

The effects of prosody on segmental variation

Kyuchul Yoon*

Department of Linguistics
The Ohio State University
1712 Neil Avenue, Columbus, Ohio
43220, USA

kyoon@ling.osu.edu

Keywords: prosody, Korean

Abstract

This paper is a production study on the effects of Korean prosody on two voiceless coronal fricatives /s^h/ and /s*/. The target segments were embedded in four prosodic positions: initial to the Intonational Phrase or the Accentual Phrase, and medial to the Accentual Phrase or the Prosodic Word. Acoustic measurements showed that although there are segmental differences associated with the /s^h/ versus /s*/ contrast, these differences vary in magnitude in different prosodic positions, confirming that segmental properties are affected by prosodic categories. This suggests that speech synthesizers should take the prosodically conditioned segmental variations into consideration and pay more attention to subsegmental variation.

1 Introduction

It is well known that speech segments show variation according to position in a word or stress foot. In earlier studies of allophony, focus has been on word-level environments, e.g. whether the target segment is word initial or medial, or whether the syllable containing the segment is stressed or not. It is true that speech segments show systematic variation relative to their word-level position and the location of stress, but as will be shown later, the range of such variation is not limited to word-level prosody.

There is increasing evidence that what is true at the word level is also true at higher levels. Depending on their level in the prosodic hierarchy of a language, speech segments vary. Prosodic categories such as accentual phrases in Korean have been shown to act as the domain of application of such rules as post-obstruent tensing and vowel shortening (Jun 98). Other works have demonstrated that segmental properties of Korean coronal stops (Cho & Keating 01) and fricatives (Kim

01) are affected by higher prosodic domains. The prosodic effects on segments for English have been studied as well ((Pierrehumbert & Talkin 92) for /h/ and glottal stop and (Smith 97) for American /z/).

This paper is a study of the production of the two Korean coronal fricatives /s^h/ (aspirated) and /s*/ (tense) when embedded in different prosodic categories. Two questions are considered: 1) Is the role of segmental properties that distinguish one fricative from the other invariant across different prosodic positions? 2) Do the acoustic cues that distinguish the two fricatives in isolation contribute to the same extent to the distinction of the fricatives when embedded in different prosodic positions? By examining the two fricatives in different positions in the Korean prosodic hierarchy, we will get some insight as to how segmental properties adapt to different prosodic positions. In addition, by looking more closely at the segmental variations of /s^h/, whose noise interval is known to contain aspiration, we will be able to gain more insight into prosodically-conditioned subsegmental variation.

1.1 Prosodic hierarchy of Seoul Korea

Figure 1 is an illustration of the prosodic hierarchy of Seoul Korean. There are two tonally defined prosodic categories above the level of the prosodic word (PW) in Korean, i.e. the intonational phrase (IP) and the accentual phrase (AP) (Jun 93). The arrows indicate the four prosodic positions and these prosodic positions will be used as place holders for the two target fricatives in the experiment.

2 Method

2.1 Stimuli

A novel approach was used to devise the carrier sentences. Rather than having the speakers repeat citation form sentences, we have created carrier “conversations” where two speakers take

I would like to thank Mary Beckman, Chris Brew, Shari Speer, the members of Phonies at OSU (especially David Odden, Tsan Huang, Giorgos Tserdanelis, and Jeff Mielke), Sunhee Lee, Soyoung Kang and Misun Seo for their comments, and six Korean speakers for their help in recording

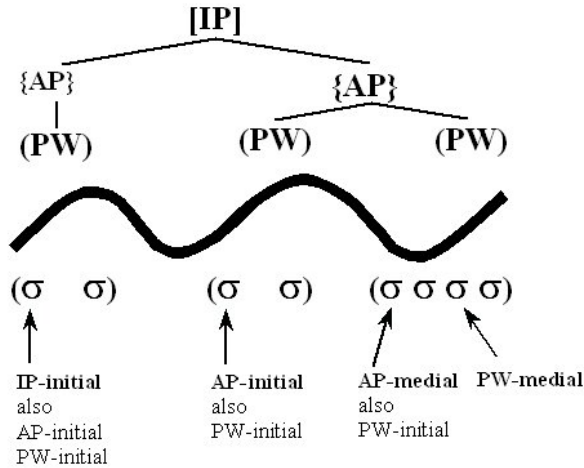


Figure 1: Prosodic hierarchy of Seoul Korean. Arrows indicate the positions where target segments will be embedded in carrier conversations. Solid wavy line indicates the schematized pitch contour of an utterance.

turns to simulate a conversation. I believe that this approach is more natural than simple citation form utterances if natural corpora are not available.

Two prosody-carrier conversations were devised to simulate situations where the desired prosody would be generated. The target fricative segments were embedded in (1) IP-initial, (2) AP-initial (but not IP-initial), (3) AP-medial (but PW-initial), and (4) PW-medial position.

In order to put the target segments in AP-initial versus AP-medial position, two strategies were employed, one involving corrective contrast, and the other involving the form *nuka*, which can be either an interrogative pronoun (‘who’) or an indefinite pronoun (‘someone’), depending on the phrasing of the pronoun-verb sequence. In both cases, pitch tracks were made to verify that the speakers actually produced the desired phrasings.

The first strategy, adapted from (Jun *et al.* 97), involves a short conversation, where a mom asks her daughter something and the daughter misunderstands and says something else, which her mom corrects.

The second strategy uses two different types of questions; a *wh*-question and a *yes/no* question containing the indefinite pronoun *nuka*. Speakers use different phrasing (Jun & Oh 96) to disambiguate the two types of sentence, putting an AP boundary after *nuka* for *yes/no* questions but grouping following material into the same AP as *nuka* for the other type of question.

2.2 Subjects & recording

Six native speakers of Seoul Korean (three male and three female) participated in the recording. Three groups, each consisting of two speakers, enacted the two prosody-carrier conversations until each pair produced ten (five for each member of the group) clear repetitions of each sentence. The total number of tokens was 6 (subjects) x 2 (prosody-carrier conversations) x 2 (fricatives) x 4 (prosodic positions) x 5 (repetitions) = 480 tokens.

2.3 Measurements

The following measurements were made for the target segment and adjacent vowels: (1) durations of the four relevant sections, i.e. fricative, aspiration noise segment, and preceding and following vowel segment, and (2) harmonic ratio (as a measurement of voice quality) in the following vowel.

2.3.1 Duration of fricative, aspiration & adjacent vowels

Standard criteria for segmenting were generally used. The most difficult segmentation point - namely drawing the dividing line between the fricative and aspiration noise - was determined as follows. In brief, we have written a Praat (version 4.0) script that will repeat what a trained phonetician would do for a set of spectrograms from a particular speaker.

First, spectra were made every 5 ms with a 5 ms-Hamming window from the start of fricative to the vowel onset. Each spectrum was divided into a high frequency band and a low frequency band across a variable (but, once determined, consistent for each speaker) frequency value that had been pre-determined by examining a good selection of spectra for each speaker. In most cases, the frequency value corresponded to the lower cutoff frequency of the fricative segment.

The amplitude values within each of the two bands were summed. The ratio of high-frequency band to low-frequency band, $\sum E_H / \sum E_L$, was used as an indicator of the relative ratio of the turbulence energy generated by air passing through the coronal constriction to the energy generated by air passing through the open glottis. This ratio was examined for each successive spectrum starting at the beginning of the fricative noise until the ratio fell below the criterion value (See Figure 2); that point can be taken as reflecting the shift in the location of the noise source. We can use that

point as the end of frication and start of aspiration. This is illustrated in Figure 2 for a sample token by Speaker f1.

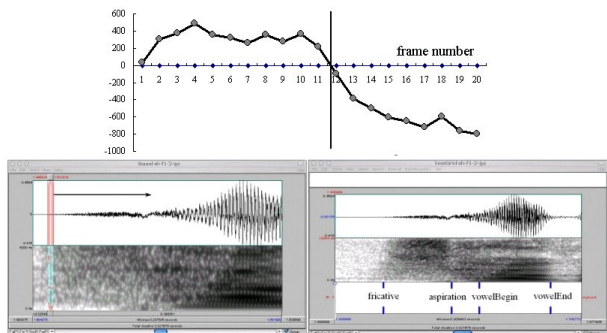


Figure 2: (Top panel) Plot of $\sum E_H / \sum E_L$ change for successive frames of spectra. The vertical line indicates the point where $\sum E_H / \sum E_L$ crosses the criterion value, whose x intercept was taken as the end of fricative and start of aspiration. (Bottom left panel) A slice was selected for spectral analysis. A series of spectra were made in the direction of the arrow. (Bottom right panel) After a series of comparisons of the ratio, a dividing line (the second vertical line in the bottom tier) was drawn.

As shown in Figure 2, after scanning through the spectra, the script inserted a dividing line into a labeling tier.

2.3.2 Harmonic ratio (H1-H2) in following vowel

To assess voice quality, another series of spectra was made every 10 ms from the start of the vowel in question with a 512-point (23 ms for a sampling rate of 22,050 Hz) Hamming window. The series continued until the right edge of the Hamming window crossed the vowel offset. The amplitude difference between the first two harmonics (H1-H2) was obtained for each of the spectra and its time course was plotted.

3 Results

3.1 Harmonic ratio

The change of harmonic ratio (H1-H2) for Speaker f1 in the second prosody-carrier conversation is given in Figure 3.

In most cases, it appears to be the first half of the vowel that displays a clear harmonic ratio difference. $/s^h/$ tokens have a higher harmonic ratio at the beginning of the following vowel whereas $/s^*/$ tokens have a lower harmonic ratio. These plots agree with observations from other studies ((Cho *et al.* 00); (Kim 01)). They reported that H1-H2 values at the vowel onset were positive for

$/s^h/$, reflecting breathy voice quality, and negative for $/s^*/$, reflecting pressed voice quality.

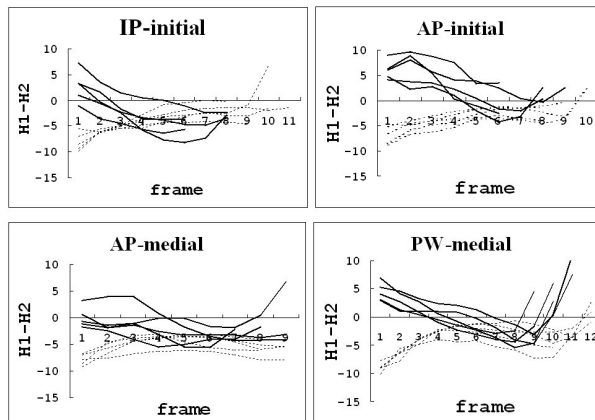


Figure 3: Harmonic ratio (H1-H2 in dB) change for Speaker f1 in the second prosody-carrier conversation. Frame interval is 10 ms. Solid lines for $/s^h/$ and dotted lines for $/s^*/$.

Unlike earlier studies that showed clear harmonic ratio differences between vowels following aspirated obstruents and tense obstruents, this study shows quite a few cases where this is not true (See Figure 4).

This suggests that although harmonic ratio difference reflects the different voice quality of the vowels following the two fricatives, it is not an invariant parameter for $/s^h/$ versus $/s^*/$ contrast in all prosodic positions.

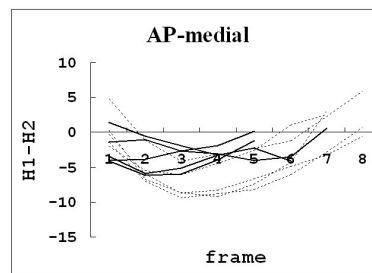


Figure 4: Harmonic ratio (H1-H2 in dB) change of subject f3 in her second conversation.

3.2 Duration

The average durations in the first prosody-carrier conversation of Speaker f3 are given in Figure 5. For the set of $/s^h/$'s in the four prosodic positions, it appears that the fricative and aspiration noise durations tend to decrease from IP-initial to PW-medial. For this speaker, the fricative noise duration roughly divides IP-initial $/s^h/$ from the rest of $/s^h/$'s. Taking the aspiration noise duration and following vowel duration into account,

the three /s^h/s in positions other than the one in IP-initial position can also be divided into two groups, AP tokens versus PW-medial token. Following vowel duration is shorter for AP tokens than for PW-medial tokens.

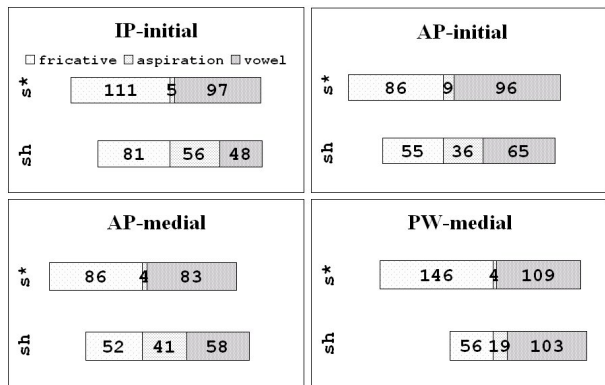


Figure 5: Average duration of fricative, aspiration, and vowel segment of five repetitions of target utterances in the first prosody-carrier conversation for Speaker f3. Segments are arranged in reference to the aspiration noise onset.

For /s*/ tokens in the four prosodic positions, the fricative noise duration was longest for PW-medial tokens. This is consistent with a previous study (Oh & Johnson 97), which analyzed tense stops as geminate, and among “degeminated” positions, the duration was longest in IP-initial position. Given this, one way of characterizing the biggest effect of prosodic position would be: “tense fricatives are geminate PW-medially and non-tense fricatives are aspirated IP-initially”.

If we look at the two fricatives /s^h/ and /s*/ in each prosodic position, fricative noise duration and aspiration noise duration seem to play important roles. The difference in fricative noise duration is greatest in PW-medial position. Here the /s^h/ versus /s*/ contrast appears to be mainly a function of fricative noise duration.

It seems that the decrease in the aspiration noise duration of /s^h/ toward the PW-medial position was compensated for by an increase in the fricative noise duration of /s*/ in that same position. In addition, the decrease of fricative noise duration in the /s^h/ toward the PW-medial position ends up augmenting the long fricative noise duration of /s*/ in PW-medial position, a possible ‘dual’ action, making the difference in fricative noise duration a more convincing acoustic parameter for the contrast in PW-medial position.

Overall, the differentiation of the two fricatives is maintained throughout the four prosodic posi-

tions despite changes in the acoustic parameters involved.

Therefore, we can tentatively hypothesize that while a primary cue is the amount of aspiration noise for the distinction of /s^h/ and /s*/ in IP-initial position, confirming my earlier study (Yoon 99), a better differentiated cue for the distinction between the /s^h/ and /s*/ in PW-medial position is the fricative noise duration.

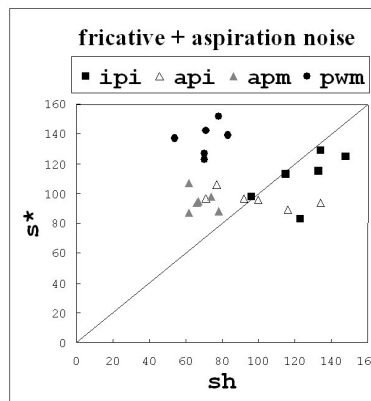


Figure 6: Scatterplot of noise duration (fricative + aspiration duration) for all subjects.

The duration of the whole noise (by adding the fricative interval to the aspiration interval) is plotted in Figure 6. In IP-initial position, the noise duration for /s^h/ is greater while in PW-medial position, the duration for /s*/ is longer. The noise duration for AP-initial and AP-medial lies somewhere in between. Whole noise duration appears to be effective in signalling prosodic positions as well.

Statistical analyses were performed based on two factor ANOVAs (by fricative type and prosodic position) with the significance level of $p < 0.001$. Fricative duration, aspiration duration, whole noise duration, and following vowel duration all showed significant main effects.

4 Conclusion

In summary, in all prosodic positions, /s^h/ had a shorter fricative noise interval but a longer period of aspiration than /s*/. /s^h/ had a higher harmonic ratio at the beginning of a following vowel, suggesting breathy voice, and /s*/ had a lower harmonic ratio, suggesting pressed voice. For /s^h/ and /s*/ in prosodic positions, the difference in aspiration noise duration was larger in IP-initial position, whereas PW-medial /s^h/ was hardly aspirated at all. Difference in fricative noise dura-

tion was largest in PW-medial position, whereas IP-initial /s*/ is nearly as short as /s^h/.

In terms of “intra-segmental” and “inter-segmental” differences, the fricatives in different prosodic positions displayed characteristics that appear to signal their prosodic location by means of durational differences. From these observations, we can say that for /s*/, it is fricative noise duration that appears to be primarily responsible for both “intra-segmental” (e.g. /s*/ in each of the four prosodic positions) and “inter-segmental” (e.g. distinction of /s^h/ and /s*/ in all prosodic positions) differences whereas for /s^h/ it is aspiration noise duration and to a lesser degree fricative noise duration. In a different perspective, one can say that if a prosodic position (e.g. PW-medial position) is not favorable for a certain acoustic cue (e.g. aspiration noise duration), one of the other cues (e.g. fricative noise duration) comes into play and maintains the segmental distinction.

Although duration appears to cue different prosodic positions, we should also pay attention to subsegmental duration. Take, for example, the durations of /s^h/ in IP-initial and PW-medial positions in Figure 5. The durations of the fricative and aspiration intervals of a PW-medial segment are 69% and 34% the durations of the corresponding intervals of an IP-initial segment. Uniform stretching or shrinking of /s^h/ based on one or the other would result in distortion of the modified sound in concatenative synthesis. The situation would be the same even if we regarded the aspiration interval as part of the following vowel. The dynamic property of the fricative spectrum as reflected in the high-to-low frequency band ratio in Figure 2 is another indication that a simple durational model is inappropriate to model the dynamic variation of the turbulence energy. All of these factors suggest that a simple segment-based durational approach has a lot to overcome in modeling the effects of prosody on segments.

A recent study (Cho *et al.* 02) showed that the /s^h/ in Korean is likely to be voiced intervocalically. It reported that about 47% of the tokens had fully voiced /s^h/ and about another 40% of the tokens were voiced over more than half the frication period. In our study, only 23% of the tokens were voiced in intervocalic position over more than half the fricative noise interval. Among these, only one token had fully voiced /s^h/.

54% of the tokens were voiced for about a quarter of the frication period, and the rest of them (27%) were not voiced at all. In the context of concatenative synthesis, this implies that, at least for Korean coronal fricatives, simple durational manipulation would not be enough to model this seemingly “optional” intervocalic voicing.

This study confirms the well-known fact that any speech synthesizer, if it wishes to sound natural, should take into consideration the prosodic position each speech segment occupies before it tackles the process of actual synthesis. These findings have further significance in that they add one more case to the increasing body of evidence that prosody has segmental correlates as well as tonal correlates. If tonal correlates contribute to the parsing of prosody (Beckman 96), segmental correlates may also help listeners parse the overall prosody of an utterance.

References

- (Beckman 96) Mary Beckman. The parsing of prosody. *Language and Cognitive Processes*, 11(1):17–67, 1996.
- (Cho & Keating 01) Taehong Cho and Patricia A. Keating. Articulatory and acoustic studies of domain-initial strengthening in Korean. *Journal of Phonetics*, 29, 2001.
- (Cho *et al.* 00) Taehong Cho, Sun-Ah Jun, and Peter Ladefoged. Acoustic and aerodynamic correlates of Korean stops and fricatives. *UCLA Working Papers in Phonetics*, 99, 2000.
- (Cho *et al.* 02) Taehong Cho, Sun-Ah Jun, and Peter Ladefoged. Acoustic and aerodynamic correlates of Korean stops and fricatives. *Journal of Phonetics*, 30(2):193–228, 2002.
- (Jun & Oh 96) Sun-Ah Jun and Mira Oh. A prosodic analysis of three types of wh-phrases in Korean. *Language and Speech*, 39(1):37–61, 1996.
- (Jun 93) Sun-Ah Jun. *The Phonetics and Phonology of Korean Prosody*. PhD thesis, The Ohio State University, 1993.
- (Jun 98) Sun-Ah Jun. The accentual phrase in the Korean prosodic hierarchy. *Phonology*, 15(2):189–226, 1998.
- (Jun *et al.* 97) Sun-Ah Jun, Mary Beckman, S. Niimi, and M. Tiede. Electromyographic evidence for a gestural-overlap analysis of vowel devoicing in Korea. *Journal of Speech Sciences*, 1:153–200, 1997.
- (Kim 01) Sahyang Kim. The interaction between prosodic domain and segmental properties: domain initial strengthening of fricatives and post obstruent tensing rule in Korean. MA thesis, UCLA, 2001.
- (Oh & Johnson 97) Mira Oh and Keith Johnson. A phonetic study of Korean intervocalic laryngeal consonants. *Journal of Speech Sciences*, 1:83–102, 1997.
- (Pierrehumbert & Talkin 92) Janet Pierrehumbert and David Talkin. Lenition of /h/ and glottal stop. *Papers in Laboratory Phonology II*, pages 90–116, 1992.
- (Smith 97) C.L. Smith. The devoicing of /z/ in American English: effects of local and prosodic context. *Journal of Phonetics*, 25:471–500, 1997.
- (Yoon 99) Kyuchul Yoon. A study of Korean alveolar fricatives: An acoustic analysis, synthesis, and perception experiment. In *Mid-America Linguistics Conference Papers*, 1999.