Journal of Phonetics 46 (2014) 128–146

Contents lists available at ScienceDirect

Journal of Phonetics

journal homepage: www.elsevier.com/locate/phonetics

Prosodic strengthening on the /s/-stop cluster and the phonetic implementation of an allophonic rule in English

Taehong Cho [a,*], Yoonjeong Lee [b], Sahyang Kim [c]

[a] Hanyang Phonetics and Psycholinguistics Lab, Department of English Language and Literature, Hanyang University, 222 Wangsimni-ro, Seongdong-gu, Seoul 133-791, Republic of Korea
[b] Department of Linguistics, University of Southern California, USA
[c] Department of English Education, Hongik University, Seoul, Republic of Korea

ARTICLE INFO

Article history:
Received 31 July 2013
Received in revised form
15 June 2014
Accepted 24 June 2014
Available online 2 August 2014

Keywords:
Prosodic strengthening
/s/-stop sequence
Allophonic rule
Prosodic boundary
Prominence
English
Phonetic feature enhancement
Laryngeal gesture
Articulatory Phonology

ABSTRACT

This acoustic study investigates effects of boundary and prominence on the temporal structure of **s#CV** and **#sCV** in English, and on the phonetic implementation of the allophonic rule whereby a voiceless stop after /s/ becomes unaspirated. Results obtained with acoustic temporal measures for /sCV/ sequences showed that the segments at the source of prosodic strengthening (*i.e.*, /s/ in **#sCV** for boundary marking and the nucleus vowel for prominence marking) were expanded in both absolute and relational terms, whereas other durational components distant from the source (*e.g.*, stop closure duration in **#sCV**) showed temporal expansion only in the absolute measure. This suggests that speakers make an extra effort to expand the very first segment and the nucleus vowel more than the rest of the sequence in order to signal the pivotal loci of the boundary *vs.* the prominence information. The potentially ambiguous **s#CV** and **#sCV** sequences (*e.g.*, *ice#can vs. eye#scan*) were never found to be neutralized even in the phrase-internal condition, cuing the underlying syllable structures with fine phonetic detail. Most crucially, an already short lag VOT in **#sCV** (due to the allophonic rule) was shortened further under prosodic strengthening, which was interpreted as enhancement of the phonetic feature {voiceless unaspirated}. It was proposed that prosodic strengthening makes crucial reference to the phonetic feature system of the language and operates on a phonetic feature, including the one derived by a language-specific allophonic rule. An alternative account was also discussed in gestural terms in the framework of Articulatory Phonology.

1. Introduction

One of the important goals of linguistic phonetics is an understanding of how abstract representations of sounds that are rather coarsely specified by the phonology of the language are phonetically shaped in a flow of speech. Numerous phonetic studies over the past several decades have unequivocally demonstrated that one of the pivotal factors that influence the phonetic shaping of individual segments is the prosodic structure of an utterance. Prosodic structure has been assumed to serve as a frame for articulation (*e.g.*, Beckman, 1996; Keating & Shattuck-Hufnagel, 2002), and to reflect *boundary marking* for grouping prosodic constituents (*i.e.*, *delimitative* function) and *prominence marking* for signaling information locus in the utterance by stressing particular prosodic constituents (*i.e.*, *culminative* function). Prosodic structuring of an utterance therefore modulates phonetic realization of individual segments not only at the phonetic level (determining their phonetic details), but also at the phonological level (constraining application of phonological rules).

The present study continues to explore the phonetics-prosody interface by examining how segments in **sCV** sequences ('C' = a voiceless stop) in English are realized along the temporal dimension as a function of marking prosodic boundary *vs.* prominence. It specifically examines **#sCV** (with the /s/-stop onset cluster as in 'eye#scan') and **s#CV** (with the post-lexically created /s/-stop cluster

* Corresponding author. Tel.: +82 2 2220 0746; fax: +82 2 2220 0741.
E-mail addresses: tcho@hanyang.ac.kr (T. Cho), yoonjeol@usc.edu (Y. Lee), sahyang@hongik.ac.kr (S. Kim).

http://dx.doi.org/10.1016/j.wocn.2014.06.003

as in ‘ice#can’). This will allow us to observe detailed phonetic manifestation of prosodic boundary and prominence on the same segmental sequences with different syllable structures as well as prosodic conditioning of the phonetic implementation of the phonological rule that a stop becomes unaspirated after /s/ in the **sCV** sequence.

1.1. Prosodic strengthening

Fine-grained but systematic phonetic variation of individual speech sounds as a function of prosodic structuring has often been discussed in terms of ‘prosodic strengthening’, which can be defined as spatio-temporal expansion of segments associated with important prosodic landmark locations such as edges of a prosodic domain and stressed/accented syllables (see Fletcher, 2010; Cho, 2011 for a review). The edge effects, known as final lengthening at the right edge and domain initial strengthening at the left edge of a prosodic domain, have been the loci of phonetic investigation by many researchers in efforts to illuminate how final and initial segments are realized in various contextually-adjusted phonetic forms to signal the boundary (prosodic grouping) information of prosodic structure (*e.g.*, Byrd & Saltzman, 2003; Byrd, Krivokapić, & Lee, 2006; Byrd, Lee, & Campos-Astorika, 2008; Cho, 2004, 2006, 2008; Cho & Keating, 2001, 2009; Fougeron, 2001; Fougeron & Keating, 1997; Keating, Cho, Fougeron, & Hsu, 2003; Krivokapić, 2007; Krivokapić & Byrd, 2012; *inter alia*). Another type of prosodic strengthening is prominence-induced strengthening driven by accent/stress marking, which renders a particular prosodic unit (usually a syllable or a word) phonetically more salient than other units in a phrase (Beckman, Edwards, & Fletcher, 1992; Cho, 2006; Cho & Keating, 2009; de Jong, 1995, 2004; Fowler, 1995; Lehiste, 1970; *inter alia*). This latter type of prosodic strengthening is often assumed to be linked with enhancement of distinctive features, bringing about maximizing phonemic (and lexical) contrasts, and is used, especially when realized with focus, as a diagnostic for what phonetic content is used to mark phonemic contrast in a given language (de Jong, 2004; de Jong & Zawaydeh, 2002).

Despite the growing body of studies on prosodic strengthening as discussed above, however, we are still left with a number of questions to be answered. Some of the questions are generally centered around the issues regarding how precisely the scopes (or the domains) of boundary *vs.* prominence marking are determined and how they are constrained by various factors such as allophonic rules of a given language and syllable structure. The goal of the present study is to explore these questions by investigating how the acoustic temporal realizations of individual segments in **sCV** sequences are modulated by boundary- and prominence-induced prosodic strengthening factors; how the distribution of temporal effects of prosodic strengthening is further conditioned by syllable structure (**s#CV** *vs.* **#sCV**)[1]; and how prosodic strengthening relates to phonetic implementation of a language-specific allophonic rule in English in connection with enhancement of phonetic feature. The acoustic temporal measures to be explored for the **sCV** sequence include /s/-duration, stop closure duration, VOT and vowel duration. Specific research questions that the present study particularly aims to answer are discussed in the following section.

1.2. Research questions

1.2.1. How does the boundary effect interact with the prominence effect on the temporal realization of #sCV?

Byrd and Choi (2010), in an electromagnetic midsagittal articulometer (EMMA) study, examined the boundary effect on #CCV sequences in English, showing that the stop as the second member of **#sC** may undergo domain-initial lengthening, but not as robustly as when it was initial in **s#C**. This was interpreted as supporting the predictions made by the theory of the π-gesture—*i.e.*, the boundary-induced lengthening effect is strongest at the boundary and becomes gradually weaker as a function of the segment's proximity to the boundary (*e.g.*, Byrd & Saltzman, 2003; Byrd et al., 2006; Krivokapić & Byrd, 2012; *cf.* Cho & Keating, 2009). The prominence effect, on the other hand, is assumed to be centered around the vowel (the nucleus) and to possibly spread leftward to adjacent segments in a gradually attenuating fashion (*e.g.*, Turk & White, 1999; White & Turk, 2010). Byrd and Choi, however, did not systematically take into account strengthening effects arising from prominence, leaving it unclear how the reported boundary effect would interact with prominence. We therefore extend Byrd and Choi by examining boundary-prominence interactions along the acoustic-temporal dimension.

While it is an open question exactly how boundary interacts with prominence in the temporal realization of **#sCV**, some predictions can be made as follows. Considering the segment's proximity to the boundary as discussed above, the first member of the cluster should demonstrate the strongest boundary lengthening effect, and the effect is expected to be *prominence-independent* (*i.e.*, regardless of whether the target-bearing word is accented or not). This would be because /s/ is initial (**#sCV**) and at the same time is farthest away from the nucleus vowel, the source of prominence-induced strengthening. Compared to /s/, the stop ‘C’, the second member of the cluster should then show a relatively weaker boundary effect as it is farther away from the boundary and at the same time closer to the locus of prominence (the nucleus vowel), being more vulnerable to the prominence effect. Given that the boundary effect on the consonantal articulation is often obscured by the prominence effect (Cho & Keating 2009; Cho, Lee, & Kim, 2011), we expect the boundary-induced lengthening effect on the stop in **#sCV** to be *prominence-sensitive*, such that it may be weakened or unobservable under accent, while it may surface robustly when it is free of the influence of prominence (*i.e.*, when unaccented). Lastly, the boundary lengthening effect on the following vowel is expected to be weakest or completely absent as the vowel in **#sCV** is not only farther away from the boundary but it also becomes the locus of prominence.

[1] The syllable structure factor is confounded with the location of the word boundary, but these factors cannot be teased apart as the location of the word boundary determines a simplex (s#C) vs. a complex (#sC) onset.

1.2.2. How does prosodic strengthening modulate the phonetic implementation of the allophonic rule by which a voiceless stop after /s/ becomes unaspirated?

Regarding prosodic modulation of phonological features in phonetic implementation, quite a few studies have suggested that prominence-induced strengthening has an effect of maximizing phonemic contrasts through phonetic featural enhancement in a language-specific way (*e.g.*, Cho, 2006; de Jong, 1995, 2004; de Jong & Zawaydeh, 2002). Cho and McQueen (2005) further explored the issue of language-specific prosodic enhancement of consonantal voicing features by examining how voiceless unaspirated stops in Dutch are phonetically realized in prosodic strengthening environments. They showed that the unaspirated stop (with a short lag VOT) was produced with an even shorter VOT both when it was stressed/accented and when it was domain-initial. Based on these findings that did not follow the principle of phonetic polarization along the VOT continuum (*i.e.*, a longer VOT for a voiceless stop would maximize its contrast with its voiced counterpart), they proposed that prosodic strengthening operates on a language-specific *phonetic feature* which is {−spread glottis} (or {voiceless unaspirated}) for the Dutch voiceless stop and {+spread glottis} (or {voiceless aspirated}) for the English voiceless stop, resulting in asymmetric strengthening effects on VOT—*i.e.*, shortening in Dutch and lengthening in English. Note that curly brackets '{}' are used here to refer to phonetic features. The phonological feature such as [+/−voice] may be used to refer to the phonological contrast between voiced and voiceless stops within each language, whereas phonetic features such as {voiceless unaspirated} or {voiceless aspirated} are required to explain the difference in phonetic implementation in voiceless stops between languages (*e.g.*, Cho & Ladefoged, 1999; Keating, 1984, 1990).

The present study extends Cho and McQueen's study to a case in English, in which the phonetic feature {voiceless unaspirated} (henceforth {vl. unasp.})[2] is used for the allophonic variant of the voiceless stop that is derived from a phonological (allophonic) rule in **#sCV** context. Two possibilities will be construed as working hypotheses as laid out below.

The first possibility is that VOT will be lengthened in prosodic strengthening environments as a consequence of the strengthened laryngeal abduction gesture. This will be referred to as *the passive VOT lengthening hypothesi*s. Under this hypothesis, VOT is taken not as something that is directly modulated by prosodic strengthening, but as an output variable stemming from the magnitude of laryngeal abduction (see below for further discussion on this point). The allophonic rule that makes the stop unaspirated after /s/ may be understood as the inevitable consequence of laryngeal settings during the production of the /s/-stop cluster. Despite its separate segmental makeup, the tautosyllabic /s/-stop cluster is known to be produced with a single peak of glottal abduction (*e.g.*, 'eye#scan'), although separate glottal abduction peaks are usually observed with the heterolexical /s/-stop cluster that spans a lexical boundary (*e.g.*, 'ice#can') (*e.g.*, Yoshioka, Löfqvist, & Hirose, 1981). Crucially, for **#sC** the peak is known to be aligned to the first member (/s/) of the onset cluster, and by the time the following stop is released, the glottis is already at the final phase of the glottal abduction–adduction cycle, which is responsible for unaspiratedness (*i.e.*, a short lag VOT) of the stop (see Löfqvist, 1995 and Hoole, 1999 for reviews).

The size of a laryngeal abduction gesture is an important determinant for VOT[3]—*i.e.*, the larger the laryngeal abduction gesture, the longer the VOT (*e.g.*, Cooper, 1991; Goldstein, 1992; Pirrehumbert & Talkin, 1992). Goldstein (1992), for example, noted that '[t]he size (and timing) of a laryngeal gesture coordinated with an oral closure will determine the stop's voice-onset time (VOT)…' (p. 212). This observation was further supported by results of an articulatory (transillumination) study by Cooper (1991), which showed a positive relationship between the size of glottal opening and VOT. A possible explanation is as follows. When the vocal folds are wider apart, the displacement to be made from abduction to adduction will also be larger, which, all else being equal, requires a longer time for the vocal folds to be adducted for voicing. Moreover, a more widely open glottis is likely to allow for a larger amount of airflow through the glottis when the stop is released. This will increase the impedance to the glottal adduction, possibly causing some delay in allowing an adequate transglottal pressure for the initiation of the vocal fold vibration. As a result, an elongated VOT lag is likely to occur. It follows then that VOT for the unaspirated stop after /s/, all else being equal, may well be elongated in an environment in which the magnitude of glottal abduction increases—*i.e.*, as a necessary consequence of the strengthened glottal abduction gesture. It is a widely held view that the glottal abduction gesture is indeed strengthened (with an increased magnitude) in prosodically strong conditions (*e.g.*, Cooper, 1991; Goldstein, 1992; Pierrehumbert & Talkin, 1992; see Fougeron, 1999, for a review), which may account for both boundary- and prominence-induced lengthening of VOT for singleton voiceless stops commonly observable across languages (*e.g.*, Cho & Keating, 2001, 2009; Keating, et al., 2003). By the same logic, we might expect the laryngeal gesture for the /s/-stop cluster to be strengthened, which, all else being equal, would have an effect of lengthening (rather than shortening) VOT in both domain-initial and accented syllables.

An alternative possibility is that VOT will be directly modulated to be shortened in prosodic strengthening environments in a way to enhance the phonetic feature {vl. unasp.}. A voiceless stop assigned with the phonetic feature {vl. unasp.} is phonetically implemented with a short lag VOT (*e.g.*, Keating, 1984, 1990). It is therefore reasonable to assume that enhancement of {vl. unasp.} involves shortening of VOT which reinforces the phonetic content of the feature and the phonetic clarity of the segment associated with the feature. This will be called *the phonetic feature enhancement hypothesis*. As discussed above, Cho and McQueen (2005) showed that Dutch voiceless stops are produced with shortened VOTs in both domain-initial and accented syllables. Given that VOT may be influenced by the magnitude of the glottal abduction, one way of explaining the shortening of VOT might be that it stems from weakening of the glottal abduction gesture. Such a laryngeal reduction, however, runs counter to general prosodic strengthening

[2] The phonetic feature {vl. unasp.} was originally used in Keating (1984), which was later replaced by {-spread glottis} in Keating (1990) in order to capture the laryngeal configuration at the time of the stop release. In the present study, however, we will use {vl. unasp.} as employed in Cho and Ladefoged (1999) whose phonetic realization is more transparently translated along the VOT continuum.

[3] VOT may also be conditioned by timing between peak glottal abduction and the stop release. This, however, appears to be less relevant in the case of the /s/-stop cluster as peak glottal abduction is generally tightly aligned with /s/ (Browman & Goldstein, 1986; Goldstein, 1992; Hoole, 1999).

patterns observed in the literature. The most parsimonious account then seems to be that VOT is directly modulated by prosodic strengthening, still allowing for strengthening of the glottal abduction gesture as a characteristic of prosodic strengthening. VOT can be indeed shortened by virtue of an enhancement of the phonetic feature {vl. unasp.} even in an environment in which the glottal abduction gesture is strengthened. This is made possible by adopting the notion of 'Articulatory VOT,' which was proposed to account for a wide range of VOT for the same voiceless stop category observable across languages (Cho & Ladefoged, 1999; see also Ladefoged & Cho, 2001). It was defined as the articulatory timing between the stop release gesture and the laryngeal gesture responsible for vocal fold vibration. While 'Articulatory VOT' adopted the notion of gestural timing advanced in Articulatory Phonology (Browman & Goldstein, 1986, 1990, 1992), departing from Articulatory Phonology (in which VOT is rather passively determined as a result of the laryngeal abduction-addition cycle), Cho and Ladefoged proposed that the timing of the laryngeal gesture responsible for initiating voicing is something that can be controlled by the speaker. In this way, VOT may be fine-tuned in a language-specific way (and encoded in the phonetic grammar of the language), accounting for seemingly arbitrary, but linguistically relevant, distributions of VOT across languages. In the same fashion, VOT may be modulated directly by prosodic strengthening in a way to enhance the phonetic feature {vl. unasp.}, resulting in shortening VOT as in Dutch even when the glottal gesture is strengthened. Likewise, if prosodic strengthening operates on the phonetic feature {vl. unasp.} which is allophonically derived by the language-specific allophonic rule in English, VOT in **#sCV** is expected to be shortened under prosodic strengthening in much the same way as the phonetic feature for voiceless stops was enhanced in Dutch.

1.2.3. How does the temporal organization of #sCV differ from that of s#CV?

Comparing the durational patterns in **#sCV** with those in **s#CV** (*e.g.*, 'eye#scan' vs. 'ice#can') especially in the phrase-internal context will allow us to observe the extent to which the two sequences may be temporally neutralized, given the possibility that /s/ in **s#CV** is syllabified into the following onset. While the domain of syllabification in English is generally thought to be within a prosodic word (Kahn, 1976; Nespor & Vogel, 1986), the present study allows us to explore the (re)syllabification issue in connection with the domain of the phonological (allophonic) rule that makes a voiceless stop unaspirated after /s/. If /s/ were syllabified into the following onset across a lexical boundary, the rule may also apply to the heterolexical **s#C** sequence, resulting in VOT shortening for the following stop comparable to that in **#sC**. But it is less likely so when we consider aforementioned differential laryngeal settings for **#sC** (with one glottal abduction peak) *vs.* **s#C** (with separate glottal abduction peaks). Based on the laryngeal characteristics, Browman and Goldstein (1986) proposed that only one glottal abduction gesture may be associated with the tautosyllabic (and tautolexical) /s/-stop cluster (**#sC**) in word-initial position in English, while the heterolexical /s/-stop cluster (**s#C**) that spans a lexical boundary may have two abduction gestures, one coordinated with the midpoint of frication for /s/ and the other with the release of the stop (*the lexically specified laryngeal gesture hypothesis*). Under this hypothesis, the lexically-specified dual glottal abduction gestures for the heterolexical cluster are to be preserved on the surface even in the potentially neutralizing phrase-internal condition. In such a case, the stop in **s#C** will still be produced with some degree of aspiration, which will be reflected in a substantially longer VOT for the stop in **s#C** compared to the case with **#sC**.

The comparison of prosodic strengthening effects on **#sCV** *vs.* **s#CV** will also contribute to a further understanding of various other aspects of boundary- *vs.* prominence-induced strengthening in connection with syllable structure. First, we will be able to examine how the temporal structure of preboundary /s/ in **s#CV** (compared to /s/ in **#sCV**) is influenced by accentuation of the following syllable across a prosodic boundary. Some previous studies have suggested that accentual lengthening may be extended leftwards across a syllable boundary and possibly across a lexical word boundary, even though its effect would be more attenuated compared to accentual lengthening within a syllable (*e.g.*, Turk & White, 1999; see also White & Turk, 2010 and Cho, Kim, & Kim, 2013, for related discussion). It would be therefore interesting to test the lexical boundary effect on the leftward accentual lengthening in **s#C** *vs.* **#sC** particularly in the potentially neutralizing Word boundary condition. Another question to be explored is whether there is an asymmetric boundary effect on the final /s/ in **s#CV** *vs.* the initial /s/ in **#sCV**. As both /s/s are immediately adjacent to the boundary, one may predict no difference in boundary lengthening between them (*e.g.*, Byrd, Lee, Riggs, & Adams, 2005), but the difference between preboundary *vs.* postboundary /s/s may come about, given that the word-level onset/coda asymmetry has often been observed in the literature (*e.g.*, Byrd, 1996; Keating, Wright, & Zhang, 1999) and given that /s/ is likely to be longer in phrase-final position than in any other position in the utterance as noted by Klatt (1976).

1.2.4. To what extent can temporal variation of #sCV under prosodic strengthening be understood in terms of relational invariance?

Finally, the present study explores the nature of temporal variation arising with prosodic strengthening in terms of the maintenance of temporal relation of the segment duration relative to the total **#sCV** sequence duration across different prosodic conditions. Some researchers have proposed that speakers tend to make efforts to preserve the segments' pronunciation 'norm' across conditions that affect the temporal realization of the segment (*e.g.*, Solé, 2007; Smiljanić & Bradlow, 2008). For example, Smiljanić and Bradlow (2008) found that although absolute VOT may vary across speaking styles in both English and Croatian, its proportion to the entire stop duration remained stable. In a similar vein, Solé (1997) proposed that 'cues that are deliberately controlled by the speaker to convey linguistic information to the listener should be present in the same proportion across durational differences in rate, stress, or syllable types' (p. 305). In the present study we test this *relational invariance hypothesis* by exploring to what extent temporal variation with prosodic strengthening can be understood as maintaining such relational invariance. To test this, we will examine how prosodic strengthening affects the temporal realization of **#sCV** (with a complex onset) in relational terms (*i.e.*, the magnitude of the duration change relative to the duration change of the entire **#sCV** sequence) in comparison with patterns obtained with absolute measures. Under *the relational invariance hypothesis*, it is predicted that any durational component for the **#sCV** sequence (*e.g.*, /s/-duration, stop closure duration, VOT, vowel duration) that is to undergo temporal modification in the absolute measure will turn out to remain unchanged when it is considered in the relational measure. Alternatively, however, durational effects that may go beyond the maintenance of relational invariance may be observable if speakers are to make additional efforts, for example, for exaggerating the boundary effect on the very initial segment and the accent effect on the vowel which is the locus of prominence. In such a case, shortening of VOT (predicted by the phonetic feature enhancement) may also exceed the level of maintaining relational invariance, which may be seen as support for the view that VOT is directly modulated by prosodic strengthening.

2. Method

2.1. Participants and recording

Ten native speakers of American English, five females and five males, participated in the experiment. To reduce possible influence of dialectal differences, our speaker pool was narrowed to the speakers from the Midwest and the West Coast, excluding those from the East Coast and the southern parts of the United States. The participants, all in their 20s, were either exchange students or English teachers working in Seoul at the time of recording. All participants were paid for their participation and were naïve as to the purpose of the present study. The speech data were recorded in a soundproof booth at the Hanyang Phonetics and Psycholinguistics Lab, with a Tascam HC-P2 digital recorder and a SHURE KSN44 condenser microphone at a sampling rate of 44 kHz.

2.2. Test sentences and procedure

The test sequence **sCV** contained a voiceless stop for 'C' varying in place of articulation (/p,t,k/) in English. It was embedded in a two-word sequence, in which /s/ was either word-final as in ‘ice#can’ (/s#k/) or word-initial as in ‘eye#scan’ (/#sk/). Table 1 shows a list of two-word sequences in pairs (/s/-final *vs.* /s/-initial) tested in the experiment: for each /s/-stop sequence (with three places of articulation) two pairs were included, giving a total of six pairs.

The two-word sequences were then placed in carrier sentences in such a way that three experimental factors were manipulated: (1) Prosodic boundary (Intonational Phrase (IP) vs. Word (Wd) boundaries), (2) Accent (Accented *vs.* Unaccented), and (3) Syllable structure (**s#CV** *vs.* **#sCV**). Table 2 shows an example set of /sk/-bearing test sentences. Two test sentences were created for each prosodic condition so that speakers could contrast the differences between the two sentences and produce the intended prosodic structure. For example, to induce an accented condition, speakers were asked to produce two sentences in a row (as in Table 2a), so that the underlined stop-initial word ‘CAN’ in the second sentence could be accented with contrastive focus, being contrastive with ‘PAN’ in the first sentence. For the unaccented condition (as in Table 2b), the underlined critical word ‘can’ in the second sentence carried old information (as it was repeated), so that it was likely to be unaccented, while the contrastive focus was placed somewhere else in the sentence. Note that in order to examine the accent effects of the critical (second) word without being influenced by the prominence of the immediately preceding context word, the first word in the two-word sequence (‘ice’ in ‘ice#can’) was controlled to be unaccented, again by placing another contrastive focus somewhere else before the first word.

To obtain different boundary conditions (IP *vs.* Word), two sentence types were employed. As in Table 2a,b,e,f, for the IP boundary condition, a complex sentence was used which contained two clauses (*e.g.*, *After they say* ‘*ice*’, # ‘*can again*’ *will be the next phrase to say*), so that speakers could naturally be guided to put a phrase boundary between the critical two words at the juncture of the two clauses. For the Wd boundary condition (Table 2c,d,g,h), the two-word sequence was placed inside an infinitive clause (*e.g.*, *To say* ‘*ice can again*’ *with me is going to be easy*), which would help speakers to produce the critical two words connectedly without putting a prosodic phrase boundary between them.

Before data collection, speakers had a short practice session by reading out the test sentences, in order to be familiarized with intended renditions for various prosodic conditions. In the experiment, the participants were presented with test sentences on a computer screen and asked to read the sentences aloud as comfortably and naturally as possible. The entire corpus was repeated three times in a randomized order. In total, 1440 tokens were collected and analyzed in the present study (2 prosodic boundaries (IP *vs.* Wd) × 2 accent conditions (Accented *vs.* Unaccented) × 2 syllable structures (/s/-final, **s#C** *vs.* /s/-initial, **#sC**) × 3 places of articulation (/sp/, /st/, *vs.* /sk/) × 2 comparison pairs for each condition (*e.g.*, ‘ice#can’–‘eye#scan’ and ‘rice#cone’–‘rye#scone’) × 3 repetitions × 10 speakers).

Table 1
A list of two-word pairs with **sCV** test sequences.

Place of articulation for C	s#CV–#sCV
/sp/	‘dice # pot’–‘dye # spot’
	‘lace # pin’–‘lay # spin’
/st/	‘peace # tax’–‘pea # stacks’
	‘base # tone’–‘bay # stone’
/sk/	‘ice # can’–‘eye # scan’
	‘rice # cone’–‘rye # scone’

Table 2
An example set of test sentences containing s#CV in IP/Accented (a), IP/Unaccented (b), Wd/Accented (c), Wd/Unaccented (d) conditions, and #sCV in IP/Accented (e), IP/Unaccented (f), Wd/Accented (g), Wd/Unaccented (h) conditions. The critical two-word sequences are underlined, and accented items are in bold uppercase letters.

(Vs#CV)
(a) IP/Accented (s#C)
After **THEY** say 'ice', '**PAN** again' will be the next phrase to say.
But after **WE** say <u>'ice', # '**CAN**</u> again' will be the next phrase to say.

(b) IP/Unaccented (s#C)
After **THEY** say 'ice', 'can again' will be the **NEXT** phrase to say.
But after **WE** say <u>'ice', # 'can</u> again' will be the **FINAL** phrase to say.

(c) Wd/Accented (s#C)
To say 'ice **PAN** again' with me is going to be **DIFFICULT**.
But to say <u>'ice # **CAN**</u> again' with me is going to be **EASY**.

(d) Wd/Unaccented (s#C)
To say 'ice can again' with **JOHN** is going to be **DIFFICULT**.
But to say <u>'ice # can</u> again' with **ME** is going to be **EASY**.

(V#sCV)
(e) IP/Accented (#sC)
After **THEY** say 'eye', '**SPAN** again' will be the next phrase to say.
But after **WE** say <u>'eye', # '**SCAN**</u> again' will be the next phrase to say.

(f) IP/Unaccented (#sC)
After **THEY** say 'eye', 'scan again' will be the **NEXT** phrase to say.
But after **WE** say <u>'eye', # 'scan</u> again' will be the **FINAL** phrase to say.

(g) Wd/Accented (#sC)
To say 'eye **SPAN** again' with me is going to be **DIFFICULT**.
But to say <u>'eye # **SCAN**</u> again' with me is going to be **EASY**.

(h) Wd/Unaccented (#sC)
To say 'eye scan again' with **JOHN** is going to be **DIFFICULT**.
But to say <u>'eye # scan</u> again' with **ME** is going to be **EASY**.

2.3. Measurements

In order to assess how the temporal structure of the **sCV** sequence is modified by prosodic strengthening factors (boundary and prominence) in different syllable structures, all the durational components of the sequence were measured: */s/-duration*, *Stop closure duration*, *VOT* and *V-duration*.

2.3.1. /s/-duration

The duration of /s/ was measured as indicated by high frequency noises (usually distributed above 5000 Hz) in the spectrogram and visually confirmed by the presence of aperiodic noises in the waveform. It will show how the boundary-induced lengthening effect differs between /s/-final (preboundary) and /s/-initial (postboundary) conditions, and how the duration of /s/ in the potentially neutralizing phrase-internal (Wd boundary) condition varies as a function of whether /s/ is underlyingly tautosyllabic with the following stop or not (**s#C** *vs.* **#sC**). In particular, /s/ in **#sC**, being initial, is expected to undergo the most robust boundary-induced lengthening independent of prominence (the *prominence-independent* boundary effect), and /s/ in the (phrase-internal) Wd condition is expected to show some durational difference in **s#C** and **#sC** if the former is not fully resyllabified.

2.3.2. Stop closure duration (CD)

CD for the stop in the /s/-stop sequence was taken from the cessation of the preceding aperiodic noises (*i.e.*, the preceding /s/) to the beginning of the stop burst seen in the spectrogram, which corresponded to a complete silence seen in the waveform. For **#sCV**, CD was measured in both the IP and the Wd boundary conditions, but for **s#CV**, it was measured only in the Wd boundary condition because it was difficult to be disentangled from a pause that might arise at an IP boundary. While this measure will show how stop closure duration is modified in various prosodic and syllabic structure conditions, it will be particularly interesting to examine the effect of its proximity to the source of the strengthening: Boundary-induced lengthening of CD in **#sCV** may be observable only when unaccented, showing the *prominence-sensitive* boundary effect as it is not strictly at the boundary while being adjacent to the source of the prominence effect.

2.3.3. Voice onset time (VOT)

VOT for the stop was measured from the time of the acoustic release burst to the onset of voicing in the following vowel. Prosodically-conditioned variation in VOT will not only help understanding the temporal modification of the **sCV** sequence, but it will also be used as a measure for testing enhancement patterns of the stop voicing feature—*i.e.*, whether it will be shortened, enhancing the hypothesized phonetic feature {vl. unasp.} (*the phonetic feature enhancement hypothesis*), or lengthened due to strengthening of the glottal abduction gesture (*the passive VOT lengthening hypothesis*). It will also be important to test whether VOT will indeed be substantially longer in **s#C** than for **#sC** even in the potentially neutralizing Wd boundary condition (as predicted by *the lexically specified laryngeal gesture hypothesis*—*i.e.*, dual gestures are specified for the heterolexical **s#C** cluster and a single gesture for the tautolexical **#sC** cluster).

2.3.4. Total voiceless interval (TVI)

TVI was calculated as the sum of voiceless durational components (/s/-duration + CD + VOT). (As was the case with CD, TVI was not measured for **s#CV** with an IP boundary due to its inseparability from a possible pause.) TVI was used as a measure of voicelessness, which is assumed to be strengthened in prosodic strengthening environments presumably to enhance the CV contrast (*e.g.*, Pierrehumbert & Talkin, 1992). TVI may also be taken as an indirect measure for estimating the magnitude of glottal abduction. As discussed in Section 1, the period of the laryngeal abduction–adduction cycle may be positively correlated at least in part with the magnitude of the largyneal abduction gesture. Furthermore, the voiceless interval for the consonant cluster may be generally maintained during the laryngeal abduction–adduction cycle, and voicing initiates upon or near the completion of this cycle (see Hoole, 1999, for a related discussion). It is therefore plausible that the total voiceless interval of the voiceless consonant cluster is elongated along with an increased magnitude of laryngeal abduction (independently of whether VOT turns out to be shortened *vs.* lengthened in **#sC** under prosodic strengthening). It is also expected to be longer for **s#C** than for **#sC** (testable in the Wd condition) if two abduction gestures are realized for the heterolexical **s#C** sequence as opposed to a single gesture for the tautolexical **#sC,** as predicted by *the lexically specified laryngeal gesture hypothesis.*

2.3.5. V-duration

The duration of the vowel in the postboundary word was measured from the onset to the offset of voicing for the vowel. V-duration was included to test to what extent the temporal structure of the vowel after the test consonant sequence is influenced by boundary- *vs.* prominence-induced strengthening. The most robust lengthening effect on V is expected to stem from accent (which is the locus of prominence) while the boundary effect is expected to be minimal as V is distant from the boundary. The boundary effect on V, if it exists, is expected to be smaller in **#sCV** than in **s#CV** as V is separated from the boundary by two segments in **#sCV** but by one in **s#CV**.

2.3.6. Relational measures

For the **#sCV** sequence (with a complex onset) for which CD was measurable in the acoustic dimension, the results obtained with absolute measures will be compared with those in relational terms, expressed as percent proportion of each absolute measure relative to the entire **#sCV** sequence duration in order to test *the relational invariance hypothesis* as discussed in Section 1. The relational measures will be indicated by adding '%' as in %-*/s/-duration,* %-*CD,* %-*VOT,* and %-*V-duration*. Under *the relational invariance hypothesis*, any temporal variation due to prosodic strengthening observable in absolute measures will turn out to be invariant in relational measures. Alternatively, durational effects may go beyond the maintenance of relational invariance if speakers make additional efforts for exaggerating temporal effects on particular durational components (*e.g.*, at the source of prosodic strengthening or to enhance a phonetic feature). (Note that the #CV sequence will not be examined in relational terms as the stop closure duration for a phrase-initial C cannot be reliably measured due to its inseparability from a possible pause.)

2.4. Statistical analyses

The systematic influence of Boundary, Prominence and Syllable structure factors on the acoustic realization of **sCV** was statistically evaluated, based on repeated measures analyses of variance (RM ANOVAs). The data were averaged over repetitions across items with different stops in order to provide each speaker's representative value per condition. The RM ANOVAs performed in the present study employed three within-subject factors: Boundary (IP *vs.* Wd), Accent (Accented *vs.* Unaccented), and Syllable structure (**s#CV** *vs.* **#sCV**). Whenever there was an interaction between factors, posthoc *t*-tests were carried out in order to understand the source of the interaction. Regardless of interaction effects, however, additional t-tests for within-factor comparisons were performed with the critical Wd boundary data. Although some of these additional posthoc analyses may not be justifiable especially when there was no relevant between-factor interaction involving Boundary, it was done in order to take a closer examination of the temporal organization of a potentially neutralizing pair ('ice#can' vs. 'eye#scan') (see Ott & Longnecker, 2001 for the necessity of such planned comparisons). It turned out that there was only one critical case in which the results of such further analyses were worth reporting, for which some words of caution were provided. Effect sizes were estimated by conducting η^2 (eta^2)

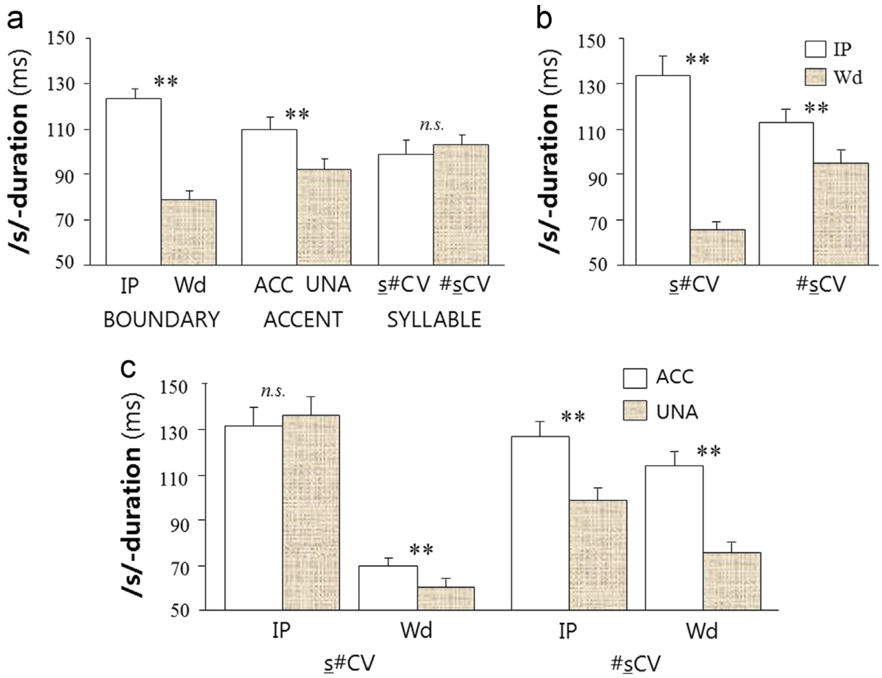

Fig. 1. /s/-duration: Main effects (a), Boundary × Syllable interaction (b), and Boundary × Accent × Syllable interaction (c). ('*' refers to $p<.05$ and '**', $p<.01$.)

analyses (Sheskin, 2000), which were often useful in discussing the relative robustness of observed effects. In all cases, *p*-values less than .05 were considered significant.

3. Results

In this section, we will first report results on effects of Boundary (IP *vs.* Wd), Accent (Accented *vs.* Unaccented) and Syllable (**#sCV** *vs.* **s#CV**) on absolute measures (/s/-duration, Stop Closure Duration, VOT, and Vowel Duration), and then results on effects of Boundary and Accent on relational measures for the **#sCV** sequence.

3.1. Effects of boundary, accent and syllable structure

3.1.1. /s/-duration

Both Boundary and Accent showed main effects on /s/-duration. As shown in Fig. 1a, /s/ was longer for IP than for Wd ($F[1,9]=128.6$, $p<.001$), and longer in the accented than in the unaccented condition ($F[1,9]=140.4$, $p<.001$). Crucially, there was no significant interaction between Boundary and Accent, showing a predicted *prominence-independent* boundary effect on the very initial segment /s/ at the source of the boundary strengthening while being distant from the source of prominence.

There was no main effect of Syllable, but both Boundary and Accent interacted with Syllable ($F[1,9]=30.9$, $p<.001$; $F[1,9]=83.7$, $p<.001$, respectively). The Boundary by Syllable interaction showed a greater degree of /s/ lengthening IP-finally than IP-initially: As can be seen in Fig. 1b, the boundary lengthening effect on /s/ was asymmetrically larger domain-finally than domain-initially (**s#CV**: mean diff. 68 ms, $t(9)=9.5$, $p<.001$, $\eta^2=.91$; **#sCV**: mean diff. 18 ms, $t(9)=4.2$, $p<.005$, $\eta^2=.67$).

The Accent by Syllable interaction showed that the accent effect was further conditioned by /s/'s lexical (syllable) affiliation. That is, a significant accentual lengthening was found on /s/ in **#sCV** ($t(9)=11.1$, $p<.001$, $\eta^2=.94$), but not on /s/ across a lexical boundary in **s#CV** ($t(9)=1.7$, *n.s.*). However, although there was no further significant interaction involving Boundary, when considered only in the critical Wd condition (in which **s#CV** and **#sCV** are potentially ambiguous), the accent effect on /s/ turned out to be significant across a lexical boundary in **s#CV** as shown in Fig. 1c. As predicted, this leftward spreading of accentual lengthening across a Wd boundary, though not very robust, can be interpreted as being substantially attenuated compared to the tautosyllabic accentual lengthening in **#sCV** (**s#C**: mean diff. 9.1 ms, $t(9)=4.3$, $p<.01$, $\eta^2=.68$; **#sC**: mean diff. 38.7 ms, $t(9)=130.29$, $p<.001$, $\eta^2=.94$). The only condition that showed no accentual lengthening on /s/ was in **s#CV** across an IP boundary which completely blocked the leftward spreading of accentual lengthening (as shown in the left of Fig. 1c).

3.1.2. Stop closure duration (CD)

3.1.2.1. CD in the Wd condition. As noted earlier, because a pause and CD were acoustically inseparable in **s#CV** (#=IP), effects of Accent and Syllable were tested only in the Wd condition (without the Boundary factor). As shown in Fig. 2a, CD was longer when accented *vs.* unaccented, showing accentual lengthening of the stop ($F[1,9]=54.2$, $p<.001$). As for the syllable structure effect, CD

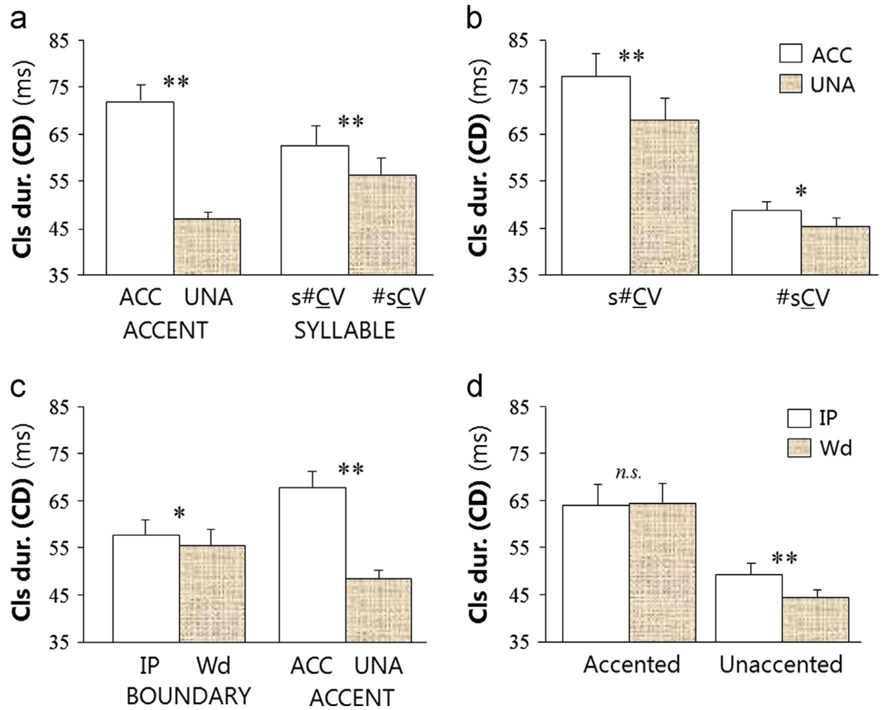

Fig. 2. Stop closure duration (CD): Main effects only with Wd (a), Accent × Syllable interaction only with Wd (b), main effects only for #sCV (c), and Boundary × Accent interaction for #sCV (d). ('*' refers to $p<.05$ and '**', $p<.01$.)

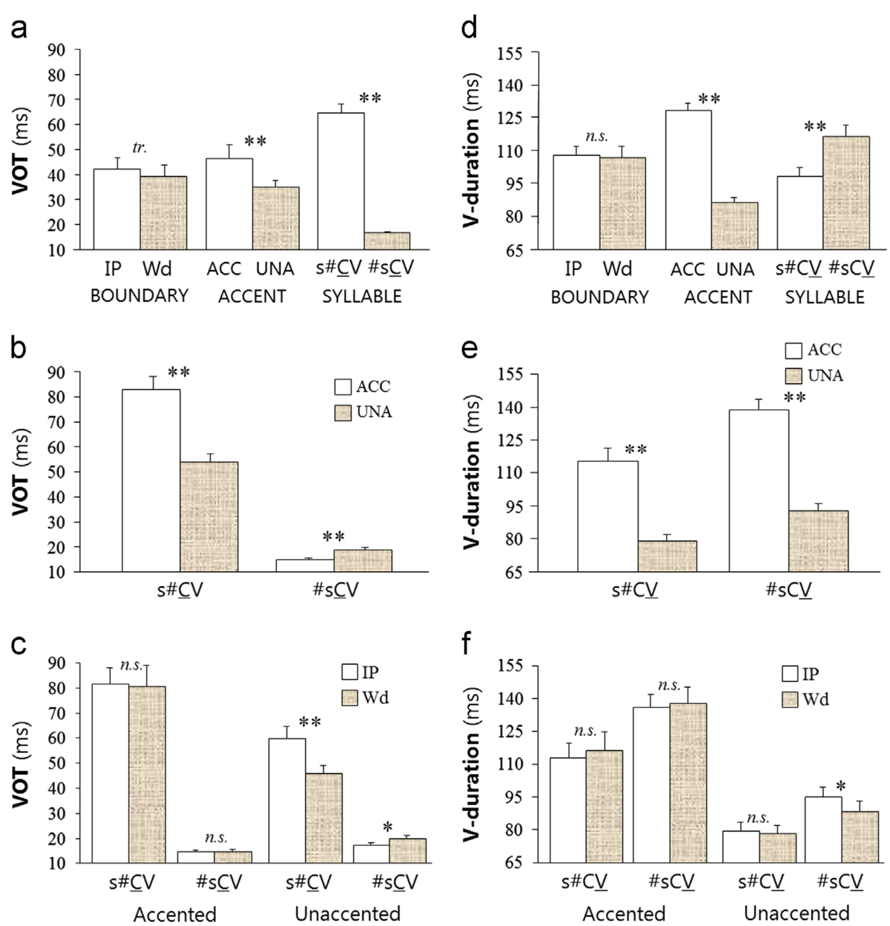

Fig. 3. Main effects and interactions on VOT (a–c) and on postboundary V-duration (d-f). ('*' refers to $p<.05$; '**', $p<.01$; and '*tr*', $p<.06$).

was also longer in the simplex onset (**s#CV**) than in the complex onset (**#sCV**) even in the (potentially neutralizing) Wd boundary condition ($F[1,9]=24.4$, $p<.001$), showing no evidence for resyllabification of /s/ in **s#CV**. (Note that there was a significant interaction between Accent and Syllable ($F[1,9]=6.2$, $p<.05$), but as can be seen in Fig. 2b the interaction was simply due to the fact that the effect of Accent was more robust in **s#CV** than in **#sCV**.)

*3.1.2.2. CD in **#sCV**.* The stop closure data were further analyzed with a two-way RM ANOVA (with Boundary and Accent) just in the **#sCV** condition (in which CD was measurable in both IP and Wd conditions, thus making it possible to test the Boundary effect). As shown in Fig. 2c, CD as the second member of the cluster (**#sCV**) was longer after an IP than after a Wd boundary ($F[1,9]=10.1$, $p<.05$), and longer in the accented than in the unaccented condition ($F[1,9]=24.4$, $p<.001$). However, Boundary interacted with Accent ($F[1,9]=18.6$, $p<.005$) such that the boundary-induced lengthening was limited to the unaccented condition ($t(9)=5.2$, $p<.005$) (Fig. 2d)—*i.e.*, as was predicted, the stop (being not initial and next to the source of prominence in **#sC̲V**) underwent a boundary-induced lengthening only when it was free of the influence of prominence (a *prominence-sensitive* boundary effect).

3.1.3. VOT

All three factors influenced VOT. As shown in Fig. 3a, VOT was longer IP-initially than Wd-initially ($F[1,9]=5.1$, $p=.05$), longer when accented *vs.* unaccented ($F[1,9]=26.4$, $p<.001$); and shorter in **#sCV** than in **s#CV** ($F[1,9]=99.8$, $p<.001$). Crucially, however, Boundary and Accent each interacted with Syllable, showing shortening *vs.* lengthening of VOT as a function of the syllable structure (**s#C** *vs.* **#sC**).

As can be seen in Fig. 3b, the Accent by Syllable interaction ($F[1,9]=48.6$, $p<.001$) was due to the fact that VOT was significantly *longer* under accent in the simplex onset (**s#C̲V**) condition ($t(9)=6.1$, $p<.001$), but significantly *shorter* in the complex onset (**#sC̲V**) condition ($t(9)=-5.58$, $p<.001$). Although the shortening was made by a small amount (mean diff. 3 ms), it was significant in line with the prediction made by *the phonetic feature enhancement hypothesis—i.e.*, VOT is modulated to be shortened to enhance the phonetic feature {vl. unasp.}. It should be worth noting that although the mean difference was very small, the accent-induced shortening effect was consistent across speakers. Examination of individual speakers' data revealed that all ten speakers showed a numerically shorter VOT: The mean difference was less than 3 ms for five speakers (ranging from 1 ms to 3 ms), and greater than 3 ms for five speakers (ranging from 3 ms to 8 ms). The Boundary factor also revealed a similar asymmetric effect on VOT as reflected in a three-way interaction ($F[1,9]=8.9$, $p<.05$). As shown in the left of Fig. 3c, there was no Boundary effect on VOT when the syllable was accented. But when it was *unaccented*, as shown in the right of Fig. 3c, Boundary *did* influence VOT, showing a *prominence-sensitive* boundary effect. Crucially, the effect was asymmetric as a function of the syllable structure: VOT was *longer* for IP *vs.* Wd (mean diff. 13 ms, $t(9)=5.18$, $p<.01$) in the simplex onset (**s#CV**), but it was *shorter* in the complex onset (**#sCV**) condition (mean diff. 2 ms, $t(9)=-2.87$, $p<.05$), which again can be interpreted as being consistent with *the phonetic feature enhancement hypothesis*. (Note, however, that a gestural account alternative to the phonetic feature enhancement hypothesis will be discussed in the discussion section.) Again, the boundary-induced shortening effect in the complex onset condition was consistent across speakers: nine out of ten speakers showed a numerically shorter VOT: The mean difference was less than 3 ms for seven speakers (ranging from .5 ms to 3 ms) and greater than 3 ms for two speakers (4 ms and 7 ms).

Furthermore, as was the case with /s/-duration, the potentially ambiguous pair of **s#CV** *vs.* **#sCV** in the Wd boundary condition was significantly differentiated by VOT (*i.e.*, longer in **s#CV** *vs.* **#sCV**) in both accented and unaccented conditions ($t(9)=15.3$; $t(9)=6.3$, respectively, both at $p<.01$; compare gray bars in Fig. 3c between **s#CV** and **#sCV**). Crucially, VOT for the stop in the heterolexical **s#C** cluster was long enough (61.4 ms) to be considered categorically aspirated (even in the Wd condition in which /s/ in **s#C** may be resyllabified with the following stop). The longer *vs.* shorter VOT patterns in **s#C** *vs.* **#sC** were consistent with *the lexically specified laryngeal gesture hypothesis—i.e.*, dual laryngeal gestures *for **s#C*** (rendering the stop aspirated) *vs.* a single gesture for **#sC** (rendering the stop unaspirated).

3.1.4. Vowel duration (in #(s)CV̲)

There were significant main effects of Accent and Syllable on V-duration. As can be seen in Fig. 3d, it was longer when accented than unaccented ($F[1,9]=73.8$, $p<.001$), and it was longer after a complex onset in **#sCV̲** than after a simplex onset in **s#CV̲** ($F[1,9]=54.2$, $p<.001$). As shown in Fig. 3e, Accent interacted with Syllable ($F[1,9]=17.8$, $p<.01$) due to the fact that, while the accentual lengthening effect on the vowel was observed in both syllable structure conditions, its effect size was larger in **#sCV** (mean diff. 44.5 ms, $t(9)=10.2$, $p<.001$, $\eta^2=.92$) than in **s#CV** (mean 35.6 diff. ms, $t(9)=6.98$, $p<.001$, $\eta^2=.84$).

The Boundary factor, on the other hand, did not yield a main effect ($F[1,9]<1$), but there was some trend towards its interaction with Syllable ($F[1,9]=3.72$, $p<.09$). Further posthoc analyses, as shown in Fig. 3f, revealed that the vowel *did* undergo a significant boundary-induced (IP-initial) lengthening only when the vowel was *unaccented* in **#sCV̲** ($t(9)=2.5$, $p<.05$) (again showing a *prominence-sensitive* boundary effect), but not in **s#CV̲** even though V in the latter condition (**s#CV̲**) was closer to the boundary than in **#sCV̲**. (This ran counter to the prediction that the closer the segment is to the boundary, the larger the boundary effect.)

3.1.5. Total voiceless interval (TVI)

3.1.5.1. TVI in the Wd condition. Results of a two-way ANOVA (run only in the Wd condition due to inseparability between a possible pause and the stop closure in the IP condition) showed a significant main effect of both Accent and Syllable. As can be seen from Fig. 4a, TVI was significantly longer in the accented than in the unaccented condition ($F[1,9]=113.35$, $p<.001$), in line with the

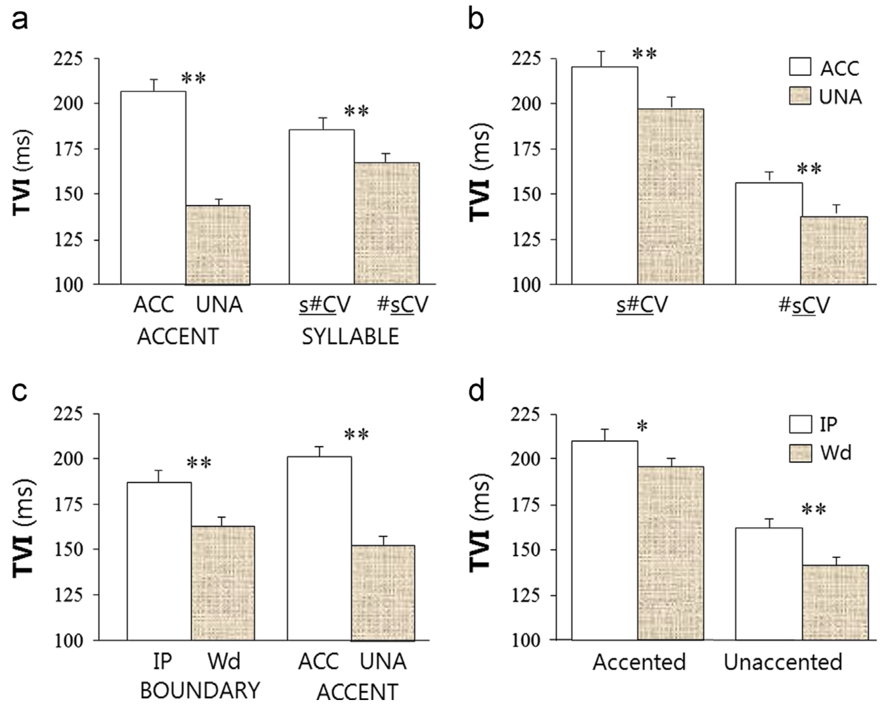

Fig. 4. Total Voiceless Interval (TVI: /s/+closure duration+VOT): Main effects only with Wd (a), Accent × Syllable interaction only with Wd (b), main effects for #sCV (c), and Boundary × Accent interaction for #sCV (d). ('*' refers to $p<.05$ and '**', $p<.01$.)

prediction that the glottal abduction gesture is globally strengthened in prosodic strengthening environments, resulting in longer TVI. And TVI was significantly longer in **s#C** than in **#sC** (F[1,9]=22.86, $p<.001$, $\eta^2=.72$). As shown in Fig. 4b, the substantially longer TVI in **s#C** *vs.* **#sC** (even when the unaccented **s#C** was compared to the accented **#sC**) is consistent with *the lexically specified laryngeal gesture hypothesis*—*i.e.*, dual *vs.* single laryngeal gesture for **s#C** *vs.* **#sC**.

3.1.5.2. TVI in ***#sCV****.* Results of a two-way ANOVA (run only for **#sCV** again due to the possibility that **s#CV** may include a pause) showed a significant main effect of Boundary and Accent. As can be seen from Fig. 4c, TVI was longer for IP than for Wd (F[1,9]= 19.72, $p<.005$), and it was longer when accented *vs.* unaccented (F[1,9]=143.71 $p<.001$), which was again consistent with the assumption that the glottal abduction gesture is strengthened by both boundary and prominence. There was a significant Boundary × Accent interaction (F[1,9]=26.44, $p<.001$). As can be seen in Fig. 4d, it was due to the fact that the effect size of the boundary on TVI was substantially larger in the unaccented condition (mean diff. 26.1 ms, $t(9)=30.55$, $p<.001$, $\eta^2=.77$) than in the accented condition (mean diff. 12.3 ms, $t(9)=7.93$, $p<.05$, $\eta^2=.47$), again showing a *prominence-sensitive* boundary effect.

3.2. Absolute vs. relative durational effects on the #sCV sequence

Thus far we have reported results with respect to how three factors (*i.e.*, Boundary, Accent and Syllable) interactively influence temporal variables in absolute terms. In this section, we continue to explore effects of Boundary and Accent, but this time just on the **#sCV** sequence in order to test *the relational invariance hypothesis*—*i.e.*, speakers preserve the segments' pronunciation 'norm' across conditions that affect the temporal realization of the segment, so that the durational components that may show temporal modification in absolute measures will remain unchanged when they are considered in relational measures. Any effect to be found in relational measures will indicate that speakers make extra efforts to expand the temporal component exceeding the level of maintaining the relational invariance. The results in absolute measure (reported in the preceding subsection) will be compared with those in the relational measure, expressed as percent proportion of each absolute measure relative to the entire **#sCV** sequence duration. Fig. 5 provides a summary of means across conditions in both absolute and relative measures for the comparison purpose.

3.2.1. %-/s/-duration in #sCV

As can be seen in Fig. 5A(a), a two-way ANOVA with %-/s/-duration in relative measure yielded a significant Boundary effect (F[1,9]=21.9, $p<.001$): /s/ was proportionally larger IP-initially than Wd-initially regardless of accent conditions. This suggests that the boundary-induced temporal expansion of /s/ being initial in **#sC** is more than just maintaining relational invariance, showing *relational expansion*. Unlike Boundary, however, Accent did not yield a main effect on %-/s/-duration (F[1,9]<1) (compare ACC and UNA in Fig. 5A(a)), although absolute /s/-duration increased significantly under accent. In other words, the proportion of /s/-duration relative to the entire **#sCV** sequence remained unchanged between the accented and the unaccented conditions (35.4% *vs.* 35.2%), showing *relational invariance.*

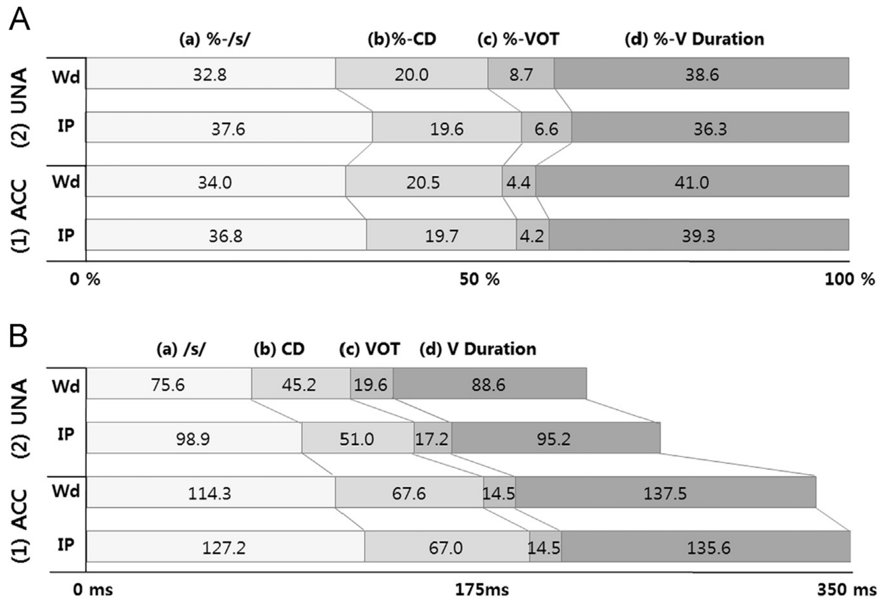

Fig. 5. Durational distributions in #sCV in relational measure (A) *versus* absolute measure (B). Relational measures are expressed as percent proportion of each absolute measure relative to the entire #sCV sequence.

3.2.2. %-CD (stop closure duration) in #sCV

%-CD was not sensitive to either Boundary or Accent conditions (no effects of Boundary and Accent) as can be seen in Fig. 5A(b). That is, although the absolute CD increased significantly under accent and IP-initially (Fig. 5B(b)), it remained unchanged in proportion to the entire **#sCV** sequence duration (all pairwise comparisons: $t(9)<1$, $p>.1$), indicating that the prosodic modification of 'non-initial' stop closure duration in **#sCV** is *relationally invariant.*

3.2.3. %-VOT in #sCV

There was a main effect of Boundary on %-VOT: It was *smaller* for IP than for Wd ($F[1,9]=8.2$, $p<.05$). But as was the case with absolute VOT measure, Boundary interacted with Accent ($F[1,9]=38.6$, $p<.001$), such that %-VOT was *smaller* for IP (*vs.* Wd) only in the *unaccented* condition (mean diff. 2.1%, $t(9)=-4.1$, $p<.01$). This means that shortening of VOT due to Boundary exceeded the level of relational invariance at least in the unaccented syllable, showing *relational reduction.* Accent also yielded a main effect ($F[1,9]=91$, $p<.001$), showing an accent-induced *relational reduction* of %-VOT in both IP and Wd conditions (IP, mean diff. 2.3%, $t(9)=-10.72$, $p<.001$, $\eta^2=.93$; Wd, mean diff. 4.2%, $t(9)=-8.7$, $p<.005$, $\eta^2=.89$). The VOT shortening effects that exceeded the maintenance of relational invariance under prosodic strengthening were in line with *the phonetic feature enhancement hypothesis* which postulates that VOT is directly modulated by boundary and prominence.

3.2.4. %-V-duration

%-V-duration showed significant main effects of Boundary and Accent ($F[1,9]=11.3$ and $F[1.9]=13.1$, respectively, both at $p<.01$), but unlike the results with the absolute V-duration measure, the two factors produced opposite patterns: %-V-duration was significantly *smaller* for IP *vs.* Wd, but it was significantly *larger* when accented than when unaccented. In other words, while Accent gave rise to lengthening of the vowel in both absolute and relative measures, showing more than just maintaining the relational invariance, Boundary showed a different pattern: the vowel (in **#sCV**) was proportionally *shortened* IP-initially, even in the unaccented condition in which the absolute vowel duration was significantly longer IP-initially (compare Fig. 5A(d) and Fig. 5B(d)). (This will be discussed in connection with the relation between VOT and the following vowel.)

3.2.5. Summary of relative durational effects in #sCV

For the boundary effect, %-/s/-duration (being initial immediately after the boundary) showed a lengthening effect that exceeded the level of maintaining relational invariance while the boundary effect on CD for the non-initial stop turned out to be relationally invariant. For the accent effect, %-V-duration (being at the source of prominence) showed an accentual lengthening effect that went beyond the maintenance of relational invariance, while accentual effect on /s/ and CD that were not immediately at the source of prominence turned out to be relationally invariant. In short, the temporal expansion at the source of prosodic strengthening was more than just for maintaining relational invariance, whereas the relational invariance was observed for the durational components that were not strictly at the source of prosodic strengthening. Furthermore, prosodic strengthening effects on VOT (by boundary and prominence) also showed a pattern of *relational reduction* (again beyond the maintenance of relational invariance), reinforcing the prediction made by *the phonetic feature enhancement hypothesis.*

4. Discussion

4.1. Prominence-independent and prominence-sensitive boundary effects on consonants

One of the fundamental research questions of the present study was concerned with how the boundary effect would interact with the prominence effect on the **#sC** sequence. As discussed in Section 1, results of the present study suggest that boundary effects on the domain-initial segments, also known as domain-initial strengthening (DIS) effect (*e.g.*, Cho & Keating, 2009; Fougeron & Keating, 1997), could be divided into two cases, depending on the segment's proximity to the boundary (*e.g.*, Byrd & Saltzman, 2003; Byrd et al., 2006) and to the locus of prominence (the nucleus vowel). The *prominence-independent* DIS effect was observed most clearly when the segment was strictly domain-initial. In **#sCV**, the initial segment /s/ was found to be longer IP-initially across accent conditions. The *prominence-sensitive* DIS effect, on the other hand, was observed when a durational component was not strictly initial so that it became less influenced by the boundary but more vulnerable to the prominence effect. In such a case, the effect was generally limited to when the target-bearing syllable was free of the influence of prominence—*i.e.*, in an unaccented condition. Stop closure duration in **#sCV** and VOT[4] in **#sCV** (and in **s#CV**) showed such a pattern. Note that a similar, robust (prominence-independent) boundary effect on the initial consonant of the German consonant cluster was reported in Bombien, Mooshammer, Hoole, and Kühnert (2010) and Bombien, Mooshammer, and Hoole (2013). They also showed that the boundary effect on the second consonant in German was very small (limited to a certain cluster type) or nonobservable, being more subject to the lengthening effect of lexical stress, though the accent factor was not considered in their studies.

The prominence-sensitive boundary effect, however, may not appear to be entirely compatible with findings of some previous studies on English (Choi, 2003; Cole, Kim, Choi, & Hasegawa-Johnson, 2007) which often failed to show a reliable boundary effect on VOT for English stops even in an unaccented condition. As noted by Cho and Keating (2009), however, such inconsistent findings may be due to the fact that the presence or absence of accentuation on the test word was not fully factored in. By investigating the data with controlled prominence factors, Cho and Keating (2009) demonstrated exactly the same prominence-sensitive boundary effect on VOT as the present study showed with the **s#CV** condition (see also Cho, 2006, 2008; Kim & Cho, 2011, for similar prominence-sensitive boundary effects on articulatory measures). The interaction of DIS effects with prominence leads to a question regarding its generalizability to other languages. A recent study by Cho et al. (2011) on Korean (which does not employ lexical stress in its prominence system) indeed showed that realization of VOTs for initial aspirated stops is also modulated by phrasal accent in much the same way as in English, showing a prominence-sensitive boundary-induced lengthening of VOT (*i.e.*, only when unaccented). Thus, the results of both present and previous studies taken together imply that the boundary effect is modulated by the prominence system of the language in a cross-linguistically similar way, regardless of whether the prominence system is complicated by lexical stress or not. More studies are certainly called for in order to understand at what level of detail the relationship between boundary and prominence can be taken to be cross-linguistically applicable and in what aspects it is further constrained by a language-specific prosodic system.

4.2. Boundary effect on the vowel

In addition to the prominence-sensitive DIS effect on stop closure duration and VOT, a potentially prominence-sensitive boundary effect was also observed on V-duration which showed a noteworthy difference due to syllable structure. In **s#CV**, V-duration showed no boundary effect, which was largely consistent with the previous observation that the boundary strength has little influence on the temporal realization of the following vowel in English. It might be presumably because, as Barnes (2002) suggested, the vowel in English is reserved for prominence (or stress) marking along the temporal dimension (but see Cho & Keating, 2009, and Cho et al., 2011 for related discussion from different perspectives). Interestingly, however, a boundary effect on the vowel arose in **#sCV** (when it was unaccented), demonstrating a case in which the boundary-induced lengthening effect *does* spread into the following vowel.

A question then arises as to whether the temporal structure of the domain-initial vowel in **#sCV** can be seen as being directly modulated by boundary strength. One might posit that, although the vowel may be restricted for prominence marking in English, the restriction may become loose when the vowel is unaccented, so that the boundary effect may permeate into the following vowel in the complex onset condition (**#sCV**). This possibility, however, becomes less plausible, when we consider an asymmetrical fact that the boundary effect was not observed in the simplex onset condition (**s#CV**). Given that V in **#sCV** is more distant from the boundary, being separated by two segments than V in **s#CV** in which V is separated from the boundary by one segment, the boundary effect on the vowel, if it existed, would have been more readily found with **s#CV** rather than with **#sCV**. But the opposite was what we found. At first glance, this unexpected pattern appears to take support away from the general assumption that the boundary effect becomes attenuated as the segment gets farther away from the boundary (*e.g.*, Byrd & Saltzman, 2003; Byrd et al., 2006; Krivokapić & Byrd, 2012; *cf.* Cho & Keating, 2009). However, the lengthening of the vowel in **#sCV** may be interpreted as being due to a trade-off (inverse) relationship between VOT and the following vowel: the shorter the VOT, the longer the following vowel. (An additional analysis of regression lent support to this possibility (with VOT as the predictor, and the vowel duration as the dependent variable;

[4] VOT may be viewed as being initial from the articulatory point of view as it can be considered to be part of the laryngeal gesture that is always initial regardless of the segmental makeup. However, VOT is a non-initial durational component when considered in the liner acoustic dimension, and given the possibility that VOT may be seen as an independent feature as will be discussed below.

$R^2=0.33$, $p<.0001$)).[5] From the kinematic point of view, the vocalic gesture in (C)CV articulation starts well before the consonant release (*e.g.*, Browman & Goldstein, 1990, 1992), so that the period of VOT overlaps with the vocalic movement duration for the vowel. Much of the acoustic temporal effect on the vowel is then likely to be saturated with, or masked by, the period of VOT, in such a way that the longer the VOT, the shorter the acoustic vowel duration and *vice versa* (see Cho et al., 2011 for related discussion). Thus, the boundary effect becomes less evident in the following acoustic vowel duration when it has an effect of lengthening VOT as in #CV, whereas the boundary effect appears to be reflected in the vowel duration when VOT is shortened in **#sCV**. It therefore appears that the boundary-related lengthening of the vowel observed with **#sCV** is a result of the allophonically-driven shortening of VOT after /s/ rather than a direct consequence of the boundary effect on the vowel. (See Section 4.6 for related discussion on the prosodic-gesture model.)

4.3. Prominence marking in s#CV vs. #sCV

One of the most clear effects of accent (*i.e.*, prominence marking) associated with both **s#CV** and **#sCV** was accentual lengthening of the nucleus vowel across boundary conditions.[6] This was consistent with previous findings in English (*e.g.* Cambier-Langeveld & Turk, 1999; Cho & Keating, 2009; de Jong, 2004; Turk & White, 1999; White & Turk, 2010). VOT, on the other hand, showed a sharp contrast in accent effects as a function of syllable structure: accent caused lengthening of VOT in **s#CV**, but shortening of VOT in **#sCV**. As was the case with the boundary effect, the voiceless aspirated stop in **s#CV** became more aspirated under accent, while the allophonically unaspirated stop in **#sCV** became more unaspirated. (See Sections 4.5 and 4.6 for related discussion on this point.) Another noteworthy finding was that a leftward accentual lengthening was observed on /s/ even across a lexical word boundary (**s#CV**) in the Wd boundary context. In line with a possibility discussed in Turk and White (1999), the leftward spreading was more attenuated across a lexical boundary (**s#CV**) than within a word (**#sCV**). Moreover, the fact that the leftward spreading was completely blocked across an IP boundary demonstrated that the leftward spreading of accentual lengthening is further modulated by the boundary strength: it permeated through a Wd boundary but not thorough an IP boundary.

4.4. Fine phonetic differences between s#CV and #sCV in the word boundary condition

Another important question of the present study was how the temporal organization of **#sCV** would differ from that of **s#CV** in the potentially neutralizing phrase-internal (Wd boundary) context in which /s/ in **s#CV** may prefer to be resyllabified according to the maximal onset principle (*e.g.*, Clements & Keyser, 1983). Not a single temporal measure tested in the present study showed any evidence for **s#CV** to be neutralized with **#sCV** along the temporal dimension even in the Wd boundary condition. First, the vowel was longer in **#sCV** than in **s#CV** in line with a general observation that a vowel tends to be longer after a stop with a short lag VOT than with a long lag VOT (*e.g.*, Lisker & Abramson, 1967). Second, /s/ was *longer* word-initially (in **#sCV**) than word-finally (in **s#CV**) which was consistent with a word-level asymmetry observed in the literature (*e.g.*, Oller, 1973; Keating et al., 1999; see also Byrd, 1996 for articulatory evidence).[7] Finally, VOT in **s#CV** was still too long (61.1 ms) to be categorized as unaspirated, indicating no application of the phonological rule, and hence no restructuring of syllable structure. The results therefore indicate that underlying syllable structures are differentiated in fine phonetic detail even in the potentially ambiguous phrase-internal condition, which is likely to help listeners with lexical segmentation of otherwise ambiguous sequences (*e.g.*, 'ice#can' *vs.* 'eye#scan') (see Shatzman & McQueen, 2006 for related discussion on Dutch data; Cutler, 2012, chap. 7 for a review).

Another noteworthy point is that the substantially longer VOT in **s#C** (enough to be categorized as aspirated) was observed in a context that might cause shortening of VOT after /s/. This was consistent with the prediction made by *the lexically specified laryngeal gesture hypothesis* (*i.e.*, single *vs.* dual laryngeal gestures for **#sC** *vs.* **s#C** (Browman & Goldstein, 1986, Goldstein, 1992). For the longer VOT in **s#C**, the underlyingly specified dual laryngeal abduction gestures (one for the coda /s/ and one for the onset 'C' in **s#C**) are preserved on the surface even in the phrase-internal context in which the cohesiveness of the two gestures (or the gestural bonding strength) may increase (Browman & Goldstein, 2000). The second laryngeal abduction gesture that is assumed to be aligned with the release of the stop is responsible for the longer VOT, rendering the stop aspirated. On the other hand, for the shortening of VOT in **#sC**, only one gesture is lexically specified, being aligned with /s/, such that when the stop is released, the glottis is roughly at the final phase of the abduction–adduction cycle, accounting for the unaspiratedness of the stop. The results obtained with Total Voiceless Interval provided further support. It was predicted to be longer in **s#C** with underlying two laryngeal abduction gestures than in **#sC** with a single gesture. This was exactly what we found with the TVI measure taken in the Wd context. Under this account, the allophonic rule that makes a stop unaspirated after /s/ in the tautosyllabic condition can be understood not as

[5] As suggested by a reviewer, further analyses of the data in the unaccented condition were conducted on Total Vowel Duration (VOT plus V-duration). Results of a two-way RM ANOVA showed a significant interaction between Boundary and Syllable ($F[1,9]=17.89$, $p<.005$). In **#sCV**, there was no boundary effect ($t(9)=2.54$, $p>.1$) in line with the discussed reverse relationship between VOT and V-duration, whereas a significant boundary effect was found in **#CV** (mean diff. 14 ms, $t(9)=11.96$, $p<.01$). The discrepancy between **#sCV** and **#CV** again appears to stem at least in part from differential laryngeal settings, such that the laryngeal abduction gesture aligned with the stop release in **#CV** induces a further elongation of VOT in IP-initial position, contributing to the longer Total Vowel Duration. In contrast, the shortening of VOT (at least partly associated with the fact that the laryngeal abduction gesture is aligned with /s/) appears to have offset the lengthening effect on V-duration, yielding the null effect when VOT and V Duration were combined. This is also consistent with the boundary proximity effect: the longer Total Vowel Duration in **#CV** may come from the fact that V is more proximal to the boundary in **#CV** than in **#sCV**.

[6] Results of ANOVAs run on Total Vowel Duration (VOT+V-duration) also showed a main effect of Accent (longer when accented) in both **#sCV** and **s#CV** conditions ($F[1,9]=102.22$, $p<.001$).

[7] Note that this word-level asymmetric pattern was reversed at a phrase-level (with an IP boundary), such that final /s/ tended to be longer than initial /s/, being consistent with the previous observation that /s/ is longer in phrase-final position than in any other position in the utterance (*e.g.*, Klatt, 1976).

a result of an idiosyncratic language-specific allophonic rule, but as a necessary consequence of the dynamic relationship between the laryngeal abduction gesture that is aligned (roughly) with the midpoint of /s/ and the release gesture of the stop.

4.5. Relational invariance vs. relational expansion of prosodic strengthening in #sCV

The present study also explored how prosodic strengthening would be realized in **#sCV** (with the complex onset) in both absolute and relational measures—*i.e.*, when the magnitude of the duration change was compared to the duration change of the entire **#sCV** sequence. The combined examination of absolute and relational measures showed that some durational measures which had shown significant prosodic strengthening effects in the absolute measure turned out to remain unchanged in the relational measure. For example, /s/-duration in **#sCV** which had shown a significant accentual lengthening effect in the absolute measure was found to remain unchanged with respect to accent in the relational measure. Stop closure duration in the absolute measure also showed some significant lengthening effects of both Accent and Boundary, but not in the relational measure. As discussed at the outset of the paper, such a maintenance of relational invariance across prosodic conditions may be interpreted as coming from the speaker's effort to preserve the segment's pronunciation 'norm' across conditions that affect the temporal realization of the segment (*e.g.*, Smiljanić & Bradlow, 2008) or to maintain a constant perceptual distance between sounds (*e.g.*, Solé, 1997). (See also Port & Dalby, 1982; Miller, Green, & Reeves, 1986; Kessinger & Blumstein, 1997; Boucher, 2002 for related findings and discussion.)

Our results, however, showed at least two cases of more than just maintaining the relational invariance—*i.e.*, the boundary effect on /s/-duration and the accent effect on V-duration. Recall that the boundary-induced lengthening of initial /s/ and the accent-induced lengthening of the vowel in the absolute measure were found to be proportionally larger as well (relative to the lengthening of the **#sCV** sequence).[8] These two lengthening effects therefore present cases with more than just a preservation of the pronunciation 'norm' across different conditions. They in fact showed *relational expansion* along the temporal dimension. (Note that VOT in **#sCV** showed shortening under prosodic strengthening in both absolute and relative terms, which will be discussed in the next section.) In her discussion on the relational invariance, Solé (1997) proposed that the speaker controls speech timing and maintains the same proportion across durational differences in rate, stress, or syllable types in order to convey linguistic information to the listener. Departing from this assumption, the observed patterns of relational expansion suggest that the speaker *does* make an extra effort to expand the very first segment and the nucleus vowel more than the rest of the sequence presumably to signal the pivotal loci of the boundary *vs.* the prominence information in the continuously changing stream of speech. This, of course, does not mean that other durational components are not subject to temporal expansion due to boundary or accent, but their temporal expansion tends to be proportional to the sequential expansion (*e.g.*, /s/-duration and stop closure duration under accent), maintaining relational invariance as proposed by Solé (1997) and Smiljanić and Bradlow (2008). However, the two distinct localized effects (*i.e.*, the boundary effect on /s/ and the accent effect on the nucleus vowel) appear to illuminate the underpinnings of how the two important functions of prosodic structuring, boundary-marking and prominence-marking, are encoded separately in speech production process (see Keating & Shattuck-Hufnagel, 2002; Cho & Keating, 2009, for related discussion).

4.6. Prosodically-conditioned phonetic implementation of an allophonic rule: a featural account vs. a gestural account

Another important finding regarding the temporal structure of **#sCV** was variation of VOT which reflects how prosodic strengthening exercises its leverage on the phonetic implementation of the allophonic rule whereby a voiceless stop becomes unaspirated after /s/ in English. As discussed in Sections 1 and 4.4, the allophonic rule can be understood as a consequence of a particular laryngeal setting in which the laryngeal abduction peak is tightly aligned with the frication of /s/. Based on this assumption, two possibilities were initially thought of as competing hypotheses. The first hypothesis was *the passive VOT lengthening hypothesis* which predicted lengthening of VOT under prosodic strengthening as the laryngeal abduction gesture was expected to be strengthened, which, all else being equal, would cause an elongation of VOT. The second hypothesis was *the phonetic feature enhancement hypothesis* under which a short lag VOT in **#sCV** was expected to be even more shortened if VOT is directly modulated by prosodic strengthening to enhance the phonetic feature {vl. unasp.}.

Results of the present study indeed showed shortening of VOT under prosodic strengthening in line with the prediction made by *the phonetic feature enhancement hypothesis.* Under accent, not only did the short lag VOT for the stop in **#sC** become even shorter in the absolute measure, but also it was proportionally further reduced in the relational measure regardless of the boundary size. Likewise, boundary strength showed a similar VOT shortening effect at least in the unaccented condition. The VOT shortening, however, is not attributable to reduction of the laryngeal gesture as reflected in Total Voiceless Interval (TVI): TVI was found to be consistently longer in prosodic strengthening environments in **#sC** (and in **s#C**) even when VOT was shortened, indicating that the laryngeal abduction gesture was indeed strengthened under prosodic strengthening as previously assumed (Cooper, 1991; Pierrehumbert & Talkin, 1992; Goldstein, 1992; see Fougeron, 1999, for a review). Most crucially, the lengthening of TVI, but the shortening of VOT that exceeded the maintenance of relational invariance can be interpreted as supporting the view that VOT is directly modulated by prosodic strengthening. As was discussed in Section 1, the direct modulation of VOT is in principle possible if

[8] The relational expansion of the vowel duration under accent disappears when Total Vowel Duration (including VOT) is taken into account. This would be primarily because of the counter effect of VOT shortening (relational reduction) due to the application of the allophonic rule. Under an assumption that the primary culminative function (marking prominence) of prosodic strengthening is to make the acoustically most sonorous part of the syllable (the voiced part of the nucleus vowel) perceptually more salient than the other segments, the relational expansion argument may still hold for the 'voiced' vowel duration.

one adopts the notion of ‘Articulatory VOT’—that is, the timing between the gesture for the stop release and the gesture responsible for the initiation of voicing may be directly modulated as proposed by Cho and Ladefoged (1999) and Ladefoged and Cho (2001). What would then cause shortening of VOT under prosodic strengthening?

Following Cho and McQueen (2005), we propose that VOT shortening for a voiceless stop can be explained in terms of the enhancement of language-specific phonetic feature. Cho and McQueen showed that VOT for an unaspirated /t/ in Dutch was not lengthened, counter to the prediction by the polarization principle (*i.e.*, longer VOT for a voiceless stop would maximize its contrast with its voiced counterpart). It was instead found to be shortened under prosodic strengthening, which was interpreted as a result of enhancement of a language-specific phonetic feature {-spread glottis} (*e.g.*, Keating, 1990), which is equivalent to {voiceless unaspirated} used in Keating (1984) and Cho and Ladefoged (1999). Cho and McQueen also suggested that a voiceless stop in #CV in English is specified with {+spread glottis} (again equivalent to {voiceless aspirated}) whose enhancement can account for lengthening of VOT under prosodic strengthening. This was exactly what we found with **s#CV** in the present study. In the case of the **#sCV** context examined in the present study, however, the voiceless stop undergoes an allophonic rule which has an effect of assigning the stop with the phonetic feature {voiceless unaspirated}.[9] We suggest that, as was the case in Dutch, it is this phonetic feature that undergoes featural enhancement under prosodic strengthening, resulting in the shortening of VOT. The shortened VOT in both Dutch and the case of **#sCV** in English is made possible under prosodic strengthening by modulating ‘Articulatory VOT’ even when the laryngeal gesture is expected to be strengthened. Just as VOT for the same voiceless stop category is fine-tuned in a language-specific way, giving rise to seemingly arbitrary but linguistically relevant distributions of VOT across languages (as proposed by Cho & Ladefoged, 1999), so can it be modulated directly by the phonetic grammar of the language in such a way to implement prosodic strengthening by making reference to the phonetic feature system of the language. We therefore propose that prosodic strengthening operates at the phonetic level on a language-specific phonetic feature with phonetic content, regardless of whether the feature is underlyingly associated with a segment (in the case of Dutch) or derived by an allophonic rule (in the case of English for the **#sCV** sequence).

One remaining issue to be discussed is in what aspect the observed VOT shortening can be considered to be a case of phonetic feature enhancement. The term ‘enhancement’ has been used in different contexts. It may refer to a case in which redundant features serve to enhance the auditory salience of distinctive features (*e.g.*, the protrusion of the lips for [round] may serve to enhance the backness of the tongue for [back]) (*e.g.*, Stevens, Keyser, & Kawasaki, 1986; Stevens & Keyser, 2010). It is also used when distinctive features are phonetically implemented in a *hyperarticulated* way to maximize phonological contrast (and therefore lexical contrast) (*e.g.*, de Jong, 1995), usually showing a polarization of phonetic values along the phonetic dimension in which phonological distinctions are made (*e.g.*, the English vowel /ʊ/ is more posterior (de Jong, 1995), as opposed to /i/ which is more anterior (Cho, 2005) in accented than in unaccented syllables). The term ‘phonetic’ feature enhancement was used by Cho and McQueen (2005), who proposed that the enhancement of the phonological distinction under prosodic strengthening may not simply be obtained by the principle of contrast maximization along the phonetic dimension (*i.e.*, the polarization hypothesis), but it may be determined by how a language-specific ‘phonetic’ feature is phonetically implemented. If prosodic strengthening makes direct reference to the phonetic feature, the enhancement of the phonetic feature may occur even when the phonetic feature is allophonically determined. Of course, while the enhancement of an allophonic feature would not necessarily maximize phonological contrast between phonemes, the allophonic information may be jointly specified with either some other segment (*e.g.*, /s/ in the /sC/ sequence in this case) or with the prosodic environment, so that the enhanced allophonic information may still reinforce lexical contrast in a particular segmental and prosodic context. Furthermore, as briefly discussed in Section 4.4, it may still serve to enhance an allophonic cue that may help listeners with lexical segmentation of otherwise ambiguous sequences (*e.g.*, ‘ice#can’ *vs.* ‘eye#scan’) (see Cutler, 2012, chap. 7 for a review). Recall, however, that although the shortening of VOT under prosodic strengthening was significant and consistent across speakers, its magnitude was rather small (3 ms on the average). While there is considerable evidence in the literature that prosodically-induced fine-grained phonetic detail is generally available to the listener in speech comprehension (*e.g.*, Salverda, Dahan, & McQueen, 2003; Cho, McQueen, & Cox, 2007; Kim & Cho, 2013), it remains to be seen to what extent such a subtle VOT shortening effect in **#sCV** in English is indeed exploited by listeners in speech comprehension.

Our claim that prosodic strengthening makes crucial reference to the phonetic feature system, however, is not the only way of interpreting the shortening of VOT. An alternative (gestural) account can be thought of in the framework of Articulatory Phonology (AP). In the current model of AP, voicing is not specified at the gestural level, but rather it occurs as a result of default laryngeal settings that assume the glottal adduction (Browman & Goldstein, 1986, 1992, see Pouplier, 2011 for a review). In such a system, two possibilities can be thought of that might account for VOT shortening under prosodic strengthening. The first possibility is to view variation of VOT as determined by two factors: the magnitude of the laryngeal abduction and its timing with the oral gesture (Browman & Goldstein, 1986; Goldstein, 1992). For the /s/-stop cluster, however, the magnitude of the laryngeal gesture becomes a more important determinant for VOT as the timing of the laryngeal abduction gesture is tightly aligned with /s/ (for the necessity to initiate and maintain turbulent noise for the fricative). In this case, all else being equal, shortening of VOT is expected to occur most likely when the magnitude of the laryngeal gesture is reduced. As was discussed above, however, the reduction of the laryngeal duration is an unlikely scenario under prosodic strengthening.

[9] Although the allophonic rule may be phonetically grounded (due to a particular laryngeal setting for the /s/-stop sequence), it is still feasible that the recurring pattern of a shortened VOT can be phonologized in the grammar of the language (*e.g.*, Keating, 1990; Kingston & Diehl, 1994), so that the feature system assigns {-spread glottis} or {vl. unasp.} for the stop in the /s/-stop context.

A more plausible possibility is then to view shortening of VOT as being attributable to a delayed timing of the stop release gesture under prosodic strengthening. It can be either passively achieved when the stop closure duration is extensively lengthened as a result of prosodic strengthening on the stop or actively achieved if the stop release is deliberately delayed under prosodic strengthening. In either way, VOT shortening can be seen as a consequence of an expansion of closure duration due to prosodic strengthening, rather than as something that can be directly controlled. This gestural account has a further advantage of explaining VOT shortening in connection with the general predictions made by the theory of the π-gesture (Byrd, 2006; Byrd et al., 2000, 2006; Byrd & Saltzman, 2003). Under the influence of the π-gesture, articulatory gestures at the prosodic boundary are expected to be lengthened. VOT lengthening of voiceless aspirated stops that have been observed in domain-initial positions (and in the case of **s#C** in the present study) can be seen as being attributable to the π-gesture's influence that has an effect of lengthening (strengthening) the laryngeal abduction gesture that is aligned with the release of the stop. As for the shortening of VOT, however, the π-gesture's influence may not seem to provide an immediate explanation as it generally predicts lengthening rather than shortening of temporal components. But under the gestural account discussed above, a further shortening of an already shortened VOT for the stop after /s/ can still be attributable to the expansion of the oral stop closure under the π-gesture's influence, which in turn has an effect of delaying the time of the release. In such a case, the laryngeal gesture itself is also expected to be lengthened by the π-gesture, but its effect is likely to be negligible by the time the following consonant is released because the laryngeal gesture is tightly aligned with /s/ in the /s/-stop cluster.

Thus far we have discussed two possible accounts from different theoretical perspectives (in terms of the phonetic feature system *vs.* the gestural system) for shortening of VOT in the /s/-stop cluster under prosodic strengthening. Each account may have its own advantage over the other. For example, under the gestural account, VOT does not need to be controlled by the system as it is determined dynamically by the timing between the laryngeal abduction gesture and the delayed stop release. It also provides a unified account for lengthening *vs.* shortening of VOT under the influence of π-gesture. However, the critical assumption that the delayed stop release (due to prosodic strengthening on the stop) would cause the shortening of VOT is undermined, when we consider the fact that the glottal abduction–adduction cycle for the fricative-stop cluster is often observed to be completed well before the release (Hoole, 1999; Löfqvist, 1995). If this were the case, a delay of the stop release due to consonantal strengthening would not contribute to further shortening of VOT under prosodic strengthening. A challenge for a gestural account as sketched above would therefore be to devise a way to specify a gesture responsible for initiating voicing at the gestural level, such that the intergestural timing between the gesture for the stop release and the voicing gesture (exactly in the way as Articulatory VOT was defined) can be modulated directly by the system. That way, it may account for VOT variation under prosodic strengthening, while at the same time providing a gestural explanation for the language-specifically determined range of VOT that brings about a range of VOT across languages. The proposed featural account, on the other hand, can successfully explain both the prosodically-modulated VOT within a language and cross-linguistic variation of VOT, overcoming the potential weakness that arises with the gestural account—*i.e.*, the phonetic feature is assigned with a modal value of Articulatory VOT that is determined by the phonetic grammar of the language and is further modulated by prosodic strengthening.

5. Conclusion

The results of the present study have added to a growing body of literature on the phonetics-prosody interface, by showing that the temporal structure of the **sCV** sequences is systematically modulated by prosodic strengthening that serves dual functions of marking prosodic structure—*i.e.*, boundary marking (delimitative function) and prominence marking (culminative function). In general, both types of prosodic strengthening gave rise to temporal expansion of durational components, but the effect was largest when the segment was at the source of strengthening: the initial segment for boundary marking and the nucleus vowel for prominence marking. These segments were temporally expanded in both absolute and relational measures, showing 'relational expansion' relative to the entire sequence. In contrast, other durational components that were not at the source showed temporal expansion in the absolute measure, but not in the relational measure, showing relational invariance. The boundary effect on the initial segment was prominence-independent, and the effect on the non-initial segment (*e.g.*, the stop in **#sCV**) was prominence-sensitive—*i.e.*, the effect was found only when the segment was free of the prominence influence (when unaccented), showing an interaction between boundary and prominence markings. It was also found that not a single durational measure showed any evidence that **s#CV** is neutralized with **#sCV** even in the phrase-internal condition, showing that underlying syllable structures are cued in fine phonetic detail.

A particularly novel finding of the present study was the opposite direction of VOT variation for the voiceless stop in **s#CV** *vs.* **#sCV**—*i.e.*, VOT was longer in **s#CV**, but shorter in **#sCV** under prosodic strengthening. The asymmetrical temporal modification was interpreted as showing an enhancement of phonetic features {voiceless aspirated} and {voiceless unaspirated} (or {+spread glottis} and {-spread glottis}), respectively. It was proposed that prosodic strengthening makes crucial reference to the phonetic feature system of the language and operates on a phonetic feature that categorizes the stop as aspirated (*e.g.*, {voiceless aspirated}) and that is derived by an allophonic rule (*e.g.*, {voiceless unaspirated}). Articulatory VOT (Cho & Ladefoged, 1999) was proposed to be directly modulated by prosodic strengthening to enhance the phonetic feature in much the same way as VOT under the same phonological category can be fine-tuned in a language-specific way across languages. Although the asymmetrical pattern can be interpretable dynamically as a consequence of laryngeal strengthening (for VOT lengthening) *vs.* the delayed stop release due to consonantal strengthening (for VOT shortening), it was proposed that the laryngeal gesture responsible for voicing should be specified at the gestural level whose timing with the stop release may be directly modulated to account for observable VOT variation within and across languages.

Acknowledgments

We thank the speakers who participated as subjects in our experiment. We are also grateful to reviewers and the editor, Ken de Jong for constructive comments on an earlier version of this paper. Part of the data reported here was used in the second author's MA thesis at Hanyang University (Seoul, Korea). This work was supported by the research fund of Hanyang University (HY-2013) to the first author (T. Cho).

References

Barnes, J. A. (2002). *Positional neutralization: A phonologization approach to typological patterns* (Unpublished Ph.D. dissertation). Berkeley: University of California.
Beckman, M. E. (1996). The parsing of prosody. *Language and Cognitive Processes*, *11*, 17–67.
Beckman, M. E., Edwards, J., & Fletcher, J. (1992). Prosodic structure and tempo in a sonority model of articulatory dynamics. In: G. Docherty, & B. Ladd (Eds.), *Papers in laboratory phonology II: Gesture, segment, prosody* (pp. 68–86). Cambridge: Cambridge University Press.
Bombien, L., Mooshammer, C., Hoole, P., & Kühnert, B. (2010). Prosodic and segmental effects on EPG contact patterns of word-initial German clusters. *Journal of Phonetics*, *38*, 388–403.
Bombien, L., Mooshammer, C., & Hoole, P. (2013). Articulatory coordination in word-initial clusters of German. *Journal of Phonetics*, *41*, 546–561.
Boucher, V. (2002). Timing relations in speech and the identification of voice-onset times: A stable perceptual boundary for voicing categories across speaking rates. *Perception and Psychophysics*, *6*, 121–130.
Browman, C. P., & Goldstein, L. (1986). Towards an articulatory phonology. *Phonology Yearbook*, *3*, 219–252.
Browman, C. P., & Goldstein, L. (1990). Tiers in articulatory phonology, with some implications for casual speech. In: J. Kingston, & M. E. Beckman (Eds.), *Papers in laboratory phonology I: Between the grammar and the physics of speech* (pp. 341–376). Cambridge: Cambridge University Press.
Browman, C. P., & Goldstein, L. (1992). Articulatory phonology: An overview. *Phonetica*, *49*, 155–180.
Browman, C. P., & Goldstein, L. (2000). Competing constraints on intergestural coordination and selforganization of phonological structures. *Bulletin de la Communication Parlée*, *5*, 25–34.
Byrd, D. (1996). Influences on articulatory timing in consonant sequences. *Journal of Phonetics*, *24*, 209–244.
Byrd, D. (2000). Articulatory vowel lengthening and coordination at phrasal junctures. *Phonetica*, *57*, 3–16.
Byrd, D. (2006). Relating prosody and dynamic events: Commentary on the papers by Cho, Navas, and Smiljanić. In: L. Goldstein, D. H. Whalen, & C. T. Best (Eds.), *Laboratory phonology, Vol. 8: Varieties of phonological competence* (pp. 549–561). Berlin/New York: De Gruyter Mouton.
Byrd, D., & Choi, S. (2010). At the juncture of prosody, phonology, and phonetics—The interaction of phrasal and syllable structure in shaping the timing of consonant gestures. In: C. Fougeron, B. Kühnert, M. D'Imperio, & N. Vallée (Eds.), *Laboratory phonology*, 10 (pp. 31–59). Berlin & New York: Mouton de Gruyter.
Byrd, D., Krivokapić, J., & Lee, S. (2006). How far, how long: On the temporal scope of prosodic boundary effects. *Journal of the Acoustical Society of America*, *120*, 1589–1599.
Byrd, D., Lee, S., & Campos-Astorika, R. (2008). Phrase boundary effects on the temporal kinematics of sequential tongue tip consonants. *Journal of the Acoustical Society of America*, *123*, 4456–4465.
Byrd, D., Lee, S., Riggs, D., & Adams, J. (2005). Interacting effects of syllable and phrase position on consonant articulation. *Journal of the Acoustical Society of America*, *118*, 3860–3873.
Byrd, D., & Saltzman, E. (2003). The elastic phrase: Modeling the dynamics of boundary-adjacent lengthening. *Journal of Phonetics*, *31*, 149–180.
Cambier-Langeveld, T., & Turk, A. (1999). A cross-linguistic study of accentual lengthening: Dutch *vs.* English. *Journal of Phonetics*, *27*, 255–280.
Cho, T. (2004). Prosodically-conditioned strengthening and vowel-to-vowel coarticulation in English. *Journal of Phonetics*, *32*, 141–176.
Cho, T. (2005). Prosodic strengthening and featural enhancement: Evidence from acoustic and articulatory realizations of /≙,i/ in English. *Journal of the Acoustical Society of America*, *33*, 121–157.
Cho, T. (2006). Manifestation of prosodic structure in articulatory variation: Evidence from lip kinematics in English. In: L. Goldstein, D. H. Whalen, & C. T. Best (Eds.), *Laboratory phonology, Vol. 8: Varieties of phonological competence* (pp. 519–548). Berlin/New York: De Gruyter Mouton.
Cho, T. (2008). Prosodic strengthening in transboundary V-to-V lingual movement in American English. *Phonetica*, *65*, 45–61.
Cho, T. (2011). Laboratory phonology. In: N. C. Kula, B. Botma, & K. Nasukawa (Eds.), *The continuum companion to phonology* (pp. 343–368). London/New York: Continuum.
Cho, T., & Keating, P. A. (2001). Articulatory and acoustic studies on domain-initial strengthening in Korean. *Journal of Phonetics*, *29*, 155–190.
Cho, T., & Keating, P. A. (2009). Effects of initial position *versus* prominence in English. *Journal of Phonetics*, *37*, 466–485.
Cho, T., Kim, J., & Kim, S. (2013). Preboundary lengthening and preaccentual shortening across syllables in a trisyllabic word in English. *Journal of the Acoustical Society of America*, *133*(5), EL384–EL390.
Cho, T., & Ladefoged, L. (1999). Variation and universals in VOT: Evidence from 18 languages. *Journal of Phonetics*, *27*, 207–229.
Cho, T., Lee, Y., & Kim, S. (2011). Communicatively driven *versus* prosodically driven hyper-articulation in Korean. *Journal of Phonetics*, *39*, 344–361.
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*, 121–157.
Cho, T., McQueen, J. M., & Cox, E. (2007). Prosodically driven phonetic detail in speech processing: The case of domain-initial strengthening in English. *Journal of Phonetics*, *35*, 210–243.
Choi, H. (2003). Prosody-induced acoustic variation in English stop consonants. In M. J. Solé, D. Recasens, & J. Romero (Eds.), *Proceedings of the 15th international congress of phonetic sciences* (pp. 2661–2664). Barcelona, Spain.
Clements, G. N., & Keyser, S. J. (1983). *CV phonology: A generative theory of the syllable*. Cambridge, MA: The MIT Press.
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*, 180–209.
Cooper, A. (1991). *Glottal gestures and aspiration in English* (Ph.D. dissertation). Yale University.
Cutler, A. (2012). *Native listening: Language experience and the recognition of spoken words*. Cambridge, MA: The MIT Press.
de Jong, K. J. (1995). The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation. *Journal of the Acoustical Society of America*, *97*, 491–504.
de Jong, K. J. (2004). Stress, lexical focus, and segmental focus in English: Patterns of variation in vowel duration. *Journal of Phonetics*, *32*, 493–516.
de Jong, K. J., & Zawaydeh, B. A. (2002). Comparing stress, lexical focus, and segmental focus: Patterns of variation in Arabic vowel duration. *Journal of Phonetics*, *30*, 53–75.
Fletcher, J. (2010). The prosody of speech: Timing and rhythm. In: W. J. Hardcastle, J. Laver, & F. E. Gibbon (Eds.), *The handbook of phonetic sciences* (second edition). Oxford: Blackwell.
Fougeron, C. (1999). Prosodically conditioned articulatory variations: a review. UCLA working papers in phonetics (Vol. 97, pp. 1-74).
Fougeron, C. (2001). Articulatory properties of initial segments in several prosodic constituents in French. *Journal of Phonetics*, *29*, 109–135.
Fougeron, C., & Keating, P. A. (1997). Articulatory strengthening at edges of prosodic domains. *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, & J. J. Raphael (Eds.), *Producing speech: Contemporary issues: For Katherine Safford Harris* (pp. 355–373). New York: AIP Publishing.
Goldstein, L. (1992). Comments on chapters 3 and 4. In: G. Docherty, & D. R. Ladd (Eds.), *Laboratory phonology, Vol. 2: Gesture, segment, prosody* (pp. 120–124). Cambridge: Cambridge University Press.
Hoole, P. (1999). Laryngeal coarticulation. Section A: Coarticulatory investigations of the devoicing gesture. In: W. H Hardcastle, & N. Hewlett (Eds.), *Coarticulation: Theory, data and techniques* (pp. 105–121). Cambridge University Press.
Kahn, D. (1976). *Syllable-based generalization in English phonology* (Ph.D. dissertation). MIT.
Keating, P. A. (1984). Phonetic and phonological representation of stop consonant voicing. *Language*, *60*, 286–319.
Keating, P. A. (1990). Phonetic representations in a generative grammar. *Journal of Phonetics*, *18*, 321–334.
Keating, P. A., Cho, T., Fougeron, C., & Hsu, C. (2003). Domain-initial strengthening in four languages. In: J. Local, R. Ogden, & R. Temple (Eds.), *Laboratory phonology, Vol. 6: Phonetic interpretation* (pp. 145–163). Cambridge: Cambridge University Press.
Keating, P. A. & Shattuck-Hufnagel, S. (2002). *A prosodic view of word form encoding for speech production*. UCLA working papers in phonetics (Vol. 101, pp. 112–156).
Keating, P. A., Wright, R., & Zhang, J. (1999). *Word-level asymmetries in consonant articulation*. UCLA Working Papers in Phonetics (Vol. 97, pp. 157–173).
Kessinger, R. H., & Blumstein, S. E. (1997). Effects of speaking rate on voice-onset time in Thai, French and English. *Journal of Phonetics*, *25*, 143–168.

Kim, S., & Cho, T. (2011). Effects of prosodic boundary *versus* accent in the articulation of English /æ/ in #VC and #CVC. In *Proceedings of the 17th international congress of phonetics sciences* (*ICPhS*) (pp. 1082–1085). Hong Kong, China.
Kim, S., & Cho, T. (2013). Prosodic boundary information modulates phonetic categorization. *Journal of the Acoustical Society of America*, *134*(1), EL19–EL25.
Kingston, J., & Diehl, R. L. (1994). Phonetic knowledge. *Language*, *70*, 419–454.
Klatt, D. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America*, *59*, 1208–1221.
Krivokapić, J. (2007). *The planning, production, and perception of prosodic structure* (Ph.D. dissertation). University of Southern California.
Krivokapić, J., & Byrd, D. (2012). Prosodic boundary strength: An articulatory and perceptual study. *Journal of Phonetics*, *40*, 430–442.
Ladefoged, P. & Cho, T. (2001) Linking linguistic contrasts to reality: The case of VOT. In: N. Gronnum & J. Rischel (Eds.), Travaux Du Cercle Linguistique De Copenhague, vol. XXXI. (To Honour Eli Fischer-Forgensen.) (pp.212-223) C.A. Reitzel, Copenhagen.
Lehiste, I. (1970). *Suprasegmentals*. Cambridge: MIT Press.
Lisker, L., & Abramson, A. S. (1967). Some effects of context on voice onset time in English stops. *Language and Speech*, *10*, 1–28.
Löfqvist, A. (1995). Laryngeal mechanisms and interarticulator timing in voiceless consonant production. In: F. Bell-Berti, & L. J. Raphael (Eds.), *Producing speech: Contemporary issues for Katherine Safford Harris* (pp. 99–116). Woodbury, NY: AIP Press.
Miller, J. L., Green, K. P., & Reeves, A. (1986). Speaking rate and segments: A look at the relation between speech production and speech perception for the voicing contrast. *Phonetica*, *43*, 106–115.
Nespor, M., & Vogel, I. (1986). *Prosodic phonology*. Dordrecht: Foris.
Oller, D. K. (1973). The effect of position in utterance on speech segment duration in English. *Journal of the Acoustical Society of America*, *54*, 1235–1247.
Ott, R. L., & Longnecker, M. (2001). *An introduction to statistical methods and data analysis* (5th ed.). Pacific Grove, CA: Duxbury Publishing.
Pierrehumbert, J., & Talkin, D. (1992). Lenition of /h/ and glottal stop. In: G. Docherty, & D. R. Ladd (Eds.), *Laboratory phonology, Vol. 2: Gesture, segment, prosody* (pp. 90–117). Cambridge: Cambridge University Press.
Port, R., & Dalby, J. (1982). C/V ratio as a cue for voicing in English. *Journal of the Acoustical Society of America*, *69*, 262–274.
Pouplier, M. (2011). The atoms of phonological representations. In M. van Oostendorp, K. Rice, B. Hume, & C. Ewen (Eds.), The Blackwell companion to phonology (pp. 107–129). Wiley-Blackwell.
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*, 51–89.
Shatzman, K. B., & McQueen, J. M. (2006). Segment duration as a cue to word boundaries in spoken-word recognition. *Perception & Psychophysics*, *68*, 1–16.
Sheskin, D. J. (2000). *Handbook of parametric and nonparametric statistical procedures* (2nd ed.). Boca Raton: Chapman & Hall/CRC.
Solé, M. J. (2007). Controlled and mechanical properties in speech: A review of the literature. In: M. J. Solé, P. Beddor, & M. Ohala (Eds.), *Experimental approaches to phonology* (pp. 302–321). Oxford: Oxford University Press.
Smiljanić, R., & Bradlow, A. R. (2008). Stability of temporal contrasts across speaking styles in English and Croatian. *Journal of Phonetics*, *36*, 91–113.
Stevens, K. N., Keyser, S. J., & Kawasaki, H. (1986). Toward a phonetic and phonological theory of redundant features. In: J. S. Perkell, & D. H. Klatt (Eds.), *Invariance and variability in speech processes* (pp. 426–463). Hillsdale, NJ: Lawrence Erlbaum Associates.
Stevens, K. N., & Keyser, S. J. (2010). Quantal theory, enhancement and overlap. *Journal of Phonetics*, *38*, 10–19.
Turk, A. E., & White, L. (1999). Structural influences on accentual lengthening in English. *Journal of Phonetics*, *27*, 171–206.
White, L. S., & Turk, A. (2010). English words on the Procrustean bed: Polysyllabic shortening reconsidered. *Journal of Phonetics*, *38*, 459–471.
Yoshioka, H., Löfqvist, A., & Hirose, H. (1981). Laryngeal adjustments in the production of consonant clusters and geminates in American English. *Journal of the Acoustical Society of America*, *70*, 1615–1623.