The Role of Post-lexical Intonational Patterns in Korean Word Segmentation

Sahyang Kim*

ABSTRACT

The current study examines the role of post-lexical tonal patterns of a prosodic phrase in word segmentation. In a word spotting experiment, native Korean listeners were asked to spot a disyllabic or trisyllabic word from twelve syllable speech stream that was composed of three Accentual Phrases (AP). Words occurred with various post-lexical intonation patterns. The results showed that listeners spotted more words in phrase-initial than in phrase-medial position, suggesting that the AP-final H tone from the preceding AP helped listeners to segment the phrase-initial word in the target AP. Results also showed that listeners' error rates were significantly lower when words occurred with initial rising tonal pattern, which is the most frequent intonational pattern imposed upon multisyllabic words in Korean, than with non-rising patterns. This result was observed both in AP-initial and in AP-medial positions, regardless of the frequency and legality of overall AP tonal patterns. Tonal cues other than initial rising tone did not positively influence the error rate. These results not only indicate that rising tone in AP-initial and AP_final position is a reliable cue for word boundary detection for Korean listeners, but further suggest that phrasal intonation contours serve as a possible word boundary cue in languages without lexical prominence.

Keywords: word segmentation, prosody, post-lexical intonation, Korean

I. Introduction

1.1 Prosody and Word Segmentation

Early research on the role of prosody in word segmentation focused on lexical level prosody, especially lexical stress. Cutler and Norris (1988), using a word spotting experiment, showed that native listeners of English were faster in detecting a monosyllable real word (e.g., *mint*) in a disyllable speech string with strong-weak stress pattern (e.g., [mɪnt↔f]) than that with strong-strong stress pattern (e.g., [mɪntef]). English listeners' exploitation of this stress pattern cue for word segmentation seems to be attributed to its input frequency. Cutler and Carter (1987) found that 90% of open-class English words (from a dictionary with 33,000 entries) begin with strong syllables (including secondary as well as primary stress). Another language that reveals a similar metrical segmentation pattern as English is Dutch. As for input frequency, 90% of Dutch words start with a

* Department of English Education, Hong-Ik University

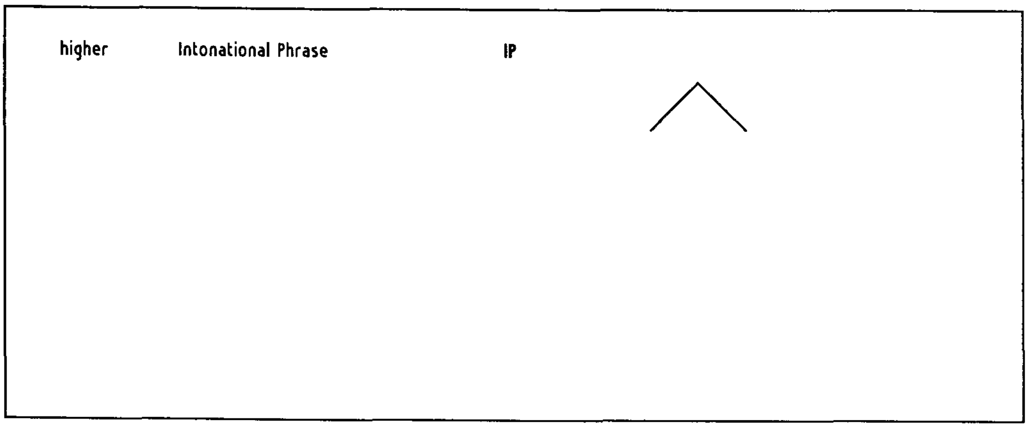
stressed syllable (Quene, 1992). Dutch listeners also showed the same processing pattern as English listeners (Vroomen & de Gelder, 1995). Overall, these studies suggest that listeners' segmentation decision largely depends on a language-specific characteristic of the prosody (in this case, frequent lexical stress pattern) of a given language.

Recently, studies have turned their attention to the influence of higher-level prosodic unit on word recognition and segmentation. Christophe and her colleagues (Christophe, Gout, Peperkamp & Morgan, 2003; Christophe, Peperkamp, Pallier, Block & Mehler, 2004; Gout, Christophe & Morgan, 2004) investigated whether prosodic boundary affects on-line lexical search in adults and infants' speech processing. They found that listeners were confused by local ambiguity when there was no Phonological Phrase boundary in a segmental context with potential ambiguity, but that they were *not* confused by local ambiguity when there was a Phonological Phrase boundary in the same context. The results suggest that the ability to detect prosodic structure may ease listeners' task in lexical segmentation, by helping listeners to reduce the number of segmentation hypotheses, i.e., by limiting lexical search only within a prosodic phrase.

Then, questions arise as to which cues make the detection of prosodic boundary possible and directly affect speech segmentation and processing. Research on speech production has revealed that prosodic structure can be manifested by various acoustic correlates in phrase edges, such as phrase final lengthening (Klatt, 1975; Cooper & Paccia-Cooper, 1980; Beckman & Edwards, 1990; Edwards, Beckman & Fletcher, 1991; Wightman, Shattuck-Hufnagel, Ostendorf & Price, 1992, *inter alia*), phrase boundary tone (Bolinger, 1970; Pierrehumbert, 1980, *inter alia*), and domain initial strengthening (Pierrehumbert & Talkin, 1991; Jun, 1996; Fougeron & Keating, 1997 ; Cho & Keating, 2001; Cho 2001, *inter alia*), in many languages. Which of these prosodic cues helps listeners' lexical segmentation? Do they differ in their degree of contribution to the segmentation process or not? Is the exploitation of specific prosodic cues in speech segmentation language-specific or universal? One way to approach these issues would be to test the effect of individual phrasal cues in different languages and make cross-linguistic comparisons. The current study attempts to provide partial answers to these questions by investigating whether post-lexical prosody, especially intonational patterns, can directly influence on-line word segmentation, and whether the frequency of intonational patterns affects lexical segmentation as the frequency of a stress pattern does. The language of particular interest in this study is Korean. Not having any lexically specified prosody, Korean allows us to observe the pure effect of post-lexical intonation in lexical segmentation without any complication from the interaction between lexical level and post-lexical level prosody. Before presenting specific research questions, I will provide a short description on Korean prosody first in the following section.

1.2 Korean Intonation

Intonational pattern of Seoul Korean (henceforth Korean) is mainly determined by tonal shape of the Accentual Phrase (henceforth AP), according to an intonation-based prosodic model of Korean (Jun 1996, 2000) shown in.



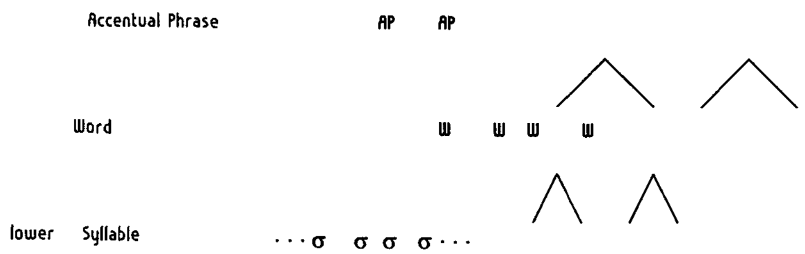


Figure 1. Intonation-based Korean Prosody (adopted from Jun 2000)

In this model, the AP is the unit that dominates one or more prosodic word and is dominated by an Intonational Phrase. 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. Other than this underlying pattern, at least 14 different tonal patterns have been attested so far. Half of the fourteen attested AP tonal patterns starts with a L tone (LHLH, LHHH, LLLH, LH, LHL, LHLL, LL), and the other half starts with a H tone (HHLH, HH, HLH, HL, HHL, HLL, HHLL). Jun (2000)'s model views that 'phonetic' variants of 'phonological' THLH pattern are derived from overshoot or undershoot of individual tone. Some apparent factors that contribute to tonal variation are the number of syllables within an AP and speech rate. Thus far, there has been no evidence showing that these phonetic variants represent contrastive meanings, or that the occurrence of certain patterns is predictable. Unlike IP boundaries where audible pause and final lengthening occur, there is no audible pause between AP boundaries. AP-final lengthening is still controversial (Jun, 1996) did not find AP-final lengthening, while Cho and Keating (2001) observed small but significant lengthening at the same position).

Although there are numerous varieties, not all the AP tonal patterns are equally frequent. Kim (2004)'s transcription of Korean intonational patterns in read speech and radio drama showed that about 85% of AP's end with the final H tone, and that about 88% of AP-initial *multisyllabic content words start with a rising (LH) tone when words' onset is not a tense or aspirated* consonant. The study also revealed that most AP's contained 3-4 syllables, that the observed AP's did not exceed 7 syllables, and that AP's contain 1.13 content words in average, which did not differ much from 1.2 obtained in another study based on intonational transcription of read speech (Jun & Fougeron, 2000).

1.3 Research Questions

Given that an AP contains about one content word in average, and that AP-final lengthening is not too robust, it is very likely that intonational patterns by themselves can play a role as one of the most effective cues for lexical segmentation in Korean. Therefore, the primary hypothesis of this study is that Korean listeners would take advantage of AP tone patterns in order to detect phrase boundary, and at the same time, segment words. An on-line word spotting task was performed to examine which of the possible Korean intonational patterns are directly relevant to lexical segmentation. For this purpose, target words occurred with various intonational patterns in carrier speech stream.

First, in order to investigate whether the phrase-final tone of AP would affect word segmentation or not, target words were inserted in both AP-initial and AP-medial positions. It was hypothesized that AP-initial words would be much easier to segment than AP-medial words. When a Korean word is at the beginning of an AP, but not at the beginning of an IP at the same time, the word will usually follow the phrase-final H tone of the preceding AP. The final H tone of an AP usually marks the end of a syntactic phrase and hence, that of a word. Thus, it is likely that the final H tone would simultaneously indicate the upcoming onset of another word.

Phrase-final H tone is known to facilitate native French (Bagou, Fougeron & Frauenfelder, 2002) and Korean listeners' (Kim, 2004) word segmentation ability in the context of artificial language learning. However, there has been no evidence showing that French or Korean listeners could also utilize this cue in on-line word segmentation. If this cue is used for on-line word segmentation task, Korean listeners' detection of target words would be faster when they are in AP-initial position than in AP-medial position.

Second, the current study tested whether AP-initial rising intonation can positively affect word segmentation. The rising tone is the most frequent tonal pattern at AP-initial position, and words would most frequently co-occur with a rising tone pattern when they appear in AP-initial position (Kim, 2004). If listeners were sensitive to this co-occurrence frequency, they could make use of this information for segmentation. It was hypothesized that listeners would detect words faster when the words occur with rising tone (e.g., LH) than with non-rising tone (e.g., LL).

In off-line perception experiments, Welby (2003) found that French listeners' perception of word boundary depended on the presence of phrase-initial rising intonation, and that French listeners relate an early rise to the content word boundary. At least on the surface, Korean intonation and French intonation look similar. They both have phrasal-level prominence without lexical-level prominence such as lexical stress, and the smallest prosodic phrase (the Accentual Phrase, in Jun's Korean model (1996, 2000), and in Jun & Fougeron's French model (2000)) also has a similar default tone pattern, with two consecutive rising tones (LHLH). However, the two languages differ in terms of the location of function/content words and tonal alignment. In French, an early rise is optional, and the L tone in the initial rise usually falls on a function word. Given that it is a head-initial language, the H tone following the initial L tone usually falls on the onset of a content word, when it is realized. Welby (2003)'s experiments show that French listeners are sensitive to this alignment of early rise with the onset of content words. However, since Korean is a head-final language, a prosodic phrase (AP) usually starts with a content word and ends with a function word (e.g., Noun-Nominative marker). H tone in the AP-initial rise does not cue the beginning of a content word in Korean. Rather, AP-initial multisyllabic content words would have H tone (either the peak or interpolation) in word-medial position when words are composed of three or more syllables, or in word_final position when words are disyllabic.

One assumption that can be drawn by the results of previous research is that there is a general tendency that word segmentation is influenced by perceptual prominence present at the onset of a content word (as in the cases of the trochaic stress pattern in English and Dutch and a H tone of early rise in French). However, if Korean listeners do use a rising intonation that does not have a H tone fall at the onset of a content word, it would support the idea that lexical segmentation is affected by language-specific intonational pattern, even if it does not necessarily give a perceptual benefit at the onset of a content word.

2. Experiment

2.1 Word spotting task

The word spotting task has been extensively used in psycholinguistic studies on speech segmentation and word recognition since Cutler and Norris (1988)'s initial work. The procedure for the current experiment followed similar steps to that for previous studies, but with some variations in stimuli and procedure. Unlike previous studies where listeners were typically asked to detect a monosyllabic real word (e.g., *mint*) from a disyllabic speech stream (e.g., *mintef*), the present study extended the number of syllables in the speech stream as well as in the embedded target words. The listeners who participated in this experiment were exposed to speech streams that consisted of twelve syllables and contained disyllabic or trisyllabic target words. In previous studies, listeners were asked to press a button as soon as they hear the word embedded in the string, and then to say the word aloud. The reaction time was measured by button pressing. In the current experiment, however, reaction time was measured by voice activation without button press, in order to avoid potential reaction time delay between button press and spell-out of target words.

2.2 Stimuli

2.2.1 Word Selection

Sixteen disyllabic nouns and sixteen trisyllabic nouns were selected as target words. In addition, four factors were controlled in the process of target word selection.

First, words that contained only CV syllables were selected as targets. CVC syllables were excluded since the final consonant of a CVC syllable might affect the segmental property of a following consonant, which, in turn, might influence the recognition of target words. Thus, the disyllabic target words had a CV.CV structure and the trisyllabic target words had a CV.CV.CV structure.

Second, word-initial segments were controlled, because it has been observed that in Korean, aspirated and tense consonants trigger a high tone in AP-initial position, while other consonants maintain a low tone in the same position (Jun, 1996, 2000). Eight out of sixteen disyllabic target words and five out of sixteen trisyllabic target words contained either a nasal stop (/m/, /n/) or a glide (/j/) as the word-initial consonant, and the rest of the targets contained a lenis obstruent (/t/, /k/, /p/, /t∑/) as the word initial segment.

Third, the frequency of the target words was controlled based on two frequency databases: Frequency Analysis of Korean Morpheme and Word Usage (Kim & Kang, 2000) and KAIST Concordance Program (KCP) Online Demo Version (KAIST, 1999). Disyllabic target words were in the high frequency range, but trisyllabic target words were in the low frequency range. The frequency range for trisyllabic target words was set lower since most of the high frequency trisyllabic words had another word embedded in them, which was not desirable for the current experimental task.

The last criterion considered was familiarity. As there was no word familiarity database for Korean, I conducted a familiarity survey with 100 words, including seventy target word candidates for this experiment. Forty native Korean speakers whose age range was similar to that of potential experiment participants (20 to 35-year-olds) took part in the survey. None of the survey respondents participated in the current experiment. The respondents were asked to grade each word based on the scale from 1 (not familiar at all) to 5 (very familiar). The scores were averaged, and words that had at least an average score of 3 (familiar) were selected as targets.

There was the same number of filler words as target words for each syllabic category (i.e., 16 disyllabic filler words and 16 trisyllabic filler words). The filler words were not controlled as strictly as the targets. The word selection criteria and target word list are summarized in Appendix 1, and detailed frequency and familiarity rating information about the target words is given in Appendix 2.

2.2.2 Locations of Words within a Carrier Speech Stream

A nonsense carrier speech stream was composed of three APs, with each AP containing four syllables; hence, a speech stream had twelve syllables. No consecutive syllables, other than the target word, formed a word in Korean. All the syllables in the carrier speech streams had CV structures only. In the target-bearing streams, each target word was inserted into the second AP of a carrier speech stream. The first and the third APs of the target-bearing stream had the default AP tone contour, LHLH. The filler-bearing streams contained a filler word either in the first or the third AP. The APs that did not contain the filler word had various tonal patterns other than LHLH.

There were two locations of the target/filler words within an AP: an AP-initial condition and an AP-medial condition. For the AP-initial condition, the target/filler words started at the beginning of an AP, and thus aligned with AP boundary. In the case of the AP-medial condition, the target/filler words started at the second syllable of an AP.

2.3 Tonal Contrasts and Hypotheses

Among the seven attested surface AP tonal patterns beginning with a low tone (LHLH, LHHH, LLLH, LH, LHL, LHLL, LL), the current experiment employed three tonal patterns, LHLH, LHHH, and LLLH. Since the APs in the experiment were composed of four syllables, with each syllable being mapped to one tone, the AP tonal patterns that did not have four tone targets (LH, LHL, LL) were excluded. LHLL was not included, because AP-final L tone could be confusable with IP-boundary tone and because it is infrequent (Kim, 2004). In addition to the three attested patterns, the experiment included one unattested tonal condition, LLAH. A indicates the rising interpolation from the L on the second syllable to the H on the fourth syllable. LLAH is different from LLLH, and not a variant of LLLH, in that the third syllable of LLAH is higher than the preceding L, whereas the third syllable of LLLH is level to the preceding L. As will be explained below, this pattern was created in order to see the effect of rising intonation in non-AP-initial position.

Both AP-initial and AP-medial conditions included three out of the four tonal patterns selected for the study. For the AP-initial condition, the tonal patterns LHLH, LHHH, and LLLH were used. In the AP-medial condition, only two of these patterns (LHLH and LLLH) were employed. The AP-medial condition included the tonal pattern LLAH, instead of the LHHH pattern. Thus, each target word had six conditions based on the location within the AP and the tonal pattern. Appendix 3 shows the tonal contrasts and location of a target word in a target bearing stream.

First, AP-initial and AP-medial positions were compared in order to see the effect of the presence of AP-final H tone preceding a target word. It was hypothesized that listeners would be faster in detecting words in AP-initial position than those in AP-medial position, because of the beneficial role of the AP-final H tone that precedes AP-initial target words. Since the target words were inserted in the second AP of the speech stream, this cue was always present for the target words in AP-initial position, but absent for those in AP-medial position.

Second, rising tonal patterns were distinguished from 'non-rising' patterns in both AP-initial and AP-medial positions. It was hypothesized that the rising tone accompanying word onsets could be used as a reliable cue for word segmentation compared to a non-rising tonal pattern. In order to test this hypothesis, the rising (LH) tonal pattern in *LH*LH and *LH*HH was contrasted with the non-rising (LL) tonal pattern in *LL*LH in AP-initial position. Further, the rising (LA) pattern in L*LA*H was contrasted with the non-rising (HL, LL) patterns in L*HL*H and L*LL*H in AP-medial position. Unlike the other patterns employed for the experiment, L*LA*H is an unattested tone pattern in Korean. However, this pattern was adopted in order to observe the effect of a rising tone on word segmentation in a different context, namely, AP-medial position.

Third, *LH*LH and *LH*HH were contrasted within the AP-initial condition. The distinction was made so as to see the role of tonal pattern frequency. As aforementioned, *LH*LH is a default AP tonal pattern, and *LH*HH is less frequent than *LH*LH. Thus, if the overall frequency of AP tonal patterns affects the segmentation, the target words in *LH*LH would be detected faster than those in *LH*HH.

Finally, L*HL*H was contrasted with L*LL*H for the AP-medial condition. These are attested AP patterns, and the target words with these AP tonal patterns bear either a falling tone (HL) or a level low tone (LL). It was anticipated that listeners' performance would be better when a word bears a level low tone (LL) than when it has a falling tone (HL), because the former simply entails the absence of a facilitative cue whereas the latter carries the tone pattern which is the exact opposite to the facilitative cue, i.e., a rising tone.

2.4 Recording and Manipulation of Stimuli

In order to avoid any potential effect from semantic and syntactic context, the nonsense speech streams were used for the experimental stimuli. The 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. Speech streams were produced as a single Intonational Phrase, and there were no audible pauses at AP boundaries.

Each speech stream was normalized to the average value of duration (Mean = 2.504 *sec*, S.D = 0.004 *sec*) and the tonal patterns were also modified when necessary in order to make sure that each stream had the intended tonal contrast. F0 was manipulated using the PSOLA (pitch-synchronous overlap and add) technique, with Praat software.

2.5 Phonetic Measurements of the Stimuli

Since each target word for the experimental task was produced naturally within a given tonal context, there was a high probability that a target word could have different phonetic values for both segmental and suprasegmental properties, depending on the context that it was embedded in. In order to examine the potential influence of phonetic details of stimuli on the results of the perception experiment, the durations of target-bearing APs, target words, and the first syllable of the target words were measured. Other parameters included the average amplitude of the first syllable of the target words, the closure duration of initial consonant of the target words (for oral and nasal stops only), lag VOT and RMS burst energy of the word-initial oral stop, and the nasal energy minimum of the word-initial nasal consonants.

A series of Friedman's test, a rank-based nonparametric equivalent of repeated measures, were performed. Significant differences among the tonal patterns were found for the amplitude (X^2 (5) = 70.55, p < .001) and duration of the word initial syllables (X^2 (5) = 23.06, p < .001) and for the VOT values of the word initial stop consonants (X^2 (5) = 62.71, p < .0001). These parameters, however, did not show the same tendency. On the one hand, word initial syllables were louder and longer when they were in AP-medial position than in AP-initial position. On the other hand, VOT values of word initial stop consonants were higher in the AP-initial tonal patterns than in the AP-medial tonal patterns. If acoustic cues present in the signal were to affect listeners' performance on a perceptual test, these three parameters could potentially influence the results of current segmentation task.

2.6 Procedure

Ninety native speakers of Seoul Korean participated in the experiment. They were born and raised in Korea and studying at University of California, Los Angeles or Stanford University at the time of the study. None of them were familiar with the purpose of the study and none had hearing difficulties. They were paid for their participation.

Every target and filler word carried six different tonal patterns as described in the previous section. 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. Every listener heard eight practice items before she/he was exposed to one of the six main experimental lists.

Participants were tested individually in a sound-attenuated booth. The PsyScope software package and a CMU button box were used for stimulus presentation and reaction time (henceforth RT) recording. Listeners heard the stimuli on a Macintosh computer through a pair of headphones at a comfortable volume. They were informed, via oral and written instructions, that the words that they had to spot were disyllabic or trisyllabic nouns of Korean, and they were asked to say the word out loud as soon as they spotted a word from the given speech stream. After the instructions, the participants could start the practice session by pressing a button. The RT measure was activated by the participant's voice through a desktop microphone that was located directly in front of the participant. The recorded RT was later adjusted such that RT was the duration between the offset of the word and the onset of the participants' verbal responses. Participants' oral responses were recorded into another Macintosh computer using the SoundEdit program. Missing or incorrect responses were assessed by listening to the participants' responses during the experimental session, and they were scored as errors.

3. Results

Missing items and incorrect responses were treated as errors, and they comprised 34.8% of the obtained data. RT values that did not fall within two standard deviations of the mean for each subject were 3.9% of all responses, and they were also treated as errors. Thus, the overall error rate was 38.7%.

For the statistical data analysis, mixed model analyses were employed, rather than repeated measures (RM) ANOVA. This analysis is more appropriate for analyzing the current results than an ANOVA analysis would be, because it does not assume homogeneity of variance, constant covariances, and sphericity. Further, the mixed model analysis is better in capturing statistical effects than a RM ANOVA analysis when there are missing data points (see Quene & van den Bergh (2004) for more information on mixed model analysis). The fact that this analysis can adeptly handle missing data points was a considerable advantage for the current study because there were a sizable number of missing data points for the RT analysis. The response data were submitted to linear mixed model analyses using SPSS v.11.5, with participants as the subject variable and items as the repeated variable.

In general, most of the main effects and interactions found in the error rates data were also found in the RT data. The mixed model analysis showed that there was a highly significant effect of *Tone patterns* in both error rates (F (5, 2791) = 187.64, $p < .0001$) and RT's (F (5, 1679) = 9.43, $p < .0001$).

<Figure 2> and <Figure 3> illustrate the mean error rates and RT's of six different tonal conditions, respectively. They clearly show that error rates were higher and RT's were faster in each of the three initial-tonal contours compared to the three medial-tonal contours in both disyllabic and trisyllabic words. Note that the two figures also display a visible discrepancy between the disyllabic words and trisyllabic words in all six tonal patterns.

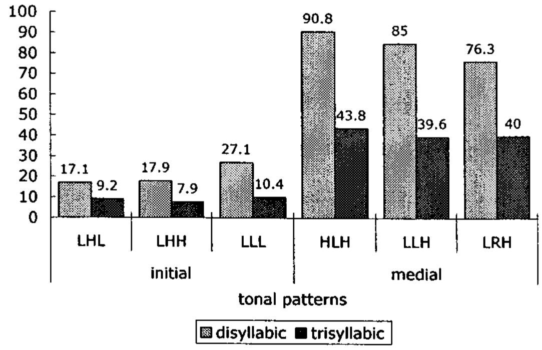


Figure 2. Mean error rate (%) in six tonal patterns

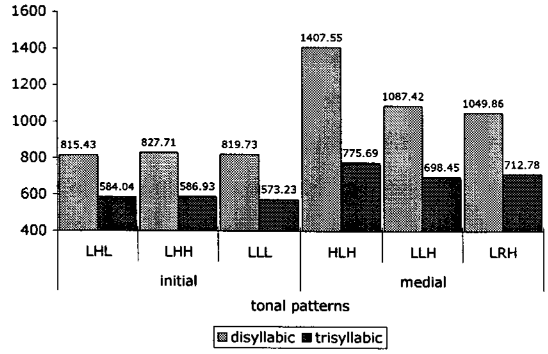


Figure 3. Mean RT's (ms) in six tonal patterns

Detailed statistical investigations showed that the *Location* effect (*viz.*, AP-initial *vs.* AP-medial positions) was large and significant in both error rates ($F(1, 2799) = 916.65$, $p < .0001$) and RT's ($F(1, 1472) = 123.69$, $p < .001$). The mean error rate in AP-medial condition (62.6%) was four times higher than in AP-initial condition (14.9%), and the target words were detected faster in AP-initial position than in AP-medial position (811.7ms, SD=443.2).

The effect of the *Number of Syllables* in the target words, which was observed in <Figures 2 and 3> above, also turned out to be highly significant. Listeners made more errors while spotting disyllabic words than trisyllabic words ($F(1, 2789) = 245.02$, $p < .0001$). The mean error rate among disyllabic words was 52.4%, but 25.1% among trisyllabic words. RT was also significantly faster ($F(1, 1491) = 70.60$, $p < .001$) in trisyllabic (646.8 ms, SD=433.8) than disyllabic words (932 ms, SD=600.3).

There was an interaction between *Location* and *Number of Syllables* ($F(1, 2810) = 115.43$, $p < .0001$) for the error rate, $F(1, 1685) = 12.35$, $p < .0001$ for RT), and between *Tone patterns* and *Number of Syllables* ($F(5, 2790) = 24.93$, $p < .0001$, for the error rate, $F(5, 1674) = 5.51$, $p < .0001$ for RT). Recall that three tonal patterns were under AP-initial position and the other three patterns were under AP-medial position. Since the tonal condition factor was embedded in the AP location factor, the interaction between the two could not be obtained.

Subsequent mixed model analyses were performed separately on the four conditions, which were the combinations of the syllable count of words and the location of words within AP (*viz*., disyllabic AP-initial and -medial, trisyllabic AP-initial and -medial conditions).

First of all, it should be noted that the RT data did not reveal as much of a tonal pattern effect as error rate data did in these analyses. A significant RT difference was found only in disyllabic AP-medial condition. Thus, the analyses reported below are based only on error rates, unless specified otherwise. In the disyllabic AP-initial condition, the tone effect was significant (F (2, 639) = 4.67, p = .01). Further *post-hoc* pairwise comparisons showed that the error rates for ***LH***H and ***LH***LH were significantly lower than those for ***LL***LH (p < .01, p = .012, respectively). There was no statistically significant difference between ***LH***H and ***LH***LH. In the disyllabic medial condition, the tone effect was highly significant (F (2, 639) = 10.12, p < .0001). The error rates for L***HL***H and L***LL***H were significantly higher than those for L***LH***H (p < .0001, p < .01, respectively). Although L***HL***H showed higher error rates than L***LL***H, the difference between the two was not statistically significant. RT difference was significant in this condition (F(2, 100) = 4.44, p=.014): listeners were slower in spotting words in L***HL***H than L***LL***H (p = .018) and L***LH***H (p < .01). There was no significant tone pattern effect in trisyllabic words, regardless of their location within the AP.

Recall that the phonetic measurements of the stimuli showed significant differences in terms of three parameters: amplitude was louder and duration was longer in the word medial syllables than in the word initial syllables; VOT was longer in the word initial syllables than in the word medial syllables. Among these three parameters, only VOT values proved to be consistent with the results of word spotting task: listeners' responses were faster and more accurate when they spotted words in AP-initial position than those in AP-medial position. Thus, a correlation analysis was performed in order to investigate if there was a correlation between the error rate and RT and VOT values. Pearson's correlation between the error rates and VOT values was r(1185) = -.33, p<0.001, which means that there was a weak but significant negative correlation between them. This result indicates that the error rates increased as VOT values decreased. No significant correlation was observed between RT and VOT values.

Finally, a significant consonant effect was observed (F (1, 2789.000) = 9.574, p < .01). The error rates were higher when the target word began with a sonorant consonant (nasals and glides, 42.1%) than with a non-sonorant consonant (stops and affricates, 36.4%). There was an interaction between consonant type and tone patterns (F (12, 1006.480) = 220.616, p < .0001). The RT results also revealed a significant consonant effect, but only in the disyllabic initial condition (F (1, 493.655) = 9.026, p < .01). Participants were faster in detecting words with a non-sonorant onset than a sonorant onset.

4. Discussion

4.1 Positional Effect in Word Segmentation

Listeners spotted words more accurately in AP-initial position than in AP-medial position, and RT was also faster in AP-initial than in AP-medial position. Within the category of rising pitch patterns (*LH*LH, *LH*HH, L*LH*H), listeners performed significantly better when the words were in AP-initial position (*LH*LH, *LH*HH) than when they were in AP-medial position (L*LH*H). The same positional effect was also observed when the words occurred with non-rising pitch patterns. Listeners were better at detecting words with *LL*LH tone in AP-initial position than with L*LL*H in AP-medial position. Further, the results showed that even the non-favorable tonal pattern in initial position (*LL*LH) was significantly more helpful to the listeners than the favorable tonal pattern in medial position (L*LH*H), which implies that there was a positional effect which exerted more power than rising intonation during the word segmentation task. The results strongly support the claim that words that are aligned with prosodic boundaries are favored in the word recognition process (Salverda, Dahan & McQueen, 2003).

When words are aligned with prosodic boundary, word segmentation can be benefited by acoustic correlates of prosodic boundaries. In the current study, the target words were always in the second AP in speech streams composed of three APs. Thus, the words that were inserted in AP-initial position were always preceded by another AP. Since the stimuli of the current experiment were read naturally, the phrase final syllable of the first AP in the stream had all the prosodic information appropriate to an AP boundary such as AP-final rising tone. The result is concordant with Kim (2004)'s findings which showed that AP-final rising tone could aid segmentation. The results of current study provide evidence that this cue can be actively used in on-line word segmentation, as well. It seems to play a role as a marker of the end of the first AP, and at the same time, the beginning of the following AP, and thus, the following word. That is to say, final rising tone can be a reliable segmentation cue since it allows listeners to anticipate the beginning of a new word even before the word onset is actually perceived.

There is another cue that is linked to the beginning of an AP that contains a target word. It is well known that consonants are articulatorily strengthened in a domain-initial position (Cho & Keating, 2001; Fougeron & Keating, 1997). These studies implied that acoustic correlates of domain-initial strengthening could make it easier for listeners to find the beginning of speech units, and hence, help the segmentation process of spoken input. In fact, a recent study by Cho, McQueen, & Cox (*in press*) provided direct evidence in support of this conjecture. Their study showed that English listeners are sensitive to the acoustic manifestation of domain initial strengthening present in the IP, and that they can use such phonetic information in speech segmentation.

Although the Korean AP is a smaller prosodic unit than the English IP, it is plausible to expect a similar perceptual effect of domain initial strengthening in Korean, because Korean showed stronger and more consistent initial strengthening effects than other languages such as English, French, and Taiwanese (Keating, Cho, Fougeron, & Hsu, 2003). Unlike Cho et al. (in press)'s study, the target words for this study were put in the same prosodic phrase, but in different positions. Word initial stop consonants showed longer VOT in AP initial position than AP medial position, and a correlation between VOT and listeners' error rates was observed. Furthermore, their performance was much better, that is, more accurate and faster, when the initial consonant was a non-sonorant consonant than when it was a sonorant consonant. This result indicates that lower level acoustic processing, that is, a salient perceptual cue such as the VOT of the at word initial segment, can influence the speed of lexical access.

Therefore, it seems reasonable to conclude that acoustic correlates of domain-initial strengthening is at least partially responsible for listeners' preference for AP-initial words. A strong consonant appearing at phrase initial position would trigger activation of words starting with that consonant. Therefore, detection of phrase-initial words could directly make use of such activation, whereas detection of phrase-medial words would rather be interfered by the words activated as such at the beginning of a phrase.

This result suggests that both an intonational pattern (i.e., final H tone that demarcates AP's) and phonetic manifestation of prosody (i.e., domain-initial strengthening) can aid listeners' word segmentation by giving out prosodic boundary information. However, the current study was not able to show the degree to which these cues contribute to word segmentation. Further studies with cross-spliced stimuli that can allow us to observe the impact of individual cues will be needed to explore this issue.

4.2 Effects of Tonal Patterns in Word Segmentation

The statistical analyses on the individual tonal patterns indicate that listeners used cues from various tonal patterns in a selective way during the word segmentation task. Since the trisyllabic target words did not show any effect of the different tonal patterns, I will mainly discuss the results from disyllabic target words.

4.2.1 Rising vs. Non-rising

The rising pattern imposed upon the word played a beneficial role in on-line word segmentation in Korean. The words with rising tonal contours (***LH***LH, ***LH***HH in AP-initial position; L***LH***H in AP-medial position) were spotted more accurately than the ones without a pitch rise (***LL***LH in AP-initial position; L***HL***H, L***LL***H in AP-medial position) both in AP-initial and AP-medial positions. Recall that Kim's (2004) corpus study showed that the rising tone pattern co-occurs most frequently with multisyllabic content words, when the word-initial syllable onset is not a tense/aspirated consonant, and that about 78% of AP-medial words started with a rising tone and had their word offsets aligned with AP-final syllables. The listeners' behavior observed here, therefore, indicates that language-specific prosodic property and its frequency rate affect the speed of word segmentation in on-line processing. This further confirms that listeners made use of lexical stress in English and Dutch or a H peak aligned with the initial syllable of a content word in French not because they are perceptually salient, but because they are one of the language-specific prosodic aspects to a given language.

Although the occurrence of a rising contour in the middle of an AP (i.e., L***LH***H) has not been attested in Seoul Korean, the unacceptability of that AP tonal pattern did not interfere with listeners' segmentation process. It is possible that the result was caused by listeners' misparse of the intonation pattern. As they were hearing speech signal unfolds in time, they might have perceived the local rising tone (LH) that spans over a disyllabic word as an AP-final H tone. That is, listeners could have interpreted the tonal pattern of target bearing AP as LLH before they heard the 'real' AP-final H tone in LLHH. It may be partially due to the aforementioned fact that 78% of AP-medial words started with a rising tone and had their word offsets aligned with AP-final syllables (Kim, 2004). Listeners' might have assumed that the H tone in L***LH***H was an indicator of the end of an AP, and the presence of such phrase-final cue in their mind could have aided word segmentation.

4.2.2 AP-initial LHLH vs. AP-initial LHHH

In AP-initial position, the error rates and RT for the two rising tonal patterns, *LH*LH and *LH*HH, were not different from each other. This suggests that a falling tone after the target word does not necessarily aid segmentation by marking an AP boundary after the target word, and further implies that once the word is recognized in one way or another, the tonal pattern that follows the recognition point does not facilitate segmentation. It is similar to the result of Cho et al (*in press*)'s study in that potential segmentation cues have no effect on on-line lexical segmentation once a word is recognized.

4.2.3 AP-medial LHLH vs. AP-medial LLLH

In AP-medial position, the RT for medial L*HL*H was significantly slower than that for L*LL*H. Error rates for medial L*HL*H were also higher than those for medial L*LL*H, though the difference was not significant. The significant RT difference between the two medial non-rising tone patterns indicates that the negative tonal cue in L*HL*H (falling, as opposed to rising) affected listeners' segmentation process in a negative way. It further implies that tonal contours can have both a facilitative and a disruptive influence on segmentation.

Compared with L*LH*H, both L*HL*H and L*LL*H did not contribute to increasing listeners' accuracy. This indicates that the frequency of the overall intonational pattern is less important than a specific intonational pattern that is imposed on a word itself.

4.3 Effect of Syllable Count on Word Segmentation of Korean

The results of the current experiment clearly revealed that the overall performance of listeners differed depending on the syllable counts of the target words. Error rates were significantly lower and RT was faster in trisyllabic target words than in disyllabic target words. Furthermore, the tonal pattern effect also showed distinct patterns between disyllabic and trisyllabic conditions. No tonal pattern 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 much 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 the Korean lexicon itself. Most Korean trisyllabic words, within a certain frequency range, happen to contain a string of syllables which sound exactly like existing disyllabic words. For instance, the noun *moseori* [mosɤɾi] 'edge' contains a disyllabic word *seori* [sɤɾi] 'frost', and *tarimi* [taɾimi] 'iron' contains *tari* [taɾi] 'leg'. The current experiment avoided such words, in order to be faithful to the 'no word-within-word' constraint for the word spotting task, but this resulted in a set of trisyllabic target words which were phonologically distinctive. As a result, the trisyllabic target words did not have dense phonological neighborhoods. Hence, we could assume that the discrepancy between the trisyllabic target words and the disyllabic target words might follow from the difference in neighborhood density.

Neighborhood density refers to the number of words which are phonologically similar to a given word. Luce and Pisoni (1998) have reported that listeners are sensitive to neighborhood density during on-line word recognition tasks, such that words in high density neighborhoods were responded to more slowly than those in low density neighborhoods. In order to examine if the neighborhood account could fit the current results, I performed a *post-hoc* analysis on the neighborhood density of each target word by calculating the number of phonological neighbors of 32 target words (16 disyllabic and 16 trisyllabic) based on the 77,180 Korean words listed in Kim and Kang (2000)'s corpus. A phonological neighbor (i.e., phonologically similar word) was defined by an addition, deletion, or substitution of a segment to a target word regardless of the location of a segment within a word. Thus, for instance, a word *napi* [nabi] 'butterfly' has neighbors such as *napip* [nabip] 'payment', *nap* [nap] 'lead', *mapi* [mabi] 'paralysis' etc. The calculated result showed that disyllabic targets had 24.5 phonological neighbors on average (S.D.= 7.66), whereas trisyllabic targets had average 2.06 phonological neighbors (S.D= 1.65). The difference between the number of neighbors of disyllabic words and those of trisyllabic words were, of course, highly significant (F (1, 30) = 131.199, $p < .001$). Further, there was a highly significant correlation between the number of neighbors for the target words and the number of errors for each target word ($r(30) = 0.581$, $p < .001$). Pearson's correlation was also significant between the number of neighbors for the target words and the reaction time for the target words (r (2878) = 0.267, $p < .001$). This indicates that the number of neighbors of target words affected listeners' segmentation ability. Overall, the results of these statistical analyses seem to provide enough evidence for the neighborhood density account.

The better performance of listeners for trisyllabic target words also seems to suggest that the effect of word frequency is less crucial than the effect of number of competitors, i.e., neighborhood density, when it comes to word recognition. It is well-known that high frequency words are recognized faster than low frequency words. In the current study, although trisyllabic words were in a lower frequency range, they still drew faster and more accurate responses from the listeners than the target disyllabic words which were in a higher frequency range. This conforms to previous studies which have claimed that the frequency effect can be eliminated altogether if the number of neighbors is controlled.

Therefore, based on the number of phonological neighbors of target words, we can conclude that the trisyllabic target words used in this experiment had minimal lexical competition due to sparser neighborhoods, as a result of which, listeners were perhaps able to detect these words equally well and fast, regardless of the tonal pattern superposed upon them.

5. CONCLUSION

This study performed a word spotting experiment in order to investigate whether post-lexical intonational patterns of a prosodic phrase can affect listeners' on-line word segmentation. The results show that intonation patterns and their frequency play various roles in word segmentation. Listeners spotted more words when they occurred in AP-initial position than when they occurred in AP-medial position, suggesting that the final rising tone of the preceding AP could be reliably exploited as a cue for segmentation, with other acoustic correlates of edges of a prosodic phrase. It is likely that this tone allowed the listeners to

correctly anticipate the upcoming word before its onset had actually been perceived. Results further show that listeners were much faster in detecting trisyllabic words than disyllabic words, and that there was no tonal effect observed in trisyllabic words. This result was accounted for by neighborhood density. Trisyllabic target words for the experiment did not have many neighbors, hence the listeners were able to detect these words more easily regardless of the tonal patterns imposed upon them. This, in turn, suggests that neighborhood effect could weigh more than post-lexical intonation patterns in word segmentation. In the analysis of disyllabic word spotting, the results indicate that error rates were significantly lower for rising tonal patterns than for non-rising patterns, regardless of the position of the target words within an AP and regardless of the frequency and legality of the overall AP tonal patterns that were presented. Finally, a falling tone over target words turned out to be a negative tonal cue. These findings suggest that a rising tone superposed upon words, but not the overall AP tonal pattern itself, can serve as a word segmentation cue in Korean, a language that lacks lexical prominence. The results further suggest a possibility that even non-lexical prosodic information can be stored and can affect word segmentation when the information frequently co-occurs with words.

References

Bagou, O., Fougeron, C. & Frauenfelder, U. 2002. "Contribution of prosody to the segmentation and storage of "words" in the acquisition of a new mini-Language." *Proceedings of Speech Prosody*, 159-162.

Beckman, M. & Edwards, J. 1990. "Lengthening and shortening and the nature of prosodic constituency." In J. Kingston and M. Beckman (eds.), *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech.* New York: CUP. pp. 152-178.

Bolinger, D. 1970. "Relative Height." In P. Leon (Ed.), *Prosodic Feature Analysis*, Montreal: Marcel Didier. pp. 109-127.

Cho, T. 2001. "Effects of morpheme boundaries on intergestural timing: evidence from Korean." *Phonetica* 58, 129-162.

Cho, T. & Keating, P. 2001. "Articulatory strengthening at the onset of prosodic domains in Korean." *Journal of Phonetics* 28,155-190.

Cho, T., McQueen, J. & Cox, E. (in press). "Prosodically driven phonetic detail in speech processing: The case of domain-initial strengthening in English." *Journal of Phonetics* 33(2), 121-157.

Christophe, A., Gout, A., Peperkamp, S. & Morgan, J. 2003. "Discovering words in the continuous speech stream: the role of prosody." *Journal of Phonetics* 31, 585-598.

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

Cooper, W. & Paccia-Cooper, J. 1980. *Syntax and Speech.* Cambridge, MA. Harvard University Press.

Cutler, A. & Carter, D. 1987. "The predominance of strong initial syllables in the English vocabulary." *Computer Speech and Language* 2, 133-142.

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

Edwards, J., Beckman, M. & Fletcher, J. 1991. "Articulatory kinematics of final lengthening." *Journal of the Acoustical Society of America* 89, 369-382.

Forster, K. & Chambers, S. 1973. "Lexical access and naming time." *Journal of Verbal Learning and Verbal Behavior* 12, 627-635.

Fougeron, C. & Keating, P. 1997. "Articulatory strengthening at edges of prosodic domains." *Journal of the Acoustical Society of America* 101, 3728-3740.

Gout, A., Christophe, A. & Morgan, J. 2004. "Phonological phrase boundaries constrain lexical access: II. Infant data." *Journal of Memory and Language* 51, 548-567.

Havens, L. L. & Foote, W. E. 1963. "The effect of competition on visual threshold and its independence of stimulus frequency." *Journal of Experimental Psychology* 65, 6-11.

Hood, J. D. & Poole, J. P. 1980. "Influence of the speaker and other factors affecting speech intelligibility." *Audiology* 19, 434-455.

Jun, S. 1996. *The Phonetics and Phonology of Korean Prosody: intonational phonology and prosodic structure*, New York, NY: Garland Publishing Inc.

Jun, S. 2000. "K-ToBI (Korean ToBI) labelling conventions: Version 3." *Speech Sciences* 7, 143-169.

Jun, S. & Fougeron, C. 2000. "A Phonological model of French intonation." In A. Botinis (Ed.), *Intonation: Analysis, Modeling and Technology*, Kluwer Academic Publishers, pp. 209-242.

KAIST. 1999. KAIST Concordance Program Online Demo Version. http://csfive.kaist.ac.kr/kcp/

Keating, P, Cho, T., Fougeron, C. & Hsu, C. 2003. "Domain-initial articulatory strengthening in four languages." In J. Local, A. Ogden, A. Temple (Eds.), *Papers in Laboratory Phonology* 6, pp. 143-161.

Kim, H.-G. & Kang, B.-M. 2000. *Frequency Analysis of Korean Morpheme and Word Usage*. Seoul: Institute of Korean Culture, Korea University.

Kim, S. 2004. *The Role of Prosodic Phrasing in Korean Word Segmentation*. Doctoral dissertation, UCLA.

Klatt, D. H. 1975. "Vowel lengthening is syntactically determined in a connected discourse." *Journal of Phonetics* 3, 129_140.

Luce, P. A. 1986. "Neighborhoods of words in the mental lexicon." Bloomington, Indiana: Indiana University, Psychology Department, Speech Research Laboratory.

Luce, P. A. & Pisoni, D. B. 1998. "Recognizing spoken words: The neighborhood activation model." *Ear and Hearing* 19, 1-36.

Norris, D. (1986). "Word recognition: context effects without priming." *Cognition* 22(2), 93-136.

Pierrehumbert, J. 1980. *The Phonology and Phonetics of English Intonation*. Doctoral dissertation, MIT

Pierrehumbert, J. & Talkin, D. 1991. "Lenition of /h/ and glottal stop." In *Papers in Laboratory Phonology II*, Cambridge UK: Cambridge University Press, pp. 90-117.

Pisoni, D. B., Nusbaum, H., Luce, P.A. & Slowiaczek, L. M. 1985. "Speech perception, word recognition, and the structure of the lexicon." *Speech Communication* 4, 75-95.

Rubenstein, H., Garfield, L. & Millikan, J. 1970. "Homographic entries in the internal lexicon." *Journal of Verbal Learning and Verbal Behavior* 9, 487-494.

Quene, H. 1992. "Integration of acoustic-phonetic cues in word segmentation." In M. Schouten (Ed.), *The Auditory Processing of Speech: From Sounds to Words*, pp. 349-355.

Quene, H. & van den Bergh, H. 2004. "On multi-level modeling of data from repeated measures designs: A tutorial." *Speech Communication* 43, 103-121.

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

Vroomen, J. & de Gelder, B. 1995. "Metrical segmentation and lexical inhibition in spoken word recognition." *Journal of Experimental Psychology: Human Perception and Performance* 23, 710-720.

Welby, P. 2003. "The slaying of lady mondegreen, being a study of French tonal association and alignment and their role in speech segmentation." Doctoral dissertation. The Ohio State University.

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

received: January 29, 2007
accepted: March 10, 2007

▲ Sahyang Kim
Dept. of English Education, Hong-Ik University
Mapo-Ku Sangsoo-Dong 72-1, Seoul 121-791
Tel: +82-2-320-1180
E-mail: sahyang@gmail.com

APPENDIX I. Frequency and familiarity of target words

	phonetic transcription	gloss	Freq.count (Kang & Kim 2000)	Freq.count (Kaist)	Freq.rank (Kaist)	Familiarity
Disyllabic Target Words	[na.bi]	butterfly	47	496	2638	3.3
	[ma.ru]	floor	61	413	3044	3.7
	[mo.t∑a]	hat	73	572	2377	4.5
	[m℘.pi]	head	635	5809	320	4.5
	[na.ra]	nation	1885	5918	317	4.1
	[no.pE]	song	477	1017	1478	4.7
	[na.mu]	tree	571	2305	698	4.5
	[t∑a.ju]	freedom	695	24	23021	4.1
	[ta.pi]	leg/bridge	387	2425	667	4.5
	[ko.gi]	meat	176	770	1846	4.7
	[pa.t∑i]	pants	71	612	2242	4.6
	[ki.do]	prayer	95	243	4686	4.4
	[pa.da]	sea	469	1867	841	4.3
	[ku.du]	shoes	147	407	3082	4.5
	[k℘.pi]	street	517	3031	562	4.4
	[ju.pi]	glass	92	317	3771	3.9
Trisyllabic Target Words	[mu.d℘.gi]	pile	29	91	9780	3
	[na.nu.gi]	division	1	n/a	n/a	3.4
	[na.d\.pi]	outing	26	48	15105	3
	[nu.d℘.gi]	rag	11	98	9297	2.5
	[to.tʰo.pi]	acorn	15	80	10664	3
	[to.kE.bi]	elf	26	123	7836	3.6
	[mE.t*u.gi]	grasshopper	18	33	19033	3.4
	[t∑E. t∑ʰE.ki]	sneeze	13	28	20924	4.1
	[tu.k*℘.bi]	toad	7	52	14299	3.1
	[ko.gu.ma]	sweetpotato	13	48	15125	4
	[pa.gu.ni]	basket	33	178	5942	3.8
	[to.pa.t∑i]	bellflower	10	54	13955	3.2
	[tu.d℘.t∑i]	mole	12	43	16198	3
	[to.ga.ni]	pot	8	29	20678	3.1
	[ki. p℘.ki]	wildgeese	13	52	14322	3
	[po.t*a.pi]	bundle	14	196	5519	3.4

APPENDIX 2. Target word location and tone patterns

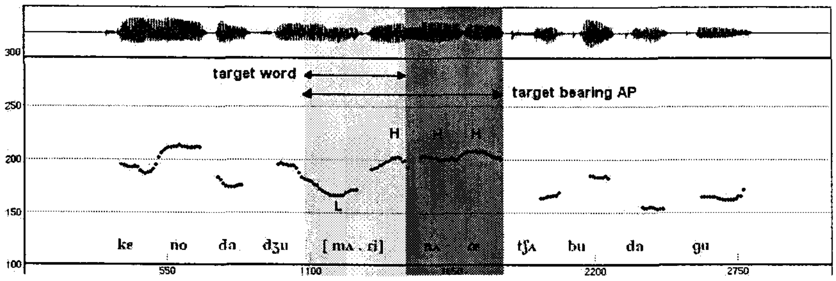


Figure 1. meori [m℘pi] 'head' in AP initial LHLH pattern

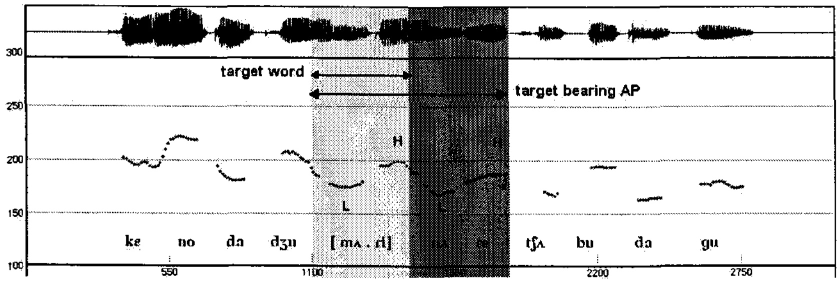


Figure 2. meori [m℘pi] 'head' in AP initial LHHH pattern

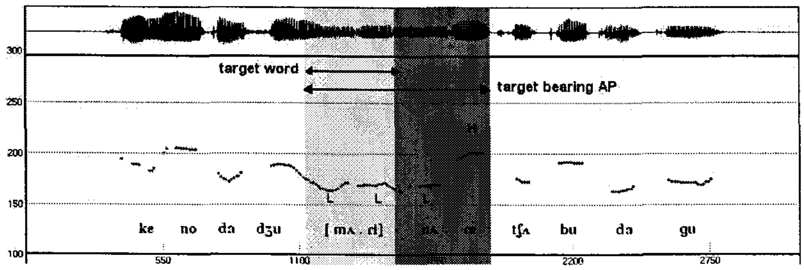


Figure 3. meori [m℘pi] 'head' in AP initial LLLH pattern

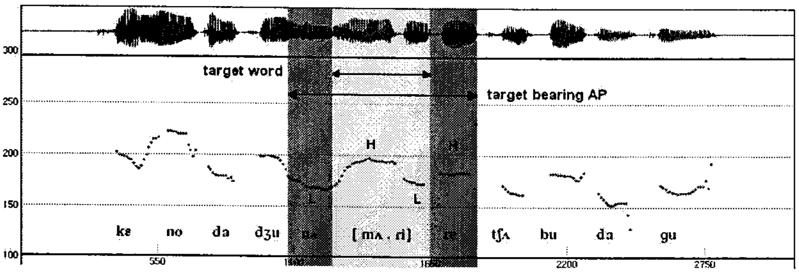


Figure 4. meori [m℘pi] ‘head’ in AP medial LHLH pattern

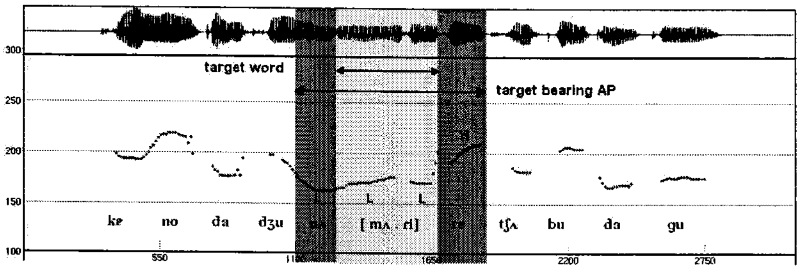


Figure 5. meori [m℘pi] ‘head’ in AP medial LLLH pattern

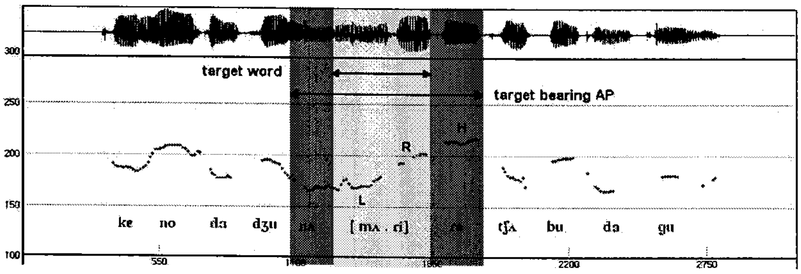


Figure 6. meori [m℘pi] ‘head’ in AP medial LLRH pattern