A Production and Perception Experiment of Korean Alveolar Fricatives*

Kyuchul Yoon**

ABSTRACT

Korean has two types of voiceless alveolar fricatives: a non-tense fricative /s^h/ and a tense fricative /s'/. Twenty native speakers of Korean produced five pairs of isolated words containing word initial /s^{h}V/ and /s'V/ sequences where V was any one of five (/a, e, i, o, u/) of Korean vowels. Acoustic measures such as duration, fricative noise prominent frequency, energy change of following vowel, and fundamental frequency at vowel onset were examined. Results showed that among the parameters, aspiration noise duration of /s'/ in mid and low vowel contexts was less than 21 ms. In a perception experiment, where only the aspiration noise interval of the /s^h/ tokens was incrementally reduced, some listeners shifted perception from /s^h/ to /s'/.

Keywords: Korean, Fricative, Aspiration, Perception

1. Introduction

Korean has two types of voiceless alveolar fricatives. The slightly glottalized one /s'/ will be called tense and the non-tense one /s^h/ will be called aspirated. Park (1999) looked at the perceptual role of the aspiration noise segment following Korean alveolar fricatives by incrementally removing the aspiration and the beginning of the following vowel and found out that approximately 25 ms of aspiration along with the onset of following vowel was necessary for the perception

---

* I'd like to thank Frances Ingemann, Michael M.T. Henderson, and Clifton Pye for their advice and support for this work. I'd also like to thank Joong-Sun Sohn for his help and my Korean friends who participated in the recording.

** Department of Linguistics, University of Kansas, Lawrence, Kansas, USA

of /s^h/. Although many researchers have examined Korean alveolar fricatives acoustically and perceptually, few of them explicitly looked at the role of aspiration noise period.

The goal of this study was to examine acoustic parameters such as duration, prominent frequency of the fricative noise, energy change of following vowel, and fundamental frequency at vowel onset and see what parameters contribute to /s^h/ versus /s'/ contrast in isolated utterances. This does not imply that any one of the parameters contributes more, or less, to the overall realization of the fricatives. Rather, it implies that speech segments are never produced out of their acoustic context. Therefore, by looking at the acoustic parameters in their segmental contexts, we will be able to better understand the speech segments. Based on the results of the production experiment, a perception experiment was performed and the perceptual contribution of the aspiration noise segment was assessed.

2. Production Experiment

2.1 Method

This section examines the acoustic differences between the two Korean fricatives /s^h/ and /s'/ embedded word initially in isolated words.

Twenty male native speakers of Korean, all with normal speech and hearing, were selected for this study. The speakers were either graduate or undergraduate students at the University of Kansas. The mean age was 29.2 years with a standard deviation of 5.6 years. Nine spoke the Seoul dialect, six the Kyungsang dialect, four the Chungcheong dialect, and one the Cholla dialect. Although dialectal differences exist in pronouncing aspirated and tense alveolar fricatives in normal conversation, no noticeable dialectal differences were observed by the author during the recording sessions, perhaps because the speakers were asked to be careful and conscious of their pronunciation. All the speakers were able to pronounce both types of fricatives with equal fluency.

The test words were five minimal pairs in which each pair had alveolar fricatives before different vowels: /a, e, i, o, u/. These words were embedded and presented in Korean orthography in a randomized list of seventy-two words containing Korean alveolar fricatives in various vowel contexts

/s^hata/ – /s'ata/	: *buy – cheap*
/s^heta/ – /s'eta/	: *count – strong*
/s^hi/ – /s'i/	: *poem – seed*
/s^hota/ – /s'ota/	: *soda (a drink) – stung*
/s^huta/ – /s'uta/	: *chitchat – make (a mess)*

The utterances were recorded on a Marantz® PMD 201 tape-recorder using Maxwell® type II cassette tape and a Sony® F500S Dynamic Microphone. The recordings were done in a quiet room. 200 tokens (20 speakers × 10 tokens) were digitized at a sampling rate of 22 kHz using Multi-Speech. Six tokens were not measured because the recordings were poor. The tokens that had to be discarded were /s^hi/ and /s'i/ from Speakers 5 and 14 and /s^huda/ and /s'uta/ from Speaker 6. The total number of tokens that were analyzed, therefore, is 194.

For each token, the initial syllable was segmented into three parts: fricative noise, aspiration noise, and vowel. For fricative and aspiration noise segments, measures of duration, energy, and spectral peaks were measured. For the vowel segment, measures of duration, formant frequency, and energy change and fundamental frequency at voicing onset were determined.

For duration measurements, standard criteria were used. Determining the start of aspiration noise segment was more difficult. Usually the fricative noise segment blends into the aspiration noise segment. Since drawing the dividing line between the fricative and aspiration noise was done manually, it is possible that measurement errors were made. However, in most cases the energy distribution of the fricative and aspiration noise was distinct enough for the dividing line to be drawn manually. This can be improved in the future by having a computer script do the dividing on the basis of energy distribution, thus making the measurement errors as consistent as possible. Figure 1 illustrates the measurement procedure, where vertical lines mark the segmentation into fricative, aspiration, and vowel.

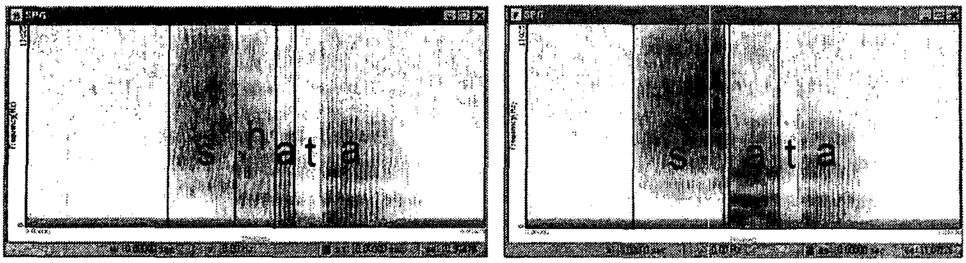

Figure 1. Spectrogram of /s^hata/ *buy* (upper panel) and /s'ata/ *wrap* (lower panel) showing the location of measurement points for the fricative, aspiration noise, and vowel.

For spectrographic analysis, a 323 Hz wide-band filter with a Blackman window was used with a display range from 0 to 11 kHz and a pre-emphasis factor of 0.8. 10 ms of frame length and frame advance were selected for energy contour analysis. An energy contour analysis was performed for each segment by obtaining the root mean square and standard deviation for an entire segment. For the energy change at the beginning of the vowel, 10 ms of frame length and frame advance were selected and the energy was measured at voice onset, 10 ms, and 20 ms after the voice onset. When the frame length was longer than the target segment, it was shortened to cover only the target segment. For measuring fundamental frequency, a spectrogram with a filter of 31.54 Hz bandwidth was produced, displaying from 0 to 1 kHz. In the spectrogram window, the fifth harmonic was counted and measured at voicing onset.

2.2 Results

2.2.1 Duration

The duration of each segments averaged across twenty speakers is given in Figure 2. A negative value in the aspiration noise segment means that voicing starts even before the frication ends. Fricative noise duration of /s'/ was generally longer than /s^h/ except for tokens containing /s^hi/ and /s'i/. The analysis of variance indicated that the effect of fricative type on the fricative noise duration was statistically significant ($F = 44.4$; $df = 1, 17$; $p < .05$). However, averaged data is misleading because individual data showed many exceptions.

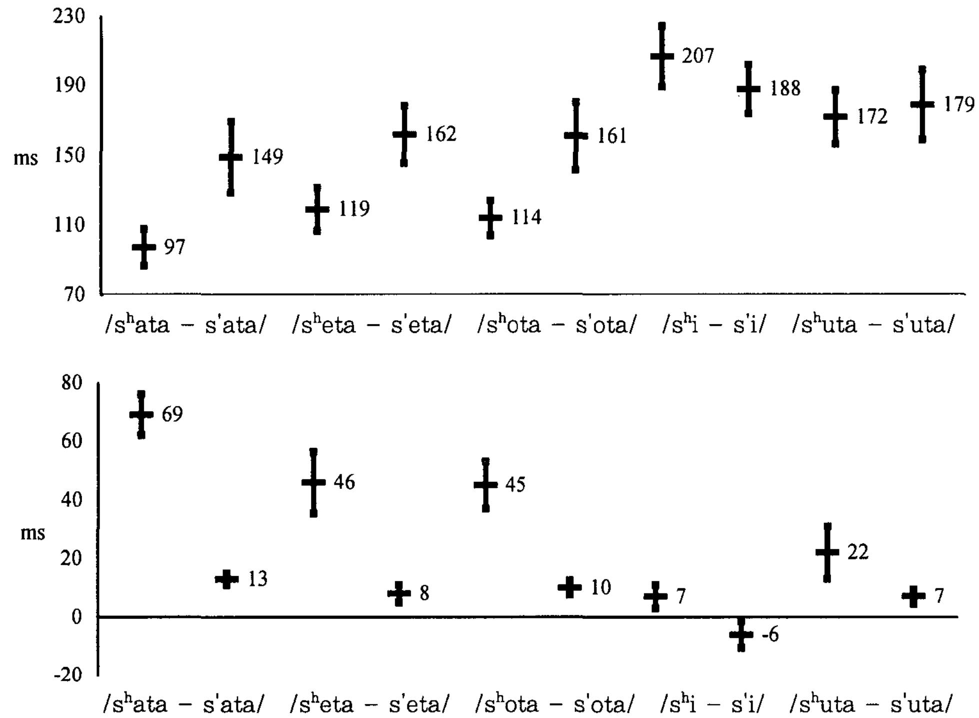

Figure 2. Average duration of fricative (upper panel) and aspiration noise segment (lower panel). Horizontal lines indicate average values, and vertical lines standard deviations.

As was seen in Figure 2, the difference of aspiration noise duration for all word pairs was the greatest for the word pairs containing low and mid vowels: /s^hata – s'ata/, /s^heta – s'eta/ and /s^hota – s'ota/. There was no exception to this pattern for these three word pairs in the individual data. In all vowel contexts, the aspiration noise duration of /s'/ was never more than 21 ms compared with /s^h/. This agrees with the findings by Park (1999) that less than 20 ms of aspiration with the fricative noise segment caused listeners to perceive /s^h/ in the vowel /a/ environment. The analysis of variance also indicated that the differences of aspiration noise duration were statistically significant ($F = 262.6$; $df = 1, 16$; $p < .05$). From the fact that there was no exception to this pattern, the aspiration noise duration could be seen as one of the important parameters contributing to the /s^h/ versus /s'/ contrast.

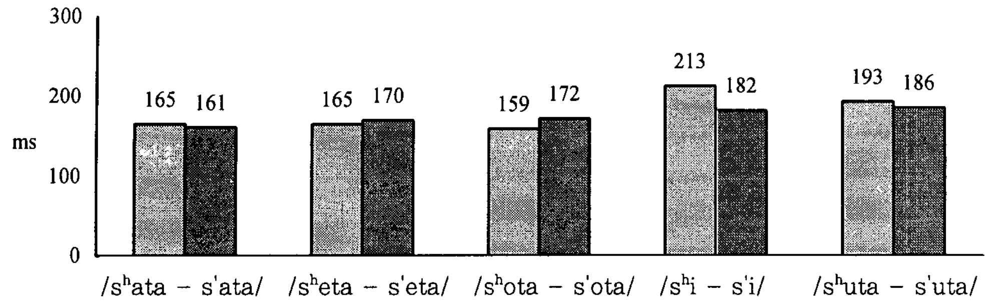

Figure 3. Average duration of frication plus aspiration.

Figure 3 shows the average duration of frication plus aspiration. The analysis of variance indicated that the differences of the duration between the two types of frricatives were not significant ($F = 1.6$; $df = 1, 17$; $p > .05$).

Vowel duration was usually longer for vowels following /s'/, but individual variation was too great to draw any consistent pattern (See Figure 4). The analysis of variance indicated that the differences of vowel duration between aspirated and tense fricatives were statistically significant ($F = 9.1$; $df = 1, 16$; $p < .05$).

One of the reasons for the much longer duration of /i/ is that while other vowels were in the first syllable of a two-syllable word, /i/ was in a monosyllabic word and as a result had final lengthening.

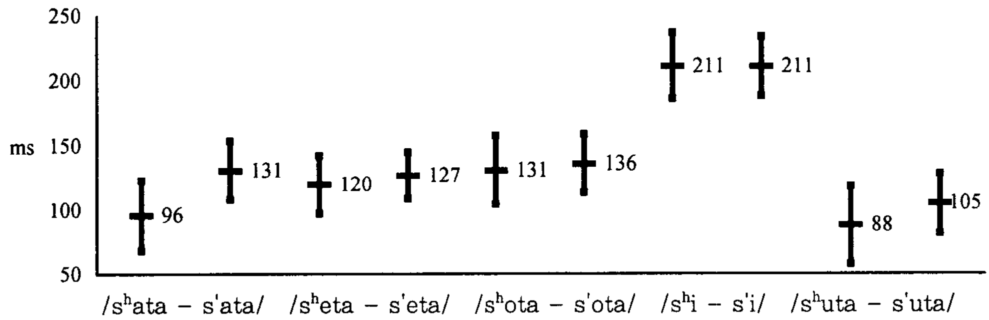

Figure 4. Average duration of vowel segment. Horizontal lines indicate average values, and vertical lines standard deviations.

2.2.2 Energy

The values in Figure 5 represent the energy level of the fricative noise segment relative to the average RMS energy of the following vowel taken as a whole. Since absolute dB values are affected by such non-linguistic factors as the loudness of the speaker's voice, distance from the microphone, and recording settings, relative values within each token are more useful than absolute values. The 0 dB line in Figure 5 represents the reference vowel energy. The energy of the fricative noise segment of the words with /s'/ is usually lower than that of /s^h/ with reference to the vowel energy. Conversely, it could be stated that vowels following /s'/ have greater energy than vowels following /s^h/ provided that the frication energy is the same. The analysis of variance indicated that the differences of relative frication energy between /s^h/ and /s'/ were statistically significant ($F = 12.1$; $df = 1, 16$; $p < .05$). However, individual data show many exceptions to this pattern. Therefore, the energy of the fricative noise segment is not an invariant parameter to aspirated/tense distinction.

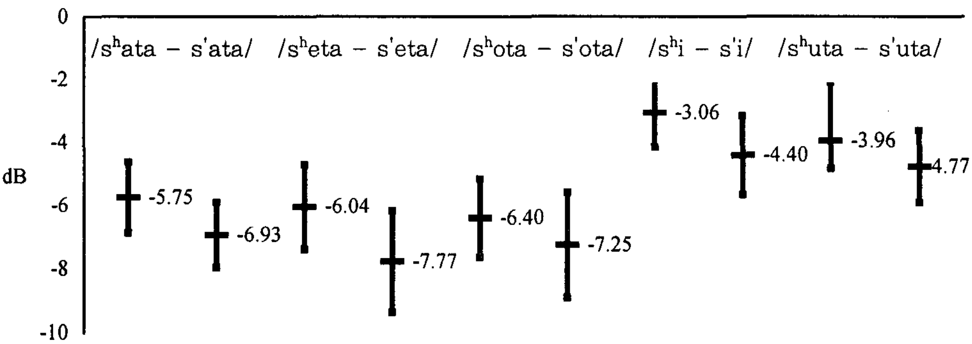

Figure 5. Average energy difference between frication and vowel segments relative to vowel energy. The values represent E(F)–E(V), where E(F) is energy of the frication and E(V) energy of the vowel. Horizontal lines indicate average values, and vertical lines standard deviations

The values in Figure 6 represent the energy level of the aspiration noise segment relative to the average RMS energy of the following vowel taken as a whole. In general, the difference in energy between aspiration and vowels is greater for the non-tense fricatives than for the tense fricatives. However, the analysis of variance indicated that the differences of relative aspiration energy between the two types of fricatives were not statistically significant ($F = 0.6$; $df = 1, 17$; $p > .05$).

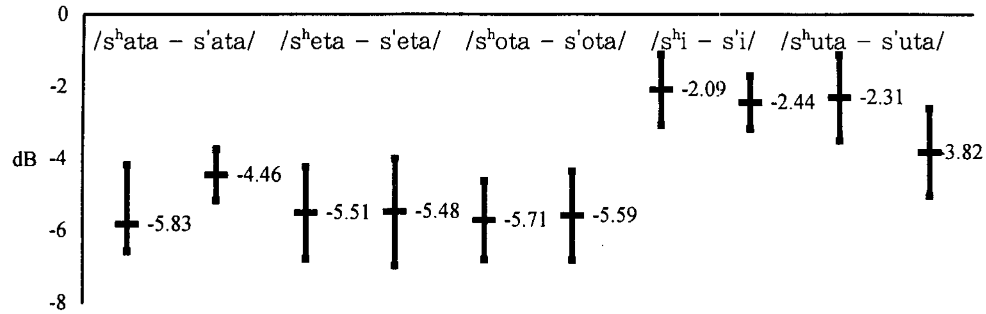

Figure 6. Average energy difference between aspiration and vowel segments relative to vowel energy. The values represent E(A)−E(V), where E(A) is energy of aspiration and E(V) energy of vowel. Horizontal lines indicate average values, and vertical lines standard deviations.

2.2.3 Prominent frequency of fricative noise

Prominent frequency will be defined as the frequency of the spectral peak with the highest energy. The average prominent frequencies of the fricative noise segment is given in Figure 7.

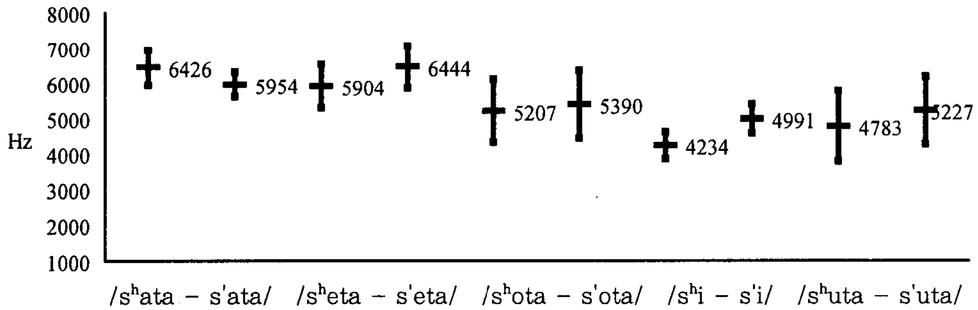

Figure 7. Prominent frequency of fricative noise segment of the words with both types of fricatives. Prominent frequency represents the spectral peak with the greatest energy. Horizontal lines indicate average values, and vertical lines standard deviations.

Although the prominent frequency was usually higher for /s'/ except for the first word pair /sʰata – s'ata/, it was not necessarily true in the individual data. Since the individual data did not show any consistent pattern, it was not possible to determine any difference in the prominent frequency of the fricative noise segment for both fricatives.

It seems that the prominent frequency does not play any significant role in the aspirated/tense distinction. Statistical analysis indicated that the prominent frequency differences between aspirated and tense fricatives were not significant ($F = 0.9$; $df = 1, 17$; $p > .05$).

2.2.4 Energy change of following vowel

The average energy change relative to overall vowel energy (seen as 0 dB) in the first 20 ms of the vowel following the fricatives is given below in Figure 8. The result agrees with what Han and Weitzman (1970) found in their study on Korean stop consonants. They found that the tense stops have a rapid build-up of energy in the onset of the following vowel.

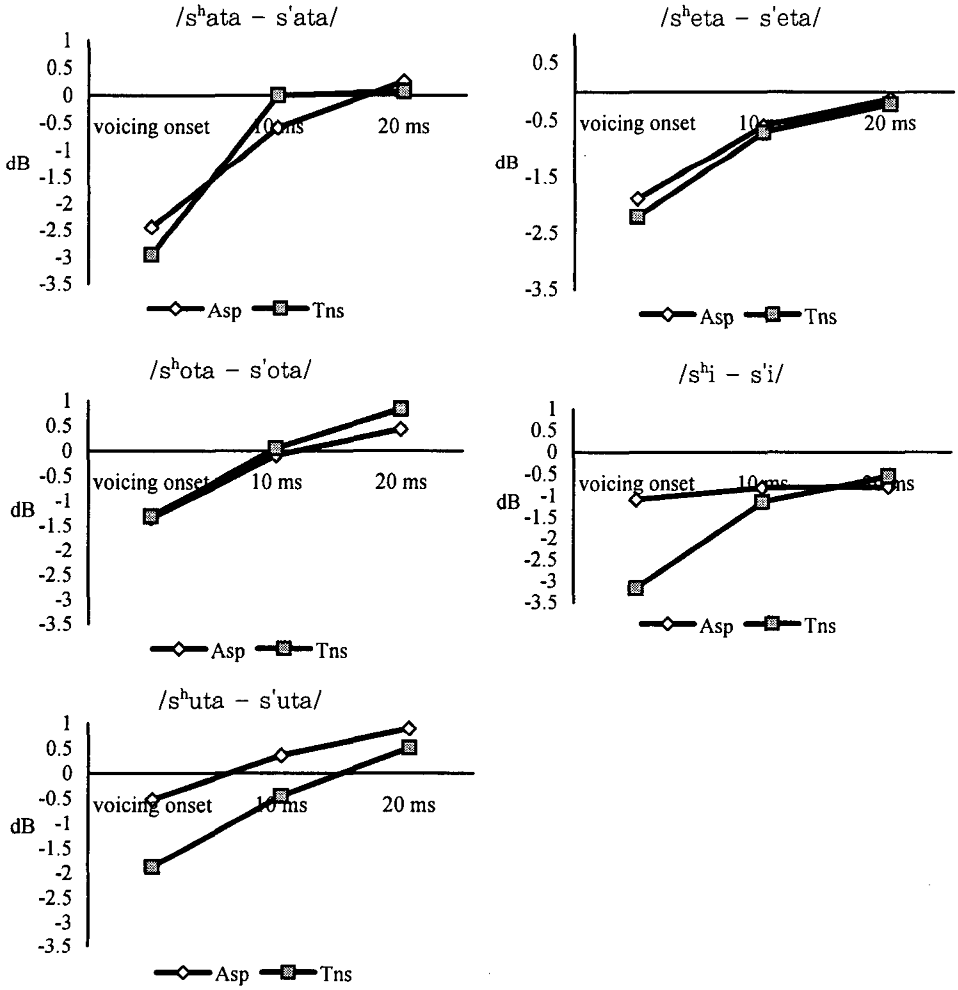

Figure 8. Average energy change in the vowel segment of each word pair. Asp and

Tns stand for the energy change of the words with an aspirated and a tense fricative respectively. The 0 dB line represents the reference average vowel energy.

To a greater or lesser degree, the words with /s'/ show greater rate of energy build-up in the first 20 ms of the following vowel. This is true for both the first 10 ms and 20 ms interval provided that both intervals start from the voicing onset.

However, for all vowel contexts, there are many counterexamples to the general pattern found in Figure 8. Statistical analysis also indicated that the differences in energy change between the two types of fricatives were not significant (F = 0.6; df = 1, 15; p 〉 .05). Overall, although the energy build-up parameter following the voicing onset up to the point of about 20 ms was generally rapid for the words with tense fricatives, it was not always true.

2.2.5 Fundamental frequency at vowel onset

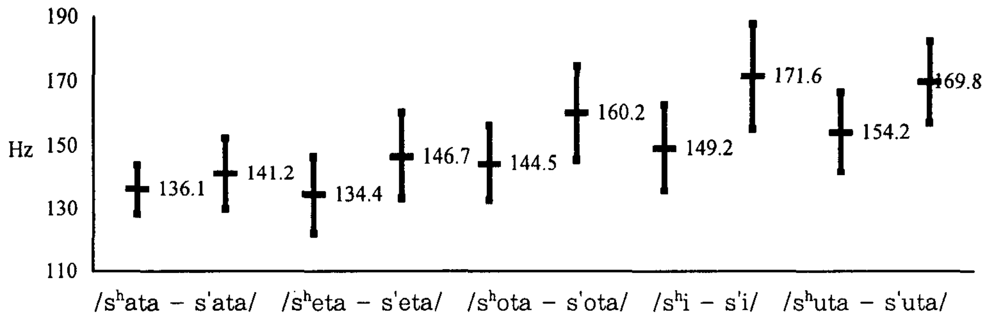

Figure 9. Fundamental frequency at voicing onset for the words with both types of fricatives. Horizontal lines indicate average values, and vertical lines standard deviations.

As Figure 9 shows, the fundamental frequency at voicing onset of the following vowel is generally higher for /s'/, although, as the standard deviation indicates, the variation is too great. The analysis of variance indicated that the differences of the fundamental frequency at voice onset between the two types of fricatives were statistically significant (F = 13.1; df = 1, 16; p 〈 .05). These findings agree with those of Kim (1965), Han and Weitzman (1970), and Hardcastle (1973). They suggested that the fundamental frequency of the vowel following tense stops is higher than that of the vowel following the corresponding lax stops. As was the case with the energy change above, however, counterexamples are found for all word pairs, indicating that the fundamental frequency parameter is not invariant.

Of all the parameters we looked at, only the duration of the aspiration noise segment seemed to be invariant. The aspiration noise segment was shorter in duration by 21 ms for /s'/ in low and mid vowel contexts for all speakers.

3. Perception Experiment

3.1 Method

Since the production study indicated that the aspiration noise duration was the most reliable parameter, the author and another Korean speaker listened to all twenty tokens of /s^hata/ twice, first time with intact aspiration interval and the second time after the removal of the aspiration noise. For some of the tokens, perception shifted from /s^hata/ to /s'ata/.

Utterances from four of the speakers whose tokens caused perception shift and utterances from four other speakers whose recordings did not cause perception shift were selected for the perception experiment. The aspiration noise segment was removed in 10 ms increments starting at the center to avoid removing any acoustically significant information that might be found near the fricative noise and the vowel onset.

The prepared stimuli were duplicated eight times and presented to listeners in random order for identification.

In addition, a synthesized CV syllable /sa/ (using Klatt Synthesizer) with and without aspiration was also duplicated eight times and presented to listeners for identification.

Twenty native speakers of Korean (eleven males and nine females), all with normal speech and hearing, served as listeners. All were either graduate or undergraduate students at the University of Kansas. The mean age and standard deviation were 28.5 and 5.7 years.

Session 1 – 4: The tokens of the first four sessions were prepared by removing sections in 10 ms increments from the center of the aspiration. Depending on the aspiration noise duration in the original utterance, 5 to 10 tokens with different durations were prepared for further duplication. These tokens were duplicated 8 times and evenly divided into four sessions. Natural /s^hata/ and /s'ata/ utterances from the eight speakers (2 x 8 = 16) were added to each of the sessions in order to compare judgments of unaltered speech with those from which aspiration had been removed. With 16 natural utterances included, each session had 144 tokens presented in random order. The total number of items to be identified in the four sessions was 576 (144 x 4 = 576).

Session 5: The syllable /s^{h}a/ was synthesized in three segments: fricative, aspiration noise, and vowel segment based on measurements taken from the production part of this study. The segments were 150 ms, 70 ms. and 200 ms long respectively. Some modifications of measured values were made to make the syllable sound more natural or to accommodate to the limitations of the synthesizer For the values used for parameter settings, refer to Yoon (1999).

The syllable was synthesized with and without the 70 ms aspiration segment. Wave forms and spectrograms of the two synthesized stimuli are presented in Figure 10.

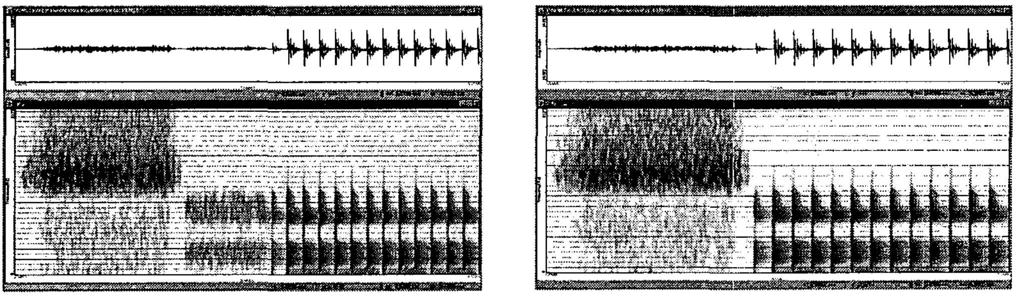

Figure 10. Spectrograms of a synthetic /s^{h}a/ created with Klatt Synthesizer (left panel), and /sa/ with its aspiration removed (right panel).

The perception experiment for the synthesized syllable was not focused on where the shift in perception starts, but on whether the shift in perception to its tense counterpart occurs at all. The syllable with and without aspiration was duplicated eight times, resulting in a total of 16 tokens. Only synthetic stimuli were presented in a session given after the four sessions of modified natural tokens. The total number of tokens from spoken and synthetic stimuli was 592 (144 x 4 + 16 = 592).

The 592 tokens described above were presented to the twenty native speakers of Korean for identification. Each listener was tested individually using the Auditory Stimulus Preparation and Presentation software from Kay Elemetrics. Each listener sat in front of a computer, was given instructions, and listened to the tokens that came from a speaker hooked up to the CSL Model 4300B from Kay Elemetrics Co.

3.2 Results

3.2.1 Natural stimuli

As expected, reduction in the aspiration noise duration shifted perception from /s^h/ to /s'/. A beginning point of perception shift will be defined as the place where the majority of the responses start to show fewer than 100% [s^hata] responses and a complete perception shift is defined as the place where most listeners give fewer than 25% [s^hata] responses.

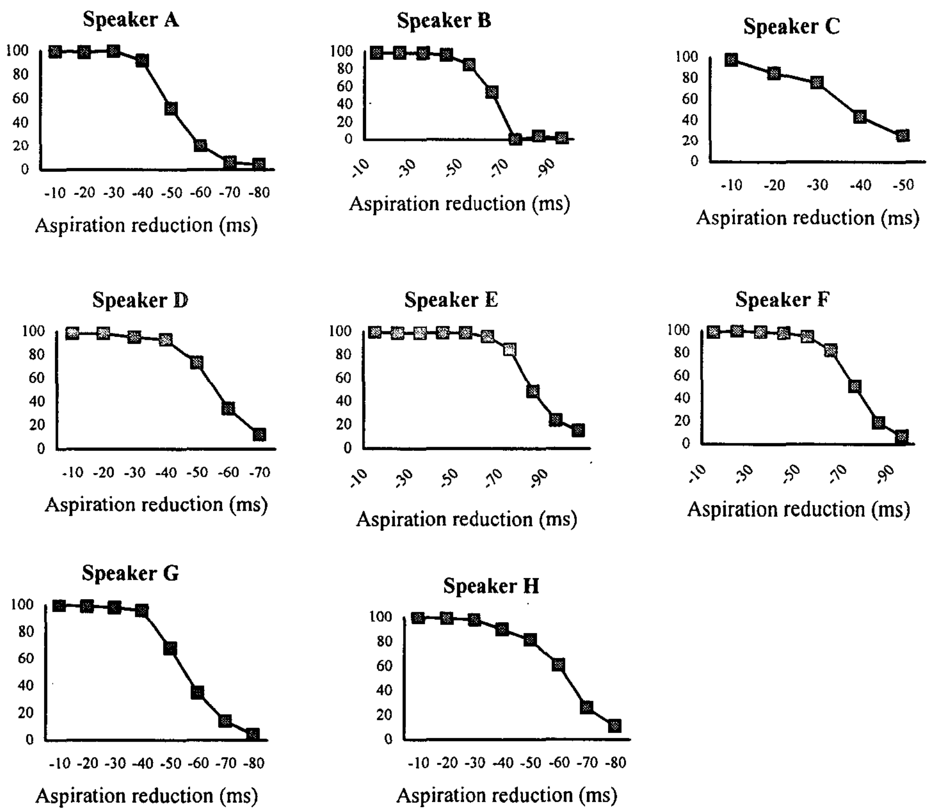

Figure 11. Average percent /s^ha/ scores (vertical axis) of all listeners for each speaker.

All listeners correctly identified all 16 presentations of the natual utterances. Most listeners reported a complete perception shift when the aspiration reduction was greater than 70 ms. Some listeners required less aspiration noise for perception shift despite their longer aspiration duration in their actual production. This implies that there might be some other parameters, besides aspiration noise duration, that influence perception shift.

Figure 11 shows that, depending on the speakers, an aspiration noise duration of 11 to 37 ms or less in the manipulated natural utterances caused perception to start shifting from aspirated to tense. In other words, less than 37 ms of aspiration causes listeners to judge the tokens as belonging to the tense category.

3.2.2 Synthetic stimuli

Responses to synthetic stimuli in session 5 are given in Figure 12.

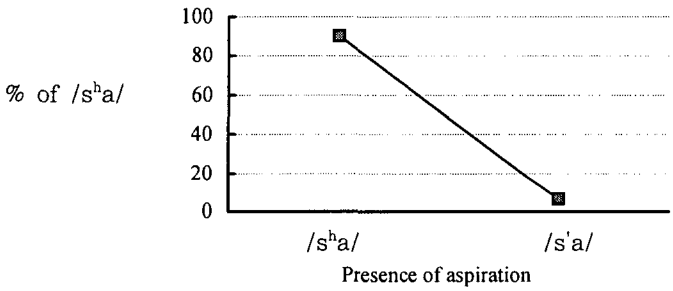

Figure 12. Average responses for synthetic /sa/ with or without aspiration.

Depending on the presence of the aspiration, listeners' responses fell into two groups. When the tokens had aspiration, they were perceived as containing non-tense fricatives, but when they did not, they were reported as containing tense fricatives.

4. Summary and Conclusion

The production part of this study showed that only the duration of the aspiration noise segment in mid and low vowel contexts was an invariant acoustic cue to the aspirated/tense distinction of the voiceless Korean alveolar fricatives. This was true not only in terms of overall statistical significance, but also in terms of every instance of individual measurements. In general, the aspiration noise duration of the words with tense fricatives was much shorter (less than 21 ms) compared to that of the words with non-tense fricatives.

In perception experiment, when presented with either synthetic or natural tokens of varying aspiration noise duration, most listeners experienced perception shift from aspirated to tense when aspiration was either completely absent or less than 37 ms. However, the perception experiment has a limitation in that it was performed only with the low vowel /a/.

A comparison between the amount of aspiration in the perception test and the actual aspiration from the natural utterances leads us to suspect that other parameters or synergistic effect of other parameters may exist. It was sometimes the case that perception did not shift from aspirated to tense even when the same amount of aspiration as for the tense fricative was present. In cases such as this, more aspiration reduction was necessary for the expected perception shift, indicating that the aspiration noise duration was not the only parameter for the aspirated/tense distinction. This study is limited in the sense that the fricatives were embedded as word initial segments in isolated words. The role of aspiration duration along with other acoustic parameters is being studied with the fricatives embedded in utterances controlled for prosodic positions.

References

Han, M. S. & R. S. Weitzman. 1970. "Acoustic Features of Korean /P, T, K/, /p, t, k/, and /p^h, t^h, k^h/." *Phonetica,* 22, 112-28.

Hardcastle, W. J. 1973. "Some Observations on the Tense-Lax Distinction in Initial Stops in Korean." *Journal of Phonetics,* 1, 263-72.

Kim, Chin-woo. 1965. "On the Autonomy of the Tensity Feature in Stop Classification (with Special Reference to Korean Stops)." *Word,* 21, 339-59.

Park, Hansang. 1999. "The Phonetic Nature of the Phonological Contrast between the Lenis and Fortis Fricatives in Korean." In *Proceedings of the XIVth International Congress of Phonetic Science*, 425-428.

Yoon, Kyuchul. 1999. *A Study of Korean Alveolar Fricatives: An Acoustic Analysis, Synthesis, and Perception Experiment.* MA Thesis. University of Kansas.

Received: July 14, 2002.
Accepted: August 24, 2002.

▲ Kyuchul Yoon
2048 Park Run Dr. #E
Columbus OH 43220
Tel: +1-614-442-1960
E-mail: kyoon@ling.osu.edu