Acquisition of English fricatives by Korean ESL speakers: durational aspect*

Sun-Kwang Bae
Simon Fraser University

Using eight Korean ESL learners' production data for English / tʃ, dʒ, s, ʃ, z / in the word-initial position of nonsense words, two hypotheses were tested: i.e., first, whether there is any difference in the production duration of the fricatives compared with a native speaker group (two participants); second, whether there is a correlation between the duration of fricative production and native speakers' perceptions of the goodness of the productions. For the perception test, the five segments produced by a total of ten participants were presented to two phonetically trained judges in a CV context. The first hypothesis was partially confirmed by the acoustical analysis, in that the Korean ESL speakers' production duration of /z/ was significantly shorter than that of native speakers. The second hypothesis was also confirmed, in that the native judges' goodness rating for the production data showed significant association with the duration.

1 Introduction

For Korean ESL learners, whose L1 utilises syllable timing for its metrical structure, difficulties may be expected in acquiring the syllable length related aspects of a stress-timed language. For example, Ueyama (1999) studied her Japanese subjects' difficulties in acquiring the durational contrast between stressed and unstressed syllables, especially for reduced vowels. Acquiring durational variations of English consonants may also be one of the challenges for Korean ESL learners, especially in syllable-initial stressed position, as their L1 exhibits a shorter duration in the syllable onset position (Ko, 1998: p.17). In this respect, English fricatives in the syllable-initial stressed position may pose an obvious challenge, not only because they belong to a longer class of consonants, but also because temporal cues are relevant in their differentiation (e.g., Flege, 1984; Flege and Hillenbrand, 1986). This challenge, which seems less obvious as the durational variations do not serve a contrastive purpose in the system of phonemic oppositions, may well result in the native English speakers' perception of foreign accented speech in Korean speakers' production or even in a misinterpretation of the sound.

This paper tests two main hypotheses related to this challenge, using Korean ESL speakers' English consonant production data and native English speakers' perception data

* This paper is based on a term-project of the Applied Phonetics course run by Dr Murray J. Munro at Simon Fraser University. I am grateful for his guidance and comments, but any remaining errors are of course my responsibility. I am also grateful to each of the anonymous participants who volunteered for this project.

for the respective segments. The first working-hypothesis is that *there is difference in the production duration of English fricatives between native speakers of English and Korean ESL speakers*. That is, it is expected that the duration of Korean ESL speakers' productions of English fricatives (and, perhaps, English consonants in general) in the syllable-initial stressed position is shorter than that of native English speakers. In this study, the segment duration is measured not only for its absolute duration, but also in relative terms within a word (CVCV) context to take into account individual differences in speaking rate.

The second working-hypothesis is that *there is a correlation between the production duration of fricatives and native speakers' perception on their goodness evaluation*. It is expected that native speakers of English may evaluate the shorter-duration fricatives (and other consonants as well) as less native-like than their longer counterparts. In the analysis of the perception task data, the relative duration is not taken into account, as it is irrelevant to control individual speaking rate for the perception task of CV context playback (see the method section on perception task for detail).

2 Method

2.1 Production experiment

Speakers. Two native speakers of English (NE, both female) and eight Korean ESL speakers (NK, 3 male and 5 female) from the Greater Vancouver area volunteered for this study. The NE speakers were all born in Canada, and the NK speakers either had arrived in Canada or had lived in other English speaking countries more than a year (one in the US and the other in Australia) between the ages of 2 and 27. The mean length of residency in an English speaking environment of the NK speakers was 7.7 years, ranging from two months to 15 years. Except for one NK participant, all of the speakers were enrolled in the Canadian formal education system: seven were undergraduate students at Simon Fraser University (SFU) and two were high school students in British Columbia. The NK speakers were further divided into two groups, based on their age of first exposure to English (see Table 1 for summary characteristics of the groups). The criterion used for grouping in this study was age 12 in their onset exposure to L2 environment. All of the LO NK speakers, however, had also been exposed to English in a foreign-language classroom setting from the age of 12.

Table 1
Summary characteristics of the participants: group means (and standard deviations)

	N	Age	Age of Arrival	Length of Residency	Gender
NE speakers	2	23 (1.15)	-	23 yrs. (1.15)	2 females
EO NK speakers	4	20.5 (2.88)	7.5 (3.74)	12 yrs. (2.00)	3 females 1 male
LO NK speakers	4	24.75 (10.21)	20.5 (5.58)	3.35 yrs. (3.39)	2 females 2 males

Note: NE - native English; EO NK - early-onset native Korean; LO - late-onset.

Recordings. Speech samples were recorded in a sound-treated room in the Phonetics Research Lab at SFU, using a Pioneer PDR-509 CD recorder, a MB QUART K800 microphone attached to a headset and a RadioShack audio mixer(Model: 32-8208). A two-syllable nonsense word-list that contained full canonical configurations of all syllable-initial English consonants in the vowel context of /_a_i/ was constructed, and the speakers repeated the nonsense words in a carrier sentence, "I said _____ and left the room," after listening to the model sentences through headphones. A written version of the word list was also provided to level out differences in listening comprehension ability (see Appendix for the word list). Each of the words in the carrier sentence was presented twice to repeat. For the NK groups, full configurations of Korean consonants were also elicited in a similar manner. This paper, however, focuses only on the English fricatives and affricates for further analysis.

Measurements. The CD recordings of the sentences were re-sampled at 22.1 KHz with 16 bits resolution using the SoundEdit 16. Using the same program, the duration of the first consonantal segments in the target words was measured to the nearest .001 s. or 1 ms., and the duration of the target words was also measured for VCV, excluding the first segment, to calculate the relative duration of the segments. All the measurements were based on waveform analysis and wide-band spectrograms.

As a method of checking measurement reliability, 10 percent of all the tokens were randomly sampled and re-measured. Among the 10 fricative and affricate segments, only five segments, which showed little variation in the repeated measurements, were selected and used for further analysis in the production portion of this study. The five segments selected were / tʃ, dʒ, s, ʃ, z /. The mean of absolute differences between the original measurements and the re-measurements was 1.6 ms (s.d. 1.6; range 0 to 4.7) for the first segments in isolation and 1.5 ms (s.d 1.9; range 0 to 5.2) for the rest of the VCV's.

2.2 Perception task

Judges. Two NE female graduate students in the Department of Linguistics at SFU volunteered to rate the stimuli. Both judges reported normal hearing.

Stimuli. For the perception test, two hundred tokens of the 10 word-initial fricative and affricate segments were prepared. The use of the first CV was preferred to the whole word, as it has fewer potential distracters than a whole word, without losing the segmental properties of the consonants - transition, for example. The CV's were cut after the syllable peak: i.e., just before the transition of the vowel to the following consonants was evident. The judges were reminded to rate only the initial C in each case, although they were presented within the CV context. The stimuli were played back to the judges in random order using custom playback software.

Task. The judges rated each segment on a five-point scale, ranging from 1, 'a very good exemplary segment,' to 5 'a very bad segment.' The judges had one demonstration session to familiarise them with the task. The stimuli for the demonstration session consisted of a random selection of 10 per cent of the tokens of the actual task. Twenty tokens of 10 segments, a total of 200 tokens, were rated segment by segment. That is, the judges had a total of 10 sessions, each session consisting of rating 20 tokens of one

segment. The judges were informed in advance about which segment they would be rating. Again, only the five segments, for which the duration was reliably measured, were used for further analysis.

Inter-judge reliability was evaluated with the perception scores of 100 tokens of the five segments. For the two judge's scores a Pearson correlation coefficient was calculated, yielding a modest .724. Although this inter-rater reliability is not so high, such a correlation coefficient is not unusual for such perception tasks (Hatch and Lazaraton, 1991: p.534). There was no significant difference in the average time spent for a session between the two judges (Judge 1 - mean: 81.7 seconds, s.d.: 6.2; Judge 2 - mean: 85.6 seconds, s.d.: 16.4)

3 Results

3.1 Analysis of production data

Overall duration. The three group means of the absolute duration (AD) and the relative duration ratio (RD = C / VCV * 100) of all the five segments are reported in Table 2. The early-onset (EO) NK group produced slightly longer segments (127.3 ms) than the late-onset (LO) NK group (116.2 ms) or the NK group (116.3 ms). However, ANOVA statistics showed that the three speaker groups showed no statistically significant differences in their overall production of the five segments in terms of duration: F (2, 97) = .81, p = .45 for the AD; and F (2, 97) = 1.55, p = .22 for the RD.

Table 2
Duration of the word-initial segments: mean-duration of all the segment pooled

	N	Mean	Std. Deviation	F (df)		Sig.
Absolute duration (ms.)				.810	(2, 97)	.448
NE speaker group	20	116.3	35.4			
EO NK speaker group	40	127.3	44.5			
LO NK speaker group	40	116.2	43.8			
Relative duration (%) (C/VCV*100)				1.554	(2, 97)	.217
NE speaker group	20	26.3	8.9			
EO NK speaker group	40	31.1	13.3			
LO NK speaker group	40	27.2	11.3			

Duration of affricate and fricative. This time, the five segments are pooled into affricates and fricatives, and their respective group means for the AD and the RD are reported in Table 3 with respective ANOVA statistics. Again, there is no significant difference in the affricate and the fricative production as pooled between the three speaker groups.

Table 3
Duration of the word-initial segments: mean-duration of the affricates and fricatives Pooled

AFFRICATES	N	Mean	Std. Deviation	F (df)		Sig.
Absolute duration (ms.)				.248	(2, 37)	.782
NE speaker group	8	92.1	31.8			
EO NK speaker group	16	94.3	31.1			
LO NK speaker group	16	100.8	32.0			
Relative duration (%) (C/VCV*100)				.177	(2, 37)	.838
NE speaker group	8	21.3	8.6			
EO NK speaker group	16	22.4	9.5			
LO NK speaker group	16	23.7	10.0			

FRICATIVES	N	Mean	Std. Deviation	F (df)		Sig.
Absolute duration (ms.)				2.004	(2, 57)	.144
NE speaker group	12	132.4	28.5			
EO NK speaker group	24	149.3	35.2			
LO NK speaker group	24	126.5	47.0			
Relative duration (%) (C/VCV*100)				3.006	(2, 57)	.057
NE speaker group	12	29.7	7.7			
EO NK speaker group	24	36.9	12.3			
LO NK speaker group	24	29.5	11.7			

Segment by segment. Table 4 reports the mean production duration of each segment of the three speaker groups with respective ANOVA statistics. Among the five segments, the duration of /z/ appears to show significant difference between the three speaker groups: F (2, 17) = 5.63, p = .01 for the AD; and F (2, 17) = 4.06, p = .04 for the RD. The Tukey's HSD post-hoc test at the .05 level reveals that the LO NK speaker group produces /z/ with significantly shorter duration than the EO NK group: p = .01 for the AD, and p = .03 for the RD. There appears to be a statistically significant difference neither between the NE group and the LO NK group (p = .12 for the AD, and p = .45 for the RD) nor between the NE group and the EO NK group (p = .85 for the AD, and p = .53 for the RD).

Table 4
Duration of the word-initial segments: mean-duration by each segment

	N	Mean	Std. Deviation	F (df)		Sig.
/tʃ/ Absolute Duration				1.035	(2,17)	.377
NE group	4	121.2	7.2			
EO NK group	8	118.4	18.6			
LO NK group	8	128.7	12.0			
/dʒ/ Absolute Duration				.302	(2,17)	.744
NE group	4	63.1	7.3			
EO NK group	8	70.1	19.8			
LO NK group	8	73.0	25.1			
/s/ Absolute Duration				1.974	(2,17)	.170
NE group	4	132.8	9.9			
EO NK group	8	160.2	25.8			
LO NK group	8	145.4	24.8			
/ʃ/ Absolute Duration				1.274	(2,17)	.305
NE group	4	163.7	13.4			
EO NK group	8	179.6	22.6			
LO NK group	8	161.6	28.1			
/z/ Absolute Duration (ms.)				5.629	(2,17)	.013
NE group	4	100.8	8.3			
EO NK group	8	108.2	22.0			
LO NK group	8	72.3	25.9			
/z/ Relative Duration (%)				4.125	(2,17)	.036
NE group	4	21.0	2.6			
EO NK group	8	25.6	8.5			
LO NK group	8	15.9	6.2			

3.2 Analysis of perception task data

The speaker group difference. Means of the rating scores -- 1 for 'very good segment' and 5 for 'very bad segment' -- given to each of the speaker groups by the judges are reported in Table 5. The NE speaker group received the best scores from both of the judges and the EO NK speaker group gets the next, as it would be expected. Even though the magnitude of the scores given by each judge was different, the tendency of the scores showed a similar pattern as evidenced by the correlation reported earlier: thus, the scores of the both judges are averaged for further analysis - the goodness rating scores (GRS). Note, however, that the differences in the rating scores between the three groups are not statistically significant.

Table 5
Each judge's goodness rating scores by the speaker group

	N	Mean	Std. Deviation
Judge1			
NE group	20	2.25	1.25
EO NK group	40	2.70	1.34
LO NK group	40	2.95	1.50
Judge 2			
NE group	20	1.50	1.05
EO NK group	40	1.93	1.49
LO NK group	40	2.05	1.71
Index [(Judges 1+2)/2]			
NE group	20	1.88	1.10
EO NK group	40	2.31	1.33
LO NK group	40	2.5	1.46

Segment difference. Among the five segments, /z/ received the worst scores for the all three groups, follows by /dʒ/, and so forth (see Table 6): F (4, 95) = 19.93, p = .00. The Tukey's HSD post-hoc test at the .05 level reveals that the voiced segments of /z/ and /dʒ/ were rated significantly worse than the other voiceless segments, and /z/ was rated significantly worse than /dʒ/. There was no statistically significant difference in the GRS among the voiceless segments.

Table 6
Goodness rating scores by the segment

	N	Mean	Std. Deviation	F(df)		Sig.
/tʃ/	20	1.70	.52	19.925	(4, 95)	.000
/dʒ/	20	2.85	1.23			
/s/	20	1.40	.70			
/ʃ/	20	1.73	.70			
/z/	20	3.83	1.56			

Table 7
Tests of between the segment and the group effects: two-way ANOVA

Source	df	Mean Square	F	Sig.
Intercept	1	447.227	430.604	.000
Speaker Group Effect	2	2.609	2.512	.087
Segment Effect	4	17.609	19.955	.000
Interaction Effect	8	.616	.593	.781

Between-Subject Factors: Descriptive Statistics

	N	Mean	Std. Deviation
/tʃ/			
NE group	4	1.375	.250
EO NK group	8	1.875	.583
LO NK group	8	1.688	.530
/dʒ/			
NE group	4	2.625	.479
EO NK group	8	3.063	1.425
LO NK group	8	2.750	1.363
/s/			
NE group	4	1.125	.250
EO NK group	8	1.188	.259
LO NK group	8	1.750	1.000
/ʃ/			
NE group	4	1.250	.289
EO NK group	8	1.813	.704
LO NK group	8	1.875	.791
/z/			
NE group	4	3.000	1.780
EO NK group	8	3.625	1.553
LO NK group	8	4.438	1.558

The GRS of each segment were further broken down by speaker groups, and a two-way ANOVA was carried out to examine any interaction effects between the segment effects and the speaker group effects. These results are summarised in Table 7. The *speaker-group*segment* interaction term is not significant ($p = .78$), so the speaker-group and segment can be assumed to be consistent across levels of the other factor. This is apparently so, as the mean scores broken down by the segment and the speaker-group show a consistent pattern across the speaker groups within each segment. The main effect for segment is statistically significant ($p = .00$), but not the effect of speaker-group ($p = .09$).

Segment duration and perception scores. To examine whether the GRS are related to the duration of the segment production, a series of Pearson's correlation analysis were performed: i.e., between the duration of the two affricates and the GRS, and between the duration of the three fricatives and the GRS. The correlation coefficient between the fricative duration and the GRS is a moderate -.566 ($p = .00$), indicating that the longer the duration, the less (i.e., the better) the GRS. There is, however, no significant relationship between the affricate duration and the GRS ($r = -.236$, $p = .14$).

4 Discussion

The first working hypothesis that the NK speaker groups may produce English fricatives with shorter duration than the NE speaker group is only partially confirmed in this study. That is, there were no significant differences in the production duration (in terms of the absolute or relative duration) between the NE speaker group and the NK speaker groups for the fricatives (/s, ʃ, z/) as a whole. However, it is notable that the late-onset NK speaker group produces /z/ with a significantly shorter duration than the other two groups, and this conforms to the expectation. The early-onset NK speaker group is not significantly different from the NE speaker group in the production duration of /z/.

Why is there the difference? It can be speculated that the LO NK speakers, whose L1 lacks the voiced fricative in their phonemic inventory, may have substituted a similar segment in their L1, a lax unaspirated alveo-palatal affricate (Schmidt, 1996: p.3206). The lack of [+voice] feature in their obstruent system may also have posed difficulties in extending the duration.

The second working hypothesis, that native speakers' perception of the goodness of fricatives is related to production duration, is confirmed in this study. First of all, it is noteworthy that the goodness of the affricates is not significantly associated with the duration, indicating that other acoustic cues rather than temporal cues may have been utilised for their perception of goodness. There is a clear indication that temporal cues are utilised in the perception of fricatives, on the other hand. Second, the degree of association between the goodness rating and the fricative duration is a modest one, indicating that there may be some additional acoustic properties that are relevant for the perception of goodness. We cannot rule out, furthermore, a possibility that some other acoustic properties confound the association of the goodness rating and the duration.

Another pattern emerged in analysis of perception task data is that some segments - voiceless ones in this study - get a consistently better score than the other segments across the speaker groups. It is not clear why this is so. If this cannot be attributed to an artefact of experimental design -- i.e., the judges may have assigned the better scores altogether for the better produced segment on average across the speaker groups, it can be speculated that voiced segments in sentence level utterances tend to be less accurate than their voiceless counterparts if they are heard in isolation.

5 Conclusion

This exploratory paper examined a durational aspect of the acquisition of English fricatives, which may pose an inherent difficulty for the ESL learners whose L1 is a syllable-time language. Clearly, further research is needed to establish the durational aspect as one of the acoustical correlates of fricative acquisition among L2 speakers, especially in relation to other spectral correlates such as the location of spectral peaks. Nonetheless, the evidence provided so far calls for further research in the area of L1 syllable structure transference among the L2 learners. For example, timing of complex onset acquisition can also be an interesting topic to pursue in terms of the syllable-onset duration acquisition.

Another topic of interest may be to examine whether a symmetric pattern exists across the syllable-timed and stress-timed languages. That is, is it easier for the stress-

time language speakers to acquire the syllable-time language in terms of their metrical structure? Or is it more difficult or is it the same? Examination of the bilinguals or the early-onset L2 groups' performance in their comparable L1 production may provide at least a partial answer for this question: i.e., by examining whether the acquisition of L2 adversely affects the learners' L1 system. More complete answer can be achieved by examining a balanced set of cross-linguistic L2 acquisition data.

References

Flege, J. 1984. The effect of experience on Arabs' perception of the English /s/ vs. /z/ contrast. *Folia Linguist.* 18, 117-138.

Flege, J. E. & Hillenbrand, J. 1986. Differential use of temporal cues to the /s/-/z/ contrast by native and non-native speakers of English. *Journal of the Acoustical Society of America.* 79, 508-517.

Hatch, E. & Lazaraton, A. 1991. Assumptions of statistical test. Ch.18. *The Research Manual: Design and Statistics for Applied Linguistics.* Boston: Heinle & Heinle Publishers.

Ko, D. H. 1998. *Phonetics Studies in North Korea* [in Korean]. Seoul: Hanguk Munhwa-sa.

Schmidt, A. M. 1996. Cross-language identification of consonants. Part 1. Korean perception of English. *Journal of the Acoustical Society of America.* 99, 3201-3211.

Ueyama, M. 1999. Durational reduction in L2 English produced by Japanese speakers. *ICPhS* 99, San Franciso, 567-570.

Appendix Word list

vally [váli]	laffy [láfi]	rajee [rádʒi]	jappy [dʒápi]	chammie [tʃámi]
hatty [háti]	shanny [ʃáni]	zabby [zábi]	sadee [sádi]	fathee [fáði]
thagi [ðági]	thaky [θáki]	natchie [nátʃi]	mathie [máθi]	garee [gári]
kazie [kázi]	dasee [dási]	tavy [távi]	bashie [báʃi]	pahee [páhi]