The Role of Prosodic Boundary Cues in Word Segmentation in Korean

Sahyang Kim*

ABSTRACT

This study investigates the degree to which various prosodic cues at the boundaries of prosodic phrases in Korean contribute to word segmentation. Since most phonological words in Korean are produced as one Accentual Phrase (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 utilizing an artificial language learning paradigm revealed that cues conforming to the aforementioned prosodic characteristics of Korean facilitated listeners' word segmentation. Results also indicated that duration and amplitude cues were more helpful in segmentation than pitch. Nevertheless, results did show that a pitch cue that did not conform to the Korean AP interfered with segmentation.

Keywords: word segmentation, prosody, Korean

1. Introduction

To process spoken language, listeners must be able to segment the speech stream into discrete meaningful units. In order to accomplish this task successfully, they must be capable of detecting the correct boundaries for each of these units. Although segmentation is achieved without any conscious effort on the part of the listener, the task is by no means trivial. Numerous studies that have sought to identify those cues which may facilitate word segmentation have revealed that accurate segmentation may be aided by multiple means, such as phonotactic cues (McQueen, 1998; Jusczyk, Luce, & Charles-Luce, 1994), transitional probabilities (Saffran, Newport & Aslin, 1996; Johnson & Jusczyk, 2001), rhythmic cues (Cutler, Mehler, Norris & Segui, 1986; Cutler & Norris, 1988; Cutler & Otake, 1994) and various acoustic cues (Nakatani & Dukes, 1977; Quene, 1993). These studies have also provided evidence that most cues for segmentation, such as lexical, prosodic, and phonotactic cues, are language-specific.

Recently, several studies have focused on exploring the role of post-lexical prosody in

* Department of Audiology and Speech-Language Pathology, Wayne State University

lexical segmentation and access (Christophe, Peperkamp, Pallier, Block, & Mehler, 2004; Gout, Christophe, & Morgan, 2004). For example, in Christophe et al.'s 2004 study, adult French listeners were presented with sentences that contained a local ambiguity (e.g., [d'un chat grincheux]PP 'of a grumpy cat', ambiguous with the word chagrin) and sentences that did not (e.g., [d'un chat drogué] PP 'of a doped cat', no ambiguity). The listeners were asked to detect the word chat 'cat', and reaction times for the word-spotting task were significantly slower for the ambiguous tokens than for the non-ambiguous tokens. The experimenters went on to examine whether or not the local ambiguity effect would hold if the two syllables spanned a Phonological Phrase (e.g., ⋯chat] PP [grimpai⋯, with potential ambiguity vs. ⋯chat] PP [dressait ⋯, without potential ambiguity). In this case, though, there was no difference in reaction times between the two conditions. Moreover, detection of the target words was much faster when there was a Phonological Phrase boundary between the two syllables. Based on these findings, the researchers concluded that Phonological Phrase boundaries constrain on-line lexical access.

This study attempts to determine which attributes of the speech signal aid listeners' word segmentation by virtue of facilitating the perception of phrase boundaries. In particular, the current study examines whether or not the three acoustic cues associated with Accentual Phrase boundaries in Korean (i.e., phrase-initial strengthening, phrase-final high pitch, and phrase-final lengthening) facilitate lexical segmentation in Korean, and, if they do, whether the degree of contribution of each cue to word segmentation process differs. The Accentual Phrase (henceforth AP) in Korean is similar to the Phonological Phrase examined in Christophe et al.'s study, in that it is a prosodic level that lies in between a prosodic word and an Intonational Phrase. It was hypothesized that detection of the AP would directly help word segmentation in Korean, because it has been reported that the AP usually contains one or two content words in Korean (Jun & Fougeron, 2000; Kim, 2004). The current study also seeks to determine whether word segmentation would be affected by only those acoustic cues relevant to language-specific prosody, or whether this process could be impacted by any perceptually salient and reliable acoustic cue. To this end, this study investigates whether an acoustic cue that is perceptually salient enough but does not conform to the language-specific prosodic pattern of Korean (i.e., phrase-initial high pitch) can help listeners' lexical segmentation or not.

2. Experiment

In order to test the role of various acoustic cues in speech segmentation, this experiment employed an artificial language learning paradigm (Saffran et al., 1996; Bagou, Fougeron & Frauenfelder, 2002). The methodology for this paradigm dictates that the experimenter should create a simple artificial language composed of a small set of novel words. These words must not be morphologically or semantically related to the subjects' native language. The experimenter must then construct a lengthy stream of speech by randomly concatenating these words. The experiment itself consists of a 'learning phase', during which subjects are passively exposed to the concatenated stream of speech for about 20 minutes, and a 'testing phase', during which subjects are asked to identify the words that comprise the artificial language.

In Saffran et al. (1996)'s seminal work, this method was employed in order to determine whether or not transitional probability facilitates word segmentation.

Transitional probability shares the same basic concept as conditional probability, in that it considers the probability of an event given that another event has occurred. The standard equation for transitional probability is shown in (1):

$$P\ (Y|X) = \text{Frequency of } XY\ /\ \text{Frequency of } X \tag{1}$$

The equation in (1) computes the probability of Y given X (denoted as P (Y|X)), which is equal to the probability of the co-occurrence of XY over the probability of the occurrence X. In natural languages, transitional probability of two adjacent syllables tends to be higher within words than between words. For example, consider a word *tiger*. The occurrence of *ger* given ti would be higher than that of most other random syllables, such as *ber*, *der*, and so forth. This means that the transitional probability of *ti-ger* would be higher than that of *ti-ber* or *ti-der*. Consequently, if *ti* is followed by one of these random syllables, listeners would assume a word boundary between the two syllables because of the lower transitional probabilities. Saffran et al. (1996) found that listeners were able to segment words in an artificial language by relying solely on transitional probability information. They further showed that listeners' performance was enhanced when a prosodic cue (final lengthening) was added to the speech input.

As revealed in the above study, one of the advantages of the paradigm of artificial language learning is that an experimenter can control the acoustic and/or positional attributes and transitional probability, allowing them to be independently evaluated, without interference from confounding factors that arise when working with natural languages. Another benefit of this method is that the results of the experiment allow inferences to be made about the effect of segmentation cues on language learning itself, because the experimental procedure requires listeners to store the words they extract from the speech stream in lexical memory.

2.1 Stimuli

The artificial language used for this experiment contained four consonants (/p/, /t/, /k/, /m/) and four vowels (/a/, /i/, /u/, /e/). 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*. None of these are real words in Korean, and they do not possess any semantic or morphological content.

2.2 Recording of stimuli

A female native speaker of Korean (the author) produced each CV syllable 10 times. In order to avoid coarticulation effects across the syllables, each was produced independently, following a pause. Of the 10 repetitions, only one was chosen for use 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 for all the syllables. The normalized syllables were concatenated to make six trisyllabic words. Then, the words were concatenated in random order without a pause between words, yielding a speech stream of approximately 10 minutes in length. Each word occurred 144 times in the speech stream. None of the words ever occurred twice in a row in the speech stream, and there was never a pause between the words. Transitional probabilities of adjacent syllables within words ranged from 0.5 to 1, and those across words ranged from 0.05 to 0.33. Acoustic parameters were manipulated when necessary, depending on experimental conditions (see section 2.3. for details). Praat software was used for the concatenation and sound manipulation processes.

2.3 Experimental conditions

The experiment had five prosodic conditions. The ***No prosody*** 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 the speech stream relying solely on transitional probabilities.

There were three conforming prosodic conditions: ***Duration***, ***Amplitude***, and ***Pitch final***. These conditions imitated three prosodic characteristics of Korean (AP-final lengthening, AP-initial amplitude effect, and AP-final pitch rise), which were observed in various speech production studies.

In the ***Duration*** condition, which imitated AP-final lengthening, the final syllable of each word was lengthened by 30%. Although it was reported in Jun's studies (1993, 1995a) that AP-final lengthening was not found in Korean, two subsequent studies observed the lengthening effect. Cho and Keating (2001) and Oh (1998) found that there was a small but significant durational difference between AP-final and word-final (but non-AP-final) vowels. Note that the two speech production studies have presented different increase rate in terms of Korean AP-final lengthening. Cho and Keating (2001) found AP-final vowel lengthening of 50% in read-speech data from 3 speakers. Oh's (1998) study, which was based on read-speech data from 6 speakers reading 25 sentences five times, reported a 12% increase in vowel length at the

AP-final position. The rate of lengthening that the current study adopted was similar to the average rate of lengthening found in these two studies.

Amplitude cues seem to mark AP-initial boundary in Korean, as well. Jun (1995b) showed that when a trisyllabic reiterative word with a CV syllable (e.g., 'mamama') was in sentence-medial position, forming one AP, the mean amplitude (RMS) of the first syllable was higher than that for the second syllable, although it was similar to that for the third syllable. This pattern was observed strongly when speakers produced the word in loud voice. High amplitude in the third syllable was accompanied by a pitch cue, but the first syllable only had a strong intensity value, not a high pitch value. In the ***Amplitude*** condition, amplitude of the first syllable of each word was increased by 10%, imitating AP initial strengthening. Since an estimate of the appropriate rate of amplitude increase was not obtainable from any previous production studies, this rate was simply based on subjective evaluation of saliency by the author.

Phrase-final high tone is a part of pitch rising (LH) that consists a default tonal pattern of a Korean AP. The canonical tonal pattern of the AP includes phrase-initial pitch rising and phrase final pitch rising (viz., LHLH; see Jun, 1993). Although a low tone can be aligned with AP-final position, it happens only rarely. A recent corpus study (Kim, 2004) showed that 84% of IP-medial AP's end with a high tone. Furthermore, Jun (1995b) found that when reiterated nonsense words composed of light (CV) syllables form one AP in IP-medial position, the very last syllable of a word (i.e., the 2nd syllable in disyllabic words, the 3rd syllable in trisyllabic words) shows the highest F0 value. In the ***Pitch final*** condition, the F0 of the word final syllable was 13% higher than the others, imitating AP final pitch rising. This rate was obtained from the difference in F0 between low tone and high tone in thirty-three bisyllabic words at AP-initial position, which showed the AP-initial rising tone pattern (data from Kim, 2004). The degree of F0 increase in this condition was also similar to that obtained from a quantitative study conducted by Jun (1995b), which showed that the average F0 value of word-final (and AP-final, simultaneously) syllables was approximately 11.6% higher than that of the penultimate syllables.

Finally, ***Pitch initial*** condition had a high pitch on the first syllable of every word (13% increase rate, as in ***Pitch final*** condition). In Korean, the laryngeal feature of the AP-initial segments determines the AP-initial tone (see Jun, 1993; Jun, 2000). A tense or an aspirated consonant ([+stiff vocal cords]) induces a high tone in the AP-initial position. When a lenis or a nasal consonant is in the AP-initial position, as in the current experiment, the consonant bears a low tone. Therefore, ***Pitch initial*** condition does not conform to the prosodic patterns of the AP.

It was expected that three conforming prosodic conditions (***Duration***, ***Amplitude***, ***Pitch final***) would help listeners to segment words from speech stream. If it is the case, listener's performance in the three conforming conditions will be better than that in ***No prosody*** condition. If the exploitation of a prosodic cue in word segmentation is controlled by language-specific characteristics of one's native language (as are the exploitation of other cues for segmentation such as metrical patterns (Cutler, Demuth & McQueen, 2002; Cutler & Norris, 1988; Cutler & Otake, 1994) and phonotactics (Weber, 2001), it is possible that the non-conforming prosodic condition (***Pitch initial***) would hinder speech segmentation because of the conflict between the distributional cue (i.e., transitional probability) and the prosodic cue (word-initial high pitch, rather than word-final high pitch). Alternatively, if any prosodic property can aid segmentation, as long as the cue is perceptually salient enough to draw listeners' attention and is appropriately combined with the correct distributional information, then, the ***Pitch initial*** condition would help Korean listeners' lexical segmentation.

2.4 Procedure

There were twelve participants for each of the five prosodic conditions (***No prosody***, ***Duration***, ***Amplitude***, ***Pitch final***, and ***Pitch initial***). Each of the sixty participants was a native speaker of Seoul Korean, and each was paid for participating. 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 nonsense 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, no information about the number of words in the language was given to subjects.

The speech stream was presented to subjects using the Sound Edit program in a Macintosh computer, and subjects heard the concatenated sounds through headphones. They were asked to adjust the volume to the most comfortable level. The learning phase lasted approximately 11 minutes for the ***Duration*** condition and approximately 10 minutes 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 "non-word" test strings. The trisyllabic non-word test strings were comprised of three part-word and three non-word strings that had not been included in the original word list for the artificial language. The part-words contained a two-syllable string from the end of a real word of the language concatenated with an additional syllable. All of the three part-word strings used in the testing phase occurred 42 times each during the learning phase. Non-word strings were composed of syllables that were used in the learning phase, but their syllables occurred in a sequence that the subjects never heard during the learning phase. Thus, the transitional probabilities for these 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 of which was a test string. There was an 800 ms inter stimuli interval between the two strings in a pair. All of the strings presented in the testing phase had the same prosodic cues as those presented in the learning phase. After listening to the two alternatives, subjects were asked to identify which of the two strings was a word from the artificial language. Subjects entered their responses by pressing the '1' key on the keyboard if the first string presented was thought to be the real word from the artificial language or the '0' key if the second string presented was thought to be the real word from the artificial language.

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

3. Results

The average percentage of correct identification was 56.6% (mean raw score 20.4 out of 36, SD = 4.4) for the ***No prosody*** condition, which was higher than chance level (50%). One third of the participants in this condition, however, performed at chance or below. The ***Duration*** condition showed the highest average percentage for correct identification with 83% (mean score 29.9, SD = 4.7), followed by 76.3% (mean score 27.5, SD = 5) for the ***Amplitude*** condition. All of the participants for these two conditions performed above chance levels. In the ***Pitch*** final condition, all of the listeners performed above chance, except for one, whose performance was at chance level. The average correct percentage for this condition was 67.5% (mean score 24.3, SD = 4.6). The average correct percentage for the ***Pitch*** initial condition was 49.9% (mean score 18, SD = 5.1), which was the lowest level of performance of all the conditions. Half of the participants in this condition performed at chance or below. These results are illustrated in Figure 1.

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

As described above, each syllable used for the stimuli was produced naturally, rather than through speech synthesis. Thus, although every effort was made to normalize the acoustic values of the tokens, some phonetic variation was unavoidable. For instance, there were differences in consonant and vowel lengths between the lenis consonants and the nasal consonant, and some minor deviations from mean values existed in the stimuli. The purpose of the within subject analysis was to ensure that there were no artifacts in the results that were attributable to the method by which the stimuli were created.

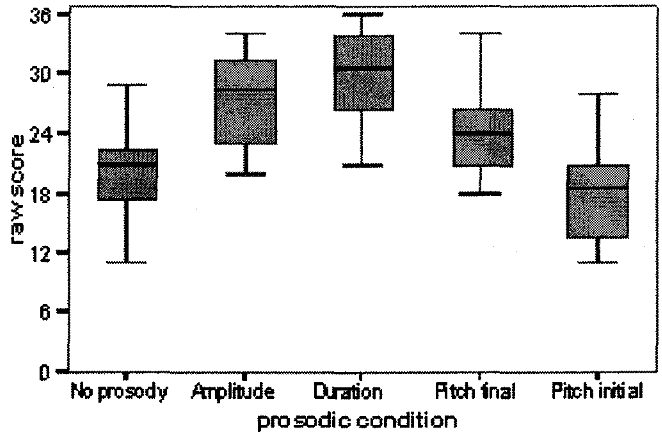

Figure 1. Correct identification (raw score) in five prosodic conditions. The raw score of 18 is chance level (50%). The black line within each box indicates the mean raw score of the corresponding prosodic condition

In this RM ANOVA analysis, the sphericity assumption was met, and, thus, degree of freedom was not adjusted. The results showed that there was no main effect of *word*, suggesting that all of the words were equally difficult for the listeners to process. 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.*

Another RM ANOVA was performed with *test strings* (part-words and non-words) as a within subject factor and *prosodic condition* as a between subject factor, in order to investigate whether the listeners' performance varied with the transitional probabilities of the test strings (i.e., part-words, with lower transitional probabilities than words vs. non-words, with zero transitional probabilities). There was no significant difference in the correct response rate between part-word strings and non-word strings. Again, the *prosodic condition* effect was significant ($F(4, 55) = 12.665$, $p < .001$), and no interaction was found between *test strings* and *prosodic condition.*

In these two repeated measures ANOVAs, the effect of the between subject factor, *prosodic condition,* was highly significant, indicating that the presence or absence of prosodic cues affected segmentation and storage of the new words. Pair-wise comparisons showed that listeners performed significantly better in the ***Duration***, ***Amplitude***, and ***Pitch final*** conditions than they did in the ***No prosody*** condition ($p < .001$, $p = .001$, $p = .04$, respectively). There was no difference between the ***Duration*** and ***Amplitude*** conditions. However, listeners' performance was significantly better in the ***Duration*** condition than in the ***Pitch final*** condition ($p = .006$). The ***Amplitude*** and ***Pitch final*** conditions were not significantly different from one other. Finally, listeners' performance was significantly worse in the ***Pitch initial*** condition than in the ***Duration***, ***Amplitude***, and ***Pitch final*** conditions ($p < .001$, $p < .001$, $p = .002$, respectively), but there was no difference between the ***Pitch initial*** and ***No prosody*** conditions.

4. Discussion

The average performance of the listeners of the ***No prosody*** condition (56.6 %) was slightly higher than chance level. However, Korean listeners who participated in this experiment did not perform as well as English listeners in Saffran et al. (1996) study or French listeners in the Bagou et al. (2002) study. English listeners showed a 65% average correct performance rate when the testing phase had real words vs. part word pairs and a 76% average correct performance rate when the testing phase had real word vs. non-word pairs in the condition where no prosodic information was available. French listeners also showed about 65% mean performance when there was no prosodic cue in the speech input.

No conclusive explanation can be provided to account for this difference, because a direct comparison between Korean, English and French listeners was not undertaken, and the experimental stimuli and procedures were not identical across the studies. Nevertheless, one plausible reason for the low performance of the Korean listeners', in comparison to the French or English listeners of previous studies, could be the lack of allophonic cues available to the listeners in the stimuli for this experiment. In Korean, lenis stop voicing always occurs within an AP, and, thus, within a word (Jun 1993). However, in the artificial language used in this study, all of the lenis stops were voiceless, because every syllable was produced independently. This pattern contradicts the allophonic variation of lenis stop in Korean, and may, therefore, have interfered with listeners' ability to encode the artificial language's words. That is, the nature of the acoustic input may have resulted in a higher overall level of difficulty for the language-learning task in this study, as compared to previous investigations. However, additional experiments are required to validate this hypothesis.

The results revealed that the three conditions conforming to the prosodic characteristics of the Korean AP, i.e., ***Duration*** (AP-final lengthening), ***Amplitude*** (AP-initial strengthening) and ***Pitch final*** (AP-final high tone), facilitated word segmentation, since performance was significantly better in these conditions than in the baseline ***No prosody*** condition. Thus, we can conclude that each conforming prosodic cue, when combined with distributional information (transitional probability), can efficiently aid speech segmentation of Korean.

The statistical difference observed between the ***Duration*** and ***Pitch final*** conditions also suggests that the degree to which these cues contribute to word segmentation differs. Specifically, the 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 several ways. One possible interpretation involves Korean's prosodic structure. That is, it is possible to surmise that the observed difference was caused by the fact that the lengthened duration was sufficient to mark a prosodic unit that is, in fact, higher than the AP in Korean, i.e., the Intonational Phrase, whereas the heightened F0 value was sufficient only to mark the AP. It is well-known that speakers' speech production is governed by the prosodic structure, and that the acoustic and articulatory features are enhanced or reduced depending on the level of the phrase in the prosodic hierarchy (Cho & Keating, 2001; Fougeron & Keating, 1997; Keating, Cho, Fougeron & Hsu, 2004). If the acoustic properties from speech production form part of the basis for speech perception (Klatt, 1976), it is conceivable that the prosodic structural information realized in the speech input could be exploited in speech perception as well. 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 would rely simply on the 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 respect to the current experiment, it may be the case that the degree of lengthening (30%) was psycho-acoustically stronger than that of F0 increase (13%), thereby causing a larger perceptual effect for duration than for pitch raising. If the second interpretation is correct, then the difference in facilitation would not necessarily imply any additional knowledge on the part of the listeners of the acoustic features of Korean prosodic structure.

Additional experiments are required in order to decide which of these two interpretations is accurate. However, before any additional perception experiments are conducted, it will be necessary to obtain sufficient quantitative production data to establish the accurate full range of acoustic values for each of the relevant prosodic cues in the various prosodic positions.

The non-conforming condition in this experiment, the ***Pitch initial*** condition with higher pitch on the first syllable, showed a mean correct rate of 49.9%, with half of the participants performing below chance level. This indicates that this cue adversely impacted listeners' ability to segment words in the artificial language. Listeners' performance in this condition was poorer than the performance of listeners in the ***No prosody*** condition where no prosodic cue was available at all; however, the two conditions were not significantly different. This prosodic cue is present in some languages (for example, most content words have a trochaic stress pattern in English and Dutch), but it is not a part of the native prosody of Korean. Additionally, this condition violates a tone-segment mapping rule of Korean (Jun, 2000; Jun, 2004), by mapping a lenis/nasal consonant to a high tone. Thus, it seems likely that listeners may have attempted to match the high F0 to the last syllable of a word rather than the first syllable of a word. This attempt would lead to a mismatch between the two different sources of information, i.e., the distributional information present in the stimuli and the prosodic information, and, consequently, would account for the fact that listeners performed at chance levels in the ***Pitch initial*** condition.

The comparison between the ***Pitch final*** and ***Pitch initial*** conditions indicates that the same prosodic cue can have a 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 a particular language. Further, the result suggests that it is not the case that listeners automatically respond to any perceptually salient prosodic cue for speech segmentation. Rather, this result reveals listeners' strong dependency on the cues that conform to the prosodic structure of their native language for word segmentation.

5. Conclusion

This study examined whether and how acoustic correlates of post-lexical prosodic structure are exploited in lexical segmentation, using an artificial language learning paradigm. The results of the experiment indicate that phrase-level prosodic properties can play an important role in word segmentation and as short-term storage of lexical items. All of the prosodic properties that are associated with the AP (i.e., the duration cue at the end of the phrase, the amplitude cue at the beginning of the phrase, and the pitch cue at the end of the phrase) facilitated segmentation, while a consistent perceptible cue that does not conform to Korean AP properties did not aid the process. These results therefore suggest that the post-lexical prosodic cues that facilitate word segmentation are not merely any predictable acoustic cues, but rather they are the language-specific prosodic patterns and structure of the given population of listeners.

References

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

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

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

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

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.

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

Cutler, A., Demuth, K. & McQueen, J. M. 2002. "Universality versus language-specificity in listening to running speech." *Psychological Science*, 13, 258-262.

Fougeron, C. & Keating, P. A. 1997. "Articulatory strengthening at edges of prosodic domains." *Journal of 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*, 547-567

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

Jun, S. 1993. *The Phonetics and Phonology of Korean Prosody*. Unpublished Doctoral dissertation, The Ohio State University.

Jun, S. 1995a. "Asymmetrical prosodic effects on the laryngeal gesture in Korean." In B. Connell & A. Arvaniti (Eds.), *Phonology and phonetic evidence: papers in laboratory phonology IV*. 235-253. Cambridge, United Kingdom. Cambridge University Press.

Jun, S. 1995b. *A Phonetic Study of Stress in Korean*. Paper presented at the 130th meeting of the Acoustical Society of America, St. Louis, Mo.

Jun, S. 2000. "K-ToBI (Korean ToBI) Labelling Conventions", Ver. 3.1.

Jun, S. 2004. "Korean Intonational Phonology and Prosodic Transcription." in S. Jun (ed.), *Prosodic Typology and Transcription: A Unified Approach*. Oxford University Press.

Jun, S. & Fougeron, C. 2000. "A phonological model of French intonation." in A. Botnis (ed.), *Intonation: Analysis, Modelling and Technology*. Kluwer.

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

Keating, P. A., Cho, T., Fougeron, C. & Hsu, C. 2004. "Domain-initial articulatory strengthening in four languages." *Laboratory Phonology*, 6, 143-161.

Kim, S. 2004. "The role of prosodic phrasing in word segmentation of Korean." Unpublished doctoral dissertation, University of California, Los Angeles.

Klatt, D. H. 1976. "Linguistic uses of segmental duration in English: Acoustic and perceptual evidence." *Journal of Acoustical Society of America*, 59, 1208-1221.

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

Nakatani, L. H. & Dukes, K. D. 1977. "Locus of segmental cues for words juncture." *Journal of Acoustical Society of America*, 62 (3), 714-719.

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

Quene, H. 1993. "Segment durations and accent as cues to word segmentation in Dutch." *Journal of Acoustical Society of America*, 94 (4), 2027-2035.

Saffran, J., E. Newport, L. & Aslin, R. N. 1996. "Word segmentation: the role of distributional cue." *Journal of Memory and Language*, 35(4), 606-621.

Weber, A. 2001. *Language-specific listening: the case of phonetic sequences*. Unpublished Ph.D. Dissertation, Max Planck Institute for Psycholinguistics, Nijemegen, the Netherlands.

received: January 30, 2006
accepted: March 12, 2006

▲ Sahyang Kim
Dept. of Audiology and Speech-Language Pathology, Wayne State University
207 Rackham Memorial Building, 60 Farnsworth Ave., Detroit, MI 48201., U.S.A.
Tel: +1-248-990-4391
E-mail: sahyang@gmail.com