The role of post-lexical tonal contours in word segmentation

Sahyang Kim

University of California, Los Angeles

E-mail: sahyang@ucla.edu

ABSTRACT

The current study examines the role of non-lexical tonal patterns of a prosodic phrase in word segmentation. Seventy-two native Seoul Korean speakers participated in the word spotting experiments. The results showed that listeners spotted more words in phrase initial than in phrase medial position, and that their error rates were significantly lower in rising tonal patterns than in non-rising patterns, regardless of the position of the target words within an AP, and regardless of the frequency and legality of AP tonal patterns. Tonal cues other than pitch rising did not influence the error rate. These results indicate that rising pitch in AP initial position is a reliable cue for word boundary detection for Korean listeners. These findings further suggest that phrasal tone contours serve as a possible word boundary cue in languages without lexical prominence.

1.   INTRODUCTION

To process spoken language, listeners must be able to divide the stream of speech into smaller discrete units. In order to do this job successfully, they must be capable of detecting the correct boundaries for each unit. As noted in many previous studies, accurate segmentation can be aided by multiple means, such as phonotactic cues [1], [2], transitional probabilities [3], [4], and rhythmic cues [5], [6], [7]. The current study concentrates on the role of prosody in word segmentation.

There has been extensive cross-linguistic research into the effect of prosody on prelexical segmentation. It has been claimed that listeners exploit a 'metrical segmentation strategy' for initial segmentation [6], [8], and that the strategy that listeners employ is language specific [7], [9], [10], [11]. Previous studies on word segmentation have focused, however, on languages with lexical prosodic information, such as stress or pitch accent. This study aims to look for a potential prosodic segmentation cue in Korean, a language without any lexically specified prosody.

The prosodic unit that may help segmentation process of Korean, if any, would be the Accentual Phrase (henceforth AP). The AP is the most basic tonally defined prosodic unit in Korean prosodic hierarchy [12], and it is dominated by an Intonational Phrase (henceforth IP). The AP could potentially provide strong cues to aid in word segmentation, since it is marked by distinctive tonal patterns, and it usually contains no more than two content words in it. The underlying tonal pattern of an AP is THLH. The initial tone (T) is realized by H when AP initial segment is aspirated or tensed, and by L elsewhere. Although THLH is the most common pattern of the AP, other intonational contours may also occur. In general, an AP starts with a rising pitch, and the H in the rising pattern is aligned to the second syllable of an AP. In normal speech rate, AP contains one or two content words and a case marker or a postposition. It is reported that Korean APs contain 1.2 content words on average [13]. AP-final syllables do not usually show phrase final lengthening, and there is no pause between AP boundaries [12].

The purpose of the present study is to examine whether the prosodic properties of a post-lexical intonational unit can aid segmentation in the same way that lexical prosodic units do, by investigating the effect of various tonal contours of AP in word segmentation.

2.   EXPERIMENT

Materials

A word spotting task was employed for the current study. The target words were sixteen disyllabic (CV.CV) and sixteen trisyllabic nouns (CV.CV.CV). The frequency and familiarity of target words were controlled. Each target word was inserted into a speech stream that was composed of three APs of four syllables. Speech streams were produced as a single Intonational Phrase, and there were no audible pauses in AP boundaries. The syllables in a speech stream were ordered in such a way that no consecutive syllables, other than the target word, formed a word in Korean. In the target-bearing streams, the target words were inserted in the second AP of the streams with varying tonal patterns. The first and third AP of the target-bearing stream had LHLH tone contour. In addition to 32 target words, 16 disyllabic and 16 trisyllabic filler words were included in the experiment. The filler words were real Korean nouns as well. In order to increase the variety of contexts and to prevent the listeners from predicting the possible locations of the target words, the filler-bearing streams contained a filler word either in the first or the third AP, and the APs that did not contain the filler word had various tonal patterns other than LHLH.

There were three tonal patterns for AP initial position and

medial position, respectively, as shown in Table 1.

AP initial	AP medial
L H L H	L *H L* H
L H H H	L *L R* H [1]
L L L H	L *L L* H

Table 1. Tonal patterns and location of real words in each condition. Bold and italicized characters indicate the location of disyllabic words and underlines indicate that of trisyllabic words.

The tonal patterns used for the experiment were LHLH, LHHH, LLLH, and LLRH. LLRH was an illegal AP tone pattern in Korean, whereas the others were legal. LHLH was employed in both AP initial and medial positions since it is the most common AP pattern of Korean. LHHH was selected to see the effect of an H tone after the uniqueness point, in contrast to the L tone in LHLH. LLLH was chosen to compare the effect of initial rising against that of non-rising. Finally, the illegal pattern LLRH was employed only in the AP medial position, in order to compare the effect of rising tone in different contexts.

Every target and filler word was embedded in six different tonal pattern contexts, shown in Table 1. The experiment thus had six lists, arranged such that each listener heard every word just once, in one of the six tonal contours. Each list contained 32 target strings and 32 filler strings in a pseudo-random order, and no two stimuli with the same tonal pattern were presented in a row. Half of the participants were given the lists in one pseudo-random order, and the rest of the participants were given the lists in reverse order.

The non-sense speech streams for the experimental stimuli were recorded onto Digital Audio Tape in a sound-attenuated booth by a female native speaker of Seoul Korean (the author), and digitized at a sampling rate of 22kHz. The duration of the each speech stream and tonal pattern were modified with Praat software, when necessary.

Participants and Procedure

Seventy-two native speakers of Seoul Korean, born and raised in Korea and studying at UCLA or Stanford University at the time of the study, participated in the experiment. They were paid for their participation. Participants were tested individually in a sound-attenuated booth. The PsyScope software package and a CMU button box were used for stimulus presentation and RT recording. They were informed that the words that they had to spot were disyllabic or trisyllabic nouns of Korean. They were asked to say the word out loud as soon as they spotted a word

[1] R indicates rising on the third syllable.

from the given speech stream. Reaction time (henceforth RT) measure was activated by listener's voice, and the oral responses were recorded into a Macintosh computer using the program SoundEdit. Missing or incorrect responses were reported as errors. Every listener heard eight practice items before she/he was exposed to one of the six main experimental lists.

3. RESULTS

The recorded RTs were adjusted to measure from target word offset to the beginning of listeners' utterance. RTs outside the range of 4000 ms were treated as errors. The missing rate was 57.6%, the rate of incorrect response was 7.5%, and the rate of RTs over 4000 ms was 1.8%. Thus, the overall error rate was 66.9%. Because of this very high error rate (indicating a lot of missing data for the RT analyses), error rate was taken as the primary dependent measure for the experiment.

The results showed that there was an effect of word length on error rates. Listeners missed more disyllabic words than trisyllabic words. An ANOVA performed on error rate revealed significant differences between the two syllabic categories ($F_{(1, 2302)} = 185.655$, $p < .0001$). An effect of AP location was also found. The result of the ANOVA showed that the words located in AP initial position were easier to spot than those in AP medial position ($F_{(1, 2302)} = 742.888$, $p < .0001$). Listeners were significantly better in spotting words in each of the intial three tonal contours compared to the medial three tonal contours.

Subsequent ANOVAs were performed separately for the disyllabic initial and medial targets, and for the trisyllabic initial and medial targets. Mean error rate for the separate conditions are shown in Figure 1 and Figure 2.
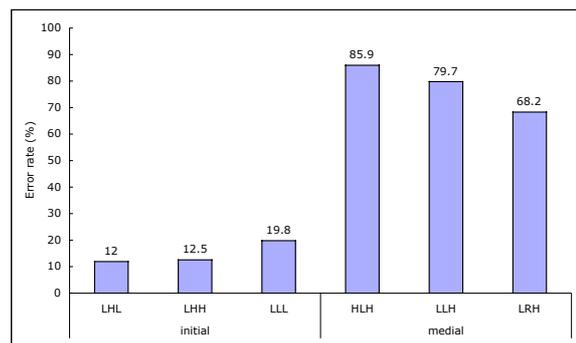

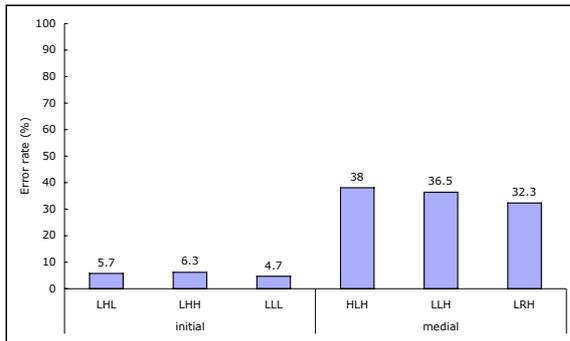*Figure 1. Error rates of disyllabic words for six different tonal patterns*

Figure 2. Error rates of trisyllabic words for six different tonal patterns

In the disyllabic initial condition, the tone effect was marginally significant (F (2, 573) = 2.927, *p* = .05). However, *post-hoc* analysis showed that the error rates in LHH and LHL were significantly lower than that of LLL (*p* < .05, in both cases). There was no statistically significant difference between LHH and LHL. In the dysillabic medial condition, the tone effect was significant (F (2, 573) = 9.253, *p* < .0001). The error rates of HLH and LLH were significantly higher than that of LRH (*p* < .0001, *p* < .01, respectively). Although HLH showed higher error rates than LLH, the difference between the two was not significant. There was no significant tone pattern effect in trisyllabic words, regardless of their location within AP.

An ANOVA performed on mean RT showed that there were significant effects of syllable count as well as location of word in AP. RT was significantly faster for trisyllabic words than disyllabic words (F (1,1532) = 67.889, *p* < .0001), and for initial position than medial position (F (1, 1532) = 116.895, *p* < .0001). RTs were significantly faster in each of the three initial tonal contours compared to the three medial tonal contours in both disyllabic and trisyllabic words.

	AP initial			AP medial		
	LHL	LHH	LLL	HLH	LLH	LRH
2syll	817.7	759.1	872.5	1468.6	1331.4	1382.2
3syll	546.6	585.3	583.2	818.3	713.7	742.5

Table 2. Mean RTs (ms) in different tonal conditions.

As expected, RT did not reveal as much effect of tone patterns as error rates. ANOVAs performed separately depending on the number of syllable and location of AP did not reveal any significant effect on RT. A *post-hoc* analysis showed that listeners were faster in spotting words in LHL than LLL (*p* < .05) with regard to disyllabic words in initial position. No other significant difference among tone patterns was found.

4. SUMMARY AND DISCUSSION

The results revealed a strong effect of number of syllables in target words. Error rates were significantly lower and word spotting was faster in trisyllabic target words than in disyllabic target words. Furthermore, no tone contour effect was found in detecting trisyllabic target words, either in error rate or in RT measures. This suggests that it was relatively easy for the listeners to spot trisyllabic words without relying on tonal patterns, and that trisyllabic target words showed a ceiling effect.

This result seems to be an artifact produced by the initial target word selection procedure for the experiment and by the nature of Korean lexicon itself. In order to perform a successful word spotting task, one should not select a word which embeds another word as a target. However, most Korean trisyllabic words within a certain frequency range happened to contain a string of syllables which sounded exactly like existing disyllabic words, such as *seori* 'frost' in *moseori* 'angle' and *tari* 'leg' in *tarimi* 'iron'. The experimental design was faithful to the 'no word-within-word' constraint despite the difficulty, and as a result, most of the selected trisyllabic target words turned out to be quite distinctive by the time the second syllable was presented. In other words, the trisyllabic target words had early uniqueness points and did not have many neighbors, and this seems to be the reason why the trisyllabic words showed a ceiling effect. Listeners were able to identify the trisyllabic target words correctly and were ready to respond even before they heard the whole sound string of the words. The results imply that post-lexical tone pattern would be less crucial in word recognition than an instrinsic lexicl property as early uniqueness point.

The results also showed that there was a positional effect in segmentation. Listeners spotted words more reliably in AP initial position than in AP medial position, and RT was also faster in AP initial than in medial position. This effect held not only when the average values of the two conditions were compared, but also when each individual tonal pattern in the conditions were compared. Listeners were significantly better when the words with the rising pitch patterns were AP initial (*LH*LH, *LH*HH) than when they were AP medial (L*LR*H). The same effect was also observed when the words occurred with non-rising pitch patterns. Listeners were better at detecting words with *LL*LH tone in initial than with L*LL*H in medial position. Further, the results showed that non-favorable tonal pattern in initial position (*LL*LH) was significantly more helpful to the listeners than favorable tonal pattern in medial position (L*LR*H). These results indicate that there was a general AP initial advantage across the tonal contours.

It is well known that segmental phonetic properties vary depending on prosodic context. Studies on speech production of various languages, including on Korean [14], have shown that the segments were stronger in terms of articulatory and acoustic magnitude in domain initial position

than in non-initial position. Although it has been hypothesized that listeners would get some perceptual advantage from this effect, there has been little direct evidence in support of this. For this experiment, the target words were produced as naturally as possible in the given prosodic contexts. Thus, it is likely that the segments were produced with more strength in the AP initial than in the medial positions, and this affected listeners' performance. The current result is a piece of evidence revealing that listeners actively make use of the subtle phonetic cues existing in domain initial positions when they process spoken language. The phonetic cues at prosodic edges make it easier for listeners to find the boundaries for speech units, and hence help the segmentation process of spoken input.

The overall results on the individual tonal patterns indicate that listeners use cues from various tonal patterns in a selective way. The error rate and RT of initial *LH*L and *LH*H patterns were not different from each other. It suggests that the falling to the low tone after the target word does not necessarily aid segmentation by marking discontinuity from a given word. It also implies that once the word is recognized in one way or another, listeners do not pay attention to the following tonal patterns. The error rate of medial L*HL*H was higher than that of medial L*LL*H, which indicates that negative tonal cue (falling, as opposed to rising) affect listeners' segmentation process in a destructive way. It further implies that the tonal contours can have both a facilitative and a disruptive influence on segmentation.

In disyllabic target words, the words with rising tonal contours (LH / LR) were spotted more reliably and faster than the ones with no pitch rising (LL), both in AP initial and medial positions. Although rising contours in the middle of APs do not occur in Seoul Korean, the unacceptability of the given tonal pattern did not interfere listeners' segmentation process — and in fact, it aided the process. If the listeners had relied more on the familiarity and acceptability of the tonal patterns, they should have been better with L*LL*H or L*HL*H tone than L*LR*H. Therefore, the result suggests that rising pitch in word initial position is a reliable cue for word segmentation for Seoul Korean listeners.

To summarize, the results of this study suggest that word segmentation is aided by prosodic information represented in subtle phonetic differences of spoken language, that a post-lexical tonal pattern marking a phrasal prosodic unit can help word segmentation just as well as lexical prosody can, and that rising AP-initial pitch is a reliable cue for word boundary detection for Korean listeners.

ACKNOWLEDGEMENTS

REFERENCES

[1] J.M. McQueen, "Segmentation of Continuous Speech Using Phonotactics," *Journal of Memory and Language*, vol. 39, pp.21-46, 1998.

[2] P.W. Jusczyk, P.A. Luce and J. Charles-Luce, "Infants' sensitivity to phonotactic patterns in the native language," *Journal of Memory and Language*, vol. 33, pp.630-645, 1994.

[3] J.R. Saffran, E.L. Newport and R.N. Aslin, "Word segmentation: The role of distributional cues," *Journal of Memory and Language*, vol. 35, pp.606-621, 1996.

[4] E. Johnson and P.W. Jusczyk, "Word segmentation by 8-month-olds: when speech cues count more than statistics," *Joural of Memory and Languages*, vol.44, pp.548-567, 2001.

[5] A. Cutler, J. Mehler, D. Norris, and J. Segui, "The syllable's differing role in the segmentation of French and English," *Journal of Memory and Language*, vol.25, pp.385-400, 1986.

[6] A. Cutler and D. Norris, " The role of strong syllables in segmentation for lexical access," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 14, pp.113-121, 1988.

[7] A. Cutler and T. Otake, "Mora or phonemes? Further evidence for language-specific listening," *Journal of Memory and Language*, vol. 33, pp.824-844, 1994.

[8] A. Cutler, "Exploiting prosodic probabilities in speech segmentation," in *Cognitive Models of Speech Processing*, G.T.M. Altmann, Ed., Cambridge MA: MIT Press, 1990.

[9] N. Sebastian-Galles, E. Dupoux, J. Segui and J. Mehler, "Contrasting syllabic effects in Catalan and Spanish," *Journal of Memory and Language*, vol. 31, pp.18-32, 1992.

[10] J. Vroomen, M. van Zon and B. de Gelder, "Cues to speech segmentation: evidence from juncture misperceptions and word spotting," *Memory and Cognition*, vol. 24, pp.744-755, 1996.

[11] T. Otake, G. Hatano, A. Cutler and J. Mehler, "Mora or syllable? Speech monitoring and lexical access," *British Journal of Psychology*, vol. 72, pp.471-477, 1993.

[12] S.-A. Jun, "The Accentual Phrase in the Korean prosodic hierarchy," *Phonology* 15, pp.189-226, 1998.

[13] S.-A. Jun and C. Fougeron, "A Phonological model of French intonation," *Intonation: Analysis, Modeling and Technology*, A. Boinis, Ed., pp. 209-242, Dordrecht: Kluwer Academic Publishers, 2000.

[14] T. Cho and P. Keating, "Articulatory and acoustic studies on domain–initial strengthening in Korean," *Journal of Phonetics*, vol. 29, pp. 155-190, 2001.