

Initial tones and prominence in Seoul Korean*

Ken de Jong

Department of Linguistics, The Ohio State University

Abstract

The present paper reports on an investigation of the informal observation that Seoul Korean nouns have an initial prominence and often have an initial prominence entirely due to an initial intonational prominence, or could other aspects of the signal be involved? second, are initial tones best considered pitch accents (prominence-lending tones associated with a particular syllable) or are they better seen as phrase tones (markers of the edge of a prosodic group) Two subjects recited various sentences at different rates with different levels of overall emphasis. Word initial syllables are not longer than their medial counterparts; the only durational lengthening found is the consonant initial to the phrase. F₀ measurements on initial nouns of various lengths suggests that initial high tones are underlyingly always present, but are obscured in short noun phrases by a following, phrase final tonal prominence. These results suggest that Korean nouns do not have initial stress, but rather occur in prosodic phrases which are marked in part by an initial high tone.

Introduction

Temporal duration and intonational accents have been noted to act together in prosodic system to lend prominence to a particular syllable (Lehiste, 1970; Beckman, 1986). For example, in English, it has been shown that the location of word stress can be cued both by intonational and durational differences (Fry, 1958; Bolinger, 1958, 1965; Lieberman, 1960).

A similar situation might be true of Seoul Korean. Native speakers have suggested informally that there might be stress on the initial syllable of nouns; though this claim does not seem to surface in the literature. Furthermore, Lee (reported on in this volume) has also found initial intonational prominences. Figure 1 is an f₀ track from this study of a sentence with a five syllable noun in subject position. The feature of interest is the initial f₀ maximum, which is realized on the initial syllable of the subject noun. These f₀ maxima appear in all of repetitions of this sentence.

However, if the sentence also contains an adjective previous to the noun, the peak occurs over the adjective, and no peak then appears on the first syllable of the following noun. Thus, the peak location is not a property of the word, but is more likely a property of some larger prosodic domain.

This paper deals with two related questions. First, is the perception of initial stress simply due to the intonational prominence illustrated in Figure 1? Or is there some other aspect of the speech signal which is involved in the perception of stress?

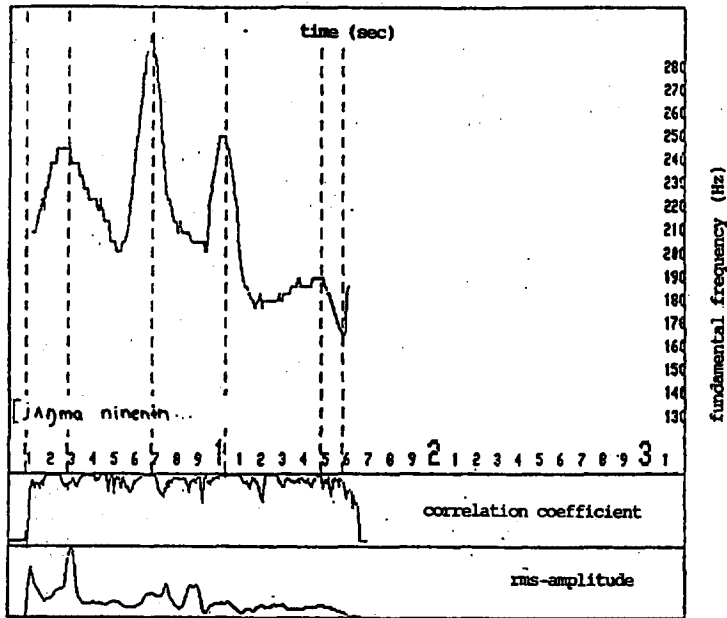


Figure 1. An f_0 trace of a Seoul Korean sentence with a five syllable noun in subject position. The first two f_0 maxima occur over the subject. The first occurs over the initial syllable of the subject-noun, the second, more prominent peak over the subject marker, /nin/.

Stress in English, besides being involved with the occurrence of pitch accents, has its own correlates, duration, amplitude, and more distinct formant structures (e.g., as shown in Fry, 1958; Harris, 1978; Huss, 1978; and Summers, 1987). Second, is the pitch peak apparent in Figure 1 indicative of a prominence lending pitch accent, such as those found in English (Bolinger, 1965; Pierrehumbert, 1980)? To obtain an answer to the first question, I measured the durations of segmentally identical CV sequences placed in word medial and word initial positions. If Korean does have word initial stress, one would expect consistent durational differences between all word initial syllables and medial syllables. If, however, there is no syllabic stress, there would be no differences between word medial and word initial syllables.

A third and a fourth possibility also exist. The durational difference may be a correlate of a pitch accent. Perhaps the f_0 peak is indicative of a pitch accent which has durational lengthening associated with it. Pitch accents in Swedish and English have such lengthening associated with them, (Stålhammar, et al. 1973; Nord, 1986, and Engstrand, 1988 for Swedish; Beckman, Edwards, and Fletcher, 1992, de Jong, 1991, for English).

Alternatively, durational differences may be an edge marker of some prosodic unit not isomorphic with the word, for example, an accentual or intonational phrase of the type proposed in Pierrehumbert (1980), Selkirk (1980), and Pierrehumbert and Beckman (1988). I am assuming in this scenario that the initial f_0 maximum

illustrated in Figure 1 also marks the edge of this domain. Thus, if either of these possibilities is correct, word initial syllables should only be longer than word medial syllables when they are phrase initial as well. When the noun is preceded by an adjective, the lengthening should occur on the first syllable of the adjective, not on the first syllable of the noun.

To obtain an answer to the second question, I must address two additional questions. First, what is the relationship between the tone and the segmental material with which it appears? Pitch accents in English are phonologically associated with a stressed syllable. Thus, they seem to be temporally aligned with the sonority peak of the stressed syllable (Silverman and Pierrehumbert, 1990). Second, can the initial pitch peak be construed as the property of a prosodic domain? Pitch specifications, called phrase tones or phrase accents, can appear as (usually peripheral) markers of a prosodic domain. English pitch accents contrast with phrase tones by not being associated with a prosodic domain. They simply appear optionally before a nuclear accent in an intonational or intermediate phrase (Pierrehumbert, 1980; Pierrehumbert and Beckman, 1988).

With regard to this second issue, the initial intonational peaks found by Lee occurred consistently in the five syllable tokens of which Figure 1 is an example. However, as can be seen in Figure 2, the initial peak did not always appear in three syllable tokens. Thus, the pitch peak might be phonologically optional, just as English pitch accents are. In English, prenuclear accents tend to appear near the beginning of an intonational phrase, especially when the nuclear accent is placed somewhat later in the phrase. With much shorter phrases, the subject will often forgo the prenuclear accents.

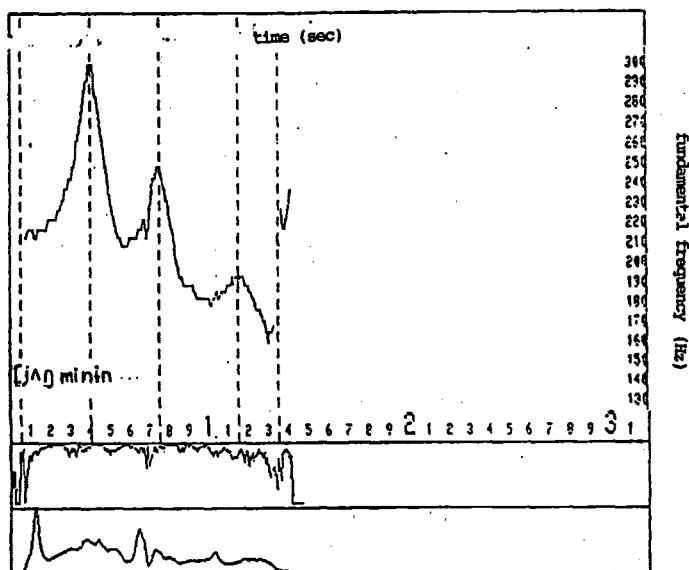


Figure 2. An f_0 trace of a Seoul Korean sentence with a three syllable subject. The first maximum occurs over the subject marker, /nin/, which appears immediately following the noun.

However, there is another possible interpretation of the Korean patterns. One might claim that the initial pitch peaks are indicative of phrase tones which occur consistently at the beginning of some prosodic domain. To do so, one must say that the initial peak does exist underlyingly even in the utterance shown in Figure 2. However, it does not appear on the surface due to the proximity of the following, much higher tonal prominence on /nin/.

The initial tonal prominence appears clearly in the utterance shown in Figure 1, because the high tone on /nin/ is too far away temporally to obscure the initial peak. If this is, in fact, what is occurring here, one should find a continuum from clear f0 peaks to completely obscured peaks, with intermediate forms with partially obscured peaks. If, however, the optional pitch accent analysis is correct, there should be no transitional forms either the accent is there, or it is not.

In order to answer these questions about the status of the initial pitch peak, a series of sentences were recorded similar to those recorded for the durational analysis, except that the subject nouns varied in length. Since the utterances in Lee's corpus consistently have a major intonational prominence on the subject marker appended to the right edge of the subject noun, this corpus shows the behavior of the initial tonal maximum as the following tonal prominence becomes temporally closer. Of interest is the alignment of the initial f0 maximum with respect to the initial syllable of the noun, and the height of the initial f0 peak as the following f0 peak draws closer. In order to explore a range of prosodic possibilities, emphasis and rate were also manipulated explicitly.

Methods

Two speakers of the Seoul dialect of Korean, one male (HKK) and one female (MRO), were recorded reading a corpus of sentences written in Korean script in an anechoic chamber. Both speakers were graduate students in linguistics, but were naive to the objectives of the experiment. Each of the sentences contained the syllable placed in various morphosyntactic positions, either in the subject or in the object, either initial to its noun phrase or preceded by an adjective, either medial or initial to a word. Except for the target syllables placed noun phrase initially, all of the targets followed an /i + nasal/ sequence. /sək/ appeared twice in each sentence, making four sentence types in all, which are given in Table 1. Each sentence was repeated five times in the corpus. Subject HKK read the corpus four times, each at a different rate or emphasis; first, 'normal,' then, 'slowly and distinctly,' 'fast,' and finally 'quick and emphatically.' Subject MRO read the corpus five times, 'normal,' 'slow', 'quickly and excitedly,' 'slowly and emphatically,' and finally 'fast.' The words and glosses are given in Table 1.

Table 1. Corpus for duration measurements

1. Noun phrase initial
 - a: [kimsəksan-i-ninsəkjəŋtʰəl-i-lilmanat-ta]
'Kim Suksan met Suk Yungchul'
 - b: [səkjəŋtʰəl-i-ninkimsəksan-i-lilmanat-ta]
'Suk Yungchul met Kim Suksan'
2. Noun phrase medial
 - a: [məttʰin kimsəksan-i-nin məttʰin səkjəŋtʰəl-i-lilmanat-ta]
'Good looking Kim Suksan met good looking Suk Yungchul'
 - b: [məttʰin səkjəŋtʰəl-i-nin məttʰin kimsəksan-i-lilmanat-ta]
'Good looking Suk Yungchul met good looking Kim Suksan'

The sentences were then digitized and durational measurements were made on the target syllables using a waveform editor. The target syllable duration was measured as the time from the onset of perceptible frication for the /s/ to the last nonsinusoidal pitch pulse at the offset of the vowel. The acoustic duration of the vowel was also measured as the time from the first glottal pulsation without high frequency frication to the last nonsinusoidal pitch pulse at the offset of the vowel. For a reference to factor out differences due to rate, the duration of the verb /mannat/ was also measured. Its duration was the time from the onset of sinusoidal pulsation for the /m/ to the last perceptible pitch pulse at the offset of the second /a/.

The two subjects were again recorded reading a corpus of sentences in which the subject noun varied in length from three to five syllables. Also, the sentences differed in that half contained an adjective previous to the noun, and half did not. The sentence types are listed in Table 2. Subjects HKK and MRO were instructed to read through the corpus four and five times, respectively, varying rate and emphasis as is described above for the durational corpus. The token sentences were then digitized and analyzed for f0 using an autocorrelation routine on a PC6300 under MSDOS.

Table 2. Corpus for f0 measurements

No. of syll.

| | |
|---|--|
| 3 | [jægmi-ninjəga-liltʃoo-hæ] 'Youngmee likes Young-Ah' |
| 4 | [jægman-i-ninjəga-liltʃoo-hæ] 'Youngman likes Young-Ah' |
| 5 | [jægman-i-ne-ninjəga-liltʃoo-hæ] 'Youngman's people like Young-Ah' |
| 6 | [uəhən jægmi-ninjəga-liltʃoo-hæ] 'Graceful Youngmee likes Young-Ah' |
| 7 | [uəhən jægman-i-ninjəga-liltʃoo-hæ] 'Graceful Youngman likes Young-Ah' |
| 8 | [uəhən jægman-i-ne-ninjəga-liltʃoo-hæ] 'Graceful Youngman's people like Young-Ah' |

First, to examine the alignment of the initial tonal peak with respect to the initial syllable, and to determine the temporal proximity of the following intonational prominence on /nin/, the time of the initial intonational peak, the time of the major peak on /nin/, the onset of the utterance (taken as the onset of the first perceptible pitch pulse in the /j/), and the offset of the first syllable (taken as an amplitude minimum characteristic of the offset of the /ŋ/) were noted. An illustration of these measurements is given in Figure 3.

To examine the relative height of the f0 peaks, three f0 values were noted. The value of the initial f0 peak was taken as the first reliable f0 maximum, i.e. the first f0 value in or after the first vowel, which was higher than subsequent values. The value of the final f0 peak on /nin/, and the time and value of the lowest f0 value between the two peaks which occurred during a vowel were taken.

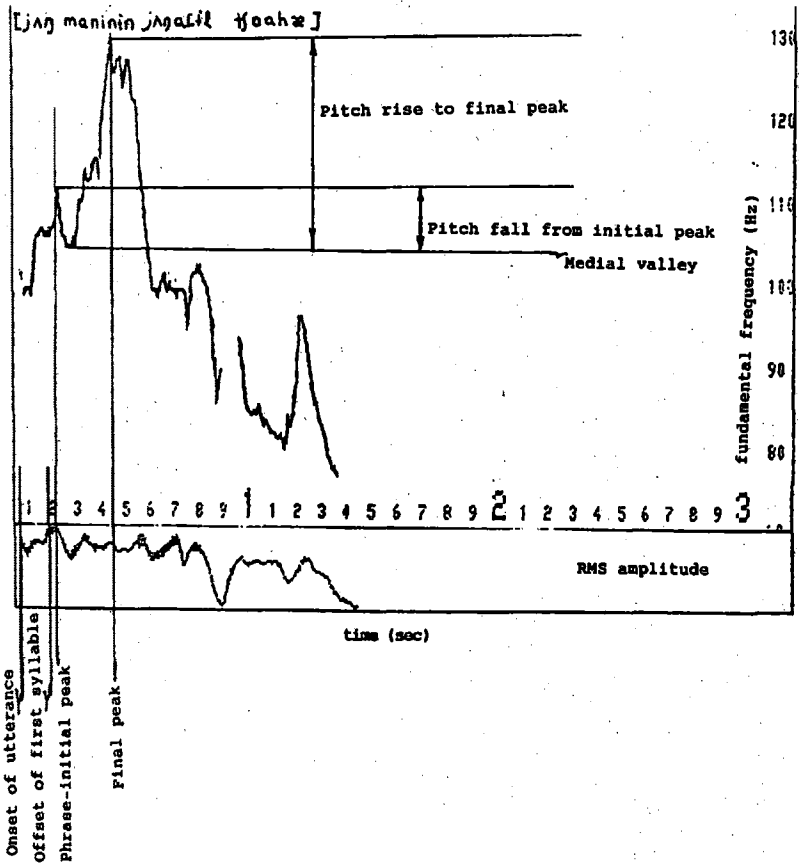


Figure 3. An f_0 trace and intensity envelope for one of the token sentences with the temporal and f_0 measures used to analyze the status of the initial f_0 peak.

Results

Figure 4 shows that phrase initial tokens both in subject and object position are longer at a given rate than are medial tokens in the same sentences for subject HKK. The same difference appears for subject MRO, however not as consistently. Figure 5 shows that both subjects have a durational difference between word initial and word medial tokens of the same sentence also when there is an adjective preceding the word in the phrase. Word initial syllables tend to be longer than word medial tokens, regardless of the linguistic context of the word.

Closer inspection of these durational differences reveals that they are not the same as the durational differences correlated with stress in English. If one

measures only vowel durations, the picture is somewhat different. Figure 6 shows that, for both speakers, there are no clear differences in vowel duration between word initial and medial syllables. At any given rate, word initial vowels and word medial vowels have the same duration. Thus, the word initial durational differences found in syllable durations seem to be localized to some extent in the initial consonant.

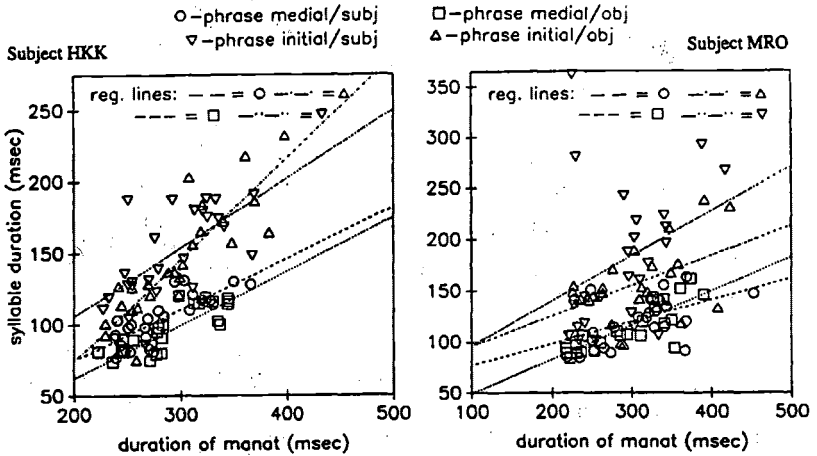


Figure 4. The duration of word initial and medial syllables occurring both in phrase initial and phrase medial positions plotted against the duration of the following verb, 'mannat' to normalize for rate. That phrase initial tokens tend to be longer than phrase medial tokens is evident in the correlation line for the phrase initial tokens being much higher in the plot than that for the medial tokens.

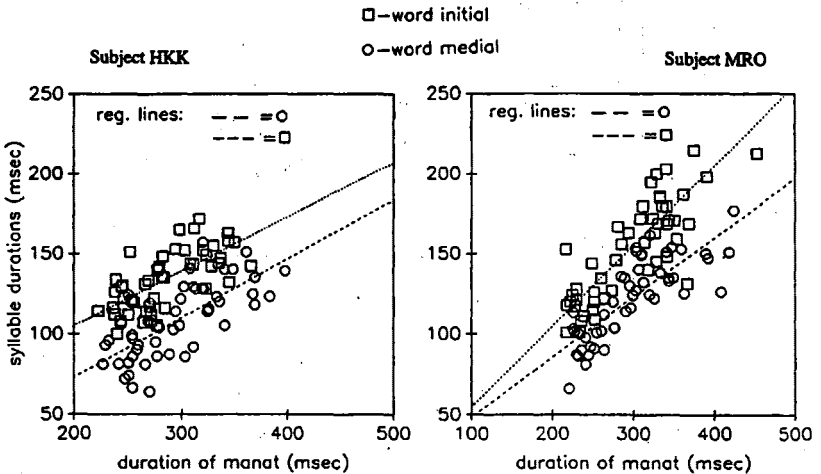


Figure 5. The duration of word initial and medial syllables occurring in phrase medial nouns plotted, as in Figure 4, against the duration of the following verb to normalize for rate.

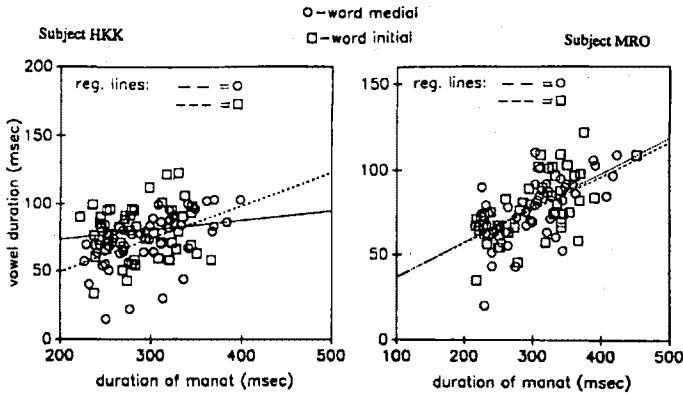


Figure 6. The duration of the vowel in initial and medial syllables occurring in phrase initial nouns plotted against the duration of the following verb to normalize for rate. Of note here is the fact that the regression lines for the two conditions lie atop one another. Patterns for phrase medial nouns were similar.

Turning to the analysis of the second corpus, the initial pitch peak appeared clearly in many of the two subject's utterances. The left half of Figure 7 shows similar patterns to those of Lee. A pitch peak appears in the tokens with more syllables, but not in tokens with fewer syllables. These contours are examples of MRO's renditions of the 'fast and emphatic' reading of the corpus. An investigation of the traces on the right, which are examples of her 'slowly and distinctly' reading of the corpus, shows a somewhat different pattern. All three of these traces have a contour similar to that in Figure 1. The initial peak is apparent even in the trisyllabic 'slowly and distinctly' tokens. The same was true of HKK's tokens.

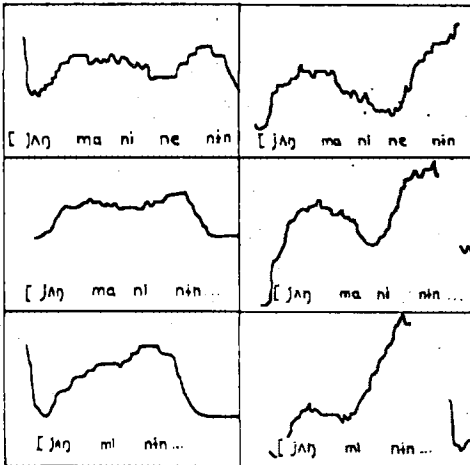


Figure 7. F0 tracks for six of MRO's utterances. Tokens vary from five syllable nouns at the top to three syllable nouns at the bottom. Tokens to the left are 'fast and distinct,' while tokens to the right are 'slow and emphatic.' Time scale and frequency scale vary from panel to panel.

I turn now to the alignment of the initial f_0 peak. There is a fairly tight alignment of the initial tone with the end of the initial syllable for both subjects. Figure 8 shows a fairly restricted range of variation in tone position. Though the tone does not always appear over the initial syllable (as is shown by positive latencies in Figure 8), it never appears more than 50 ms. past the offset of the first syllable for HKK. Except for a few of MRO's "normal" renditions, her tokens also cluster around the offset of the first syllable (within 100 ms. or so). A few of the tokens also come somewhat earlier. MRO occasionally has early peaks in her "slow" renditions. Roughly half of HKK's slow and emphatic tokens also have an unusually early pitch peak, along with tokens from other conditions where the following f_0 maximum occurs very closely following the initial peak (shown to the left in Figure 8).

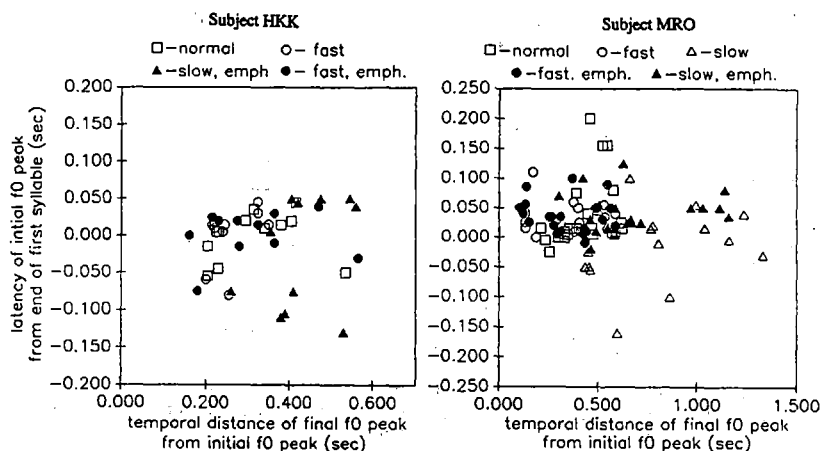


Figure 8. The latency of the initial f_0 maximum from the offset of the first syllable of the phrase plotted against the latency of the f_0 maximum on /nin/ from the initial f_0 maximum.

The latter tokens may be reflecting tonal repulsion, which has also been shown to exist in English (Silverman and Pierrehumbert, 1990). The occurrence of the tones on or after the offset of an initial syllable has also been noted of English pitch accents (also in Silverman and Pierrehumbert, 1990). There seems to be some tendency for f_0 peaks to be repulsed away from utterance initial edges. Thus, the tight clustering of f_0 peaks around the offset of the initial syllable suggests that the initial pitch peak is associated with the initial syllable just as English pitch accents are associated with particular syllables.

Turning to the prominence of the pitch peak, Figure 9 shows that there is a fairly strong correlation between the prominence of the initial peak, measured as the amount of subsequent fall in f_0 , and the proximity of the later peak (HKK, Pearson's $r = 0.737$; MRO, $r = 0.689$). In addition to the reduction of fall from the initial high apparent in HKK's tokens, MRO's slower tokens with more than 500 ms. between the two tonal peaks show little variation in the height of fall from the initial high. The difference in the patterns for these two speakers seems to have arisen because HKK simply did not have tokens long enough to show initial highs without the effect of the following high. The leveling off of fall from the initial

high for MRO's extremely long tokens shows a consistency in the height of the initial high when the effect of the following high is eliminated.

The behavior of the subject final high is in striking contrast to the initial high. Figure 10 shows that, for HKK, bringing the two tones closer together has no appreciable effect on the amount of rise to the subject final high. However, there is a relationship between the temporal distance between the two peaks and the amount of rise to the final peak in MRO's tokens.

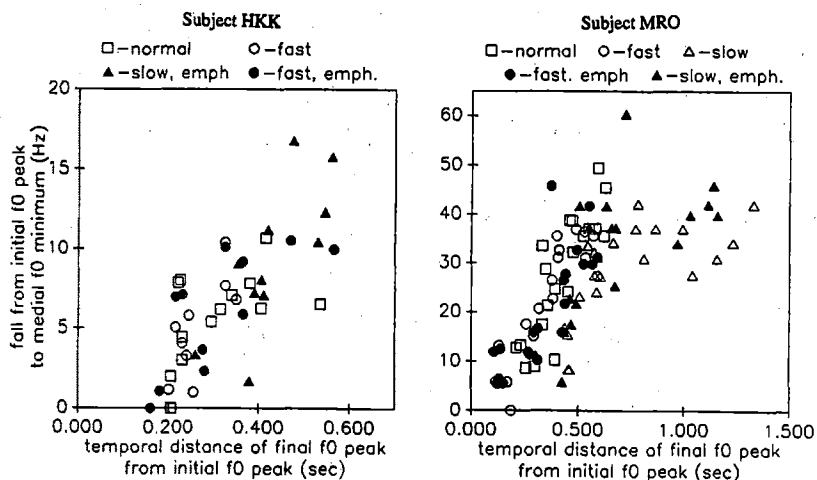


Figure 9. The amount of f0 fall (in Hz) from the initial f0 maximum to the lowest f0 before the rise to the subject final f0 maximum on /nin/ plotted against the latency (in sec.) of the f0 maximum on /nin/ from the initial f0 maximum.

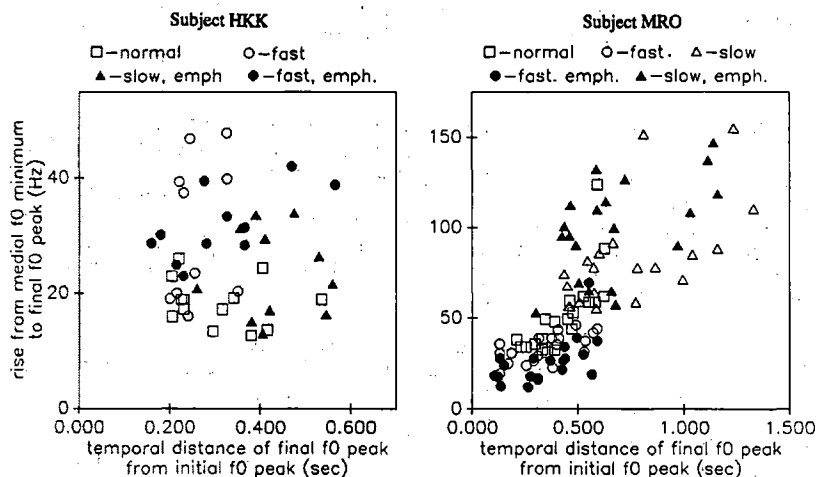


Figure 10. The amount of rise from the f0 minimum following the initial f0 maximum to the subject final f0 maximum on /nin/ plotted against the latency of the f0 maximum on /nin/ from the initial f0 maximum.

Looking closer at MRO's tokens, there is clear tendency toward separation between tokens in different conditions. MRO performed much greater changes in pitch range between the five readings of the corpus than did HKK. So, some of the correlation of pitch rise and time between f_0 peaks seems to be due to concurrent effects of the stylistic manipulations on both rate (affecting time between f_0 peaks) and pitch range (affecting amount of pitch rise). However, not all of the relationship is due to condition differences, because splitting the tokens by condition of the corpus does not eliminate the relationship. A positive correlation is found across the tokens within each condition (normal, $r = 0.746$; slow, $r = 0.605$; slow, emphatic, $r = 0.545$; fast, $r = 0.678$; fast, emphatic, $r = 0.570$).

These correlations disappear when one does not consider the amount of fall from the initial high. As can be seen in Figure 11, if one looks at the difference in the height of the two peaks, any relationship with the distance between the two peaks essentially disappears. Thus, it seems that the depth of the valley is the crucial portion of the f_0 rise that is being affected by changes in the temporal distance between the two peaks.

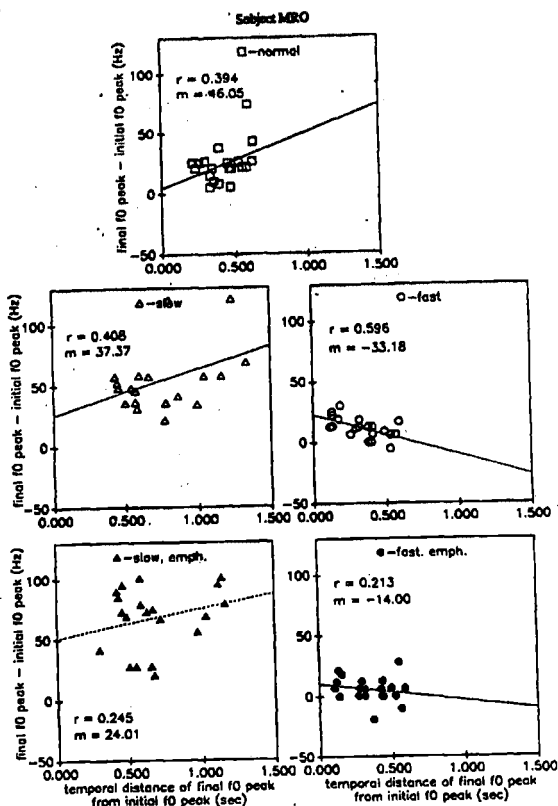


Figure 11. The difference between the value of the initial f_0 maximum and subject final f_0 maximum on /ni n/ plotted against the latency of the f_0 maximum on /ni n/ from the initial f_0 maximum. Tokens from MRO shown here are broken down by rate and emphasis condition.

The pattern, then, of the present data can be schematized as in Figure 12. The subject final high is obscuring the initial peak as it draws closer in time. This being the case, the disappearance of the peak in very short subject phrases under fast conditions is not positive evidence against the phrase accent analysis of the initial tones.

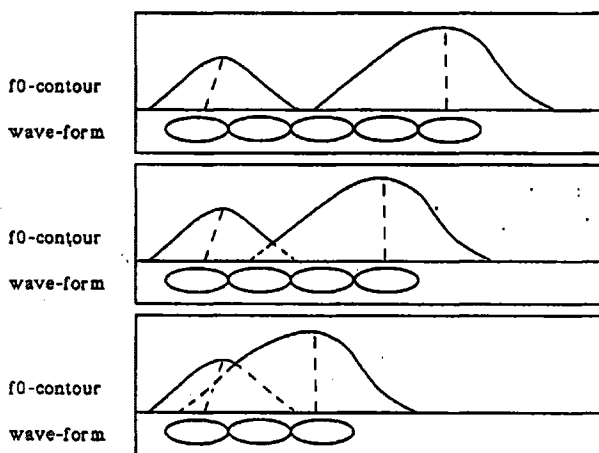


Figure 12. A schematic illustration showing the behavior of the initial and subject final tones as the two tones draw closer together in time.

Discussion

Word initial syllables in Seoul Korean tend to be longer than word medial syllables, regardless of the position of the word in the sentence. Thus, there seems to be a left edge durational effect for words. It is clear that this difference cannot be associated with the appearance of the initial tonal prominence illustrated in Figure 1, since it also appears in nouns preceded by an adjective which do not bear an initial tone.

However, it is probably wrong to see these durational differences as being identical to those associated with stress in English. The distribution of durational differences within the syllable is somewhat different being localized rather strongly in the initial consonant. This pattern of localization suggests that the durational differences may be an "edge effect," similar to final lengthening, and not a "head effect," such as those associated with greater prominence. Beckman, Edwards, and Fletcher (1992) have shown that strikingly different articulatory mechanisms are associated with the two different kinds of lengthening; accentual lengthening arises from a change in the opening and closing gestures of the stressed syllable, while final lengthening tends to affect more the closing movement at the end of the vowel (see Summers, 1987, for a similar finding for consonant associated lengthening versus stress). Thus, the durational effects found here seem to be a case of "initial lengthening."

Some of the f0 analyses reported here suggest that the initial tone is similar to the prominence lending pitch accents in English. The peak tends to be closely

aligned in time with the offset of the initial syllable. This association, thus, could partially account for the perception of 'stress' on the initial syllable of nouns..

However, other f0 analyses reported here suggest that the initial peak is not to be seen as a prominence lending pitch accent. Unlike prenuclear accents in English, the peak always occurs underlyingly on initial syllables, and is not optional. The peak consistently appears when the final f0 peak is relatively far away in all renditions of the corpus for both speakers. Between these examples and those in which no initial peak is apparent, there is a continuous and predictable string of intermediate forms varying in the depth of the valley intervening between the peak and its right neighbor.

Conclusion

Cross-linguistically, prosodic effects which appear very similar on the surface, might turn out to have very different roles in the prosodic systems of the two languages. Only measurement of the right aspects of the speech production activity and signal emitted across appropriate linguistic and stylistic manipulations can tease out these differences. Of such an analysis of Seoul Korean, this paper is but a small part.

Acknowledgments

I thank Sook-hyang Lee and Sun-Ah Jun for constructing the corpora and assisting in the elicitation procedures described herein.

* An earlier version of this paper was presented at the 119th meeting of the Acoustical Society of America, May, 1990.

References

- Beckman, M.E., J. Edwards, and J. Fletcher. 1992. "Prosodic structure and tempo in a sonority model of articulatory dynamics." In Docherty, G. and D.R. Ladd, eds. *Papers in Laboratory Phonology II: Gesture, Segment Prosody*. Cambridge University Press.
- Bolinger, D. 1958. "A theory of pitch accent in English." *Word*, 7: 199-210.
- Bolinger, D. 1965. *Forms of English: Accent, Morpheme, Order*. Harvard University Press.
- Bruce, G. 1977. *Swedish Word Accents in Sentence Perspective*. (Travaux de L'institut de Linguistique de Lund) CWK Gleerup.
- de Jong, K.J. 1991. *The Oral Articulation of English Stress Accent*. Unpublished dissertation. Columbus, Ohio. Ohio State University.
- Edwards, J., M.E. Beckman, and J. Fletcher. 1991. "The articulatory kinematics of final lengthening." *Journal of the Acoustical Society of America*, 89(1): 369-382.
- Engstrand, O. 1988. "Articulatory correlates of stress and speaking rate in Swedish CV utterances." *Journal of the Acoustical Society*, 88(5): 1863-1875.
- Fry, D.B. 1958. "Experiments in the perception of stress." *Language and Speech*, 1: 126-152.
- Harris, K.S. 1978. "Vowel duration changes and its underlying physiological mechanisms." *Language and Speech*, 21(4): 354-361.
- Huss, V. 1978. "English word stress in the postnuclear position." *Phonetica*, 35: 86-105.

- Lehiste, I. 1970. *Suprasegmentals*. Cambridge, MA.: MIT Press.
- Nord, L. 1986. "Acoustic studies of vowel reduction in Swedish." *STL Quarterly Progress Status Report*, 1986. (4): 19-36.
- Pierrehumbert, J.B. 1980. *The Phonology and Phonetics of English Intonation*. Unpublished dissertation. Cambridge, MA.: MIT.
- Pierrehumbert, J.B., and M.E. Beckman. 1988. "Japanese Tone Structure". *Linguistic Inquiry*, Monograph 15. Cambridge, MA: MIT Press.
- Selkirk, E.O. 1980. "The role of prosodic categories in English word stress." *Linguistic Inquiry*, 11: 563-605.
- Silverman, K. and J.B. Pierrehumbert. 1990. "The timing of prenuclear high accents in English." In Kingston, J., and M.E. Beckman, eds. *Papers in Laboratory Phonology I: Between the Grammar and the Physics of Speech*. Cambridge, MA. Cambridge University Press.
- Stålhammar, U., I. Karlsson, and G. Fant. 1973. "Contextual effects on vowel nuclei." *STL - Quarterly Progress Status Report*. 1973. (4): 1-18.
- Summers, W.V. 1987. "Effects of stress and final consonant voicing on vowel production: Articulatory and acoustic analyses." *Journal of the Acoustical Society of America*, 82: 847-863.