Trace of F2 Peaks as a Quantitative Descriptor of Aspiration

HANSANG PARK
The University of Texas at Austin

0. Introduction

Phonation types of obstruents have been qualitatively, or phonologically, described in terms of voicing (voiced vs. voiceless), aspiration (aspirated vs. unaspirated), and laryngealization (laryngealized vs. unlaryngealized). Phonation types of obstruents have also been quantitatively described since Lisker and Abramson (1964) described phonation types of stops in terms of voice onset time (VOT). Although VOT was accepted as the single best metric for the phonation type of stops, and a huge amount of research has been conducted in terms of VOT for stops of various languages, affricates and fricatives have seldom been described in terms of VOT. Phonation types of fricatives may not have been investigated because the variation of phonation types is minor in fricatives. However, aspirated alveolar fricatives do occur in Burmese, Karen, Mazahua (Maddieson 1984). Korean has a laryngealized alveolar fricative as well as an aspirated alveolar fricative. Affricates may not have been investigated precisely because they show almost the same pattern of phonation types as stops across languages. However, it would be more desirable to investigate phonation types of all the classes of obstruents in a unified way. On the other hand, VOT does not provide any information about what happens in the phase of consonantal articulation showing different VOT values, since VOT is only a temporal descriptor of phonation types. It may be claimed that obstruents have been sufficiently classified in terms of VOT. However, the traditional qualitative method of classification in terms of voicing, aspiration, and laryngealization is useful only for taxonomic purposes. In this sense, we need to seek a method of quantitative description of phonation types in spectral domain as well as time domain.

The present study aims at a quantitative description of phonation types of fricatives in the spectral domain. Aspiration can be best described in the spectral domain, since we can see most important characteristics of aspiration in this domain. Aspiration is acoustically characterized by the absence of a voice bar, a severe reduction of energy in the lower frequencies, and a formant transition in the higher frequencies. The absence of voice bar represents the absence of vocal fold vibration. A severe reduction of energy in the lower frequencies results from coupling of the supraglottal and subglottal cavities. A formant transition in the higher frequencies is considered to be an acoustic correlate of the frication noise generated in the glottis. This study attempts to quantitatively describe a characteristic of aspiration occurring in the formant transition, assuming that the trace of the amplitude of F2 peaks may serve as an indicator of aspiration. This study also tries to quantify spectral characteristics of phonation types by calculating the difference in amplitude of F2 peaks between different time points. These attempts are useful in that they provide a nonlinear description of phonation types, that is, both temporal and spectral description of phonation types, unlike VOT, which provides only temporal information.

1. Method

1.1 Subject

Four subjects, two for each language, were recruited for the present study. All the subjects were native speakers of each language. English subject LH is a male in his mid thirties from Washington DC and ST a male in his mid twenties from Austin, Texas. Dialectal difference was disregarded, since no significant difference was observed with respect to phonation type across dialects. Both Korean subjects, SL and HS, were males in their mid twenties from Standard Korean area. They were born to parents also speaking Standard Korean. It had been less than 1 year since they had come to the United States. None of the subjects had a history of any speech disorder at the time of recording.

1.2 Material

The tokens used in the experiment were Korean and English alveolar fricatives preceded and followed by a low vowel /a/, that is, /as^{h}a/ and /as*a/ for Korean and /asa/ and /aza/ for English. Korean fricatives contrast in aspiration, not in voicing, whereas English fricatives contrast in voicing, not in aspiration. The Korean tokens were real words, while the English tokens nonsense words. Tokens were produced in isolated forms. Each token was repeated 10 times, such that the total number of the recorded materials was 80: 2 languages x 2 speakers x 2 fricatives x 10 repetitions.

1.3 Recording and Measurement

Subjects were asked to read the materials. Before recording, English subjects were asked where stress would fall in VCV tokens. They agreed that stress would fall on the second syllable. The recording was conducted in Phonetics Lab in the Department of Linguistics at the University of Texas at Austin. Subjects' voices were recorded with a Marantz Superscope tape recorder. Sound files were created on Soundscope at a sampling rate of 22,050 Hz.

Duration of the segments, F0, and amplitude of F2 peaks were measured for each token. The duration of the preceding vowel, the fricative, and the subsequent vowel were measured. Proportion of each segment as well as the duration of the whole token was calculated. Four points were measured to obtain duration of the tokens: the onsets and offsets of the preceding and subsequent vowels. Duration of the preceding vowel, the fricative, and the following vowel was calculated by subtracting each measured time values. The onset of the preceding vowel and the offset of the following vowel were determined by locating the first and last time points showing the first and last pitch track in F0 plot created by autocorrelation method with a frame advance of 5 ms. As a result, a series of ordered pairs of <time, F0> was obtained every 5 ms. The time value in the F0 plot was used to determine the offset of the preceding vowel and the onset of the following vowel. The offset of the preceding vowel and the onset of the following vowel were determined by selecting the time value, among the time values obtained in the method mentioned above, which was closest to the one showing a striking difference in the waveform shape, referring to voice onset. In case of voiced fricative, only the waveform shape was consulted.

F0 was measured in 4 different time points: the onsets and offsets of the preceding and subsequent vowels. F0 values of the onset of the preceding vowel and the offset of the subsequent vowel were measured at the time point showing the first and last pitch track in the F0 plot. F0 values of the offset of the preceding vowel and the onset of the subsequent vowel were measured respectively at the time points of 10 ms before the offset of the preceding vowel and 10 ms after the onset of the subsequent vowel in the F0 plot.

Amplitude of F2 peaks were measured in 20 ms Hamming windows on Narrowband FFT at an interval of 5 ms. The center of the window was set to the time point in the F0 plot. Narrowband FFT was created in Sound Scope with a FFT point of 1024 and a filter of 45 Hz. Neither 6dB preemphasis nor Smoothing was selected. Amplitude of F2 peaks was measured from 100 ms before the offset of the preceding vowel to 100 ms after the onset of the following vowel. The highest peak between 1,000 and 2,000 Hz was considered to be the F2 peak. The range covers the range of the F2 formant transition of vowel /a/ as well as the range mentioned in Soli (1981) who has reported that the F2 peak occurred between 1,500 and 2,000 Hz in frication noise and formant transition.

2. Results

2.1 Duration

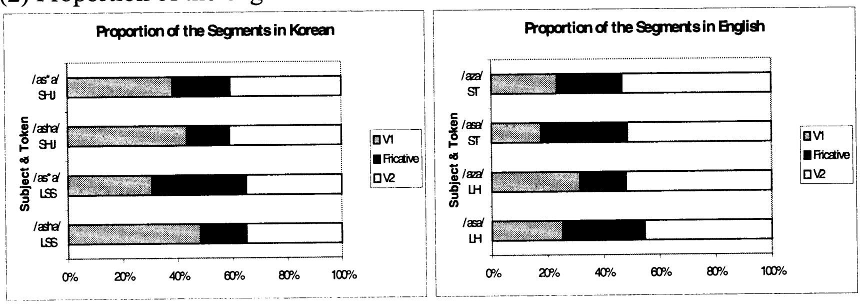
In this section, the temporal structure of the tokens will be provided. A one-way ANOVA was conducted to test for significant differences across tokens and subjects in each language. Level of significance was set to .05 in all statistical analyses.

The mean duration of the preceding vowel (V1), the fricative, the subsequent vowel (V2), and the whole token is given in (1) below. Proportion of the three segments was also calculated by dividing the duration of each segment by the duration of the whole token. The proportion of the segments is illustrated in (2) below.

(1) Mean Duration of the Whole Token and Its Segments

			V1		Fricative		V2		Total	
			ms	%	ms	%	ms	%	ms	%
Korean	LSS	/as^{h}a/	252	48	88	17	183	35	523	100
		/as*a/	269	31	190	34	193	35	552	100
	SHJ	/as^{h}a/	255	43	92	16	241	41	588	100
		/as*a/	231	38	127	21	244	41	602	100
English	LH	/asa/	106	25	123	30	188	45	417	100
		/aza/	129	32	67	16	211	52	407	100
	ST	/asa/	104	18	182	31	300	51	586	100
		/aza/	139	24	139	23	316	53	594	100

(2) Proportion of the Segments



First, the mean duration of the whole token was compared. As can be seen in (1), /as*a/ is longer than /as^{h}a/ in both Korean subjects. /as*a/ is longer than /as^{h}a/ by about 30 ms. in LSS, while by about 15 ms in SHJ. On the other hand, SHJ showed a longer duration in both /as*a/ and /as^{h}a/ than LSS. A one-way ANOVA was conducted to test for a significant difference between tokens for each Korean subject. The dependent variable was the duration of the whole token and the independent variable was token. There was a significant difference between tokens in LSS ($F = 8.357$, $p = .010$) but no significant difference in SHJ ($F = 2.064$, $p = .168$). Another one-way ANOVA was conducted to test for a significant difference between subjects for each token. The dependent variable was the duration of the whole token and the independent variable was subject. There was a significant difference between subjects for both /as^{h}a/ ($F = 71.972$, $p = .000$) and /as*a/ ($F = 17.406$, $p = .001$). For English, /asa/ was longer than /aza/ in LH but the other way in ST. /asa/ was longer than /aza/ by about 10 ms in LH, while the reverse was true by about 8 ms in ST. On the other hand, ST showed a longer duration in both /asa/ and /aza/ than LH. A one-way ANOVA was conducted to test for a significant difference between tokens for each subject. There was no significant difference between tokens for both LH ($F = .441$, $p = .515$) and ST ($F = .177$, $p = .679$). Another one-way ANOVA was conducted to test for a significant difference between subjects for each token. There was a significant difference between subjects for both /asa/ (F = *205.338*, p = *.000*) and /aza/ (F = *78.333*, p = *.000*).

Second, the proportions of the preceding vowel, the fricative, and the subsequent vowel were compared. Unlike the duration of the whole token, it may be meaningless to compare the absolute duration across tokens and subjects. As shown in (2), the proportion of the subsequent vowel shows little difference across tokens but some difference across subjects. On the other hand, the fricative shows a substantial difference across tokens for both subjects but moderate difference for /as^{h}a/ and little for /as*a/ between subjects. Statistical analyses were conducted only for the proportion of the fricative, since our main concern lies in the proportion of the fricative. A one-way ANOVA was conducted to test for a significant difference in the proportion of the fricative for Korean. There was a significant difference in the proportion of the fricative between tokens for both LSS (F = *239.864*, p = *.000*) and SHJ (F = *29.413*, p = *000*), while no significant difference between subjects for /as^{h}a/ (F = *2.542*, p = *.128*) but a significant difference for /as*a/ (F = *99.828*, p = *.000*). A significant difference test was also conducted for English. There was a significant difference in the proportion of the fricative between tokens for both LH (F = *64.366*, p = *.000*) and ST (F = *39.718*, p = *000*), while no significant difference between subjects for /asa/ (F = *1.063*, p = *.316*) but a significant difference for /aza/ (F = *22.353*, p = *.000*).

2.2 F0

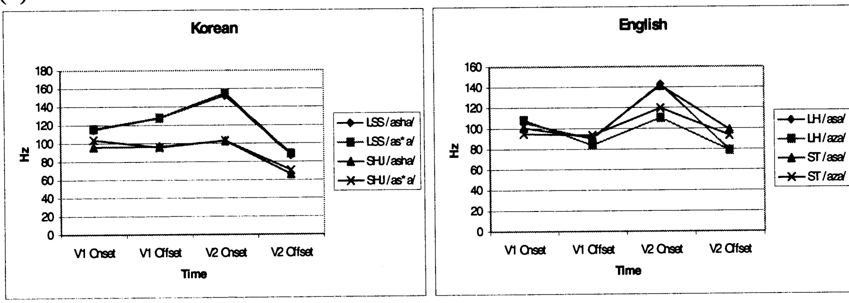
In this section, the F0 at 4 different time points: the onsets and offsets of the preceding and subsequent vowels will be provided. The mean F0 at the 4 time points is given in (3) and also illustrated in (4) below.

As can be seen in (4), the mean F0 is generally higher in the onset of the subsequent vowel than in any other time point for Korean subjects. The only exception is /as*a/ in SHJ but the difference between the V1 onset (104 Hz) and the V2 onset (103 Hz) is minor. The mean F0 values of both tokens are almost the same for each subject.

(3) Mean F0 at the 4 Time Points of the Token

			V1 Onset	V1 Offset	V2 Onset	V2 Offset
Korean	LSS	/as^{h}a/	115	128	153	87
		/as*a/	116	128	155	89
	SHJ	/as^{h}a/	96	97	103	66
		/as*a/	104	96	103	71
English	LH	/asa/	106	90	143	79
		/aza/	108	84	110	79
	ST	/asa/	101	91	142	99
		/aza/	95	94	120	94

(4) Illustration of the Mean F0 at the 4 Time Points of the Token



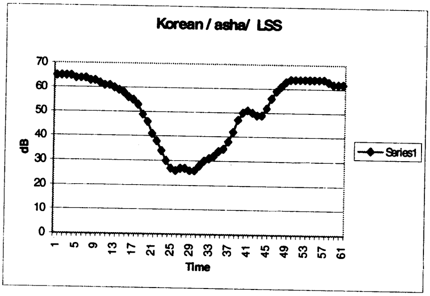
Statistical analysis was performed only for the onset of the subsequent vowel, since our main concern lies in the F0 in the onset of the subsequent vowel. A one-way ANOVA test for Korean showed that there was no significant difference between tokens for both LSS (*F = 1.035, p = .322*) and SHJ (*F = .032, p = .859*), while a significant difference between subjects for both /as^{h}a/ (*F = 757.955, p = .000*) and /as*a/ (*F = 484.365, p = .000*). A one-way ANOVA test for English showed that there was a significant difference between tokens for both LH (*F = 63.876, p = .000*) and ST (*F = 25.115, p = .000*), while no significant difference between subjects for /asa/ (*F = .067, p = .799*) but a significant difference for /aza/ (*F = 14.115, p = .001*).

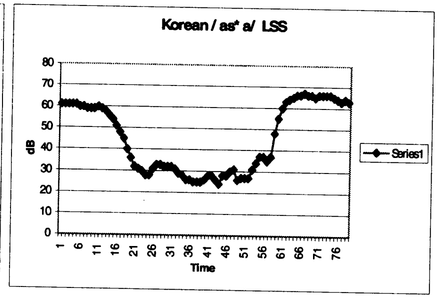
2.3 Trace of F2 peaks

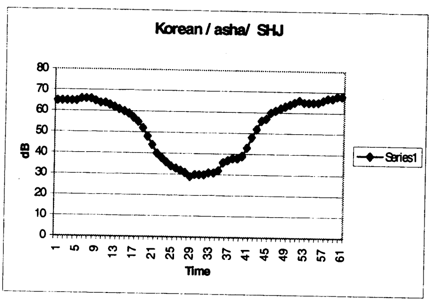
In this section, the trace of F2 peaks will be discussed. The F2 peaks of one of the tokens for each subject in each language are illustrated in (5). The figures given in (5) show the trace of F2 peaks from 100ms before the offset of the preceding vowel to 100ms after the onset of the subsequent vowel. The vowels have higher amplitude values than the fricative. There seems to be no substantial difference between the preceding vowel and the subsequent vowel in Korean, while the following vowel has much higher amplitude than the preceding vowel in English, which seems to reflect that stress falls on the second syllable. A very low amplitude at the first time point in English subject ST' /asa/ means that the preceding vowel has the onset of the vowel around that time point.

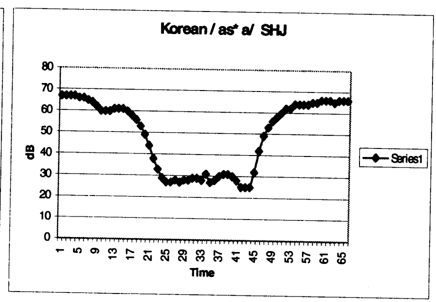
We need to focus on the transition from the preceding vowel to the fricative and from the fricative to the subsequent vowel. We can observe an abrupt decrease at the onset of the fricative and an abrupt increase at the offset of the fricative for Korean /as*a/, so that we can see a clear cut between the vowels and fricative. The overall shape can be compared a wide valley with cliffs on either side. On the other hand, we can observe a gradual decrease at the onset of the fricative and a gradual increase at the offset of the fricative for Korean /as^{h}a/, so that we cannot identify the boundary between the vowels and the fricative. The overall shape can be compared to a narrow valley with less steep slope.

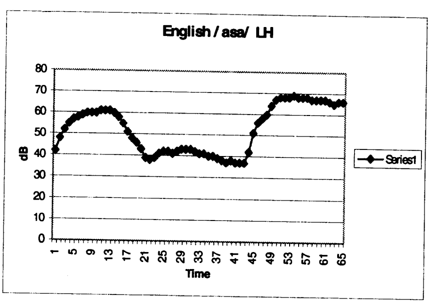
(5) Trace of F2 Peaks

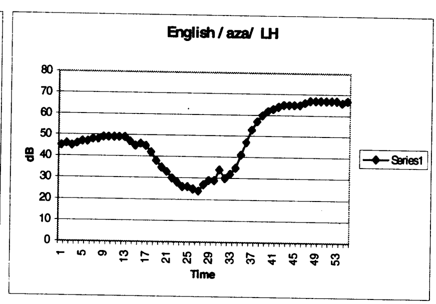


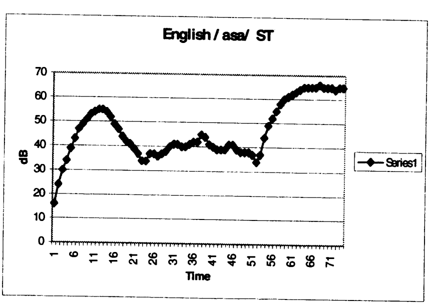


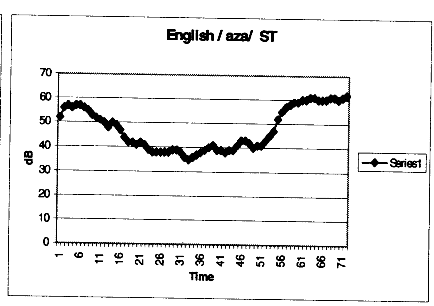












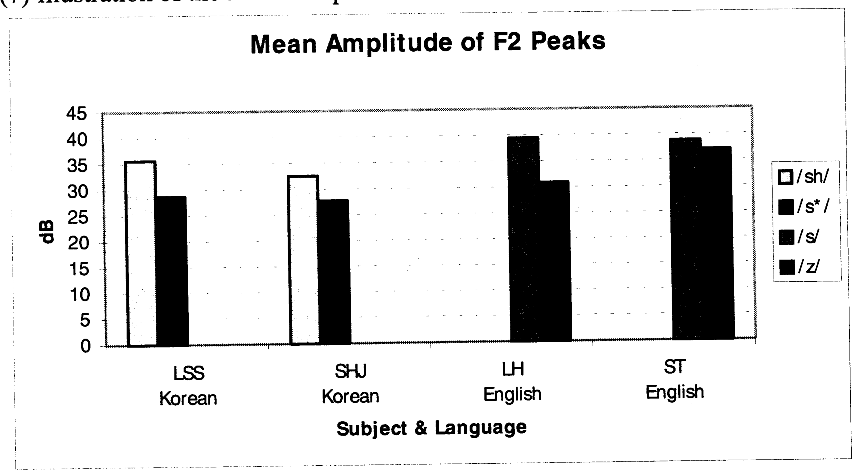
English /asa/ is similar to Korean /as*a/ in the overall pattern of the trace of F2 peaks, while English /aza/ to Korean /as^{h}a/. We can also see a substantial difference in the overall amplitude of F2 peaks of the fricative between English /asa/ and /aza/, which can be compared to elevation of the bottom of the valley.

The present study attempts to quantify the amplitude of F2 peaks (L2) of the fricative, or the elevation of the valley bottom, and the increment of L2 between two time points, that is, slope. First, mean L2 of the fricative was calculated by dividing sum of the L2 values of each window in the fricative by the number of windows. The mean amplitude of F2 peaks of the fricative is given in (6) and also illustrated in (7) below.

(6) Mean Amplitude of F2 Peaks of the Fricative

	Korean		English		Diff.
	LSS	SHJ	LH	ST	
/s^{h}/	35.7	32.5			3.2
/s*/	28.7	27.8			0.9
/s/			39.4	38.8	0.6
/z/			30.8	37.0	6.2
Diff.	7.00	4.7	8.6	1.8	

(7) Illustration of the Mean Amplitude of F2 Peaks of the Fricative



As seen in (7), the mean L2 is higher in /s^{h}/ than in /s*/ for both Korean subjects. On the other hand, it is higher in /z/ than in /s/ for both English subjects. Statistical analyses were performed to test for significant differences between fricatives for each subject of each language. A one-way ANOVA test for Korean showed that there was a significant difference between fricatives for both LSS ($F = 72.248$, $p = .000$) and SHJ ($F = 78.868$, $p = .000$), while a significant difference between subjects for /s^h/ (*F = 25.205, p = .000*) and no significant difference for /s*/ (*F = 1.575, p = .226*). A one-way ANOVA test for English showed that there was a significant difference between fricatives for both LH (*F = 103.179, p = .000*) and ST (*F = 4.719, p = .043*), while no significant difference between subjects for /s/ (*F = .750, p = .398*) but a significant difference for /z/ (*F = 44.642, p = .000*).

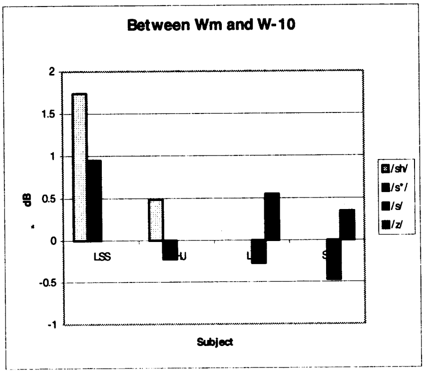
Second, two kinds of mean increment of L2 were calculated: mean increment of L2 between the window at mid-time point of the fricative (Wm) and the window at the time point of 10 ms prior to the offset of the fricative (W-10), and that of L2 between W-10 and the 8th window backward from the offset of the fricative, that is, the time point 50 ms prior to the offset of the fricative (W-50). The reason for selecting W-10 is that that time point is the center of the last window containing no or least fragment of the following vowel. The reason why increment of L2 between Wm and W-10 is calculated is simply that we can see L2 transition of the latter half of the fricative comparison. On the other hand, the reason why increment of L2 between W-50 and W-10 is calculated is that we can observe most prominent L2 transition between the two time points. The mean increments of L2 between Wm and W-10 and between W-50 and W-10 were calculated by dividing the increment of L2 between the two time points by the number of the windows between the two time points, respectively. The present study quantifies only the increment in CV transition. The same method can be applied to the decrement in VC transition.

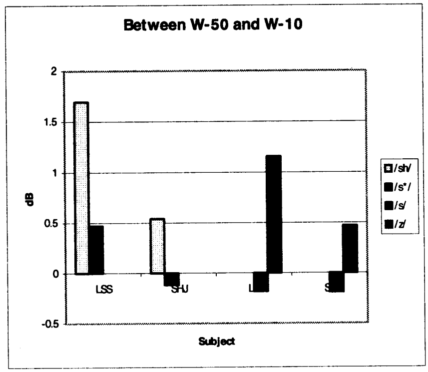
The mean increments between W-10 and Wm and between W-10 and W-50 are given in (8) and also illustrated in (9) below. First, the mean increment of L2 between Wm and W-10 was compared. For Korean, /s^h/ shows a higher value than /s*/ for both subjects. For English, /z/ shows a higher value than /s/ for both subjects. Difference in the mean increment of L2 between fricatives is almost the same in all the subjects regardless of language. Second, the mean increment of L2 between W-50 and W-10 was compared. For Korean, /s^h/ also shows a higher value than /s*/ for both subjects. For English, /z/ shows a higher value than /s/ for both subjects. Difference in the mean increment of L2 between W-50 and W-10 is not the same in all the subjects but the difference in one of the subjects is almost double that in the other in each language.

(8) Mean Increment of Amplitude of F2 Peaks of the Fricative

		Wm and W-10					W-50 and W-10				
		LSS	SHJ	LH	ST	Diff	LSS	SHJ	LH	ST	Diff
Korean	/s^h/	1.74	0.49			1.25	1.69	0.54			1.15
	/s*/	0.95	-0.23			1.18	0.47	-0.12			0.59
English	/s/			-0.28	-0.48	0.2			-0.18	-0.19	0.01
	/z/			0.55	0.35	0.2			1.15	0.47	0.68
Diff		0.79	0.72	0.83	0.83		1.22	0.66	1.33	0.66	

(9) Illustration of the Mean Increment of Amplitude of F2 Peaks of the Fricative





Statistical analyses were performed to test for significant differences between fricatives for each subject of each language. First, a significant difference test was conducted for the increment of L2 between Wm and W-10. A one-way ANOVA test for Korean showed that there was a significant difference between fricatives for both LSS ($F = 27.876, p = .000$) and SHJ ($F = 10.730, p = .004$), and a significant difference between subjects for both /s^h/ ($F = 19.731, p = .000$) and /s*/ ($F = 12.731, p = .002$). A one-way ANOVA test for English showed that there was a significant difference between fricatives for both LH ($F = 12.072, p = .003$) and ST ($F = 9.230, p = .007$), while no significant difference between subjects for both /s/ ($F = .002, p = .963$) and /z/ ($F = 2.690, p = .118$). Second, a significant difference test was conducted for the increment of L2 between W-50 and W-10. A one-way ANOVA test for Korean showed that there was a significant difference between the fricatives for both LSS ($F = 10.345, p = .005$) and SHJ ($F = 15.350, p = .001$) and a significant difference between subjects for both /s^h/ ($F = 27.795, p = .000$) and /s*/ ($F = 37.511, p = .000$). A one-way ANOVA test for English showed that there was a significant difference between the fricatives for both LH ($F = 15.937, p = .001$) and ST ($F = 14.042, p = .001$), while no significant difference between subjects for both /s/ ($F = 1.205, p = .287$) and /z/ ($F = .690, p = .417$).

3. Discussion

We investigated temporal and spectral characteristics of the fricatives of Korean and English in VCV context. First, an examination of the duration of the whole token showed that Korean /as*a/ was significantly longer than /as^{h}a/ for one of the Korean subjects and that there was no significant difference between /asa/ and /aza/. On the other hand, an examination of the proportion of the segments showed that Korean /s*/ is significantly longer than /s^h/ and English /s/ than /z/ without a significant difference in the proportion of the subsequent vowel. The difference in the duration of the whole token between subjects is not a serious issue, since an absolute duration of the whole token could be quite variable. However, the significant difference in the proportion of the fricative in both languages leads us to conclude that it may be an important indicator of the difference between the homorganic fricatives in VCV context in each language.

Second, F0 was higher in the onset of the subsequent vowel than in the other points for both languages. On the other hand, an investigation of the F0 in the onset of the subsequent vowel demonstrated that there was no significant difference between /s^h/ and /s*/ while a significant difference between /s/ and /z/. The inter-speaker difference is not an important issue, since F0 is often quite variable across speakers. In Korean, F0 in the onset of the vowel following either /s^h/ or /s*/ was higher, which was also noted in (1993) and Ahn (1999). In English, F0 in the vowel onset after /s/ was significantly higher than after /z/, which may be an important attribute distinguishing the homorganic fricatives in English.

Third, an examination of spectral characteristics of the trace of F2 peaks demonstrated interesting results. We could observe a gradual increase of L2 in Korean /as^ha/ and an abrupt increase of L2 in Korean /as*a/ and a gradual increase of L2 in English /aza/ and an abrupt increase of L2 in English /asa/. Quantification of the trace of F2 peaks showed that Korean /s^h/ was significantly higher than /s*/ in the mean L2 of the fricative and the mean increment and English /z/ than /s/. Difference in the mean increment between fricatives was constant for both languages. A significantly higher values of both metrics in Korean /s^h/ is due to a higher level of L2 around the offset of the fricative. The higher L2 around the offset of the fricative is an indicator of aspiration and the fricatives were distinguished with a significant and constant difference. However, it is not sure that the same is true for English fricatives, even though trace of F2 peaks also served as an important cue to distinguish /s/ and /z/. Whether it is aspiration or some other characteristic that causes a higher value of the L2 around the offset of the fricative fricatives remains to be studied.

4. Conclusion

It has been made apparent that fricatives can be distinguished in terms of temporal and spectral cues, such as proportion of the fricative in VCV context, F0, and trace of F2 peaks. It has been claimed that trace of F2 peaks serves as an important cue to distinguish fricatives and the spectral attributes in the trace of F2 peaks around the offset of the fricative is an indicator of aspiration. It is notable that Korean /s*/ is very similar to English /s/ in all the aspects discussed in the present study. A study of a language with three series of fricatives (voiced, voiceless unaspirated, and voiceless aspirated), if any, would enable us to describe phonation types of fricatives more satisfactorily. The metrics proposed in the present work must be attested through a study of other classes of sounds, in various contexts, and in different languages with a sufficient number of subjects. Those works remain to be done.

References

Ahn, Hyunkee. 1999. Post-release phonatory processes in English and Korean: Acoustic correlates and implications for Korean phonology. PhD dissertation, The University of Texas at Austin.

Jun, Sun-A. 1993. The phonetics and phonology of Korean prosody, PhD dissertation, Ohio State University.

Lisker, Leigh and Arthur Abramson. 1964. A cross-language study of voicing in initial stops: Acoustical measurement. Word **20**: 384-422.

Maddieson, Ian. 1984. *Patterns of sounds*. Cambridge: Cambridge University Press.

Soli, Sigfrid. 1981. Second formants in fricatives: Acoustic consequences of fricative-vowel coarticulation. Journal of the Acoustical Society of America **70**: 976-984.

Department of Linguistics
Calhoun Hall 501
The University of Texas at Austin
Austin, TX 78712-1196

phans@mail.utexas.edu