Article

Effects of rhythm and phrase-final lengthening on word-spotting in Korean

Jeon, Hae-Sung and Arvaniti, Amalia

Jeon, Hae-Sung ORCID: 0000-0001-7536-5571 and Arvaniti, Amalia (2017) Effects of rhythm and phrase-final lengthening on word-spotting in Korean. The Journal of the Acoustical Society of America, 141 (6). pp. 4251-4263. ISSN 0001-4966

It is advisable to refer to the publisher's version if you intend to cite from the work.
http://dx.doi.org/10.1121/1.4983178

The Journal of the Acoustical Society of America 141 (6), 4251-4263, accepted on 26th April 2017, published on-line 8 June;
doi: http://dx.doi.org/10.1121/1.4983178

Title(s): Effects of rhythm and phrase-final lengthening on word-spotting in Korean

Running title: Prosody effects on Korean word-spotting

The names and affiliations of the authors:

Hae-Sung Jeon

HJeon1@uclan.ac.uk

School of Language and Global Studies

Faculty of Culture and the Creative Industries

University of Central Lancashire

Preston PR1 2HE, UK

+44 (0) 1772 89 3143

Amalia Arvaniti

a.arvaniti@kent.ac.uk

English Language and Linguistics, SECL

University of Kent

Canterbury CT2 7NF, UK

+44 (0) 1227 827734

The Journal of the Acoustical Society of America 141 (6), 4251-4263, accepted on 26th April 2017, published on-line 8 June;
doi: http://dx.doi.org/10.1121/1.4983178

Abstract

A word-spotting experiment was conducted to investigate whether rhythmic consistency and phrase-final lengthening facilitate performance in Korean. Listeners had to spot disyllabic and trisyllabic words in nonsense strings organized in phrases with either the same or variable syllable count; phrase-final lengthening was absent, or occurring either in all phrases or only in the phrase immediately preceding the target. The results show that, for disyllabic targets, inconsistent syllable count and lengthening before the target led to fewer errors. For trisyllabic targets, accuracy was at ceiling, but final lengthening in all phrases reduced reaction times. The results imply that both rhythmic consistency (i.e. regular syllable count) and phrase-final lengthening play a role in word-spotting and, by extension, in speech processing in Korean, as in other languages. However, the results also reflect the language specific role of prosodic cues. First, the cues here were used primarily with disyllabic targets, which were cognitively more demanding to process partly due to their high phonological neighborhood density. Second, the facilitating effect of rhythmic consistency was weak, possibly because strict consistency is not present in spoken Korean. Overall, rhythmic consistency facilitated spotting when targets mapped onto phrases, confirming the importance of phrasal organization in Korean speech processing.

The Journal of the Acoustical Society of America 141 (6), 4251-4263, accepted on 26th April 2017, published on-line 8 June;
doi: http://dx.doi.org/10.1121/1.4983178

I. INTRODUCTION

In order to comprehend speech, listeners need to segment spoken utterances into words. A general consensus from psycholinguistic studies is that word segmentation is an active process of continuous hypothesis testing in search of the word in the mental lexicon that best matches the incoming acoustic information (see Cutler, 2012, Chap.1 for an overview). As clear acoustic cues to word boundaries are not always present in speech, listeners are aided by other cues, such as semantic context, phonotactics, coarticulatory effects, and so on (Davis *et al*., 2002; Mattys *et al*., 2005).

A growing body of research demonstrates the important role of speech prosody in the task of word segmentation; for instance, several studies show that lexical access occurs within the phonological phrase, a linguistic unit typically larger than the word and demarcated by acoustic disjuncture (e.g., Christophe *et al*. 2004 on French; see Shattuck-Hufnagel & Turk, 1996 for a review on prosodic organization). In addition, research shows that *rhythmic consistency* (e.g., the regular grouping of syllables into prosodically defined groups of equal length) can aid word segmentation by building up listeners' expectations on when future word boundaries will occur in speech (e.g., Dilley & McAuley, 2008, Morrill *et al*. 2014 in English). The overall role of prosody has been reported for a variety of languages, such as French (Christophe *et al*. 2004), Dutch (Salverda *et al*. 2003), and Italian (Shukla *et al*. 2007), but studies have also shown that the effects of prosody vary by language (see e.g., Welby, 2007, and Spinelli *et al*. 2010 on French, and Warner *et al*. 2010 on Japanese).

The central question of the present study was whether rhythmic consistency (i.e., the regular recurrence of prosodic events) and/or phrase-final lengthening would aid listeners during word-spotting in Seoul Korean (henceforth Korean). Previous work on English provides evidence that rhythmic consistency at the utterance-level (derived from regular foot structure) shapes listeners' expectations about the location of upcoming word boundaries (Dilley & McAuley, 2008; Brown *et al.*, 2011; Morrill *et al.*, 2014). Korean is of interest in this respect because it does not have lexical stress like English, or other word-level phenomena that would help listeners build up metrical expectations. The smallest prosodically demarcated linguistic unit in Korean is the *Accentual Phrase*, which is typically larger than a word (see Section I. B, and Jeon, 2015a for a review). Therefore, it was expected that both rhythmic consistency at the phrasal level and local effects indicating phrasing would aid word-spotting.

To test this hypothesis, we carried out a word-spotting experiment in which the listeners' task was to recognize a real word embedded in a nonsense string of syllables prosodically organized in ways that tapped into both global rhythmic organization and local phrasing cues (see McQueen, 1996 and Cutler, 2012, sect. 3.3 for an overview of the word-spotting paradigm). The study examines the role of recurrent prosodic events (i.e., of consistent rhythmic structure), a factor that was not investigated in previous research on Korean involving the word-spotting paradigm (e.g. Kim, 2004; Kim & Cho, 2009). It also differs from the work of Dilley and colleagues (Dilley & McAuley, 2008; Brown *et al.*, 2011; Morrill *et al.*, 2014, inter alia) who do examine the role of rhythmic structure in word segmentation, but do so using strings of syllables that can be *grouped*

in a number ways to create words (see I.A for details); this optionality was not available to our participants who could spot only one word in each string.

A. Prosodic cues to word boundaries

The view adopted here is that prosodic cues are broadly classified into two categories: global rhythmic cues and local cues to prosodic boundaries (Fletcher, 2010; Nolan & Jeon, 2014; White, 2014). Listeners can exploit both types as discussed below.

Previous work demonstrates that the regular recurrence of prosodic events creates a rhythmic structure which acts as a global cue allowing listeners to anticipate the timing of upcoming word onsets (Barnes & Jones, 2000; Jones *et al.,* 2006; inter alia; cf. Arvaniti, 2009). For example, Dilley and McAuley (2008), Dilley *et al*. (2010), Brown *et al*. (2011), and Morrill *et al*. (2014) show that in English repeating patterns of pitch or duration influence processing, in that listeners can use them to anticipate foot structure in the upcoming string and group syllables into words accordingly. For instance, in Dilley & McAuley (2008), the experimental stimuli were 8-syllable sequences of English that could be grouped into different words; the listeners' task was to determine how the last four syllables were grouped, e.g., in *channel dizzy foot note book worm,* the last four syllables could be grouped as either *footnote bookworm* or as *foot notebook worm*. The acoustic characteristics of the last three syllables were kept constant, but either the pitch or the duration of the preceding syllables was varied to be, for example, low-high-low-... or long-short-long, etc. Dilley & McAuley (2008) showed that the grouping of the last four syllables was affected by these patterns. For instance, when the first five syllables had a low-high-low-high-low pitch pattern and the last three syllables had high-low-high, native English speakers grouped the last two syllables

together, i.e., *footnote bookworm*, since the entire string was thus evenly divided into four rhythmically consistent groups, [low-high]-[low-high]-[low-high]-[low-high]. On the other hand, when the first five syllables had a high-low-high-low-high pattern followed again by high-low-high, listeners grouped the last four syllables as *foot notebook worm*, i.e., [high-low]-[high-low]-[high]-[high-low]-[high], in order to preserve the established [high-low] grouping as much as possible.

In contrast to the global patterns which help listeners anticipate events, local prosodic cues at linguistic unit boundaries help listeners interpret the organisation of the incoming signal by directing attention to changes in particular parameters (White, 2014). Local cues such as pitch movements and lengthening at prosodic boundaries are observed across languages (see Fletcher, 2010, and Jun, 2014, for reviews). For example, intonational phrases are cross-linguistically associated with acoustic disjuncture, which can be manifested by means of a pause, the presence of specific boundary tones, or significant phrase-final lengthening (Shattuck-Hufnagel & Turk, 1996). Listeners exploit these local cues to identify the boundaries of prosodic units; for instance, Welby (2007) and Spinelli *et al*. (2010) found that French speakers use a local pitch rise as a cue to the boundary between content and function words. Local cues are exploited by listeners in artificial language learning tasks as well: Tyler & Cutler (2009) have shown that English, French and Dutch speakers use lengthening as a cue to phrase-finality and this helps them detect upcoming word onsets. Kim *et al*. (2012) report similar results with speakers of Dutch and Korean.

In summary, both global and local prosodic cues can affect word segmentation. While local effects at phrase boundaries appear to be robust and have been reported

for a variety of tasks and languages, the global effect of rhythm has only been shown for English, which has word-internal metrical structure and relatively regular rhythm both at the lexical and postlexical levels (Hayes, 1995; Tilsen & Arvaniti, 2013). The way global cues operate, however, may be language-specific, and little is known about how they operate in languages without a prominence structure at the lexical level. This is the focus on the present paper.

B. Prosodic cues to utterance organisation in Seoul Korean

As noted earlier, Korean does not have stress or other word-level phonological properties akin to stress. Its prosody is organized at the phrasal level: utterances are grouped into *Intonational Phrases* (IPs), which are in turn exhaustively parsed into *Accentual Phrases* (APs) (Jun, 1998; 2000; 2005). The AP typically consists of a content word followed by one or more particles (although the particles tend to be omitted in colloquial speech). On average, the AP has 3–4 syllables and 1.14–1.2 content words (Kim, 2004). The AP is primarily demarcated by its pitch contour which is phonologically represented as (T)HLH (Jun, 1998; 2000; 2005). The initial tone T is determined by the laryngeal feature of the AP-initial segment; when it is an aspirated or fortis consonant (e.g., /pʰ/, /p*/), or a voiceless fricative (e.g., /s/, /s*/) then the initial tone tends to be H; otherwise it is L. All four tones in the phonological representation are realised when the AP has four or more syllables; shorter APs have more variable pitch contours. Jun (2000) reports 14 surface tonal patterns for the AP, including LH, LHH, HH, and HL. Some of these patterns are more frequent than others; e.g., in Kim's (2004) radio corpus, 80% of disyllabic and trisyllabic APs in which the initial segment induced L had the LH pattern. To our knowledge, this tonal variability is unconnected to

the morphological structure of the AP, such as the presence or location of particles relative to content words (unlike in French, discussed in Section I.A).

Due to the frequent occurrence of LH in the AP, Jun (2014) classifies Korean as a language with strong tonal *macro-rhythm*, i.e., a language which shows regularity in pitch contours at the phrasal level (in Korean, the AP). Additional regularities in the realization of the AP have also been reported: Cho & Keating (2001) have shown that the AP left edge is associated with articulatory strengthening; Tilsen & Arvaniti (2013) have found that despite the lack of word-level stress, Korean exhibits supra-syllabic amplitude alternations comparable in frequency to those of stress feet in English; given that the typical AP is 3-4 syllables long, as noted above, these alternations are likely to map onto the AP level. Taken together these findings suggest that rhythmic consistency at the AP level may be a useful cue for listeners in speech processing. On the other hand, reports on the presence of phrase-final lengthening associated with the AP are inconsistent: Jun (1995) found no lengthening, but Cho & Keating (2001) and Oh (1998) report significant AP-final lengthening.

As for the IP, it is right-demarcated by a monotonal or multitonal boundary tone (e.g. L%, H%, LH%, LHL%) signalling pragmatic meaning. In addition, right IP boundaries show substantial and consistent final lengthening (e.g., Jun, 1998; 2005): IP-final syllables are 1.6-1.8 times longer by comparison to IP-medial syllables (see Jeon, 2015a for a review).

In terms of speech processing, the above-mentioned cues to Korean prosodic structure have the following consequences. Overall, consistent and temporally

predictable markers of AP boundaries should be useful to listeners in detecting word onsets since, as noted above, content words tend to be aligned with the left edge of their AP and the AP may serve as the basis for rhythm in Korean. The findings of earlier studies also suggest that listeners are likely to rely more on tonal than durational cues in detecting AP boundaries, probably because AP-final lengthening is not consistently present (Jeon & Nolan, 2010). At the IP level, on the other hand, production data suggest that both tonal and durational cues are reliably available to listeners and both should help with word-spotting: these cues signal the end of an IP, and therefore the onset of a new IP and a new AP and, by extension, of a new word as well. These predictions were generally supported by the word-spotting experiments of Kim (2004) and Kim & Cho (2009). Kim (2004) found that performance in word-spotting improved when the target word was aligned with the left edge of an AP marked by a local tonal change (H#L, where # denotes the AP boundary), compared to when the target word was AP-medial. Kim & Cho (2009) showed that this finding applies regardless of the tonal structure in the middle of the AP the target word is part of. IP-level final lengthening, on the other hand, had a facilitating effect only when the IP-level tonal cues were unclear, e.g., when the tones across an IP-level boundary were H#HL, L#LH, or H#HL (Kim & Cho 2009).

In addition, in both Kim (2004) and Kim & Cho (2009), word-spotting was more accurate with trisyllabic than disyllabic targets. This could be attributed to the fact that the two sets of targets differed in phonological neighborhood density (PND; Luce & Pisoni, 1998). One reason for this difference was that trisyllabic words with a nested disyllabic word had to be excluded (e.g., /mosʌli/ ‘edge’ which includes /sʌli/ ‘frost); as a

result of this restriction common to all word-spotting experiments, the trisyllabic targets in Kim (2004) and Kim & Cho (2009) had low PND. In Kim (2004), responses to trisyllabic targets aligned with the left edge of the AP were at ceiling. On the other hand, in Kim & Cho (2009), disyllabic and trisyllabic targets were affected by the prosodic context in a similar fashion. The reason for the discrepancy is not clear, but it is possible that the ceiling effect in Kim (2004) was due to a clear local pitch cue (H#L) which was always present prior to the target, while in Kim & Cho (2009) the AP-final tone preceding the target varied (L# or H#).

In summary, the existing studies indicate that local pitch changes at prosodic boundaries (H#L) are crucial for speech processing in Korean, while phrase-final lengthening has a facilitating effect only when the pitch cue is insufficient (Kim & Cho, 2009). While these local cues have been amply investigated, there has been little research on the role of global rhythmic properties in Korean. The central question that motivated the experiment presented here is whether, on the one hand, phrase-final lengthening and anticipation established by rhythmic consistency (strong *macro-rhythm*) would facilitate word-spotting.

C. Hypotheses

It was hypothesised that accuracy and reaction times in word-spotting would improve with both the globally regular phrase structure and with phrase-final lengthening, a local cue. Specifically, it was predicted that consistency in the syllable count of consecutive prosodic phrases, either APs or IPs (see Section II. A. 3), combined with a repeating tonal pattern, would enable listeners to anticipate the

upcoming prosodic structure and thus the timing of word onsets. Further, it was hypothesised that phrase-final lengthening occurring in all phrases in the string would both signal phrase-finality and allow more processing time, thereby improving performance in comparison to lack of lengthening or its presence only immediately prior to the target word. Finally, the number of syllables per target was expected to affect the results in a similar manner to Kim (2004) and Kim & Cho (2009); namely we expected to find that trisyllabic targets would be spotted more accurately and faster than disyllabic ones.

II. EXPERIMENT

Listeners heard nonsense strings of syllables in which a real Korean noun was embedded; they had to press a button as soon as they heard the word and say it aloud. Listeners did not know beforehand what word would appear; therefore, their task was to recognise and segment the word from the nonsense string (see McQueen, 1996 and Cutler, 2012, sect. 3.3 for an overview of the word-spotting paradigm). The strings were organized into prosodic phrases all demarcated by F0 in the same way (see Section II. A. 3); the phrases would be interpreted as either APs or IPs depending on the presence of final lengthening: when final lengthening was present, the resulting right boundary would be perceived as an IP boundary; otherwise it would be perceived as an AP-boundary (see Section II. A. 3). In the remainder of the paper, we distinguish between APs and IPs where relevant; when the distinction is not necessary the term *phrase* is used to refer to APs and IPs as a group.

A. METHOD

1. Participants

Seventy-eight students from Hanyang University in Seoul, Korea participated in the experiment. They were native Seoul Korean speakers, born and educated in Seoul with no self-reported impairments in hearing or speaking. Although English is taught at school in Korea and people are exposed to English in daily life, none of the participants was highly proficient in any language other than Korean. None of the participants had lived outside Korea for more than 18 months. Their age ranged from 19 to 29 years. Six participants were excluded from analysis: two were removed due to technical errors; another four turned out not to meet the recruitment criteria. The results reported here are based on data from 72 participants (38 females and 34 males).

2. Design

The experimental factors were TARGET TYPE (disyllabic, trisyllabic), RHYTHMIC CONSISTENCY (consistent, inconsistent) and LENGTHENING (no, pre-boundary, pre-target). TARGET TYPE refers to the number of syllables in the target words, which were either disyllabic or trisyllabic nouns. RHYTHMIC CONSISTENCY refers to the number of syllables in each phrase: in the consistent condition, all phrases in a string had the same number of syllables; in the inconsistent condition, the syllable count differed in consecutive phrases prior to the target. LENGTHENING refers to the presence (or absence) of phrase-final lengthening: in the no lengthening condition, no phrase-final lengthening occurred in the strings; in the pre-boundary lengthening

condition, all phrase-final syllables were lengthened, while in the pre-target lengthening condition, only the syllable immediately prior to the target word was lengthened.

3. Experimental stimuli

a. Targets in nonsense strings

The targets included 48 test targets, 48 fillers and 12 practice items. Twenty four disyllabic and twenty four trisyllabic nouns were selected as test targets from a database of frequently used Korean words (NIKL, 2005). Their PND was calculated based on Holliday *et al*. (2016). On average, the PND of disyllabic test targets was 39.79 (sd = 18.61), and that of trisyllabic test targets was 2.17 (sd = 1.67); the reasons for the difference are similar to those discussed in Section I. B. Despite the high error rates with disyllabic words in previous studies (e.g., around 50% in Kim & Cho, 2009), and the large differences in PND, both disyllables and trisyllables were included to keep listeners from searching only for words with a certain number of syllables. All test targets consisted of CV syllables; the fillers and practice items consisted of CV or V syllables. Fillers and practice items had similar PND to test targets (for disyllabic fillers, mean = 37.42, sd = 16.50; for trisyllabic fillers, mean = 1.9, sd = 1.09; for disyllabic practice items, mean = 42.67, sd = 11.57; for trisyllabic practice items, mean = 3.6, sd = 2.19). All test targets began with either a lenis or a nasal consonant inducing AP-initial L tone so that the F0 contour of all prosodic phrases could be controlled.

Each target was embedded in a nonsense string of syllables. In order for experimental strings to be 15 syllables long in total, the nonsense strings used with trisyllabic targets were 12 syllables long, while those used with disyllabic targets were

13 syllables long (see Table I). The mean duration of the 15-syllable strings was 3.2 s (sd = 0.2 s). All syllables in the nonsense strings were of CV structure with an onset that triggers AP-initial L. The strings were sequences of the syllables /kɑ, pɑ, tɑ, tɕɑ, mɑ, nɑ, lɑ, kɛ, pɛ, tɛ, tɕɛ, mɛ, nɛ, lɛ, ki, pi, ti, tɕi, mi, ni, li, ko, po, to, tɕo, mo, no, lo, ku, pu, tu, tɕu, mu, nu, lu, kʌ, pʌ, tʌ, tɕʌ, mʌ, nʌ, kɯ, pɯ, tɯ, tɕɯ, mɯ, nɯ, lɯ/. The order of the syllables in each string was pseudo-randomized so that none of the sequences formed a real Korean word apart from the targets; there were no pairs of identical nonsensical strings. Some syllables in the strings are particles or monosyllabic words, e.g., /kɯ/ 'that'. This was unavoidable, as monosyllabic morphemes are abundant in Korean. Although this means that listeners could spot monosyllabic items in the strings, string composition was not a confounding factor, because each string of syllables was used in all prosodic conditions as discussed further in Section II. A. 3. b.

The test targets always began on the tenth syllable in their string, e.g., /tɛpʌnutodipomɛtɯtu**tɕɛtɕʰɛki**pipumɯ/ (test target 'sneeze' in bold). In the resynthesis procedure described in Section II. A. 3. b, the strings were divided into phrases all of which had the same LH F0 contour. The phrases were 3 to 5 syllables long, a length which, as noted earlier, corresponds to that typically observed in APs (Kim, 2004, chap. 3). For the consistent condition, the strings were divided into five phrases of three syllables each, i.e., 3-3-3-**3**-3, where the phrase including the test target is marked in bold. For the inconsistent condition, the strings were divided into four phrases, and the phrases prior to the test target varied in syllable count, i.e., 4-5-**3**-3. As a result of this structure, the test targets were always in a phrase with three syllables; consequently,

The Journal of the Acoustical Society of America 141 (6), 4251-4263, accepted on 26th April 2017, published on-line 8 June;
doi: http://dx.doi.org/10.1121/1.4983178

each trisyllabic test target formed a phrase by itself, whereas the disyllabic test targets were always followed by an additional syllable in the same phrase.

TABLE I. Structure of stimuli. The syllables in boldface were lengthened. Phrase boundaries are indicated by #. The targets are in square brackets.

LENGTHENING	**Test-target-bearing string, consistent RHYTHMIC CONSISTENCY**
no	σ1 σ2 σ3 # σ4 σ5 σ6 # σ7 σ8 σ9 # $_{Tri}$[$_{Di}$[σ10 σ11] σ12] # σ13 σ14 σ15
pre-boundary	σ1 σ2 **σ3** # σ4 σ5 **σ6** # σ7 σ8 **σ9** # $_{Tri}$[$_{Di}$[σ10 σ11] **σ12**] # σ13 σ14 **σ15**
pre-target	σ1 σ2 σ3 # σ4 σ5 σ6 # σ7 σ8 **σ9** # $_{Tri}$[$_{Di}$[σ10 σ11] σ12] # σ13 σ14 σ15
	Test-target-bearing string, inconsistent RHYTHMIC CONSISTENCY
no	σ1 σ2 σ3 σ4 # σ5 σ6 σ7 σ8 σ9 # $_{Tri}$[$_{Di}$[σ10 σ11] σ12] # σ13 σ14 σ15
pre-boundary	σ1 σ2 σ3 **σ4** # σ5 σ6 σ7 σ8 **σ9** # $_{Tri}$[$_{Di}$[σ10 σ11] **σ12**] # σ13 σ14 **σ15**
pre-target	σ1 σ2 σ3 σ4 # σ5 σ6 σ7 σ8 **σ9** # $_{Tri}$[$_{Di}$[σ10 σ11] σ12] # σ13 σ14 σ15
	Filler-bearing string, consistent RHYTHMIC CONSISTENCY
no	σ1 σ2 σ3 # $_{Tri}$[$_{Di}$[σ4 σ5] σ6] # σ7 σ8 σ9 # σ10 σ11 σ12 # σ13 σ14 σ15
pre-boundary	σ1 σ2 **σ3** # $_{Tri}$[$_{Di}$[σ4 σ5] **σ6**] # σ7 σ8 **σ9** # σ10 σ11 **σ12** # σ13 σ14 **σ15**
pre-target	σ1 σ2 **σ3** # $_{Tri}$[$_{Di}$[σ4 σ5] σ6] # σ7 σ8 σ9 # σ10 σ11 σ12 # σ13 σ14 σ15
	Filler-bearing string, inconsistent RHYTHMIC CONSISTENCY
no	σ1 σ2 σ3 $_{Tri}$[$_{Di}$[σ4 # σ5] σ6] σ7 σ8 σ9 # σ10 σ11 σ12 # σ13 σ14 σ15
pre-boundary	σ1 σ2 σ3 $_{Tri}$[$_{Di}$[**σ4** # σ5] σ6] σ7 σ8 **σ9** # σ10 σ11 **σ12** # σ13 σ14 **σ15**
pre-target	σ1 σ2 σ3 $_{Tri}$[$_{Di}$[**σ4** # σ5] σ6] σ7 σ8 σ9 # σ10 σ11 σ12 # σ13 σ14 σ15

The Journal of the Acoustical Society of America 141 (6), 4251-4263, accepted on 26th April 2017, published on-line 8 June;
doi: http://dx.doi.org/10.1121/1.4983178

TABLE II. Structural interpretation of each experimental condition (test-target-bearing strings only). The phrase including the target is marked in bold. Audio files of an example stimulus /tɛpʌnutotipomɛtɯtu**tɕɛtɕʰɛki**pipumɯ/ (target "sneeze" in bold) are provided.

LENGTHENING	**RHYTHMIC CONSISTENCY**	
	consistent	inconsistent
no	[[3]$_{AP}$ [3]$_{AP}$ [3]$_{AP}$ [**3**]$_{AP}$ [3]$_{AP}$]$_{IP}$ (Mm. 1. 273 KB)	[[4]$_{AP}$ [5]$_{AP}$ [**3**]$_{AP}$ [3]$_{AP}$]$_{IP}$ (Mm. 4. 273 KB)
pre-target	[[3]$_{AP}$ [3]$_{AP}$ [3]$_{AP}$]$_{IP}$ [[**3**]$_{AP}$ [3]$_{AP}$]$_{IP}$ (Mm. 2. 280 KB)	[[4]$_{AP}$ [5]$_{AP}$]$_{IP}$ [[**3**]$_{AP}$ [3]$_{AP}$]$_{IP}$ (Mm. 5. 280 KB)
pre-boundary	[[3]$_{AP}$]$_{IP}$ [[3]$_{AP}$]$_{IP}$ [[3]$_{AP}$]$_{IP}$ [[**3**]$_{AP}$]$_{IP}$ [[3]$_{AP}$]$_{IP}$ (Mm. 3. 307 KB)	[[4]$_{AP}$]$_{IP}$ [[5]$_{AP}$]$_{IP}$ [[**3**]$_{AP}$]$_{IP}$ [[3]$_{AP}$]$_{IP}$ (Mm. 6. 301 KB)

In addition to the test-target-bearing strings, 48 filler-bearing strings were created. The filler-bearing strings included 24 disyllabic and 24 trisyllabic fillers; they began on the *fourth* syllable in each string to distract listeners from the fact that all test targets started on the tenth syllable. Unlike the test targets which aligned with the left edge of a phrase in both the consistent and inconsistent conditions, the fillers in the inconsistent condition straddled a phrase boundary (see Table I). In addition, the experiment included twelve practice items, six with disyllables and six with trisyllables. In the practice strings, half of the targets (three disyllables and three trisyllables) began on the fourth syllable, and the other half on the tenth syllable.

b. Recording and resynthesis of the strings

Author HJ, a female native speaker of Seoul Korean and trained phonetician, recorded the original utterances for resynthesis. Each of the syllables appearing in the stimuli was embedded as the second syllable in the second AP in the carrier sentence /ikʌsɯn#ɑ__ɑ#imnitɑ/ (‘this is a____a’; ‘this thing’ + topic marker + ‘a__a’ (nonword) + ‘be’ + polite sentence ender). The target syllable was preceded and followed by /ɑ/, a low back vowel in Korean (Shin, 2015), since the syllables spoken in this context were judged to be the most natural-sounding when resynthesized. The speaker read all sentences at a comfortable speaking rate several times, paying attention to produce AP boundaries at the desired locations. The recording was made in a sound-attenuated booth at the Phonetics Laboratory, University of Cambridge with a Nagra ARES-BB+ digital recorder and a Sennheiser MKH 40 cardioid microphone (manufactured in Germany). The sampling rate was 44.1 kHz with 16-bit quantification.

Praat (Boersma & Weenink, 2014) was used for all editing and resynthesis processes. In order to create the base utterances for further resynthesis, syllables without irregular pulses were extracted from the original recordings (irregular pulses were excluded so as not to degrade the quality of resynthesized materials). When the syllable had /ɑ/ and was followed by /ɑ/ (in the /#ɑ__ɑ#/ carrier phrase), the dip in the amplitude envelop and/or the onset of creaky voice seen in the spectrogram was used as a marker of the boundary between the two adjacent /ɑ/ vowels. The selected syllables were concatenated to create 54 15-syllable strings (24 for test targets, 24 for fillers, and 6 for practice; see below). This process ensured that no coarticulatory cues were available in the strings. The PSOLA method was used to resynthesize these base

utterances so as to have flat F0 of 200 Hz; this frequency was chosen because it was close to the speaker's typical F0 in IP-medial vowels carrying a L-tone. Afterwards each string was edited to remove any clipped sections resulting from the concatenation process.

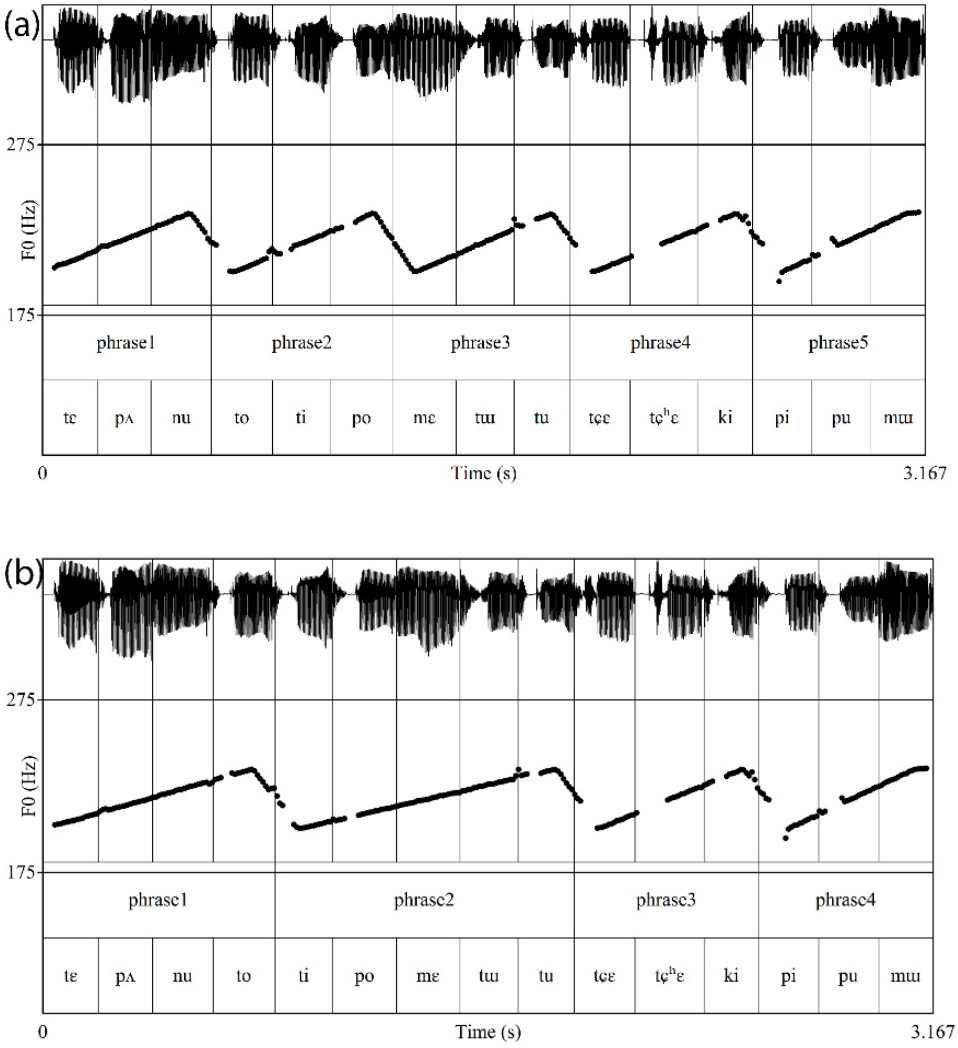
Each of the 54 strings was further manipulated to create two versions, a 3-3-3-3-3 phrasing and a 4-5-3-3 phrasing for the consistent and inconsistent conditions of RHYTHMIC CONSISTENCY respectively. F0 was manipulated to group the syllables into phrases. F0 was specified to be 200 Hz in the middle of the first vowel in each phrase and 240 Hz in the middle of the last vowel; the F0 between these two targets was interpolated (see Fig. 1). Consequently, all phrases had a LH pitch contour which frequently occurs in the AP (see Section I. B).

A corollary of the differences in phrasing was that the number of phrases preceding the target differed between the two RHYTHMIC CONSISTENCY conditions. As illustrated in Figure 1, in both conditions the location of the test target was constant, but the strings in the consistent condition had three phrases before the test target, while the inconsistent strings had two. As a result, the former showed more F0 excursions before the test target than the latter. This difference creates a possible confound between rhythmic consistency, reflected in the number of syllables per phrase, and complexity, reflected in the F0 contours. Despite the confound, the 4-5-3-3 phrasing was used for the following reason: if the number of syllables in the string and the location of the test target were to stay constant, the alternative phrasings for the inconsistent condition would include 2-, 3- and 4-syllable phrases before the target (e.g., 2-3-4-**3**-3); this, however, would introduce a trisyllabic phrase early in the string (before

The Journal of the Acoustical Society of America 141 (6), 4251-4263, accepted on 26th April 2017, published on-line 8 June;
doi: http://dx.doi.org/10.1121/1.4983178

the target), giving rise to a rhythmic structure that might not have been sufficiently different from that of the consistent condition.

FIG.1. Example waveforms and F0 tracks for (a) consistent (3-3-3-3-3 phrasing) and (b) inconsistent (4-5-3-3 phrasing) syllable counts in the no lengthening condition. In panel (a), the test target is in phrase 4; in panel (b) it is in phrase 3.



The resulting 108 strings with resynthesised F0 contours were further manipulated for lengthening. For the no lengthening condition, duration was left unchanged. For the pre-boundary lengthening condition, all phrase-final vowels were lengthened by a factor of 1.7, which is within the range of IP-final lengthening reported

for Korean (see Section I. B; the same factor was employed in Kim & Cho, 2009). For the pre-target lengthening condition, *only* the vowel preceding the target was lengthened by a factor of 1.7. In both conditions, lengthening resulted in vowel duration increasing by 82 ms on average (sd = 17 ms) relative to no lengthening. Thus, the six experimental conditions (RHYTHMIC CONSISTENCY [consistent, inconsistent] × LENGTHENING [no, pre-boundary, pre-target]) represent different types of phrasal organization (see Table II)[1]. Specifically, in the no lengthening condition, we expected participants to interpret the strings as sequences of APs; in the pre-boundary lengthening condition, strings should be interpreted as IP sequences instead; in the pre-target lengthening condition, the expected interpretation was that the target word was preceded by an IP boundary, while the other boundaries would be interpreted as AP-level boundaries.

In the test-target-bearing strings, there were no potentially inhibitory prosodic properties, such as acoustic ambiguity at the phrase boundary, targets straddling a prosodic boundary, or F0 contours unattested in Korean. In contrast, the fillers in the inconsistent condition straddled a phrase boundary (Section II. A. 3). Given the results of Kim (2004), it was expected that these fillers would be particularly challenging for the participants.

The resynthesis process yielded 576 strings in total for the main experiment and 72 practice strings (6 practice strings with disyllabic targets and 6 with trisyllabic targets × 2 RHYTHMIC CONSISTENCY × 3 LENGTHENING). The 576 strings included 288 test-target-bearing strings and 288 filler-bearing strings (24 strings with disyllabic targets and 24 strings with trisyllabic targets × 2 RHYTHMIC CONSISTENCY × 3

LENGTHENING). A Latin square design was used: the strings were divided into 6 lists, each of which had 12 strings for the practice session (2 strings from each experimental condition) and 96 strings for the main experiment (16 strings from each experimental condition); half of these (N = 48) were filler-bearing strings.

4. Experimental procedure

The participants were tested individually in a quiet room in the Hanyang Phonetics & Psycholinguistics Laboratory, Seoul. They were randomly assigned to one of the six lists mentioned above. Stimuli presentation and data collection were performed using PsychoPy2 ver.1.83.01. Participants heard the stimuli on a desktop PC through a pair of Microsoft headphones with a noise-cancelling microphone attached. Participants were told that they would hear nonsense strings with a real Korean word in each string. They were asked to press the spacebar on the keyboard with their preferred hand as quickly as possible when they spotted a real word and to say the word aloud. After recording their verbal response, they had to press the spacebar to proceed to the next trial. Before the main experiment, each participant was presented with 12 practice strings. The presentation order of all strings was randomised for each participant. The strings in the main experiment were divided into three blocks and participants took 1 min compulsory breaks after each block (i.e., every 32 trials) for a total of two breaks. The experiment took approximately 15 minutes to complete. Reaction times, for key-presses only, were recorded from each target word offset. After the experiment, all participants completed a written questionnaire on their linguistic background and were given a 5,000 KRW (approximately 4.5 USD) voucher as remuneration for their time and effort.

B. Results

The aim of the analysis was to investigate whether word-spotting was affected by prosody. To this end, we analysed both accuracy, as is typically done in spotting studies (McQueen, 1996), and reaction times (RTs) *to correct responses*. RTs are often not analysed in spotting experiments because they tend to be long and variable; e.g. in Kim (2004), mean RTs ranged from 573 to 1408 ms depending on experimental conditions. Long and variable RTs suggest that the experimental task is cognitively demanding and that listeners may identify the target after the string offset. Further, while in other processing tasks a trade-off may be observed between accuracy and speed (e.g., Heitz, 2014), RTs and error rates in word-spotting experiments tend to show similar patterns (e.g., Kim, 2004; Kim & Cho, 2009). Nevertheless, we examined both measurements here, on the assumption that accuracy results reflect the prosodic parameters required for successful word-spotting, while RTs can reveal factors facilitating the process. Although the fillers were included as distractors in the experiment, we also analysed accuracy for them, since responses to fillers could shed light on the role of the local pitch cue (see Section II. A. 3).

1. Accuracy

Participants' verbal responses were coded as 0 (correct) or 1 (error). Missing and incorrect responses were both treated as errors; no further analysis on error types was conducted as the majority were missing responses (81% for disyllabic and 71% for trisyllabic test targets).

TABLE III. The effect of fixed factors TARGET TYPE (disyllabic, trisyllabic), RHYTHMIC CONSISTENCY (consistent, inconsistent) and LENGTHENING (no, pre-boundary, pre-target), as results of model comparisons. The dependent variable was accuracy (0 correct, 1 error). Significance codes *** p < 0.001, ** p < 0.01, * p < 0.05

	df	χ^2
TARGET TYPE	1	46.85***
RHYTHMIC CONSISTENCY	1	14.24***
LENGTHENING	2	1.1
TARGET TYPE × RHYTHMIC CONSISTENCY	1	5.42*
TARGET TYPE × LENGTHENING	2	7.15*
RHYTHMIC CONSISTENCY × LENGTHENING	2	0.94
TARGET TYPE × RHYTHMIC CONSISTENCY × LENGTHENING	2	3.78

Mixed-effects logistic regression models were constructed in R (R Core Team, 2015; lme4 package, Bates *et al*., 2015) for the responses to the test targets. The modelling process began with the full model including the fixed factors TARGET TYPE (disyllabic, trisyllabic), RHYTHMIC CONSISTENCY (consistent, inconsistent), and LENGTHENING (no, pre-boundary, pre-target). PARTICIPANT and ITEM were treated as random factors with random intercepts.

The effects of the fixed factors were established by model comparisons, using log-likelihood χ^2 tests (see Tables III for a summary). Only statistically significant results are discussed. There were statistically significant main effects of TARGET TYPE (χ^2 (1) = 46.85, $p < 0.001$) and RHYTHMIC CONSISTENCY (χ^2 (1) = 14.24, $p < 0.001$). Disyllabic test targets (error rate, mean = 62.62, sd = 28.90) were significantly more likely to lead to an error than trisyllabic test targets (mean = 12.91, sd = 18.10). For RHYTHMIC CONSISTENCY, the error rate was higher in the consistent condition (mean = 39.93, sd = 35.80) than in the inconsistent condition (mean = 35.59, sd =

33.32). The two-way interactions TARGET TYPE x RHYTHMIC CONSISTENCY (χ^2 (1) = 5.42, $p < 0.05$) and TARGET TYPE x LENGTHENING (χ^2 (2) = 7.15, $p < 0.05$) were statistically significant. Since the difference between the two levels of TARGET TYPE (disyllabic, trisyllabic) was substantial, and TARGET TYPE interacted both with RHYTHMIC CONSISTENCY and LENGTHENING, separate models were constructed for each TARGET TYPE. This allowed us to simplify the models and better explore the interactions involving the two prosodic cues, which were the most interesting for our purposes.

As summarised in Table IV, the final model for each TARGET TYPE included the fixed factors RHYTHMIC CONSISTENCY (consistent, inconsistent) and LENGTHENING (no, pre-boundary, pre-target), with PARTICIPANT and ITEM as random factors. The RHYTHMIC CONSISTENCY × LENGTHENING interaction was excluded from the model for disyllabic targets as it did not improve fit (χ^2 (2) = 3.21, ns).

For trisyllabic test targets, none of the factors in the model was statistically significant. For the disyllabic test targets, the consistent condition was more likely to lead to an error (inconsistent as reference level, est = 0.55, SE = 0.12, z = 4.6, $p < 0.001$; see Table IV for a model summary and Table V for error rates). Pre-target LENGTHENING was less likely to lead to an error in comparison to no lengthening (est = -0.28, SE = 0.14, z = -0.2, $p < 0.05$); the model with pre-target LENGTHENING as the reference level confirmed that the only significant difference was between no lengthening and pre-target lengthening (pre-boundary lengthening, est = 0.27, SE = 0.14, z = 1.21, ns). That is, accuracy in disyllabic targets significantly improved only in pre-target lengthening as compared to no lengthening, but there was no statistically

significant difference between either pre-boundary and pre-target lengthening, or between pre-boundary and no lengthening.

TABLE IV. Logistic regression model summary for accuracy (reference level = inconsistent RHYTHMIC CONSISTENCY, no LENGTHENING) for each TARGET TYPE. The dependent variable was coded as 0 (correct) and 1 (error). Significance codes *** p < 0.001, ** p < 0.01, * p < 0.05

	TARGET TYPE					
	disyllabic (n = 1752)			trisyllabic (n = 1752)		
	est.	SE	z	est.	SE	z
intercept	0.63	0.30	2.09*	-2.71	0.34	-7.98***
RHYTHMIC CONSISTENCY: consistent	0.55	0.12	4.63***	0.04	0.16	0.23
LENGTHENING: pre-boundary	-0.11	0.14	-0.75	-0.08	0.20	-0.42
LENGTHENING: pre-target	-0.28	0.14	-1.96*	0.29	0.19	1.53

TABLE V. Means and standard errors of error rates (%) derived from participant means and presented separately for each level of RHYTHMIC CONSISTENCY and LENGTHENING.

			TARGET TYPE			
			disyllabic		tridisyllabic	
			Mean	SE	Mean	SE
RHYTHMIC CONSISTENCY	test	inconsistent	58.22	1.99	12.96	1.26
		consistent	67.01	1.9	12.85	1.21
	filler	inconsistent	98.5	0.5	93.75	0.82
		consistent	82.87	1.35	28.47	1.54
LENGTHENING	test	no	64.76	2.3	12.5	1.48
		pre-boundary	63.02	2.45	11.46	1.32
		pre-target	60.07	2.47	14.76	1.7

	filler	no	92.19	1.29	61.46	3.17
		pre-boundary	89.41	1.49	61.11	3.12
		pre-target	90.45	1.42	60.76	3.06

Error rates were higher for the fillers (mean = 75.90, SE = 1.11) than the test targets (mean = 37.76, SE = 1.18). Filler error rates were particularly high in the *inconsistent* condition, in contrast to the results from the (disyllabic) test targets (see below). Specifically, in the inconsistent condition (in which the fillers straddled the phrase boundary), listeners missed 98.5% of the disyllabic and 93.75% of the trisyllabic fillers (Table V). The error rates were substantially lower in the consistent condition (at 82.87% for disyllables and 28.47% for trisyllables). Mixed-effects logistic models were constructed with the data from both fillers and test targets; the fixed factors were TEST-FILLER (test, filler), TARGET TYPE (disyllabic, tri- syllabic), RHYTHMIC CONSISTENCY (consistent, inconsistent), and LENGTHENING (no, pre-boundary, pre-target); the random factors were PARTICIPANT and ITEM. The model comparisons revealed two statistically significant interactions: TEST-FILLER × TARGET TYPE (χ^2 (1) = 11.14, p<0.001] and TEST-FILLER × TARGET TYPE × RHYTHMIC CONSISTENCY [χ^2 = (2)12.41, p<0.01]. In the follow up models, constructed for each TARGET TYPE of the fillers, the RHYTHMIC CONSISTENCY effect was statistically significant for both disyllabic and trisyllabic fillers, in that listeners' performance was worse in the inconsistent than the consistent condition for both filler types (disyllabic fillers, est. = -3.1, SE = 0.33, z = -9.5, p < 0.001; trisyllabic fillers, est. = -4.99, SE = 0.25, z = -20.2, p < 0.001).

The Journal of the Acoustical Society of America 141 (6), 4251-4263, accepted on 26th April 2017, published on-line 8 June;
doi: http://dx.doi.org/10.1121/1.4983178

2. Reaction Times for the Correct Responses

The analysis of RTs was based on data from the test targets only. The RTs to the fillers were not analysed, since the results would not be reliable due to the high error rates with fillers (see Section II. B.1). RTs longer than 2000 ms were treated as outliers and removed from the analysis, leaving 72% of the correct responses to the disyllabic test targets and 89% of the responses to the trisyllabic test targets (disyllabic, mean = 1130.18 ms, SE = 15.43; trisyllabic, mean = 954.71 ms, SE = 11.01). Since RTs showed a positively skewed distribution (disyllabic, median = 1053.82 ms; trisyllabic, median = 826.53 ms), as is typical for RT data (Ratcliff, 1993), using mean-based measures such as standard deviations to identify and remove outliers was deemed inappropriate. The cut-off value 2000 ms was chosen, as it avoided these problems and retained RTs that fell within the temporal window for simultaneous processing (cf. 'psychological present' in Fraisse, 1984).

TABLE VI. Mean RTs and standard errors (in ms) for correct responses to test targets only. The statistics were calculated from all data points and presented separately for each level of RHYTHMIC CONSISTENCY and LENGTHENING (disyllabic, n = 469; trisyllabic, n = 1360).

		TARGET TYPE			
		disyllabic		trisyllabic	
		Mean	SE	Mean	SE
RHYTHMIC CONSISTENCY	inconsistent	1133.7	20.57	970.28	15.81
	consistent	1125.6	23.39	939.04	15.33
LENGTHENING	no	1148.1	25.97	981.99	18.86
	pre-boundary	1094.2	27.03	881.33	18.38
	pre-target	1146.5	26.92	1001.7	19.55

Due to the large difference in error rates, the RT dataset for disyllabic test targets was substantially smaller than that for trisyllabic targets. For this reason, mixed-effect models were constructed separately for each TARGET TYPE as in the accuracy analysis; the models included the fixed factors RHYTHMIC CONSISTENCY (consistent, inconsistent) and LENGTHENING (no, pre-boundary, pre-target), and the random factors PARTICIPANT and ITEM with random intercepts. The effects of the fixed factors were established by model comparisons, using log-likelihood χ^2 tests (see Table VII).

For the disyllabic test targets, neither RHYTHMIC CONSISTENCY nor LENGTHENING had a statistically significant effect on RTs. For the trisyllabic test targets, there was a statistically significant effect of LENGTHENING (χ^2 (2) = 31.93, $p < 0.001$; see Table V). Tukey contrast tests (multcomp package, Hothorn *et al*., 2008) revealed that RTs were significantly faster in the pre-boundary lengthening condition than in both the no lengthening and the pre-target lengthening conditions (Table VI). That is, final-lengthening in all phrases helped listeners respond faster to trisyllabic words, despite the ceiling effect observed in the accuracy results. RHYTHMIC CONSISTENCY, on the other hand, did not affect the RTs of trisyllabic targets.

III. DISCUSSION

The results demonstrate that the effects of prosody on word segmentation differed depending on target type (disyllabic vs. trisyllabic); at the same time, rhythmic consistency (manifested in syllable count across prosodic phrases) and phrase-final lengthening affected listeners' performance without interacting. The accuracy analysis revealed that listeners relied on rhythmic consistency and lengthening cues in an

unexpected way: the consistent condition adversely affected accuracy with disyllabic test targets but was beneficial with fillers; pre-target LENGTHENING had a facilitating effect with disyllabic test targets, but no effect on trisyllables. The experimental conditions did not strongly affect RTs; the only facilitating effect was observed with pre-boundary LENGTHENING for trisyllabic test targets. The effects of target type and prosodic context are discussed further in Sections III. A and III. B respectively.

TABLE VII. Results of model comparisons for RTs (top) and mixed-effect model summary (bottom). The fixed factors are RHYTHMIC CONSISTENCY and LENGTHENING; reference level = inconsistent RHYTHMIC CONSISTENCY, no LENGTHENING. Significance codes *** p < 0.001, ** p < 0.01, * p < 0.05. Both results and summary are presented separately for each TARGET TYPE.

	TARGET TYPE					
	disyllabic (n = 469)			trisyllabic (n = 1360)		
	df	χ^2		df	χ^2	
RHYTHMIC CONSISTENCY	1	0.30		1	3.01	
LENGTHENING	2	2.35		2	31.93***	
RHYTHMIC CONSISTENCY × LENGTHENING	2	2.26		2	2.33	
Model summary	est.	SE	t	est.	SE	t
intercept	1203.65	39.88	30.19	1045.59	47.77	21.89
RHYTHMIC CONSISTENCY: consistent	14.91	27.11	0.55	-22.29	12.86	-1.73
LENGTHENING: pre-boundary	-51.59	34.47	-1.50	-75.57	15.70	-4.81

LENGTHENING: pre-target	-17.75	33.22	-0.53	3.71	15.80	0.24

TABLE VIII. Results of Tukey contrast tests on factor LENGTHENING for RTs of trisyllabic test words only. Significance codes *** p < 0.001, ** p < 0.01, * p < 0.05

	est.	SE	z
no lengthening vs. pre-boundary	75.57	15.70	4.81***
pre-target vs. pre-boundary	79.29	15.83	5.01***
pre-target vs. no lengthening	3.71	15.80	0.24

A. Target type

In the present experiment, spotting a trisyllabic word in a nonsense string was a relatively easy task for listeners regardless of prosodic context. This result is comparable to that of Kim (2004), who also found that accuracy was at ceiling with trisyllabic targets. Disyllabic targets, on the other hand, posed a challenge both in Kim (2004) and the present study, resulting in error rates higher than 50% (Table V).

The difference between the two TARGET TYPEs is probably related to the lower PND of the trisyllabic targets (Section II. A. 3. a). However, it is also likely that in the present experiment the difference also had to do with phrasal structure. Specifically, trisyllabic targets exactly matched the (always trisyllabic) phrases in which they were placed; by virtue of this structure, the trisyllabic targets mapped onto phrases and were tonally demarcated both at onset and offset, by #L and H# respectively (cf. Kim, 2004). The lack of such mapping for disyllabic targets may, in turn, explain the high error rates

for them: unlike trisyllables, disyllabic targets were always followed by a nonsense syllable in their phrase (see Table I), and thus their offset did not coincide with the end of a phrase and was not tonally demarcated in any way. Although emphasis is typically placed on the role of the left edge of prosodic constituents in marking word onsets, it is plausible that word and phrase mapping (or lack thereof) plays an important role in Korean. As noted earlier, word-spotting is cognitively demanding and assumed to take place after a word is heard; this suggests that listeners may retrieve a target after the carrier string offset. If so, then Korean listeners may have had greater difficulty spotting a target when its offset did not correspond to the phrase's right edge and was not tonally demarcated in any way (as happened with the disyllabic targets in the present experiment). The importance of tonal demarcation is also reflected in the results from fillers that straddled a phrasal boundary, which had the highest error rates in the entire experiment.

Therefore, it seems that the difference in listeners' performance between disyllabic and trisyllablic test targets was due to a combination of factors. For the trisyllabic test targets, low PND and the clear mapping between the word and its phrase marked with a local pitch cue probably offered additive advantages to listeners to the extent that other prosodic cues were not required for successful spotting. On the other hand, both RHYTHMIC CONSISTENCY and LENGTHENING were used to spot disyllabic test targets. That is, when listeners were faced with the more cognitively demanding task of recognising disyllabic words with high PND and when these words were not clearly demarcated by pitch *at both phrasal onset and offset*, other prosodic cues came to the fore. The role of these cues is discussed in Section III.B.

B. Prosodic context and word-spotting

We expected that consistent rhythmic structure (in the shape of a stable syllable count across prosodic phrases) would allow listeners to form expectations about the location of upcoming phrase boundaries and thus of word onsets since the two coincided (with the exception of the fillers, which are discussed below). This hypothesis was only partially supported. If global rhythmic consistency had facilitated word recognition in a straightforward manner, the consistent condition would have provided an ideal context for spotting, offering high predictability for word onset timing. This did not apply to trisyllabic targets, most likely because accuracy for them was at ceiling, as discussed in Section III. A. The rhythmically consistent condition resulted in higher accuracy for fillers, though surprisingly, it *negatively* impacted accuracy with disyllabic test targets. There are a number of possible explanations for this unexpected finding.

First, as noted in Section III. A, the repeating pattern of trisyllabic phrases in the consistent condition may indeed have created expectations of rhythmic consistency, making it difficult to spot disyllabic targets in trisyllabic phrases. On the other hand, the inconsistent condition did not lead listeners to create expectations of rhythmic consistency in terms of phrase size, or of a match between phrase size and target; hence, for those stimuli there was no violation of expectations that could exert an inhibitory effect. In this sense, the results do suggest that rhythmic consistency could create expectations about prosodic structure and that such expectations could have a facilitating effect *under the right conditions*.

A second possible reason may have to do with prosodic complexity. Recall that all experimental strings had 15 syllables and the test target always started on the 10th

syllable; in the consistent condition the preceding 9 syllables formed three phrases, but in the inconsistent condition they formed only two. Since each phrase had a LH F0 contour, listeners heard more F0 rises over shorter phrases in the consistent condition (see Fig. 1). Consequently, the rhythmically consistent strings may have been perceived as more complex in terms of pitch than the inconsistent strings. It is possible that this perceived complexity increased the cognitive load and further inhibited the spotting of the disyllabic words which generally posed a challenge to listeners.

In addition to the effects of rhythmic consistency, we expected that local phrase-final lengthening would facilitate spotting, particularly when present in all prosodic phrases. This expectation was also only partially met. For the disyllabic test targets, only lengthening occurring immediately before the target reduced errors, whereas lengthening in all phrases significantly reduced RTs for trisyllabic targets only. These differences further indicate that the extent to which local cues are exploited depends on the cognitive load of the task. Pre-target lengthening provided less structural complexity, i.e., only one clear IP-level boundary prior to the test target (see Table II). This lack of complexity probably helped with the spotting of disyllabic targets in comparison to the no-lengthening condition, which would have been interpreted as a sequence of APs: pre-target lengthening clearly signalled the presence of a prosodic boundary and thus of the upcoming word onset. In contrast, for trisyllabic test targets, it was the presence of repeated IP-level lengthening that resulted in shorter RTs. This was possibly because the lengthening provided additional processing time for participants, as predicted, thereby speeding up spotting; however, the possibility that this type of lengthening was

treated as a global cue that reinforced expectations about the match between targets and phrases cannot be dismissed.

Overall, the results indicate the role of prosody during word-spotting but also highlight the difficulty in determining one-to-one mapping between prosodic cues and their consequences for speech processing. In sum, the prosodic cues tested here were used primarily in challenging conditions, i.e., for the spotting of disyllabic test targets. With trisyllabic test targets, which were overall easier to spot, recurring phrase-final lengthening – which can be interpreted as a series of cues that help create expectations about prosodic structure – was exploited.

C. Word-spotting in Korean

The results of the present experiment are in line with those of previous studies of Korean and other languages which demonstrate that prosody plays a significant role in speech processing.

First, the results agree with previous studies on the cues to prosodic boundaries in Korean (Kim & Cho, 2009) and other languages (Tyler & Cutler, 2009; Kim, *et al*. 2012) in showing that lengthening is interpreted as a cue to phrase-finality. Further, the present study indirectly agrees with previous studies on the role of local pitch cues in Korean (Jeon & Nolan, 2010; Kim, 2004; Kim & Cho, 2009). This conclusion is based on the overall results, which showed that spotting was faster and more accurate when there was a match between words and phrases (which were always tonally demarcated), as happened with the trisyllabic test targets. Additional evidence on this point comes from the fillers. In the inconsistent condition, the fillers straddled a prosodic

phrase boundary and therefore they presented a mismatch between the target and the prosodic phrase, in that the local pitch cue indicated that a phrase boundary occurred within the target. Accuracy with fillers was significantly lower in the inconsistent condition, showing that word recognition occurs within the prosodic phrase (e.g. Christophe *et al*. 2004 on French; Kim 2004 on Korean).

Further, the present results indicate cross-linguistic differences in the importance of specific prosodic cues in speech processing. Specifically, our hypothesis was that since Korean lacks word-internal metrical structure, while phrasal and word onsets tend to coincide, listeners' expectations about the timing of word onsets were likely to depend on the presence of phrasal cues (cf. the role of stress in processing in English in which words tend to begin with a stressed syllable; see Cutler, 2012, chap. 4). As it turned out, this expectation did not translate into a strong need for consistent rhythm across phrases. This is very different from what is observed in English with respect to foot structure (the closest equivalent to the Korean AP), which is critical both for grammatical word-level phenomena (Hayes, 1995) and for speech processing (Quené & Port, 2005 and references therein; Dilley & McAuley, 2008; Brown et al., 2011; Morrill et al., 2014).

The difference may be related to the variability of the Korean AP and the extent to which this variability shapes speaker expectations about rhythmic consistency. For instance, Kim's (2004) analysis of 3,085 APs from radio speech suggests that the APs of a given utterance are likely to be of different sizes. In Kim (2004), the number of syllables in the AP ranged from one to seven; although trisyllabic APs were the most common, comprising approximately 50% of all APs in her corpus, it is inevitable that

they would be interspersed with APs of a different size. This lack of consistency in AP length is also reflected in *sijo*, the most 'rhythmic' Korean verse: there is commonly variation in the number of syllables in each linguistic unit in the written form similar to the AP, and the syllable count within this unit is not strictly regulated (see Jeon, 2015b).

The role of pitch cues and the presence of variability in phrase length, together with the arguments presented in Section III. A, can explain why rhythmic consistency at the phrasal level did not have a strong facilitating effect in the present experiment. At the same time, however, the role of listeners' expectations on rhythmic consistency remains evident: listeners were less accurate or slower when there was a mismatch between the target and rhythmic structure, either because targets straddled phrasal boundaries (as did fillers in the inconsistent condition) or because targets were not isomorphic to phrases (as were disyllabic targets); when all cues were present in sync (as with trisyllabic test targets) performance was at ceiling.

IV. CONCLUSIONS

The present study demonstrated that prosodic conditions, both global and local, affected word-spotting in Korean, albeit not always as expected. When the task was challenging, as when spotting disyllabic targets, lower prosodic complexity and a local lengthening cue assumed more importance than rhythmic consistency. Nevertheless, rhythmic consistency – whether manifested in the AP length or in the recurrence of phrasal boundaries – did play a role in generating expectations during word-spotting. This was directly indicated by the RTs for trisyllabic test targets, which were faster when lengthening was present at all phrasal boundaries; it was also indirectly reflected in the

inhibitory effect that rhythmic consistency had on disyllabic test targets (which did not fit expectations that words to be spotted and their phrases would be isomorphic). Overall, the results testify to the role and complex interactions of local and global prosodic cues during speech processing. They further show that these prosodic effects, though observed cross-linguistically, carry different weight depending on the language. For Korean, pitch remains a primary local cue; at the same time, the usefulness of rhythmic consistency, recurring phrase-final lengthening, and the matching of words and phrases, lend credence to the importance of phrasal organization for processing in Korean.

ACKNOWLEDGEMENTS

We thank Taehong Cho for giving us access to the Hanyang Phonetics and Psycholinguistics Laboratory. We also thank Yuna Baek, Jiyoun Choi, Jiyoung Jang, Hyojin Kim, Miru Lee, and Jinhee Park for their assistance in conducting the experiment, and Bob Ladd for his helpful comments on the manuscript. This study was supported by an Academy of Korean Studies Grant (Grant No. AKS- 964 2014-R-13) to A. A. and H.-S. J. This support is hereby gratefully acknowledged.

APPENDIX

The test targets in Korean orthography together with phonemic transcription, gloss and phonological neighborhood density (see Table IX).

The Journal of the Acoustical Society of America 141 (6), 4251-4263, accepted on 26th April 2017, published on-line 8 June;
doi: http://dx.doi.org/10.1121/1.4983178

TABLE IX. The test targets in Korean orthography together with phonemic transcription, gloss and phonological neighborhood density.

Disyllabic test targets				**Trisyllabic test targets**			
word	phonemic transcription	gloss	PND	word	phonemic transcription	gloss	PND
가게	/kɑkɛ/	shop	60	가자미	/kɑtɕɑmi/	flounder	1
가구	/kɑku/	household	54	구더기	/kutʌki/	maggot	4
가지	/kɑtɕi/	branch	61	개나리	/kɛnɑli/	forthysia	8
개미	/kɛmi/	ant	36	고구마	/kokumɑ/	sweet potato	1
고개	/kokɛ/	hill	55	그래프	/kɯlɛpʰɯ/	graph	1
고기	/koki/	meat	90	나누기	/nɑnuki/	division	1
구두	/kutu/	shoes	13	나들이	/nɑtɯli/	outing	2
그네	/kɯnɛ/	swing	28	나머지	/nɑmʌtɕi/	remainder	1
나라	/nɑlɑ/	country	23	도가니	/tokɑni/	crucible	1
나무	/nɑmu/	tree	27	도라지	/tolɑtɕi/	bellflower	3

노래	/nolɛ/	song	20	도토리	/totʰoli/	acorn	1
다리	/tɑli/	leg	45	두루미	/tulumi/	crane	1
도로	/tolo/	road	29	누더기	/nutʌki/	rag	3
도시	/tosi/	city	45	두더지	/tutʌtɕi/	mole	1
모기	/moki/	mosquito	60	마누라	/mɑnulɑ/	wife	2
무게	/mukɛ/	weight	30	마스크	/mɑsɯkʰɯ/	mask	4
무대	/mutɛ/	stage	23	메뚜기	/mɛt*uki/	grasshopper	1
배구	/pɛku/	volleyball	41	미나리	/minɑli/	water parsley	4
바다	/pɑtɑ/	sea	67	바구니	/pɑkuni/	basket	4
바지	/pɑtɕi/	trousers	39	바나나	/pɑnana/	banana	1
부모	/pumo/	parent	24	보따리	/pot*ɑli/	package	1
조카	/tɕokʰɑ/	nephew	16	자투리	/tɕɑtʰuli/	leftover	3
주부	/tɕupu/	housewife	24	주머니	/tɕumʌni/	pocket	2
지구	/tɕiku/	earth	45	재채기	/tɕɛtɕʰɛki/	sneeze	1

ENDNOTES

1. In the strings, not all acoustic cues to AP or IP boundaries were present at both edges of phrases, since we wished to focus on the durational cue demarcating the right edges prior to the target word. For example, articulatory strengthening at the left edge of the IP (Cho & Keating, 2001) was not present in the stimuli and the IP-level final lengthening was not present in the no-lengthening condition.

REFERENCES

Arvaniti, A. (**2009**). “Rhythm, timing and the timing of rhythm,” Phonetica **66,** 46–63.

Barnes, R., and Jones, M. R. (**2000**). “Expectancy, attention, and time,” Cognit. Psychol. **41,** 254–311.

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

Boersma, P., and Weenink, D. (**2014**). “Praat: Doing Phonetics by Computer, ver. 5.4.04”. Amsterdam, the Netherlands, http://www.praat.org/ (Last viewed 31 December, 2014).

Brown, M., Salverda, A. P., Dilley, L. C., and Tanenhaus, M. K. (**2011**). “Expectations from preceding prosody influence segmentation in online sentence processing,” Psychon. Bull. Rev., **18,** 1189–1196.

Cho, T., and Keating, P. A. (**2001**). “Articulatory and acoustic studies on domain-initial strengthening in Korean,” J. Phonet. **29,** 155–190.

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(4),** 523–547.

Cutler, A. (**2012**). *Native Listening: Language Experience and the Recognition of Spoken Words* (MIT Press, Cambridge, MA), Chap. 1 (pp. 1-30), sect. 3.3 (pp. 75-87) and Chap. 4 (pp. 117-153).

Davis, M. H., Marslen-Wilson, W. D., and Gaskell, M. (**2002**). "Leading up the lexical garden path: Segmentation and ambiguity in spoken word recognition," J. Exp. Psychol. Hum. Percept. Perform. **28,** 218–244.

Dilley, L. C., and McAuley, J. D. (**2008**). "Distal prosodic context affects word segmentation and lexical processing," J. Mem. Lang. **59,** 294–311.

Dilley, L. C., Mattys, S. L. & Vinke, L. (**2010**). "Potent prosody: Comparing the effects of distal prosody, proximal prosody, and semantic context on word segmentation," J. Mem. Lang. **63**, 274–294.

Fletcher, J. (**2010**). "The prosody of speech: Timing and rhythm," in the *Handbook of Phonetic Sciences*, edited by W. J. Hardcastle, J. Laver, and F. E. Gibbon, 2 ed. (Wiley-Blackwell, New York), pp. 523–602.

Fraisse, P. (**1984**). "Perception and estimation of time," Ann. Rev. Psychol. **35**, 1-36.

Hayes, B. (**1995**). *Metrical Stress Theory: Principles and Case Studies* (University of Chicago Press, Chicago, IL), pp. 24-85 and pp. 367-402.

Heitz, R. P. (**2014**). "The speed-accuracy tradeoff: history, physiology, methodology, and behaviour," Front. Neurosci. **8(150)**, 1–19.

Holliday, J. J., Turnbull, R. and Eychenne, J. (**2016**). "K-SPAN (Korean Surface Phones and Neighborhoods) ver. 1," UiT Open Research Data Dataverse, URL http://dx.doi.org/10.18710/TWM79F, last access on 21 Aug. 2016.

Hothorn, T., Bretz, F., and Westfall, P. (**2008**). "Simultaneous inference in general parametric models," Biometrical J., **50(3),** 346–363.

Jeon, H.-S. (**2015a**). "Prosody," in *Handbook of Korean Linguistics*, edited by L. Brown, and J. Yeon (Wiley-Blackwell, New York), pp. 41–58.

Jeon, H.-S. (**2015b**). Rhythm in Korean verse, sico, In the Proceedings of the 18th International Congress of Phonetic Sciences, 129, pp.1-5, Glasgow, UK, 10-14 August.

Jeon, H.-S., and Nolan, F. (**2010**). "Segmentation of the Accentual Phrase in Seoul Korean," in the Proceedings of Speech Prosody 2010, vol. 100023, pp. 1–4, Chicago, USA, 11-14 May.

Jones, M. R., Johnston, H. M., and Puente, J. (**2006**). "Effects of auditory pattern structure on anticipatory and reactive attending," Cognit. Psychol. **53,** 59–96.

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 & A. Arvaniti (Cambridge University Press, Cambridge, UK), pp. 235–253.

Jun, S.-A. (**1998**). "The Accentual Phrase in the Korean Prosodic Hierarchy," Phonology **15(2),** 189–226.

Jun, S.-A. (**2000**). "K-ToBI (Korean ToBI) labelling conventions, ver. 3.1," UCLA Working Papers in Phonetics **99**, 149–173.

Jun, S.-A. (**2005**). “Korean intonational phonology and prosodic transcription,” in *Prosodic Typology*, edited by S.-A. Jun (Oxford University Press, Oxford, UK), pp. 201–229.

Jun, S.-A. (**2014**). “Prosodic typology: by prominence type, word prosody, and macro-rhythm,” in *Prosodic Typology II: The Phonology of Intonation and Phrasing*, edited by S.-A. Jun (Oxford University Press, UK), pp. 520–540.

Kim, S. (**2004**). *The Role of Prosodic Phrasing in Korean Word Segmentation*, Unpublished PhD thesis, UCLA.

Kim, S., Broersma, M., and Cho, T. (**2012**). “The use of prosodic cues in learning new words in an unfamiliar language,” Stud. Second Lang. Acquisit. **34,** 415–444.

Kim, S. and Cho, T. (**2009**). “The use of phrase-level prosodic information in lexical segmentation: Evidence from word-spotting experiments in Korean,” J. Acoust. Soc. Am. **125(5),** 3373–3386.

Luce, P. A., and Pisoni, D. B. (**1998**). “Recognizing spoken words: The neighbourhood activation model,” Ear. Hearing **19,** 1–36.

Mattys, S., White, L., and Melhorn, J. (**2005**). “Integration of multiple speech segmentation cues: A hierarchical framework,” J. Exp. Psychol. Gen. **134,** 477–500.

McQueen, J. (**1996**). “Word spotting,” Lang. Cognit. Proc. **11(6),** 695–699.

Morrill, T. H., Dilley, L. C., and McAuley, J. D. (**2014**). “Prosodic patterning in distal speech context: Effects of list intonation and f0 downtrend on perception of proximal prosodic structure,” J. Phonet. **46,** 68–85.

National Institute of Korean Language (NIKL) (**2005**), Frequency Research Report in Contemporary Korean 2 (in Korean), URL: www.korean.go.kr. Last access on 26 Aug. 2016.

Nolan, F., and Jeon, H.-S. (**2014**). "Speech rhythm: a metaphor?," Phil. Trans. R. Soc. B., 369, 20130396: 1-11.

Oh, M. (**1998**). "The prosodic analysis of intervocalic tense consonant lengthening in Korean," Japanese/Korean Linguistics **8,** 317–330.

Quené, H., and Port, R. F. (**2005**). "Effects of timing regularity and metrical expectancy on spoken-word perception," Phonetica **62,** 1–13.

Ratcliff, R. (**1993**). "Methods for dealing with reaction time outliers," Psychol. Bull., **114**, 510-032.

R Core Team (**2015**). "R: A language and environment for statistical computing," R Foundation for Statistical Computing, Vienna, Austria. URL: http://www.R-project.org/, last view date 10 Dec. 2015.

Salverda, A., 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.

Shattuck-Hufnagel, S., and Turk, A. E. (**1996**). "A prosody tutorial for investigators of auditory sentence processing," J. Psycholinguist. Res. **25,** 193–247.

Shin, J. (**2015**). "Vowels and consonants," in *Handbook of Korean Linguistics*, edited by L. Brown, and J. Yeon (Wiley-Blackwell, New York), pp. 3–21.

Shukla, M., Nespor, M., and Mehler, J. (**2007**). "An interaction between prosody and statistics in the segmentation of fluent speech," Cognit. Psychol. **54(1),** 1–32.

Spinelli, E., Grimault, N., Meunier, F., and Welby, P. (**2010**). “An intonational cue to word segmentation in phonemically identical sequences,” Atten. Percept. Psychophys. **72(3)**, 775–787.

Tilsen, S. & Arvaniti, A. (**2013**). “Speech rhythm analysis with decomposition of the amplitude envelope: Characterizing rhythmic patterns within and across languages,” J. Acoust. Soc. Am. **134,** 628–639.

Tyler, M. D., and Cutler, A. (**2009**). “Cross-language differences in cue use for speech segmentation,” J. Acoust. Soc Am. **126(1)**, 367–376.

Warner, N., Otake, T., and Arai, T. (**2010**). “Intonational structure as a word-boundary cue in Tokyo Japanese,” Lang. Speech **53(1),** 107–131.

Welby, P. (**2007**). “The role of early fundamental frequency rises and elbows in French word segmentation,” Speech Commun. **49,** 28–48.

White, L. (**2014**). “Communicative function and prosodic form in speech timing,” Speech Commun. **63–64,** 38–54.