

음운 구조가 한국어 단어 분절에 미치는 영향

김 사 향
홍익대학교 영어교육과

The role of prosodic phrasing in Korean word segmentation

Sahyang Kim
Dept. of English Education, Hongik University
E-mail : sahyang@hongik.ac.kr

Abstract

The current study investigates the degree to which various prosodic cues at the boundaries of a prosodic phrase in Korean (Accentual Phrase) contributed to word segmentation. Since most phonological words in Korean are produced as one AP, it was hypothesized that the detection of acoustic cues at AP boundaries would facilitate word segmentation. The prosodic characteristics of Korean APs include initial strengthening at the beginning of the phrase and pitch rise and final lengthening at the end. A perception experiment revealed that the cues that conform to the above-mentioned prosodic characteristics of Korean facilitated listeners' word segmentation. Results also showed that duration and amplitude cues were more helpful in segmentation than pitch. Further, the results showed that a pitch cue that did not conform to the Korean AP interfered with segmentation.

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. Although segmentation is accomplished without any conscious effort on the

part of the listener, the task is by no means trivial. Numerous studies that have sought to illuminate facilitating word segmentation cues revealed that accurate segmentation could be aided by multiple means, such as phonotactic cues [1], [2], transitional probabilities [3], [4], rhythmic cues [5], [6], [7], and various acoustic cues [8], [9].

The current study investigated whether the acoustic cues accompanying Accentual Phrase boundaries facilitate word segmentation in Korean, and if they do, whether the degree of contribution of each cue to segmentation differs. The Accentual Phrase was taken as a reference prosodic unit, because the AP is more or less the same in size as a word in Korean [10]. Therefore, it was assumed that the detection of the AP would naturally help detection of words. This project also explored whether word segmentation would be affected only by the acoustic cues that are associated with Korean AP, or just by any salient and reliable acoustic cue that could draw people's attention.

2. Experiment

In order to test the role of various acoustic cues in speech segmentation, this experiment employed the artificial language learning method [2], [3].

2.1. Stimuli

The artificial language used for the current experiment had four consonants (/p/, /t/, /k/, /m/) and four vowels (/a/, /i/, /u/, /E/). The combination of the eight segments resulted in 16 distinct CV syllables. These syllables were combined to make six trisyllabic words: *putaki*, *makute*, *tipemu*, *kapitu*, *mikepa* and *kumepi*. These words did not have any semantic or morphological content and none of the words were real words in Korean.

2.2. Recording and manipulation of stimuli

A female native speaker of Korean (the author) produced each CV syllable 10 times. In order to avoid coarticulation effect across syllables, each syllable was produced independently, following a pause. Out of the 10 tokens, only one token was chosen as a stimulus for a CV syllable. The selected tokens were manipulated such that the duration, amplitude and pitch of each individual syllable were normalized to the average value of all the syllables. The normalized syllables were concatenated to make words. Then, the words were concatenated in random order without a pause between words and yielded about a 10-minute speech stream. Each word occurred 144 times in the speech stream. One word never occurred twice in a row in the given speech stream, and there was no pause between the words in any part of the speech stream. Transitional probabilities (= frequency of XY / frequency of X) within words ranged from 0.5 to 1, and those across words ranged from 0.05 to 0.33.

The experiment had five prosodic conditions. *Noprosody* condition was the baseline condition where all syllables had the same duration, pitch, and amplitude. In this condition, listeners were supposed to extract words from speech stream relying solely on transitional probabilities. There were three conforming prosodic conditions, i.e. *Duration*, *Amplitude*, and *Pitchfinal*, all of which conformed to the prosodic characteristics of Korean. The final syllable of each word was lengthened by 30% in *Duration* condition, imitating AP final lengthening. In *Amplitude* condition, the first

syllable of each word had increased amplitude (10% increase rate), imitating APinitial strengthening. In *Pitchfinal* condition, f₀ of the word final syllable was 13% higher than the others, imitating APfinal pitchrising. Finally, there was non-conforming condition of *Pitchinitial*, which had a high pitch on the first syllable of every word (13% increase rate).

The Praat software was used for the concatenation and the sound manipulation processes.

2.3. Procedure

Sixty native speakers of Seoul Korean participated in the experiment. Thus, there were twelve participants for each prosodic condition. The experiment was composed of a learning phase and a testing phase. All subjects were run individually in a soundproof booth.

During the learning phase, subjects heard a speech stream from one of the five conditions. They were told that they would hear a speech stream from a simple artificial language which was composed of concatenated non-sense words, and that there would be no pause between words. They were informed that their task was to extract trisyllabic words from the speech stream. However, the information about the number of words in the language was not given to subjects. The speech stream was presented to subjects using SoundEdit program in a Macintosh computer, and subjects heard the concatenated sounds through a headphone. They were asked to adjust the volume at the most comfortable level. The learning phase lasted approximately 11 min. for *Duration* condition, and 10min. for the other conditions.

In the testing phase, there were 36 forced choice pairs that were made from the combination of the six trisyllabic real-words of the artificial language and six trisyllabic test strings. The trisyllabic test strings were composed of three part-word and three non-word strings that had not been included in the original word list of the artificial language. A part-word contained the final two-syllable string from a real word of the language and an additional syllable. All three part-word strings used in the testing phase occurred 42 times each during the

learning phase. Non-word strings were composed of the syllables that were used in the learning phase, but had the sequence of syllables that subjects had never heard during the learning phase. Thus, transitional probabilities of non-words were zero.

In each trial, subjects heard a pair of trisyllabic strings, one of which was a real word in the artificial language and the other was a test string. There was 800 ms inter stimuli interval between the two strings in a pair. All the strings presented in testing phase had the same prosodic cue as in the learning phase. After listening to the two-alternatives, subjects were asked to choose which of the two strings was a word from the artificial language. Subjects entered their response by pressing the '1' key on the keyboard if the string presented first was thought to be an answer, or the '0' key on the keyboard if the string presented second was thought to be an answer.

The test stimuli were presented and the responses were collected using Psyscope software in a Macintosh computer.

3. Results

The average percentage of correct identification was 56.6 % (mean raw score 20.4 out of 36, SD = 4.4) for *NoProsody* condition, which was higher than chance level (50%). One third of the participants in this condition, however, showed at chance or below chance performance. *Duration* condition showed the highest average percentage of 83% (mean score 29.9, SD=4.7), followed by 76.3% (mean score 27.5, SD=5) for *Amplitude* condition. All the participants for these two conditions showed above chance performance. In *Pitchfinal* condition, the listeners showed above chance performance except for one participant who showed at chance performance. The average correct percentage for this condition was 67.5% (mean score 24.3, SD=4.6). The average correct percentage for *Pitchinitial* condition was 49.9% (mean score 18, SD=5.1), which was the lowest of all the conditions. Half of the participants in this condition showed at chance or below chance performance.

A repeated measures analysis of variance (RM ANOVA) was performed on the obtained data using SPSS v.11, with *word* as a within subject factor and *prosodic condition* as a between subject factor.

In this RM ANOVA analysis, the sphericity assumption was met, and thus degree of freedom was not adjusted. The result showed that there was no main effect of *word*, which suggests that all the words were equally easy or equally hard to the listeners. There was a significant effect of *prosodic condition* ($F(4,55) = 12.665, p < .001$), but there was no interaction between *word* and *prosodic condition*.

Pair-wise comparisons showed that listeners were significantly better in *Duration*, *Amplitude*, and *Pitchfinal* conditions than in *NoProsody* condition ($p < .001, p = .001, p = .04$, respectively). There was no difference between *Duration* and *Amplitude* conditions. However, listeners' performance was significantly better in *Duration* condition than in *Pitchfinal* condition ($p = .006$). *Amplitude* and *Pitchfinal* conditions were not different from each other. Finally, listeners' performance was significantly worse in *Pitchinitial* condition than in *Duration*, *Amplitude*, and *Pitchfinal* condition ($p < .001, p < .001, p = .002$, respectively). But there was no difference between *Pitchinitial* and *NoProsody* conditions.

4. Discussion

The results revealed that three conditions conform to the prosodic characteristics of the Korean AP, i.e., *Duration* (AP-final lengthening), *Amplitude* (AP-initial strengthening) and *Pitchfinal* (AP-final high tone), facilitated word segmentation more than *NoProsody* condition. Thus, we can conclude that each conforming prosodic cue, when combined with distributional information, can efficiently aid speech segmentation of Korean.

The statistical difference observed between *Duration* and *Pitchfinal* conditions suggests that the degree of contribution of each facilitating cue to word segmentation differed. Durational cue contributed significantly more to segmentation than F0 rise in the same syllable location, i.e. the

word-final position. This result can be interpreted in a couple of different ways. One possible interpretation for the difference could resort to prosodic structure account. That is, the observed difference was caused by the fact that the lengthened duration was long enough to mark a prosodic unit that is higher than the AP in Korean, i.e. Intonational Phrase, whereas the heightened F0 value only marked the AP. This interpretation implies that listeners have tacit knowledge of how an utterance is parsed into prosodic phrases, what kind of acoustic cues are associated with various prosodic units, and how the different degrees of an acoustic cue are related to different levels of prosodic phrases in the prosodic hierarchy. The second interpretation could simply rely on physical strength of each cue. Given that the two acoustic cues help speech segmentation, it is possible to consider that the stronger (i.e. longer duration and higher pitch) the acoustic cue is, the more helpful the cue is to segmentation. With regard to the current experiment in particular, it might be the case that the degree of lengthening (30 %) was psycho-acoustically stronger than that of F0 increase (13 %), and hence the former resulted in larger perceptual effect than the latter. If so, the difference does not necessarily have to do with the features of prosodic structure.

Future experiments are required in order to determine which of the two interpretations is correct. Also, before any further perception experiment is conducted, it will be necessary to obtain a sufficient amount of quantitative data so as to establish the accurate range of acoustic values of each prosodic cue in different prosodic positions.

A non-conforming condition (*Pitchinitial* condition) with pitch cue on the first syllable showed 49.9% of mean correction rate, with the half of the participants scored less than chance level, which indicates that this cue affected listeners negatively. This result was worse than that of *NoProsody* condition, where there was no prosodic cue present, though the two conditions were not significantly different. Although it does not seem to be implausible to learn a new pattern of rhythm which was combined with the existing distributional

cue within 10 minutes (which could have aided speech segmentation had it happened), listeners were notable to learn this unfamiliar pitch pattern within the given time. Recall that even if this can be a possible prosodic cue in some languages, it is not a part of native prosody of Korean. In addition, this condition violated tone-segment mapping rules of Korean by mapping a lenis/nasal consonant to a high tone. Thus, it is very likely that listeners would have tried to match high F0 to the last syllable of a word rather than the first syllable of a word. This attempt could have caused the mismatch of the two different sources of information, i.e. the distributional information present in the stimuli and the prosodic information, which consequently brought about even lower performance scores in *Pitchinitial* than *NoProsody* condition.

The comparison between *PitchFinal* and *PitchInitial* conditions indicates that a prosodic cue can have different function in terms of speech segmentation, depending on the location at which it is realized. The location of a prosodic cue is not randomly determined, but controlled by the prosodic characteristics of one language. Further, the result suggests that it is not the case that listeners automatically respond to any salient prosodic cue for speech segmentation. Rather, this result reveals listeners' strong dependency on the prosodic structure of their native language in word segmentation.

5. Summary

The results of the current experiment indicate that even though there is no word-level prosodic prominence in Korean, phrase-level prosodic properties can play an important role in word segmentation, given that the size of the AP is similar to that of a word. All the prosodic properties that are related to the AP, that is, duration cue at the end of the phrase, amplitude cue at the beginning of the phrase, and the pitch cue at the end of the phrase, helped the segmentation, while the cue that is against the AP characteristics did not aid the process.

6. References

- [1] Jusczyk, P.W., Luce, P.A., and Charles-Luce, J., "Infants' sensitivity to phonotactic patterns in the native language," *Journal of Memory and Language*, 33, 630-645, 1994.
- [2] Saffran, J., Newport, E.L., and Aslin, R.N., "Word segmentation: the role of distributional cue", *Journal of Memory and Language*, 35(4), 606-621, 1996.
- [3] Johnson, E. and Jusczyk, P.W., "Word segmentation by 8-month-olds: when speech cues count more than statistics," *Journal of Memory and Languages*, 44, 548-567, 2001.
- [4] Cutler, A., Mehler, J., Norris, D., and Segui, J., "The syllable's differing role in the segmentation of French and English," *Journal of Memory and Language*, 25, 385-400, 1986.
- [5] Cutler, A., and Norris, D., " The role of strong syllables in segmentation for lexical access," *Journal of Experimental Psychology: Human Perception and Performance*, 14, 113-121, 1988.
- [6] Cutler, A., and Otake, T., "Mora or phonemes? Further evidence for language-specific listening," *Journal of Memory and Language*, 33, 824-844, 1994.
- [7] Nakatani, L.H., and Dukes, K.D., "Locus of segmental cues for words juncture", *Journal of Acoustical Society of America*, 62 (3), 714-719, 1977.
- [8] Quene, H., "Segment durations and accent as cues to word segmentation in Dutch", *Journal of Acoustical Society of America*, 94(4), 2027-2035, 1993.
- [9] Jun, S. and Fougeron, C., "A phonological model of French intonation", in A. Botnis (ed), *Intonation: Analysis, Modelling and Technology*. Kluwer. 2000.
- [10] Bagou, O., Fougeron, C., and Frauenfelder U.H., "Contribution of prosody to the segmentation and storage of words in the acquisition of a new mini-language", *Speech Prosody Proc.*, 159-162, 2002.