The use of phrase-level prosodic information in lexical segmentation: Evidence from word-spotting experiments in Korean

Sahyang Kim
Department of English Education, Hongik University, Seoul 121-791, Korea

Taehong Cho[a)]
Department of English Language and Literature, Hanyang University, Seoul 133-791, Korea

(Received 4 May 2008; revised 14 February 2009; accepted 20 February 2009)

This study investigated the role of phrase-level prosodic boundary information in word segmentation in Korean with two word-spotting experiments. In experiment 1, it was found that intonational cues alone helped listeners with lexical segmentation. Listeners paid more attention to local intonational cues (…H#L…) across the prosodic boundary than the intonational information within a prosodic phrase. The results imply that intonation patterns with high frequency are used, though not exclusively, in lexical segmentation. In experiment 2, final lengthening was added to see how multiple prosodic cues influence lexical segmentation. The results showed that listeners did not necessarily benefit from the presence of both intonational and final lengthening cues: Their performance was improved only when intonational information contained infrequent tonal patterns for boundary marking, showing only partially cumulative effects of prosodic cues. When the intonational information was optimal (frequent) for boundary marking, however, poorer performance was observed with final lengthening. This is arguably because the phrase-initial segmental allophonic cues for the accentual phrase were not matched with the prosodic cues for the intonational phrase. It is proposed that the asymmetrical use of multiple cues was due to interaction between prosodic and segmental information that are computed in parallel in lexical segmentation. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3097777]

PACS number(s): 43.71.Sy, 43.71.Es, 43.66.Lj [AJ] Pages: 3373–3386

I. INTRODUCTION

In order to comprehend spoken language successfully, listeners must be able to segment the stream of speech into individual words. The lexical segmentation process, however, is by no means trivial. Not only is there no invariant acoustic cue that consistently signals word boundaries, but there also exist multiple layers of phonetic and phonological variation within and across words which add complexity to the process of word boundary search. A long-standing question in the field of speech comprehension has therefore been how listeners find word boundaries successfully, given lack of consistent acoustic cues (see McQueen, 2005 for a review). One approach to lexical segmentation is to consider the process as a consequence of lexical competition (e.g., Marslen-Wilson and Welsh, 1978; McClelland and Elman, 1986; Norris, 1994). In lexical competition, a set of cohort competitors whose acoustic onsets are matched with the input is initially activated. Competitors are then inhibited as soon as they mismatch the input, eventually leaving a single candidate as the winner in the competition. As lexical competition ends, a search for the word boundary is also finalized, and so is lexical segmentation. Lexical competition mechanisms relying only on phonemic representations, however, could result in ambiguous parsing for a given speech stream (e.g., *I scream* vs *ice cream*, Lehiste, 1960), especially when no other semantic and/or pragmatic context is available.

A large body of recent psycholinguistic research has shown that such ambiguity can be resolved by fine-grained phonetic information in the speech input, suggesting that lexical segmentation is modulated by subphonemic information. A well-known case is the use of subtle durational difference in lexical processing. For example, ambiguous Dutch sequences due to resyllabification (e.g., *diep # in* vs *die # pin*, Quené, 1993) and lexically ambiguous sequences interpretable as two words or as one word (e.g., *two lips* vs *tulips*, Gow and Gordon, 1995) are both reliably differentiated by subtle durational cues. Similarly, temporary ambiguity that arises due to an initially embedded word in a longer word (e.g., *cap* in *captain*) is also resolved by the word's duration (Salverda *et al.*, 2003). Further evidence for the exploitation of phonetic details can be found in the use of word-internal coarticulatory information (e.g., Dahan *et al.*, 2001), assimilatory information (e.g., Gow, 2002), and phonetic differences due to the syllable structure (e.g., Tabossi *et al.*, 2000). Although these studies have successfully demonstrated that lexical segmentation is modulated by fine-grained phonetic cues that signal lexical boundaries, their focus of attention has been only on the use of the phonetic cues that are mainly associated with low-level linguistic structures of an utterance, especially in the syllable or the word levels.

[a)]Author to whom correspondence should be addressed. Electronic mail: tcho@hanyang.ac.kr

Some studies have suggested that prosodic information about lexical stress, which is generally expressed by prosodic cues such as pitch, duration, and amplitude (Lehiste, 1970), is exploited in lexical segmentation. English listeners, for instance, tend to segment words based on the strong-weak lexical stress pattern, treating the strong (stressed) syllable as the beginning of a word with a stressed syllable (Cutler and Butterfield, 1992; Cutler and Norris, 1988). Fragment priming experiments in Spanish (Soto *et al.*, 2001) showed that when the stress of the spoken fragment (e.g., *prin*) is matched with the stress of the visual target (e.g., *PRINcipe* "prince" vs *prinCIpio* "beginning"), recognition of the target word is facilitated, but inhibition occurs when stress is mismatched. Note that these studies are also confined to word-level prosodic effects on lexical processing.

Our understanding has therefore been limited with respect to how phonetic information of higher-level structure is used in lexical segmentation. In particular, although it has been well established that an utterance is produced with phonetic markers of high-level prosodic structure (e.g., Beckman, 1996; Keating and Shattuck-Hufnagel, 2002), possible roles of its acoustic consequences in lexical processing have not been fully understood. Recent studies have therefore attempted to expand their scope of investigation of lexical segmentation cues, exploring how high-level prosodic information of a given utterance influences lexical segmentation process (Cho *et al.*, 2007; Christophe *et al.*, 2004; Shukla *et al.*, 2007; Welby, 2007).

A. Prosodic structure and its importance for lexical processing

A tenet of prosodic phonology is that an utterance is produced with prosodic structure which is assumed to be organized in such a way that prosodic constituents of different sizes are hierarchically nested (see Shattuck-Hufnagel and Turk, 1996 for a review). According to a model of prosodic organization in English (Beckman and Pierrehumbert, 1986), for example, the prosodic structure consists of the syllable, the prosodic word (PW), the intermediate phrase (ip), and the intonational phrase (IP) (see Nespor and Vogel, 1986; Selkirk, 1984 for similar prosodic structural views). These prosodic domains are assumed to be strictly layered, such that a prosodic domain of one level is exhaustively parsed into constituents of the immediately next-lower level (Selkirk, 1984). The prosodic structure is known to be marked by various prosodic cues such as pause (e.g., Gee and Grosjean, 1983; Krivokapić, 2007), phrase-final lengthening (e.g., Edwards *et al.*, 1991; Wightman *et al.*, 1992), intonation (e.g., Beckman and Pierrehumbert, 1986; Ladd, 1996), and domain-initial strengthening (e.g., Cho and Keating, 2001; Fougeron and Keating, 1997; Keating *et al.*, 2003).

The structural view of prosody generally assumes that prosodic structure is a crucial element of speech production and comprehension processes (e.g., Beckman, 1996). A general hypothesis is that if the speaker produces an utterance based on prosodic structure generated online (Keating and Shattuck-Hufnagel, 2002), its acoustic consequence should be exploited by listeners in speech comprehension. Christophe *et al.* (2004), for example, demonstrated that lexical segmentation is modulated by prosodic structure. In their word-monitoring experiments, a local ambiguity (e.g., [d'un **chat grin** cheux] "of a grumpy cat," ambiguous with *chagrin*) slowed down the detection of a target word (e.g., *chat*, "cat"), but the ambiguity effect disappeared when the ambiguity-creating sequence (e.g., *chat grin*) spanned a prosodic phrase boundary. Shukla *et al.* (2007) also showed that Italian listeners were better at recognizing words that were internal to a prosodic phrase than the same syllable sequences spanning a phrase boundary. Strong articulation of segments after a prosodic boundary (i.e., domain-initial strengthening) is also known to help listeners to recognize the word before the prosodic boundary as the domain-initial strengthening serves as a cue to the beginning of a new word (Cho *et al.*, 2007).

These empirical findings demonstrate that prosodic structure, as phonetically manifested in the speech input by combination of various prosodic cues such as pause, intonation, and duration, plays an important role in lexical segmentation. What is not yet clear, however, is exactly what kind of phonetic information of the prosodic structure is exploited by the listener in lexical processing. The present study therefore explores effects of two major prosodic cues, intonation and duration, in two word-spotting experiments to see how independently or collectively these cues are used in lexical segmentation in Korean.

The prosodic model of Seoul Korean that we adopt for our study is the one proposed by Jun (1993, 1995, 2000), which is by far the most widely adopted model in the literature. The prosodic hierarchy consists of the syllable, the PW, the accentual phrase (AP), and the IP. It also assumes the strict layer hypothesis (Selkirk, 1984), so that the edges of IP always coincide with the edges of APs, which in turn coincide with the edges of PW. The Korean prosodic model differs from the English prosodic model discussed above in that it assumes the AP between the IP and the PW. Unlike the English IP, the AP is not marked by a noticeable phrase-final lengthening (Jun, 1993, 1995; cf. Cho and Keating, 2001), and its intonational structure is independent of word-level prosody, as Seoul Korean does not have any word-level prosody such as lexical stress, pitch accent, and tone (Jun, 1993).

The Korean AP is intonationally defined, having default initial and final rising intonation patterns at the edge (i.e., #LH…LH#, where "#" refers to an AP boundary). However, the intonation system interacts with AP-initial segmental information, such that an AP that starts with an aspirated or a tense consonant is associated with #HH (but otherwise with #LH, including AP that starts with a vowel). As the AP is produced with no discernible final lengthening at the end, substantial final lengthening is associated only with the IP (Jun, 1993; Chung *et al.* 1996). With respect to boundary tones that mark the end of an IP, various intonation patterns have been identified such as L%, H%, LH%, and HL% ("%" refers to an IP-final tone), all of which occur in the IP-final syllable (Jun, 2000). When an AP is located IP-finally, the AP-final tone is overridden by the IP-final boundary tone.

AP-initial tones, however, are preserved regardless of the position of an AP within an IP since the IP does not have any initial boundary tone.

Some researchers have suggested that AP-internal intonational structure in Seoul Korean may be further constrained by the structure of the AP-initial syllable (with or without a coda consonant) and the vowel quantity (long vs short) (e.g., Lim and de Jong, 1999; Park, 2004). The vowel quantity distinction, however, is not maintained anymore in Seoul Korean, and underlying intonational structure of AP assumed by Jun (1993) is still supported by the statistics of corpus studies (Jun and Fougeron, 2000; Kim, 2004). Kim's (2004) study, which transcribed Korean intonation patterns in read speech and radio drama, showed that when AP-initial consonants were not aspirated or tense, about 88% of AP-initial multisyllabic content words started with a rising (LH) tone and about 85% of APs ended with a final H tone. Crucially, this frequency pattern was observed regardless of the AP-initial syllable structure. Other tonal patterns such as #LL, #HL, or #HH did occur, but with a very low frequency (9% for #LL, 2% for HH, and 1% for #HL). The notion of AP which is intonationally-defined as a prosodic unit has been further supported by segmental phonology—i.e., AP serves as an application domain of some phonological rules. Jun (1993), for example, showed that the lenis stop intervocalic voicing rule (where a lenis stop becomes voiced between vowels) applies within an AP, although voiced variants do occur sometimes in AP-initial position (cf. Cho and Keating, 2001).[1] Other phonological rules operating within an AP include post-obstruent tensing (a lenis stop becomes tense after an obstruent: Jun, 1998), lateralization (/n/ becomes [l] after /l/: Kim, 2000) and *n*-insertion (/n/ is inserted stem-initially in stems that begin with /i/ or /j/ when it is preceded by a stem or a prefix ending with a consonant: Kim, 2000).

To recap, we selected Seoul Korean as our target language because it is prosodically interesting in two ways. First, it does not have any word-level prosody (Jun, 1995), so the intonation structure of an utterance is determined solely at the phrase level without any influence from word-level prosody. Second, the medium-sized phrase AP is not accompanied by substantial final lengthening at the end (Jun, 1993, 1995). These unique properties of Korean prosody allow us to observe the role of intonational cues of AP in lexical segmentation without any confounding effects from word-level prosody and other domain-edge phenomena.

There have, in fact, been attempts to understand the role of AP-initial (postboundary) tones in lexical segmentation. For example, Warner *et al.* (2009) and Kim (2004) showed that speakers of Japanese and Korean, respectively, can use the AP-initial rising intonational cue in online lexical processing. In both studies, however, potential prosodic cues (domain-initial strengthening and boundary-adjacent lengthening) other than AP-initial intonational cues were not completely eliminated. The present study is therefore the first that controls the experimental condition in such a way that the sole effect of intonation patterns of AP in online segmentation can be observed without any confounding effects that would otherwise stem from other possible phonetic cues available at prosodic boundaries. In Sec. I B, we will discuss specific research questions that are to be addressed in the present study.

B. Research questions

In the present study, two word-spotting experiments are carried out to examine the use of prosodic information in lexical segmentation. In experiment 1, we explore how intonational cues alone influence lexical segmentation and in experiment 2, we add phrase-final lengthening as an additional cue to understand how single vs multiple prosodic cues are processed by listeners in lexical segmentation.

In experiment 1, both preboundary and postboundary AP intonational sequences are considered. As for AP-initial (postboundary) tones, we compare the effects of four different intonation patterns superimposed upon target words: frequent #LH vs infrequent #LL, #HH, and #HL. (They are used for disyllabic target words; see below for tonal descriptions of trisyllabic target words.) As for AP-final (preboundary) tones before the target word, two intonation patterns, frequent H# vs infrequent L#, are used. Crucially, all the consonants used as initial segments of the target words are either lenis stops or nasals, which are extracted from AP-medial position in order to eliminate other potential prosodic cues such as domain-initial strengthening cues that might affect the listener's performance (Cho *et al.*, 2007).

We specifically ask how the frequency of intonational cues influences listeners' performance in lexical segmentation. It is well established that frequency influences lexical access in terms of word frequency (Forster and Chambers, 1973; Norris, 1986) and sequential probabilities (e.g., Saffran, *et al.*, 1996; Vitevitch, *et al.*, 1997). The statistics of stress patterns within the vocabulary (i.e., stress tends to fall on initial syllables in English and Dutch: Cutler and Carter, 1987; Schreuder and Baayen, 1994) also influences lexical access—i.e., listeners tend to put a word boundary before a stressed syllable (Cooper *et al.*, 2002; Cutler and Butterfield, 1992; Vroomen and de Gelder, 1995). It is therefore reasonable to hypothesize that the frequency of intonation patterns also influences detection of a phrase boundary which coincides with a lexical boundary, such that listeners will perform better with target detection with frequent intonation patterns for AP than less frequent intonation patterns.

Another important question is whether the preboundary and the postboundary intonation patterns are exploited independently by the listener. Previous studies have focused on the role of within-phrase prosodic cues to a prosodic boundary. For instance, Welby (2007) and Warner *et al.* (2009) examined the role of phrase-initial prosodic cues, and the work of Christophe *et al.* (2004) was based on the assumption that listeners use within-phrase prosodic information to terminate their lexical search before a prosodic boundary. It is then possible to posit that intonation patterns within each prosodic phrase may play an independent role in speech processing. Under this hypothesis, the detection of target words is expected to be more facilitated with the frequent tonal pattern (#LH) than with the infrequent patterns (#LL, #HH, and #HL), regardless of whether the preboundary tone is H#

or L#. Likewise, the frequent AP-final rising tone (with H#) is expected to signal the end of AP, which, at the same time, indicates that what is coming up is the beginning of another AP. The presence of the frequent AP final tone (H#) before the boundary is therefore expected to help listeners' recognition of the postboundary target word, regardless of the postboundary tones. Alternatively, however, if one of the goals of prosodic structuring in speech production were to mark prosodic boundaries, which would be eventually available to the listener in speech comprehension, the crucial prosodic information might be present locally at the prosodic boundary (just before and after it), with H#L being the most frequent pattern. If so, there would be an interaction between the preboundary and the postboundary intonational effects, in such a way that target detection would be facilitated as long as the local boundary condition (H#L) is met. It is then expected that #LH and #LL, which differ in frequency but both meet the locality condition, would show similar facilitatory effects on the target detection as long as the preboundary tone is H#.

Examining intonational effects on lexical processing in Korean also allows us to address the general vs language-specific perceptual role of a high (H) pitch element in lexical segmentation. Intonational cues are often associated with an H pitch element or an F0 rise. Warner *et al.* (2009) therefore suggested that F0 rise is perceptually salient, and it would facilitate detection of syllables marked by it. Under this assumption, an H tone on the initial syllable of the target word would facilitate lexical segmentation regardless of listeners' linguistic background. Interestingly, however, Korean has an H tone, but it frequently falls on the second syllable of AP, providing a counter example to this assumption. Based on the results of the present study, we will address this issue, especially in terms of how the language-specific distribution of H in Korean bears on the issue of language-specific vs cross-linguistic use of the perceptually salient F0 rise in lexical processing.

In experiment 2, a final-lengthening factor is added. As in experiment 1, two preboundary tones (H# vs L#) at the preboundary syllables are used. For the postboundary tones, for the sake of simplicity, just two extreme conditions from experiment 1 are used: #LH (the most frequent) and #HL (the least frequent).

As phrase-final lengthening is known to be another important boundary-marking phonetic event that co-occurs with boundary-marking tones (e.g., Edwards *et al.*, 1991; Wightman *et al.*, 1992), some researchers have suggested that listeners make use of subtle phonetic differences and compute prosodic structure of an utterance during online word recognition (e.g., Christophe *et al.*, 2004; Salverda *et al.*, 2003; Shatzman and McQueen, 2006). We therefore test how intonational effects interact with final lengthening cues. Important questions to be addressed are how cumulative multiple prosodic cues facilitate lexical processing and how the cumulative effect is constrained by a mismatch between segmental and prosodic cues.

Spitzer *et al.* (2007) showed that the level of intelligibility gets lower when more cues of lexical stress are missing in the speech signal, suggesting that available cues are used in a cumulative way. It has been also proposed that listeners make immediate use of any available cues in order to modulate the activation of lexical competitors (Donselaar *et al.*, 2005), and all the available information is used (Norris *et al.*, 1997). Cumulative boundary cues could therefore be a very effective tool to modulate lexical search: Listeners would be able to guess the location of the end of a prosodic phrase with more certainty when the final lengthening cue is added to the intonational cue. It is therefore hypothesized that the addition of substantial final lengthening will augment the effect of intonational cues, facilitating listeners' target detection in a cumulative way.

The presence of substantial lengthening at the end of a phrase would, in fact, give rise to a percept of an IP boundary before the target word. However, adding the lengthening cue to the speech input would not make a seamless IP boundary percept in the current experiment because a mismatch would arise between segmental and prosodic information. Recall that in the present study, we use speech materials extracted from phrase-internal (AP-medial) position in order to eliminate potential prosodic cues at domain edges. IP-initially, consonant durations (including VOTs for lenis stops) are longer and lenis stops are always voiceless (no application of lenis stop voicing rule across an IP boundary). Yet, in experiment 2, initial consonants of the target words lack such domain-initial strengthening cues, including lenis stop voicing cues. That is, the consonants that listeners hear are not consistent with IP-initial, but compatible with IP-medial position. Such a mismatch between prosodic and segmental information might hinder the detection of the target words (e.g., Cho *et al.*, 2007; Salverda *et al.*, 2003).

There are therefore two competing hypotheses. On the one hand, if the cumulative effect takes precedence in lexical segmentation process, the presence of an additional phrase-final lengthening would augment listener's performance with the target detection. On the other hand, if the effect of the mismatch between prosodic and segmental cues comes into play, it would at least suppress the cumulative effect (showing no further facilitation with the presence of phrase-final lengthening) or under its strongest influence, it could override effects of both intonational and lengthening cues.

In experiments 1 and 2, we therefore test various hypotheses in order to understand the relationship between high-level prosodic information and lexical segmentation. Building on our knowledge about the relationship will also have an important implication for existing models of speech segmentation such as TRACE (McClelland and Elman, 1986), shortlist (e.g., Norris, 1994; Norris and McQueen, 2008), the distributed cohort model (Gaskell and Marslen-Wilson, 1997), and the hierarchical model (Mattys *et al.*, 2005) because these models do not take high-level prosodic information into account.

II. EXPERIMENT

In experiment 1, we carried out a word-spotting experiment in Korean in order to explore how prosodic structural information manifested in intonation influences lexical seg-

TABLE I. Intonation patterns of carrier strings and target words. (# indicates an assumed phrase boundary; underlined syllables in bold indicate target words.)

(a) A seven-syllable carrier string with a disyllabic target word					
First two syllables		Preboundary syllable	Disyllabic target word (postboundary word)	Last two syllables	
σ1	σ2	σ3	**σ4 σ5**	σ6	σ7
L	L	L #	**LH**	L	L
		H #	**LL**		
			HH		
			HL		
(b) An eight-syllable carrier string with a trisyllabic target word					
First two syllables		Preboundary syllable	Trisyllabic target word (postboundary word)	Last two syllables	
σ1	σ2	σ3	**σ4 σ5 σ6**	σ7	σ8
L	L	L #	**LLH**	L	L
		H #	**LLL**		
			HHL		
			HLH		

mentation. We tested whether intonation patterns that are frequently associated with AP boundaries would facilitate detection of lexical boundaries.

A. Method

1. Participants

Ninety-six student participants from Hanyang University in Seoul were paid for their participation. They were all native speakers of Seoul Korean who were born and raised in the Seoul metropolitan area. They were divided into eight groups of 12, according to experimental conditions that will be described below.

2. Materials

24 disyllabic and 24 trisyllabic Korean words were selected and inserted, respectively, in seven- and eight-syllable nonsense carrier strings. Other than the target word, no consecutive syllables in a carrier string formed a word in Korean. Target words and target-bearing carrier strings were composed of open syllables (CV) only. 14 out of 24 disyllabic words and 15 out of 24 trisyllabic words started with an oral lenis stop (/p/, /t/, or /k/), and the rest started with a nasal stop (/m/ or /n/). Lenis stops and nasals were chosen because they are known to be associated with the AP-initial rise tone (#LH) (Jun, 1993, 2000), which was confirmed by Kim's (2004) corpus study. The list of target words and carrier strings is given in the Appendix.

As shown in Table I, the initial syllable of a target word was always the fourth syllable of carrier strings, such that each target word was preceded by three syllables and followed by two syllables. We used two- and three-syllable target words to avoid any potential performance bias that comes from a fixed number of syllables of the target word. This also allows us to see potential effects of the syllable count, as the number of syllables in target words was found to affect listeners' word segmentation (Kim, 2004). The carrier strings with a disyllabic target word had seven syllables [Table I(a)] and those with a trisyllabic target word had eight syllables in total [Table I(b)]. Each syllable in a carrier string was associated with one tone, which was either H or L. The first two syllables and the last two syllables of carrier strings were controlled with an L tone. In order to observe preboundary tone effects on word segmentation, the syllable that immediately preceded the target word in each carrier string (viz., σ3 #, where # refers to the intended phrase boundary) had a preboundary tone of either H# (frequent) or L# (infrequent).

For intonation patterns on target words, one of the four postboundary tones was used. The frequent pattern for disyllabic target words was a rising tone with #LH and the infrequent patterns were #LL, #HH, and #HL. When the target word was trisyllabic, three tonal elements were needed for each word. The four intonation patterns employed were #LLH (frequent), #LLL, #HHL, and #HLL (infrequent). They were then matched with the infrequent disyllabic intonation patterns in the first two tonal elements: #LL vs #LLL, #HH vs #HHL, and #HL vs HLL. For the sake of simplicity, we will use the disyllabic intonation patterns (#LH, #LL, #HH, and #HL) throughout the paper when we refer to the intonation patterns of the target words. An important note here concerns the predicted effect of the infrequent #LL for disyllabic targets vs the frequent #LLH for trisyllabic targets. One of the hypotheses for the frequency effect was that the frequent AP-internal tonal pattern #LLH for trisyllabic targets would help listeners to detect the targets. However, given that the frequent #LLH shares the same initial #LL sequence with the infrequent #LL for a disyllabic target, one might wonder how listeners are expected to treat #LL as an infrequent cue for disyllabic targets but as part of the frequent cue (#LLH) for a trisyllabic target. In the experiment, an equal number of disyllabic and trisyllabic targets was employed. Listeners, therefore, did not know *a priori* whether the target word would be disyllabic or trisyllabic. As the second L tone of #LL is being heard, the likelihood for the target to be disyllabic would become weakened because #LL as a whole is not a frequent tone for disyllabic words. As the third tone H is being heard, however, it is matched with the frequent #LLH pattern for trisyllabic words, and therefore the likelihood for the target to be trisyllabic would increase. In this way, if the frequency of the phrase-internal tonal patterns would play a role in lexical segmentation, the LL portion would not work in favor of disyllabic targets, but it would help detect trisyllabic targets in the form of #LLH, even if listeners did not know beforehand the number of syllables for the targets.

Since each target word appeared in eight different intonational conditions (2 preboundary tones × 4 postboundary tones), there were eight experimental lists. In addition to the experimental items, 36 disyllabic and 36 trisyllabic Korean words were selected as fillers and were included in the experimental lists. The location of a filler word in filler-bearing strings, however, differed from that of a target word in experimental target-bearing strings in order to avoid the potential bias caused by the fixed location of both the experimental target and filler words. As with the target-bearing strings, two preboundary tones and four postboundary tones were

TABLE II. Mean error rates (%) in experiment 1 with three factors, syllable count, preboundary tone, and postboundary (target word) tone. (Standard errors are included in parentheses.)

No. of syllables in target words	Preboundary tone	Postboundary (target word) tone			
		#LH	#LL	#HH	#HL
2	H#	38.5	45.1	54.5	62.1
		(2.1)	(2.2)	(2.5)	(2.5)
	L#	48.6	54.8	49.6	62.1
		(2.5)	(2.4)	(2.6)	(2.8)
3	H#	15.2	19	24.6	28.8
		(1.9)	(2)	(2.2)	(2.4)
	L#	22.2	23.2	23.6	31.2
		(2.2)	(2.2)	(2.1)	(2.6)

evenly distributed among filler-bearing strings. Additional eight words were selected and inserted in carrier strings for practice items.

Eight experimental lists were arranged such that each subject heard every word just once in one of the eight intonational conditions. Each list contained 48 target-bearing strings and 72 filler-bearing strings in a pseudo-random order, and the filler strings appeared in the same order in all eight lists. There were no two stimuli with the same intonation pattern presented in a row.

To ensure that the carrier string did not have any other confounding prosodic factors on crucial syllables, such as phrase-final lengthening on the preboundary syllable (σ3#) and acoustic consequences of domain-initial strengthening on the postboundary syllable (the initial syllable of target words; #σ4), the following recording and splicing procedure was performed.

Each seven-syllable string (with a disyllabic target word) was divided into three four-syllable chunks ([σ1-σ2-σ3-σ4], [σ3-**σ4-σ5**-σ6], and [σ4-σ5-σ6-σ7]). (Here the underlined syllable symbols in bold were the ones that were actually used for target words.) A female native speaker of Seoul Korean who was naive about the experiment's purpose produced each chunk separately multiple times with flat intonation and with as consistent speaking rate as possible. The recording was made in a sound-proof booth onto a TASCAM HD-P2 digital recorder at the sampling rate of 44 kHz. Among the multiple tokens, the best tokens were selected that did not contain any unintended prosody, agreed by transcriptions of the two authors. The recorded strings were then spliced to be used for the actual speech input string as follows. The first three syllables of a seven-syllable carrier string were spliced from [σ1-σ2-σ3-σ4] so that the preboundary syllable (σ3#) of the actual input string (σ1σ2σ3#**σ4σ5**σ6σ7) did not involve phrase-final lengthening. Target words (σ4σ5) were spliced from [σ3-**σ4-σ5**-σ6] so that the initial syllable (#σ4) of target words did not have characteristics of domain-initial strengthening.[2] Finally, the last two syllables of the actual input string were the last two syllables of [σ4-σ5-σ6-σ7]. A similar procedure was employed to build eight-syllable strings with trisyllabic target words. Filler-bearing strings and practice items were also recorded and spliced in the same manner. The splicing was made at zero crossings, using PRAAT.

After splicing was completed, the pitch of each carrier string was manipulated using the pitch-synchronous overlap and add technique with PRAAT software. The speaker's flat intonation was regarded as a default L tone (143 Hz), and pitch was raised when H tone was required in the experimental setting. F0 minima for an L tone and F0 maxima for an H tone were aligned with the midpoint of the vowel of the target syllables. The rates of pitch rising were 1.23 times at the preboundary syllable (σ3#) and 1.16 times at the target-word final syllable (i.e., σ5 in disyllabic words and σ6 in trisyllabic words). The rates were determined by the rate of AP-final and AP-initial pitch rising from other recordings of the speaker with natural sentences.

3. Procedures

Subjects were tested individually in a sound-attenuated room. Stimuli presentation and data collection were performed by NESU software and a button box (www.mpi.nl/world/tg/ experiments/nesu.html). Subjects heard the stimuli on a PC through a pair of headphones at a comfortable volume. They were told that they would hear a list of nonsense strings and were instructed that they should spot a real Korean word in each string. They were asked to press a button with their preferred hand as quickly and as accurately as possible when they spotted a real word, and then to say the word aloud. Each subject was presented with eight practice items and was then given one of the eight experimental lists. There were 120 strings in each list. Subjects heard 70 strings (for approximately 12 min), took a 1-min break, and then continued with the rest of the strings (for approximately 8 min). An experimenter was always in the room with a subject during the experimental session and monitored the subject's missing or incorrect responses. Subjects had to detect words in both target-bearing and filler-bearing strings, but their responses for filler-bearing strings were not analyzed.

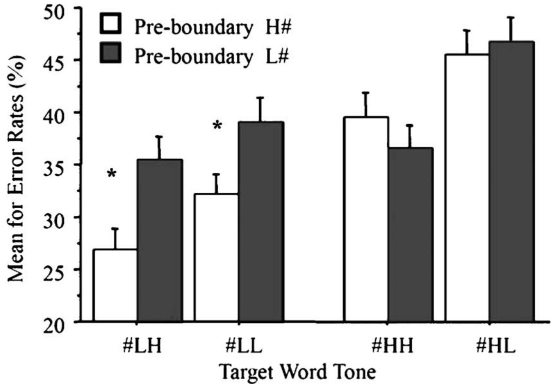

FIG. 1. Error rates in different tone conditions with an interaction between preboundary tone and postboundary (target word) tone. (* refers to $p<0.05$ in *posthoc* analyses.) The postboundary tones are #LH (#LLH), #LL (#LLL), #HH (#HHL), and #HL (#HLL), where non-parenthetical tones are for disyllabic words and parenthetical tones are for trisyllabic words. Note that the syllable count factor did not interact with either preboundary tone or postboundary tone.

B. Results

Reaction time (RT) was the duration between the offset of the target word and button press. Missing items, incorrect responses, and RT over 1500 ms were treated as errors. Mean error rates are summarized in Table II.

The error rates were very high especially when target words were disyllabic, as shown in Table II. Over 50% of target words (51.9% error rates on average) were missed in the disyllabic conditions and 23.5% in the trisyllabic conditions. Because overall error rates were very high, the latency analyses were not entirely reliable, although the latency results were by and large comparable with the accuracy results. In the present study, we will therefore report just results of the accuracy analyses. For the accuracy analyses, the error rates were submitted to repeated measures analyses of variance (ANOVAs) with the factors syllable count (number of syllables in target words; 2 vs 3), preboundary tone (frequent H# vs infrequent L#), and postboundary (=target word) tone (frequent #LH vs infrequent #LL, #HH, and #HL).

All three factors showed significant main effects. Subjects were more accurate when target words were trisyllabic than when they were disyllabic ($F1[1,88]=442.25$, $p<0.001$; $F2[1,46]=14.56$, $p<0.001$). Target words were detected with significantly lower error rates when preboundary tone was H# (frequent) than when it was L# (infrequent) ($F1[1,88]=9.84$, $p<0.005$; $F2[1,46]=5.12$, $p<0.05$). The postboundary tone effect was significant ($F1[3,264]=29.11$, $p<0.001$; $F2[3,138]=15.01$, $p<0.001$), and Bonferroni *posthoc* tests showed an overall pattern of #LH<(#LL=#HH)<#HL. #LH was significantly different from the other three conditions ($p<0.05$ both by-subjects and by-items), showing the lowest error rates. The two infrequent intonation patterns #LL and #HH were also significantly different from the least frequent pattern #HL ($p<0.05$ both by-subjects and by-items), while #LL and #HH were not significantly different from each other.

There was an interaction between preboundary tone and postboundary tone ($F1[3,264]=5.73$, $p<0.005$; $F2[3,138]=5.52$, $p<0.005$), as illustrated in Fig. 1. The effect of preboundary tone was significant only when postboundary tone was #LH or #LL (with #LH, $F1[1,95]=9.73$, $p<0.005$, $F2[1,47]=18.81$, $p<0.001$; with #LL, $F1[1,95]=6.84$, $p<0.05$, $F2[1,47]=8.86$, $p<0.01$), showing that listeners' better performance with preboundary tone H# was reliable only when postboundary tone started with #L (viz., #LH or #LL). Likewise, the effect of postboundary tone was reliable only when preboundary tone was H#. In other words, when preboundary tone was H#, the order of error rates by postboundary tone was (#LH=#LL)<(#HH=#HL) ($p<0.01$ by-subjects, $p<0.05$ by-items), but when preboundary tone was L#, the order of error rates by postboundary tone was (#LH=#LL=#HH)<#HL ($p<0.01$ by-subjects, $p<0.01$ by-items), with #LL being no different from #HL. There was no difference between #LH and #LL regardless of preboundary tone ($p>0.05$ both by-subjects and by-items).

C. Summary and discussion

In experiment 1, a general finding was that the preboundary and postboundary tones that are used most frequently to mark AP boundaries in speech production indeed help listeners to recognize words: The detection accuracy was higher when the preboundary tone ended with the frequent H# (vs L#) and when the postboundary tone started with the frequent #L (#LH, #LL vs #HH, #HL). This supports the basic hypothesis that the frequency of intonation patterns for AP is exploited by listeners, such that frequent intonation patterns facilitate lexical segmentation.

The results, however, showed interactions between the preboundary and the postboundary tones. The frequent preboundary H# improved listeners' word-spotting performance, but this effect held only when the postboundary tone (on the target word) also started with the frequent intonational element #L, but with no difference between #LH (frequent) and #LL (infrequent). Likewise, the most frequently occurring postboundary tone #L was found to be useful only when the preceding (preboundary) tone was also most frequent (i.e., with H#), but again with no difference between #LH (frequent) and #LL (infrequent). That is, as long as the first element of the postboundary tone was #L, listeners' performance was not influenced by the second tonal element of the postboundary intonational sequence (#LH, #LL). This supports the hypothesis that what is important is whether adjacent intonational elements form an H#L sequence locally across the boundary, rather than the individual patterns of preboundary and postboundary intonational sequences. Insofar as the boundary-spanning local information (H#L) was available to the listener, what comes after that does not seem to influence listeners' boundary detection in a noticeable way.

One might wonder why the second element of the postboundary intonational sequence (i.e., H in the frequent #LH) does not contribute to lexical segmentation, especially when the local intonational condition (an H#L sequence) is met. One possible explanation for this is as follows. Upon hearing the preboundary H#, listeners would start to entertain the possibility of a phrase boundary, and the subsequent postboundary #L would confirm their boundary decision (as #L is the typical AP-initial tone). Lexical search would then be initiated and cohort competitors would be simultaneously activated from the point when listeners make a decision about where to put a boundary. Whichever tone pattern follows the postboundary #L (e.g., #LH vs #LL), the same members of cohort are already activated, and therefore the effect of the second intonational element would be the same on all the cohort competitors. This is indeed in line with the explanation of Cho *et al.* (2007). Domain-initial strengthening cues immediately after a prosodic boundary (e.g., lengthened VOTs) would not necessarily facilitate the recognition of the postboundary word because the heightened acoustic clarity of domain-initial position would be beneficial to not only the target word but also all other cohort competitors that share the same initial consonant.

Finally, results showed that listeners were more accurate in segmenting words when they were trisyllabic than when they were disyllabic. As in Kim (2004), this can be accounted for by the different neighborhood density of target words (e.g., Luce, 1986; Luce and Pisoni, 1998)—i.e., trisyllabic words used in the present study had less neighboring words to compete against than disyllabic words had. A *posthoc* analysis on the neighborhood density supports this account. The number of phonological neighbors of 48 target words was calculated based on Kim and Kang (2004)'s corpus, which had 550,000 listed words. A phonological neighbor (i.e., a phonologically similar word) was defined by an addition, deletion, or substitution of a segment to a target word regardless of the location of a segment within a word. The calculated result showed that disyllabic targets had about 15 times more phonological neighbors than trisyllabic targets (mean, 49, s.d., 22.2 vs mean, 3.3, s.d., 2.1, respectively). The syllable count effect did not interact with intonational effects, suggesting its independence.

III. EXPERIMENT 2

In experiment 2, we tested how intonational effects observed in experiment 1 would interact with final lengthening cues. Important questions were whether multiple prosodic cues (intonation and duration) would cumulatively facilitate lexical processing, and how the cumulative effect would be constrained by a mismatch between segmental and prosodic cues—i.e., an IP boundary percept created just by final-lengthening and intonational cues at the boundary mismatched with domain-initial strengthening cues (including lenis stop's allophonic cues).

A. Method

1. Participants

Ninety-six student participants from Hanyang University were paid for their participation. They were all native speakers of Seoul Korean born and raised in the Seoul metropolitan area and had not participated in experiment 1. They were divided into eight groups of 12, according to the experimental conditions that will be explained below.

2. Materials

We employed two degrees of lengthening (lengthening vs no lengthening) and two preboundary tones (H# vs L#) at the preboundary syllables. For the postboundary tones, for the sake of simplicity, we included just two extreme conditions from experiment 1: #LH, which showed the most robust effect, and #HL, which showed the smallest effect. For the lengthening condition, the vowels of the preboundary syllables were lengthened, resulting in about 1.7 times of the original syllable's vowel length. The range of syllable durations was 178–259 ms in the no lengthening condition and 256–386 ms in the lengthening condition. The rate of lengthening was determined based on multiple speakers' mean values reported by Chung *et al.* (1996)—i.e., IP-final syllables are about 1.7 times longer than non-final syllables. (Note that unlike this lengthening manipulation, the pitch manipulation in experiment 1 was based on the talker's own recordings as pitch range is talker-specific, generally constrained by the talker's physiological characteristics.) PRAAT was used for duration manipulation. The two authors, as trained Korean intonation transcribers, agreed that the lengthened versions of H# and L# gave IP-final percepts as H% and L%, respectively. Target words (both disyllabic and trisyllabic), filler words, and segmental order of carrier strings were the same as those used for experiment 1. In total, there were two levels of lengthening (lengthening vs no lengthening), two preboundary tones (H# vs L#), and two postboundary tones [#LH vs #HL: Recall that tonal markings of IP-initial APs are the same as those of IP-internal APs (Jun 1993, 2000)]. Thus, each target word appeared in eight different combinations of these factors ($2\times2\times2$), yielding eight experimental lists.

As in experiment 1, eight experimental lists were arranged such that each subject heard every word just once, in one of the eight conditions. Each list contained the same number of target- and filler-bearing strings, and they were presented in the same pseudo-random order as in experiment 1. No two stimuli with the same prosodic condition were presented in a row.

3. Procedures

The procedure of experiment 2 was the same as that of experiment 1.

B. Results

Missing items, incorrect responses, and RTs over 1500 ms were treated as errors. Mean error rates are summarized in Table III. As was the case with experiment 1, only the results of the accuracy analyses are reported here, as the RT data were not entirely reliable due to high error rates, especially in disyllabic conditions (mean error rate, 48.7%).

The error rates were submitted to repeated measures ANOVAs with the factors syllable count (2 vs 3), preboundary tone (frequent H# vs infrequent L#), postboundary (=target word) tone (frequent #LH vs infrequent #HL), and lengthening (lengthening vs no lengthening).

TABLE III. Mean error rates (%) in experiment 2 with four factors, syllable count, preboundary tone, postboundary (target word) tone, and lengthening. (Standard errors are included in parentheses.)

No. of syllables	Preboundary tone	Postboundary #LH		Postboundary #HL	
		Lengthening	No lengthening	Lengthening	No lengthening
2	H	39.2 (2.4)	30.5 (2.6)	53.4 (2.6)	59.3 (2.4)
	L	39.9 (2.5)	44.4 (2.4)	57.6 (2.4)	65.2 (2.2)
3	H	13.8 (1.7)	11.8 (1.3)	15.9 (1.7)	23.2 (2.2)
	L	11.1 (1.5)	18.7 (1.9)	23.9 (2)	29.5 (2.3)

All four factors showed significant main effects. Error rates were significantly lower in the trisyllabic condition than in the disyllabic condition ($F1[1,88]=614.74$, $p<0.001$; $F2[1,46]=19.84$, $p<0.001$); when preboundary tone was H# than when it was L# ($F1[1,88]=26.41$, $p<0.001$; $F2[1,46]=20.96$, $p<0.001$); and when postboundary tone was #LH than when it was #HL ($F1[1,88]=167.70$, $p<0.001$; $F2[1,46]=37.86$, $p<0.001$). The effects of the three factors (syllable count, preboundary tone, and postboundary tone) were therefore in line with the results found in experiment 1. The lengthening effect was also significant ($F1[1,88]=11.54$, $p<0.005$; $F2[1,46]=5.169$, $p<0.05$), showing that listeners detected target words more accurately when the preboundary syllable was lengthened than when it was not.

There was an interaction between syllable count and postboundary tone ($F1[1,88]=24.14$, $p<0.001$; $F2[1,46]=5.3$, $p<0.001$). *Posthoc* analyses showed that the interaction stemmed from the differential effect size of postboundary tone depending on the syllable count. The error rates were lower for #LH than for #HL in both disyllabic ($F1[1,95]=95.92$, $p<0.001$; $F2[1,23]=24.57$, $p<0.001$) and trisyllabic conditions ($F1[1,95]=38.57$, $p<0.001$; $F2[1,23]=13.61$, $p=0.001$), but the effect was greater for disyllabic target words (mean difference: 20.4%, $eta^2=0.502$) than for trisyllabic ones (mean difference: 9.3%, $eta^2=0.289$) at $p<0.01$ both by-subjects and by-items.

There were also interactions between the lengthening factor and the intonational factors. Lengthening interacted with preboundary tone ($F1[1,88]=7.14$, $p<0.01$; $F2[1,46]=6.92$, $p<0.05$) and with postboundary tone ($F1[1,88]=8.08$, $p<0.01$; $F2[1,46]=5.42$, $p<0.05$). *Poshoc* analyses showed that the lengthening effect was reliable only when the preboundary tone was the infrequent L# ($F1[1,95]=13$, $p<0.001$; $F2[1,47]=10.96$, $p<0.005$) and when the target word tone was the infrequent #HL ($F1[1,95]=15.88$, $p<0.001$; $F2[1,47]=8.7$, $p<0.01$). However, there was also a significant three-way interaction ($F1[1,88]=8.09$, $p<0.01$; $F2[1,46]=6.29$, $p<0.05$). As shown in Fig. 2, the interaction came from the fact that, although lengthening reduced error rates in L#HL, H#HL, and L#LH (all at $p<0.05$ both by-subjects and by-items), the opposite was true for H#LH ($p<0.05$ both by-subjects and by-items): While the presence of final lengthening was generally helpful in most cases, its presence made it harder for listeners to detect target words with the frequent intonation pattern H#LH.

C. Summary and discussion

The results of experiment 2 confirmed the effects of syllable count, preboundary tone, and postboundary tone, consistent with those in experiment 1. In addition, the main effect of lengthening suggests that substantial phrase-final lengthening can also serve as a helpful segmentation cue for Korean listeners. The interaction between lengthening and preboundary tone, however, revealed that the presence of phrase-final lengthening was not useful when the preboundary tone was a frequent AP-final marking tone (H#). It was useful only when the preboundary tone was infrequent L# for AP-final marking. It appears that, when the preboundary L# gives rise to misparsing errors (as L# is likely to be perceived as the AP-initial tone) as observed in experiment 1, final lengthening becomes operative and helps listeners to reduce such misparsing errors.

The three-way interaction in the error rate analysis showed a more detailed interaction between lengthening and intonational cues. In the less frequent tone conditions (H#HL, L#LH, and L#HL), the detection performance was

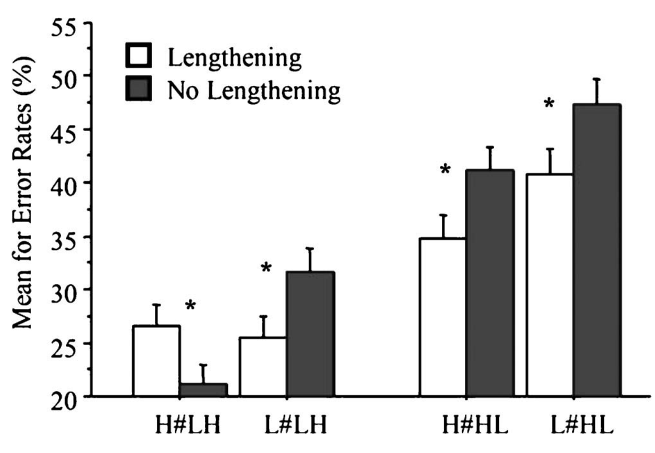

FIG. 2. Error rates with lengthening effects in different intonational conditions. (* refers to $p<0.05$ in *posthoc* analyses.)

better with than without lengthening. But the opposite was true for the most frequent tone condition, H#LH. In this condition, listeners' accuracy was significantly better without lengthening. This result works against our initial hypothesis that listeners would make use of available prosodic cues in a cumulative way in detecting a prosodic boundary, but it appears to be in favor of the alternative hypothesis that the mismatch between prosodic and segmentation information would take precedence over the advantage of the additional final-lengthening cue (see Sec. IV for further discussion on this point).

Finally, it is worth addressing an issue about detection error rates that were found to be very high in both experiments, especially with disyllabic target. Although error rates are generally high in word-spotting tasks (e.g., McQueen, 1996), one might be concerned that such high error rates could reduce the interpretability of findings of the present study. Statistical analyses, however, suggest that our results are still reliably interpretable. The main findings in detection accuracy were statistically robust in both by-subjects and by-items analyses. More crucially, error rates with trisyllabic targets were far lower than those with disyllabic targets, but there were no interactions at all between crucial prosodic factors and the syllable count. That is, patterns found with disyllabic targets were statistically the same as those with trisyllabic targets, indicating that findings of the present study were not biased due to high error rates with disyllabic targets.

IV. GENERAL DISCUSSION

This study examined the role of prosodic cues in online word segmentation of Korean. The prosodic cues under investigation in two word-spotting experiments were language-specific intonational cues and a phrase-final lengthening cue.

The results of experiment 1 showed that listeners detected the target word better when accompanied by intonation patterns frequently associated with AP boundaries (forming an H#L sequence locally at AP boundary). They suggest that intonation patterns which are frequently used in marking an AP-boundary in speech production are indeed exploited by listeners in processing the speech signal. The results also revealed an interesting interaction between the preboundary and the postboundary intonation patterns. The facilitatory frequency effect of the postboundary #L (for both #LH and #LL) was robust only when the preceding tone was also frequent (with H#). Likewise, the facilitatory frequency effect of the preboundary H# was reliable only when the following (postboundary) tone was also frequent (with #L). Moreover, although #LH is far more frequent than #LL, the difference was not significant insofar as the local H#L boundary condition was met. This suggests that listeners process the intonational cues across the prosodic boundary rather than focusing on tonal intonation in preboundary and postboundary positions independently. It also indicates that the local intonation pattern (H#L) across the prosodic boundary is the most crucial information that the listener makes use of in detecting a possible word boundary, but when such a local condition is not met, listeners make further use of other available intonational information (e.g., the second element of the postboundary intonational sequence).

It should be noted that in the present study, we used target words that started with either a lenis stop or a nasal because these consonants (among many others, including vowels) are known to be characteristically associated with an AP-initial rising tone (#LH) (Jun, 1993, 2000). (Recall that the initial tone becomes #HH only when the initial segment is an aspirated or a tense consonant.) Although the tone-segment interaction is thought to be a phonological process (Jun, 1993), it is not inviolable: Other tonal variants do occur, though with low frequency (Jun, 2000; Jun and Fougeron; Kim, 2004). Still, any observed effects of AP-initial tones may not be seen as purely intonational, to the extent that the intonation effect cannot be separated from the effect of the tone-segment interaction. However, purely intonational effects do exist with the preboundary (AP-final) tone, which is not constrained by any segmental information. Taken together, the observed effects can be viewed "intonational" insofar as listeners make use of available intonational information in lexical segmentation, regardless of whether the information is purely intonational or attributable to a result of the complex tone-segment interaction.

In experiment 2, we explored how the intonational cues for AP would interact with the substantial final-lengthening cue that might arise with IP. The results showed that Korean listeners' lexical segmentation of postboundary words is robustly influenced by lengthening and AP-boundary marking intonational cues, but the effects are not entirely cumulative: Preboundary lengthening helps the detection of the following word across the boundary, but only when boundary-adjacent intonational cues are not frequent. When final lengthening was combined with the most frequent boundary-adjacent intonation pattern (H#LH), however, a poorer performance was observed. This suggests that an additional prosodic cue (in this case, phrase-final lengthening) does not always operate in favor of lexical segmentation. Then, a question arises: Why does the presence of final lengthening cue yield asymmetrical results, especially showing the unexpected pattern when intonational cues are frequent for AP boundary percept?

We propose that this is due to a mismatch between segmental and prosodic information in our stimuli. While processing the incoming speech signal, listeners are likely to be able to predict what comes next, based on what they have already heard. Preboundary (phrase-final) lengthening should give a phrase boundary percept to listeners, just as preboundary H# does. When an IP boundary is hypothesized due to substantial (IP-induced) final lengthening, the listener would expect that what comes after the assumed IP boundary would be another IP. If the forthcoming segmental materials across the boundary are perceived as containing phonetic information appropriate for that IP-initial position (e.g., IP-initial strengthening cues), the listener's initial boundary decision would eventually be confirmed.

As discussed in the Introduction, however, appropriate domain-initial strengthening cues for IP-initial position were not included in our speech materials to avoid any potential confounding effects from other possible prosodic cues. In particular, allophonic variation in lenis stops was not matched with IP: Lenis stops are always voiceless when in IP-initial position (a type of domain-initial strengthening) (Cho and Keating, 2001; Jun, 1993), but our speech materials contained the lenis stops with voicing as appropriate for IP-medial rather than IP-initial position. The substantial final lengthening cue which would give rise to an IP-final boundary percept is not matched with AP-induced segmental cues, lacking domain-initial strengthening cues (including allophonic voicing cues) for IP-initial position. Such a mismatch may therefore hinder the segmentation process. This idea can be further explained in terms of the prosody analyzer account of speech perception proposed by Cho *et al.* (2007) and Salverda *et al.* (2003). The prosody analyzer account assumes that suprasegmental and segmental information are processed in parallel. When the computed prosodic boundary based on suprasegmental information is matched with prosodically-driven segmental information for that boundary, lexical segmentation is facilitated, whereas a mismatch between them is predicted to hinder lexical segmentation.

This also has another implication for the use of multiple phrase-level prosodic cues in lexical segmentation process. Previous studies have suggested that available lexical segmentation cues are used immediately in an exhaustive and cumulative way (e.g., Spitzer *et al.*, 2007; Donselaar *et al.*, 2005; Norris *et al.*, 1997). However, our results suggest that not all the available segmentation cues are necessarily exploited by the listener in a straightforwardly cumulative way. Instead, in detecting prosodic boundaries, listeners appear to use available prosodic cues with differential weighting: When the local boundary-marking information is not sufficient to finalize the decision about the location of a prosodic boundary (as with L#L), additional information (such as preboundary lengthening and the second element of the post-boundary intonational sequence, as in #LH) is further utilized by the listener. The relative use of available phrase-level prosodic cues is reminiscent of the hierarchical model of speech segmentation proposed by Mattys *et al.* (2005). It assumes that available segmentation cues are weighted, such that higher order knowledge (e.g., lexical information and contextual information) takes precedence over sublexical information. The sublexical cues then become operative in a non-optimal communicative condition so that segmental cues are first called upon and word-internal metrical prosodic cues are used as a last resort. Although the model does not deal with how phrase-level prosodic cues are utilized in lexical segmentation, multiple phrase-level prosodic cues may well be used with relative weights. For example, the local boundary-marking intonational cues (H#L) may well be weighted above other boundary-adjacent prosodic cues.

We are now left with a question about how the use of phrase-level prosodic information in lexical segmentation may be incorporated into existing models of speech segmentation such as the hierarchical model (Mattys *et al.*, 2005), or the shortlist model (Norris, 1994; Norris and McQueen, 2008). While addressing this issue is beyond the scope of the present study, one possible way is to allow a module such as the prosody analyzer whose function is to compute prosodic boundaries using multiple prosodic cues, along with their relative weights. The prosodic structure computed in this way can then serve as part of the high order knowledge in the hierarchical model. Or, it can be checked against computed lexical boundaries in a model like the shortlist model, such that lexical competition is further modulated by the alignment between the computed prosodic structure and the potential lexical boundary.

Finally, the results of the present study have implications for the language-specific vs cross-linguistic use of perceptual salient rising F0 (or high pitch) in lexical segmentation. Given that an F0 rise is a potential prosodic cue in lexical segmentation in various languages, Warner *et al.* (2009) suggested that F0 rise is perceptually salient, and the presence of F0 rise word-initially is likely to facilitate lexical segmentation cross-linguistically. Our results, however, showed the opposite: Korean listeners found it difficult to detect the target word when it starts with H surrounded by L tones as in L#HL.

We propose that although substantial F0 rise may be universally perceptually salient, how such an intonational element is aligned with segmental content is determined by language-specific intonational phonology. The intonational element of H in Korean, for example, is aligned with either the second or the final syllable of AP (i.e., #LH⋯LH#), and thus H on the first syllable of the target word hinders lexical segmentation process. But in French, H is phonologically specified to be aligned with the first syllable of the content word in AP (Welby, 2007), and thus it facilitates lexical segmentation process in a way that conforms to French intonational phonology. Shukla *et al.* (2007) showed that Italian listeners were able to exploit Japanese IP boundaries in word segmentation and claimed that prosody contains universal cues for lexical segmentation. Again, what is "universal" here may be that some prosodic cues are used cross-linguistically: Italian listeners are likely to have exploited Japanese IP boundaries not because the particular prosodic cues that were available to them were universal, but because Italian and Japanese IP happen to share some prosodic cues. Our interpretations are in line with previous studies on word segmentation in general: The set of segmentation cues is language-universal, but the detailed manifestations of individual segmentation cues that listeners exploit are language-specific. It is well known that listeners use phonotactic, allophonic, and various other cues for word segmentation, but the exploitation of those cues is sensitive to the phonological constraints and structure of a given language (Cutler *et al.*, 2002; Cutler and Norris, 1988; Cutler and Otake, 1994; Mehler *et al.*, 1981; Sebastián-Gallés *et al.*, 1992; Weber, 2001). Likewise, the claim that prosodic boundaries constrain online lexical search (Christophe *et al.*, 2004; Shukla *et al.*, 2007) may be applicable cross-linguistically, but the way the prosodic boundary is phonetically manifested in speech production and the way that listeners exploit the prosodic boundary cues in speech comprehension must be language-specific.

V. CONCLUSION

The present study investigated the role of phrase-level prosodic cues in word segmentation of Korean. In experiment 1, we found that listeners make use of a local intonation pattern at a prosodic (AP) boundary (i.e., H#L) in online lexical segmentation, even when other important boundary-marking cues (such as domain-initial strengthening and final lengthening) are not present in the speech signal. The locality condition of H#L at the boundary suggests that listeners generally pay more attention to information straddling the prosodic boundary rather than the global intonational contour within a phrase, indicating that the boundary detection is crucial in lexical segmentation. In experiment 2, an additional final lengthening cue was found to help listeners with lexical segmentation when intonation patterns are not frequent for marking a prosodic boundary. However, when the lengthening cue was combined with the most frequent intonation pattern of H#L, creating a percept of IP, the mismatch between prosodic cues (appropriate for IP) and domain-initial strengthening cues (including allophonic cues of lenis stops, appropriate for IP-medial position) appears to take precedence over the cumulative effect. The hypothesized cumulative effect would work only if the computed prosodic boundary is matched with expected segmental allophonic variation. A follow-up study is necessary to corroborate this claim, in order to examine the interaction between effects of prosodic variation and allophonic variation in lexical segmentation. It also remains to be seen how the use of phrase-level prosodic information in spoken word recognition may be implemented within current models of speech segmentation. But following the prosody analyzer account, we propose that prosodic information is computed in parallel with segmental information, and lexical segmentation is modulated by the interaction between the two kinds of information. More generally, our study builds up on the growing body of psycholinguistic research which highlights the important roles that prosody plays in both speech production and speech comprehension: Speakers generate a prosodic structure online in which a given utterance is organized into prosodic units, and its expected acoustic-phonetic cues are in turn exploited by listeners in online lexical segmentation.

ACKNOWLEDGMENTS

This work was supported by the Korea Research Foundation Grant funded by the Korean Government (MOEHRD) (Contract No. KRF-2007-332-A00102). We thank the reviewers and the associate editor for their constructive comments; Adam Albright for his suggestions; and Eunbin Lee, Jiyoung Park, Soomin Yoo, and Yuna Hur for their assistance with data collection.

APPENDIX: LIST OF TARGET WORDS AND CARRIER STRINGS

The following shows the list of disyllabic and trisyllabic target words and carrier strings.

a. Disyllabic Target Words

Target Words (Phonemic Transcription)	Glosses	Carrier Strings (Phonemic Transcription)
/ka.wi/	scissors	/ma.pɛ.tʃo.**ka.wi**.nɛ.li/
/kɛ.mi/	ant	/tʃʌ.ku.ta.**kɛ.mi**.sʌ.tu/
/kʌ.li/	street	/na.so.lɛ.**kʌ.li**. tʃʌ.p^{h}u/
/ko.kɛ/	hill	/ta.sʌ.ni.**ko.kɛ**.tʃu.pi/
/ko.ki/	meat	/tu.tʃɛ.ma.**ko.ki**.na.po/
/ku.tu/	shoes	/po.na.tʃa.**ku.tu**.mi.pɛ/
/ki.to/	prayer	/tʌ.so.nʌ.**ki.to**.ju.ma/
/na.ra/	nation	/nu.jʌ.ko.**na.la**.kɛ.tʃi/
/na.mu/	tree	/tʃo.la.su.**na.mu**.kʌ.to/
/no.lɛ/	song	/pɛ.mi.tʃʌ.**no.lɛ**.pa.ki/
/ta.li/	leg/bridge	/ku.tʃʌ.mo.**ta.li**.pɛ.la/
/to.si/	city	/tsi.ma.pɛ.**to.si**.mʌ.tsɛ/
/tu.pu/	tofu	/tʃo.nu.sɛ.**tu.pu**.nɛ.ma/
/ma.lu/	floor	/tʃu.ka.pʌ.**ma.lu**.ki.ta/
/mʌ.li/	head	/tɛ.tʃu.pi.**mʌ.li**.tʃo.hɛ/
/mo.ki/	mosquito	/ta.mi.tʃʌ.**mo.ki**.pɛ.ki/
/mo.tʃa/	hat	/tʃʌ.pɛ.**mo.tʃa**.so.tu/
/mu.kɛ/	weight	/pa.mʌ.lo.**mu.kɛ**.nʌ.sɛ/
/mi.lɛ/	future	/kjo.tʃʌ.tu.**mi.lɛ**.po.ku/
/mi.so/	smile	/tʃa.pɛ.lo.**mi.so**.nʌ.ku/
/pa.ta/	sea	/pu.sɛ.tʃi.**pa.ta**.mʌ.tʃu/
/pa.tʃi/	pants	/ma.jʌ.ku.**pa.tʃi**.ni.mɛ/
/pu.tʃa/	a rich person	/ti.lɛ.mi.**pu.tʃa**.lʌ.ko/
/pi.nu/	soap	/pa.tɛ.mʌ.**pi.nu**.mɛ.tʃʌ/

b. Trisyllabic Target Words

Target Words (Phonemic Transcription)	Glosses	Carrier Strings (Phonemic Transcription)
/ko.ku.ma/	sweet potato	/no.tʃu.pʌ.**ko.ku.ma**.li.tʃa/
/ku.tʌ.ki/	maggot	/ko.ta.po.**ku.tʌ.ki**.pa.tɛ/
/ki.lʌ.ki/	wild goose	/mu.tʃa.nʌ.**ki.lʌ.ki**.no.kɛ/
/na.nu.ki/	division	/ta.lu.tʃo.**na.nu.ki**.mʌ.no/
/na.tɨ.li/	outing	/pi.ma.ko.**na.tɨ.li**.ku.tʃʌ/
/na.mʌ.tʃi/	remainder	/mɛ.pi.tʃu.**na.mʌ.tʃi**.mi.sʌ/
/no.ta.tʃi/	a gold mine	/tʃa.pʌ.ku.**no.ta.tʃi**.ma.la/
/nu.tʌ.ki/	rag	/mi.ta.po.**nu.tʌ.ki**.tʃu.mi/
/tʌ.tɨ.mi/	antenna	/mo.na.tʃu.**tʌ.tɨ.mi**.pʌ.sɛ/
/to.ka.ni/	pot	/ma.la.pe.**to.ka.ni**.mʌ.ha/
/to.kɛ.pi/	elf	/nu.tʃɛ.**to.k*ɛ.pi**.ma.sʌ/
/to.ka.tʃi/	bellflower	/kʌ.pi.tʃo.**to.la.tʃi**.pʌ.ta/
/to.t^{h}o.li/	acorn	/tʃa.mi.kʌ.**to.t^{h}o.li**.pa.tɛ/
/tu.k*ʌ.pi/	toad	/mo.la.tʃɛ.**tu.k*ʌ.pi**.ta.tʃo/
/tu.tʌ.tʃi/	mole	/ni.thɛ.mʌ.**tu.tʌ.tʃi**.la.kɛ/
/tu.lu.mi/	crane	/ka.tʌ.pi.**tu.lu.mi**.po.tʃu/
/ma.nu.la/	wife	/pɛ.li.no.**ma.nu.la**.tʃi.pʌ/
/mɛ.t*u.ki/	grasshopper	/tʃɛ.ni.ka.**mɛ.t*u.ki**.tu.pa/

b. Trisyllabic Target Words (cont.)

Target Words (Phonemic Transcription)	Glosses	Carrier Strings (Phonemic Transcription)
/mjʌ.nɨ.li/	daughter-in-law	/ti.kɛ.no.**mjʌ.nɨ.li**.pa.tʃu/
/mu.tʌ.ki/	pile	/nu.mi.pa.**mu.tʌ.ki**.sʌ.nɛ/
/pa.ku.ni/	basket	/pɛ.ta.tʃo.**pa.ku.ni**.mʌ.la/
/po.t*a.li/	bundle	/ku.mo.sʌ.**po.t*a.li**.tɛ.mjʌ/
/tʃɛ.tʃ^{h}ɛ.ki/	sneeze	/tu.pa.mʌ.**tʃ.tʃ^{h}ɛ.ki**.pa.tʃu/
/tʃu.mʌ.ni/	pocket	/tʃo.ti.**tʃu.mʌ.ni**.mo.li/

[1]Intervocalic voicing of lenis stops is generally held to apply within the AP; however, even within the AP, it does not apply in every single instance. A reviewer suggested that this might weaken the argument that AP is a phonologically-motivated categorical prosodic unit. However, the phonological nature of AP cannot be determined simply by the observed gradient nature of voicing within an AP. This is because a process may be gradient or probabilistic, even if the necessary pre-conditions for application (i.e., internal to the AP) are categorically present (see Zsiga, 1995; Cohn, 1998; Fourakis and Port, 1986 for discussion on phonetic vs phonological processes). Jun (1995) indeed noted this and suggested that the lenis stop intervocalic voicing rule may be seen as a gradient phonetic process. More importantly, however, whether AP is phonetic or phonological is beyond the scope of the present study. What is critical in the present study is that the most extensively received model (Jun, 1993, 2000) is adopted here as a framework within which effects of intonation and durational cues on lexical segmentation can be tested: Even if the model dividing Korean phrases into a small number of discrete categories such as AP and IP were not theoretically impregnable, the results of our experiments could easily be interpreted in a model that makes use of a more flexible prosodic model since our study actually measures the effects of acoustic properties (lengthening and intonation), not of phrase structures like "AP" or "IP" directly.

[2]The results of splicing were checked by the two authors to ensure that no discernible splicing artifact remained in the speech signal. Nevertheless, as pointed out by a reviewer, it is possible that subtle discontinuities may remain and serve as a segmentation cue (Johnson and Jusczyk, 2001; Mattys, 2004). However, because the same spliced token was used in all test conditions, any observed differences between prosodic conditions (e.g., frequent vs infrequent intonational conditions) must be attributed to the prosodic manipulation, not to splicing artifacts.

Beckman, M. E. (**1996**). "The parsing of prosody," Lang. Cognit. Processes **11**, 17–67.

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

Cho, T., and Keating, P. (**2001**). "Articulatory and acoustic studies of domain-initial strengthening in Korean," J. Phonetics **29**, 155–190.

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

Christophe, A., Peperkamp, S., Pallier, C., Block, E., and Mehler, J. (**2004**). "Phonological phrase boundaries constrain lexical access: I. Adult data," J. Mem. Lang. **51**, 523–547.

Chung, K., Chang, S., Choi, J., Nam, S., Lee, M., Chung, S., Koo, H., Kim, K., Kim, J., Lee, C., Han, S., Oh, M., Song, M., Hong, S., and Jee, S. (**1996**). *A Study of Korean Prosody and Discourse for the Development of Speech Synthesis/Recognition System* (KAIST Artificial Intelligence Research Center, Daejon, Korea).

Cohn, A. (**1998**). "The phonetics-phonology interface revisited: Where's phonetics?," in Texas Linguistics Society 1998 Conference Proceedings, pp. 25–40.

Cooper, N., Cutler, A., and Wales, R. (**2002**). "Constraints of lexical stress on lexical access in English: Evidence from native and non-native listeners," Lang Speech **45**, 208–228.

Cutler, A., and Butterfield, S. (**1992**). "Rhythmic cues to speech segmentation: Evidence from juncture misperception," J. Mem. Lang. **31**, 218–236.

Cutler, A., and Carter, D. M. (**1987**). "The predominance of strong initial syllables in the English vocabulary," Comput. Speech Lang. **2**, 133–142.

Cutler, A., Demuth, K., and McQueen, J. M. (**2002**). "Universality versus language specificity in listening to running speech," Psychol. Sci. **13**, 258–262.

Cutler, A., and Norris, D. (**1988**). "The role of strong syllables in segmentation for lexical access," J. Exp. Psychol. Hum. Percept. Perform. **14**, 113–121.

Cutler, A., and Otake, T. (**1994**). "Mora or phoneme? Further evidence for language-specific listening," J. Mem. Lang. **33**, 824–844.

Dahan, D., Magnuson, J. S., Tanenhaus, M. K., and Hogan, E. M. (**2001**). "Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition," Lang. Cognit. Processes **16**, 507–534.

Donselaar, W. V., Koster, M., and Cutler, A. (**2005**). "Exploring the role of lexical stress in lexical recognition," Q. J. Exp. Psychol. **58**, 251–274.

Edwards, J., Beckman, M. E., and Fletcher, J. (**1991**). "Articulatory kinematics of final lengthening," J. Acoust. Soc. Am. **89**, 369–382.

Forster, K., and Chambers, S. (**1973**). "Lexical access and naming time," J. Verbal Learn. Verbal Behav. **12**, 627–635.

Fougeron, C., and Keating, P. A. (**1997**). "Articulatory strengthening at edges of prosodic domains," J. Acoust. Soc. Am. **101**, 3728–3740.

Fourakis, M., and Port, R. (**1986**). "Stop epenthesis in English," J. Phonetics **14**, 197–221.

Gaskell, M. G., and Marslen-Wilson, W. D. (**1997**). "Integrating form and meaning: A distributed model of speech perception," Lang. Cognit. Processes **12**, 613–656.

Gee, J. P., and Grosjean, F. (**1983**). "Performance structures: A psycholinguistic and linguistic appraisal," Cogn. Psychol. **15**, 411–458.

Gow, D. W. (**2002**). "Does English coronal assimilation create lexical ambiguity?," J. Exp. Psychol. Hum. Percept. Perform. **28**, 163–179.

Gow, D. W., and Gordon, P. (**1995**). "Lexical and prelexical influences on word segmentation: Evidence from priming," J. Exp. Psychol. Hum. Percept. Perform. **21**, 344–359.

Johnson, E. K., and Jusczyk, P. W. (**2001**). "Word segmentation by 8-month-olds: When speech cues count more than statistics," J. Mem. Lang. **44**, 548–567.

Jun, S.-A. (**1993**). "The phonetics and phonology of Korean prosody," Ph.D. dissertation, The Ohio State University, Columbus, OH.

Jun, S.-A. (**1995**). "Asymmetrical prosodic effects on the laryngeal gesture in Korean," in *Papers in Laboratory Phonology IV: Phonology and Phonetic Evidence*, edited by B. Connell and A. Arvaniti (Cambridge University Press, Cambridge), pp. 235–253.

Jun, S.-A. (**1998**). "The accentual phrase in the Korean prosodic hierarchy," Phonology **15**, 189–226.

Jun, S.-A. (**2000**). "K-ToBI (Korean ToBI) labelling conventions (Version 3.1)," http://www.linguistics.ucla.edu/people/jun/ktobi/K-tobi.html (Last accessed 16 March 2009).

Jun, S.-A., and Fougeron, C. (**2000**). "A phonological model of French intonation," in *Intonation: Analysis, Modelling and Technology*, edited by A. Botinis (Kluwer Academic, Boston, MA), pp. 209–242.

Keating, P., Cho, T., Fougeron, C., and Hsu, C. (**2003**). "Domain-initial strengthening in four languages," in *Papers in Laboratory Phonology VI*, edited by J. Local, R. Ogden, and R. Temple (Cambridge University Press, Cambridge), pp. 145–163.

Keating, P., and Shattuck-Hufnagel, S. (**2002**). "A prosodic view of word form encoding for speech production," UCLA Working Papers in Phonetics **101**, 112–156, University of California, Los Angeles, CA.

Kim, S.-J. (**2000**). "Accentual effect on segmental phonological rules in Korean," Ph.D. dissertation, University of North Carolina, Chapel Hill, NC.

Kim, S. (**2004**). "The role of prosodic phrasing in Korean word segmentation," Ph.D. dissertation, University of California, Los Angeles, CA.

Kim, H.-G., and Kang, B.-M. (**2004**). *Frequency Analysis of Korean Morpheme and Word Usage II* (Institute of Korean Culture, Korea University, Seoul, Korea).

Krivokapić, J. (**2007**). "Prosodic planning: Effects of phrasal length and complexity on pause duration," J. Phonetics **35**, 162–179.

Ladd, D. R. (**1996**). *Intonational Phonology* (Cambridge University Press, Cambridge).

Lehiste, I. (**1960**). "An acoustic-phonetic study of internal open juncture," Phonetica **5**, 1–54.

Lehiste, I. (**1970**). *Suprasegmentals* (MIT Press, Cambridge, MA).

Lim, B. J., and de Jong, K. J. (**1999**). "Tonal alignment in Seoul Korean," J. Acoust. Soc. Am. **106**(4), 2152.

Luce, P. A. (**1986**). "Neighborhoods of words in the mental lexicon," Re-

search on Speech Perception Technical Report No. 6, National Institutes of Health, Bloomington, IN.
Luce, P. A., and Pisoni, D. B. (**1998**). “Recognizing spoken words: The neighborhood activation model,” Ear Hear. **19**, 1–36.
Marslen-Wilson, W., and Welsh, A. (**1978**). “Processing interactions during word-recognition in continuous speech,” Cogn. Psychol. **10**, 29–63.
Mattys, S. L. (**2004**). “Stress versus coarticulation: Toward an integrated approach to explicit speech segmentation,” J. Exp. Psychol. Hum. Percept. Perform. **30**, 397–408.
Mattys, S. L., White, L., and Melhorn, J. F. (**2005**). “Integration of multiple speech segmentation cues: A hierarchical framework,” J. Exp. Psychol. Gen. **134**, 477–500.
McClelland, J. L., and Elman, J. L. (**1986**). “The TRACE model of speech perception,” Cogn. Psychol. **18**, 1–86.
McQueen, J. M. (**1996**). “Word spotting,” Lang. Cognit. Processes **11**, 695–699.
McQueen, J. M. (**2005**). “Speech perception,” in *The Handbook of Cognition*, edited by K. Lamberts and R. Goldstone (Sage, London), pp. 255–275.
Mehler, J., Dommergues, J.-Y., Frauenfelder, U. H., and Segui, J. (**1981**). “The syllable’s role in speech segmentation,” J. Verbal Learn. Verbal Behav. **20**, 298–305.
Nespor, M., and Vogel, I. (**1986**). *Prosodic Phonology* (Foris, Dordrecht).
Norris, D. (**1986**). “Word recognition: Context effects without priming,” Cognition **22**, 93–136.
Norris, D. (**1994**). “Shortlist: A connectionist model of continuous speech recognition,” Cognition **52**, 189–234.
Norris, D., and McQueen, J. M. (**2008**). “Shortlist B: A Bayesian model of continuous speech recognition,” Psychol. Rev. **115**, 357–395.
Norris, D., McQueen, J. M., Cutler, A., and Butterfield, S. (**1997**). “The possible-word constraint in the segmentation of continuous speech,” Cogn. Psychol. **34**, 191–243.
Park, K. C. (**2004**). “The phrase-initial high in Korean,” Studies in Phonetics, Phonology and Morphology **10**, 203–223.
Quené, H. (**1993**). “Segment durations and accent as cues to word segmentation in Dutch,” J. Acoust. Soc. Am. **94**, 2027–2035.
Saffran, J. R., Newport, E. L., and Aslin, R. N. (**1996**). “Word segmentation: The role of distributional cues,” J. Mem. Lang. **35**, 606–621.
Salverda, A. P., Dahan, D., and McQueen, J. M. (**2003**). “The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension,” Cognition **90**, 51–89.
Schreuder, R., and Baayen, R. H. (**1994**). “Prefix stripping re-revisited,” J. Mem. Lang. **33**, 357–375.
Sebastián-Gallés, N., Dupoux, E., Segui, J., and Mehler, J. (**1992**). “Contrasting syllabic effects in Catalan and Spanish,” J. Mem. Lang. **31**, 18–32.
Selkirk, E. (**1984**). *Phonology and Syntax: The Relation Between Sound and Structure* (MIT Press, Cambridge).
Shattuck-Hufnagel, S., and Turk, A. (**1996**). “A prosody tutorial for investigators of auditory sentence processing,” J. Psycholinguist. Res. **25**, 193–247.
Shatzman, K. B., and McQueen, J. M. (**2006**). “Segment duration as a cue to word boundaries in spoken-word recognition,” Percept. Psychophys. **68**, 1–16.
Shukla, M., Nespor, M., and Mehler, J. (**2007**). “An interaction between prosody and statistics in the segmentation of fluent speech,” Cogn. Psychol. **54**, 1–32.
Soto, S., Sebastián-Gallés, N., and Cutler, A. (**2001**). “Segmental and suprasegmental mismatch in lexical access,” J. Mem. Lang. **45**, 412–432.
Spitzer, S. M., Liss, J. M., and Mattys, S. L. (**2007**). “Acoustic cues to lexical segmentation: A study of resynthesized speech,” J. Acoust. Soc. Am. **122**, 3678–3687.
Tabossi, P., Collina, S., Mazzetti, M., and Zoppello, M. (**2000**). “Syllables in the processing of spoken Italian,” J. Exp. Psychol. Hum. Percept. Perform. **25**, 758–775.
Vitevitch, M. S., Luce, P. A., Charles-Luce, J., and Kemmerer, D. (**1997**). “Phonotactic and syllable stress: Implications for the processing of spoken nonsense words,” Lang Speech **40**, 47–62.
Vroomen, J., and de Gelder, B. (**1995**). “Metrical segmentation and lexical inhibition in spoken word recognition,” J. Exp. Psychol. Hum. Percept. Perform. **23**, 710–720.
Warner, N., Otake, T., and Takayuki, A. (**2009**). “Intonational structure as a word boundary cue in Japanese,” Lang Speech in press.
Weber, A. (**2001**). “Language-specific listening: The case of phonetic sequences,” Ph.D. dissertation, Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands.
Welby, P. (**2007**). “The role of early fundamental frequency rises and elbows in French word segmentation,” Speech Commun. **49**, 28–48.
Wightman, C., Shattuck-Hufnagel, S., Ostendorf, M., and Price, P. (**1992**). “Segmental durations in the vicinity of prosodic phrase boundaries,” J. Acoust. Soc. Am. **91**, 1707–1717.
Zsiga, E. C. (**1995**). “An acoustic and electropalatographic study of lexical and post-lexical palatalization in American English,” in *Phonology and Phonetic Evidence: Papers in Laboratory Phonology IV*, edited by B. Connell and A. Arvaniti (Cambridge University Press, Cambridge), pp. 282–302.