplitude of the burst for voiceless plosives did not increase but remained constant across IP-initial and IP-medial positions. Nevertheless, since there have been some studies that also obtained similar findings (i.e., that the RMS amplitude of the burst for voiceless plosives was lower than that for voiced plosives, e.g., Cole et al., 2007, and that it was lower in a larger prosodic domain than a smaller prosodic domain, e.g., Cho & Keating, 2009; Kuzla & Ernestus, 2011), more investigation is needed to find out what the RMS amplitude of the burst reflects in voiceless and voiced plosives and whether and how it is used to distinguish voiceless and voiced plosives in different prosodic positions.

An enhancement of the distinction of place of articulation was observed in the spectral peak of the burst and F2 at the onset of the following vowel in both read and interactive speech, suggesting paradigmatic contrast enhancement. These results should be interpreted with caution because prosodic strengthening was contingent on the vowel and speech style context. Nevertheless, when there was prosodic strengthening, the directionality of the enhancement reflects the general patterns of the phonetic realization of plosives observed in previous studies (e.g., Blumstein & Stevens, 1980; Dorman, Studdert-Kennedy, & Raphael, 1977; Edwards, 1981; Fant, 1973; Fischer-Jørgensen, 1954; Halle, Hughes, & Radley, 1957; Keating, Byrd, Flemming, & Todaka, 1994; Keating & Lahiri, 1993; Repp & Lin, 1988; Smits, 1996; Winitz,

Scheib, & Reeds, 1972; Zue, 1976). More specifically, only in interactive speech, the spectral peak of the burst for plosives at bilabial, alveolar, and velar places of articulation in the [ɪ] vowel context was enhanced such that that for the alveolar place increased the spectral peak of the burst in IP-initial position compared to IP-medial position. In the previous studies, the general patterns of spectral peak of the burst in different places of articulation show that the spectral peak of the burst is at low frequencies (below 1kHz) for labial place of articulation and at high frequencies (above 3kHz) for dental or alveolar plosives. Velar plosives are at mid frequencies with a wide range (between 1kHz and 4kHz) contingent on the vowel context such that the spectral peak of the burst is found in the F3 ~ F4 region in a front vowel context and in the F2 region in a non-front vowel context (e.g., Dorman, Studdert-Kennedy, & Raphael, 1977; Edwards, 1981; Fant, 1973; Fischer-Jørgensen, 1954; Halle, Hughes, & Radley, 1957; Keating, Byrd, Flemming, & Todaka, 1994; Keating & Lahiri, 1993; Repp & Lin, 1988; Winitz, Scheib, & Reeds, 1972; Zue, 1976). Based on these observations, in the [ɪ] vowel context, which is a front vowel context, the spectral peak for velar plosives is close to that for alveolar plosives. Thus, as the present study has found, heightening the spectral peak of the burst for alveolar plosives is one way to enhance the distinction between alveolar and velar plosives, and, in turn, between them and bilabial plosives in IP-initial position compared to IP-medial position.

In terms of F2 at the onset of the following vowel, the distinction between plosives at bilabial, alveolar, and velar places of articulation was enhanced in IP-initial position compared to IP-medial position in both read and interactive speech (i.e., regardless of speech style), suggesting paradigmatic contrast enhancement. The enhancement of the distinction between places of articulation was only found in the [ʌ] vowel context where all three plosives at bilabial, alveolar, and velar places of articulation were distinguished by F2 at the onset of the following

vowel in IP-initial position, whereas the distinction between plosives at alveolar and velar places of articulation disappeared in IP-medial position. Possible patterns of F2 at the onset of the following vowel [ʌ] for bilabial, alveolar, and velar plosives were inferred from the patterns of F2 in the back vowel [u] context for bilabial, alveolar, and velar plosives (i.e., [bu, du, gu]) from Blumstein and Stevens' studies (e.g., Blumstein & Stevens, 1980; Stevens & Blumstein, 1978) in which alveolar and velar voiced plosives that were followed by the back vowel [u] were realized in a similar range of F2 (approximately 1400Hz ~ 1600Hz) whereas bilabial voiced plosives were realized in a lower F2 range (lower than 1000Hz). In fact, as described above, we observed these patterns in IP-medial position, and the distinction between alveolar and velar plosives was enhanced in IP-initial position. It should be noted that, unlike other acoustic correlates for plosives in the present study (i.e., VOT, spectral peak of the burst), the enhancement of F2 at the onset of the following vowel between bilabial, alveolar, and velar plosives was found in both read and interactive speech. This might occur for two different reasons. First, unlike VOT that can potentially be used to increase both phonological contrast and consonantality (compared to vowels), F2 at the onset of the following vowel can be mainly used to increase phonological contrast, particularly distinguishing place of articulation because fronting and backing of the tongue does not contribute to the decrease or increase of sonority much. Thus, there might be less constraints on how F2 at the onset of the following vowel can be used to strengthen speech sounds in relation to linguistic function in domain-initial positions in both read and interactive speech. Second, unlike spectral peak of the burst that was measured in the initial part of the consonants (i.e., the burst part), F2 was measured at the onset of the following vowels [ɪ] and [ʌ] all of which are lax vowels. Lax vowels are more susceptible to coarticulation than tense vowel

so F2 that was measured at the onset of the following vowel might be more likely to reflect subtle influence from the preceding consonants in read speech as well as in interactive speech.

Taken together, spectral measures suggest that prosodic strengthening has the function of paradigmatic contrast enhancement also in the dimension of place of articulation. Although prosodic strengthening was not found in the [ɪ] vowel context for the spectral peak of the burst and in the [ʌ] vowel context for F2 at the onset of the following vowel, place of articulation was well distinguished by spectral peak of the burst and by F2 at the onset of the following vowel (with a main effect of place of articulation) across prosodic positions in both cases. Based on this observation, we speculate that speakers may enhance the contrast between places of articulation in IP-initial position compared to IP-medial position in a context where the contrast can be acoustically obscured, and less so in a context where the contrast is acoustically distinctive.

Nasals

Prosodic strengthening was not found on all acoustic correlates of nasals that were examined in the present study, showing that prosodic boundary influences not all relevant acoustic correlates. First, nasal duration has often been measured as an indication of nasality in previous studies, but has shown somewhat inconsistent findings in that some studies observed shorter nasal duration (e.g., Cho & Keating, 2009) and others observed longer nasal duration in domain-initial positions than in domain-medial positions (e.g., Fougeron & Keating, 1997). The present study was more in line with the latter such that nasal duration increased in IP-initial position compared to IP-medial position in both read and interactive speech. Given that lengthening nasal duration is lengthening the duration of the nasal murmur, longer nasal duration in IP-initial position than in IP-medial position can be interpreted as increasing nasality

(compared to other oral consonants), suggesting paradigmatic contrast enhancement. It is noteworthy that the lengthening of nasal duration in IP-initial position compared to IP-medial position appears to be affected by simple locational differences in the present study. Prosodic strengthening driven by prosodic boundary was greater when a nasal consonant was located earlier in a sentence compared to when it was located later. This shows that the segmental duration might be subject to prosodic strengthening driven by prosodic boundary but also to general articulatory declination over a sentence. Nevertheless, it is important that prosodic strengthening with longer nasal duration was still observed in IP-initial position compared to IP-medial position.

Next, prosodic strengthening on the F1 bandwidth was only found in read speech, with the F1 bandwidth decreasing in IP-initial position compared to IP-medial position. The F1 bandwidth that can reflect the amount of sound energy is expected to be wider for nasals compared to non-nasals due to damping in the vocal tract that has greater surface area and volume including both nasal and oral cavities (e.g., Fujimura, 1962; Johnson, 1997, 2003; Pruthi & Espy-Wilson, 2003; Styler, 2017). Based on this, a wider F1 bandwidth is expected when nasality increases in IP-initial position compared to IP-medial position. However, the present study observed a narrower (decreased) F1 bandwidth in IP-initial position compared to IP-medial position, which cannot be understood as an increase in nasality that makes nasals different from non-nasals, thus refuting paradigmatic contrast enhancement. The results are instead more in line with syntagmatic contrast enhancement in that the narrower F1 bandwidth can reflect nasals becoming more consonantal by having less surface area and volume that can absorb sound energy in IP-initial position compared to IP-medial position. This line of reasoning can be inferred from the fact that the closing of the nasal cavity in the production of nasal consonants

[m, n], leading to a reduction of surface area and volume in the vocal tract, results in the production of stop consonants [b, d]. Alternatively, a decreased F1 bandwidth might be a consequence of supralaryngeal articulatory force such that the IP-initial position induces muscular tension in the wall of the vocal tract that reduces the absorption of sound energy compared to the IP-medial position.

On the other hand, the maximum value of A1 was lowered in IP-initial position compared to IP-medial position in both read and interactive speech. A1 reflects the amplitude of F1 that decreases in nasals compared to oral sounds (i.e., vowels) due to the impeded airflow by the narrower opening of the nasal cavity where the constriction occurs (e.g., Fujimura, 1962; Johnson, 1997, 2003). Accordingly, the lowering of the maximum value of A1 indicates less airflow through the nasal cavity that makes nasals less sonorous in IP-initial position than in IP-medial position. Thus, nasals become more consonantal in terms of sonority in IP-initial position than in IP-medial position, suggesting syntagmatic contrast enhancement. This interpretation is in line with Fougeron (2001), where nasal flow for French nasals (measured using a Rothenberg split mask) decreased in IP-initial position compared to IP-medial position. One might wonder whether the lowering of the maximum value of A1 should be interpreted as increasing nasality because a low A1 is one of the well-known characteristics of nasal sounds. However, a low A1 (i.e., low amplitude of F1) characterizes nasal sounds when compared to oral vowels that show a higher A1 (i.e., high amplitude of F1), meaning that the lowering of A1 in IP-initial position compared to IP-medial position can be interpreted as increasing nasality when compared to oral vowels. When the contrast is made between nasals and oral consonants as in paradigmatic contrast, increasing nasality should mean more airflow going through the nasal cavity and, thereby, an increased A1 (i.e., increased amplitude of F1). In conclusion, in terms of the

maximum value of A1, the present study suggests that prosodic strengthening for nasals shows syntagmatic contrast enhancement in both read and interactive speech. Nevertheless, as for f1 bandwidth, articulatory effort as an alternative account can also explain the pattern of prosodic strengthening on the maximum value of A1. Given that the lowering of the velum is controlled by the relaxation of the levator palatini and the elevation of the velum is controlled by its contraction (Bell-Berti, 1993), articulatory force can reduce the relaxation of the levator-palatini and result in a smaller lowering of the velum (e.g., also Fougeron, 2001; Fujimura, 1990; Straka, 1963). Consequently, more impedance of the airflow can be reflected as the lowering of A1 in IP-initial position compared to IP-medial position (see the "Prosodic Strengthening in Different Speech Styles: Read vs. Interactive Speech" section for a discussion of the theoretical implications for why nasals did not pattern similarly to plosives in the patterns of prosodic strengthening in relation to linguistic function).

Finally, in the examination of F2 at the onset of the following vowel, the present study did not find the enhancement of place of articulation for nasals in either read or interactive speech. Nasals followed by the vowel [ɪ] at the bilabial place of articulation showed higher F2 at the onset of the following vowel in IP-initial position than in IP-medial position, whereas nasals at the alveolar place of articulation remained unchanged in IP-initial position compared to IP-medial position. As a result, the difference between bilabial and alveolar nasals was smaller in IP-initial position than in IP-medial position. For nasals at the bilabial place of articulation, there appears to be some tongue advancement that is reflected in higher F2 in IP-initial position than in IP-medial position. We suspect that it might be the reflection of a more fronted tongue for the following vowel [ɪ] in IP-initial position than in IP-medial position. Prosodic strengthening driven by prosodic boundary is mostly local to the initial consonant in a CV sequence but can

extend to the following vowel to a weaker extent (e.g., Byrd, 2000; Byrd, Krivokapić, & Lee, 2006; Cho & Keating, 2009). Moreover, in comparison with alveolar consonants, bilabial and velar consonants are more susceptible to coarticulation such that bilabial and velar consonants show a shallower slop of F2 transition from the offset of consonants to the vowel nucleus across manners of articulation (e.g., Sussman, Bessell, Dalston, & Majors, 1997; Sussman & Shore, 1996). Therefore, in the present study, specifically for bilabial nasals and not for alveolar nasals, F2 at the onset of the following vowel [ɪ] might become higher in IP-initial position than in IP-medial position because the F2 of [ɪ] became higher in IP-initial position than in IP-medial position.

Prosodic Strengthening in Different Speech Styles: Read vs. Interactive Speech

As we predicted, for English plosives, prosodic strengthening driven by prosodic boundary showed different patterns in relation to linguistic function depending on speech style. As previous studies that investigated prosodic strengthening in read English speech have shown, the present study finds that the patterns of prosodic strengthening for VOT in read speech support syntagmatic contrast enhancement. However, these patterns of prosodic strengthening supporting syntagmatic contrast enhancement are only evident from the VOT data. More importantly, they do not extend to interactive speech. In interactive speech where speakers and listeners exchange information, VOT was enhanced to distinguish the contrasts of voicing and place of articulation, showing paradigmatic contrast enhancement. Spectral acoustic correlates for plosives show the enhancement of the contrast between places of articulation in IP-initial position compared to IP-medial position, supporting paradigmatic contrast enhancement in both read and interactive speech. Overall, for plosives, when prosodic strengthening patterns to

enhance the contrast between neighboring consonants and vowels, it is observed in read speech. Interactive speech shows paradigmatic contrast enhancement for voicing and/or place of articulation depending on the phonological distinction that each acoustic correlate can make. These results from plosives might be reflecting the fact that read speech involves reading a given script instead of conveying meaningful information to a listener, whereas interactive speech involves interacting with and conveying meaningful information to a listener. Pauses occur less often in read speech because speakers do not have to plan speech. Speakers might just mark a prosodic juncture (i.e., where a prosodic boundary is) such as an IP boundary by enhancing the contrast between neighboring consonants and vowels without considering a listener. As a result, read speech shows syntagmatic contrast enhancement in IP-initial position compared to IP-medial position. On the other hand, without a prepared script, interactive speech needs more speech planning, resulting in more occurrence of pauses, which might already have the function of marking a prosodic juncture. In addition, the important goal of interactive speech is to deliver information. Thus, speakers might focus more on helping listeners access lexical entries without the potential lexical confusion driven by phonetic/phonemic similarities between word-initial sounds by enhancing phonological contrasts between sounds. As a result, interactive speech shows paradigmatic contrast enhancement in IP-initial position compared to IP-medial position.

However, when nasals are considered, they show different patterns from plosives regarding the linguistic function of prosodic strengthening and speech style. When there is prosodic strengthening, the acoustic correlates of nasals, except for segmental duration, show syntagmatic contrast enhancement in both read and interactive speech. A potential explanation might be found in the difference between English plosives and nasals in terms of how similar they are phonetically and phonologically to speech sounds that they can be compared with.

English plosives /p, t, k, b, d, g/ are phonetically and phonologically more similar to each other than nasals are to other consonants. In the present study, paradigmatic contrast enhancement was tested by comparing voicing and place of articulation within plosives. Thus, enhancing the phonological contrast of voicing and place of articulation in domain-initial position compared to domain-medial position might be useful for listeners in lexical access in interactive speech. However, with the opening of the nasal cavity, English nasals differ enough from non-nasal consonants. In English, unlike plosives that can be phonetically and phonologically close to another plosive enough to be perceptually confused when their acoustic correlates are adjusted, nasals are still distinctive from oral consonants in terms of their nasality as long as the airflow goes through the open nasal cavity. In this case, speakers do not necessarily have to enhance the contrast between nasals and non-nasals by enhancing its nasality to make them more distinct from non-nasals for listeners in interactive speech. Thus, nasals are mostly realized with prosodic strengthening that enhances the CV contrast in IP-initial position compared to IP-medial position.

As stated above, one exception was segmental duration of nasals that appears to show paradigmatic contrast enhancement if we consider longer nasal duration increased nasality. It raises some issue to be addressed. Segmental duration of nasals is tricky to interpret in relation to paradigmatic contrast enhancement because it does not directly characterize nasals compared to non-nasal consonants unlike VOT (the durational acoustic correlate for plosives in the present study) that is one of the characteristics of plosives. Moreover, recall that Fougeron and Keating (1997) suggests that segmental lengthening at a prosodic juncture (including pre- and post-boundary positions) does not necessarily accompany articulatory strengthening in English, implying that a different mechanism is behind the lengthening of segmental duration than the one

behind articulatory strengthening. Thus, alternatively, segmental duration at a prosodic juncture might be influenced by a different mechanism such as the influence of π-gesture rather than paradigmatic contrast enhancement. Byrd and colleagues have proposed that segments at a prosodic juncture are under the influence of π-gesture, the abstract prosodic gesture, that controls the temporal domain and lengthens the segmental duration (e.g., Byrd, 2000; Byrd, Kaun, Narayanan, & Saltzman, 2000; Byrd, Krivokapić, & Lee, 2006; Byrd & Saltzman, 1998; Byrd & Saltzman, 2003; Saltzman, 1995). However, the present study does not have enough evidence to generalize the alternative explanation. More investigation should follow on this issue in the future.

Taken together, the findings in the present study are not consistent with those of previous studies in terms of how plosives and nasals are realized in a prosodic strengthening environment regarding its linguistic function. Recall that previous studies propose that the patterns of prosodic strengthening in relation to its linguistic function in English may vary depending on the source of prosodic strengthening, that is whether the strengthening is driven by prosodic boundary or prominence. Unlike those previous studies, the present study suggests that the linguistic function of prosodic strengthening may depend on speech style (read speech vs. interactive speech) and whether speech sounds of interest are phonetically and phonologically different enough from speech sounds that they can potentially be compared with.

The present study is not without limitations. One of the limitations is that the interactive speech in the present study was not completely spontaneous, but partly scripted by presenting scenes that depicted the characters and action to be described and by presenting the written names of the animal characters. It is difficult to elicit interactive speech without any script when the goal is to elicit specific prosodic boundaries while controlling for other potential confounding

factors such as the naturalness of sentences and the location of the target word in relation to the prosodic structure of the utterance. The prosodic structure of sentences in interactive speech was intended to be similar to that in read speech so that the patterns of prosodic strengthening in these different speech styles could be compared. In addition, the task should not be the one where speakers had to memorize the names of characters, which would influence speakers' performance in the interaction. Different ways to elicit prosodic boundaries in more natural interactive speech without any script involved should be explored and tested.

Another limitation in the present study was the fact that the strengthening patterns for plosives and nasals could not be directly compared with those for the following vowels in the evaluation of syntagmatic contrast enhancement even though the account is about contrasting neighboring consonants and vowels in terms of their sonority. This comparison was possible in some previous studies such as Fougeron and Keating (1997), who compared the consonant /n/ and the following vowel /o/ in terms of linguopalatal contact that they could measure for both the consonant and following vowel. However, in the present study, it was difficult to directly compare the different acoustic correlates that characterize consonants and the following vowels. For example, VOT for plosives does not have a corresponding acoustic correlate for the following vowels that it can be directly compared with. It would make the patterns of prosodic strengthening associated with syntagmatic contrast enhancement clearer if neighboring consonants and vowels could be directly compared rather than relying on abstract vocalicity defined by sonority, which has been controversial (e.g., Harris, 2006; see also Ohala, 1974, 1990a, 1990b), but doing so is inherently difficult due to the limitation of the comparison between neighboring consonants and vowels. Despite these limitations, the present study provides better understanding of prosodic strengthening driven by prosodic boundary and

suggests that prosodic strengthening driven by prosodic boundary should be understood in the interaction with different factors that modulate its patterns regarding its linguistic function (syntagmatic vs. paradigmatic contrast enhancement).

Chapter 5: Conclusion

The present study provides comprehensive picture of boundary-induced prosodic strengthening for English plosives /p, t, k, b, d, g/ and nasals /m, n/ that differ in voicing and place of articulation by examining their acoustic correlates and evaluating their patterns of prosodic strengthening in relation to its linguistic function. The present study shows that prosodic strengthening driven by prosodic boundary patterns differently depending on the speech sounds of interest and their acoustic correlates. Prosodic strengthening driven by prosodic boundary yields both syntagmatic contrast enhancement in which the contrast between neighboring consonants and vowels is enhanced and paradigmatic contrast enhancement in which the contrast between voicing or places of articulation. Speech style is one of the factors that influence the patterns of prosodic strengthening and its linguistic function. For plosives, read speech tends to induce syntagmatic contrast enhancement, whereas interactive speech tends to induce paradigmatic contrast enhancement. For nasals, except for segmental duration, both read and interactive tend to induce syntagmatic contrast enhancement when there is prosodic strengthening possibly because nasals are already phonetically and phonologically different enough from non-nasal consonants such that they do not have to enhance their nasality in domain-initial positions compared to domain-medial positions in interactive speech as well as in read speech. The present study has the implication that prosodic strengthening driven by prosodic boundary and its linguistic function may be contingent on speech style and on whether speech sounds are phonetically and phonologically different enough from other speech sounds that they can potentially be compared with. The present study further suggests that boundary-induced prosodic strengthening and its linguistic function should be investigated with wider range of speech sounds in different speech contexts.

Future research can expand on the findings in the present study to fill gaps that could not be addressed in the present study. Since this study assumed a listener-oriented account to explain the difference in the patterns of prosodic strengthening in terms of linguistic function in read and interactive speech, it is important to investigate whether and how the linguistic function of prosodic strengthening driven by prosodic boundary influences listeners' perception of words. To illustrate, listeners' recognition of words might be faster and/or more accurate when acoustic parameters are manipulated to have the patterns of paradigmatic contrast enhancement (e.g., longer positive VOT for voiceless stops and longer prevoicing for voiced stops) than when the same acoustic parameters are manipulated to have the patterns of syntagmatic contrast enhancement (e.g., longer positive VOT for both voiceless and voiced stops). Moreover, given that durational and spectral acoustic correlates pattern differently in relation to the linguistic function of prosodic strengthening, one of the questions that can also be addressed in future research is how these acoustic correlates are used and weighted by listeners at different levels of prosodic boundaries. In the present study, speakers used durational acoustic correlates more consistently than spectral acoustic correlates to realize prosodic strengthening. Would durational acoustic correlates also be used more reliably by listeners than spectral acoustic correlates when recognizing spoken words at the edge of different prosodic boundaries? Listeners might benefit more from durational acoustic correlates than spectral acoustic correlates when recognizing spoken words at the edge of different prosodic boundaries. This question will provide us more insight into whether prosodic strengthening can be explained as a listener-oriented tactic in speech production.

Lastly, we need to further investigate prosodic strengthening on speech sounds other than plosives and nasals in read and interactive speech to gain insight as to how the linguistic function

of prosodic strengthening differs partly depending on whether the speech sounds being compared are phonetically and phonologically different enough from each other. The hypothesis should be tested with speech sounds in other languages to see how the sound system in a language and the linguistic function of prosodic strengthening interact. Furthermore, more investigation is needed to find out factors that can influence the difference in the patterns of prosodic strengthening in relation to linguistic function between plosives and nasals, such as functional load or neighborhood density.

References

Abramson, A. S., & Whalen, D. H. (2017). Voice Onset Time (VOT) at 50: Theoretical and practical issues in measuring voicing distinctions. *Journal of phonetics, 63*, 75-86.

Ayers, G. M. (1994). Discourse functions of pitch range in spontaneous and read speech. *Ohio State University Working Papers in Linguistics, 44*, 1-49.

Barry, W. J. (1995). Phonetics and phonology of speaking styles. *Proceedings of 13th International Congress of Phonetic Sciences* (Vol. 2, pp. 4-10).

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1-48. doi: 10.18637/jss.v067.i01.

Beckman, M. E., & Ayers, G. (1997). Guidelines for ToBI labelling. *The OSU Research Foundation*, *3*, 30.

Beckman, M. E. & Elam, G. A. (1997). Guidelines for ToBI labelling, version 3.0. The Ohio State University Research Foundation

Beckman, M. E., & Edwards, J. (1990). Lengthenings and shortenings and the nature of prosodic constituency. In J. Kingston & M.E. Beckman (Eds.), *Papers in Laboratory Phonology I: Between the Grammar and the Physics of Speech* (pp. 152-178)*,* Cambridge: Cambridge University Press.

Beckman, M. E., & Edwards, J. (1994). Articulatory evidence for differentiating stress categories. In P. Keating (Ed.), *Papers in laboratory phonology III: Phonological structure and phonetic form* (pp. 7-33), Cambridge: Cambridge University Press.

Beckman, M. E., & Pierrehumbert, J. B. (1986). Intonational structure in Japanese and English. *Phonology*, *3*, 255-309.

Bell-Berti, F. (1993). Understanding velic motor control: Studies of segmental context. In M. K. Huffman & R. A. Krakow (Eds.), Phonetics and Phonology, Vol. 5, *Nasals, nasalization, and the velum* (pp. 63-85), San Diego, CA: Academic Press.

Blaauw, E. (1992). Phonetic differences between read and spontaneous speech. *Proceedings of Second International Conference on spoken language processing* (pp. 751-758). Banff.

Blumstein, S. E., & Stevens, K. N. (1980). Perceptual invariance and onset spectra for stop consonants in different vowel environments. *The Journal of the Acoustical Society of America*, *67*(2), 648-662.

Bolinger, D. L. (1958). A theory of pitch accent in English. *Word*, *14*(2-3), 109-149.

Bolinger, D. (1965). Pitch accent and sentence rhythm. *Forms of English: Accent, morpheme, order*, 139-180.

Bombien, L., Mooshammer, C., & Hoole, P. (2013). Articulatory coordination in word-initial clusters of German. *Journal of Phonetics*, *41*(6), 546-561.

Boersma, P., & Weenink, D. (2018). Praat: Doing phonetics by computer (Version 6.0. 37). http://www.praat.org/

Bruce, G. (1995). Modelling Swedish intonation for read and spontaneous speech. *Proceedings of International Congress of Phonetic Sciences* (Vol. 2, pp. 28-35).

Byrd, D. (2000). Articulatory vowel lengthening and coordination at phrasal junctures. *Phonetica*, *57*(1), 3-16.

Byrd, D., Kaun, A., Narayanan, S., & Saltzman, E. (2000). Phrasal signatures in articulation. *Papers in laboratory phonology V*, 70-87.

Byrd, D., Krivokapić, J., & Lee, S. (2006). How far, how long: On the temporal scope of prosodic boundary effects. *The Journal of the Acoustical Society of America*, *120*(3), 1589-1599.

Byrd, D., & Saltzman, E. (1998). Intragestural dynamics of multiple prosodic boundaries. *Journal of Phonetics*, *26*(2), 173-199.

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

Cho, T. (2006). Manifestation of prosodic structure in articulatory variation: Evidence from lip kinematics in English. In L. Goldstein (Ed.), *Laboratory phonology 8: Varieties of phonological competence* (pp. 519-548). New York, NY: Walter De Gruyter.

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

Cho, T., & Jun, S. A. (2000). Domain-initial strengthening as enhancement of laryngeal features: Aerodynamic evidence from Korean. *UCLA working papers in phonetics*, 57-70.

Cho, T., & Keating, P. A. (2001). Articulatory and acoustic studies on domain-initial strengthening in Korean. *Journal of phonetics*, *29*(2), 155-190.

Cho, T., & Keating, P. (2009). Effects of initial position versus prominence in English. *Journal of Phonetics*, *37*(4), 466-485.

Cho, T., Lee, Y., & Kim, S. (2014). Prosodic strengthening on the/s/-stop cluster and the phonetic implementation of an allophonic rule in English. *Journal of Phonetics*, *46*, 128-146.

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.

Cho, T., McQueen, J. M., & Cox, E. A. (2007). Prosodically driven phonetic detail in speech processing: The case of domain-initial strengthening in English. *Journal of Phonetics*, *35*(2), 210-243.

Cole, J., Kim, H., Choi, H., & Hasegawa-Johnson, M. (2007). Prosodic effects on acoustic cues to stop voicing and place of articulation: Evidence from Radio News speech. *Journal of Phonetics*, *35*(2), 180-209.

Christophe, A., Peperkamp, S., Pallier, C., Block, E., & Mehler, J. (2004). Phonological phrase boundaries constrain lexical access I. Adult data. *Journal of Memory and Language*, *51*(4), 523-547.

Deese, J. (2011). Pauses, prosody, and the demands of production in language. In H. W. Dechert & M. Raupach (Eds.), *Temporal variables in speech: Studies in honor of Frieda Goldman-Eisler* (pp. 69-84). Berlin: Mouton de Gruyter.

de Jong, K. J. (1995). The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation. *The Journal of the Acoustical Society of America*, *97*(1), 491-504.

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.

de Jong, K., Beckman, M. E., & Edwards, J. (1993). The interplay between prosodic structure and coarticulation. *Language and speech*, *36*(2-3), 197-212.

Delvaux, V., & Soquet, A. (2007). The influence of ambient speech on adult speech productions through unintentional imitation. *Phonetica*, *64*(2-3), 145-173.

Dilley, L., Shattuck-Hufnagel, S., & Ostendorf, M. (1994). Prosodic constraints on glottalization of vowel-initial syllables in American English. *The Journal of the Acoustical Society of America*, *95*(5), 2978-2979.

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.

D'Imperio, M., Elordieta, G., Frota, S., Prieto, P., & Vigário, M. (2005). Intonational phrasing in Romance: The role of syntactic and prosodic structure. In S. Frota, M. Vigário, & M. J. Freitas (Eds.), *Prosodies: With special reference to Iberian languages* (pp. 59-97). Berlin/New York: Mouton de Gruyter.

Dorman, M. F., Studdert-Kennedy, M., & Raphael, L. J. (1977). Stop-consonant recognition: Release bursts and formant transitions as functionally equivalent, context-dependent cues. *Perception & Psychophysics*, *22*(2), 109-122.

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.

Edwards, T. J. (1981). Multiple features analysis of intervocalic English plosives. *The Journal of the Acoustical Society of America*, *69*(2), 535-547.

Elordieta, G., Frota, S., & Vigário, M. (2005). Subjects, objects and intonational phrasing in Spanish and Portuguese. *Studia Linguistica*, *59*(2-3), 110-143.

Fant, G. (1973). *Speech sounds and features.* MIT Press: Cambridge, MA.

Ferreira, F. (1991). Effects of length and syntactic complexity on initiation times for prepared utterances. *Journal of Memory and Language*, *30*(2), 210-233.

Ferreira, F. (1993). Creation of prosody during sentence production. *Psychological review*, *100*(2), 233-253.

Fischer-Jørgensen, E. (1954). Acoustic analysis of stop consonants. *Le Maître Phonétique*, *32*, 42-59.

Fougeron, C. (1999). Prosodically conditioned articulatory variations: A review. *UCLA working Papers in Phonetics*, *97*, 1-73.

Fougeron, C. (2001). Articulatory properties of initial segments in several prosodic constituents in French. *Journal of phonetics*, *29*(2), 109-135.

Fougeron, C., & Keating, P. A. (1997). Articulatory strengthening at edges of prosodic domains. *The journal of the acoustical society of America*, *101*(6), 3728-3740.

Fowler, C. A. (1995). Acoustic and kinematic correlates of contrastive stress accent in spoken English. In F. Bell-Berti & L. J. Raphael (Eds.), *Producing speech: Contemporary issues: For Katherine Safford Harris* (pp. 355-373). Woodbury, NY: AIP Press.

Frazier, L., Clifton Jr, C., & Carlson, K. (2004). Don't break, or do: Prosodic boundary preferences. *Lingua*, *114*(1), 3-27.

Frota, S. (2014). *Prosody and focus in European Portuguese: Phonological phrasing and intonation*. Routledge.

Flege, J. E. (1982). Laryngeal timing and phonation onset in utterance-initial English stops. *Journal of Phonetics*, *10*(2), 177-192.

Fujimura, O. (1962). Analysis of nasal consonants. *Journal of the Acoustical Society of America, 34*(12), 1865–1875.

Garellek, M. (2012). Word-initial glottalization and voice quality strengthening. *UCLA Working Papers in Phonetics*, *111*, 92-122.

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

Gee, J. P., & Grosjean, F. (1983). Performance structures: A psycholinguistic and linguistic appraisal. *Cognitive psychology*, *15*(4), 411-458.

Georgeton, L., & Fougeron, C. (2014). Domain-initial strengthening on French vowels and phonological contrasts: Evidence from lip articulation and spectral variation. *Journal of Phonetics*, *44*, 83-95.

Goldman-Eisler, F. (1968). *Psycholinguistics: Experiments in spontaneous speech*. London and New York: Academic Press.

Goldman-Eisler, F. (1972). Pauses, clauses, sentences. *Language and speech*, *15*(2), 103-113.

Halle, M., Hughes, G. W., & Radley, J. P. (1957). Acoustic properties of stop consonants. *The Journal of the Acoustical Society of America*, *29*(1), 107-116.

Harris, J. (2006). The phonology of being understood: Further arguments against sonority. *Lingua*, *116*(10), 1483-1494.

Hawkins, S., & Stevens, K. N. (1985). Acoustic and perceptual correlates of the non-nasal–nasal distinction for vowels. *The Journal of the Acoustical Society of America, 77*(4), 1560-1575.

Hayes, B. (1989). The prosodic hierarchy in meter. In P. Kiparsky & G. Youmans (Eds.), *Phonetics and Phonology, Vol 1: Rhythm and meter* (pp. 201-260). San diego: Academic Press.

Hayes, B. (1995). *Metrical stress theory: Principles and case studies*. Chicago: University of Chicago Press.

Hellmuth, S. (2004). Prosodic weight and phonological phrasing in Cairene Arabic. *Proceedings from the 40th Chicago Linguistic Society* (pp. 97-111). Chicago Linguistic Society.

Herd, W. (2020). Sociophonetic voice onset time variation in Mississippi English. *The Journal of the Acoustical Society of America*, *147*(1), 596-605.

Hirschberg, J. (2000). A corpus-based approach to the study of speaking style. In M. Horne (Ed.), *Prosody: Theory and experiment* (pp. 335-350). Springer, Dordrecht.

Hsu, C. S. K., & Jun, S. A. (1998). Prosodic strengthening in Taiwanese: Syntagmatic or paradigmatic?. *UCLA working papers in phonetics*, 69-89.

Hunnicutt, L., & Morris, P. A. (2016). Prevoicing and aspiration in Southern American English. *University of Pennsylvania Working Papers in Linguistics*, *22*(1), 24.

Jacewicz, E., Fox, R. A., & Lyle, S. (2009). Variation in stop consonant voicing in two regional varieties of American English. *Journal of the International Phonetic Association*, *39*(3), 313-334.

Johnson, K. (1997). Acoustic and Auditory Phonetics. Oxford: Blackwell.

Johnson, K. (2003). Acoustic and Auditory Phonetics. Oxford: Blackwell.

Jongman, A., Wayland, R., & Wong, S. (2000). Acoustic characteristics of English fricatives. *The Journal of the Acoustical Society of America, 108*(3), 1252-1263.

Jun, S.-A. (1993). *The Phonetics and Phonology of Korean Prosody*. Ph.D. Dissertation. Ohio State University.

Kalbertodt, J., Primus, B., & Schumacher, P. B. (2015). Punctuation, prosody, and discourse: Afterthought vs. right dislocation. *Frontiers in psychology*, *6*, 1803. doi: 10.3389/fpsyg.2015.01803

Kang, K.-H., & Guion, S. (2008). Clear speech production of Korean stops: Changing phonetic targets and enhancement strategies. *Journal of the Acoustical Society of America, 124*(6), 3909-3917.

Keating, P. A. (1984). Phonetic and phonological representation of stop consonant voicing. *Language, 60*(2) 286-319.

Keating, P. A. (1990). Phonetic representations in a generative grammar. *Journal of phonetics*, *18*(3), 321-334.

Keating, P. A. (2006). Phonetic encoding of prosodic structure. In J. Harrington & M. Tabain (Eds.), *Speech production: Models, phonetic processes, and techniques* (pp. 167-186). New York, NY: Psychology Press.

Keating, P. A., Byrd, D., Flemming, E., & Todaka, Y. (1994). Phonetic analyses of word and segment variation using the TIMIT corpus of American English. *Speech Communication*, *14*(2), 131-142.

Keating, P., Cho, T., Fougeron, C., & Hsu, C. S. (2003). Domain-initial articulatory strengthening in four languages. *Phonetic interpretation: Papers in laboratory phonology VI*, 143-161.

Keating, P., & Lahiri, A. (1993). Fronted velars, palatalized velars, and palatals. *Phonetica, 50*(2), 73-101.

Keating, P., & Shattuck-Hufnagel, S. (2002). A prosodic view of word form encoding for speech production. *UCLA working papers in phonetics*, *101*, 112-156.

Keating, P., Wright, R., & Zhang, J. (1999). Word-level asymmetries in consonant articulation. *UCLA Working Papers in Phonetics, 97*, 157-173.

Kim, M.-R., Beddor, P. S., & Horrocks, J. (2002). The contribution of consonantal and vocalic information to the perception of Korean initial stops. *Journal of Phonetics, 30*(1)*,* 77-100.

Kim, S., Kim, J., & Cho, T. (2018). Prosodic-structural modulation of stop voicing contrast along the VOT continuum in trochaic and iambic words in American English. *Journal of Phonetics*, *71*, 65-80.

Kohler, K. J. (1995). Articulatory reduction in different speaking styles. *Proceedings of the 13th International Congress of Phonetic Sciences* (Vol. 12, pp. 12-19).

Krivokapić, J. (2007). Prosodic planning: Effects of phrasal length and complexity on pause duration. *Journal of phonetics*, *35*(2), 162-179.

Krivokapic, J. (2012). Prosodic planning in speech production. *Speech planning and dynamics*, 157-190.

Krivokapić, J. (2014). Gestural coordination at prosodic boundaries and its role for prosodic structure and speech planning processes. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1658), 20130397.

Kuzla, C., & Ernestus, M. (2011). Prosodic conditioning of phonetic detail in German plosives. *Journal of Phonetics*, *39*(2), 143-155.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of statistical software*, *82*(13), 1-26.

Ladd, R., & Campbell, N. (1991). Theories of prosodic structure: Evidence from syllable duration. *Proceedings of the 12th International Congress of Phonetic Sciences* (pp. 290-293). Aix-en-Provence: University of Provence.

Ladefoged, P. (1975). *A Course in Phonetics.* New York: Harcourt, Brace, Jovanovich.

Lee, H., & Jongman, A. (2012). Effects of tone on the three-way laryngeal distinction in Korean: An acoustic and aerodynamic comparison of the Seoul and South Kyungsang dialects. *Journal of the International Phonetic Association*, *42*(2), 145-169.

Lee, H., Politzer-Ahles, S., & Jongman, A. (2013). Speakers of tonal and non-tonal Korean dialects use different cue weightings in the perception of the three-way laryngeal stop contrast. *Journal of phonetics*, *41*(2), 117-132.

Lehiste, I., Olive, J. P., & Streeter, L. A. (1976). Role of duration in disambiguating syntactically ambiguous sentences. *The Journal of the Acoustical Society of America*, *60*(5), 1199-1202.

Lenth, R. (2021). emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.5.4. https://CRAN.R-project.org/package=emmeans

Levelt, W. J. (1993). *Speaking: From intention to articulation* (Vol. 1). MIT press.

Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and brain sciences*, *22*(1), 1-38.

Liberman, M., & Prince, A. (1977). On stress and linguistic rhythm. *Linguistic inquiry*, *8*(2), 249-336.

Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, *20*(3), 384-422.

Lisker, L., & Abramson, A. S. (1967). Some effects of context on voice onset time in English stops. *Language and speech*, *10*(1), 1-28.

Nearey, T. M., & Rochet, B. L. (1994). Effects of place of articulation and vowel context on VOT production and perception for French and English stops. *Journal of the International Phonetic Association*, *24*(1), 1-18.

Nespor, M., & Vogel, I. (1986). *Prosodic Phonology*. Dordrecht: Foris Publications.

Ohala, J. J. (1974). Phonetic explanation in phonology. *Parasession on natural phonology*, 251-274.

Ohala, J. J. (1990a). Alternatives to the sonority hierarchy for explaining segmental sequential constraints." In M. Ziolkowski, M. Noske, & K. Deaton (Eds.), *CLS 26: Papers from the 26th Regional Meeting of the Chicago Linguistic Society, volume 2: the parasession on the syllable in phonetics and phonology* (pp. 319-338). Chicago: Chicago Linguistic Society.

Ohala, J. J. (1990b). There is no interface between phonology and phonetics: a personal view. *Journal of phonetics*, *18*(2), 153-171.

Ostendorf, M., Price, P. J., & Shattuck-Hufnagel, S. (1995). The Boston University radio news corpus. *Linguistic Data Consortium*, 1-19.

Pardo, J. S. (2006). On phonetic convergence during conversational interaction. *The Journal of the Acoustical Society of America*, *119*(4), 2382-2393.

Pierrehumbert, J., & Beckman, M. B. (1988). *Japanese Tone Structure*. Cambridge, Mass.: MIT Press.

Pierrehumbert, J., & Frisch, S. (1994). Source allophony and speech synthesis. *Proceedings of the* 2nd *ESCA/IEEE Workshop on Speech Synthesis*, 1-4.

Pierrehumbert, J., & Talkin, D. (1992). Lenition of /h/ and glottal stop. In G. J. Doherty & D. R. Ladd (Eds.), *Papers in Laboratory Phonology II* (pp. 90-117). Cambridge: Cambridge University Press.

Pruthi, T., & Espy-Wilson, C. (2003). Automatic classification of nasals and semivowels. *Proceedings of the 15th International Congress on Phonetic Sciences*. Barcelona, Spain.

Repp, B. H., & Lin, H. B. (1989). Acoustic properties and perception of stop consonant release transients. *The Journal of the Acoustical Society of America*, *85*(1), 379-396.

Saltzman, E. (1995). Intergestural timing in speech production: Data and modeling. *Proceedings of the XIIIth International Congress of Phonetic Sciences*, *2*, 84-91.

Salverda, A. P., Dahan, D., & McQueen, J. M. (2003). The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition*, *90*(1), 51-89.

Salverda, A. P., Dahan, D., Tanenhaus, M. K., Crosswhite, K., Masharov, M., & McDonough, J. (2007). Effects of prosodically modulated sub-phonetic variation on lexical competition. *Cognition*, *105*(2), 466-476.

Schwartz, J. L., Boë, L. J., Vallée, N., & Abry, C. (1997). The dispersion-focalization theory of vowel systems. *Journal of phonetics*, *25*(3), 255-286.

Selkirk, E. O. (1978). On prosodic structure and its relation to syntactic structure. In T. Fretheim (Ed.), *Nordic Prosody II* (pp. 111-140). Trondheim: TAPIR.

Selkirk, E. O. (1980). The role of prosodic categories in English word stress. *Linguistic inquiry*, *11*(3), 563-605.

Selkirk, E. O. (1984). *Phonology and syntax: the relation between sound and structure*. Cambridge, Mass.: MIT Press.

Selkirk, E. O. (1986). On derived domains in sentence phonology. *Phonology*, *3*, 371-405.

Selkirk, E. O. (2000). The interaction of constraints on prosodic phrasing. In M. Horne (Ed.), *Prosody: Theory and experiment* (pp. 231-261). Springer, Dordrecht.

Selkirk, E. O. (2005). Comments on intonational phrasing in English. In S. Frota, M. Vigário, & M. J. Freitas (Eds.), *Prosodies: With special reference to Iberian languages* (pp. 11-58). Berlin/New York: Mouton de Gruyter.

Shattuck-Hufnagel, S., & Turk, A. E. (1996). A prosody tutorial for investigators of auditory sentence processing. *Journal of psycholinguistic research*, *25*(2), 193-247.

Shattuck-Hufnagel, S., & Turk, A. E. (1998). The domain of phrase-final lengthening in English. *The sound of the future: A global view of acoustics in the 21st century*, 1235-1236.

Shin, S., Kim, S., & Cho, T. (2015). What is special about prosodic strengthening in Korean: Evidence in lingual movement in V# V and V# CV. *Proceedings of the 18th International Congress on Phonetic Sciences*. Glasgow, Scotland: University of Glasgow.

Shin, S., & Tremblay, A. (2018). Effect of prosodic context on lexical access: An investigation of Korean denasalization. *Proceedings of the 9th Speech Prosody Conference* (pp. 408-412). Poznań, Poland.

Shue, Y. L., Shattuck-Hufnagel, S., Iseli, M., Jun, S. A., Veilleux, N., & Alwan, A. (2010). On the acoustic correlates of high and low nuclear pitch accents in American English. *Speech Communication, 52*(2), 106-122.

Silverman, K. E., Blaauw, E., Spitz, J., & Pitrelli, J. F. (1992). Towards using prosody in speech recognition/understanding systems: Differences between read and spontaneous speech. *Proceedings of the workshop on Speech and Natural Language* (pp. 435-440). Association for Computational Linguistics.

Slifka, J. (2006). Some physiological correlates to regular and irregular phonation at the end of an utterance. *Journal of Voice, 20*(2), 171-186.

Smits, R. (1996). Context-dependent relevance of burst and transitions for perceived place in stops: it's in production, not perception. *Proceeding of Fourth International Conference on Spoken Language Processing.* ICSLP ’96 (pp. 2470-2473). Philadelphia, PA, USA.

Smith, B. L. (1978). Temporal aspects of English speech production: A developmental perspective. *Journal of Phonetics*, *6*(1), 37-67.

Stevens, K. (1998). *Acoustic Phonetics.* Cambridge: MIT Press.

Stevens, K. N., & Blumstein, S. E. (1978). Invariant cues for place of articulation in stop consonants. *The Journal of the Acoustical Society of America*, *64*(5), 1358-1368.

Straka, G. (1963). La division des sons du langage en voyelles et consonnes peut-elle être justifiée? *Travaux de Linguistique et de Littérature* (Vol. 1, pp. 17-99). Centre de Philologie et de Littératures Romanes de l'Université de Strasbourg.

Styler, W. (2017). On the acoustical features of vowel nasality in English and French. *The Journal of the Acoustical Society of America, 142*(4), 2469-2482.

Sussman, H. M., Bessell, N., Dalston, E., & Majors, T. (1997). An investigation of stop place of articulation as a function of syllable position: A locus equation perspective. *The Journal of the Acoustical Society of America*, *101*(5), 2826-2838.

Sussman, H. M., McCaffrey, H. A., & Matthews, S. A. (1991). An investigation of locus equations as a source of relational invariance for stop place categorization. *The Journal of the Acoustical Society of America*, *90*(3), 1309-1325.

Sussman, H. M., & Shore, J. (1996). Locus equations as phonetic descriptors of consonantal place of articulation. *Perception & psychophysics*, *58*(6), 936-946.

Tabain, M. (2003). Effects of prosodic boundary on /aC/ sequences: Articulatory results. *The Journal of the Acoustical Society of America*, *113*(5), 2834-2849.

Tabain, M., & Perrier, P. (2005). Articulation and acoustics of /i/ in preboundary position in French. *Journal of Phonetics*, *33*(1), 77-100.

Tremblay, A., Broersma, M., Coughlin, C. E., & Choi, J. (2016). Effects of the native language on the learning of fundamental frequency in second-language speech segmentation. *Frontiers in psychology*, *7*, 985. doi: 10.3389/fpsyg.2016.00985

Tremblay, A., Cho. T., Kim, S., & Shin, S. (2019). Phonetic and phonological effects of tonal information in the segmentation of Korean speech: An artificial-language segmentation study. *Applied Psycholinguistics, 40*(5), 1221-1240.

Turk, A. E., & Sawusch, J. R. (1997). The domain of accentual lengthening in American English. *Journal of Phonetics*, *25*(1), 25-41.

Turk, A. E., & White, L. (1999). Structural influences on accentual lengthening in English. *Journal of phonetics*, *27*(2), 171-206.

Vaissière J. (2011). On the acoustic and perceptual characterization of reference vowels in a cross-language perspective. *Proceedings of the 17th international congress of phonetic sciences*, 52-59. Hong-Kong, China.

Vanderslice, R., & Ladefoged, P. (1972). Binary suprasegmental features and transformational word-accentuation rules. *Language*, *48*(4), 819-838.

Watson, D., & Gibson, E. (2004). The relationship between intonational phrasing and syntactic structure in language production. *Language and cognitive processes*, *19*(6), 713-755.

Westbury, J. R. (1979). Aspects of the temporal control of voicing in consonant clusters in English. In *Texas Linguistic Forum Austin, Tex* (No. 14, pp. 1-304).

Wightman, C. W., Shattuck-Hufnagel, S., Ostendorf, M., & Price, P. J. (1992). Segmental durations in the vicinity of prosodic phrase boundaries. *The Journal of the Acoustical Society of America*, *91*(3), 1707-1717.

Winitz, H., Scheib, M. E., & Reeds, J. A. (1972). Identification of stops and vowels for the burst portion of /p, t, k/ isolated from conversational speech. *The Journal of the Acoustical Society of America*, *51*(4B), 1309-1317.

Zhang, C., Jepson, K., Lohfink, G., & Arvaniti, A. (2020). Speech data collection at a distance: Comparing the reliability of acoustic cues across homemade recordings. *The Journal of the Acoustical Society of America*, *148*(4), 2717-2717.

Zhao, S. Y. (2010). Stop-like modification of the dental fricative /ð/: An acoustic analysis. *The Journal of the Acoustical Society of America*, *128*(4), 2009-2020.

Zue, V. W. (1976). *Acoustic characteristics of stop consonants: A controlled study.* Ph.D. dissertation, MIT, Cambridge, MA.

Appendix A: Picture stimuli

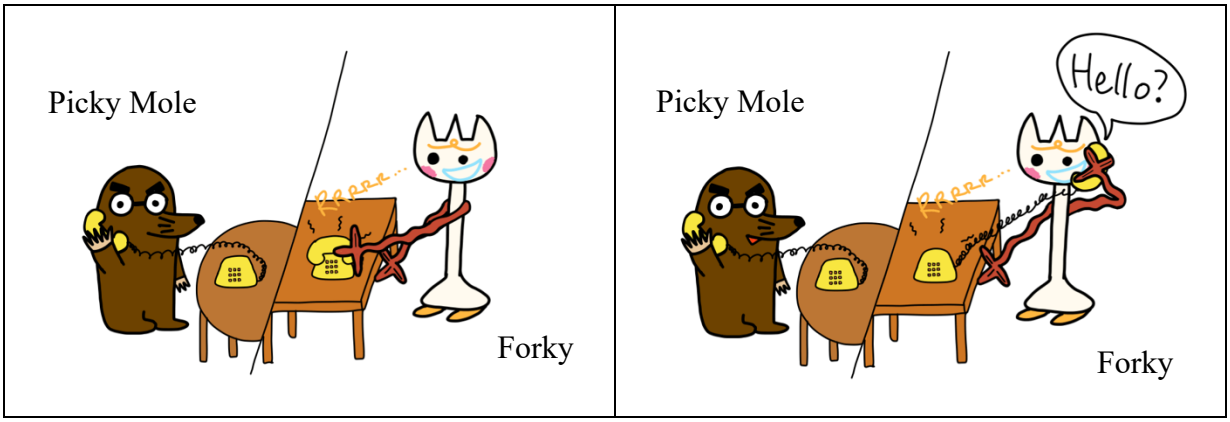


Figure A1: A scene in an experimental trial for the interactive speech task. The target word is *Picky* and located in IP-initial position. Two pictures were presented as a connected action (as a GIF).

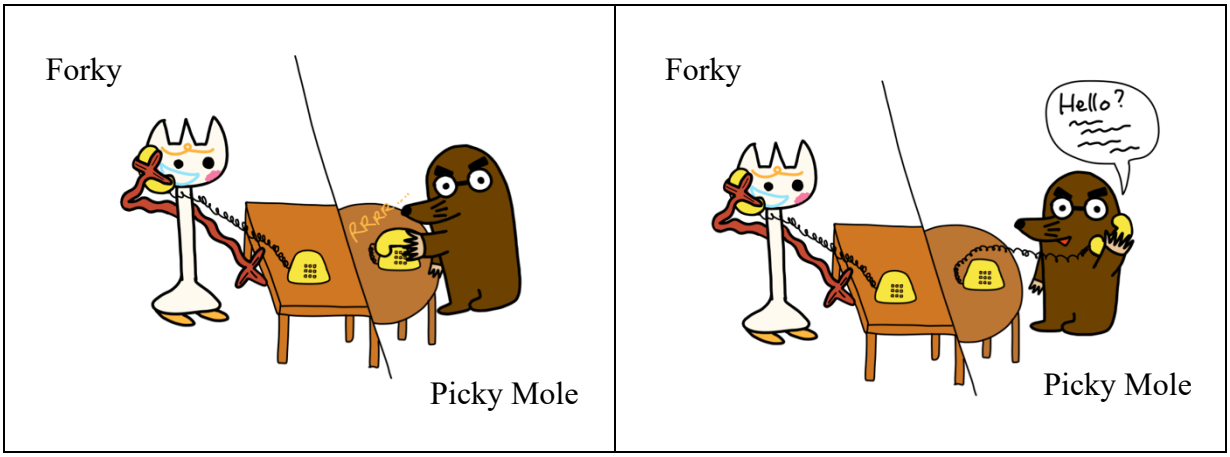


Figure A2: A scene in an experimental trial for the interactive speech task. The target word is *Picky* and located in IP-medial position. Two pictures were presented as a connected action (as a GIF).

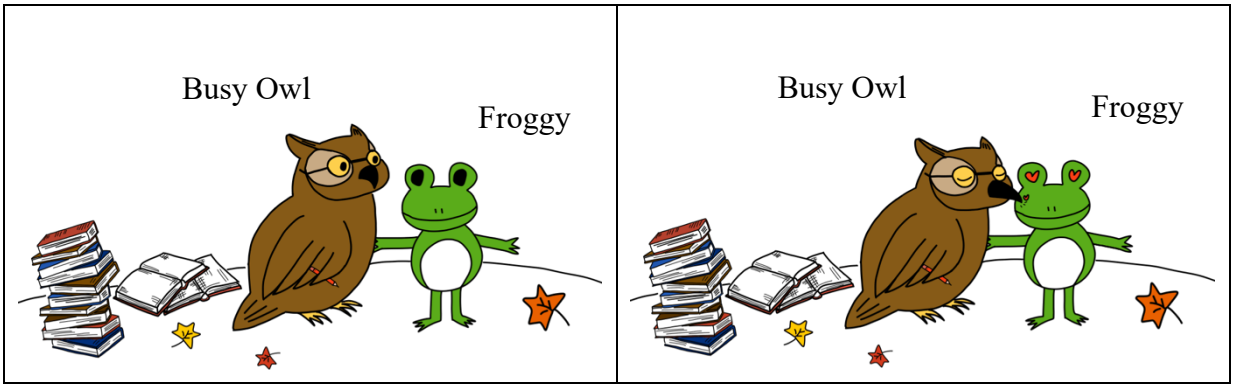


Figure A3: A scene in an experimental trial for the interactive speech task. The target word is *Busy* and located in IP-initial position. Two pictures were presented as a connected action (as a GIF).

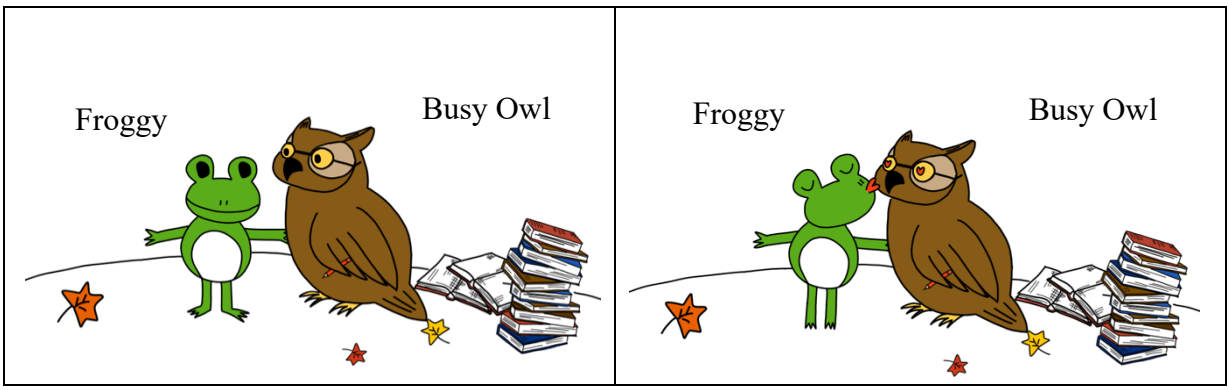


Figure A4: A scene in an experimental trial for the interactive speech task. The target word is *Busy* and located in IP-medial position. Two pictures were presented as a connected action (as a GIF).

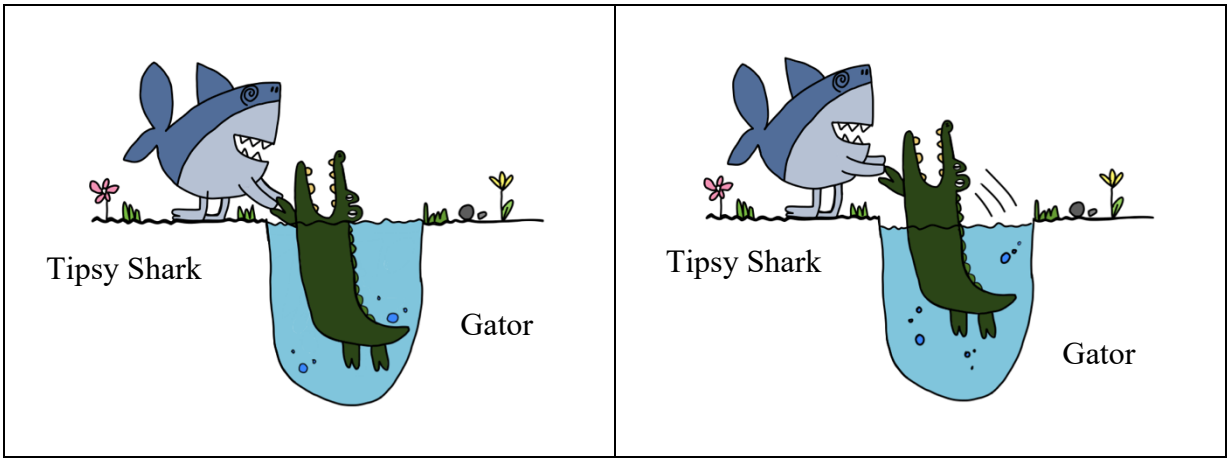


Figure A5: A scene in an experimental trial for the interactive speech task. The target word is *Tipsy* and located in IP-initial position. Two pictures were presented as a connected action (as a GIF).

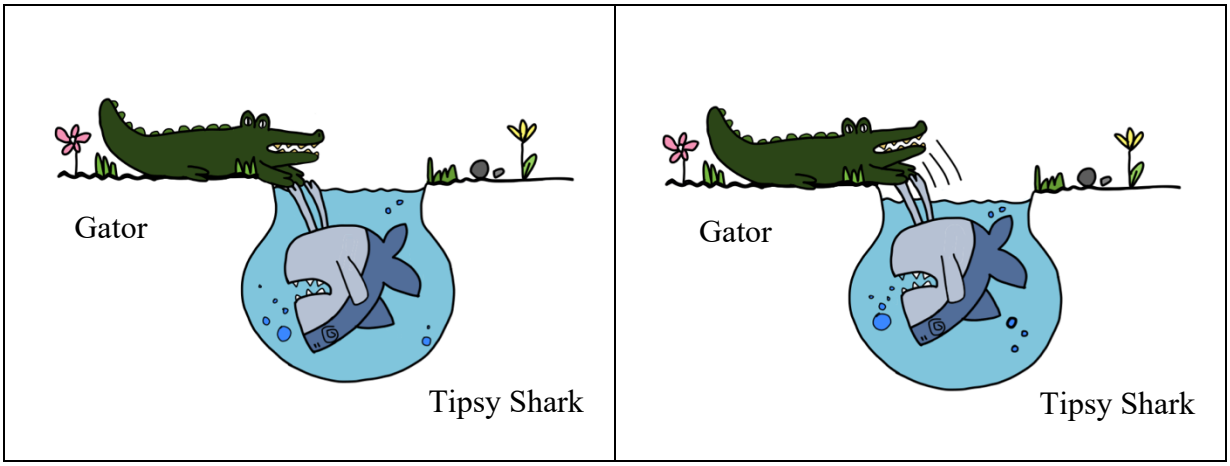


Figure A6: A scene in an experimental trial for the interactive speech task. The target word is *Tipsy* and located in IP-medial position. Two pictures were presented as a connected action (as a GIF).

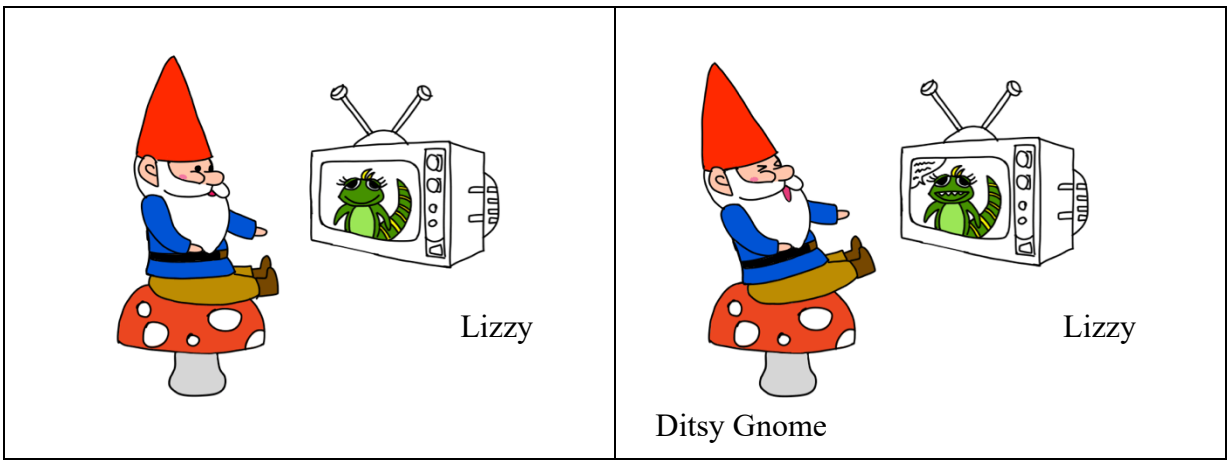


Figure A7: A scene in an experimental trial for the interactive speech task. The target word is *Ditsy* and located in IP-initial position. Two pictures were presented as a connected action (as a GIF).

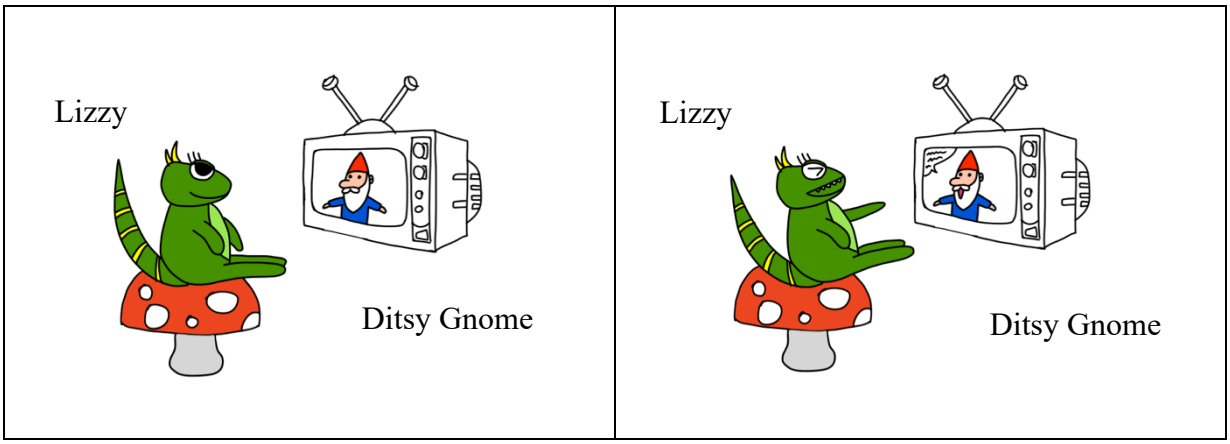


Figure A8: A scene in an experimental trial for the interactive speech task. The target word is *Ditsy* and located in IP-medial position. Two pictures were presented as a connected action (as a GIF).

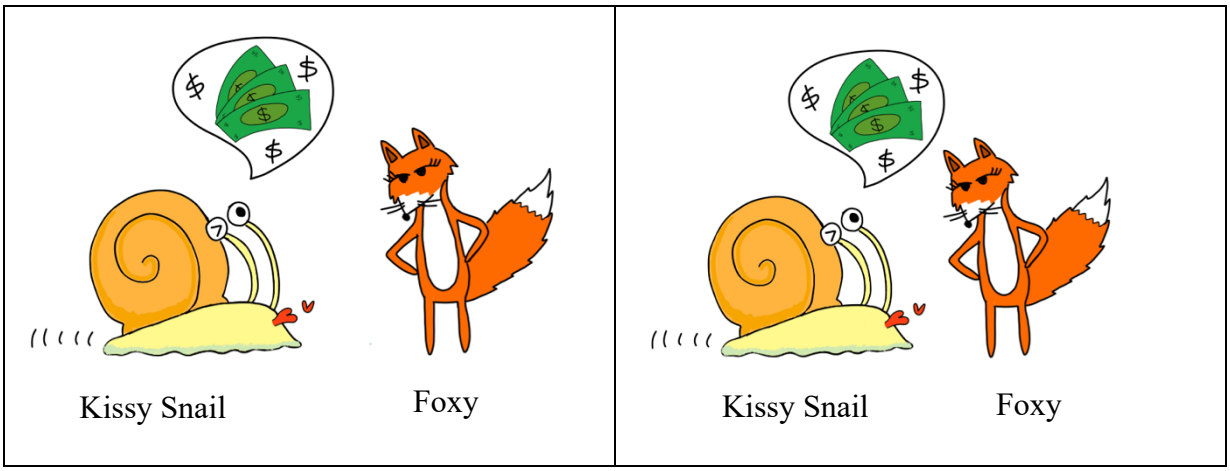


Figure A9: A scene in an experimental trial for the interactive speech task. The target word is *Kissy* and located in IP-initial position. Two pictures were presented as a connected action (as a GIF).

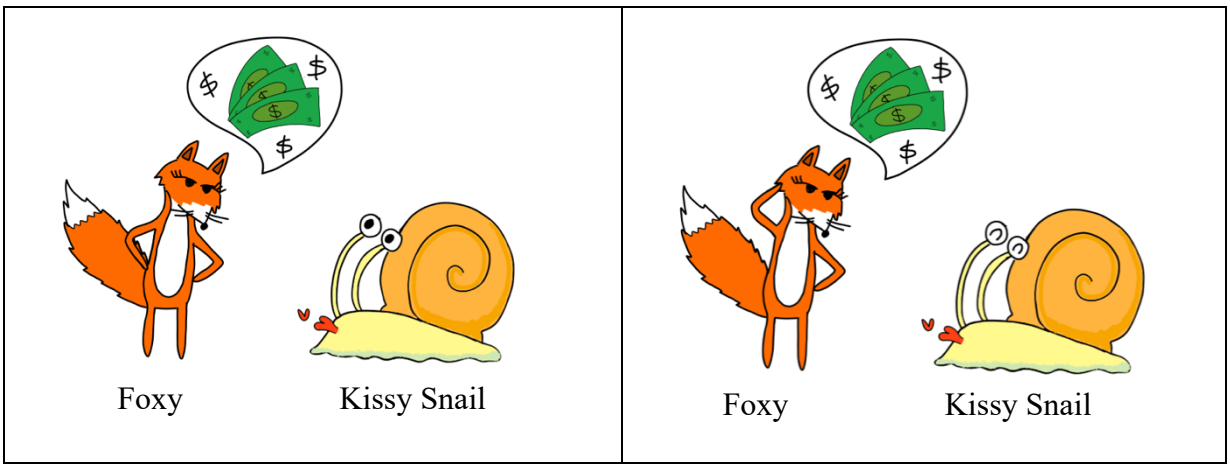


Figure A10: A scene in an experimental trial for the interactive speech task. The target word is *Kissy* and located in IP-medial position. Two pictures were presented as a connected action (as a GIF).

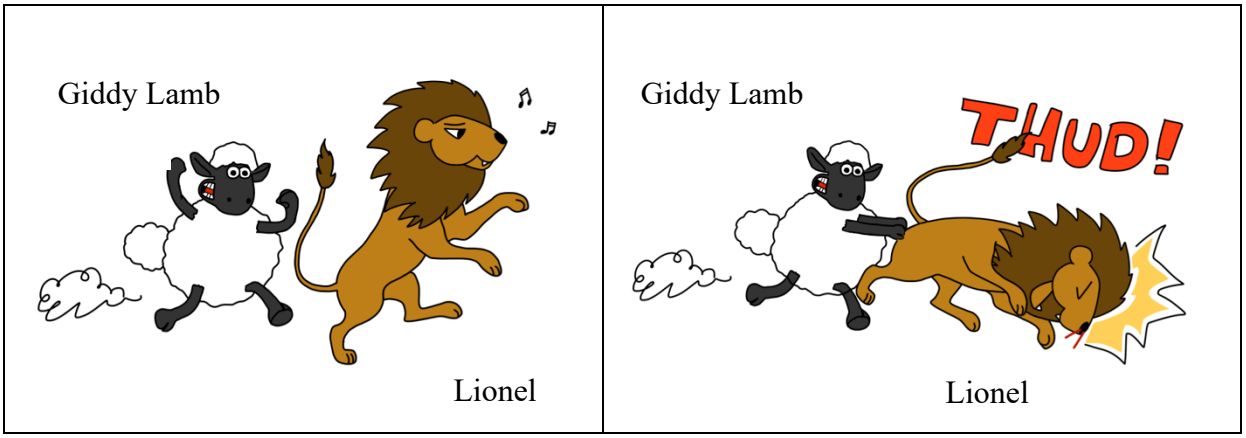


Figure A11: A scene in an experimental trial for the interactive speech task. The target word is *Giddy* and located in IP-initial position. Two pictures were presented as a connected action (as a GIF).

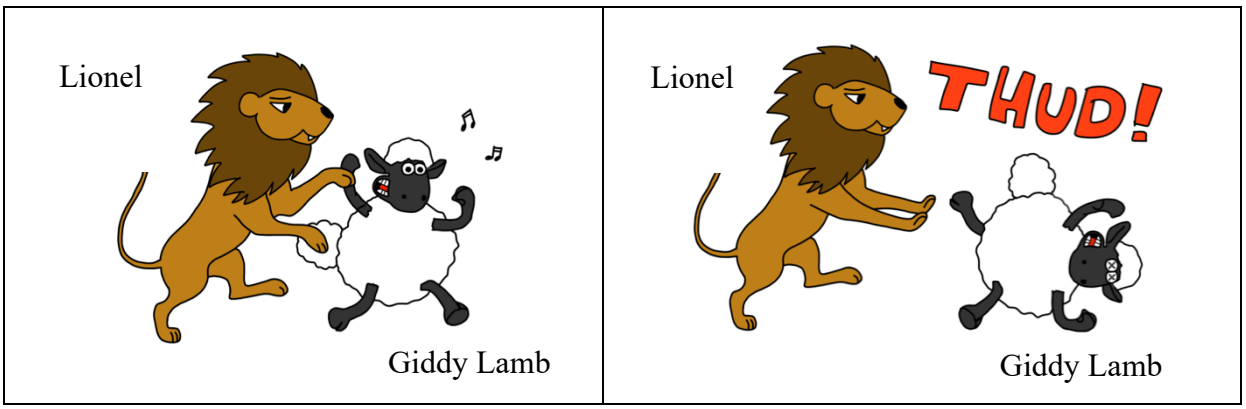


Figure A12: A scene in an experimental trial for the interactive speech task. The target word is *Giddy* and located in IP-medial position. Two pictures were presented as a connected action (as a GIF).

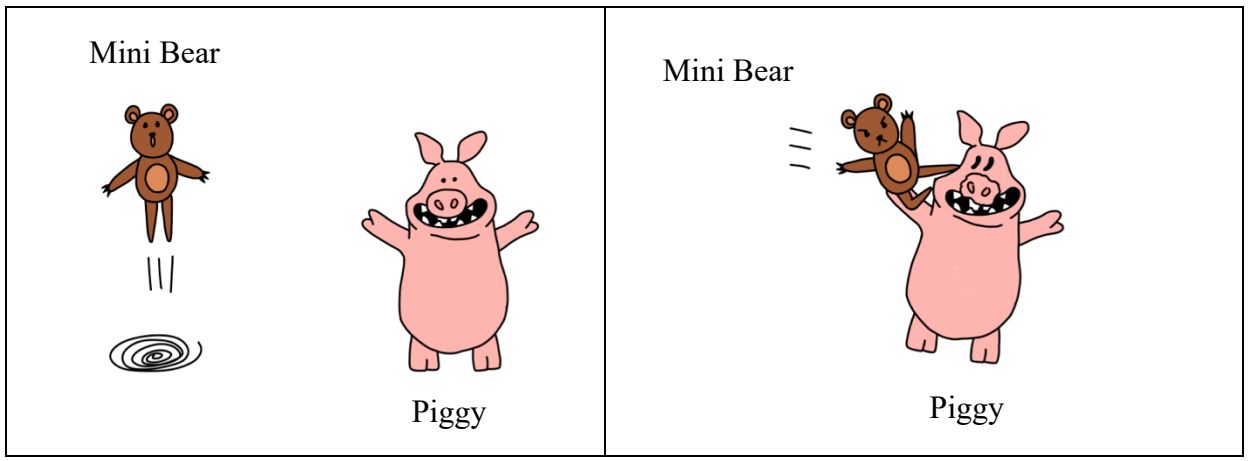


Figure A13: A scene in an experimental trial for the interactive speech task. The target word is *Mini* and located in IP-initial position. Two pictures were presented as a connected action (as a GIF).

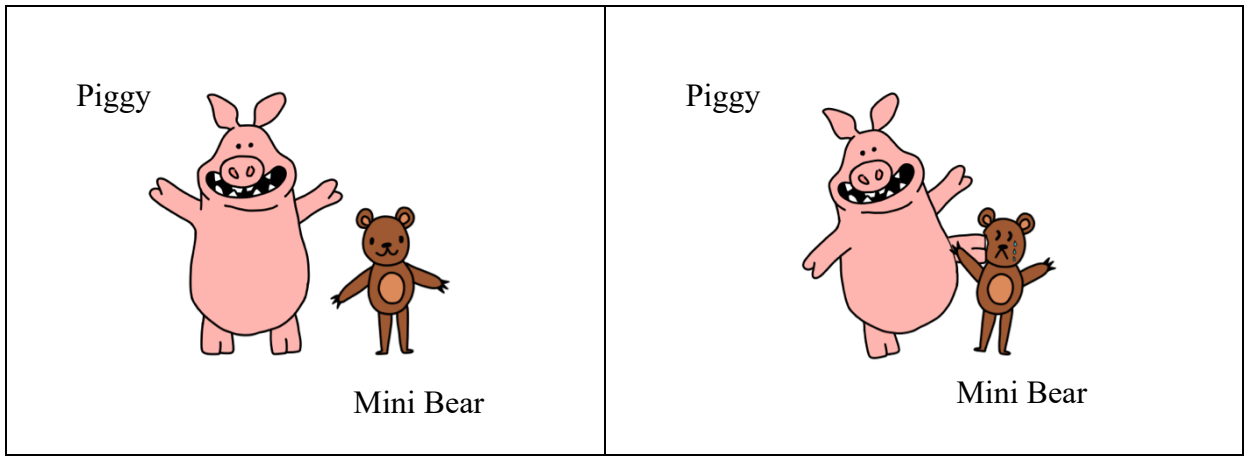


Figure A14: A scene in an experimental trial for the interactive speech task. The target word is *Mini* and located in IP-medial position. Two pictures were presented as a connected action (as a GIF).

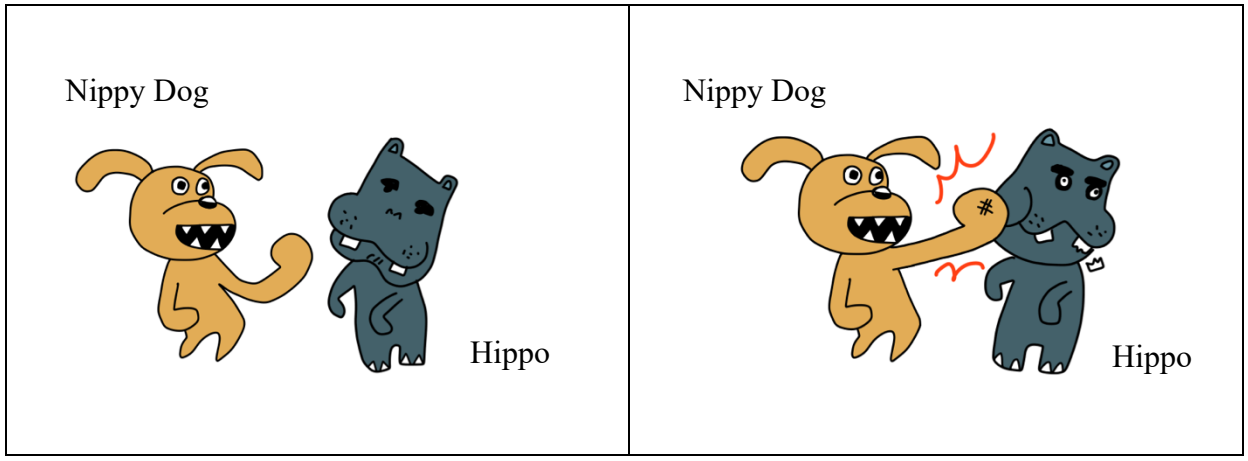


Figure A15: A scene in an experimental trial for the interactive speech task. The target word is *Nippy* and located in IP-initial position. Two pictures were presented as a connected action (as a GIF).

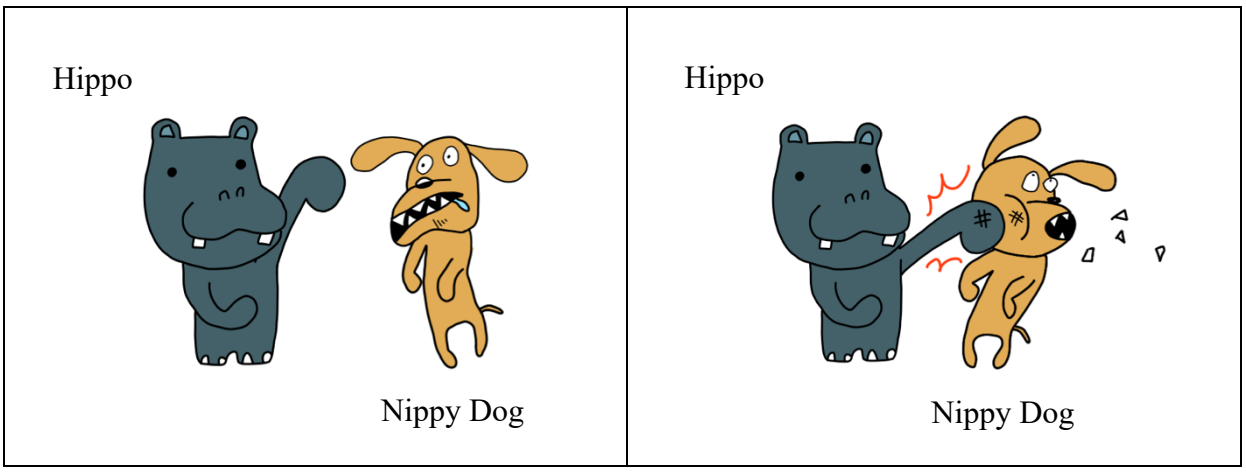


Figure A16: A scene in an experimental trial for the interactive speech task. The target word is *Nippy* and located in IP-medial position. Two pictures were presented as a connected action (as a GIF).

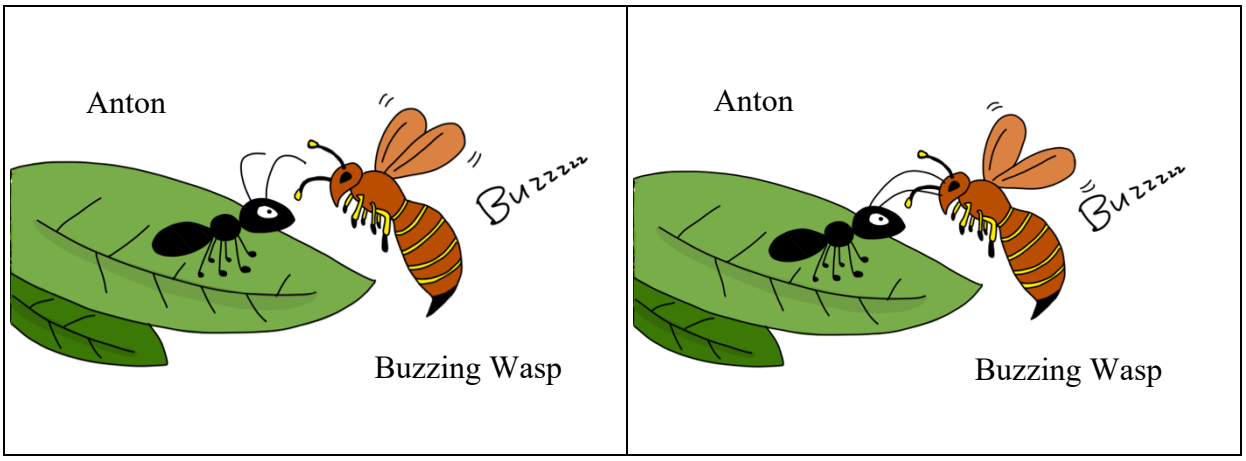


Figure A17: A scene in an experimental trial for the interactive speech task. The target word is *Buzzing* and located in IP-initial position. Two pictures were presented as a connected action (as a GIF).

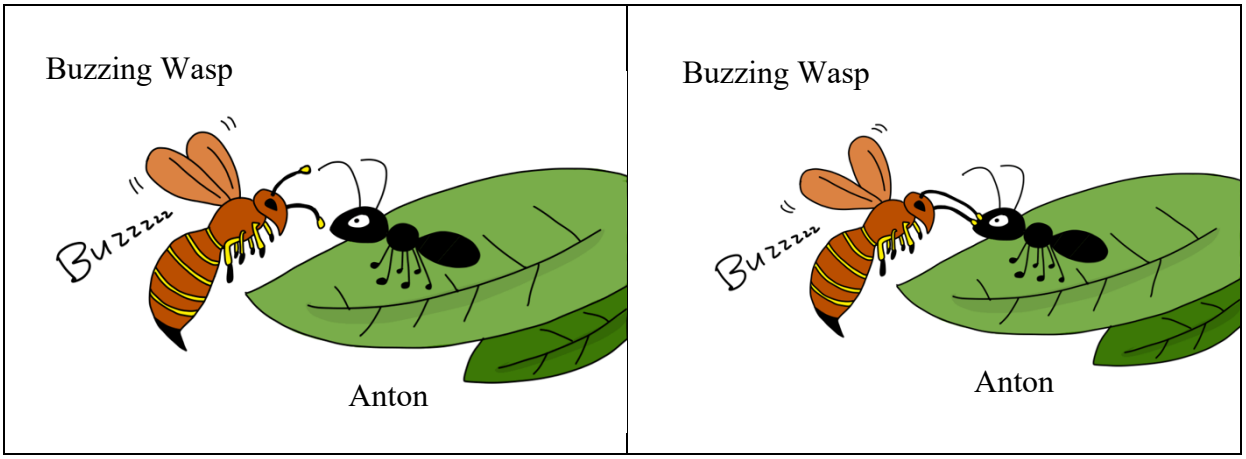


Figure A18: A scene in an experimental trial for the interactive speech task. The target word is *Buzzing* and located in IP-medial position. Two pictures were presented as a connected action (as a GIF).

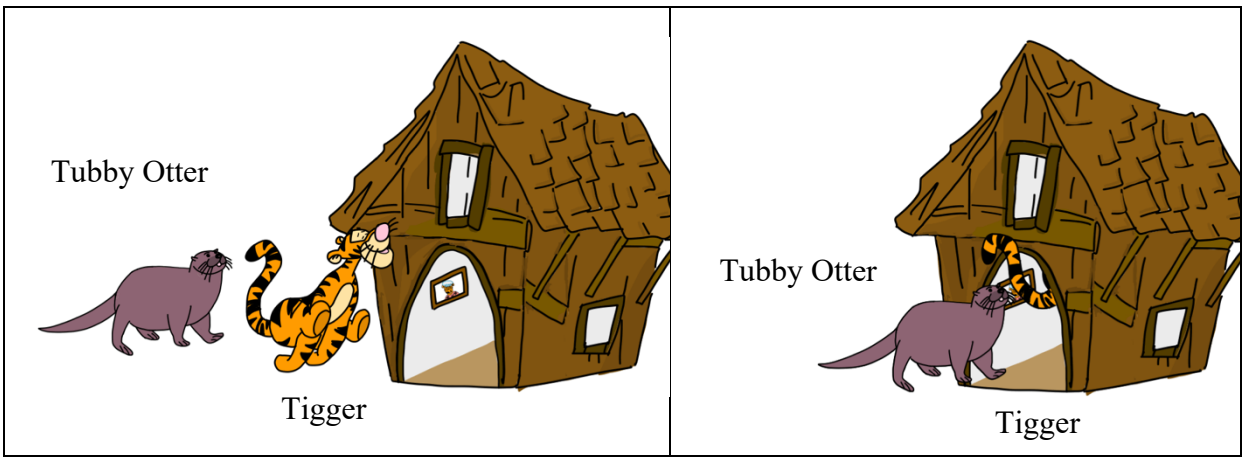


Figure A19: A scene in an experimental trial for the interactive speech task. The target word is *Tubby* and located in IP-initial position. Two pictures were presented as a connected action (as a GIF).

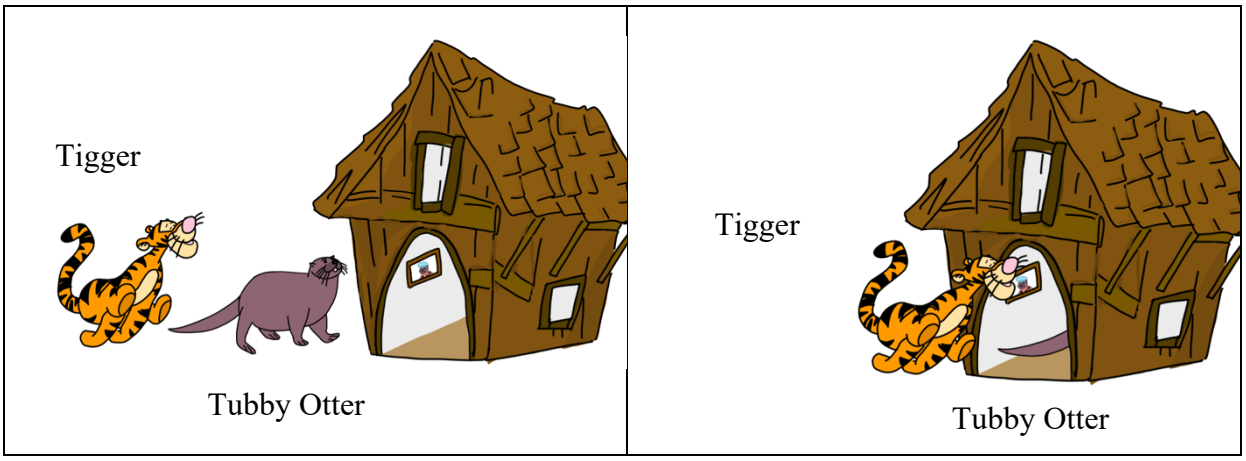


Figure A20: A scene in an experimental trial for the interactive speech task. The target word is *Tubby* and located in IP-medial position. Two pictures were presented as a connected action (as a GIF).

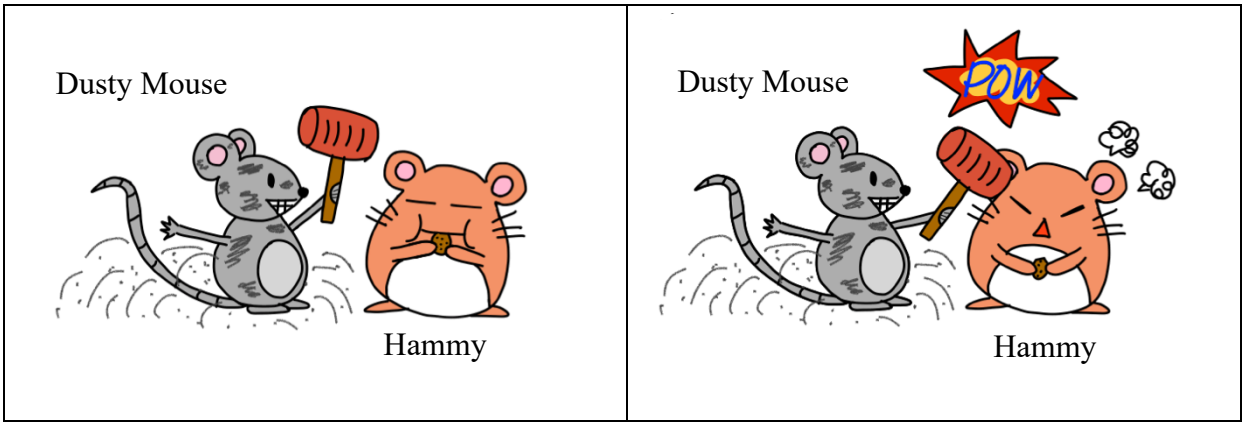


Figure A21: A scene in an experimental trial for the interactive speech task. The target word is *Dusty* and located in IP-initial position. Two pictures were presented as a connected action (as a GIF).

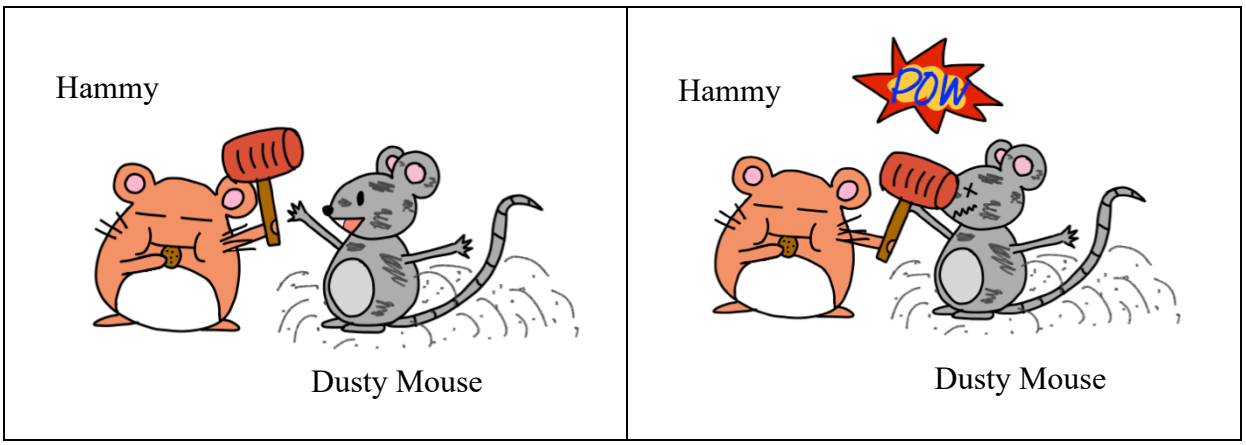


Figure A22: A scene in an experimental trial for the interactive speech task. The target word is *Dusty* and located in IP-medial position. Two pictures were presented as a connected action (as a GIF).

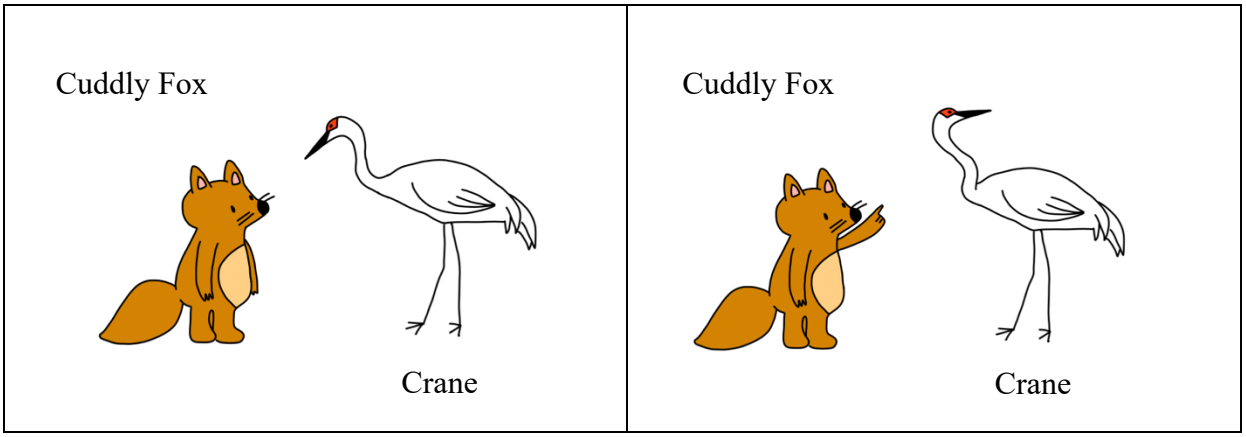


Figure A23: A scene in an experimental trial for the interactive speech task. The target word is *Cuddly* and located in IP-initial position. Two pictures were presented as a connected action (as a GIF).

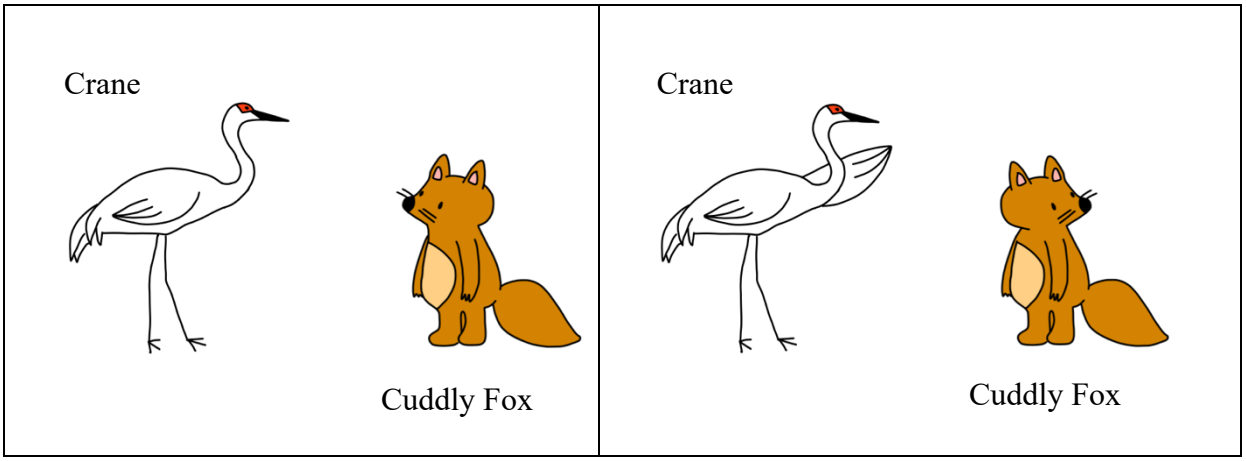


Figure A24: A scene in an experimental trial for the interactive speech task. The target word is *Cuddly* and located in IP-medial position. Two pictures were presented as a connected action (as a GIF).

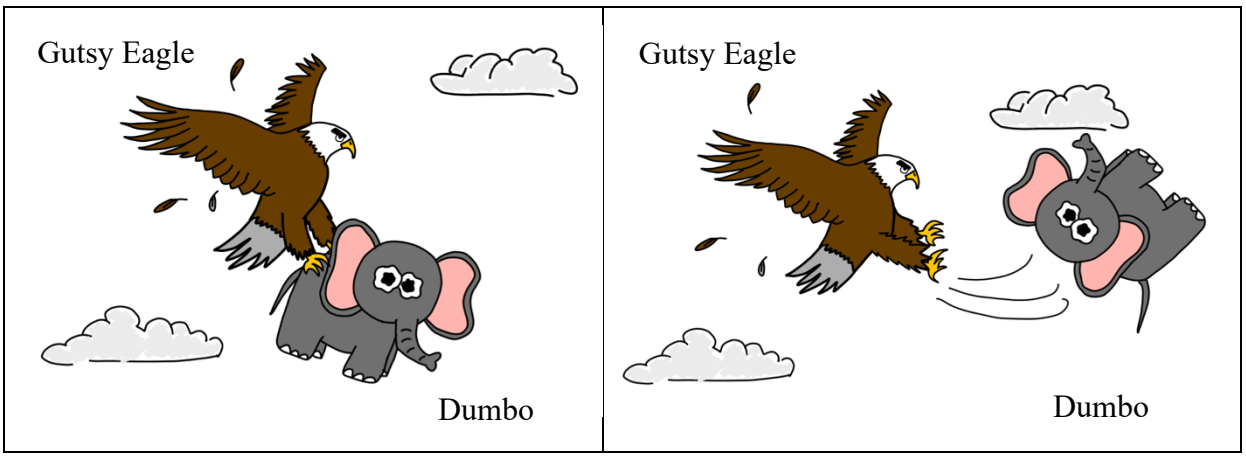


Figure A25: A scene in an experimental trial for the interactive speech task. The target word is *Gutsy* and located in IP-initial position. Two pictures were presented as a connected action (as a GIF).

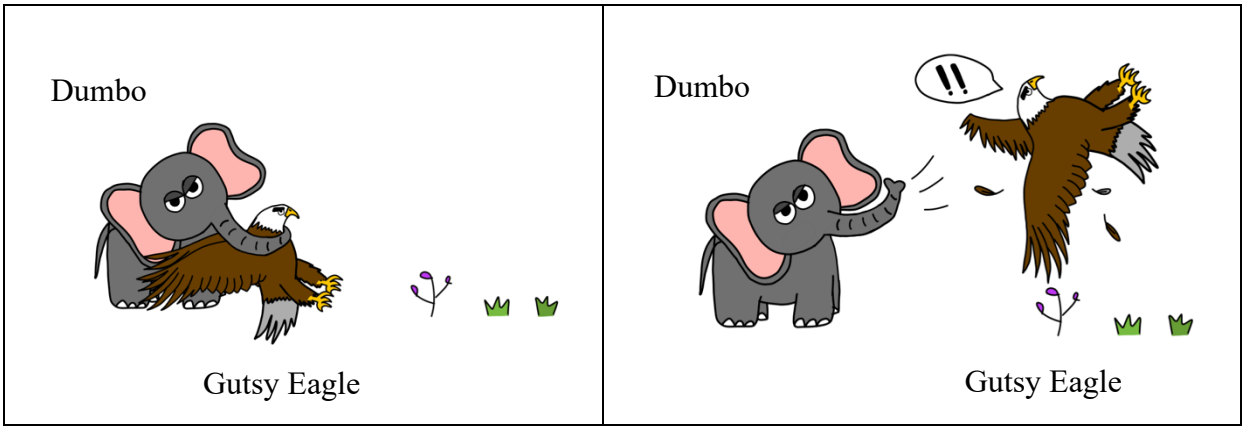


Figure A26: A scene in an experimental trial for the interactive speech task. The target word is *Gutsy* and located in IP-medial position. Two pictures were presented as a connected action (as a GIF).

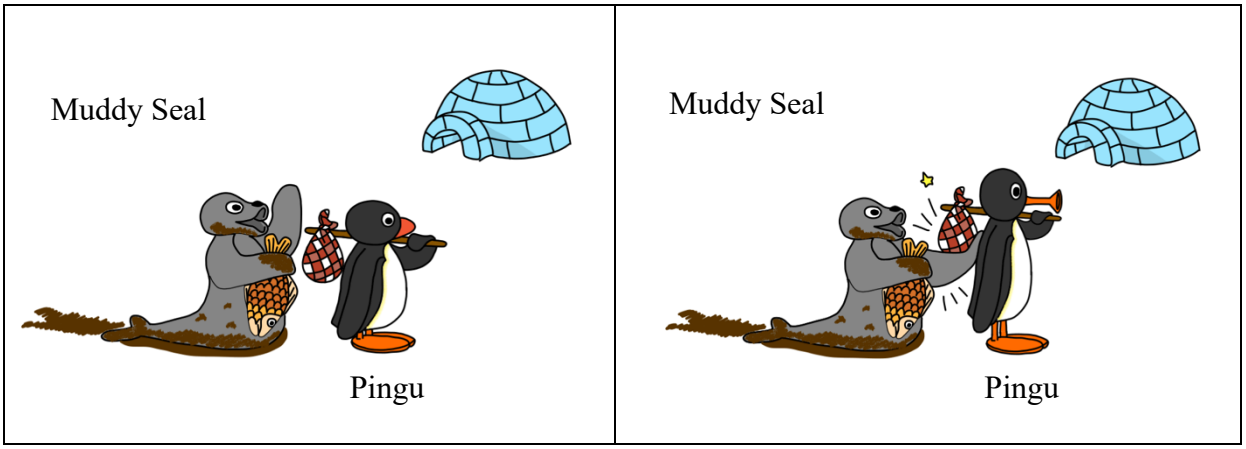


Figure A27: A scene in an experimental trial for the interactive speech task. The target word is *Muddy* and located in IP-initial position. Two pictures were presented as a connected action (as a GIF).

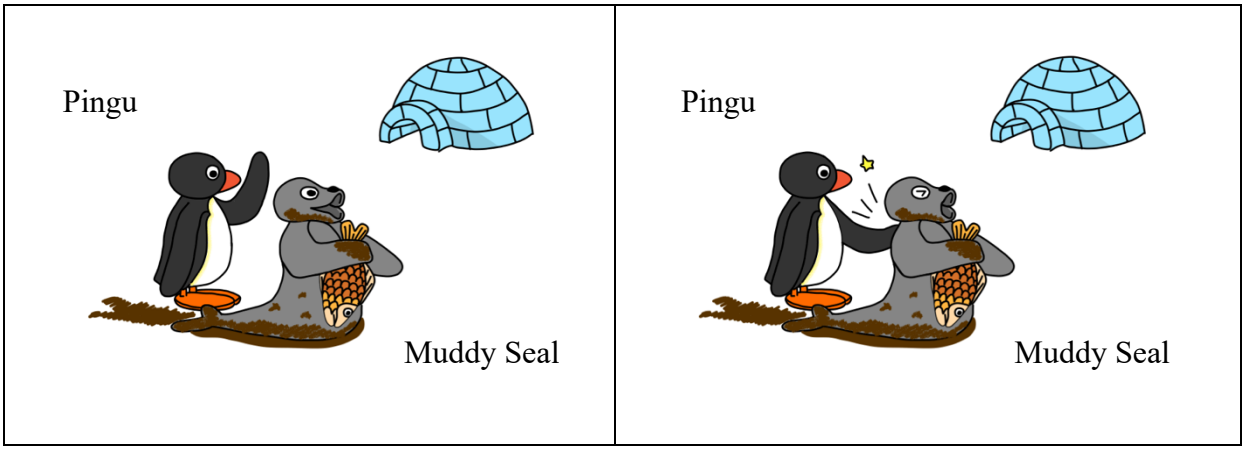


Figure A28: A scene in an experimental trial for the interactive speech task. The target word is *Muddy* and located in IP-medial position. Two pictures were presented as a connected action (as a GIF).

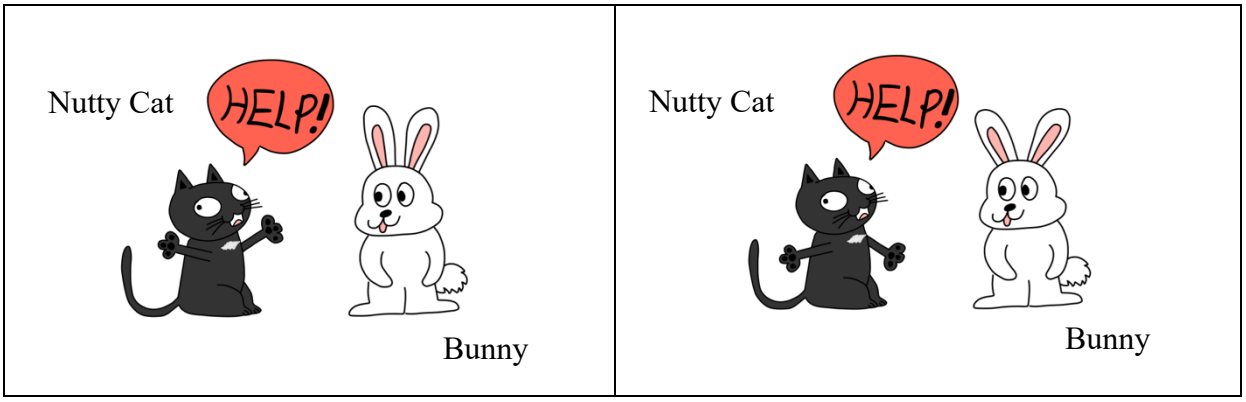


Figure A29: A scene in an experimental trial for the interactive speech task. The target word is *Nutty* and located in IP-initial position. Two pictures were presented as a connected action (as a GIF).

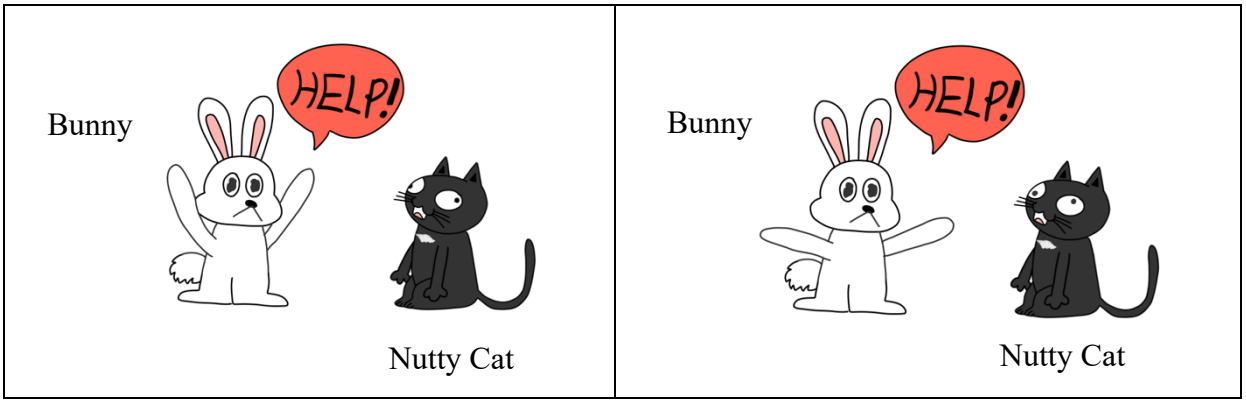


Figure A30: A scene in an experimental trial for the interactive speech task. The target word is *Nutty* and located in IP-medial position. Two pictures were presented as a connected action (as a GIF).