UNIVERSITY OF CALIFORNIA

Los Angeles

# The Role of Prosodic Phrasing in Korean Word Segmentation

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy
in Linguistics

by

Sahyang Kim

2004

The dissertation of Sahyang Kim is approved.

Patricia Keating

Colin Wilson

Jody Kreiman

Sun-Ah Jun, Committee Chair

University of California, Los Angeles

2004

*To my parents.*

# TABLE OF CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# ACKNOWLEDGEMENTS

At last I am here, closing another chapter of my life. Before moving on to the next chapter, I would like to leave a little thank-you note to everyone who helped me to complete this part of my story.

I am deeply indebted to my advisor, Sun-Ah Jun, who introduced me to phonetics and intonation. I benefited greatly from her invaluable guidance and advice at every stage of this dissertation. I wish to thank her for encouraging me, having confidence in me, and being my true mentor. She was always generous with her time, and she made me feel comfortable talking to her about any ideas I had regarding both linguistic and non-linguistic issues. I cannot thank her enough for the tremendous amount of academic counsel and moral support she provided over the past six years.

I would like to thank Patricia Keating, who has eyes that can see things that my eyes cannot. Her comments and insights improved my dissertation enormously. I thank her for her crucial input on the design of the perception experiments and for her questions on various other parts of my dissertation. I am grateful to her for generously taking the time to correct my articles. I also wish to thank her for her constructive advice and moral support in my job search.

I am grateful to Colin Wilson who supplied valuable feedback on this manuscript, as well as guidance with the experimental details. I thank him for his critical reading, his constant encouragement, and for his time. I am also grateful to Jody Kreiman for her help with statistical issues and for her encouragement through the final stages of this dissertation.

I would also like to thank other faculty members of the UCLA Linguistics Department who aided me at various points during my graduate student years: Ed Keenan, Peter Ladefoged, Anoop Mahajan, Pam Munro, Carson Schutze, and Donca Steriade.

I would like to thank Virgil Lewis, my Pima teacher, who taught me not only his language, but also the importance of many invisible things, as well.

I would like to thank Professor Young-joo Kim, my former advisor at Hong-Ik University and my role model throughout my college years. It was her 'Introduction to Language' course that led me to pursue my career in linguistics. Her direction and encouragement did not cease even after I left Korea. I thank her for all the help and support that she has given me at every crucial step in my career.

I would also like to express my gratitude to my friends and colleagues who made my life richer and more colorful in many ways.

I thank Motoko Ueyama and Marco Baroni for many adventures in La La Land during my first three years in this city and for their friendship. I thank Haiyong Liu for his moral support and for all the postcards he sent me from the various places he visited. Thanks also go to Robin Huffstutter for postcards, souvenirs and hilarious emails from Poland. I would like to thank Shabnam Shademan and David Ross for their open door, open boat, open hearts, and encouragement. I would also like to thank Adam Albright for countless instances of timely assistance with anything and everything and Angelo Mercado for *Requiem*, which kept me going during the last two months of writing this dissertation. I thank Christina Esposito and Rebecca Scarborough, who always made me smile, for their moral support. I would also like to thank Leston Buell and Harold

Torrence for their senses of humor and their *annyung* and *komapsupnidayo*. My thanks also go to Ivano Caponigro, who shares my feelings towards Caetano and 'Happy Together', and to Jie Zhang, my official BF. I am also grateful to Tim and Heidi Arbisi-Kelm, Mary Baltazani, Patrick Barjam, Roger Billery-Mosier, Lumme Erilt, Heidi Fleischaker, Alex Hurst, Brook Lillehaugen, Ying Lin, Alex Macbride, Kuniko and Jon Nielson, Marcus Smith, Minjung Son, and Mariam Sy for their moral support, camaraderie, and encouragement.

My special thanks go to Taehong Cho, who helped me in every possible way to settle in and acclimate myself to my new life in a foreign land. Without his counsel, I know that I would be driving a lemon by now, instead of my sturdy *seveki*. His constant advice, support, and encouragement also helped me overcome many obstacles during my grad student years. I am also thankful to his wife, Hyejeong Han, for her warm-heart, moral support, and delicious Korean food. I would like to let them know how grateful I am, too, for the three Christmas Eves that I was privileged to spend with their family.

I would like to thank my dearest friend, Amanda Jones. She was everything that one could expect from a friend. She always had time for me, she always lighted up my heart, and she always supported me and helped me with everything. I don't know how I could have gone through six years of turmoil without her. I also thank her and Kurt Muehlner for unforgettable Thanksgiving dinners. Thanks, Jo, for being there.

I would like to thank Heriberto Avelino, with whom I shared most of my time, the best laughs, the best fights, the best academic and non-academic discussions, and the coolest office in the ling dept. I cannot even say a word about my life at UCLA without mentioning his name. Betito was my officemate, lunch mate, dinner mate, coffee mate (!), exercise buddy, *toronja* break buddy, music buddy, movie buddy, world travel buddy, coworker, and my friend. I am grateful for the time we spent together, the space we shared together, the numerous projects on which we collaborated, and all the memories we will look back on together. I also thank Adriana Cruz for spending so much time enjoying so many adventures with me. I would also like to thank her for the interesting discussions on immigration and identity issues and for *sopa de camarones*. *Betito y Barbarita, gracias por compartir sus vidas diarias y su cultura conmigo.*

I would like to thank the people who helped me with the experiments for this dissertation. My thanks go to all the Korean subjects who participated in the experiment, to Bruce Hayes and Jill Gilkerson who generously gave me priority to use the sound booth when there were time conflicts, and to Henri Tehrani who always helped me whenever I had technical difficulties. I would like to thank Heesun Kim and Hyuckjoon Lee for agreeing to the subject pool idea, for lending me their equipment, for helping me set up the experiments, for recruiting half of the subjects, and for running my experiments at Stanford University.

My thanks also go to the staff members of the UCLA Linguistics Department who made all the bureaucratic work easier for me: Kate Goodyear, Mary Hoang, Jami Jasek, Natasha Levy, Patrick Manalastas, Kathryn Roberts, and Matt Tyler.

I would also like to thank my way-too-qualified-proofreaders: Adam, Amanda, and Christina. Thank you for your patience, time and assistance. I owe you all a lot.

Finally, I would like to thank my family and friends, who have been out of sight, but very much in mind. I am grateful to Hyejin Park for her friendship over the past 16 years. I thank her for being my private record-keeper and for standing by me all the time. I also thank Wookyung Kwon for her constant help and encouragement. Their voices over the phone were soothing enough to let me forget all my troubles.

Special thanks go to my sisters, Sagang and Sarim Kim. Sagang has lived with me in L.A. for the last three years, and despite her own busy schedule as a grad student, she has helped me enormously in so many ways. She has taken care of all the house chores, prepared my lunch while I pretended to be busy with my dissertation work, and she has put up with me in general. Sarim Kim has delivered innumerable goodies from home and always treats her poor elder sisters to fancy dinners when she comes to L.A. on business trips. For these things and more, I am grateful to my sisters.

I would like to express my deepest gratitude to my parents for their unconditional love and faith in me through all these years. I thank them for letting me choose and walk my own path. Especially, I thank my mom and my grandmothers for their endless prayers for me.

Finally, I would like to thank the UCLA Linguistics Department for financial support over the past six years. Thanks go to Professors Sun-Ah Jun and Patricia Keating for various research assistantships and to the UCLA Graduate Division for my Dissertation Year Fellowship.

# VITA

| | |
|---|---|
| September 20, 1972 | Born, Seoul, Korea |
| 1995 | B.A. in English Education<br>Hong-Ik University, Seoul, Korea |
| 1995-1998 | M.A. in English Language and Literature<br>Hong-Ik University, Seoul, Korea |
| 1998-2001 | M.A. in Linguistics<br>University of California, Los Angeles |
| 1998-2003 | Graduate Research Assistant<br>Department of Linguistics<br>University of California, Los Angeles |
| 1999-2003 | Graduate Teaching Assistant<br>Department of Linguistics<br>University of California, Los Angeles |
| 2002-2003 | UCLA Dissertation Year Fellowship |

# PUBLICATIONS AND PRESENTATIONS

Kim, Sahyang. 2000. Post obstruent tensing in Korean: Its status and domain of application. 140th Meeting of Acoustical Society of America. Newport Beach, CA.

___. 2002. Passives in Pima.Workshop on American Indigeneous Languages. Santa Barbara, CA.

___. 2002. The extension of Pima reflexive morphemes. The 7th Annual Workshop on Structure and Constituency in the Languages of the Americas (WSCLA 7). Edmonton, Canada.

___. 2003. Post obstruent tensing in Korean: Its status and domain of application. In P. Clancy (ed.), *Japanese/Korean Linguistics*, Vol 11, Center for Study of Language and Information, Stanford University, CA.

___. 2003. Domain initial strengthening of Korean fricatives. In S. Kuno et al. (eds.), *Harvard Studies in Korean Linguistics IX.* Hanshin Publishing Co. Seoul, Korea.

___. 2003. The role of post-lexical tonal contours in word segmentation. *Proceedings for 15th International Congress of Phonetic Sciences*, Barcelona, Spain.

___. 2004. The role of prosodic phrasal cues in word segmentation. 2004 Annual Meeting, Linguistic Society of America. Boston, MA.

___. , and Heriberto Avelino. 2003. An intonational study of focus and word order variation in Mexican Spanish. El Coloquio Internacional. *La tonía: dimensiones fonéticas y fonológicas*. Mexico City, Mexico.

Avelino, Heriberto and Sahyang Kim.2004. Variability and Constancy in Articulation and Acoustics of Pima Coronals. *Proceedings for 29th Berkeley Linguistics Society*, University of California at Berkeley, Berkeley, CA.

# ABSTRACT OF THE DISSERTATION

The Role of Prosodic Phrasing in Korean Word Segmentation

by

Sahyang Kim

Doctor of Philosophy in Linguistics

University of California, Los Angeles, 2004

Professor Sun-Ah Jun, Chair

This dissertation investigates the role of prosodic phrasing in word segmentation in Korean. I hypothesized that the Accentual Phrase, a post-lexical prosodic unit with phrase edge prominence, would play a crucial role in Korean word segmentation. Before exploring this hypothesis, I conducted a corpus study to corroborate existing beliefs regarding the linguistic attributes of the Korean AP. Results of the corpus study confirmed that most APs contain one content word and that 90% of content words are located in AP-initial position. Additionally, results showed that 88% of multisyllabic content words begin with a rising tone in AP-initial position, and 84% of all APs end with a final high tone. Thus, it was expected that detection of AP boundary cues and tonal characteristics would directly facilitate Korean word segmentation. Two perception experiments were performed in order to determine whether these regularities in AP tonal patterns and boundary cues help Korean listeners segment words from the speech stream. The first experiment employed a word-spotting task and examined whether some post-lexical tonal patterns of the AP facilitated segmentation more than others and whether listeners' preference for particular tonal patterns was determined by the observed frequencies of these patterns, as established by the corpus study. Results showed that the rising tonal pattern at word onset and fixed edge prominence (i.e., AP-final H tone) facilitated Korean listeners' word segmentation. The second perception experiment employed an artificial language learning paradigm and explored whether prosodic cues at the boundaries of Korean APs would affect word segmentation and whether some boundary cues would prove more influential in segmentation than others. Results of this study demonstrated that boundary cues conforming to native-language prosody (i.e., AP-final lengthening, AP-final pitch rise, AP-initial amplitude effects) facilitated word segmentation, but that a boundary cue that was regular and salient but that did not conform to native-language prosody (i.e., phrase initial pitch rise) did not facilitate segmentation. Overall, these results suggest that exploitation of post-lexical intonational cues in Korean word segmentation is determined by frequency regularities for tonal patterns imposed upon words and that exploitation of AP boundary cues for segmentation is language-specific.

# CHAPTER 1

# INTRODUCTION

## 1.1 The Word Segmentation Problem

Human beings are exposed to continuous streams of speech in their everyday language experience. In order to process and understand spoken language successfully, they must be able to segment these streams of speech into smaller discrete units by detecting the correct boundaries for each unit. Although segmentation is accomplished without any conscious effort on the part of the native listener, the task is by no means trivial. In most cases, speech input often lacks clear audible pauses between words, there is no single acoustic cue that constantly marks word boundaries, and there are always coarticulation effects when words are produced within a larger speech context. Thus, the question remains as to what kind of information helps listeners to find the correct word boundaries despite these apparent complexities in the speech input.

One approach to the word segmentation problem is to consider the process as a by-product of lexical competition, as suggested by the connectionist models of word recognition, such as TRACE (McClelland & Elman, 1986) and Shortlist (Norris, 1994). According to these models, word recognition is achieved on-line by competition between candidate words, and the word segmentation problem can be naturally and reliably resolved during the process of lexical competition. For instance, the speech input '*ice* *cream*' activates candidate words such as *eye*, *I*, *ice*, *scream*, *cream*, *ream*, and so forth. When lexical candidates overlap, such as *ice* and *scream*, they inhibit one another: hence, the parsing of [aɪskrim] into '*ice scream*' would be prevented while the parsing into '*ice cream*' would not. This treatment, however, is not enough to provide a seamless account of word segmentation. First, assuming that word boundaries emerge through competition among many activated words implies that listeners have already established their mental lexicons. Thus, the 'segmentation through competition' approach cannot explain how infants or second-language learners segment words from speech input and how they find their first word in the absence of enough top-down knowledge at the initial stage of language learning. Second, there are cases where competition by itself cannot provide any solution to ambiguity in word segmentation. For instance, '*I scream*' and '*ice cream*' (Lehiste, 1960) have the same segmental sequence, and hence, they would activate the same set of candidate words. Thus, lexical competition mechanisms relying only on phonemic representations could result in two parsing possibilities for the speech stream of [aɪskrim], but would still not be able to conclude which of the two conforms to the speaker's intention, especially when no other semantic and/or pragmatic context is given.

This indicates that in addition to making the best of the lexical competition process, adult listeners need to exploit other cues in speech. Indeed, several studies have shown that the word boundary ambiguity that arises in the above example can be solved with the help of acoustic information in the speech input (Davis, Marslen-Wilson, & Gaskell, 2002; Nakatani & Dukes, 1977; Quene, 1993). Furthermore, the integration of other segmentation cues in a word recognition model significantly improves the general performance of the model. Norris, McQueen, and Cutler (1995) demonstrated that the incorporation of a metrical segmentation cue (i.e., Metrical Segmentation Strategy (Cutler & Norris, 1988)) into a competition-based word recognition model (i.e., Shortlist (Norris, 1994)) resulted in simulation results that were closer to the empirical results gained from previous behavioral studies of word segmentation and recognition. A later study by Norris, McQueen, Cutler and Butterfield (1997) claimed that the Possible-Word Constraint, a constraint that bans parsing which leaves impossible words (e.g., unsyllabified consonants), gave more robust simulation result when implemented in the Shortlist model. They further showed that recognition was even more improved when more segmentation cues, such as pause, metrical structure, phonotactics, and acoustic cues, were available to the system.

These studies strongly suggest that the types and nature of segmentation cues that listeners exploit should be explicitly investigated independent from, as well as associated with, the lexical competition process, and these cues should be integrated with spoken word recognition models, in order to achieve the optimal result in speech processing. Moreover, the question of what kinds of information can be utilized as segmentation cues must be examined on a language-by-language basis, because the speech cues that are exploited for word segmentation, such as metrical cues and phonotactic cues, are sensitive to the phonological constraints and/or structure of a given language (Cutler, Demuth, & McQueen, 2002; Cutler & Norris, 1988; Cutler & Otake, 1994; McQueen, 1998; Mehler, Dommergues, Frauenfelder, & Segui, 1981; Sebastian-Galles, Dupoux, Segui, & Mehler, 1992; Weber, 2001). Studies of language-specific segmentation cues will elucidate how the mental lexicon starts to be established at an early stage of language acquisition, and how people parse speech input and recognize words in each individual language. The accumulation of results from such studies will also provide valuable insights into the universal properties of speech processing.

## 1.2 Research Overview

Previous studies have shown that speech input contains various cues that aid word segmentation, and that listeners do exploit these cues for word segmentation (see Section 2.1 for details). Lexical prosody is known as one of the most robust cues for segmentation. It is conceptually very plausible to hypothesize that the existence of lexical-level prominence (i.e., lexical stress, pitch accent, tone) and its acoustic correlates would aid word segmentation and recognition. Indeed, this hypothesis has been tested and proved in several languages (see reviews in Section 2.1.1). However, lexical prosody is not available in all languages; some languages simply do not have lexical-level prominence, like Korean. If a language lacks lexical prominence, what kinds of prosodic information can be used as cues for word segmentation? This question is the starting point of the current study.

The purpose of this dissertation is to seek language-specific prosodic cues that facilitate word segmentation in Korean. The first step towards this goal is to look for a word-size prosodic unit that contains sufficient prosodic information. Based on previous research on Korean prosody, I hypothesize that the Accentual Phrase, a post-lexical prosodic unit with phrase-edge prominence (see Section 2.2 for details), would be a good candidate for such a unit.

The organization of this dissertation is as follows: In order to test the validity of this initial hypothesis and to find potential cues that can efficiently aid word segmentation of Korean, a corpus study is conducted. Chapter 3 will report the results of the corpus study. The corpus study provides detailed hypotheses for two perception experiments. Chapter 4 presents a perception experiment that employs a word-spotting task (Cutler & Norris, 1988; McQueen, 1996). The study examines whether certain post-lexical tonal patterns of the AP facilitate segmentation more than others, and whether listeners' preference for particular tonal patterns is determined by their frequencies observed in the corpus study. Another perception experiment that employs an artificial language learning method (Saffran, Newport, & Aslin, 1996) will be reported in Chapter 5. The experiment explores whether the prosodic cues at the boundaries of the Korean AP affect word segmentation, and whether all the boundary cues tested are equally influential to segmentation. Finally, Chapter 6 will summarize and discuss the findings of this dissertation, state the limitations of the experiments and propose future studies.

# CHAPTER 2

# BACKGROUND

## 2.1 Cues for Word Segmentation

Researchers interested in segmentation seem to agree implicitly that there is no perfect cue for word segmentation in the speech input. That is to say, there is no ‘one and only’ cue which exerts a decisive influence on the initial segmentation process, and there is no cue that can be an absolute word-boundary indicator by being constantly available in any circumstances. Numerous studies that have sought to illuminate cues facilitating word segmentation have revealed that accurate segmentation could be aided by multiple means, such as prosodic, acoustic, phonotactic, and distributional cues.

### 2.1.1 Prosodic Cues

Lexical stress is probably the most discussed and the most widely studied word segmentation cue, within both adult and infant populations. The seminal behavioral study was performed by Cutler and Norris (1988), using a word-spotting task. In the experiment, native listeners of English were asked to detect a monosyllable real word of English embedded in a speech string that was composed of two syllables. The result showed that the listeners were faster in detecting *mint* in [mɪntəf] with strong-weak stress pattern than *mint* in [mɪntef] with two strong syllables (Cutler & Norris, 1988). They claimed that English listeners have a tendency to put a word boundary in front of a strong vowel. According to this claim, it was more difficult for the listeners to spot the word *mint* in a strong-strong syllable sequence than in a strong-weak syllable sequence since the word boundary was inserted before the strong syllable in [mɪntef] (such that the syllables are parsed as [mɪn.tef][1]), and hence [t] was treated as the onset of the following syllable. English listeners' missegmentation errors also supported this claim (Cutler & Butterfield, 1992). Both real speech perception errors and experimentally induced perception errors were shown to stem from listeners' tendency to insert word boundaries before strong syllables (e.g., 'into opposing camps' was perceived as 'into a posing camp') and to overlook the boundaries before weak syllables (e.g., 'my gorge is' was perceived as 'my gorgeous'). 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[2]. Another language that reveals a similar metrical segmentation pattern as English is Dutch. As for input frequency, 90 % of Dutch words start with a stressed syllable (Quene, 1992). Dutch listeners were faster in visual lexical decision for target words when words were embedded in a Strong-Weak stress pattern (e.g., *melk*

[1] '.' marks a syllable boundary.

[2] "Strong syllables" included secondary as well as primary stress.

‘milk’ in [mɛlkəm]) than when they were in a Strong-Strong stress pattern (e.g., *melk* in [mɛlka:m]) (Vroomen & de Gelder, 1995).

The role of lexical stress in segmentation was also emphasized in speech perception studies. Nakatani and Schaffer (1978) recorded trisyllabic adjective-noun phrases, and replaced segmental information with reiterant syllables (e.g., “new result” becomes “ma mama” and “foolproof lock” becomes “mama ma”). They tested if the suprasegmental information remaining in the input could help listeners to parse the speech input correctly. They found that stress pattern (i.e., trochaic metrical structure) and rhythm (i.e., pre-boundary duration) were useful prosodic cues for word perception in English, whereas pitch (i.e., F0) and amplitude were not. Quené (1993) presented two-word phrases with an ambiguous word boundary to Dutch listeners, and investigated the relative contributions of accentuation, duration of pivotal consonant (i.e., the consonant that can be the coda of the first word or the onset of the second word, for example, [s] in the speech string [aɪskrim] ‘ice cream’), and post-boundary vowel duration to word segmentation. Results showed that only word-initial ‘accent’ cues (including acoustic cues such as F0, intensity and duration), and the word-initial consonant duration cues were able to facilitate segmentation.

This prosodic information of lexical stress seems to be available from a very early stage of language acquisition. Previous studies have shown that infants ‘know’ the difference between native language prosody and non-native prosody (Mehler et al., 1988; Nazzi, Bertoncini, & Mehler, 1998; Nazzi, Jusczyk, & Johnson, 2000), even when they are 2 days old (Moon, Cooper, & Fifer, 1993). Studies have also shown infants' sensitivity to the language-specific stress pattern of English: 9-month-old American infants prefer to listen to a list of words with strong-weak syllables, compared to a list with weak-strong syllables, regardless of the presence or absence of segmental information (Jusczyk, Cutler, & Redanz, 1993), and they perceived novel bisyllables as cohesive when the stimuli have a trochaic pattern (Morgan, 1996). Echols, Crowhurst, and Childers (1997) found that both adults and 9-month-old American infants tended to consider a trochaic pattern coherent. Adults were better in detecting post-stress pauses than pre-stress pauses in trisyllabic nonsense words, and infants were better at recognizing a strong-weak pattern than a weak-strong pattern after they had been equally trained with three syllable weak-strong-weak words. These studies further suggested that the preference for a trochaic foot pattern could be used by infants in the English environment in segmenting words from the speech input.

Listeners are also sensitive to prosodic cues related to higher-level prosodic units and they exploit these cues for word boundary detection. The most prominent cue for word segmentation is, if present, a pause following a prosodic phrase. An audible pause can be a reliable cue for the major phrase boundary, and hence can facilitate speech segmentation. Hirsh-Pasek, Kemler Nelson, Jusczyk, Wright Cassidy, Druss, and Kennedy (1987) and Gerken, Jusczyk, and Mandel (1994) found that infants were sensitive, not only to the presence of pause itself, but to the location of the pause in the sentence. Infants would listen longer to the utterances when the pauses coincided with major syntactic phrases, than those that contained mismatched pauses. However, pause itself cannot be a very reliable cue for a word boundary because speakers usually do not put a pause after every word.

Phrase-boundary cues may or may not be as noticeable as pause, but they can facilitate speech segmentation. Christophe, Dupoux, Bertoncini, and Mehler (1994) found that French infants and adults could discriminate the difference between a bisyllable ($C_1V_1C_2V_2$) extracted from a long word (e.g., [mati] from 'cli**mati**se') and the same sequence of syllables extracted from two consecutive words (e.g., [mati] from 'panora**ma ty**pique'), where a phonological phrase boundary can possibly occur. Christophe, Mehler, and Sebastian-Galles (2001) performed a similar experiment with French newborns and adults. This time, they presented Spanish stimuli from the same environment (within and between words, with the stress on the second syllable) to newborns and found the same result as before. It is not hard to imagine that there must have been abundant prosodic cues at the prosodic boundary. In fact, they reported significant differences in $V_1$ duration and stop closure of $C_2$ in the French speech, and $C_2$ duration and pitch differences within words vs. between words in the Spanish speech.

More recently, Christophe, Gout, Peperkamp, and Morgan (2003) reported two studies that demonstrated how the presence of a phonological phrase boundary facilitated word segmentation. In the first study (Christophe, Peperkamp, Pallier, Block, & Mehler, in revision), adult French listeners were given sentences that contained a local ambiguity (e.g., [d'un **chat grin**cheux]$_{PP}$ 'of a grumpy cat', ambiguous with the word *chagrin*) and those that did not (e.g., [d'un **chat dro**gué]$_{PP}$ 'of a doped cat', no ambiguity) and asked to detect the word *chat* 'cat'. Reaction time for the word spotting task was significantly slower in the ambiguous case than in the non-ambiguous case. Then they tested if the local ambiguity effect still held when the two syllables spanned a phonological phrase (e.g., …**chat**] $_{PP}$ [**grim**pai…, with potential ambiguity *vs*. …**chat**] $_{PP}$ [**dre**ssait…, without potential ambiguity). There was no difference in reaction time between the two conditions. Moreover, detection of the target word was much faster when there was a phonological phrase boundary between the two syllables than when there was not. In the second study (Gout, Christophe, & Morgan, in revision), they found that 13-month-old American infants could not access previously learned target words when the syllables that comprised the target words were separated by a phonological phrase boundary (e.g., *paper* within a phonological phrase […paper…]$_{pp}$ *vs*. across phonological phrase boundary …pay]$_{pp}$ [per…). Based on these results, they claimed that phonological phrase boundaries constrain lexical access on-line. They further stated, “the prosodic analysis of sentences might be computed in parallel with lexical activation and recognition, in which case prosodic boundaries would be one of the cues that contribute to activation of lexical candidates”. Again, they did not delve into which cues make the perception of this boundary possible, although they acknowledged that phrase-final lengthening could be one factor.

Other studies have made direct attempts to show the effect of phrase-final lengthening on word segmentation. Saffran, Newport, and Aslin (1996), using an artificial language-learning paradigm (see Section 3.2.1 for details), found that word-final lengthening significantly increases English listeners’ segmentation ability. However, the amount of lengthening tested in their study seems to be closer to phrase-final lengthening than word-final lengthening. They lengthened the original vowel by 100 ms (Since their original syllable duration was about 277 ms[3], the increase rate is 36.1 %). The word-final lengthening data for English given in Beckman and Edwards (1990) do not seem to exceed 50 ms. Turk and Shattuck-Hufnagel (2000)'s extensive research on word-boundary-related duration patterns in English also support the phrase-final lengthening interpretation; They showed that there is no significant lengthening in non-pitch accented word-final (but not utterance final) position in English and claimed that "durational cues are unlikely to provide enough information to locate word boundaries unequivocally in all contexts (p.429)". Thus, it is likely that the degree of lengthening used in Saffran et al.'s study was perceived as phrase-final, rather than word-final lengthening.

Bagou, Fougeron, and Frauenfelder (2002), also using the same artificial language learning method as Saffran, Newport et al. (1996), tested the effect of phrase-final lengthening and phrase-final pitch rising in French. Their results showed that both cues were useful, but that the pitch cue was slightly stronger than the duration cue. They also found that the presence of both cues at the same time did not provide any cumulative effect. The correct response rate was 90% for the pitch cue, 82% for the duration cue, and 85 % for the combination of pitch and duration cues.

The influence of the post-lexical intonation pattern in word segmentation was observed in French (Welby, 2003). French intonation is characterized by primary accent on the final full syllable of a prosodic phrase and an optional early rise that is usually

[3] The exact value of syllable duration was not reported in Saffran, Newport, et al. (1996). However, they mentioned that the rate of speech was 216 syllables per minute, which allow us to infer the duration of each syllable (60 sec ÷ 216 syll= 0.277).

realized at the beginning of a content word (Jun & Fougeron, 2000, 2002; Welby, 2003). Welby (2003) showed in a production study that the low point in the early rise pitch is consistently located between function words and content words. She tested if this early rise can aid listeners' word segmentation. She presented minimal pairs or near-minimal pairs which could be parsed in two different ways, and then presented these pairs to the listeners in noise. One such example with a parsing ambiguity is the pair *le niveau de mes sénats* 'the level of my senates' and *le niveau de mécénat* 'the level of patronage'. They differed by how the last three syllables would be parsed. The results showed that the listeners chose *mécénat* when there was an early F0 rise on the first syllable, but that they chose *mes sénats* when there was no early rise. She further showed that alignment of the rise also influenced segmentation. She presented ambiguous non-word pairs such as *mes lamondines* 'my lamondines' and *mélamondine* 'mélamondine' in a carrier sentence such as following: *Et ______ pourrait être une expression utilisée par les français* 'And ____ could be an expression used by the French'. Listeners had a tendency to choose *mélamondine* when the early rise was located at the beginning of [me], but they preferred *mes lamondines* when the early rise was located at the beginning of [la]. The results of this study indicate that the non-lexical intonational pattern plays a significant role in word segmentation in French, and confirmed that French listeners relate an early rise to the content word boundary.

Overall, previous studies provide ample support that various prosodic cues pertaining to both lexical and post-lexical prosodic units significantly facilitate listeners' word segmentation.

### 2.1.2 Allophonic Cues and Coarticulation Cues

Acoustic cues in the speech input can also be a useful source of information for segmentation. As is well-known, the phonetic realization of a segment may differ depending on its location within a word. For instance, English voiceless stops are aspirated in word-initial position (e.g., pie), but not, for example, after the consonant /s/ (e.g., spy). Nakatani and Dukes (1977) showed that such allophonic variations could help English listeners find a word juncture in ambiguous speech sequences. They found that the allophonic variation of syllable-initial and syllable-final /l/ and /r/ (e.g., "we loan" *vs*. "we'll own"; "two ran" *vs*. "tour an"), and aspiration of voiceless stops could be effectively used for word boundary decision. They also found that a glottal stop and/or creaky voice was a strong cue with words that began with a stressed vowel (e.g., "no notion" *vs*. "known ocean").

Listeners are also sensitive to coarticulation information in word segmentation and recognition. Mattys (in press) showed that listeners rely heavily on coarticulation and that this cue is not only efficient, but also more robust than stress (see also Section 2.1.5). Coarticulation also seems to play an important role in lexical perception. Brown (2001) found that words with low relative frequency (i.e., low frequency words with many high frequency neighbors) were produced with more coarticulation than words with high relative frequency (i.e., high frequency words with a few low frequency neighbors), and that the existence of this coarticulation information in the speech input speeded listeners' responses during a lexical decision task.

Listeners' sensitivity to detailed segmental information in word segmentation and recognition implies that this information modulates lexical access, and hence, it should be integrated in models of spoken word processing.

Other studies have shown that, not only adults, but also infants could use detailed acoustic information in word segmentation. Jusczyk et al. (1999) familiarized infants with one item in an allophonic minimal pair (e.g., 'nitrates' or 'night rates') and tested if they could locate the familiarized target in passages. It was expected that if the infants were sensitive to allophonic variation, those who were familiarized with 'nitrates' would be more attentive to the passage with 'nitrates' than that with 'night rates'. They found that while 9-month-olds were not able to distinguish the allophonic differences, 10.5-month-olds were able to do so. These results suggest that this allophonic cue is exploited later than the stress cue is. Furthermore, Johnson and Jusczyk (2001) reported that 8-month old infants rely more on coarticulation cues than on statistical cues (see Section 2.1.4 for details) when they segment words from the speech stream. These studies suggest that both adult and infant listeners have detailed information on phonological and phonetic variations in speech input, which could be readily used during on-line word segmentation.

### 2.1.3 Phonotactic Cues

Native listeners and speakers of a language have implicit knowledge about the phonological structure and constraints of a language. This includes knowing language-specific phonotactics, that is, knowledge of legal and illegal sound sequences in the language. For instance, English does allow words to begin with the consonantal sequence /sw/, as in *sweet* and *swell*, but does not allow the same sequence at the end of words. Likewise, a sequence such as /lp/ is allowed at the end of words, as in *help* and *gulp*, but not allowed at the beginning of words. On the other hand, a sequence like /tf/ never occurs at either edge of words, yet can occur within words, as in *outfit* and *pitfall*, while a sequence like /tv/ never occurs at any position within a word. The occurrence of illegal consonantal sequences in speech input can allow listeners to know where to insert a syllable or morpheme boundary, which can often cue a word boundary. Thus, the existence of such illegal sequences in the speech input may give out strong information about the boundary cue.

There is evidence that listeners make use of such phonotactic knowledge for segmenting words from the speech stream. In a word spotting experiment, McQueen (1998) showed that Dutch listeners could detect words faster and more accurately when a word (e.g., *pil* 'pill') was embedded in a sequence that was aligned with a phonotactic boundary (e.g., /pil.vrem/; [vr] is a possible onset in Dutch) than when it was misaligned with a phonotactic boundary (e.g., /pilm.rem/; [mr] is not a possible onset in Dutch). Weber (2001) obtained a similar result from English listeners. They were faster and more accurate in detecting words that were aligned with English phonotactic boundaries than those that were misaligned with phonotactic boundaries. McQueen (1998) further claimed that this phonotactic cue can be integrated with the Possible-Word Constraint (Norris et al., 1997) and modulate the lexical competition process, just as metrical and other segmentation cues can. Indeed, PARSYN (Auer, 1993), a connectionist model of spoken word recognition which explicitly encodes segmental probabilities and segment-to-segment transition probabilities, successfully simulated results obtained from behavioral studies that showed processing time differences depending on the probabilities of phonotactic patterns. Moreover, there is a strong body of evidence showing that phonotactic probability facilitates and influences word recognition (Massaro & Cohen, 1983; Vitevitch & Luce, 1999).

Jusczyk, Friederici, Wessels, Svenkerud, and Jusczyk (1993) showed that 9-month-old infants preferred to listen to lists of words which follow the phonotactic constraints of their native language compared to a non-native language, while 6-month-olds did not show this preference. Further, Jusczyk, Luce, and Charles-Luce (1994) reported that 9-month-old American infants, but again not 6-month-olds, were sensitive to the frequency of phonotactic patterns. They attended more to the high-probability phonotactic patterns of their native language than the low-probability phonotactic patterns. Mattys and Jusczyk (2001) presented direct evidence that 9-month-old infants made use of phonotactic cues to segment words from the speech stream: infants were able to segment CVC target words more successfully when they were  inserted in a speech stream with a good phonotactic word boundary cue context than in the speech stream without the cue.

Like any other segmentation cue, a phonotactic cue is not consistently available in speech, because not all word boundaries can be identified by the existence of illegal sequences. In English, for instance, only 37 % of word boundaries could be correctly identified by phonotactic constraints (Harrington, Watson, & Cooper, 1989). However, when the cue does occur, it could be more powerful than metrical information. Although Dutch listeners were able to use this metrical information in word segmentation (Vroomen & de Gelder, 1995), the effect of metrical information was overridden by the effect of phonotactic information, when phonotactic information existed in the input (McQueen, 1998).

The use of phonotactic knowledge in word segmentation provides evidence that listeners are sensitive to the probabilities of phonemic sequences and to the positions where they can occur.

### 2.1.4 Distributional Cues

Previous research has claimed that listeners use distributional cues in the linguistic input for word segmentation. One such study focused on the function of the transitional probability of a sub-word unit, the syllable (Saffran, Aslin, & Newport, 1996; Saffran, Newport et al., 1996). They paid attention to the fact that the cohesion of syllables within a word is always stronger than between words, and applied the standard equation for transitional probability, as shown in (1):

$$(1) \qquad P\,(Y \mid X) = \frac{\text{Frequency of XY}}{\text{Frequency of X}}$$

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 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 (i.e., frequency of XY) over the probability of the occurrence X (i.e., frequency of X). In principle, X and Y can be any unit (e.g., phoneme, word, etc.), though Saffran, Newport et al. (1996) employed the basic unit of the syllable. When the first syllable of a disyllabic word was given, the probability of the occurrence of the second syllable of the word is higher than the probability of the occurrence of any other random syllable that cannot form a real word with the first syllable. For instance, in 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. 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. Note also that the first syllable of this word, *ti*, can be a word by itself in English. Therefore, in order to know the transitional probability of the word *tiger*, one should know the frequency of each syllable and be able to compute the co-occurrence frequency of two adjacent syllables.

In order to examine whether listeners can use this transitional probability information for word segmentation, Saffran, Newport et al. (1996) exposed English speaking adult subjects to synthesized speech strings of six nonsense trisyllabic words (CVCVCV) for 21 minutes. In the speech stream, there was no pause between words and there were no acoustic cues which could potentially indicate word boundaries. They then tested whether the listeners were able to extract words from a continuous speech stream. The obtained results strongly suggest that transitional probability alone could assist listeners to extract words from speech input, even in the absence of any acoustic or prosodic cue. Saffran, Aslin et al. (1996) further showed that 8-month-old infants in an American English language environment could segment words from speech streams in a similar experiment, after being exposed to the input for only two minutes. These results show that infants can use some statistical knowledge in segmentation. It further implies that 8-month-old infants have tacit knowledge of syllables, have access to syllable frequency information, and are able to do difficult computations.

Although these behavioral studies with adults and infants have provided clear evidence that people can use transitional probabilities for word segmentation, the one-way transitional probability as calculated in Saffran et al. (1996)'s research, could cause problems in processing, as Dahan and Brent (1999) pointed out. That is, calculating the probability of Y given X, without considering that of X given Y, could cause incorrect predictions during the segmentation process. Further, it seems that this cue is often weaker than other cues in terms of word segmentation (see section 2.1.5 for details).

Another research project that focused on distributional frequency is the INCDROP model (Brent, 1997, 1999). INCDROP (incremental distributional regularity optimization) is a computational model of speech segmentation, which looks for the distribution of a larger unit than a syllable, which Brent (1999) describes as a 'word-like unit'. A 'word-like unit' refers to an utterance that is surrounded by pauses and has the typical prosodic characteristics of boundaries. This model assumes that initial word segmentation starts with perceiving an utterance as a potential word and storing this unsegmented utterance as a whole. Later, when a listener hears another utterance which contains an already familiar unit, the familiar unit will be extracted from the input utterance, and the rest of it will be inferred as a new unit that may or may not be further divided. For instance, the utterance “Gotit”, when first heard, can be stored in the memory as one chunk. Later, if one is exposed to the utterance “Shegotit”, “gotit” will be recognized and thus, segmented from the utterance, and by inference, “she” will be recognized as another unit.

This strategy contains two potential problems. One is that an utterance can sometimes be too long such that it might be burdensome for infants to store the whole utterance. The other problem is that longer utterances can be segmented in many different ways. The model attempts to solve these problems by establishing the criteria shown in (2).

(2) Minimize the total length of all novel words.
Minimize the total number of all novel words.
Maximize the product of relative frequencies. (Brent, 1997)

These were set in such a way that the burden of processing new units would be minimized. This model predicts listeners would prefer segmentation that maximally satisfies these criteria.

Dahan and Brent (1999) conducted a behavioral study in order to test whether adult listeners could segment words from utterances as predicted by the segmentation criteria of the INCDROP model. In their experiment, listeners heard short utterances in isolation and long utterances that contained the short utterances. Results showed that listeners were able to extract a new word-like unit (i.e., the remainder of the long utterance after the removal of the short utterance) from a long utterance. This is predicted by the model, because this type of segmentation (compared to the one that stores the long utterance as a new unit or the one that extracts and stores other sub-parts of the long utterance as new unist) satisfies the first two criteria listed in (2). They also showed that listeners are better at extracting new word-like units when short utterances were presented at the edges of long utterances, than when they were in the middle of long utterances. This result is similar to the results obtained in an infant study, which reported that infants could learn new words faster when they were in utterance-final position rather than utterance-medial position (Cummings & Fernald, 2003). While acknowledging that this result can be explained by other perceptual factors, Dahan and Brent (1999) claimed that this could also be explained by the INCDROP model, because familiar units at the edges of utterances imply creating fewer novel units.

Although the model clearly assumes raw acoustic speech input as the source of information, it does not provide any explanation about the treatment of input. One question that can arise in this model is how infants can handle 'variability' in the input, in other words, how similar the sound of each utterance should be in order to be considered as a 'new' or 'old' unit. For instance, 'utterance' in this model is defined by a pause and the prosodic characteristics of boundaries, but prosodic aspects in the input were not considered thoroughly. One could ask how much of the prosodic information would be stored in the initial lexicon. Suppose that there are two utterances with the same segmental sequences. If the two utterances involve different patterns of lexical stress or tone, or if the two utterances involve different sentential level intonation (e.g., a declarative sentence with a falling tone and an interrogative sentence with a rising tone), will they be stored as a different unit or the same unit? Additionally, it does not consider allophonic variations or acoustic variability which can be introduced in the speech input due to the temporal location of a word within an utterance (e.g., when a word is presented in isolation or edges of an utterance vs. when it is present in the middle of the utterance).

The two approaches introduced so far are quite different in terms of directionality: Transitional probability assumes that bottom-up segmentation will help infants to find their first word (i.e., starting from a syllable to find a word) while the INCDROP model exploits a top-down approach (i.e., starting from an utterance to find a word). However, both strategies share one property in common in that they emphasize the role of probabilistic information in word segmentation. We should also note that human exploitation of transitional probabilities is not just limited to linguistic input (Saffran, Johnson, Aslin, & Newport, 1999).

### 2.1.5 Weighting of Segmentation Cues

Besides identifying individual cues that aid word segmentation, it is important to know how the individual cues are integrated and how people weigh different cues when they conflict.

Recent studies have made attempts to compare the weight of lexical stress cues to that of other segmentation cues. Norris et al. (1997) claimed that metrical cues are stronger than phonotactic cues in the context of the Possible-Word Constraint and the lexical-competition model of Shortlist (Norris, 1994), based on their simulation of word recognition in English. However, this is a rather hasty conclusion, because they did not compare the influences of individual cues. Rather, they compared a condition with silence plus phonotactic cues to a condition with silence, phonotactic cues, and metrical cues. The effect gained by subtraction of the former condition from the latter condition does not guarantee that it is solely due to the effect of metrical cues. That is, there is always a possibility that the larger effect was obtained by accumulation of all the positive cues. Indeed, in a behavioral study, McQueen (1998) presented an opposite result to what Norris et al. (1997)'s simulation revealed. He showed that Dutch listeners rely more on phonotactic cues than stress cues when both cues were present in the input. Coarticulation cues also seem to be stronger than stress cues in clear speech condition. Mattys (in press) performed a series of cross-modal fragment priming experiments and reported that when there was no noise in the auditory prime, stress cues did not have any effect on segmentation, while coarticulation cues did. However, when there was noise in the auditory prime, listeners relied more on stress cues than coarticulation cues.

Studies also suggest the possibility that the weight of each cue may change over the course of language acquisition. Infants are sensitive to phonotactic constraints by the age of nine months (Mattys, Jusczyk, Luce, & Morgan, 1999). However, unlike the adult Dutch listeners in McQueen (1998)'s study, 9-month-old American infants relied more on stress cues than phonotactic cues when the cues were pitted against each other (Mattys et al., 1999). This may be due to a language difference between English and Dutch, but it may also be due to a developmental difference. As previous studies have revealed, infants' dependency upon prevalent stress cue changes over time. 7.5-month-old infants were able to segment only the words with Strong-Weak stress pattern, but 10.5-month-old infants were able to segment both Strong-Weak and Weak-Strong stress patterns. Thus, it is possible that infants rely more on prosodic cues than phonotactic cues at an early stage of language acquisition, yet the weight of the two cues may change as infants become more familiar with the language (Jusczyk, 1999).

Johnson and Jusczyk (2001) found that infants preferred speech cues to a general cognitive strategy in word segmentation. Although 8-month-old infants were able to use the statistical cue of transitional probability for word segmentation when it was the only cue available in the input, they relied more heavily on stress and coarticulation cues than on the statistical cue when these cues conflicted in the speech input.

Overall, these results indicate that the weighting of cues for word segmentation cannot be determined absolutely. The relative dependency on a certain segmentation cue may be influenced by the speech environment that listeners are exposed to at the time of input processing, and the preference for certain cues over others may change during the course of language acquisition.

### 2.1.6 General Properties of Word Segmentation Cues and Their Implications to the Current Study

The research introduced thus far provides us with insights about the nature of the word segmentation process. First, the studies reveal that both adults and infants rely on the same segmentation cues, although their degree of dependency on each cue may vary. By the age of 7-10 months, infants are able to use various cues in speech for word segmentation (Jusczyk, 1999; Jusczyk, Cutler et al., 1993; Jusczyk, Friederici et al., 1993; Jusczyk et al., 1999; Jusczyk et al., 1994). Dependence upon these cues can be replaced by other segmentation strategies as infants acquire more knowledge of their native language. For instance, since adults' lexicons are much larger than infants', it is conceivable that adults mainly rely on lexical competition (McClelland & Elman, 1986; Norris, 1994). However, this by no means indicates that adult do not use other means for segmentation. In fact, we have seen enough evidence that their use of individual segmentation cues was exactly the same as that of infants when they were forced to segment words from a speech stream in experimental situations (Cutler & Norris, 1988; McQueen, 1998; Nakatani & Dukes, 1977; Saffran et al., 1999). Moreover, the cues that they have used for word segmentation of their native language can be activated (Weber, 2001) and persistently used (Cutler et al., 2002; Cutler & Otake, 1994), even when they process a foreign language. Therefore, although the target population of the current study is adult Korean listeners, I assume that the results of this study will also shed light on how infants born in a Korean environment solve the initial segmentation problem.

Second, studies reviewed so far underscore the effect of frequency in word segmentation. The studies that focus on distributional information such as transitional probabilities (Saffran, Aslin et al., 1996; Saffran, Newport et al., 1996) suggest that listeners have knowledge of the occurrence and co-occurrence frequencies of syllables, words, or utterances. Similarly, phonotactic constraints can also be considered to make use of similar information about segments, because ‘impossible’ phonotactics basically means ‘zero’ co-occurrence of two segments. The use of prosodic cues, such as stress pattern in English and Dutch, also depends on the frequent encounter with one pattern (e.g., trochaic) over another (e.g., iambic) in the speech input (see, for example, Cutler & Carter, 1987). Therefore, the results of these studies indicate that segmentation is governed by frequency information from the speech input. That is, people exploit various probabilistic information to find words and to locate possible word boundaries. The frequency effect seems to play a role at all levels of linguistic units (e.g., phoneme, syllable, and ‘word-like’ units). This dissertation examines whether word segmentation can also be influenced by the frequency of intonational patterns of prosodic phrases.

Third, previous studies strongly indicate that the set of segmentation cues is language universal (e.g., listeners use phonotactic, allophonic, and prosodic cues for word segmentation), but the detailed manifestations of individual segmentation cues that listeners exploit are language-specific. Although [bz] is not a phonotactically legal sequence in word initial position in English, it is a legal sequence in Polish (*UCLA Phonetic Archives*). Thus, English listeners would put a word boundary between [b] and [z] while Polish listeners might not. Allophonic cues are governed by language-specific phonological rules, and the degree of coarticulation also shows language-specific differences (Manuel, 1990; Manuel & Krakow, 1984). Listeners implicitly know such acoustic details of their native language, and this language-specific knowledge is used in on-line segmentation. Likewise, although the prosodic and intonational systems of spoken languages exploit the same set of parameters (such as pitch, duration, intensity and so forth (Pierrehumbert, 2000)), detailed phonetic realizations of prosodic system differ from language to language. It is this language-specific property of word segmentation which further calls our attention to the necessity for more cross-linguistic studies on the topic. In this context, the current study aims to specify what are the language-specific segmentation cues for Korean. The results will not only add to the body of research that attempts to find language-specific and language-universal mechanisms in speech processing, but also provide essential information for modeling spoken word recognition in Korean.

## 2.2 Korean Prosody

Like any other language, the Korean language has many regional dialects. This dissertation investigates the role of prosody in word segmentation of Korean, but the research will concentrate on one specific dialect, Seoul Korean. Seoul Korean, also known as Standard Korean (*pyo-joon-mal*), is the official dialect of South Korea. The target dialect of most research on Korean prosody, which will be introduced in this section, is also Seoul Korean. Thus, for the sake of convenience, I will refer to this dialect as ‘Korean’ throughout this dissertation.

### 2.2.1 Prominence in Korean: Lexical or Post-lexical?

Although there is a consensus among researchers that Korean has fixed prominence, it has been controversial as to whether this prominence belongs to the word-level or phrase-level.

H.-B. Lee (1974) and H.-Y. Lee (1990) claimed that Korean has word-level stress, and that prominence-lending depends on syllable weight and duration. Stress falls on the first syllable of a word if it is heavy (CVC or CVV[4]) or on the second syllable when the first syllable is light. They also argued that duration is more important than pitch in deciding where the stress is within a word. However, later studies provided evidence against word-level prominence. Jun (1995b)'s production studies demonstrated that the location of "stress" within a word changed depending on where the word was produced within a prosodic structure, confirming her earlier claim: "the tonal pattern of an AP in Korean is not specific to a lexical item but is a property of the phrase" (Jun, 1993). A production study by Lim (2001), which was based on recordings from two male native speakers of Korean,  showed that in trisyllabic words, heavy medial syllables were longer than heavy final syllables, which, in turn, were longer than heavy initial syllables. His result also revealed that the F0 value was higher in medial syllables than others. In a perception experiment, he showed that listeners were sensitive to the location of prominence: most Korean listeners who participated in his experiment perceived the medial syllable of a word as prominent. In addition, he also observed that Korean

[4] Seoul Korean does not have a vowel length distinction any more.

listeners' percept of prominence seemed to rely on the relative change in pitch movement more than on duration. These results contradict the word-level stress claim (Lee, 1974; Lee, 1990) and support the phrase-level prominence claim (Jun, 1993, 1995b, among others). However, Lim and de Jong (1999), through further analysis of Lim (2001)'s production data, found that tonal alignment of the phrase initial F0 peak was sensitive to syllable weight to a certain degree. When there was a heavy phrase-initial syllable, the F0 peak could be realized at the end of the first syllable, whereas the F0 peak was realized on the second syllable of a phrase when the phrase-initial syllable was a light syllable. This result seems to explain why syllable weight was considered a main factor of Korean stress in Lee (1974) and Lee (1990). Although more work is needed to figure out Korean tone-segment alignment, the results of these studies clearly suggest that Korean prominence is related to the intonational pattern of post-lexical prosodic phrases, rather than words.

One thing to be noted here is that phrase-level prominence in Korean does not refer to post-lexical pitch accent. In prosodic typology, there are two different types of phrasal prominence: post-lexical 'head' prominence, which is realized by marking the head of a prosodic unit (e.g., post-lexical pitch accent) and post-lexical 'edge' prominence, which is realized by marking the edge of a prosodic unit with a specific phrasal tone (Beckman, 1986; Beckman & Edwards, 1990; Hyman, 1978; Jun, 2004b; Ladd, 1996). Post-lexical prominence in Korean is the 'edge' prominence, not the 'head' prominence. The following sections will describe how the edge prominence is realized in Korean.

### 2.2.2 Korean Prosodic Hierarchy

The Korean prosodic hierarchy adopted for this study is the intonation-based prosodic model proposed by Jun (1993; 2004a). The prosodic hierarchy is composed of four prosodic levels, as illustrated in Figure 1.

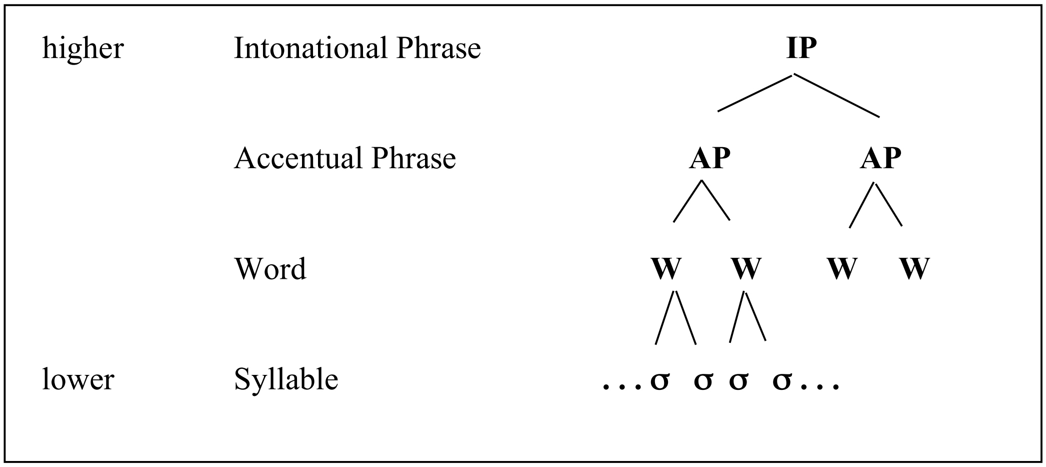

*Figure 1.Prosodic Hierarchy of Korean*

There are two prosodic phrases in this model: the Intonational Phrase (IP) and the Accentual Phrase (AP). The Intonational Phrase (IP) is the highest prosodic unit in the hierarchy.  It is often followed by a pause, and characterized by final lengthening and boundary tones (Jun, 1993, 2004a). There are nine attested IP boundary tones (L%, H%, LH%, HL%, LHL%, HLH%, LHLH%, HLHL%, LHLHL%, where the % symbol marks the right edge of the IP). These boundary tones deliver various semantic and pragmatic meanings (Jun, 2004a; Park, 2003). The Accentual Phrase (AP) is lower than the IP in the hierarchy. Unlike the IP, the AP is never followed by a pause (Jun, 2004a). The existence of phrase-final lengthening in the AP is controversial (Jun, 1993, 1995, but see also Cho & Keating, 2001; Oh, 1998). The default tonal pattern of the AP is THLH (Tone (Low or High) – High – Low – High). The initial tone (T) is realized by H when AP-initial segment is aspirated or tensed (i.e., [+stiff vocal cords]), and by L elsewhere. Thus, the AP is usually demarcated by an initial LH (or HH) and a final LH, and the edge prominence of the AP is realized by the AP-final H tone. The realization of these tones, however, can vary depending on several factors, including the number of syllables within the AP. When an AP has four or more syllables, the default tonal pattern is often realized, but when an AP has less than four syllables, the AP medial tones (i.e., H, L, or both, in L**HL**H) can be undershot. In addition, the AP-final H tone is sometimes realized as a L tone due to its interaction with adjacent tones or pragmatic factors. There are about fourteen attested AP tonal patterns (i.e., LH, LHH, LLH, HLH, HH, HL, LHL, HHL, HLL, LL, HHLH, LHLH, LHLL, HHLL), but these different patterns do not seem to be associated with contrastive meaning (Jun, 2004a). The Word level is lower in the hierarchy than the AP, but unlike the two prosodic phrases, it is not defined by a tonal patterns. The lowest level in the hierarchy is the Syllable.

### 2.2.3 Korean Accentual Phrase as a Unit for Word Segmentation

This dissertation hypothesizes that the Accentual Phrase, a post-lexical prosodic phrase, is a rhythmic unit that can facilitate word segmentation in Korean. The hypothesis is based on the following. First, the AP is the smallest unit which bears intonational prominence in Korean to which listeners are sensitive (Jun, 1995b; Lim, 2001; Lim & de Jong, 1999). Second, the AP is tonally demarcated. It has the underlying default tone pattern of THLH, although variation in this pattern is possible depending upon pragmatic, segmental, and durational factors. It is likely that the consistent rising patterns of the AP can give a sense of rhythm in speech which listeners could use as a segmentation cue. As Jun (1993) stated, the tonal events delimit a prosodic grouping of words. And needless to say, 'grouping' or 'chunking' of the speech stream by a demarcated prosodic unit can aid in speech segmentation. Third, AP boundaries match with either the beginning or end of a content word, and sometimes with both edges of a word. That is, a word never extends over an AP boundary. This implies that any cues to phrase boundaries can also be cues to word boundaries. Thus, AP-initial position can function as a very effective point from which initial lexical candidate activation process can be attempted. Finally, decisive evidence that the detection of the AP would be a useful prosodic unit for word segmentation comes from Jun and Fougeron (2000)'s study, where they observed that the Korean AP has 3.2 syllables and 1.2 content words on average. These numbers are quite low, compared to the corresponding English prosodic phrase (i.e., the intermediate Phrase), which has been reported to have 5.3 syllables (Ueyama, 1998) and 3.9 content words (Ayers, 1994) on average. Also, Schafer and Jun (2001) reported that Korean speakers tend to produce one phonological word as one AP. They found that this pattern comprise 90% of their data. Therefore, although it is true that the AP can contain more than one prosodic word in principle, the AP turns out to be quite similar in size to the word. Based on these observations, I hypothesize that the detection of AP boundaries will certainly facilitate word segmentation.

Now, with this as a working hypothesis, let us take a look at a sentence in Korean, which contains a parsing ambiguity.

(3) a. Ambiguous word boundary
Korean orthography 아버지가방에있다.
Romanized transcription Abeojikapangeissta
Phonological transcription /a.pʌ.tʃi.ka.paŋ.ɛ.is*.ta/[5]
K-ToBI Romanization abEzigabaQeiDda

b. Parsing option 1.
Korean orthography 아버지가 방에 있다.
Phonological transcription apʌtʃi-ka paŋ-ɛ is*ta
English gloss Father-NOM room-LOC is-DECL[6]
'Father is in the room.'
Phonetic/Prosodic transcription [[abʌdʒiga]$_{AP}$ [paŋeit*a]$_{AP}$]$_{IP}$

c. Parsing option 2.
Korean orthography 아버지 가방에 있다.
Phonological transcription apʌtʃI kapaŋ-ɛ is*ta
English gloss Father bag-LOC is-DECL
'(It) is in father('s) bag.'
Phonetic/Prosodic transcription [[abʌdʒi]$_{AP}$ [kabaŋeit*a]$_{AP}$]$_{IP}$
or
[[abʌdʒi]$_{AP}$ [kabaŋɛ]$_{AP}$ [it*a]$_{AP}$]$_{IP}$

The sentence in (3a), "아버지가방에있다 (/a.bʌ.dʒi.ka.paŋ.ɛ.is*.ta/)" can be interpreted in two different ways. The ambiguity solely lies in how to segment the

[5] c*: tense consonant

[6] NOM: nominative marker, LOC: locative marker, DECL: declarative marker

sentence, and what is crucial for segmentation is the fourth syllable of the sentence, *ka*. The sentence can be interpreted as “Father is in the room,” (3b), if the fourth syllable *ka* is attached to the sentence-initial noun *Abeoji* ‘father’ and functions as a nominative marker. Alternatively, the sentence can be interpreted as “It is in father’s bag,” (3c), if the fourth syllable *ka* is attached to the following syllable *pang*, because the combination of the two syllables forms a noun *kapang* ‘bag’.

In written form, this ambiguity can be easily solved by the location of space. If there is a space after *ka*, it will be interpreted as (3b), and if space is before *ka*, it will be interpreted as (3c). But, what will help to resolve this ambiguity in a spoken form? Figure 2 and Figure 3 below illustrate the waveforms and pitch tracks of the sentences in (3b) and (3c), respectively, as produced by a female native speaker of Korean (the author). In these utterances, there is no pause which can readily distinguish word boundaries[7]. It is prosodic phrasing that distinguishes these two  interpretations.

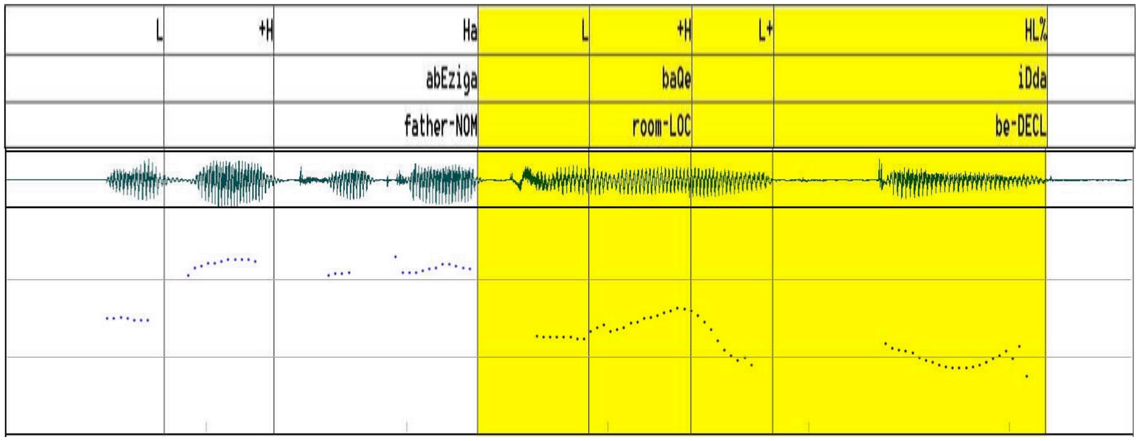

*Figure 2. Parsing Ambiguity in Korean: (3b) ‘Father is in the room’.*
*“Ha” on the top most line indicates AP-final boundary. This sentence has two APs.*

[7] Space shown in the waveforms of Figure 2 and Figure 3 is not pause but stop closure.

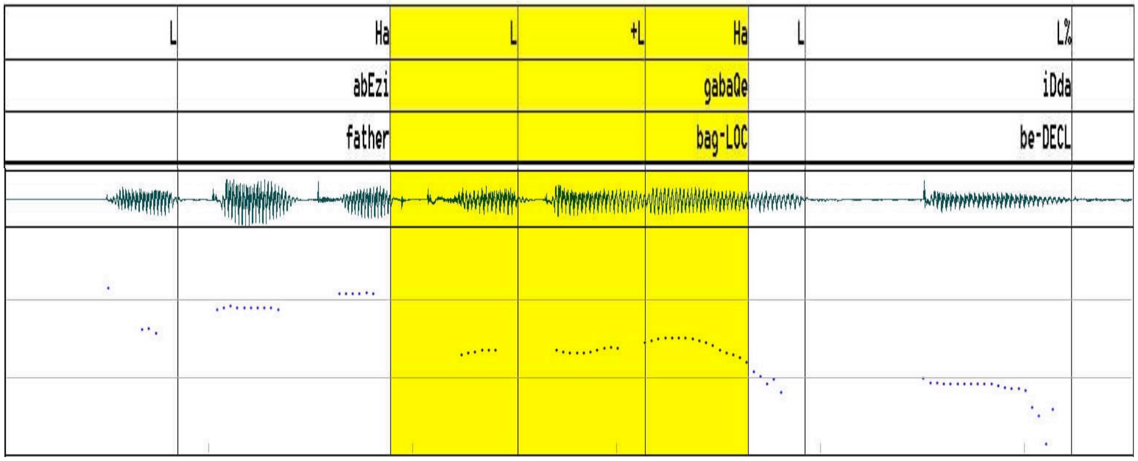

*Figure 3. Parsing Ambiguity in Korean: (3c) '(It) is in father('s) bag.'*
*"Ha" on the top most line indicates AP-final boundary. This sentence has three APs.*

Figure 2 shows that the utterance is composed of two APs ([[abʌdʒiga]$_{\mathrm{AP}}$ [paŋɛit*a]$_{\mathrm{AP}}$]$_{\mathrm{IP}}$), and the fourth syllable /ka/ (i.e., [ga]) belongs to the first AP. Figure 3 shows that the utterance is composed of three APs ([[abʌdʒi]$_{\mathrm{AP}}$ [kabaŋɛ] $_{\mathrm{AP}}$ [it*a]$_{\mathrm{AP}}$]$_{\mathrm{IP}}$), and the fourth syllable /ka/ (i.e., [ka]) belongs to the second AP. Therefore, if listeners can detect any prosodic cues or other acoustic differences that pertain to AP boundaries, the ambiguity in the speech string can be resolved.

One of the prosodic cues that characterizes the AP is pitch, especially, the AP-final H tone which is clearly visible in Figure 2 and Figure 3. Note that each AP-final syllable is clearly marked by an H tone ('Ha' in the figures) and is followed by the L tone in the next AP in both figures (unless the AP final tone is preempted by an IP boundary tone, which is marked with %). In addition, although this is not shown in the figures allophonic variation can also be a useful cue in this case, as shown in the phonetic transcriptions in (3).

The domain of the Korean Lenis Stop voicing rule is the AP (Jun, 1993). The lenis stop is usually voiceless in AP-initial position, but there is a strong tendency for the lenis stop to become voiced in AP-medial position (Cho & Keating, 2001; Jun, 1996b). Thus, as shown in the phonetic transcription in (3b), when the fourth syllable /ka/ belongs to the first AP, /k/ in *ka* would be voiced (i.e., [g]) since it is in AP-medial position and /p/ in the next syllable *pang* will remain voiceless because it is in AP-initial position of the second AP. Listeners' knowledge of this type of allophonic variation will let them insert the AP boundary between the two syllables. On the other hand, as shown in the phonetic transcription of (3c), /k/ in *ka* would remain voiceless since it is at the AP-initial position and /p/ in the next syllable *pang* would be voiced (i.e., [b]) since it is within the AP in this sentence. In this case, listeners' knowledge of allophonic variation will prevent them from inserting the boundary between the two syllables.

Choi and Mazuka (to appear) showed that young Korean children (3-5 year-olds) can exploit the AP-final high tone and lenis stop voicing cues in parsing sentences with phrasing ambiguity, such as the sentences shown in (3). They also showed that the prosodic cue is stronger than the allophonic cue in disambiguating such sentences.

The knowledge of prosodic and acoustic cues of the AP will greatly aid listeners during on-line lexical segmentation and access, as well as sentence processing, especially when it is combined with the knowledge that a word never spans over AP boundaries in Korean. Thus, in the case of (3b), for example, a lexical candidate such as *kapang* 'bag' would not be activated (or the candidate would be penalized even if it is activated), because the two syllables belong to different prosodic phrases. Likewise, a lexical candidate such as *jikap* ‘purse’ would be penalized (or not activated at all) for the same reason, in the case of (3c).

To summarize, this dissertation hypothesizes that the AP can be a reliable word boundary indicator in Korean, because the AP is approximately word-size. Further, it can provide chunking information because it has various prosodic cues and is a domain for allophonic variation. Thus, listeners’ ability to detect AP boundaries would lead to more efficient and more economic word segmentation. By reducing the number of segmentation hypotheses and by not testing segmentation units smaller than words, they can attempt a lexical search only within the prosodic phrase.

# CHAPTER 3

# INTONATIONAL PATTERN FREQUENCY AND WORD SEGMENTATION

## 3.1 Introduction

In the previous chapter, I hypothesized that the Accentual Phrase would play a crucial role in word segmentation in Korean because it is similar in size to a word (Jun & Fougeron, 2000), and because it is a tonally demarcated unit that can potentially provide a sense of rhythm in speech (Jun, 2004b). However, the data reported in Jun and Fougeron's study (2000) were quite limited in that the data were only from read-speech[8]. Also, given that there are quite a few varieties of AP tone patterns, a question still remains as to whether speech input delivers enough regularity to be found and exploited by the listeners for speech segmentation. Therefore, in order to confirm the validity of the initial hypothesis and to provide grounds for subsequent perception studies, a quantitative corpus study was undertaken.

In order to obtain speech production data for various speech styles, materials for this study were composed of recordings of read speech (henceforth Read speech corpus) and two radio shows (henceforth Radio corpus). From these production data, I collected various distributional information, including the frequency of post-lexical tone patterns of

[8] In their study, three Korean speakers read the story 'The North Wind and the Sun'.

the AP, the locations and numbers of content words within the AP, and the frequency of tonal patterns that were imposed upon content words.

The organization of this chapter is as follows: Section 3.2 will explain the nature of the corpus adopted for this study. Section 3.3 will report the results of the distributional analysis, and Section 3.4 will be a discussion of these results.

## 3.2 Method

### 3.2.1 The Corpora

The Read speech corpus was originally designed to investigate default phrasing and attachment preference for relative clauses in speech production (Jun & Kim, 2004). The list contained fifty-six sentences. Among them, thirty-two sentences had one relative clause and two NPs in them (RC-NP1-NP2), while twenty-four sentences did not have this sentential structure. The sentences with a relative clause construction were ambiguous, because the relative clause could modify either one of the two NPs. However, the results of the production study showed that the potential ambiguity did not affect speech production, in that speakers produced each phrase of the RC-NP1-NP2 sequence in one AP, about 80% of the time. The current study used a small subset of this production data for the analysis of the corpus. The data employed here were produced by two male and two female native speakers of Seoul Korean, who were graduate students at UCLA at the time of recording. The sentences were read at normal speed. The four speakers each read 56 sentences. Thus, a total of 224 sentences were analyzed.

The Radio corpus was recorded from two radio programs, *Radio Moodae* 'radio stage' and *Radio Dokseoshil* 'radio reading room', both of which aired on KBS (Korean Broadcasting System) 1 Radio. *Radio Moodae* ('radio stage') broadcasts radio drama as recorded by voice actors. *Radio Dokseoshil* ('radio reading room') introduces and reviews newly published books. The format of *Radio Dokseoshil* included an introduction of a novel by the show host, an interview with the author of the novel, and the dramatized presentation of the novel. Thus, the Radio corpus was more similar to spontaneous speech than the Read speech corpus in general, and some part of this data actually contained sentences from spontaneous conversation. The show host, interviewees, and voice actors were all Seoul Korean speakers. Two hundred sentences were extracted from various parts of two one-hour episodes.

### 3.2.2 Analysis

The recorded sentences were transcribed using K-ToBI labeling conventions (Jun, 2000), by the author.

The first step for the analysis was to locate AP boundaries and count the number of APs in each corpus. The Read speech corpus had 1592 APs and the Radio corpus had 1493 APs. Then, the numbers of syllables and content words within the APs were counted, in order to compare the current data with the read speech data reported in Jun and Fougeron (2000). Further, the corpus was annotated with intonational patterns of APs so as to examine the frequency of the attested tonal patterns. For this analysis, IP-final APs were excluded because AP-final tone is pre-empted by IP-boundary tone when an

AP is in the IP-final position. The Read speech corpus had 1203 non-IP-final APs, and the Radio corpus had 906, making the total number of APs used for the tone frequency analysis 2109. Temporal locations of content-word onsets within an AP (i.e., AP-initial, AP-medial) were marked, for the purpose of observing the co-occurrence frequency of AP onset and word onset. Finally, I transcribed the tone patterns that were imposed upon content words and observed the frequency of post-lexical tonal patterns realized on content words. Again, only non-IP-final APs were considered for this analysis. Additionally, when I transcribed the tone patterns of content words, I did not include monosyllabic content words and focused only on the post-lexical tone mapping on the first two syllables of multisyllabic content words. There were four possible combinations of tone patterns that could be imposed on the initial two syllables of words: level H (HH), level L (LL), falling (HL), and rising (LH). Although monosyllabic nouns and adverbs were not included in the tonal pattern analysis for content words, monosyllabic verbal and adjectival stems were included, because these are bound stems in Korean and hence can compose a prosodic word only when they are combined with bound suffixes.

As already mentioned, the AP-initial tone is determined by the laryngeal feature of the AP-initial segments. Thus, for the tone-pattern frequency analyses, I divided the data into two groups depending on the property of initial segments: one group had segments that induce H tone in the AP-initial position (i.e., tense and aspirated consonants), and the other group had segments that do not induce H tone in the same position (i.e., the rest of the consonants and vowels). For the sake of convenience, I will refer to the segments that consist of the former group as H tone inducers and those that consist of the latter group as L tone inducers.

## 3.3 Results

### 3.3.1 Syllable and Content Word Count within the AP

The average number of syllables and content words within the AP was calculated based on 3085 APs (including IP-final APs) from two types of corpora. The average number of syllables within the AP was 3.39, and the average number of content words within the AP was 1.14. The Read speech corpus and the Radio corpus did not differ much when these numbers were counted separately: the average number of syllables within the AP was 3.35 for the Read speech corpus and 3.43 for the Radio corpus; the average number of content words was 1.14 for the Read speech corpus and 1.13 for the Radio corpus. This is similar to the result of Jun and Fougeron (2000)'s study, which reported that the average number of syllables within the AP was 3.2 and the average number of content words was 1.2.

The distribution of data, in terms of the number of syllables and the number of content words within the AP, is shown in Figure 4 and Figure 5, respectively. Figure 4 demonstrates that most APs were composed of three or four syllables (67. 9%), and that no AP with more than 7 syllables was observed.

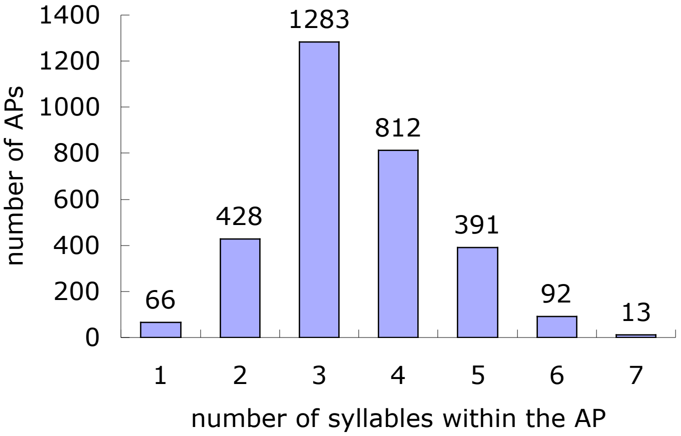

*Figure 4. Frequency distribution: number of syllables within the AP.*
*(Based on 3085 APs from both corpora)*

Furthermore, as shown in Figure 5, the majority of APs had one content word (84.2%). Only a small number of APs contained three content words (n=13) or just had function words without any content word (i.e., 0 content words in Figure 5, n=37), and no AP with more than three content words was found.

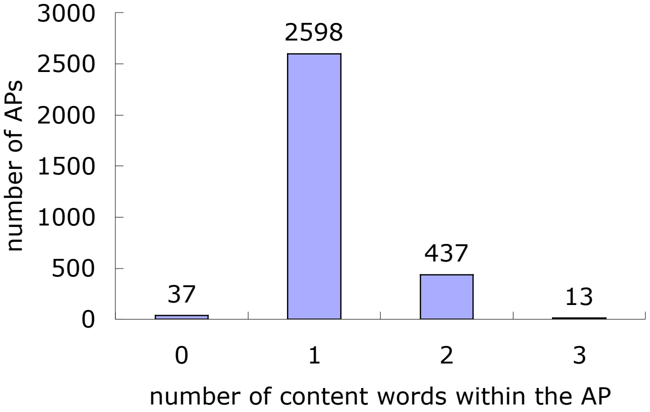

*Figure 5. Frequency distribution: number of content words within the AP.*
*(Based on 3085 APs from both corpora)*

### 3.3.2 AP tone pattern frequency

Among 2109 non-IP-final APs, 1541 APs started with an L tone inducer (i.e., lenis consonants, nasal consonants, glides, vowels), and 568 APs had an H tone inducer (i.e., tense and aspirated consonants) at the phrase-initial position. Naturally, the frequency of AP tonal patterns varied depending on what the AP-initial segment was. Table 1 illustrates the tone pattern frequency of APs with L tone inducers.

| AP Tone Patterns | n= 1541 | % |
|---|---|---|
| LH | 675 | 43.8 % |
| LHH | 307 | 19.9 % |
| LHLH | 219 | 14.2 % |
| LLH | 109 | 7.1 % |
| LHL | 69 | 4.5 % |
| LHLL | 47 | 3.0 % |
| LL | 39 | 2.5 % |
| HH | 31 | 2.0 % |
| HLH | 17 | 1.1 % |
| HHLH | 11 | 0.7 % |
| HHL | 10 | 0.6 % |
| HLL | 5 | 0.3 % |
| HL | 2 | 0.1 % |

*Table 1. AP intonational patterns with L tone inducers*

The most frequent pattern was LH, which composed 44% of the data; the second most frequent pattern was LHH (19.9%); the third most frequent pattern was the default tone pattern of LHLH (14.2%). Note also that 88.8% of the APs ended with an AP-final H tone (LH, LHH, LHLH, LLH, HH, HLH, HHLH), and 85.5% of the APs started with an AP-initial rising tone (LH, LHH, LHLH, LHL, LHLL). A small number of AP-initial high tones were observed (4.9%) even though the initial segments were L tone inducers.

Although the general tendency was the same, two corpora were a little different with regard to the distribution of the AP tone patterns, in that the Radio corpus showed more variation than the Read speech corpus. Both corpora had all seven tonal patterns starting with an L tone (viz., LH, LHH, LHLH, LLH, LHL, LHLL, LL) and, among them, the three most frequent patterns were LH, LHH and LHLH. However, while 88.9% of the APs had one of the three patterns in the Read speech corpus, only 64.8% of the APs in the Radio corpus appeared with these three patterns. In addition, the Read speech corpus had three tone patterns with H-initial tone (2.6%), but the Radio corpus had six tone patterns with H-initial tone (7.7%).

Figure 6 illustrates the distribution of three most-frequent tone patterns (i.e., LH, LHH, LHLH) in the data within each AP syllable count.

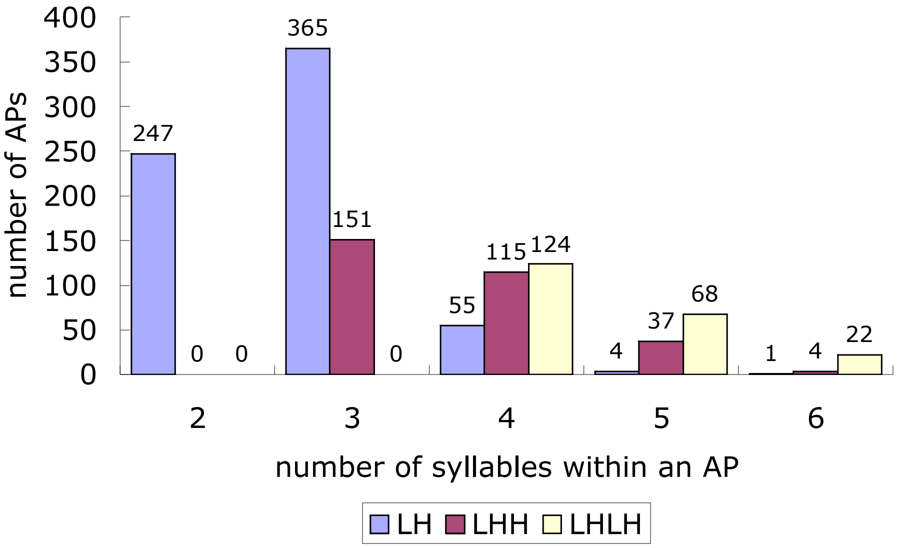

*Figure 6. The frequency of AP tonal patterns depending on syllable count (with L tone inducers)*

LH was the most frequent pattern when the APs had two or three syllables. The default pattern of LHLH was the most frequent tonal pattern when the APs had more than three syllables, which suggests that the low frequency of the default pattern in the corpus is due to the low number of syllables within the AP. However, it should be noted that the patterns of LH and LHH were still observed, with a non-ignorable proportion, in the APs with four syllables. This indicates that the undershoot of the second L tone from the LH**L**H default pattern is quite frequent[9]. On the other hand, as seen in Table 1, the undershoot of the first H tone in L**H**LH is not very frequent: LLH and LL patterns represented only 9.6% of the data.

Table 2 illustrates the tonal pattern frequency of APs with H tone inducers. The most frequent pattern was HH (37.5%), followed by the HHLH pattern (18.7%) and then

[9] LH = L (HL) H, LHH = LH (L) H

the HHL pattern (15.7%). These three tones account for 71.8% of the APs. The two corpora shared the three most frequent patterns, with the same ranking (HH > HHLH > HHL). However, the Radio corpus showed more tonal patterns than the Read speech corpus: three tone patterns that start with an L tone were observed with H inducers in Read speech, but six such patterns were observed in Radio corpus.

| AP Tone Patterns | n=568 | % |
|---|---|---|
| HH | 213 | 37.5 % |
| HHLH | 106 | 18.7 % |
| HHL | 89 | 15.7 % |
| HLH | 50 | 8.8 % |
| HLL | 45 | 7.9 % |
| HL | 27 | 4.8 % |
| LLH | 28 | 4.9 % |
| LHLH | 3 | 0.5 % |
| LHH | 3 | 0.5 % |
| LL | 2 | 0.4 % |
| LH | 1 | 0.2 % |
| LHL | 1 | 0.2 % |

*Table 2. AP intonational patterns with H tone inducers*

An AP-final H tone was frequently observed in APs with H tone inducers, as well: 71.1% of the APs had H tone at the AP-final position. An AP-final L tone was less common (28.9%), but the proportion of these patterns was twice as high as that found in the APs with L tone inducers (11.1%). This seems to be due to the 'see-saw effect' (Jun, 1996a), which describes speakers' tendency to avoid a sequence of two identical tones. Also, A small portion of the AP-initial L tones was observed (6.7%), despite of the existence of H tone inducers in the AP-initial position.

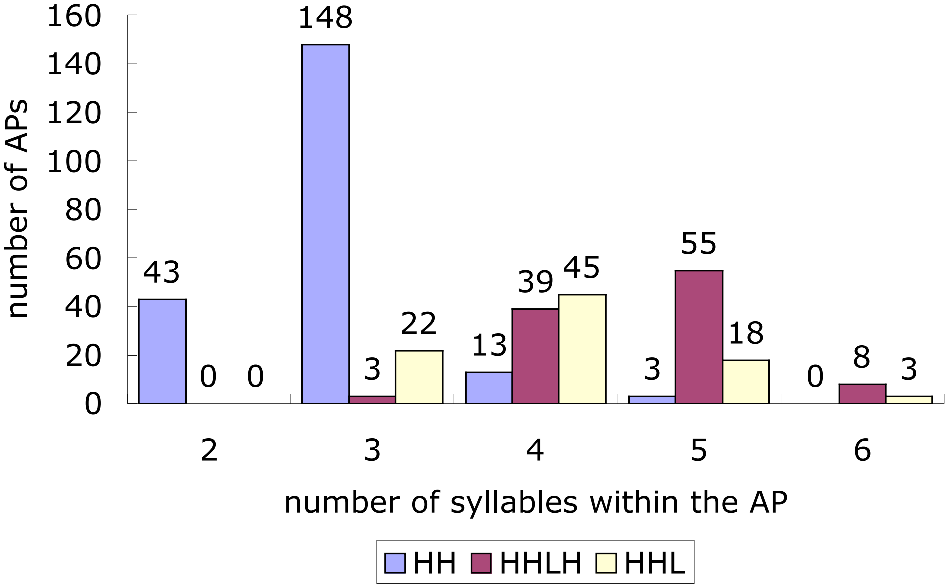

*Figure 7. The frequency of AP tonal patterns depending on syllable count (with H tone inducers).*

Again, as shown in Figure 7, the data revealed that the occurrence of the default pattern (HHLH) was frequent when the APs had more than three syllables. HH was the most frequent pattern in di- and tri-syllabic APs.

The results show that the same trend is observed with both L tone inducers and H tone inducers. That is, the tones that were most frequently realized are the AP-initial tone (either L or H) and the AP-final tone (H).

### 3.3.3 Location of Content Words within the AP

The co-occurrence frequency of AP onset and content-word onset was extremely high, as illustrated in Figure 8. The non-IP-final APs in this corpus contained 2343 content words, and, among them, 2076 occurred in AP-initial position (88.6%) and 267 words appeared in AP-medial position (11.4%). This means that most content-word onsets coincided with AP onsets. In addition, the majority of multisyllabic content words appeared in AP-initial position in the corpus. Once monosyllabic nouns and adverbs were excluded from the data set, there were 2017 multisyllabic content words, with 1811 of these located in AP-initial position (89.8%) and 206 located in AP-medial position (10.2%).

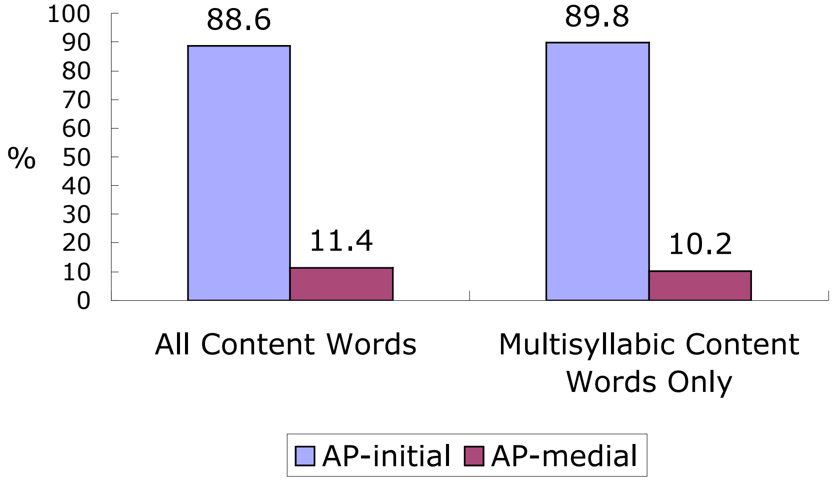

*Figure 8. Co-occurrence frequency of AP onset and content word onset*

### 3.3.4 Content Word Initial Tone Patterns

Among the 2017 multisyllabic content words, 1471 had an L inducer as a word-initial segment. 1309 words out of 1471 appeared in AP-initial position; the rest appeared in AP-medial position. As aforementioned, the word-initial tones were classified into four categories (LH, LL, HH, HL). Contour tones (LH and HL) included not only when the target syllable shows the lowest F0 point (L) or the highest F0 point (H) (i.e., Figure 9, left panels), but also when the second syllable is in the middle of rising or falling, thus not being specified with a tone (i.e., Figure 9, right panels).

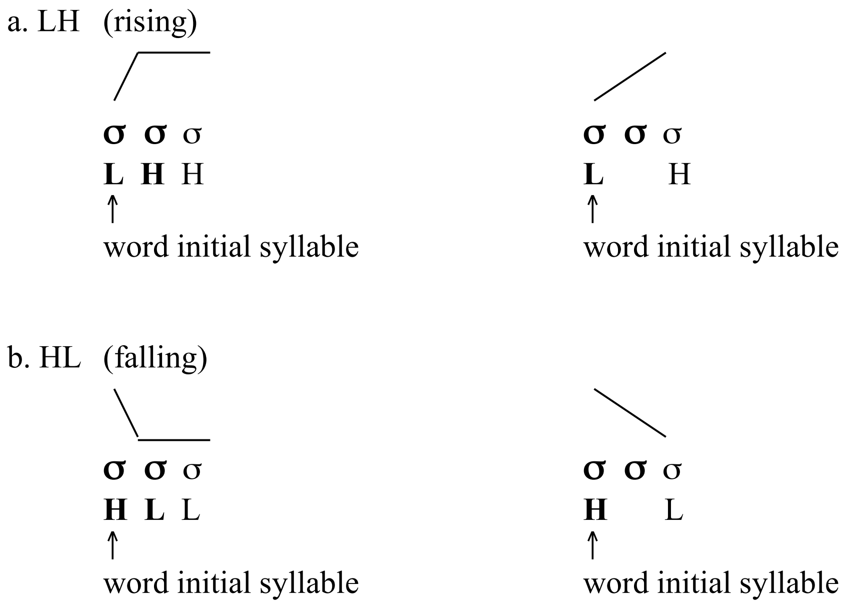

*Figure 9. Contour tones on the first two syllables of content words (a. rising, b. falling)*

The distribution of tone patterns that were realized on the first two syllables of AP-initial multisyllabic words with L inducers is illustrated in Figure 10. In the figure, ‘rising’ indicates LH tone (as shown in Figure 9a), ‘falling’ indicates HL tone (as shown in Figure 9b), and level H and level L indicate HH tone and LL tone, respectively.

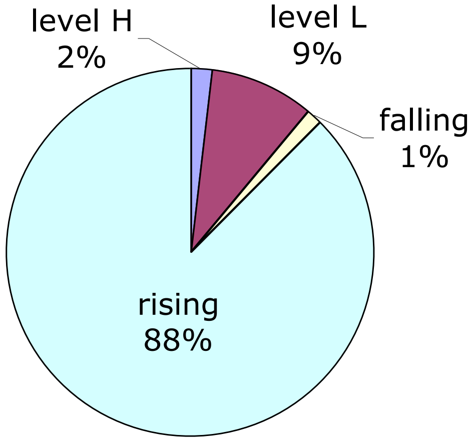

*Figure 10 . Tonal patterns of AP-initial multisyllabic content words (with L tone inducers)*

In AP-initial position, 88% of multisyllabic content words occurred with a rising tone (LH) pattern. This tendency was stronger in the Read speech corpus, where 91.5% of words carried an initial rising tone. In the Radio corpus, 82.9% of words started with a rising tone.

Among the words with an H inducer, 502 words occurred in AP-initial position while 44 occurred in AP-medial position.

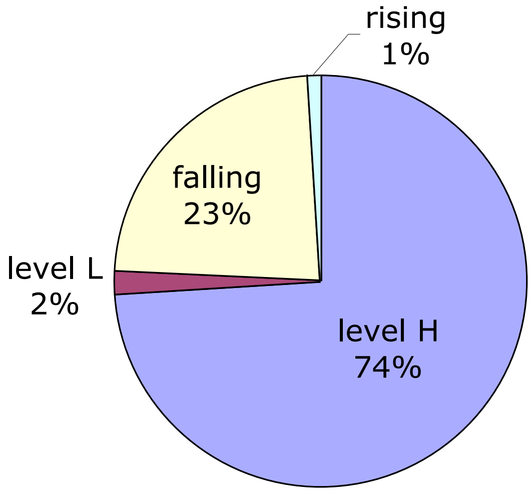

*Figure 11. Tonal patterns of AP-initial multisyllabic content words (with H tone inducers)*

As shown in Figure 11, 74% of multisyllabic content words with an H inducer had a level H tone (HH) at the beginning of word, and 23% of them had a falling tone (HL) at the same position. Rising tone (LH) was not common with the words that started with an H inducer. This tendency was a little stronger in the Read speech corpus, where 75.5% of words occurred with the level H tone, than the Radio corpus, where 70.9% of words occurred with a level H tone.

Segments with different laryngeal features do not seem to strongly induce specific tones in AP-medial position as they do in AP-initial position (Jun, 1993, 1996b). However, there is a possibility that laryngeal features of segment still affect phonetic tone realization. Thus, I will report the AP-medial words with the L tone inducers separately from those with the H tone inducers.

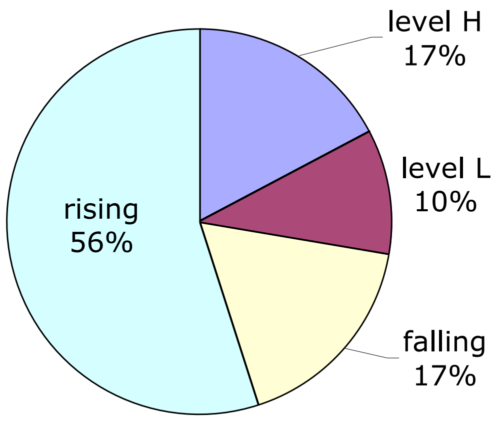

*Figure 12. Tonal patterns of AP-medial multisyllabic content words (with L tone inducers)*

Figure 12 shows the tone patterns on the first two syllables of AP-medial multisyllabic words with L tone inducers. 56% of the words started with a rising tone in AP-medial positions. Separate analysis of the two corpora revealed that 65.6% of the words started with a rising tone in the Read speech corpus, while only 41.7% of the words started with a rising tone in the Radio corpus. Further observation showed that 69 out of 89 words that started with a rising tone in AP-medial position had their word offsets aligned with AP-final syllables, indicating that the rising pattern observed in this position is the realization of AP-final rising tone (e.g., the second LH in the LH**LH** tone pattern).

Overall, regardless of the location of words within the AP, 84% (1236 out of 1471 words) of multisyllabic content words with an L inducer start with a rising tone pattern (LH) in the current data.

The number of AP-medial words with an H tone inducer was very small (n=44), but more than half of the words occurred either with level H tone (HH) or with falling tone (HL), as shown in Table 3. Word-initial rising tone was not commonly observed in this position, unlike the pattern observed with the words with an L tone inducer.

| | level H | falling | level L | rising | TOTAL |
|---|---|---|---|---|---|
| number of words | 13 | 14 | 8 | 9 | 44 |

*Table 3. Tonal patterns of AP-medial multisyllabic content words (with H tone inducers)*

### 3.4 Discussion

The data confirmed that the AP is an appropriate unit for modulating lexical access, since the average number of content words contained within an AP was not more than two, which is consistent with the previous finding of Jun and Fougeron (2000).

The most frequent AP tone patterns were LH, LHH, and LHLH for L tone inducers and HH, HHLH, and HHL for H tone inducers. The default patterns (THLH) were frequent only when the AP had more than three syllables. In general, AP-final H tone is frequently observed (more than 85% for L inducers and more than 70% for H tone inducers), which suggests that this pattern may help word segmentation by successfully marking AP boundaries.

Further, the data showed speakers' robust use of specific tonal patterns and very consistent occurrence of words in AP-initial position across different styles of speech: About 90% of multisyllabic content words appear in AP-initial position; more than 85% of multisyllabic content words begin with a rising tone when there is an L inducing segment in AP-initial position; 74% of multisyllabic content words with H tone inducers are produced with HH tone in the AP-initial position.

Taken together, these findings confirm that the AP can be an efficient unit for word segmentation, and suggest that the probabilistic information of phrasal-level intonation in the speech input (i.e., the frequency of post-lexical tonal patterns imposed upon content words and the co-occurrence frequency of AP onset and content-word onset) could be useful segmentation cues in Korean.

# CHAPTER 4

# THE ROLE OF ACCENTUAL PHRASE TONAL PATTERNS ON WORD SEGMENTATION IN KOREAN

## 4.1 Introduction

The results of the corpus study presented in the previous chapter showed that AP-final H tone and AP-initial rising tone are frequently observed in spoken Korean, and that content-word onsets are usually aligned with AP onsets. This further leads to the fact that the majority of content words start with a rising tone. In this chapter, I will present a perception experiment, which examines whether Korean listeners exploit post-lexical tonal patterns in word segmentation and whether they are sensitive to the frequency information of post-lexical tonal patterns.

As aforementioned in Chapter 2, a recent study by Welby (2003) reported that French listeners can use post-lexical intonation information for word segmentation. In particular, she found that the presence of the early rise which is often observed at the beginning of a phonological phrase facilitates segmentation, and that the location of early rise also affects word segmentation. At least on the surface, French intonation and Korean intonation are similar. They both have phrasal-level prominence (Jun & Fougeron, 2000; Welby, 2003)[10], and the smallest prosodic phrase (the Accentual Phrase, in Jun & Fougeron (2000; 2002)'s model) also has a similar default tone pattern, with two consecutive rising tones (LHLH). However, the two languages differ, for instance, in terms of the location of function/content words and tonal alignment. The L tone in the initial rise usually falls on a function word in French, given that it is a head initial language. Thus the H tone following the initial L tone usually falls on the onset of a content word. French listeners seem to be 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, as observed in Chapter 3. Thus, the 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 the H tone in word-medial position. The question arises: will this rising pattern in AP-initial position in Korean still contribute to word segmentation, as the same pattern does in French?

The current experiment will investigate: (i) whether various non-contrastive tonal patterns show different effects for word segmentation in Korean, (ii) whether frequent tonal patterns can expedite word segmentation, and (iii) whether the location of a word within the AP affects the way listeners segment words from the speech stream. Six post-lexical tonal conditions were tested in order to answer these questions.

The organization of this chapter is as follows. Section 4.2 will explain the experimental design, including the general procedure for the word spotting task (4.2.1), word selection criteria (4.2.2), tonal distribution in the speech stream for the current

[10] French stress is also controversial as to whether it is lexical or post-lexical. See Welby (2003) for detailed discussion.

experiment (4.2.3), acoustic analysis of the stimuli (4.2.4), and the detailed procedure for the present experiment (4.2.5). The results of the experiment will be reported in section 4.3, and they will be further discussed in section 4.4.

## 4.2 Experimental Design

### 4.2.1 Word Spotting

The current study employed the word spotting task, which has been extensively used in psycholinguistic studies on speech segmentation and word recognition since Cutler and Norris (1988)'s initial work. The general procedure of the task is as follows. During the experiment, participants listen to a series of speech streams, one at a time. Each stream is composed of a string of nonsense syllables that contains a real word. Listeners are supposed to detect the embedded real word from the stream. They are asked to press a button as soon as they hear the word embedded in the string, and then to say the word. Reaction time, which is usually measured from the button press, is one of the two measurements for this task. Since no information about the target words is given to participants in advance, listeners often miss quite a few words or say a word which is not the intended target. Thus, error rate, which includes the number of missed targets and incorrect responses, is also used as a measurement.

The procedure for the current experiment followed similar steps to those described above, but with some variations in stimuli and procedure. In previous studies (Cutler, Mehler, Norris, & Segui, 1992; Cutler & Norris, 1988; McQueen, Norris, &

Cutler, 1994; Vroomen & de Gelder, 1995), 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 took part in this experiment were exposed to speech streams that consisted of sixteen syllables and contained disyllabic or trisyllabic target words. Also, the current experiment measured reaction time by voice activation, without button pressing. More details of the experimental procedure will be discussed in the following sections.

### 4.2.2 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. The four selection criteria considered were as follows: syllable structure, word initial segments, frequency and familiarity of target words.

#### *Syllable Structure of Target Words*

Words that contained only CV syllables were selected as targets. Thus, the disyllabic target words had a CV.CV structure and the trisyllabic target words had a CV.CV.CV structure. Words which contained a CVC syllable anywhere were excluded as target words because a coda consonant might cause phonological and/or phonetic changes to the following onset consonants, which may potentially affect the accuracy and reaction time of the responses.

### ***Word Initial Segments of Target Words***

Word-initial segments were also controlled in the target word selection process. It has been observed in earlier studies (Jun 1993, 1996b) and confirmed in the previous chapter that in Korean, aspirated and tense consonants trigger a high tone in Accentual Phrase initial position, while other consonants maintain a low tone in the same position (Jun, 1993, 1996b). Since word-initial H tone was not the target of investigation here, words that start with an aspirated or tense obstruent were excluded from the target word selection. Consequently, 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/, /tʃ/) as the word initial segment.

### ***Frequency of Target Words***

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). Words were selected as targets only if they satisfied at least one of the following criteria set for the frequency database.

Target words were selected from the most frequent 3000 content morphemes (across grammatical categories) in the Frequency Analysis of Korean Morpheme and Word Usage (Kim & Kang, 2000) database. The criterion for the KAIST Concordance

Program (KCP) Online Demo Version (KAIST, 1999) was that disyllabic target words were to be in the high frequency range, that is, from ranking 5000 and lower (1 ~ 5000), and that trisyllabic target words were to be in the low frequency range, that is, from ranking 5001 and higher (5001 ~ ) in the database. The frequency range for trisyllabic target words was set lower since most of the high frequency trisyllabic words had another word contained in them, which was not desirable for the current experimental task. Although selecting disyllabic target words from the low frequency range could have been an option in order to balance the frequency range between two syllable count sets (i.e., disyllabic words and trisyllabic words), I did not do so because it could have made the task too difficult.

### ***Familiarity of Target Words***

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 who were living in Korea at the time when the survey was conducted and 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. The average score of the disyllabic targets was 3.9, and that of the trisyllabic targets was 3.4.

***Foil Words Selection***

There was the same number of foil words as target words for each syllabic category (i.e., 16 disyllabic foil words and 16 trisyllabic foil words). The foil words were not controlled as strictly as the targets. Unlike the target words, which only contained CV syllables, the foil words were allowed to have more than one syllabic structure. None of the foil words, however, started with aspirated or tense consonants. The frequency and familiarity of foil words were not controlled, but the familiarity rating was similar: the average score was 3.4 for disyllabic words and 3.5 for trisyllabic words. The word selection criteria are summarized in Appendix 1, and detailed frequency and familiarity rating information about the target words is given in Appendix 2 and Appendix 3.

### 4.2.3 Locations and Tonal Patterns of Words

#### 4.2.3.1 Locations of Target 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. In order to increase a variety of contexts and to prevent the listeners from predicting the possible locations of the target words, the filler-bearing streams contained a filler word either in the first or the third AP. The APs that did not contain the filler word had various tonal patterns other than LHLH. This is summarized in (4).

(4) a. Target bearing streams ($\sigma = CV$)

$$\{[\sigma\ \sigma\ \sigma\ \sigma]_{AP1}\ \ [\sigma\ \underset{\substack{\uparrow \\ \text{target word}}}{\sigma}\ \sigma\ \sigma]_{AP2}\ \ [\sigma\ \sigma\ \sigma\ \sigma]_{AP3}\}_{IP}$$

b. Filler bearing streams ($\sigma = CV$)

$$\{[\sigma\ \underset{\substack{\uparrow \\ \text{filler word}}}{\sigma}\ \sigma\ \sigma]_{AP1}\ \ \underset{\textit{or}}{[\sigma\ \sigma\ \sigma\ \sigma]_{AP2}}\ \ [\sigma\ \underset{\substack{\uparrow \\ \text{filler word}}}{\sigma}\ \sigma\ \sigma]_{AP3}\}_{IP}$$

### 4.2.3.2 Locations of Target/Filler Words within an AP

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. In the case of the AP-medial condition, the target/filler words started at the second syllable of an AP.

### 4.2.3.3 Tonal Patterns of Words

The conditions were further divided by their tonal patterns to examine if post-lexical intonational tones play a significant role in word segmentation. In this section, I will first enumerate the tonal patterns that were employed for the study, and then explain why they were selected.

Among the seven attested surface AP tonal patterns beginning with a low tone (Jun, 2000), the current experiment employed three tonal patterns, LHLH, LHHH, and LLLH[11]. In addition, the experiment included one unattested tonal condition, LLRH[12]. LLRH is not a variant of LLLH, but different from LLLH in that the third syllable of LLRH is higher than the preceding L, whereas the third syllable of LLLH is level to the preceding L  (see Figure 17 and Figure 18 in Section 4.2.4.1 for comparison). As will be explained later, this pattern was created in order to see the effect of rising intonation in non-AP-initial position.

As each AP was composed of four syllables, each tone in a tonal pattern mapped onto one syllable. 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 LLRH, instead of the LHHH pattern. Thus, each target word had six conditions based on the location within the AP and the tonal pattern, as summarized in Table 4. In Table 4, bold and italicized characters indicate the locations of the disyllabic target words and underlining indicates the locations of the trisyllabic target words.

[11] The other four of the seven patterns are LH, LHL, LHLL, LL. APs ending with a low tone were avoided because they are rare. LH was not employed, because it was not common when the AP contained more than 3 syllables (see Figure 6 in the previous chapter).

[12] R indicates rising (viz., the interpolation from the L on the second syllable to the H on the fourth syllable) on the third syllable.

| AP-initial condition | AP-medial condition |
|---|---|
| ***L H*** L H | L ***H L*** H |
| ***L H*** H H | L ***L R*** H |
| ***L L*** L H | L ***L L*** H |

*Table 4. Tonal patterns and locations of target words in each condition.*

### *Tonal contrast I: Presence vs. Absence of H from the preceding AP*

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. 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. AP-final H tone is more frequent than AP-final L tone, as shown in Chapter 3. In addition, since the final H tone of an AP usually marks the end of a syntactic phrase, it is likely that the final H tone would simultaneously indicate the upcoming beginning of another 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.

### *Tonal contrast II: Rising vs. Non-rising*

Rising tonal patterns were distinguished from 'non-rising' patterns in both AP-initial and AP-medial positions. AP-initial rising is the default AP tonal pattern of Seoul

Korean, and the corpus study in Chapter 3 confirmed that words would most frequently co-occur with a rising tone pattern when they appear in AP-initial position. If listeners were sensitive to this co-occurrence frequency, they could make use of this information for segmentation. Based on this, we could hypothesize 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 (LR) pattern in L***LR***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***LR***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. It was expected that this specific pattern would reveal whether listeners pay more attention to the local rising pattern or to the overall AP tone pattern when they are engaged in the segmentation task. If segmentation relies more on the local co-occurrence frequency of word-initial rising tone and word onset, listeners would prefer L***LR***H to L***HL***H and L***LL***H. If listeners are more attentive to the overall AP tone pattern, rather than the local rising tone pattern, they would show better performance when a word occurs in frequent and familiar AP tonal patterns, such as L***HL***H and L***LL***H, than in L***LR***H.

***Tonal contrast III: AP-initial LHLH vs. AP-initial LHHH***

***LH***LH and ***LH***HH were contrasted within the AP-initial condition. When target words are disyllabic, ***LH***LH has a L tone (the second L in ***LH***LH) after the initial rise whereas ***LH***HH has a H tone. This contrast enables us to compare the effect of a L tone to that of a H tone (the second H in ***LH***HH) after the initial rising tone. The second L in ***LH***LH can be viewed as a falling tone (from the H at the second syllable to L at the third syllable), which may indicate the offset of a word, and hence, potentially facilitate the segmentation decision. In contrast, the H tone continues even after the word offset in ***LH***HH, and thus, this pattern may not give as much word boundary information as ***LH***LH. Therefore, if the falling tone after the target word helps listeners make a faster segmentation decision by giving them a sense of discontinuity, their performance should be better in ***LH***LH than in ***LH***HH. This distinction between ***LH***LH and ***LH***HH was maintained in the trisyllabic target word items as well, but for a different reason. In the trisyllabic word 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.

***Tonal contrast IV: AP-medial LHLH vs. AP-medial LLLH***

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.

The following table summarizes these tonal patterns. The '+' symbol indicates the presence of the facilitative tonal factor described above, and the '-' symbol indicates the lack thereof.

| Position in AP | Tonal pattern | Preceding H tone | Initial rising | Frequency of overall AP pattern |
|---|---|---|---|---|
| Initial | […H]$_{AP1}$ [***L H*** L H]$_{AP2}$ [L…] $_{AP3}$ | + | + | frequent |
| | […H]$_{AP1}$ [***L H*** H H]$_{AP2}$ [L…] $_{AP3}$ | + | + | less frequent |
| | […H]$_{AP1}$ [***L L*** L H]$_{AP2}$ [L…] $_{AP3}$ | + | - | less frequent |
| Medial | […H]$_{AP1}$ [L ***L L*** H]$_{AP2}$ [L…] $_{AP3}$ | - | - | less frequent |
| | […H]$_{AP1}$ [L ***H L*** H]$_{AP2}$ [L…] $_{AP3}$ | - | - | frequent |
| | […H]$_{AP1}$ [L ***L R*** H]$_{AP2}$ [L…] $_{AP3}$ | - | + | not attested |

*Table 5. Potential facilitative and disruptive factors for the selected AP tonal patterns. AP2 contains target words.*

The results of the current experiment will depend on which of the three tonal factors weighs more in word segmentation and whether the different factors cumulatively affect word segmentation. If the influences of different tonal factors are cumulative, we may even predict the detailed ranking of tonal factors from most to least facilitative by counting the number of '+' symbols in each tonal category from Table 5. In any case, we can make at least the following predictions: if the final H tone in the AP that precedes a target word has an influence on segmentation, listeners' performance will be much better in the AP-initial conditions, i.e., **LH**LH, **LH**HH and **LL**LH, than those in the AP-medial conditions, i.e., L**HL**H, L**LL**H and L**LR**H. If the word-initial rising tone has a role in facilitating segmentation, the tone patterns with initial rising, namely, **LH**LH, **LH**HH and L**LR**H, will be more helpful to listeners than the other conditions. However, if the frequency and overall phrasal tone pattern of the AP are more crucial in segmentation, L**LR**H will not show any facilitative effect.

### 4.2.4 Stimuli

#### 4.2.4.1 Recording and Manipulation of Stimuli

The nonsense speech streams for the experimental stimuli were recorded onto Digital Audio Tape in a sound-attenuated booth by a female native speaker of Seoul Korean (the author), and digitized at a sampling rate of 22kHz. 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.

The following figures (Figure 6 ~ Figure 11) illustrate the locations and the tonal patterns of a sample disyllabic target word within a carrier speech stream. The areas with light gray color are the locations of the target word and the areas with dark gray color indicate the remaining two syllables of the target-bearing AP. The remaining syllables either follow (in case of AP-initial condition, Figure 13 ~ Figure 15) or surround (in case of AP-medial position, Figure 16 ~ Figure 18) the target word within the AP.

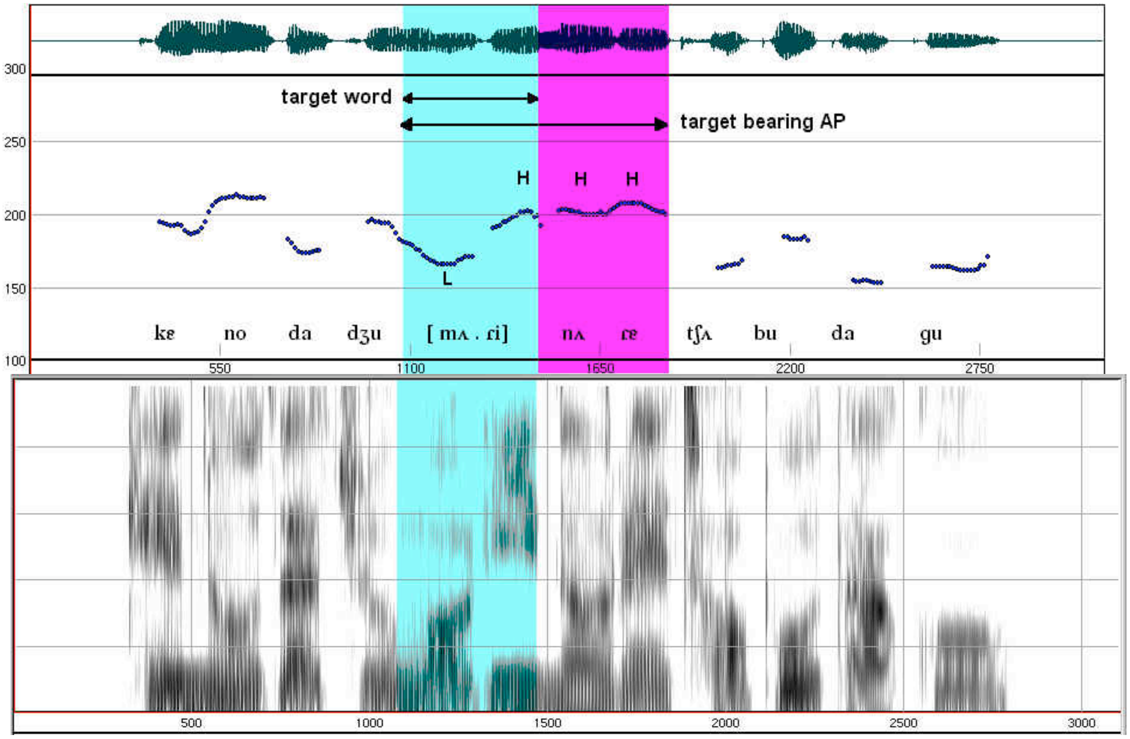

*Figure 13. meori [mʌɾi] 'head' in AP initial* ***LH****HH pattern*

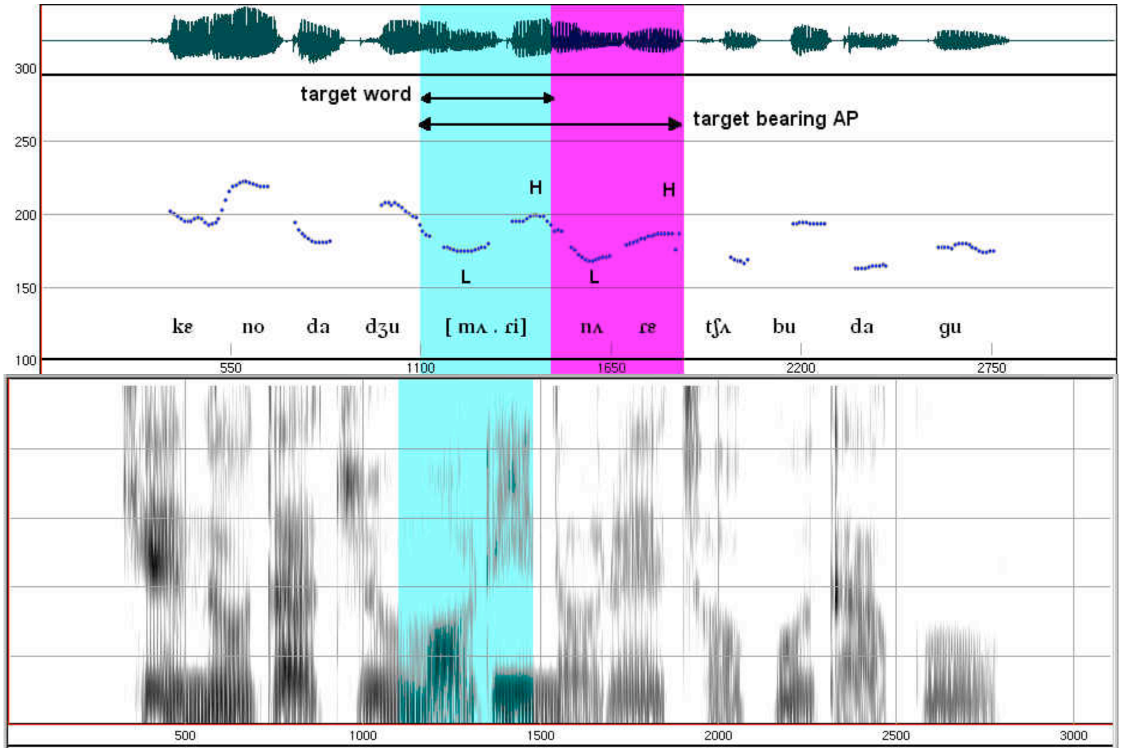

*Figure 14.* meori *[mʌɾi] 'head' in AP initial* **LH***LH pattern*

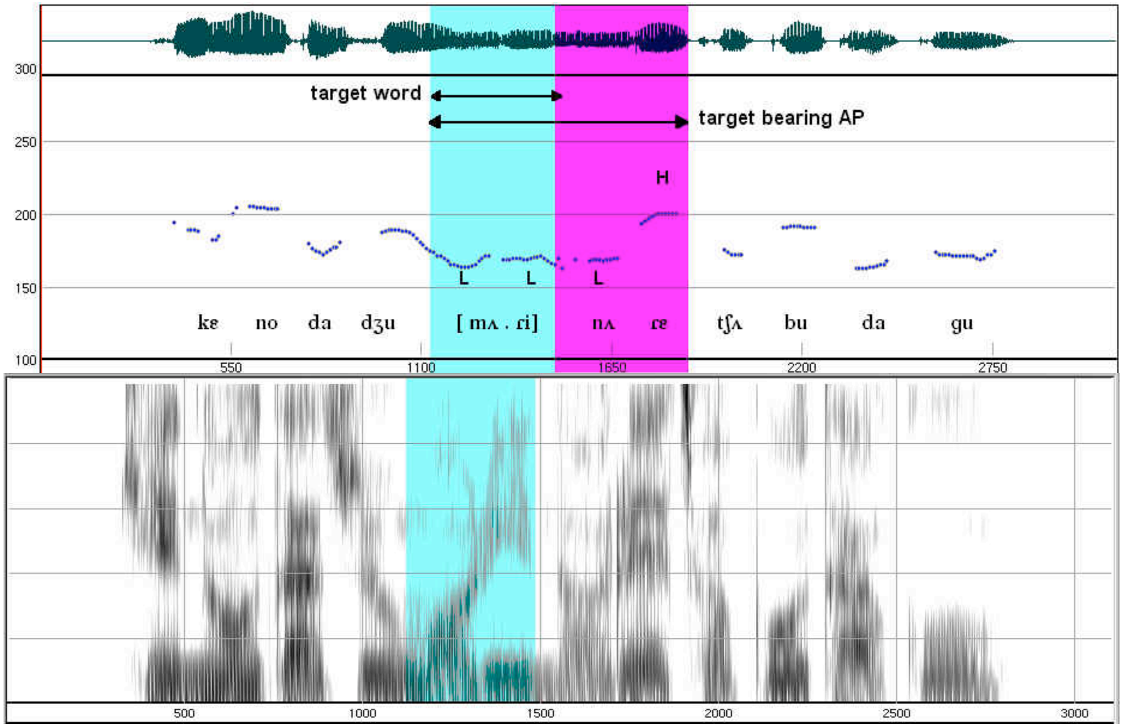

*Figure 15*. meori *[mʌɾi] ‘head’ in AP initial* **LL***LH pattern*

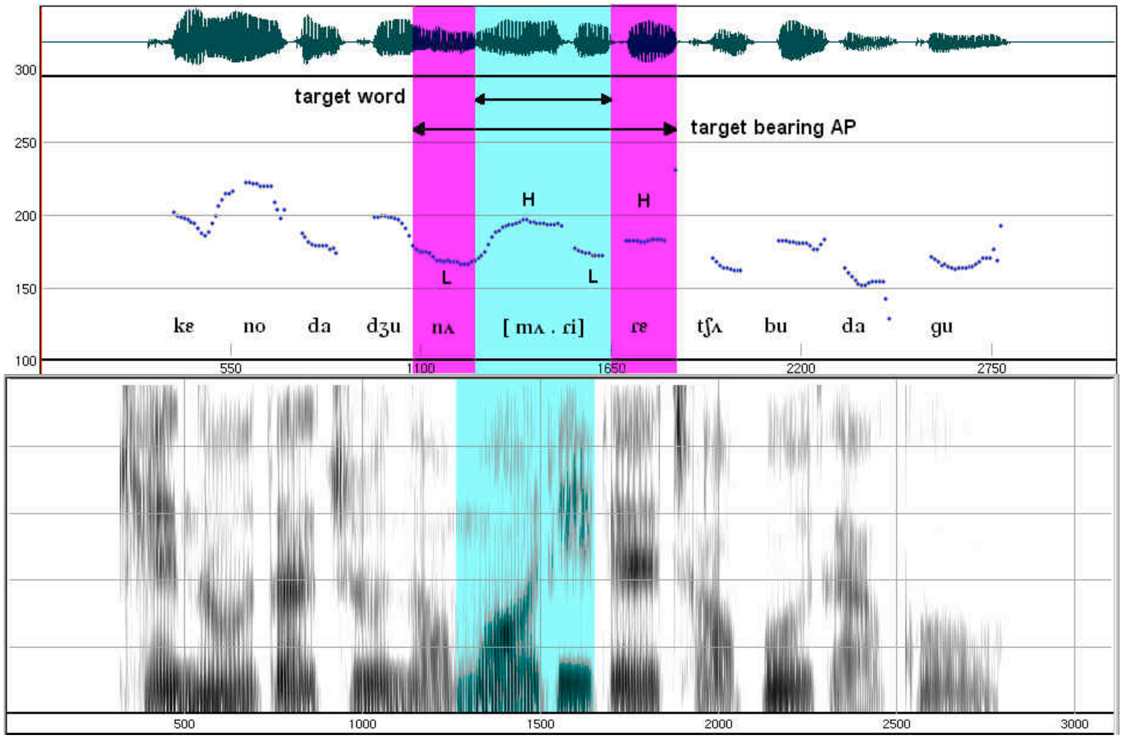

*Figure 16*. meori *[mʌɾi] ‘head’ in AP medial L***HL***H pattern*

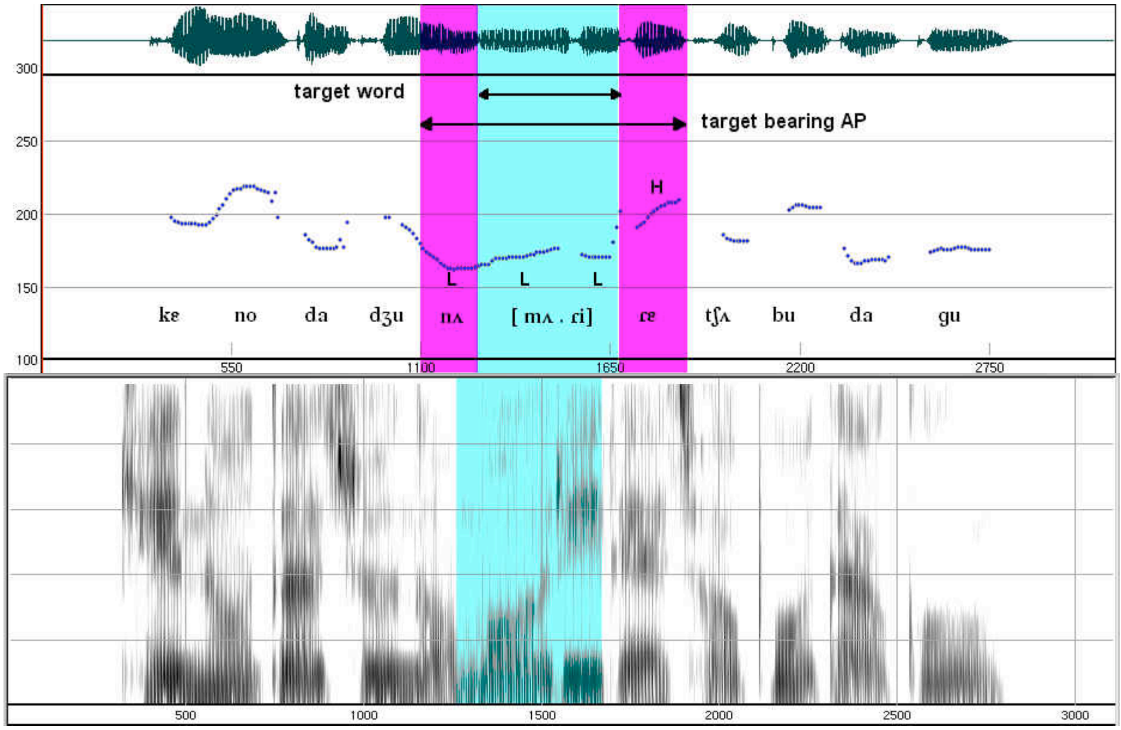

*Figure 17.* meori *[mʌɾi] 'head' in AP medial L***LL***H pattern*

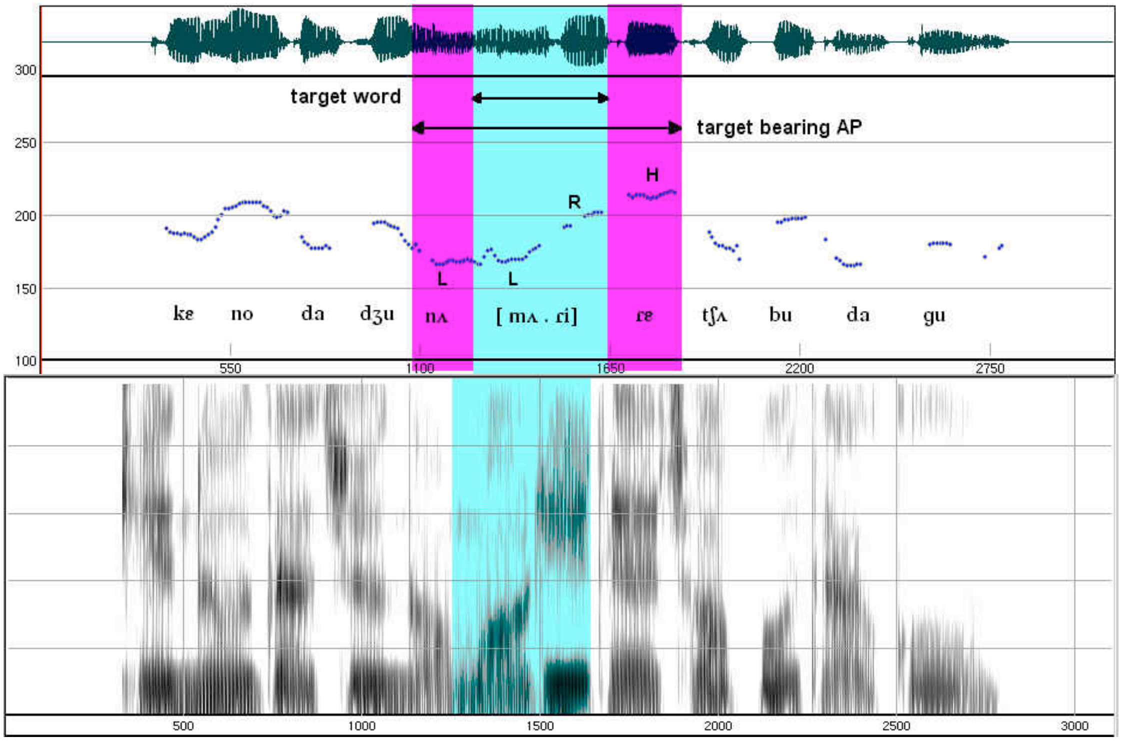

*Figure 18.* meori *[mʌɾi] 'head' in AP medial L***LR***H pattern*

#### 4.2.4.2 Phonetic measurements of the stimuli

As mentioned above, each target word for the experimental task was produced naturally within a given tonal context. This stimulus generation process implies that a target word could have different phonetic values for both segmental and suprasegmental properties other than F0, depending on the context that it was embedded in. Since the differences in these phonetic details of the stimuli may affect the results of the perception experiment, it was necessary to analyze the acoustic characteristics of the stimuli as they were produced. If it turned out that there was no other significant acoustic difference among the target words across the different tone conditions, but the listeners' reactions varied, we would be able to conclude that it is the tonal pattern that causes the different responses. Or, if there were acoustically significant and systematic differences in target words depending on their environments, the results of acoustic measurement will allow us to predict whether the acoustic cue in question could facilitate or interfere with segmenting words in specific environments. In order to examine the potential influence of stimuli on the results of the perception experiment, I measured the following acoustic parameters of the target-bearing stimuli.

1. ***AP duration***: The durations of the target-bearing APs were measured from the offset of the F2 of the vowel that directly precedes the target AP to the F2 offset of the AP-final vowel.

2. ***Word duration***: The duration of target words were also measured. The duration measures were taken from the offset of the F2 of the vowel that precede the target word to the F2 offset of the word-final vowel.

3. ***Word initial syllable duration***: The duration of the first syllable of the target words was measured. It was measured from the offset of the F2 of the preceding vowel to the offset of the vowel of the first syllable.

4. ***Word initial syllable energy***: The average energy of the first syllable of the target words was measured.

5. ***Closure duration***: The closure duration of the initial consonant of the target words was measured, except for one word which started with a glide. For oral stops, the closure duration was measured from the offset of F2 of the preceding vowel to the stop burst. For nasal stops, the closure duration was measured from the onset to the offset of nasal energy.

6. ***Voice Onset Time***: Lag VOT of the word-initial oral stop of target words was measured from the stop release to the voice onset of the following vowel. There were three AP-medial lenis stop consonants which were fully voiced; a VOT value of zero was employed for those tokens.

7. ***RMS burst energy***: The energy at the stop burst was measured from the FFT spectrum. A 10 ms window (128 points) was centered over the release of the stop consonants and the RMS value was taken over a frequency range of 1000-10000 Hz.

8. ***Nasal energy minimum***: The energy during the word-initial nasal consonants of target words was measured at the lowest point of the RMS acoustic profile, using a 10 ms window (128 points).

#### 4.2.4.3 Results of stimulus measurements

Recall that there were 32 target words, and that one speaker produced each of them with 6 different tonal patterns. Since the size of the data set available from the acoustic analysis was so limited, Friedman's test, a rank-based nonparametric equivalent of repeated measures, was employed for the statistical analysis of the measurements.

1. ***AP duration***: The durations of APs were not significantly different between the initial target bearing APs and the medial target bearing APs ($X^2$ (5) = 1.446, $p$ = 0.9192). No difference in AP duration was observed among different tonal patterns, either.
2. ***Word duration***: There was no significant durational difference between the target words in AP-initial positions and those in AP-medial positions ($X^2$ (5) = 4.566, $p$ = 0.4711). There was no effect of tonal pattern on word duration either.
3. ***Word initial syllable duration***: There was a significant difference in word-initial syllable duration across the tonal patterns ($X^2$ (5) = 23.069, $p$ < .001). As shown in Table 6, the mean duration was longer and the sum of ranks was lower for the words in AP-medial positions than the words in AP-initial positions.

|  | AP-initial |  |  | AP-medial |  |  |
|---|---|---|---|---|---|---|
|  | ***LH***HH | ***LH***LH | ***LL***LH | L***HL***H | L***LL***H | L***LR***H |
| sum of ranks | 119 | 114 | 128 | 83 | 75 | 90 |
| mean duration | 201 | 201 | 197 | 211 | 207 | 206 |

*Table 6. Sum of ranks and mean duration (ms) of word initial syllable*

Dunn's post test revealed that there was a significant difference between ***LH***HH (initial) and L***LL***H (medial) (*p* < .05), between ***LH***LH (initial) and L***LL***H (medial) (*p* < .05), between ***LL***LH (initial) and L***HL***H (medial) (*p* < .05), and between ***LL***LH (initial) and L***LL***H (medial) (*p* < .001). No other significant difference was found.

4. ***Word initial syllable amplitude***: Amplitude of word initial syllables differed significantly across the tonal patterns ($X^2$ (5) = 70.557, *p* < .001). The mean amplitude (dB) was higher and the sum of ranks was lower for the word initial syllables in AP-medial positions than those in AP-initial positions. The results are summarized in Table 7.

|  | AP-initial |  |  | AP-medial |  |  |
|---|---|---|---|---|---|---|
|  | ***LH***HH | ***LH***LH | ***LL***LH | L***HL***H | L***LL***H | L***LR***H |
| sum of ranks | 128 | 110 | 145 | 44 | 113 | 69 |
| mean amplitude | 67.1 | 67.8 | 65.3 | 72.6 | 67.9 | 71.19 |

*Table 7. Sum of ranks and mean amplitude (dB) of word initial syllable*

Dunn's post tests showed that all three tone patterns in the AP-initial condition, namely, ***LH***HH, ***LH***LH, and ***LL***LH, had significantly lower initial amplitude than the two tonal patterns in the AP-medial condition, namely, L***HL***H and L***LR***H (*p* < .01 for ***LH***LH vs. L***LR***H; *p* < .001, for the other combinations). Among AP-medial tone patterns, L***LL***H had significantly lower amplitude than both L***HL***H (*p* < .001) and L***LR***H (*p* < .01).

5. ***Closure duration***: The closure durations for both word initial oral stops and word initial nasal stops did not reveal any statistically significant difference among the tone patterns.

6. ***Voice Onset Time***: The results revealed that VOT values of the word initial lenis stop consonants were significantly different across tonal patterns ($X^2$ (5) = 62.71, $p$ < .0001). As illustrated in Table 8, the sums of ranks were lower and mean VOT durations were longer for the tonal patterns in the AP-initial condition than those in the AP-medial condition.

|  | AP-initial |  |  | AP-medial |  |  |
|---|---|---|---|---|---|---|
|  | ***LH***HH | ***LH***LH | ***LL***LH | L***HL***H | L***LL***H | L***LR***H |
| sum of ranks | 37 | 32 | 35 | 86 | 85.5 | 81.5 |
| mean duration | 52.85 | 50.45 | 51.84 | 25.42 | 27.68 | 26.22 |

*Table 8. Sum of ranks and mean VOT durations (ms) for word initial stop consonants*

Dunn's post test showed that the VOT values of the three tonal patterns in the AP-initial condition were significantly longer than those of the three tonal patterns in the AP-medial condition ($p$ < .001, in all of the pair comparisons). However, no statistically significant difference was found either within the three AP-initial tonal patterns or within the three AP-medial tonal patterns.

7. ***RMS burst energy***: There was no difference depending on the tonal patterns in RMS burst energy at the stop release.

8. ***Nasal energy minimum***: Nasal energy minimum was not significantly different across the tonal patterns.

To summarize, significant differences among the tonal patterns were found only for the amplitude and duration of the word initial syllables and for the VOT values of the word initial stop consonants. These parameters, however, do not show the same tendency. On the one hand, the results showed that word initial syllables were louder and longer when they were in AP-medial position than in AP-initial position. On the other hand, the results showed that 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. Specifically, we can anticipate the following consequences: i) if amplitude and/or duration of the word-initial syllables help word segmentation, listeners' performance should be better when words are in AP-medial position than in AP-initial position, ii) if initial consonants with a strong acoustic cue help word segmentation, listeners' performance should be better when words start with a stop consonant than a sonorant consonant.

### 4.2.5 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 which was located directly in front of the participant. Participants' oral responses were recorded into another Macintosh computer using the Sound Edit program. The Psyscope software measured RT from the onset of each stimulus to the onset of the particpants' verbal responses, but 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. This adjustment was for the benefit of overall RT analysis of target words with different numbers of syllables (without a proper adjustment, shorter RT values were expected for disyllabic words than for trisyllabic words, since the durations of disyllabic words were inherently shorter than those of trisyllabic words). Missing or incorrect responses were assessed by listening to the participants' responses during the experimental session, and they were scored as errors.

## 4.3 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 %. Since the high error rates resulted in a lot of missing data for the RT analyses, the current experiment took the error rates as the primary dependent measure. In this section, the results of error rates will be reported first, and the RT results will follow.

For the statistical data analysis, mixed model analyses were employed, rather than repeated measures (RM) ANOVA. The mixed model analysis is also known as a "multi-level model", "hierarchical linear model" or "mixed effects model" (Quene & van den Bergh, in press). Unlike RM ANOVA, this analysis combines all the random factors into a single analysis. This model 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 able to accurately capture statistical effects than a RM ANOVA analysis when there are missing data points. The fact that this analysis can adeptly handle missing data points was a considerable advantage for the current study because there was a sizable number of missing data points for the RT analysis (see (Quene & van den Bergh, in press) for more information on psycholinguistic research and mixed model 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.

### *Error Rates*

The mixed model analysis showed that there was a highly significant effect of tone (F (5, 2791.698) = 187.647, $p < .0001$). Figure 19 shows the mean error rates of six different tonal conditions.

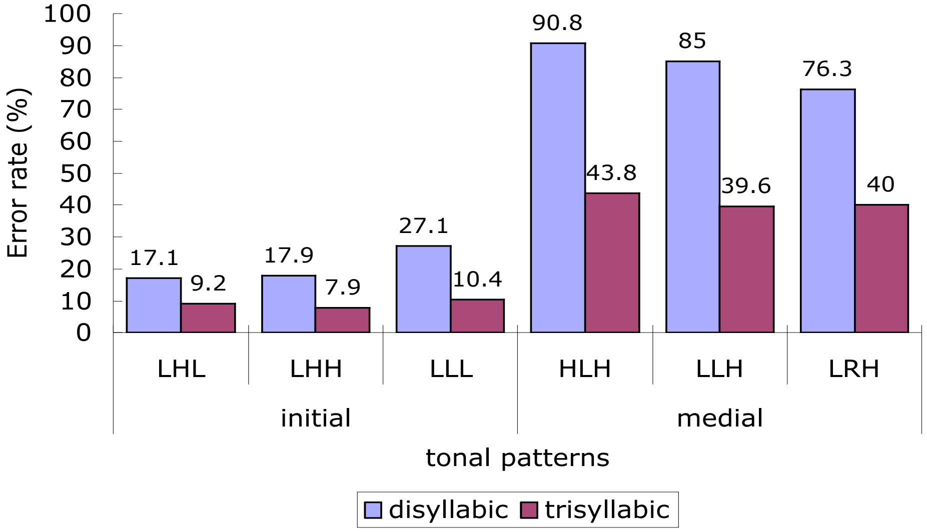

*Figure 19. Mean error rate in six tonal patterns*

The location effect was large and significant (F (1, 2799.654) = 916.654, $p < .0001$). The target words were detected with lower error rates when they were in AP-initial position than in AP-medial position, as shown in Figure 20. The error rate in AP-medial condition was four times higher than in AP-initial condition.

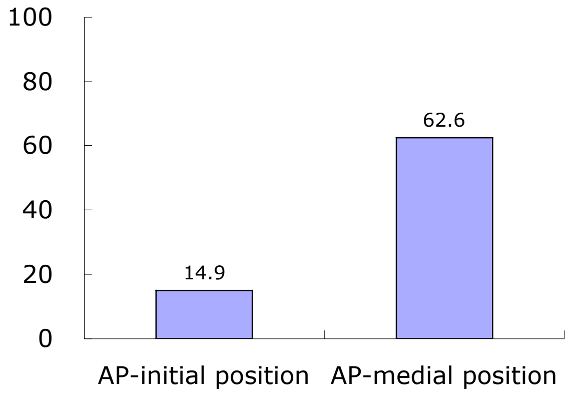

*Figure 20. Mean error rates (%) in AP-initial position and AP-medial position*

There was also a significant effect of the number of syllables in the target words. Listeners made more errors while spotting disyllabic words than trisyllabic words, and the difference was highly significant (F (1, 2789) = 245.020, *p* < .0001). As shown in Figure 21, over 50 % of target words were missed in disyllabic words, which was about twice the error rate for trisyllabic words.

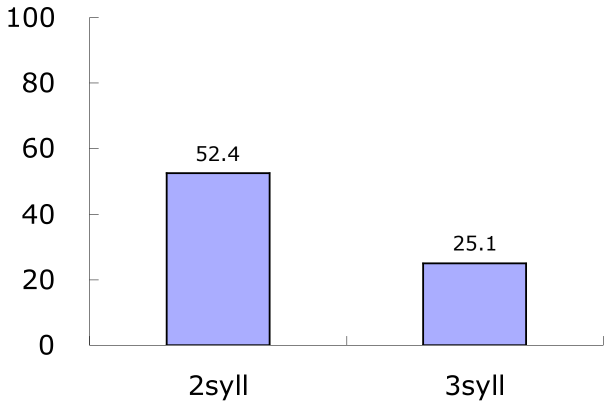

*Figure 21. Mean error rates (%) for disyllabic and trisyllabic words*

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) than with a non-sonorant consonant (stops and affricates). Figure 22 shows this contrast.

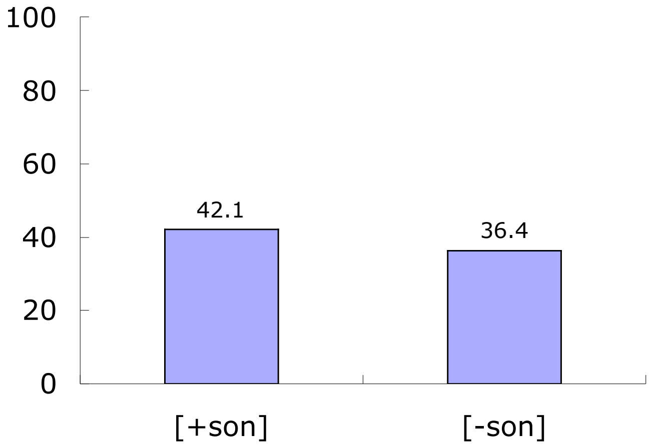

*Figure 22. Mean error rates (%) of [+son] onsets and [-son] onsets*

There was an interaction between the location within the AP and number of syllables in the target words (F (1, 2810.967) = 115.432, *p* < .0001), between the number of syllables in the target words and tone patterns (F (5, 2790.827) = 24.930, *p* < .0001), and between consonant type and tone patterns (F (12, 1006.480) = 220.616, *p* < .0001). Note 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.

Thus, subsequent mixed model analyses were performed separately for the disyllabic initial condition, disyllabic medial condition, trisyllabic initial condition and trisyllabic medial condition. In the disyllabic initial condition, the tone effect was significant (F (2, 639.223) = 4.674, *p* = .01). Further, *post-hoc* pairwise comparisons showed that the error rates for ***LH***HH and ***LH***LH were significantly lower than for ***LL***LH (*p* < .01, *p* = .012, respectively). There was no statistically significant difference between ***LH***HH and ***LH***LH. In the disyllabic medial condition, the tone effect was highly significant (F (2, 639.762) = 10.128, *p* < .0001). The error rates for L***HL***H and L***LL***H were significantly higher than those for L***LR***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. There was no significant tone pattern effect in trisyllabic words, regardless of their location within the AP. There was no consonant type effect in the disyllabic initial and disyllabic medial positions. However, a significant consonant type effect was found in the trisyllabic initial position (F (1, 678.076) = 21.370, *p*

< .0001) and in the trisyllabic medial position (F (1, 666.576) = 5.880, *p* = .016). In trisyllabic initial position, there were more errors for sonorant consonants than for non-sonorant consonants. However, in trisyllabic medial condition, the results were opposite: listeners missed more targets when the word onset was non-sonorant than when it was sonorant.

### *Reaction Time*

In general, most of the main effects and interactions found in the error rates data were also found in the RT data. A highly significant overall tone effect was found (F (5, 1679.662) = 9.430, *p* < .0001). Similar to the results from error rates, RT's were significantly faster in each of the three initial-tonal contours compared to the three medial-tonal contours in both disyllabic and trisyllabic words. Mean RT values in different tonal conditions are shown in Figure 23.

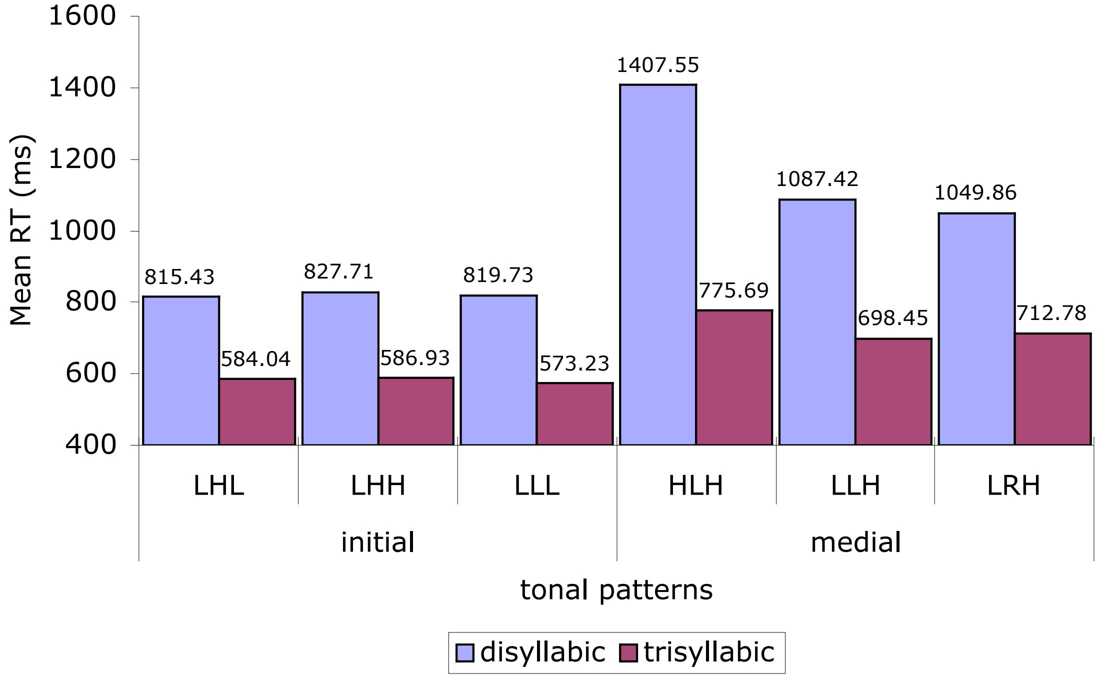

*Figure 23. Mean RTs (ms) in different tonal conditions*

A mixed model analysis performed on RT showed that there were significant effects of the location of the target word in an AP. RT was significantly faster for the words in AP-initial position than those in AP-medial position (F (1, 1472.455) = 123.697, $p < .001$), as shown in Figure 24.

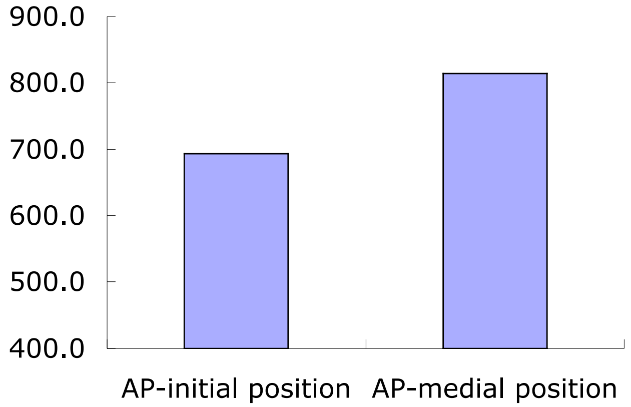

*Figure 24. Mean Reaction Time (ms) in AP-initial position and AP-medial position*

The difference in RT between disyllabic and trisyllabic words was also statistically significant (F (1, 1491.219) = 70.607, *p* < .001). Participants were faster in detecting trisyllabic words than disyllabic words, as shown in Figure 25.

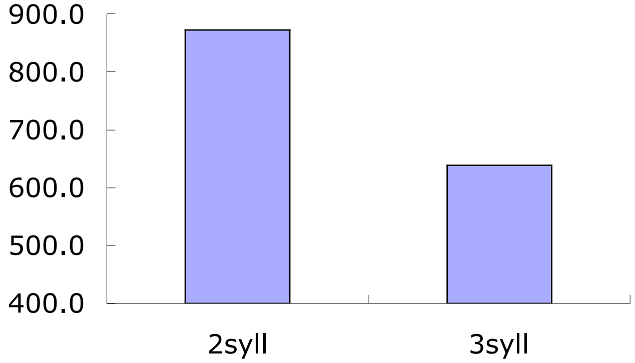

*Figure 25. Mean Reaction Time (ms) of disyllabic and trisyllabic words*

Again, a significant consonant effect was observed (F (1, 1681.761) = 17.230, *p* < .0001). Mean RT was faster when a target word started with a non-sonorant consonant (mean = 764.04 ms, SD=413.9) than with a sonorant consonant (mean = 707.76 ms, SD=407.0).

There was an interaction between the number of syllables in words and tone patterns (F (5, 1674.841) = 5.513, *p* < .0001), between the number of syllables in words and the location within the AP (F (1, 1685.261) = 12.351, *p* < .0001), and between consonant type and tone patterns (F (12, 702.489) = 75.936, *p* < .0001).

However, the RT data did not reveal as much of a tonal pattern effect as error rate data did. Mixed model analyses were performed separately on the four conditions (divided by the syllable count of words and the location of words within AP) and a significant RT difference was found only in disyllabic medial condition (F(2, 100.447) = 4.4480, *p*=.014). A *post-hoc* pairwise comparison showed that listeners were slower in spotting words in L***HL***H than L***LL***H (*p* = .018) and L***LR***H (*p* < .01). No other significant difference among tone patterns was found. The RT results revealed a consonant effect in a subset of these conditions. In the disyllabic initial condition (F (1, 493.655) = 9.026, *p* < .01), participants were faster in detecting words with a sonorant onset than a non-sonorant onset. A consonant effect was also found in the trisyllabic initial condition (F (1, 572.658) = 22.965, *p* < .0001). However, in this condition, the tendency was opposite to the disyllabic initial condition. That is, participants detected the words with a non-sonorant onset faster than the words with a sonorant onset. There was no consonant type effect in disyllabic medial and trisyllabic medial conditions.

## 4.4 Discussion

***Effect of Syllable Count on Word Segmentation of Korean***

As shown in the previous section, 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.

Unfortunately, the results seem 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* [mosʌɾi] 'edge' contains a disyllabic word *seori* [sʌɾi] '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, 1986; Luce & Pisoni, 1998). 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 (Forster & Chambers, 1973; Norris, 1986; Rubenstein, Garfield, & Millikan, 1970). In the current study, although trisyllabic words were in a lower frequency range (see 4.2.2), 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 (Havens & Foote, 1963; Hood & Poole, 1980; Luce, 1986; Pisoni, Nusbaum, Luce, & Slowiaczek, 1985).

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.

### *Differences in Onset Consonant*

The results of the study showed that there was a difference between obstruent and sonorant word-onsets. In the test stimuli, stop consonants revealed clear VOT differences depending on their location within the AP, as opposed to nasal consonants, which did not show any acoustic differences across tonal conditions. However, this difference did not affect the general results of the experiment. Sonorant and non-sonorant initial segments showed a similar tendency in terms of their interaction with various tonal patterns. That is, both of them showed an AP-initial advantage and a syllable count effect. Further, regarding the tonal pattern effect which will be discussed below in detail, they did not differ from each other. Listeners' detection ability, however, was significantly different depending on the word initial consonant. Their performance was much better, that is, more accurate and faster, when the initial consonant was an obstruent 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 word initial segment, can influence the speed of lexical access.

### *Positional Effect in Word Segmentation*

Listeners spotted words more reliably in AP-initial position than in AP-medial position, and RT was also faster in AP-initial than in AP-medial position. Within the category of rising pitch patterns (***LH***LH, ***LH***HH, L***LR***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***LR***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***LR***H). These results strongly imply that there was a positional effect which exerted more power than tonal contours during the word segmentation task.

There are at least two possible explanations for the observed preference for words that are located AP initially. The first account for the AP-initial preference is that this positional effect is due to AP-finality. In the current experiment, the target words were always in the second AP in speech streams of three APs. Thus, the words that were inserted in AP-initial position were always preceded by another AP. As the stimuli of the current experiment were read naturally, the AP-final syllable of the AP that preceded the target word had all the prosodic information appropriate to an AP boundary. If Korean listeners use post-lexical prosodic cues in word segmentation, it is very likely that they make use of this AP-final rising tone and final lengthening as a marker of the beginning of the following AP, thus, the following word. That is to say, these prosodic cues can be a reliable segmentation cue since they can allow listeners to anticipate when a new word will begin even before the beginning of the word is actually perceived.

Another possible explanation is that this result is due to a domain-initial strengthening effect (Cho & Keating, 2001; Fougeron & Keating, 1997; Keating, Cho, Fougeron, & Hsu, 2004). That is, the acoustic cue at the onset of the AP can expedite the detection of AP-initial words. There have been a great deal of speech production studies on the segmental variation caused by prosodic context. Segments were found to be stronger in articulatory magnitude in domain-initial position than in non-initial position.

This “domain-initial strengthening” effect has been widely explored in several languages, including English (Dilley, Shattuck-Hufnagel, & Ostendorf, 1996; Fougeron & Keating, 1997; Pierrehumbert & Talkin, 1992), Spanish (Lavoie, 2000), French (Fougeron, 1999), Taiwanese (Hsu & Jun, 1998), and Korean (Cho & Keating, 2001; Jun, 1993, 1995a; Kim, 2001). The effect is cumulative, and thus, speakers’ production reveals hierarchical prosodic structure to a certain degree. This kind of cue may provide listeners with the information for one of the basic functions of prosodic structure, namely, “chunking” or “grouping”, and hence, can affect listeners’ segmentation ability. That is, the strengthened acoustic cues at prosodic phrase-initial positions make it easier for listeners to find the beginning of speech units, and hence, may help the segmentation process of spoken input.  In fact, a recent study by McQueen and Cho (2003) provided direct evidence in support of this conjecture. They found that, in two-word sequences, IP-initial strengthening of the initial portion of the second word aided native listeners of American English in segmenting the first word. This suggests 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 in production studies (Keating et al., 2004).  A future experiment which can exclusively test the domain initial strengthening effect would be desirable in order to determine the effect of AP-initial strengthening in word segmentation by Korean listeners.

### *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.

#### ***a. Rising vs. Non-rising***

There was a significant difference between rising tonal patterns and non-rising tonal patterns both in AP-initial and AP-medial positions. The words with rising tonal contours (***LH***LH, ***LH***HH in AP-initial position; L***LR***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). This result is similar to that for French (Welby 2003), in that post-lexical intonation affected the speed of word segmentation. More importantly, the result strongly underscores the role of a rising pattern imposed upon the word in word segmentation in Korean. Although the occurrence of a rising contour in the middle of an AP (i.e., L***LR***H) has not been attested in Seoul Korean, the unacceptability of that AP tonal pattern did not interfere with listeners' segmentation process. It is likely that listeners perceived the local rising tone (LR) that spans over a disyllabic word as an AP-final H tone, as they were hearing speech signal unfolds in time. Recall that the previous corpus study showed that about 78% of AP-medial words started with a rising tone and had their word offsets aligned with AP-final syllables, and 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 (like the target words selected for the current experiment). Taken together, these results further suggest a possibility that frequent word-specific tone pattern information is stored in a memory, associated with individual words. This hypothesis becomes more plausible under the exemplar-based speech perception model, where each label (e.g., lexical representation) is associated with detailed memories of percepts (exemplars). According to an episodic theory of the lexicon (Goldinger, 1997), perceptual details, such as voice and intonation contour, leave a unique memory trace and can later affect the accuracy of word recognition. Church and Schacter (1994) also claimed that implicit memory for spoken words retains very specific auditory details, including intonation contour. The storage of this peripheral information should be different from the storage of lexical prosodic information such as stress or lexical tone. However, it is probable that this detailed information is stored, as words are perceived as a whole and stored as exemplars which constitute the basic substrate of the lexicon (Goldinger, 1998; Johnson, 1997).

***b. 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. That suggests that a falling tone to a low tone after the target word does not necessarily aid segmentation by marking an AP boundary after the target word. It also means that the tonal pattern cue after the target word may not give out the same perceptual information as a reinforced segmental cue. Recall that in McQueen and Cho (2003)'s study, the existence of IP-initial segments after a target word helped listeners to segment the target word faster. The result of the current study can be considered along the same line as the ceiling effect of trisyllabic target words. That is, it implies that once the word is recognized in one way or another, the tonal pattern that follows the recognition point does not facilitate segmentation.

***c. 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. This 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.

***Summary***

To summarize, the results of this experiment showed that Korean listeners are sensitive to the location of words within a prosodic phrase such that they are better at segmenting words from a speech stream when a word is in AP-initial position than in AP-medial position; that the number of phonological neighbors have greater influence on segmentation than tonal patterns do; that initial rising tone helps segmentation in Korean; and that falling tone on a word interferes with segmentation.

These results suggest that the post-lexical tonal pattern of a phrasal prosodic unit can help word segmentation just as well as lexical prosody can.

# CHAPTER 5

# THE ROLE OF PROSODIC CUES AT ACCENTUAL PHRASE BOUNDARIES ON WORD SEGMENTATION OF KOREAN

## 5.1 Introduction

The experiment that will be reported in this chapter examines whether the acoustic cues accompanying Accentual Phrase boundaries facilitate word segmentation in Korean. Since the AP can be an efficient word boundary indicator in Korean (as shown in Section 2.2.3), it is expected that the ability to detect prosodic cues at AP boundaries can help Korean listeners to extract words from a speech input.

Previous speech production studies have shown that Korean AP boundaries can be characterized by at least three prosodic cues: AP-final pitch rise, AP-final lengthening, and AP-initial amplitude effect. As mentioned in the previous two chapters, the canonical tonal pattern of a Korean AP includes a high tone at the phrase final position (Jun, 1993). Although a low tone can be aligned with AP-final position, it happens only rarely. 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.

Phrase-final lengthening can also mark AP boundaries in Korean. 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. While the findings are inconsistent regarding AP-final lengthening, studies have all agreed that phrase-final lengthening definitely exists as a property of a higher prosodic phrase (i.e., the Intonational Phrase) in Korean (Cho & Keating, 2001; Chung et al., 1996; Jun, 1993, 2000).

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, since it was the AP-final syllable, but the first syllable only had a strong intensity value, not a high pitch value.

Based on these observations, the current experiment will specifically examine whether pitch rise and final lengthening at the AP-final position would facilitate segmentation. In effect, the result of the perception experiment presented in the previous chapter (i.e., AP-initial words were easier to detect than AP-medial words) already suggests that these cues would play a significant role in word segmentation. These two cues were also known to aid word segmentation in French (Bagou et al., 2002), and in

English (Saffran, Newport et al., 1996). Further, the current study examines whether an increase in AP-initial amplitude can help word segmentation in Korean. The question that will be addressed further is whether all the prosodic cues are equally helpful in letting listeners be aware of the existence of AP boundaries and making them initiate lexical access. Thus, the effects of the three prosodic cues will be compared.

Additionally, this study will investigate whether the exploitation of prosodic cues for word segmentation is determined by the prosodic characteristics of a given language or by any perceptually salient prosodic cue, regardless of listeners' language background. In order to do so, the effect of an AP-initial H tone on word segmentation will be tested. This cue does not conform to the prosodic patterns of the AP when the AP-initial segment is not [+stiff vocal cords]. Further, cross-linguistic comparisons will be made by contrasting the result of the current experiment with the results from previous studies of other languages.

The organization of this chapter is as follows. Section 5.2 will explain the experimental method (5.2.1), introduce the prosodic conditions that were adopted for the current study (5.2.2), describe how the stimuli for the current study were made (5.2.3), discuss how the stimuli were recorded and acoustically manipulated (5.2.4), and explain the experimental procedures (5.2.5). The results of the experiment will be presented in section 5.3, and further discussed in section 5.4.

## 5.2 Experimental Design

### 5.2.1 Artificial Language Learning

In order to test the role of individual prosodic cues in speech segmentation, this experiment employed the artificial language learning method that was initiated by Hayes and Clark (1970), who tested the role of transitional probability in the segmentation of non-linguistic auditory stimuli such as glides and warbles. Saffran and her colleagues (1996) adopted this method and applied it to linguistic sounds and word segmentation. They created a simple artificial language which was composed of 6 trisyllabic words. These words did not have any syntactic or semantic content. Each syllable was synthesized separately with the same duration and F0 value, and then the syllables were put together to make a word. Then, they made a 21-minute sound sequence by concatenating the six words in random order without any pause at the word boundaries. In their experiment, subjects listened to this sound sequence during a learning phase, and then went through a testing phase, in which they had to identify the words that they had learned during the learning phase. The results of their experiment revealed that, in the absence of acoustic cues, listeners were able to learn the words of the artificial language by depending solely on distributional information in the speech stream.

One of the advantages of their method that is directly relevant to the current study is that an experimenter can control acoustic factors and/or positional factors independently, without interference from other confounding factors, given that all the syllables originally had equal duration, F0 value, and intensity. For instance, Saffran et al. (1996) themselves inspected whether word-final vowel lengthening accompanying transitional probability information could facilitate listeners' segmentation ability when compared to the condition without any acoustic cue. In the condition with word-final vowel lengthening, they were able to manipulate just the durational property of the target syllable in word-final position, with all other acoustic factors still being equal as in the baseline conditions. This ensured that they were observing the effect of vowel lengthening only, and not other factors. The same method was also adopted in Bagou et al.'s (2002) study, where they compared the contribution of phrase-final duration to that of a phrase-final pitch cue in word segmentation of French by manipulating the relevant acoustic cues. Thus, I adopted this method, considering that it can be effectively used in examining how various acoustic cues affect listeners' segmentation ability and how important each cue's contribution is. Another benefit of this method is that the result of the experiment can allow us to infer how segmentation cues affect language learning itself, because the experimental procedure forces listeners to store the words they extracted from the speech stream, although it is just for a short term.

### 5.2.2 Experimental Conditions

The acoustic parameters tested in this experiment were duration, pitch, and amplitude. In the current experiment, the stimuli were composed of six trisyllabic words (see the following section for details), and each word formed an AP by itself in the experimental conditions where prosodic cues were present. In this way, the acoustic cues for AP edges were also cues for word edges.

The basic hypothesis for this experiment was that the presence of a prosodic-acoustic cue in the input speech stream would help word segmentation and the storage of segmented words. That is, listeners would show better performance in a segmentation task when the condition with a prosodic cue was presented to them than when the condition without any such information was given, as long as the prosodic-acoustic cue in the condition conformed to the characteristics of Korean prosody. In order to test this hypothesis and to investigate which prosodic cue was most facilitative in speech segmentation of Korean, the following five prosodic conditions were created: ***No Prosody***, ***Duration***, ***Amplitude***, ***Pitch Final*** and ***Pitch Initial***. As is apparent from the name of each condition, prosodic conditions were defined by the presence and the location of one of the three prosodic cues.

The ***No Prosody*** condition was the baseline condition where only distributional information was available, without any additional prosodic cues. Since it has been shown that both adults and infants can successfully segment words from speech streams using transitional probabilities only (Johnson & Jusczyk, 2001; Saffran, Aslin et al., 1996), it was hypothesized that listeners' performance would be better than chance in this condition, although the effect might be small.

The rest of the conditions had one of the prosodic cues in addition to the distributional cue of the baseline condition. There were three "conforming conditions", which followed the prosodic characteristics of the Korean AP, as introduced in Section 5.1. They were the ***Duration*** condition, the ***Amplitude*** condition, and the ***Pitch final*** condition. The ***Duration*** condition had a lengthened syllable at the word-final (thus AP-final) position. Phrase-final duration has been known to play a facilitative role in word segmentation in other languages, such as English (Nakatani & Schaffer, 1978; Saffran, Newport et al., 1996), and French (Bagou et al., 2002).

The ***Amplitude*** condition had a syllable with increased intensity at the word-initial (thus AP-initial) position. As mentioned earlier, the effect of AP-initial amplitude is reported in Jun (1995b)'s production study. The speech environment of Jun's data (e.g., trisyllabic words, APs in the sentence medial position) is similar to the one that will be used in the current experiment. Thus, the increased intensity value of the AP-initial syllable is expected to facilitate segmentation.

The words in the ***Pitch final*** condition had an increased F0 value at the word-final (thus AP-final) syllable. Based on the observed characteristics of the Korean AP, it was hypothesized that a high tone in word-final position can mark the end of the word quite effectively, and hence will aid segmentation. A phrase-final pitch rise was shown to facilitate word segmentation in French, and its effect was even stronger than the effect of phrase-final lengthening (Bagou et al., 2002).

Finally, the experiment has a non-conforming condition, which is the ***Pitch initial*** condition. Although many studies have shown that prosodic patterns that match the characteristics of listeners' native language facilitate word segmentation, no study has been conducted to directly test the role of a post-lexical prosodic cue which does not conform to the prosodic properties of a given language.

A pitch cue was selected for the non-conforming condition, rather than duration or amplitude cues, because duration and amplitude cues may provide ambiguous information in that they can be stronger at both edges of the AP than in AP-medial positions, although the acoustic strength (e.g., length and intensity) of cues may differ depending on the location. For instance, AP-initial acoustic lengthening of onset consonants is observed, as well as pre-boundary vowel lengthening (i.e., lengthening in the AP-final position) in Korean (Cho & Keating, 2001). Similarly, a stronger amplitude cue can exist at the final, as well as, at the initial position of the AP, accompanying the high pitch cue at the AP-final position (Jun, 1995b). The pitch cue, however, does not cause this kind of ambiguity, as long as the word-initial consonant is controlled in the AP-initial position. As noted in Jun (1993; Jun, 1996b) and as observed in the current corpus study (Chapter 3), AP-initial high pitch is attested in Korean when the onset consonant is either a tense or an aspirated consonant.

In the ***Pitch initial*** condition, word-initial syllables had an increased F0 value. 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 et al., 2002; Cutler & Norris, 1988; Cutler & Otake, 1994) and phonotactics (Weber, 2001)), it is possible that this non-conforming prosodic condition would hinder speech segmentation because of the conflict between the distributional cue and the prosodic cue (word-initial high pitch, rather than word-final high pitch). If this is indeed the case, it is expected that Korean listeners would perform worse in the ***Pitch initial*** condition than they would in the other prosodic conditions, and possibly even worse than in the baseline (***No prosody***) condition. Alternatively, if any prosodic property can help 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, with its regular pitch pattern, can help Korean listeners. If this is the case, listeners' performance in this condition could be as good as their performance in other conforming prosodic conditions, and better than the baseline condition.

### 5.2.3 Stimuli

The artificial language used for the current experiment had four consonants (three lenis consonants (/p/, /t/, /k/) and one nasal consonant (/m/)) and four vowels (/a/, /i/, /u/, /ɛ/). The combination of the eight segments resulted in 16 distinct CV syllables as follows: /pa/, /pi/, /pu/, /pɛ/, /ta/, /ti/, /tu/, /tɛ/, /ka/, /ki/, /ku/, /kɛ/, /ma/, /mi/, /mu/, /mɛ/. 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. Two out of the six words had real Korean words embedded in them. They were *maku* ('carelessly' or 'horse equipment') in *makute* and *kume* ('purchase') in *kumepi*. In fact, *kumepi* ('price for purchase') itself is also a compound in Korean. But, as we will see in Section 3.3, this did not affect listeners' performance at all[13]. Each phoneme

[13] During and after the experiment, listeners did not seem to realize that these words could be real Korean words. It may be due to the fact that they were specifically told that the words that would be used in the experiment had nothing to do with any real Korean words even if they might sound similar. It may also be due to the fact that recorded words were not phonetically natural (see Section 5.2.4)

occurred 4-5 times in this word list. Two syllables, /ku/ and /pi/, occurred twice in the word list, whereas the other fourteen syllables occurred just once.

In order to construct the speech stream for the experiment, the six words were concatenated in random order without a pause between words in the following way. First, one block was made by repeating the six words in different random orders three times, thus containing eighteen concatenated words.  Six blocks were constructed in this way[14]. Since each block was composed of eighteen words, there were 108 words in the six blocks. Then, the 6 blocks were repeatedly concatenated eight times, which yielded about a 10-minute speech stream. Each word was presented 144 times (= 3 occurrences per block × 6 blocks × 8 repetitions) in the speech stream. One word never occurred twice in a row, and there were no pauses 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. Note that in Saffran, Newport, et al. (1996)'s study, the transitional probabilities within words ranged from 0.31 to 1, and those across words ranged from 0.1 and 0.2.  Thus, both the transitional probabilities of words and the difference between within- and across- words transitional probabilities were higher in this study than those in Saffran et al.'s study. The range of within- and across-words transitional probabilities did not overlap in either study.

[14] Blocks were designed as follows, where each number represents a word:
(1:  *putaki*, 2: *makute*, 3: *tipemu*, 4: *kapitu*, 5: *mikepa*, 6: *kumepi*)
BLOCK1: 123456 / 316425 / 643521
BLOCK2: 452631 / 251364 / 536142
BLOCK3: 164253 / 213564 / 526341
BLOCK4: 632145 / 315426 / 435261
BLOCK5: 652341 / 342156 / 432165
BLOCK6: 246153 / 216354 / 152643

### 5.2.4 Recording and Manipulation of Stimuli

A female native speaker of Korean (the author) produced each CV syllable 10 times. Out of the 10 recorded tokens, only one token was chosen as a stimulus for a CV syllable. During the recording, each syllable was produced independently, following a pause, in order to avoid any potential coarticulation effect. If each word was produced as a whole, this could have helped listeners tremendously by giving coarticulation information. The coarticulation cue is, as reviewed before, a very robust cue for word segmentation; it is stronger than either the transitional probability cue (Johnson & Jusczyk, 2001) or stress cues (Mattys, in press), and it can also provide prosodic structural information (Cho, 2001). Note also that the words created by concatenating independently produced syllables do not form phonetic words in that they do not show any potential allophonic variations within a word (or an AP). For example, the words, as created, do not have any voiced lenis stop which would have cued within-word cohesiveness (Jun, 1993, 1995a).

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 a word, and as described in the previous section, the words were concatenated into a 10-minute speech stream which was used as the baseline condition. Praat software was used for the concatenation and the sound manipulation processes.

Duration, amplitude, and pitch values of this ten-minute baseline speech stream were further manipulated, generating four prosodic conditions. Each prosodic cue had a different criterion in terms of the degree of manipulation. The details of the data manipulation process will be described in the following section.

### *A. No Prosody (baseline) condition*

The ***No Prosody*** condition was a prosodically neutral condition which did not have any salient prosodic cues. Thus, listeners were supposed to rely only on the transitional probability information in this condition.

Since the stimuli were constructed from naturally produced speech, the acoustic values of the syllables were not exactly the same even after the normalization process, although deviations were very small. In the baseline condition, the normalized average syllable duration was 271 ms (*SD* = 1.8 ms). In the syllables that contained oral stops, the average closure duration was 51 ms (*SD* = 1 ms), the average VOT duration was 47 ms (*SD* = 12 ms), and the average vowel duration was 173 ms (*SD* = 11 ms). In the syllables that contained nasal stops, the average consonant duration was 62 ms (*SD* = 4 ms) and the average vowel duration was 209 ms (*SD* = 5 ms). The average F0 value at vowel-medial position was 191.4 Hz (*SD* = 1.1 Hz), and the average F0 value of the whole syllable was 189.4 Hz (*SD* = 1.4 Hz). The average syllable amplitude was 67.4 dB (*SD* = 2.0 dB), and the average amplitude of the vowels was 71.1 dB (*SD* = 1.8 dB). In order to make the speech stream sound a little more natural, the amplitude was ramped over 30 ms at offset for each syllable.

For this condition and the other conditions where the syllable duration was not manipulated, the speech stream lasted for 10 minutes and 17 seconds.

### *B. Duration condition*

The ***Duration*** condition imitated AP-final lengthening. Thus, in this condition, the duration of word-final syllables was lengthened. This condition employed the degree of lengthening used in Bagou et al. (2002)'s French study, which was 30%[15]. In the current study, the rate of average lengthening was 30.6% (84 ms) of the original syllables, which produced an average syllable duration of 355 ms (*SD* = 2 ms), where the average vowel duration was 265 ms (*SD* = 9 ms).

Note that various speech production studies have presented different increase rate in terms of Korean AP-final lengthening. Jun (1993; 1995a) did not find any significant AP-final lengthening. Cho and Keating (2001) found AP-final vowel lengthening of 50% (from around 60 ms in word-final, non-AP final vowel /a/, to around 90 ms in word-final, AP final vowel /a/) 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 (85.8 ms to 96 ms) at the AP-final position. In Chung et al.'s (1996) study, where one speaker read a newspaper column (500 words, 23 sentences) with three different speech rates, AP-final syllables showed a 20% increase in syllable length on average. The degree of lengthening used in the current study is smaller

[15] Bagou et al. (2001) did not specify what the base of this amount lengthening was. However, this 30 % of final lengthening corresponded with French AP-final lengthening reported in Jun and Fougeron (2000).

than the degree observed in Cho and Keating (2001), larger than that in Oh (1998) and Chung et al. (1996), and similar to the minimum degree of lengthening in IP-final syllables in Korean as observed by Chung et al. (1996), which was a 30% increase compared to the preceding syllables. However, this percentage is much smaller than the average IP-final lengthening, which showed a 77% increase compared to the average length of non-phrase-final syllables (Chung et al., 1996), and approximately a 100% increase compared to the average length of AP-phrase final syllables (Cho & Keating, 2001).

Due to the lengthening of the word-final syllables, the duration of the whole speech stream was longer (11 minutes and 19 seconds) than the other conditions.

### *C. Amplitude condition*

The ***Amplitude*** condition imitated AP-initial strengthening. In this condition, the amplitude of word-initial syllables was increased. The amplitude of the initial syllables was strengthened by 7.1 dB, to an average of 74.5 dB (*SD* = 2.6 dB), which was a 10.5% increase from the original amplitude. This rate was simply based on subjective evaluation of saliency by the author. An estimate of the appropriate rate of amplitude increase was not obtainable from any previous production studies.

### *D. Pitch final condition*

The ***Pitch final*** condition imitated AP-final pitch rising, and hence, the F0 value of word-final syllables was increased. The average pitch of the manipulated syllables was

212.8 Hz (*SD* = 1.3 Hz) at the vowel-medial position, and the average pitch of the whole syllable was 214.6 Hz (*SD* = 1.0 Hz). This was a 24.9 Hz increase on average compared to the corresponding syllables in the ***No prosody*** condition. The increase rate was 13 %. 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[16]. The degree of F0 increase in this condition was also similar to that obtained from a quantitative study conducted by Jun (1995b). In that study, 3 speakers read sixteen words of 2-5 reiterated syllables, where each word formed one AP. The data from two trisyllabic words ('mamama' and 'tatata') 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 (from 215 Hz to 240 Hz[17]).

***E. Pitch initial condition***

Syllables with high F0 values were located in word-initial position in this condition. This pattern is not observed in Korean with AP-initial lenis and nasal consonants, which were the segments of the artificial language used in this study. These consonants occur with normal rising pattern (LH), where the highest peak is not realized on the first syllable.

The average values of increased F0 in the ***Pitch initial*** condition were about the same as those used in the ***Pitch final*** condition. The average pitch of the manipulated

[16] The data was obtained from the stimuli for the experiment that was introduced in Chapter 4.

[17] The numbers were inferred from a graph provided in Jun (1995b).

syllables was 214.2 Hz (*SD* = 2.1 Hz) at the vowel medial position, and the average pitch of the whole syllable was also 214.2 Hz (*SD* = 2.2 Hz). This was a 24.7 Hz increase on average compared to the corresponding syllables in the ***No prosody*** condition, which yielded an increase rate of 13%.

### 5.2.5 Procedure

Sixty native speakers of Seoul Korean participated in the experiment, and they received a small honorarium for their participation. 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.

A. Learning Phase

Subjects heard the 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, the information about the number of words in the language was not given to subjects. The speech stream was presented to subjects using SoundEdit on a Macintosh computer, and a SONY MDR-V200 headphone. Subjects were asked to adjust the volume to the most comfortable level. The learning phase lasted approximately 11 minutes for the ***Duration*** condition, and 10 minutes for the other conditions.

### B. Testing Phase

Thirty-six forced-choice pairs were made from combinations 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. For example, from the word '*mikepa*', the final two syllables '*kepa*' were taken, and a random syllable '*ma*' was attached to '*kepa*', thus creating a part-word '*kepama*'. 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 a sequence of syllables that subjects had never heard during the learning phase. Thus, transitional probabilities of non-words were zero.

The pairs were presented auditorily to the subjects. There was an 800 ms inter-stimuli interval between the two strings in a pair. Each trisyllabic string (real words, part-words, and non-words of the artificial language) occurred 6 times during the test 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 '1' key on the keyboard if the string presented first was their answer, or '0' key on the keyboard if the string presented second was their answer. The test stimuli were presented and the responses were collected using Psyscope software on a Macintosh computer.

All the words used in the testing phase had the same prosodic cue as in the learning phase[18]. For instance, when a subject heard the ***No prosody*** condition during the learning phase, she/he was given the strings with no prosodic cue in the testing phase, and when a subject heard the ***Pitch final*** condition during the learning phase, all the strings presented in the testing phase had high pitch on the word-final syllables.

## 5.3 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, showed at chance or below chance performance. The ***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 the ***Amplitude*** condition. All the participants for these two conditions showed above chance performance. In the ***Pitch final*** 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 the ***Pitch initial*** 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. The results are illustrated in Figure 26.

[18] The testing phase had the same prosodic cue as the learning phase, because a pilot study showed that listeners' performance was very poor across the board when the testing phase did not have any prosodic cue.

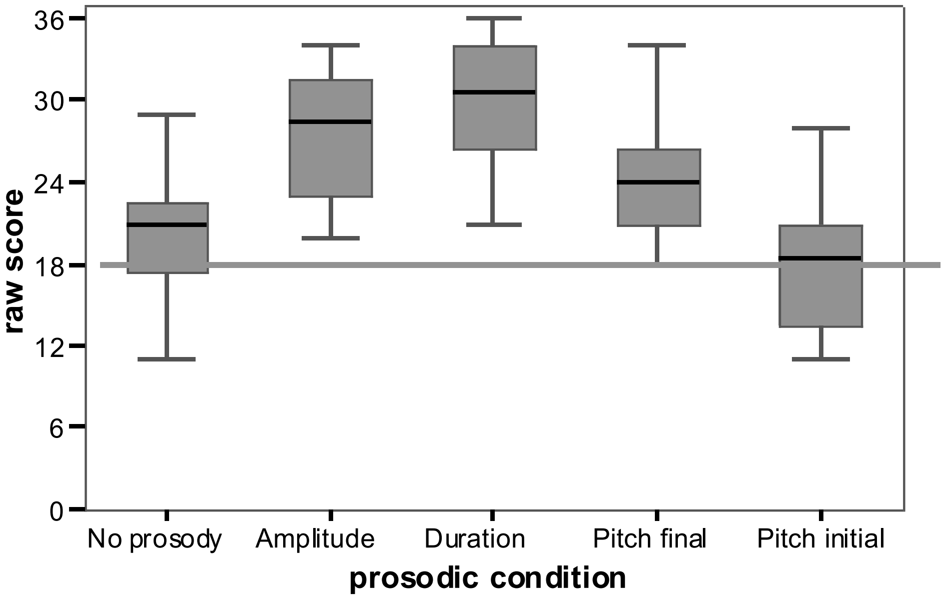

*Figure 26. Correct identification (raw score) in five prosodic conditions.*

*The grey line parallel to the raw score of 18 indicates chance level (50%). The black line within each box indicates the mean raw score for that prosodic condition.*

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.

As mentioned in 5.2.4, each syllable used for the stimuli was not synthesized but produced naturally. Thus, although an effort was made to normalize the different acoustic values, the result was not quite perfect. For instance, there were apparent differences in consonant and vowel lengths between the lenis and the nasal consonant. And a small number of deviations from the mean values always existed in the stimuli. Also, there were two real words of Korean embedded in the words of the artificial language. The purpose of the within-subject analysis was to ensure that there was no artifact in the result that was caused by the method with which the stimuli were created and by the embedded words.

In this RM, the sphericity assumption was met, and thus, the degrees of freedom were not adjusted. The result showed that there was no main effect of *word*. This result ensures that all the words were equally easy or equally hard to the listeners, and that neither small acoustic differences, nor embedded real Korean words affected listeners' performance. 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 repeated-measures ANOVA was performed with *test strings* (part words and non-words) as a within-subject factor and *prosodic condition* as a between-subject factor. This was done in order to investigate whether the listeners showed different performance depending on 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. The *prosodic condition* effect was significant (F (4, 55) = 12.665, *p* < .001), just as reported in the first RM ANOVA, but no interaction was found between *test strings* and *prosodic condition*. This result suggests that all the test strings were equally difficult in general, as well, and that the partial distributional information in the part-word strings did not bias the listeners. Although the result is not significant, an interesting tendency was observed in the data. As illustrated in Figure 27, the three conforming conditions revealed a tendency for the correct response rate to be higher for part-words than for non-words (although the difference between the two was not significant at all) whereas the ***No Prosody*** and the ***Pitch Initial*** conditions showed the opposite tendency.

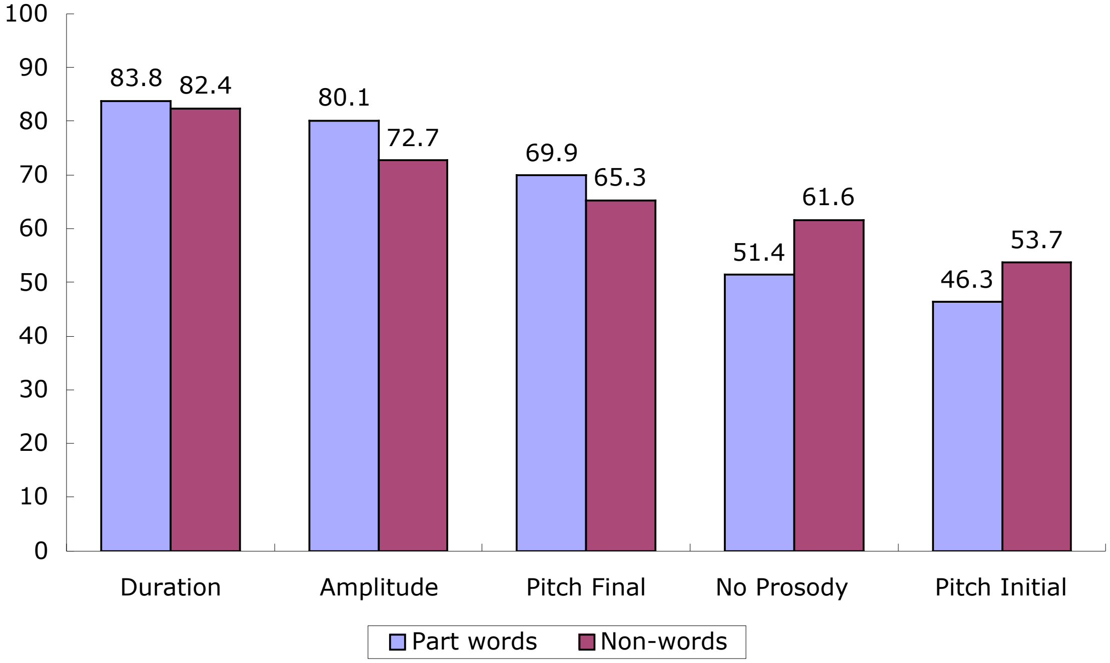

*Figure 27. Correct response rate (%) in Part-word and Non-word test strings.*

In these two repeated measures ANOVA's, 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. Planned pairwise comparisons showed that the ***Duration***, ***Amplitude***, and ***Pitch final*** conditions were significantly different from the ***No Prosody*** condition ($p$ < .001, $p$ = .001, $p$ = .04, respectively). There was no difference between the ***Duration*** and ***Amplitude*** conditions.

The ***Pitch final*** condition was significantly different from the ***Duration*** condition ($p$ = .006), but not from the ***Amplitude*** condition. Finally, the ***Pitch initial*** condition differed significantly from the ***Duration***, ***Amplitude***, and ***Pitch final*** conditions ($p < .001$, $p < .001$, $p = .002$), but not from the ***No Prosody*** condition. The results of post-hoc tests are summarized in Table 9.

| **prosodic conditions** | **statistical difference** | **no difference** |
|---|---|---|
| No prosody | Duration ($p < .001$)<br>Amplitude ($p = .001$)<br>Pitch final ($p = .04$) | Pitch initial |
| Duration | No prosody ($p < .001$)<br>Pitch initial ($p < .001$)<br>Pitch final ($p = .006$) | Amplitude |
| Amplitude | No prosody ($p = .001$)<br>Pitch initial ($p < .001$) | Duration<br>Pitch final |
| Pitch final | No prosody ($p = .04$)<br>Duration ($p = .006$)<br>Pitch initial ($p = .002$) | Amplitude |
| Pitch initial | Duration ($p < .001$)<br>Amplitude ($p < .001$)<br>Pitch final ($p = .002$) | No prosody |

*Table 9. Summary of post-hoc test results*

## 5.4 Discussion

### *The role of the distributional cue*

The average performance of the listeners on the ***No Prosody*** condition (56.6 %) did not significantly differ from chance level, which indicates that the presence of probabilistic information did not help Korean listeners segment words from the speech stream. The experiment did not succeed in replicating the results of the Saffran et al. (1996a) and Bagou et al. (2002) studies that showed that listeners could segment words from speech streams, relying only on the transitional probability cue. English listeners showed 65% average correct performance when the testing phase had real word vs. part word pairs and 76% when the testing phase had real word vs. non-word pairs (Saffran et al., 1996). French listeners also showed about 65%[19] mean performance in the condition where no prosodic information was available (Bagou et al., 2002).

No conclusive explanation can be provided about why this difference exists in the performances, because a direct comparison between Korean, English and French listeners cannot be made, given that the experimental stimuli and procedures were not the same. However, some speculations can be made in order to explain the poor performance of Korean listeners. First, the difference cannot be attributed to the length of exposure. Korean listeners in the current experiment were exposed to the artificial language for 10 minutes, which was about half the time that the English listeners in Saffran et al. (1996)'s study had been exposed to their artificial language (21 minutes), but similar to the time that the French listeners in Bagou et al. (2002)'s study had been exposed to the input (12 minutes). Second, it could be the case that the speech input was just too difficult for Korean listeners. It should be noted that experimental conditions did not have any allophonic cues or coarticulation cues in the input. Not having any coarticulation cue can obviously interfere with speech processing, but cannot be the main reason for the poor performance of Korean listeners, because English and French listeners were also deprived

[19] Bagou et al. (2002) did not provide exact percentage for this condition. The number (65 %) was inferred from a graph provided in their article.

of coarticulation cues. However, not having an allophonic cue could have been critical to Korean listeners, because the artificial language used in the present study had three lenis stops, out of four consonants. Lenis stop voicing always occurs within AP (thus within word) in Korean. However, in this experiment, all the lenis stops were voiceless since every syllable was produced independently. This goes against the phonetic/phonological pattern of Korean, and hence, probably interfered with listeners' ability to mentally form words by relying on the acoustic input. Third, it is plausible that the difference might be due to Korean listeners' less active use of transitional probability information, in comparison to listeners of other languages. Indeed, Korean participants did not show any significant difference in their performance between part-word and non-word test strings in general. Nonetheless, this by no means indicates that Korean listeners were not able to use transitional probability information in word segmentation. In the ***No Prosody*** condition, part- and non-word test strings showed a significant difference in the expected direction. There were more correct answers for non-word test strings than for part-word test strings, just like English listeners who were better on non-word test strings than on part-word test strings. It seems that Korean listeners paid more attention to the distributional information of test strings only when there was no prosodic cue available. Note, that this tendency was also observed in the ***Pitch initial*** condition, where the prosodic cue did not match with the distributional cue, although the difference between the two test string types was not significant. Neither a tendency nor a significant difference between the two test string types was observed in the other prosodic conditions. These results suggest that Korean listeners paid more attention to transitional probabilities of test strings for word segmentation when there was no facilitative prosodic cue available in the input. They paid less attention to the transitional probabilities when proper prosodic cues were available. That is, distributional cues may play a supplementary role in word segmentation in Korean.

### *Facilitating prosodic cues*

The results revealed that the ***Duration*** (AP-final lengthening), ***Amplitude*** (AP-initial strengthening) and ***Pitch final*** (AP-final high tone) conditions facilitated word segmentation more than the ***No prosody*** condition. These three conditions conform to the prosodic characteristics of the Korean AP. Thus, we can conclude that each conforming prosodic cue can efficiently aid speech segmentation of Korean.

Along with Jun (1995b)'s study, the result indicates that Korean speakers produce AP-initial syllable with greater amplitude[20] than the second syllable within the AP, and Korean listeners can use this information for word segmentation. This result suggests that an amplitude cue alone can function as a correct indicator of the beginning of the AP.

This initial amplitude effect cannot be interpreted as an acoustic correlate of word-level stress. H.-B. Lee (1974) and H.-Y. Lee (1990) claimed that Korean has word-initial stress, but only on heavy syllables. They claimed that when there is a light syllable

[20] Note that the stimuli of the previous experiment (see Ch.4) showed that AP-medial word onsets had stronger amplitude than AP-initial word onsets. This difference is due to the number of syllables within an AP. Jun (1995b)'s data showed that AP-initial syllable had greater amplitude when an AP had three syllables (especially when the sentence was produced with a stronger-than-default emphasis level), but the second syllable of an AP had greater amplitude when an AP had four or more syllables (Note that an AP had three syllables in Ch4, and four syllables in Ch5, in the current study).

in word-initial position, the stress falls on the second syllable. In this experiment, the words were composed of light syllables only. Thus, if we follow Lee (1974) and Lee (1990)'s arguments, an amplitude cue at the first syllable of the word should not be helping listeners. Thus, the observed effect is not reflecting any word-level stress of Korean.

Further, this does not seem to be due to domain-initial strengthening effect. As reviewed before, Korean shows more consistent and greater domain-initial strengthening effects than English, French and Taiwanese (Keating et al., 2004). Strong domain-initial effects in Korean suggest that Korean reinforces the beginning of the phrase. It is possible that 'intensity' signals 'reinforcement' at the phrase-initial position. However, the articulatory and acoustic cues that are known to be associated with domain-initial strengthening in Korean include a larger degree of linguo-palatal contact, longer articulatory seal duration, and longer acoustic duration of initial consonants, but not initial consonant amplitude increase (Cho & Keating, 2001). The intensity values of consonants, as measured by stop burst RMS and nasal minimum RMS, were actually lower in the consonants in higher prosodic domains than those in lower domains.

Instead, the initial amplitude effect found in the current study is due to the differences in amplitude among the vowels. The increase in initial amplitude found in Jun's study (1995b) was based on the average amplitude of the vowels. However, Cho and Keating's study (2001) did not measure the vowel amplitude effect in prosodic-phrase-initial position. Thus, we can conclude that Korean listeners are sensitive to the increase in vowel amplitude in AP-initial position, and that they make use of this boundary information in word segmentation.

The statistical difference observed between the ***Duration*** and ***Pitch final*** conditions suggests that the degree of contribution of each facilitating cue to word segmentation differed. The durational cue contributed significantly more to segmentation than the F0 rise in the same syllable location, i.e., in word-final position. The difference between the two prosodic conditions can be interpreted in a couple of different ways. One possible interpretation for the difference is a prosodic structure account. Recall that various production studies presented different degrees of lengthening with regard to AP-final lengthening (50%, 20%, 12%, 0%). The degree of lengthening (30.6 %) employed for the ***Duration*** condition turned out to be within this range. However, also recall that this rate was similar to the minimum degree in IP-final lengthening reported in Chung et al. (1996). On the other hand, the degree of increase of the ***Pitch final*** condition only corresponded to the degree of AP-final pitch rising reported in Jun (1995b). Thus, the observed difference could be 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., the Intonational Phrase, whereas the heightened F0 value only marked an AP. It is well-known that speakers' speech production is governed by prosodic structure, and that 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 et al., 2004; Wightman, Shattuck-Hufnagel, Ostendorf, & Price, 1992). The prosodic structural information realized in the speech input is exploited in speech perception, as well

(McQueen & Cho, 2003; Schafer & Jun, 2001). This interpretation implies that listeners have knowledge of how an utterance is parsed into prosodic phrases, what kind of acoustic cues are associated with each prosodic unit, 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 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 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 a 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. The prosodic structural interpretation (i.e., listeners are sensitive to prosodically-driven acoustic cues, paradigmatically as well as syntagmatically, and hence, they use this information for effective speech segmentation and word storage) can be tested, if an experiment can compare AP-level final lengthening with IP-level final lengthening. However, before any further perception experiment is conducted, it will be necessary to obtain a sufficient amount of quantitative data so as to accurately establish the range of acoustic values of each prosodic cue in different prosodic positions.

In any case, the results show that a prosodic cue which marks a phrase boundary, regardless of the level of the phrase in the prosodic hierarchy, can facilitate word segmentation by helping listeners to predict where words begin and/or end.

***Non-conforming prosodic cue***

As discussed so far, all the cues that conform to the characteristics of Korean prosody contributed to segmentation. This brings up the question of whether it is the existence of a conforming prosodic cue, or the existence of any perceptually salient prosodic cue which is prominent enough to capture listeners' auditory attention, that facilitates word segmentation. The ***Pitch initial*** condition provides a direct answer to this question. This condition, with a pitch cue on the first syllable, showed 49.9% mean correct rate, with half of the participants scoring less than chance level, which indicates that this cue affected listeners negatively. This result was worse than that of the ***No prosody*** condition, where there was no prosodic cue present, though the two conditions were not significantly different.

Recall that even if this can be a possible prosodic cue in some languages, it is not a part of the native prosody of Korean. In addition, this condition violated the tone-segment mapping rules of Korean by mapping a lenis/nasal consonant to a high tone in AP-initial position. Thus, it is very likely that listeners would have tried to match the high F0 to the last syllable of a word, rather than the first syllable of a word. This attempt would have caused a mismatch of the two different sources of information (i.e., the distributional information of segments in the stimuli and the prosodic information), which consequently brought about lower performance scores in the ***Pitch initial*** condition. We can further conjecture that, if the test strings in the testing phase had contained the non-words reflecting mismatched parsing and the real words of the artificial language, listeners would have chosen the mismatched strings over the real words. If this result had been observed, it could have provided more support for the mismatch interpretation, and further, language-specific use of boundary cues.

Compared to the ***Pitch Final*** condition, this result indicates that a prosodic cue can have different functions 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 the language.

All in all, listeners' poor performance on the ***Pitch initial*** condition 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.

### *Language-specific speech segmentation*

Previous studies of word segmentation provided quite a bit of evidence that certain cues for segmentation, such as lexical-prosodic structure (Cutler and Otake, 1992), and phonotactic constraints (McQueen, 1998; Weber, 2001), are language-specific. The results of this experiment seemed to suggest that the exploitation of post-lexical boundary cues is not an exception to the language-specific nature of segmentation cues.

Both the facilitative role of the pitch-final cue and the slight inhibitory role of the word-initial pitch cue are the by-product of the prosodic characteristics of Korean, where only an AP-final syllable, but not an AP-initial syllable, is high-pitched. These results suggest that the prosodic cues that facilitate word segmentation of a language are not any random acoustic cues, but the ones that are associated with language-specific prosodic patterns and structure.

Relying more heavily on the AP-final duration cue than the AP-final pitch cue for segmentation might be another language-specific strategy of Korean listeners, since the reverse pattern was observed for French listeners (Bagou et al., 2002). Bagou et al. (2002)'s study had a 30 % increase in both duration and F0 rise. Although they did not provide the basis for their increase rate, the rate of lengthening is similar to what has been observed AP-finally in a French production study (Jun and Fougeron, 2000). This increase rate for duration is similar to the current study, and the results of the two studies are also similar (82 % in Bagou et al.'s study, 83 % in the current study). The F0 increase rate in their study, however, is approximately twice as high as in the current study. If the F0 increase rates used in French (30%) and Korean (12%) are correctly reflecting the F0 increase rate in AP-final position in each language, the initial speculation (i.e., Korean listeners rely more on the AP-final duration cue than the AP-final pitch cue, unlike French listeners who showed the reverse pattern) could be supported. However, a direct comparison of Korean and French is not very effective at this point, because these two experiments did not use the same experimental procedures (the task, stimuli, exposure time, test strings etc.). Further research is needed to interpret the differences and to shed more light on whether the relative contribution of individual cues may also be language specific.

### *Summary*

To recapitulate, the results of the current experiment indicate that AP boundary cues 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 boundary cues that are related to the AP, that is, 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, helped segmentation, while the cue that is not an AP characteristic did not. Furthermore, the result of the current experiment, and cross-linguistic comparisons suggest that the contribution of prosodic cues to speech segmentation depends mainly on which cue is occurring where: This information hinges on the prosodic characteristics of a given language. Thus, post-lexical prosodic cues to segmentation, like other segmentation cues, are language-specific.

# CHAPTER 6

# CONCLUSION

## 6.1 Summary

This dissertation investigated language-specific prosodic cues that facilitate word segmentation in Korean. Since Korean does not have lexical prominence, I hypothesized that the Accentual Phrase, the smallest prosodic phrase that has phrasal edge prominence, would play a pivotal role in Korean word segmentation. This hypothesis was based on observations from previous studies. Although the AP can contain more than one content word in principle, the present study as well as previous ones have shown that the Korean AP is typically very similar to a content word in size. Moreover, the Korean AP contains various prosodic characteristics that listeners can exploit. Thus, it was predicted that the presence of prosodic properties that mark AP edges would directly influence word segmentation.

### *A Corpus Study*

The first part of the dissertation was a corpus study. The purpose of this study was to obtain distributional properties of the AP, including the following: the number of words and syllables it contains on average, the frequency of various AP tonal patterns, the location of content words within the AP, and the types of tonal patterns that are imposed upon content words.

This study was based on two corpora: the Read speech corpus that was obtained from production data, and the Radio corpus that was recorded from two radio programs. A total of 414 sentences from the two corpora were used for the data analysis. The results confirmed the finding from previous studies that the AP was similar to a word in size. 84% of the APs contained one content word. Further, the results showed that almost 90% of the content words are located in AP-initial position. When the AP-initial onset was not an aspirated or tense consonant, the most common AP patterns were LH, LHH, and LHLH (78%). When the AP-initial onset was an aspirated or tense consonant, the most common AP patterns were HH, HHLH, and HHL (72%). The results also revealed the following: (i) when the word-initial syllable onset is not an aspirated or tense consonant, 88% of the multisyllabic content words start with a rising tone in AP-initial position; (ii) when the word-initial syllable onset is an aspirated or tense consonant, 74% of the multisyllabic content words start with a level H tone in AP-initial position; (iii) 84.1% of the APs end with the final H tone.

### *Tonal Patterns of the AP and Word Segmentation*

The corpus study showed that there are at least two prevailing tonal patterns for APs, a word-initial rising tone (only when the word-onset is not an aspirated/tense consonant) and an AP-final high tone. A perception experiment was performed in order to examine whether the non-contrastive tonal patterns of the AP influenced word segmentation in Korean. This study particularly focused on whether or not listeners are sensitive to the regularities in frequency of post-lexical tonal patterns, as observed in the corpus study.

A word-spotting task was employed for the experiment. During the experiment, participants listened to a series of speech streams that were composed of strings of nonsense syllables containing a real word. Listeners were asked to detect the embedded real word in the given speech stream. There were 16 disyllabic and 16 trisyllabic target words, with an equal number of filler words. In this experiment, a nonsense speech stream was comprised of three APs, each of which contained four syllables. The second AP of a target-bearing stream contained the target word. Filler words were inserted either in the first or the third AP of the speech stream. In order to determine whether the AP-final tone from the preceding AP affected word segmentation, the experimental conditions were divided into two sets, the AP-initial condition and the AP-medial condition. In the AP-initial condition, the onset of the target word corresponded to the onset of the AP. In the AP-medial condition, the onset of the target word occurred at the beginning of the second syllable of the AP. Each positional condition was presented with three tonal patterns: the AP-initial condition was presented with the ***LH***LH, ***LH***HH, and ***LL***LH patterns, and the AP-medial condition was presented with the L***HL***H, L***LL***H and L***LR***H patterns. Based on the different tonal characteristics, the following hypotheses were made. First, if the final high tone of a preceding AP helps segmentation, listeners' performance will be better in the three initial tonal patterns than in the three medial tonal patterns. Second, if rising tone facilitates word segmentation, listeners' performance will be better for the tonal patterns that include a rising tone (***LH***LH, ***LH***HH in AP-initial position, and L***LR***H in AP-medial position) than for those without a rising tone, regardless of the word's position within the AP. Third, if listeners are more sensitive to the frequencies of overall AP tonal patterns than to local pitch rising, their performance will be better for frequent tonal patterns (L***HL***H, L***LL***H) than it is for the unattested tonal pattern (L***LR***H) in the AP-medial position. Fourth, if a falling tone after a word helps segmentation, listeners' performance will be better with the rising tone followed by a low tone (***LH***LH) than the rising tone followed by a high tone in the AP initial position (***LH***HH).

The results revealed that 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 reliably be exploited as a cue for segmentation. It is likely that this tone allowed the listeners to correctly anticipate the upcoming word before its onset had actually been perceived. Results further showed 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 did not have many neighbors, hence the listeners were able to detect these words more easily regardless of the tonal patterns imposed upon them. In the analysis of disyllabic word spotting, the results indicated 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. These findings suggest that a rising tone superposed upon words can serve as a word segmentation cue in Korean, a language that lacks lexical prominence. Finally, a falling tone over target words turned out to be a negative tonal cue. Thus, the results clearly suggest that even non-lexical prosodic information can be stored and can affect word segmentation when the information frequently co-occurs with words. In addition, the results show that a rising tone at word onset and an AP-final high tone have great influence on Korean word segmentation.

### *Boundary cues of the AP and Word Segmentation*

One of the most important results found in the first perception experiment was that the detection of words is faster when they were in AP-initial position than when they were in AP-medial position. As previously mentioned, a final high tone from the AP that precedes the target words must have a facilitative effect. However, it is also true that word segmentation can be benefited by other prosodic information at the phrase boundaries, when words are located in AP-initial position. The purpose of the second experiment was to investigate the role of different prosodic boundary cues in word segmentation. By employing the artificial language learning method, this experiment examined whether all the boundary cues exert the same influence, and whether the exploitation of boundary cues for word segmentation is language-specific.

The artificial language was composed of six tri-syllabic words, which were concatenated to make an 11-minute speech stream. There was no pause between the words, and the same word never occurred twice in a row. Since Korean APs typically show initial amplitude effects (Jun, 1995b), final pitch rising (Jun 1993), and final lengthening (Cho & Keating, 2001; Oh 1998), five prosodic conditions were devised for the experiment. In the ***No Prosody*** condition, all of the syllables in the tri-syllabic artificial word had the same duration, pitch, and amplitude. In the ***Duration*** condition, the third syllable of each word was lengthened, imitating AP-final lengthening. In the ***Pitch final*** condition, the f0 of the third syllable was higher than in the other syllables, imitating AP-final pitch rising. In the ***Amplitude*** condition, the first syllable of each word had increased amplitude, imitating the AP-initial amplitude effect. Finally, the ***Pitch initial*** condition had a high pitch on the first syllable of every word. This acoustic pattern did not conform to any of the prosodic characteristics of Korean APs. During the learning phase, listeners heard a speech stream from one of the five conditions. After the learning phase, there was a testing phase in which participants were asked to choose the word from a Word/Non-Word pair.

Results indicated that the ***Duration***, ***Amplitude,*** and ***Pitch final*** cues facilitated word segmentation in Korean, when compared to the ***No Prosody*** condition. This suggests that the detection of prosodic cues at AP boundaries can facilitate word segmentation beyond effect of transitional probabilities. To the contrary, correct response rate for the ***Pitch initial*** condition, which does not conform to Korean AP characteristics, did not differ significantly from that of the ***No Prosody*** condition. The data revealed a tendency for the pitch pattern conforming to the Korean AP (the ***Pitch final*** condition) to facilitate segmentation, while the non-conforming pitch pattern (the ***Pitch initial*** condition) interferes with segmentation. This result suggests that listeners are sensitive to the prosodic features of their native language and use this knowledge during the process of word segmentation.

## 6.2 General Discussion

The current study attempted to examine whether a post-lexical prosodic unit of Korean, the AP, can directly aid word segmentation, what kinds of prosodic cues of the AP help listeners to detect word boundaries, and whether the frequency of post-lexical intonational patterns affects word segmentation.

The results of the current studies strongly support the initial hypothesis that the AP can directly help word segmentation in Korean. The corpus study showed that about 90% of multisyllabic content words were located in AP-initial position in Korean, and that there were 1.15 content words in the AP on average. This means that, in Korean, most content words are located in a position where the detection of words can be benefited by various prosodic cues. If a word happens to occur at the beginning of an IP-initial AP, it is likely that there is a pause preceding the word. And as shown by many studies, a pause is an important prosodic cue for speech segmentation in general (Hirsh-Pasek et al., 1987; Gerken et al, 1994; among others). If a word happens to occur at an AP onset but IP medial, the word is preceded by another AP. At the onset of an IP-medial AP, there is no preceding pause, but the detection of the word can be facilitated by both AP-final prosodic cues of the preceding AP (e.g., final lengthening (Cho & Keating, 2001; Oh, 1998; Chung et al, 1996), AP-final high pitch (Jun, 1993)), and AP-initial cues of the word-bearing AP (e.g., amplitude increase (Jun, 1995b), domain initial strengthening of initial segments (Cho & Keating, 2001)). The artificial language learning experiment showed that each of the AP boundary cues, on its own, was powerful enough to aid word segmentation. The duration and high pitch cues in AP-final position and the amplitude cue in AP-initial position significantly improved listeners' ability to find word boundaries and learn the new words of the artificial language. The results further showed that the integration of these cues also facilitates word segmentation. Listeners detected AP-initial target words much faster than AP-medial target words, and their error rates were four times higher in AP-medial than in AP-initial position. This positional preference owing to the combination of prosodic boundary cues was more influential on word segmentation than the rising tone over content words, as demonstrated by the result that listeners' detection of words was much faster when target words were in AP-initial position with a non-rising tone than when target words were in AP-medial position with a rising tone. These results seem to support the claim that integration of multiple cues exerts more power in segmentation than individual cues (Christiansen, Allen, & Seidenberg, 1998; Norris et al., 1997).

As discussed at the beginning of this dissertation, previous studies on word segmentation suggest that the exploitation of segmentation cues relies heavily on distributional information. For instance, listeners rely on the prevailing lexical stress cue (i.e., a trochaic stress pattern in English and Dutch; Cutler & Norris, 1988; Vroomen and de Gelder, 1998), and on the frequency of segmental sequences (i.e., phonotactics; McQueen, 1998; Weber, 2001) when segmenting words from speech streams. The findings of this dissertation lead us to conclude that the exploitation of post-lexical intonation cues in word segmentation is also determined by the frequency of intonational patterns of the AP. The result of the corpus study showed that about 85% of the APs started with a rising tone (i.e., an initial L tone on the first syllable followed by a pitch rise on the second syllable). Moreover, 88% of the multisyllabic content words started with the rising tone. While the rising tone pattern was frequent on the first two syllables of APs and multisyllabic content words, a high tone was the most predominant pattern (89%) on AP-final syllables. These distributional properties of post-lexical tonal patterns in speech input had a great influence in Korean word segmentation. The word-spotting experiment showed that listeners were much faster and more accurate in detecting target words when they occurred with the frequent intonational pattern (i.e., rising) than when they occurred with an infrequent intonational pattern (i.e., non-rising) in a speech stream. The artificial language learning experiment revealed that the presence of an AP-final H tone in speech facilitated word segmentation. The tonal pattern frequency in the speech input also led to a language-specific segmentation strategy such that the extensive exploitation of a frequent tonal cue in segmentation sometimes misguided listeners. In the corpus data, only 5% of APs started with a high tone, when the AP-initial consonant was an L-inducer. This infrequent co-occurrence of a H tone on L-tone-inducing segments in AP-initial position, and the frequent occurrence of a H tone in AP-final position, resulted in listeners' poor performance in the ***Pitch Initial*** condition of the artificial language learning experiment. Korean listeners seemed to have a strong tendency to align a high pitch with a phrase- (and word-) final syllable, and not with a phrase- (and word-) initial syllable, when the first segment of a phrase (and word) is an L-tone-inducer. This tendency is, of course, due to their language-specific segmentation strategy established on the probabilistic information of the occurrence of a specific tonal pattern in a specific position within the AP.

Overall, the current findings seem to suggest that a word-size post-lexical prosodic phrase can directly influence word segmentation in Korean with the ample prosodic cues that it contains, just as lexical stress pattern directly influences word segmentation in languages with lexical-level prominence. The question that should be addressed further is how various segmentation cues (e.g., prosodic, phonotactic, allophonic, distributional cues, etc.) interact in Korean. Although this dissertation did not directly investigate the role of other segmentation cues in Korean besides prosodic cues, or the interaction among segmentation cues, the results still allow us to speculate about the relative influence of individual cues in segmentation. The results of the artificial language learning experiment showed that Korean listeners did not use distributional cues as efficiently as English and French listeners in other studies (Saffran et al, 1996; Bagou et al, 2002, respectively). This result was ascribed to the absence of an allophonic cue (i.e., lenis stop voicing) in the speech streams, which was not problematic in the two other languages. The result indicates that this specific allophonic cue might be a much stronger segmentation cue than the transitional probability cue in Korean. This seems quite plausible, given that the transitional probability cue was reported to be weaker than another phonetic cue, that is, coarticulation cue in English (Johnson & Jusczyk, 2001). The current results clearly showed that the presence of prosodic cues helped Korean listeners to successfully segment words from a speech stream, without recourse to the transitional probability cue, and in the absence of the lenis stop voicing cue. This indicates that post-lexical prosodic cues are stronger than the transitional probability cue, just as lexical prosodic cues are  (Johnson & Jusczyk, 2001). The results also indicate that the major role of prosodic cues was to help listeners to perceive the sequence of syllables as a coherent unit, even if it could not be perceived as a phonetic word due to the lack of allophonic cues. Further, prosodic cues aided the storage and segmentation of the perceived unit from a continuous speech stream. This shows that the phrase-final pitch cue (and probably other prosodic boundary cues) can signal the cohesion of the components of a perceived unit even without the help of the lenis stop voicing cue. Therefore, with the findings of Choi and Mazuka (2001), we would be able to conclude that post-lexical prosodic cues are stronger than at least a specific allophonic cue (i.e., lenis stop voicing) in word segmentation and in speech processing of Korean. At the same time, however, the results also imply the possibility that the segmental sequences of a word, which provide information about the uniqueness point and/or neighborhood density, can be more powerful in word recognition and segmentation than the intonational cue, especially when the segmental sequence cue is available earlier than the suprasegmental cue in speech signal. For instance, recall that, in the word-spotting experiment, the detection of trisyllabic words was much faster than that of disyllabic words, and that there was no tonal pattern effect in the detection of trisyllabic words. Most trisyllabic target words used in the experiment had their uniqueness point at the syllable onset of the second syllable. The rising tonal cue was not available yet at that point in time. However, once the presence of segmental cues suffices to modulate lexical access, listeners were able to segment faster without depending on the intonational cue. This supports the claim that "listeners exploit incoming information in speech signals in a continuous manner, making use of cues as soon as they become available in order to modulate the activation of potential candidate words that have received some support from the speech signal (Donselaar, Koster, & Cutler, in press)".

These results further suggest that listeners make use of all the information in speech signals, and tend to rely on more powerful and definite cues that are available in the input, at each time point of speech processing, as speech unfolds in time. When they encounter a cue or multiple cues at a certain time point, they make the most of the information to process the speech signal. The findings of this dissertation suggest that prosodic phrasing can directly facilitate word segmentation in Korean, by supplying ample cues in speech signals.

## 6.3 Limitations of the Current Study

One major limitation in the corpus study is the nature of the corpora. Although an effort was made to include various styles of speech, the corpus study did not include enough data from completely spontaneous speech. We would be able to get a clearer idea of the cues present in real speech, and of the way people could segment speech, by investigating conversational data.

Reaction time data has been a useful measure in psycholinguistic research because it can reflect how people process speech data in on-line tasks. However, in the word-spotting task, the reaction time data did not show any significant result. One reason that reliable reaction time data was not obtainable in this study lies in the fact that the study adopted the voice key activation method for RT measurements. A recent study by Kessler, Treiman, and Mullennix (2002) pointed out that voice RT measurements are biased by the initial phonemes of the response. They further indicated that these phoneme-based biases were large enough to cause problems in interpreting the RT measurements. I presume that such phoneme-based biases might have caused no RT effect. Before a better technique is designed, one should be cautious about using voice activated RT measurements.

Several issues were not clarified with regard to the second perception experiment of this dissertation. First, it is not certain what caused the AP-initial amplitude effect. As mentioned earlier, it cannot be explained by a word-level stress effect or by the domain-initial strengthening effect of AP-initial consonants. There is a possibility that the AP-initial amplitude effect might be due to the vowel amplitude of the AP-initial syllable. However, in order to confirm this conjecture, it is necessary to examine speech production data. Second, there was no concrete basis for choices in the acoustic cue manipulation, and hence, the results obtained from different prosodic conditions could not be compared in a fair manner. This problem was mainly caused by the fact that the existing production data were contradictory. Thus, the necessity for more broad and ample production data analysis is emphasized in order to get consistent data. Third, although I assumed that the negative effect found in the ***Pitch initial*** condition was due to listeners' tendency to match the high pitch to the word-final (and AP-final) syllable, and though this assumption does seem to be acceptable, it cannot be fully supported with the data obtained from the testing phase. If the testing phase had contained some items that could have shown listeners' misinterpretation tendency (if any), the claim could have been completely supported. Addition of such items in the testing phase is necessary for future studies, in order to exactly determine how participants are parsing the speech stream.

## 6.4 Implications and Future Directions

The results of the word-spotting experiment suggest that a rising pitch is a reliable cue for word onset for Seoul Korean listeners (regardless of the location of words within a prosodic phrase), when the word-initial syllable onset was not an aspirated or tense consonant. We know, from the corpus study, that when a word initial onset is an aspirated/tense consonant, the majority of content words occur with a level high tone. This suggests that, in the case of Korean, word-initial level H tone should also help word segmentation for the words with an aspirated/tense initial consonant. If an experiment can show this result, it will further confirm the idea that even non-lexical prosodic information can be stored and can affect speech processing when frequent enough.

This study mainly focused on the role of post-lexical prosodic cues in Korean word segmentation. Of course, in order to fully understand how Korean listeners use different segmentation cues, it is necessary to investigate the role of other segmentation cues and the weight of each cue.

One of the most interesting avenues for future research is to investigate when Korean infants acquire knowledge of post-lexical intonational patterns and prosodic boundary cues and how they use them for word segmentation. Studying infants' ability to detect a prosodic phrasing cue will further illuminate the issues of syntactic bootstrapping and speech processing in general, as well as word segmentation; Valian and Levitte (1996) showed that prosodic phrasing can help in learning the syntax of an artificial language by emphasizing structural relationships between words. In addition, perception of prosodic boundaries is known to help disambiguate ambiguous sentences in on-line sentences processing (Schafer, 1997; Schafer & Jun, 2001). Specifically in Korean, prosodic phrasing can convey semantic differences (Jun, 2004a; Jun & Oh, 1996). Thus, the acquisition of prosodic phrasing knowledge will build the foundation for further syntactic and semantic acquisition. We know from Choi and Mazuka (to appear)'s study that young Korean children can use the AP-final high pitch cue and lenis stop voicing cue in parsing sentences with ambiguous phrasing. An infant study will clearly help us see the time of initial acquisition and applications of such information. This future study, when compared with infant acquisition studies of other languages, will also shed light on when and how infants acquire the language-specific prosodic prominence parameter.

# Appendix 1. Word selection criteria for target and filler words

| | | Target words (16) | Filler words (16) |
|---|---|---|---|
| disyllabic words (32) | Kim & Kang (2000) Frequency | - High<br>- Top 9.2% of the noun freq list (Freq. Range: 7396 – 34)<br>- All of them were included in 'most frequent 3000 words' across grammatical categories | - Not strictly controlled<br>- 13 out of 16 were in top 9.2 % |
| | KCP Frequency | - Ranking 5000 and low | - Not controlled |
| | Familiarity (1-5 scale) | - 3.0 or more<br>- Average 3.9 | - Not controlled<br>- Average 3.8 |
| | Syll. Structure | CV.CV | CV.CV |
| trisyllabic words (32) | Kim & Kang (2000) Frequency | - All 16 targets were in top 9.2 % of the list | - Not controlled<br>- 9 out of 16 were in top 9.2%. |
| | KAIST Freq | - Ranking 5000 and up | - Not controlled |
| | Familiarity (1-5 scale) | - 3.0 or more<br>- Average 3.4 | - Not controlled<br>- Average 3.5 |
| | Syll. Structure | CV.CV.CV | CV.CV.CV<br>CV.CVC.CV<br>V.CV.CV. |

# Appendix 2. Frequency and Familiarity of Disyllabic Target words

| Disyllabic Target Words | phonetic transcription | gloss | Freq.count (Kang & Kim 2000) | Freq.count (Kaist) | Freq.rank (Kaist) | Familiarity |
|---|---|---|---|---|---|---|
| | [na.bi] | butterfly | 47 | 496 | 2638 | 3.3 |
| | [ma.ru] | floor | 61 | 413 | 3044 | 3.7 |
| | [mo.tʃa] | hat | 73 | 572 | 2377 | 4.5 |
| | [mʌ.ɾi] | head | 635 | 5809 | 320 | 4.5 |
| | [na.ra] | nation | 1885 | 5918 | 317 | 4.1 |
| | [no.ɾɛ] | song | 477 | 1017 | 1478 | 4.7 |
| | [na.mu] | tree | 571 | 2305 | 698 | 4.5 |
| | [tʃa.ju] | freedom | 695 | 24 | 23021 | 4.1 |
| | [ta.ɾi] | leg/bridge | 387 | 2425 | 667 | 4.5 |
| | [ko.gi] | meat | 176 | 770 | 1846 | 4.7 |
| | [pa.tʃi] | 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 |
| | [kʌ.ɾi] | street | 517 | 3031 | 562 | 4.4 |
| | [ju.ɾi] | glass | 92 | 317 | 3771 | 3.9 |

# Appendix 3. Frequency and Familiarity of Trisyllabic Target words

| | phonetic transcription | gloss | Freq.count (Kang & Kim 2000) | Freq.count (Kaist) | Freq.rank (Kaist) | Familiarity |
|---|---|---|---|---|---|---|
| Trisyllabic Target Words | [mu.dʌ.gi] | pile | 29 | 91 | 9780 | 3 |
| | [na.nu.gi] | division | 1 | n/a | n/a | 3.4 |
| | [na.dɨ.ɾi] | outing | 26 | 48 | 15105 | 3 |
| | [nu.dʌ.gi] | rag | 11 | 98 | 9297 | 2.5 |
| | [to.t$^h$o.ɾi] | acorn | 15 | 80 | 10664 | 3 |
| | [to.kɛ.bi] | elf | 26 | 123 | 7836 | 3.6 |
| | [mɛ.t*u.gi] | grasshopper | 18 | 33 | 19033 | 3.4 |
| | [tʃɛ. tʃ$^h$ɛ.ki] | sneeze | 13 | 28 | 20924 | 4.1 |
| | [tu.k*ʌ.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.ɾa.tʃi] | bellflower | 10 | 54 | 13955 | 3.2 |
| | [tu.dʌ.tʃi] | mole | 12 | 43 | 16198 | 3 |
| | [to.ga.ni] | pot | 8 | 29 | 20678 | 3.1 |
| | [ki. ɾʌ.ki] | wildgeese | 13 | 52 | 14322 | 3 |
| | [po.t*a.ɾi] | bundle | 14 | 196 | 5519 | 3.4 |

# REFERENCES

Auer, E. T. (1993). *Dynamic processing in spoken word recognition: The influence of paradigmatic and syntagmatic states.* Unpublished Doctoral dissertation, State University of New York at Buffalo.

Ayers, G. (1994). Discourse functions of pitch range in spontaneous and read speech. *Ohio State University Working Papers in Linguistics, 44*, 1-49.

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.* Paper presented at the Speech Prosody, Aix-en-Province, France.

Beckman, M. E. (1986). *Stress and Non-Stress Accent.* Dordrecht: Foris.

Beckman, M. E., & Edwards, J. (1990). Lengthenings and Shrtenings and the nature of prosodic constituency. In J. Kingston & M. E. Beckman (Eds.), *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech* (pp. 152-178). Cambridge: Cambridge University Press.

Brent, M. R. (1997). Toward a unified model of lexical acquisition and lexical access. *Journal of Psycholinguistic Research, 26*(3), 363-375.

Brent, M. R. (1999). Speech segmentation and word discovery: a computational perspective. *Trends in Cognitive Science, 3*(8), 294-301.

Brown, R. A. (2001). *Effects of lexical confusability on the production of coarticulation.* Unpublished M.A. thesis, University of California, Los Angeles.

Cho, T. (2001). *Effects of Prosody on Articulation in English.* Unpublished Ph.D dissertation, University of California, Los Angeles.

Cho, T., & Keating, P. A. (2001). Articulatory and Acoustic Studies on Domain-Initial Strengthening in Korean.

Choi, Y., & Mazuka, R. (to appear). Acquisition of Prosody in Korean. In C. Lee & Y. Kim & G. Simpson (Eds.), *Handbook of East Asian Psycholinguistics, Part III: Korean Psycholinguistics*. London: Cambridge University Press.

Christiansen, M. H., Allen, J., & Seidenberg, M. S. (1998). Learning to segment speech using multiple cues: A connectionist model. *Language and Cognitive Processes, 13*, 221-268.

Christophe, A., Dupoux, E., Bertoncini, J., & Mehler, J. (1994). Do infants perceive word boundaries? An empirical study of the bootstrapping of lexical acquisition. *Journal of Acoustical Society of America, 95*, 1570-1580.

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

Christophe, A., Mehler, J., & Sebastian-Galles, N. (2001). Perception of prosodic boundary correlates by newborn infants. *Infancy, 2*, 385-394.

Christophe, A., Peperkamp, S., Pallier, C., Block, E., & Mehler, J. (in revision). Phonological phrase boundaries constrain lexical access: I. Adult data. *Journal of Memory & Language*.

Chung, K., Chang, S., Choi, J., Nam, S., Lee, M., Chung, S., Koo, H., Kim, K., Kim, J., Lee, C., Han, S., Oh, M., Song, M., Hong, S., & Jee, S. (1996). *A study of Korean prosody and discourse for the development of speech synthesis/recognition system*. Daejon, Korea: KAIST Artificial Intelligence Research Center.

Church, B., & Schacter, D. (1994). Perceptual specificity of auditory priming: Memory for voice intonation and fundamental frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 521-533.

Cummings, A., & Fernald, A. (2003). *Is it easier for infants to learn a new word first presented in isolation or in a multiword utterance?* Paper presented at the Biennial meeting of the Society for Research in Child Development, Tampa, FL.

Cutler, A., & Butterfield, S. (1992). Rhythmic cues to speech segmentation: Evidence from juncture misperception. *Journal of Memory & Language, 31*, 218-236.

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

Cutler, A., Demuth, K., & McQueen, J. M. (2002). Universality versus language-specificity in listening to running speech. *Psychological Science, 13*(3), 258-262.

Cutler, A., Mehler, J., Norris, D., & Segui, J. (1992). The monolingual nature of speech segmentation by bilinguals. *Cognitive Psychology, 24*, 381-410.

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 phoneme? Further evidence for language-specific listening. *Journal of Memory & Language, 33*, 824-844.

Dahan, D., & Brent, M. R. (1999). On the discovery of novel wordlike units from utterances: an aritificial-language study with impliations for native-language acquisition. *Journal of Experimental Psychology, 128*(2), 165-185.

Davis, M. H., Marslen-Wilson, W. D., & Gaskell, M. G. (2002). Leading up the lexical garden path: segmentation and ambiguity in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 28*, 218-244.

Dilley, L., Shattuck-Hufnagel, S., & Ostendorf, M. (1996). Glottalization of word-initial vowels as a function of prosodic structure. *Journal of Phonetics, 24*, 423-444.

Donselaar, W. v., Koster, M., & Cutler, A. (in press). Exploring the role of lexical stress in lexical recognition. *Quarterly Journal of Experimental Psychology*.

Echols, C. H., Crowhurst, M. J., & Childers, J. B. (1997). The perception of rhtymic units in speech by infants and adults. *Journal of Memory & Language, 36*, 202-225.

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

Fougeron, C. (1999). Articulatory properties of initial segments in several prosodic constituents in French. *UCLA Working Papers in Phonetics, 97*.

Fougeron, C., & Keating, P. A. (1997). Articulatory Strengthening at Edges of Prosodic Domains.

Gerken, L. A., Jusczyk, P. W., & Mandel, D. R. (1994). When prosody fails to cue syntactic structure: Nine-month-olds' sensitivity to phonological versus syntactic phrases. *Cognition, 51*, 237-265.

Goldinger, S. D. (1997). Perception and production in an episodic lexicon. In K. Johnson & J. W. Mullennix (Eds.), *Talker variability in speech processing* (pp. 33-66). San Diego, CA: Academic Press.

Goldinger, S. D. (1998). Echoes of echose? An episodic theory of lexical access. *Psychological Review, 105*, 251-279.

Gout, A., Christophe, A., & Morgan, J. L. (in revision). Phonologicla phrase boundaries constrain lexical access: II. Infant data. *Journal of Memory & Language*.

Harrington, J., Watson, G., & Cooper, M. (1989). Word boundary detection in broad class and phoneme strings. *Computer Speech and Language, 3*, 367-382.

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.

Hayes, J. R., & Clark, H. H. (1970). Experiments on the segmentation of an artificial speech analogue. In J. R. Hayes (Ed.), *Cognition and the Development of Language* (pp. 221-234). New York: Wiley.

Hirsh-Pasek, K., Kemler-Nelson, D., Jusczyk, P. W., Wright Cassidy, K., Druss, B., & Kennedy, L. (1987). Clauses are perceptual units for young infants. *Cognition*(269-286).

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

Hsu, C.-S., & Jun, S.-A. (1998). Prosodic strengthening in Taiwanese: syntagmatic or paradigmatic? *UCLA Working Papers in Phonetics, 96*, 69-89.

Hyman, L. (1978). Word demarcation. In J. Greenberg (Ed.), *Universals of Human Language: Phonology* (pp. 443-470). Stanford: Stanford University Press.

Johnson, E. K., & Jusczyk, P. W. (2001). Word segmentation by 8-month-olds: When speech cues count more than statistics. *Journal of Memory & Language, 44*(4), 548-567.

Johnson, K. (1997). The auditory/perceptual basis for speech segmentation. *Ohio State University Working Papers in Linguistics, 50*, 101-113.

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

Jun, S.-A. (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* (pp. 235-253). Cambridge, United Kingdom: Cambridge University Press.

Jun, S.-A. (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.-A. (1996a). Influence of microprosody on macroprosody: a case of phrase initial strengthening. *UCLA Working Papers in Phonetics, 92*, 97-116.

Jun, S.-A. (1996b). The Phonetics and Phonology of Korean Prosody: Intonational Phonology and Prosodic Structure.

Jun, S.-A. (2000). K-ToBI (Korean ToBI) Labelling Conventions (Version 3.1, October 2000).

Jun, S.-A. (2004a). Korean Intonational Phonology and Prosodic Transcription, *Prosodic typology and transcription: A unified approach* (pp. 341-375). Oxford: Oxford University Press.

Jun, S.-A. (2004b). Prosodic Typology. In S.-A. Jun (Ed.), *Prosodic typology and transcription: A unified approach* (pp. 761-807). Oxford: Oxford University Press.

Jun, S.-A., & Fougeron, C. (2000). A phonological model of French intonation. In A. Botinis (Ed.), *Intonation: Analysis, modelling and technology*. Boston, MA.: Kluwer.

Jun, S.-A., & Fougeron, C. (2002). Realizations of the Accentual Phrase in French intonation. *Probus, 14*, 147-172.

Jun, S.-A., & Kim, S. (2004). *Default phrasing and attachment preference in Korean* (Unpublished manuscript): University of California, Los Angeles.

Jun, S.-A., & Oh, M. (1996). A prosodic analysis of three types of wh-phrases in Korean. *Language and Speech, 39*(1), 37-61.

Jusczyk, P. W. (1999). How infants begin to extract words from speech. *Trends in Cognitive Science, 3*, 323-328.

Jusczyk, P. W., Cutler, A., & Redanz, L. (1993). Infants' sensitivity to predominant stress patterns in English. *Child Development, 64*, 675-687.

Jusczyk, P. W., Friederici, A. D., Wessels, J., Svenkerud, V. Y., & Jusczyk, A. M. (1993). Infants' sensitivity to the sound patterns of native language words. *Journal of Memory and Language, 32*, 402-420.

Jusczyk, P. W., Hohne, E. A., & Bauman, A. (1999). Infants' sensitivity to allophonic cues for word segmentation. *Perception and Psychophysics, 61*, 1465-1476.

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

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

Keating, P., Cho, T., Fougeron, C., & Hsu, C.-S. (2004). Domain-initial articulatory strengthening in four languages, *Papers in Laboratory Phonology VI*. Cambridge, United Kingdom: Cambridge University Press.

Kessler, B., Treiman, R., & Mullennix, J. W. (2002). Phonetic Biases in Voice Key Response Time Measurements. *Journal of Memory & Language, 47*, 145-171.

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

Kim, S. (2001). *The interaction between prosodic domain and segmental properties: domain initial strengthening of fricatives and Post Obstruent Tensing rule in Korean.* Unpublished M.A. thesis, University of California, Los Angeles.

Ladd, D. R. (1996). *Intonational phonology*. Cambridge ; New York: Cambridge University Press.

Lavoie, L. M. (2000). *Phonological Patterns and Phonetic Manifestations of Consonant Weakning.* Unpublished Ph.D. dissertation, Cornell University, Ithaca, NY.

Lee, H.-B. (1974). Rhythm and Intonation of Seoul Koren [in Korean]. *Ehag Yengu (Language Research), 10*(2), 415-425.

Lee, H.-Y. (1990). *The structure of Korean prosody.* Unpublished Doctoral dissertation, University of London.

Lehiste, I. (1960). An acoustic-phonetic study of internal open juncture. *Phonetica, 5*, 1-54.

Lim, B.-J. (2001). The role of syllable weight and position on prominence in Korean, *Japanese Korean Linguistics* (Vol. 9, pp. 139-150). Stanford, CA: CSLI.

Lim, B.-J., & de Jong, K. (1999). *Tonal alignment in standard korean: The case of younger generation.* Paper presented at the Western Conference on Linguistics, University of Texas, El Paso, TX.

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.

Manuel, S. Y. (1990). The role of contrast in limiting vowel-to-vowel coarticulation in different languages. *Journal of Acoustical Society of America, 88*(3), 1286-1298.

Manuel, S. Y., & Krakow, R. A. (1984). Universal and language particular aspects of vowel-to-vowel coarticulation. *Haskins Laboratories Status Report on Speech Research, 77*(7), 69-78.

Massaro, D. W., & Cohen, M. M. (1983). Phonological constraints in speech perception. *Perception and Psychophysics, 34*, 338-348.

Mattys, S. L. (in press). Stress versus coarticulation: Towards an integrated approach to explicit speech segmentation. *Journal of Experimental Psychology: Human Perception and Performance*.

Mattys, S. L., & Jusczyk, P. W. (2001). Phonotactic cues for segmentation of fluent speech by infants. *Cognition, 78*, 91-121.

Mattys, S. L., Jusczyk, P. W., Luce, P. A., & Morgan, J. L. (1999). Phonotactic and prosodic effects on word segmentation in infants. *Cognitive Psychology, 38*(4), 465-494.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18*, 1-86.

McQueen, J. (1996). Word spotting. *Language & Cognitive Processes, 11*(6), 695-699.

McQueen, J., & Cho, T. (2003). *The use of domain-initial strengthening in segmentation of continuous English speech.* Paper presented at the International Congress of Phonetic Sciences, Barcelona, Spain.

McQueen, J., Norris, D., & Cutler, A. (1994). Competition in spoken word recognition: Spotting words in other words. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 621-638.

McQueen, J. M. (1998). Segmentation of continuous speech using phonotactics. *Journal of Memory & Language, 39*(1), 21-46.

Mehler, J., Dommergues, J.-Y., Frauenfelder, U. H., & Segui, J. (1981). The syllable's role in speech segmentation. *Journal of Verbal Learning and Verbal Behavior, 20*, 298-305.

Mehler, J., Jusczyk, P. W., Lambertz, G., Halsted, G., Bertoncini, J., & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition, 29*, 143-178.

Moon, C., Cooper, R., & Fifer, W. (1993). Two-day-olds prefer their native language. *Infant Behavior and Development, 16*, 495-500.

Morgan, J. L. (1996). A rhythmic bias in preverbal speech segmentation. *Journal of Memory & Language, 35*, 666-688.

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

Nakatani, L. H., & Schaffer, J. A. (1978). Hearing "words" without words: Prosodic cues for word perception. *Journal of Acoustical Society of America, 63*(1), 234-245.

Nazzi, T., Bertoncini, J., & Mehler, J. (1998). Language discrimination by newborns: Towards an understanding of the role of rhythm. *Journal of Experimental Psychology: Human Perception and Performance, 24*, 1-11.

Nazzi, T., Jusczyk, P. W., & Johnson, E. K. (2000). Language discrimination by English-learning 5-month-olds: Effects of rhythm and familiarity. *Journal of Memory & Language, 43*, 1-19.

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

Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition, 52*, 189-234.

Norris, D., McQueen, J., & Cutler, A. (1995). Competition and Segmentation in Spoken-Word Recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(5), 1209-1228.

Norris, D., McQueen, J. M., Cutler, A., & Butterfield, S. (1997). The possible-word constraint in the segmentation of continuous speech. *Cognitive Psychology, 34*(3), 191-243.

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

Park, M.-J. (2003). *Where prosody meets grammar and discrouse: A pragmatic interpretation of Korean prosodic boundary tones.* Unpublished Doctoral dissertation, University of California, Los Angeles.

Pierrehumbert, J. B. (2000). Tonal elements and their alignment. In M. Horne (Ed.), *Prosody, Theory and Experiment: Studies presented to Gosta Bruce*: Kluwer.

Pierrehumbert, J. B., & Talkin, D. (1992). Lenition of /h/ and glottal stop. In G. Docherty & D. R. Ladd (Eds.), *Papers in Laboratory Phonology II: gesture, segment, prosody* (pp. 90-117). Cambridge, United Kingdom: Cambridge University Press.

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

Quene, H. (1992). Integration of acoustic-phonetic cues in word segmentation. In M. E. H. Schouten (Ed.), *The auditory processing of speech: From sounds to words* (pp. 349-355).

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

Quene, H., & van den Bergh, H. (in press). On Multi-Level Modeling of data from repeated measures designs: A tutorial. *Speech Communication*.

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

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274*(1926-1928).

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition, 70*(1), 27-52.

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

Schafer, A. J. (1997). *Prosodic parsing: The role of prosody in sentence comprehension.* University of Massachusetts Amherst.

Schafer, A. J., & Jun, S.-A. (2001). Effects of accentual phrasing on adjective interpretation in Korean. In M. Nakayama (Ed.), *Sentence Processing in East Asian Languages*: CSLI publications.

Sebastian-Galles, N., Dupoux, E., Segui, J., & Mehler, J. (1992). Contrasting syllabic effects in Catalan and Spanish. *Journal of Memory and Language, 31*, 18-32.

Turk, A., & Shattuck-Hufnagel, S. (2000). Word boundary related durational patterns in English. *Journal of Phonetics, 28*, 397-440.

*UCLA Phonetic Archives*. Ladefoged, P. Available: http://hctv.humnet.ucla.edu/departments/linguistics/VowelsandConsonants/index.html.

Ueyama, M. (1998). *Speech rate effects on phrasing in English and Japanese*: University of California, Los Angeles.

Valian, V., & Levitt, A. (1996). Prosrody and adults' learning of syntactic stuructre. *Journal of Memory & Language, 35*, 497-516.

Vitevitch, M. S., & Luce, P. A. (1999). Probabilistic phohnotactics and neighborhood activation in spoken word recognition. *Journal of Memory & Language, 40*, 374-408.

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.

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

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

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