English Language & Literature Teaching, Vol. 17, No. 4 Winter 2011

An Acoustic Study of English Voiced Sibilants: Correct vs. Incorrect L2 Production

Misun Seo[*]
(Hannam University)
Jayeon Lim[**]
(University of Seoul)

Seo, Misun & Lim, Jayeon. (2011). An acoustic study of English voiced sibilants: correct vs. incorrect L2 production. *English Language & Literature Teaching, 17*(4), 251-271.

The present study analyzed Korean learners' production of English /z/-/dʒ/ and /z/-/ʒ/ contrasts in terms of native speaker judgments and acoustic measurements. Korean learner's production was judged to be either correct or incorrect by native English speakers. Correct and incorrect productions were then compared with productions of native speakers' in terms of acoustic analyses. The results indicated that Korean speakers' correct production was more similar to that of native speakers by sharing more acoustic cues. Incorrect production by Korean speakers indicated patterns either different or opposite from that of native speakers, confirming native speaker judgments. The results also revealed acoustic cues on which native speakers rely in judging L2 speech, thereby implying that the more consistent along with more number of acoustic cues used by native speakers may facilitate the acquisition of segment contrasts by L2 learners.

[L2 production/English voiced sibilants/native speaker judgments/acoustic measurements]

I. INTRODUCTION

When analyzing the L2 speech data, there have been different methods employed in its analyses. Of these, three common methods are transcription of L2 speech data, native speaker judgments, and acoustic measurements. A first method of transcribing L2 speech

---

[*] First author
[**] Correspondent author

data generally sets out to find allophonic or phonemic errors in the speech. Either trained-phoneticians or naïve L1 users participated in the procedure of transcription (Joh & Lee, 2001; Schmidt, 1996). As pointed out by Zampini (2008), a problem inherent to this method is that it can be subjective since it relies solely on transcribers.

A second method of native speaker judgments is more commonly used than the first one. Here, native speakers are asked to judge the L2 speech using an interval scale. Piske, MacKay and Flege (2001) report that intervals used in previous studies range from anywhere between three and nine, with a five-point scale being most frequent. Examples of these native speaker judgment studies have sought to measure global accentedeness, comprehensibility and/or intelligibility of L2 speech (Lim & Seo, 2011; Gass & Varonis, 1984; Hattori, 2009; Munro & Derwing, 1998; Munro, Derwing & Morton, 2006; Sung, 2006). Although the method retains some objectivity by employing an equal-appearing interval scale, inter- and intra-rater variability may still exist.

Recently, the most frequently employed method is measuring the relevant acoustic characteristics of L2 speech and comparing that with those of native speaker data (Flege, 1981; 1987; Flege & Hillenbrand, 1984; Flege, Bohn & Jang, 1997; Koo, 1997; Lee & Guion, 2008; Lee, Guion & Harada, 2006; Seo & Lim, 2010). By making comparisons between L1 and L2 data, researchers can observe different acoustic characteristics that distinguish the two. Compared with the previous two methods, this is relatively more objective (Zampini, 2008). Although this method can provide crucial information distinguishing L1 and L2 speech, it is not without a problem. When acoustic analyses of both correct and incorrect L2 speech are collapsed in statistical analyses, the patterns of the data may be obscured hence leading to a difficult interpretation. The problem was also mentioned in Leather (1999), where the difficulty of the data interpretation was pointed out in large-scale acoustic and statistical analyses.

In the present study, in assessing L2 speech, the methods of native speaker judgment in the form of identification experiment along with acoustic measurements will be employed. By employing both methods, the study aims to complement subjectivity issue in native speaker judgment data and also difficulty of data interpretation in acoustic analyses. A goal of this study is to investigate Korean learners' production of the English /z/-/d□/ and /z/-/□/ contrasts. In doing so, the study will acoustically analyze Korean learners' production judged to be correct/incorrect by native English speakers in terms of an identification experiment. The results will be compared to that of native speakers so as to compare acoustic cues used in distinguishing /z/ from /d□/ and /z/ from /□/.

Specifically, following research questions were explored in the present study:

(1) With the /z/-/d□/ and /z/-/□/ contrasts produced by Korean learners, how would native English speakers judge the pairs in terms of correctness/incorrectness?

(2) In the acoustic analyses of /z/-/dʒ/ and /z/-/ʒ/ contrasts, what would be the distinguishing acoustic cues employed in the native English speakers' production versus Korean learners' correct/ incorrect production?
(3) What would be the relationship between the results of the native speakers' identification and the acoustic analyses of /z/-/dʒ/ and /z/-/ʒ/ contrasts?

II. METHOD OF RESEARCH

1. Participants

The production experiment was conducted with eight Korean learners of English (four males and four females) who attended a certain university located in Seoul and were originally from Seoul. Also, four native speakers of English (two males and two females) participated in the same experiment to provide native speaker base-line data. They were all from North America with an average of 6.3 years of length of residence in Korea at the time of recording. Additionally, a different group of five native speakers of English (three males and two females) from North America judged the production data of Korean learners of English. They had an average of 4.8 years of length of residence in Korea and were teaching English at a certain university in Seoul at the time of data collection. All the participants were paid for their participation in the experiment.

2. Materials

For the production experiment, 36 words consisting of /z/-/dʒ/ or /z/-/ʒ/ minimal pairs were chosen. The words included minimal pairs of /z/-/dʒ/ in three different positions, (i.e., word-initial, medial and final positions) and the pairs of /z/-/ʒ/ in two positions (i.e., word-medial and final positions). Word-initial position was excluded from /z/-/ʒ/ since no minimal pairs were found (Please see Appendix for the stimuli). Four native speakers of English produced 144 tokens, and eight Korean learners of English 288 tokens. For the identification experiment, each of the five native speakers heard and judged the production of 288 tokens (36 words x 8 speakers) by Korean learners of English.

3. Procedure

1) Production Experiment

For the production experiment, participants were instructed to read each word put in a carrier sentence (i.e., Please say ___________.) as naturally as possible. The order of words was randomized for each participant. Recordings were conducted in a quiet room with a head-mounted microphone (Audio-Technica M8541) and TASCAM HD-P2 recorder at a setting of 44100 Hz and 16 bits. Three recordings were taken from each participant, of which the last one underwent acoustic analyses.

Acoustic measurements of the production data were taken in Praat. Four types of the spectral moment analysis (i.e., centroid, SD, skewness and kurtosis) were made by using FFT spectra in the frequency range from 500 to 10,000 Hz and a 40 ms Hamming window centered around the midpoint of the fricative noise. In addition, F2 and F3 were measured at the sibilant-vowel boundaries for the words containing a voiced sibilant in word-initial and medial positions, and at the vowel-sibilant boundaries for the words containing a voiced sibilant in word-final position.

2) Identification Experiment

For the identification experiment, each of the five native speakers listened to the same 288 tokens (36 words x 8 Korean learners) which were acoustically analyzed. Upon hearing each token, they were provided with two choices on a computer screen and were instructed to choose the one they thought they had heard by pressing one of the two buttons. For example, upon hearing the word '*zoo*', they were shown both '*zoo*' and '*Jew*' on a computer screen and were instructed to choose the word they heard.

III. RESULTS

1. Identification Experiment

In analyzing the results of the identification experiment, the score of 1 was provided to an accurately identified token and 0 to an inaccurately identified one. Thus, for each of the 288 tokens produced by Korean learners, the accuracy scores ranged from 0 to 5. When all the five native speakers correctly identified an aurally presented token, its accuracy score was 5, and 0 when all of the native speakers misidentified it. Based on such accuracy scores of tokens from the identification experiment, the 288 tokens produced by Korean learners were classified as either correct or incorrect productions. In classifying the production data, the 75% accuracy norm was employed following Li, Edwards and Beckman (2009), Prather, Hedrick and Kern (1975), Smit, Hand, Freilinger, Bernthal and Bird (1990) and Templin (1957). According to them, a speech sound is considered to have been mastered when it renders 75% accuracy. Following the same criterion, we determined the Korean learners to have shown correct production when at least four out of five native speakers correctly identified the token. Any tokens that were identified as correct by less than four were identified as incorrect. Table 1 below shows the finding. For example, for initial /z/ of the /z/-/dʒ/ contrast, 47.5%, or 19 out of 40 tokens were correct productions and 52.5%, or 21 out of 40 tokens were incorrect ones. The results are classified according to position since initial sibilants include CV transitions and final sibilants VC transitions. Based on the results provided in Table 1, the production data of Korean native speakers are classified as either correct or incorrect and analyzed acoustically. In doing so, the learner data are compared with that of native speakers to see whether similar or different acoustic cues are employed in correct or incorrect productions.

TABLE 1

Percentage of Correct/Incorrect Productions

contrast	segment		position		
			initial	medial	final
/z/-/dʒ/	/z/	correct	47.5% (19)	87.5% (21)	75% (30)
		incorrect	52.5% (21)	12.5% (3)	25% (10)
	/dʒ/	correct	75% (30)	66.7% (16)	82.5% (33)
		incorrect	25% (10)	33.3% (8)	17.5% (7)
/z/-/ʒ/	/z/	correct	N/A	62.5% (15)	62.5% (10)
		incorrect	N/A	37.5% (9)	37.5% (6)
	/ʒ/	correct	N/A	41.7% (10)	68.8% (11)
		incorrect	N/A	58.3% (14)	31.2% (5)

NOTE: Total number of productions = 288

Numbers in parentheses indicate raw number.

2. Acoustic analyses: /z/-/dʒ/ pair

For the tokens of the /z/-/dʒ/ pair, four fricative spectrum moments (i.e., centroid, SD, skewness and kurtosis) were measured. In addition, onset F2 and F3 frequency were measured for the tokens with sibilants in initial and medial positions. In case sibilants occur in final position, offset F2 and F3 frequency were measured. Table 2 below summarizes the definition and articulatory interpretation of each acoustic parameter.

TABLE 2

Definition of Each Acoustic Parameter (modified from Li et al. (2009))

	Acoustic parameter	Definition	Articulatory interpretation
Fricative spectrum moments	Centroid	Center of mass of the distribution (the weighted mean frequency)	Negatively correlates with the length of the front resonating cavity
	Standard deviation	Spread of the distribution (average squared distance from the centroid)	Differentiates tongue posture between apical and laminal
	Skewness	Asymmetry in the spectral shape (the difference between the spectrum below the centroid and the spectrum above the centroid)	Negatively correlates with the length of the front resonating cavity
	Kurtosis	Peakiness of the spectral shape (the average distance from the centroid raised to the fourth power, divided by the squared variance of the distribution)	Differentiates tongue posture between apical and laminal
CV (VC) transitions	Onset (Offset) F2	F2 frequency at the onset (offset) of the following (preceding) vowel	Negatively correlates with the length of the back resonating cavity
	Onset (Offset) F3	F3 frequency at the onset (offset) of the following (preceding) vowel	Negatively correlates with the degree of a lip rounding

A MANOVA was performed for native English speakers' production data of the /z/-/dʒ/ pairs in each of the three different positions. Thus, three MANOVAs were run. For Korean learners' production data, three MANOVAs were performed with the correct production data and another three MANOVAs with the incorrect production data. Dependent variables were acoustic parameters illustrated in Table 2. An independent variable was the sibilant type (i.e, /z/ or /dʒ/). Table 3 summarizes significant acoustic parameters used in differentiating /z/ from /dʒ/ in native English speakers' data and Korean learners' correct/incorrect productions, marked by 'O'. Non-use of the acoustic parameters is marked by 'X'.

TABLE 3

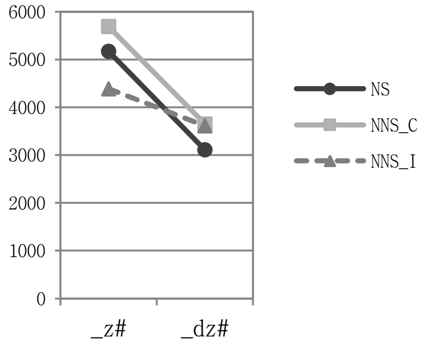
Summary of Significant Acoustic Parameters for /z/-/dʒ/

		Centroid	SD	Skewness	Kurtosis	Onset (Offset) F2	Onset (Offset) F3
Initial	NS	X	O	O	X	O	O
	NNS correct	X	O	O	X	O	X
	NNS incorrect	X	X	X	X	X	X
Medial	NS	X	O	O	X	O	X
	NNS correct	X	O	X	X	O	X
	NNS incorrect	X	X	X	X	X	X
Final	NS	O	O	X	X	O	O
	NNS correct	O	O	O	X	O	X
	NNS incorrect	X	X	X	X	X	X

Throughout the study, acoustic parameters which showed a significant difference in at least one speaker group are reported. To begin with, the acoustic cue of centroid frequency (CF) was used in differentiating /z/ from /dʒ/ in final position within native English speakers' production (df=1, F=36.289, p<.05) and non-native speakers' correct production (df=1, F=24.738, p<.05). In non-native speakers' incorrect production, centroid frequency was not an acoustic cue differentiating final /z/ (4386.4 Hz) from /dʒ/ (3612.2 Hz). As can be seen from Figure 1, both in native speakers' production and non-native speakers' correct production, the centroid frequency of /z/ (5170.7 Hz and 5689.7 Hz) was significantly higher than that of /dʒ/ (3110.9 Hz and 3649.6 Hz). In the figures throughout the study, solid lines indicate significant differences and dashed lines non-significant differences between the two sibilants.

FIGURE 1

Average CF of [z] and [dʒ] in Final

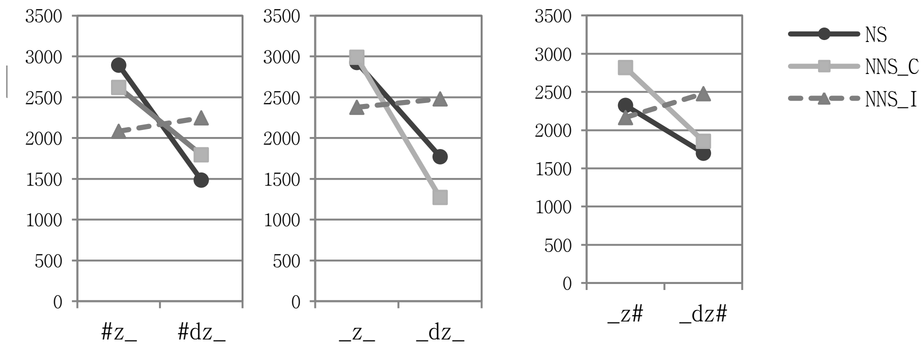


Note: dz=dʒ

Across three different positions, as illustrated in Figure 2, the acoustic cue of SD was used in differentiating /z/ from /dʒ/ in native speakers' production (df=1, F=61.244, p<.05 in initial; df=1, F=51.078, p<.05 in medial; df=1, F=6.749, p<.05 in final) and non-native speakers' correct production (df=1, F=25.324, p<.05 in initial; df=1, F=11.879, p<.05 in medial; df=1, F=32.309, p<.05 in final). SD was significantly higher with /z/ than with /dʒ/ in native speakers' production (2893.8 Hz vs. 1485.2 Hz in initial, 2926.5 Hz vs. 1771.5 Hz in medial, 2325 Hz vs. 1699.3 Hz in final) and non-native speakers' correct production (2620.6 Hz vs. 1795.5 Hz in initial, 2990.6 Hz vs. 1272.8 Hz in medial, 2820.1 Hz vs. 1855.3 Hz in final). In non-native speakers' incorrect production, SD was not a differentiating cue (2083.6 Hz vs. 2248.5 Hz in initial, 2377.8 Hz vs. 2480.5 Hz in medial, 2165.5 Hz vs. 2478.3 Hz in final).

FIGURE 2

Average SD of [z] and [dʒ] in Initial, Medial, Final

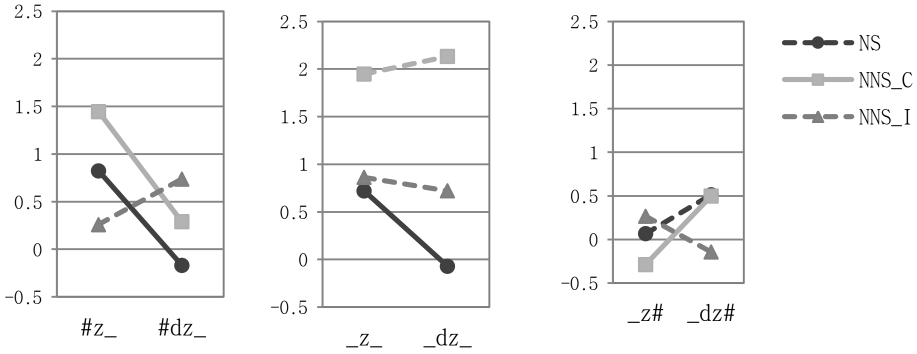


Note: dz=dʒ

The acoustic cue of skewness was a factor differentiating initial and medial /z/ from /dʒ/ with native speakers' production (df=1, F=7.737, p<.05 in initial; df=1, F=4.349, p<.05 in final). In both positions, as shown in Figure 3, the average skewness values were significantly higher with /z/ than with / dʒ/ (.8245 vs. -.1677 in initial, .7206 vs. -.0707 in medial). Skewness was an acoustic cue distinguishing initial and final /z/ from /dʒ/ in non-native speakers' correct production (df=1, F=5.739, p<.05 in initial; df=1, F=6.571, p<.05). In non-native speakers' correct production, the average skewness was significantly higher with /z/ than with /dʒ/ in initial position (1.4469 vs. .2913) and significantly lower with /z/ than with /dʒ/ in final position (-.2892 vs. .4983). In non-native speakers' incorrect production, skewness was not a differentiating factor.

FIGURE 3

Average Skewness of [z] and [dʒ] in Initial, Medial, Final

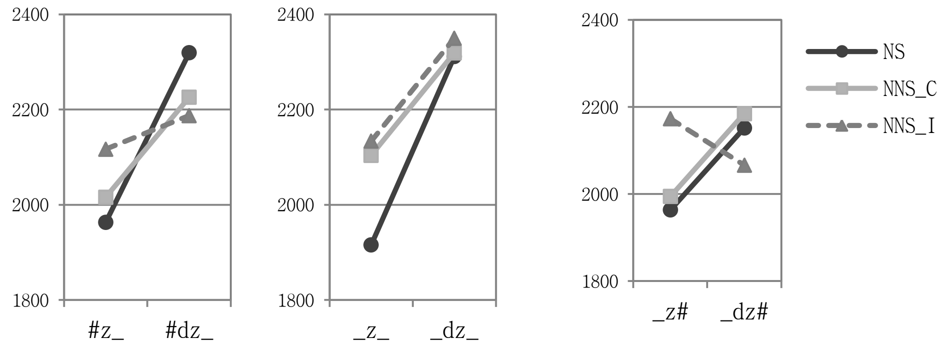


Note: dz=dʒ

Across three different positions, the acoustic cue of onset/offset F2 was employed in distinguishing /z/ from /dʒ/ in both native speakers' production (df=1, F=38.766, p<.05 in initial; df=1, F=38.154, p<.05 in medial; df=1, F=6.031, p<.05 in final) and non-native speakers' production (df=1, F=6.360, p<.05 in initial; df=1, F=4.920, p<.05 in medial; df=1, F=4.901, p<.05 in final). As can be seen from Figure 4, significantly higher F2 was attested with /dʒ/ than with /z/ in both native speakers' production (1963.1 Hz vs. 2319.4 Hz in initial, 1915.7 Hz vs. 2311.4 Hz in medial, 1963.6 Hz vs. 2151.6 Hz in final) and non-native speakers' correct production (2015.7 Hz vs. 2225.9 Hz in initial, 2103.7 Hz vs. 2318.9 Hz in medial, 1994.5 Hz vs. 2184.4 Hz in final). Onset/offset F2 was not a cue differentiating /z/ from /dʒ/ in non-native speakers' incorrect production.

FIGURE 4

Average Onset/offset F2 of [z] and [d□] in Initial, Medial, Final

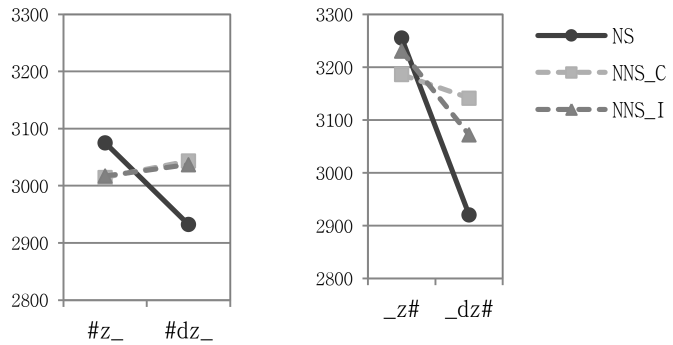


Note: dz=d□

Onset/offset F3 was used in differentiating initial and final /z/ from /d□/ by native speakers (df=1, F=4.891, p<.05 in initial; df=1, F=14.984, p<.05 in final). As shown in Figure 5, /z/ exhibited significantly higher F3 than /d□/ (3075.3 Hz vs. 2932.6 Hz in initial, 3255.8 Hz vs. 2920.6 Hz in final). In non-native speakers' correct and incorrect production, onset/offset F3 was not significantly different with /z/ and /d□/ across three different positions.

FIGURE 5

Average Onset/offset F3 of [z] and [d□] in Initial, Final



Note: dz=d□

3. Acoustic analyses: /z/-/ʒ/ pair

As with the /z/-/dʒ/ pair, four fricative spectrum moments (i.e., centroid, SD, skewness and kurtosis), onset/offset F2 and F3 frequency were measured for the /z/-/ʒ/ pair. A MANOVA was performed for native English speakers' production data of the /z/-/ʒ/ pairs in each of the two different positions, thus two MANOVAs in total. For Korean learners' production data, two MANOVAs were performed with the correct production data and another two MANOVAs with the incorrect production data. Dependent variables were acoustic parameters illustrated in Table 2 and an independent variable was the sibilant type (i.e, /z/ or /ʒ/). Table 4 summarizes significant acoustic parameters used in differentiating /z/ from /ʒ/ in native English speakers' data and Korean learners' correct/incorrect productions.

TABLE 4

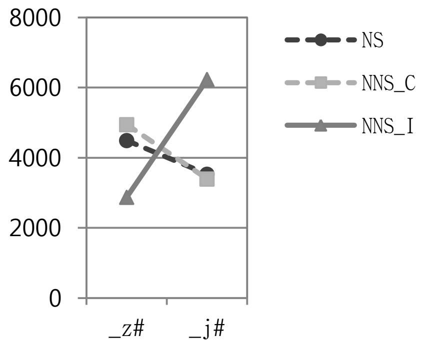
Summary of Significant Acoustic Parameters for /z/-/ʒ/

		Centroid	SD	Skewness	Kurtosis	Onset (Offset) F2	Onset (Offset) F3
Medial	NS	X	X	X	X	O	X
	NNS correct	X	X	X	X	X	X
	NNS incorrect	X	X	X	X	O	X
Final	NS	X	X	X	X	X	O
	NNS correct	X	O	X	X	X	X
	NNS incorrect	O	X	X	X	X	X

Centroid frequency was used in distinguishing final /z/ from /ʒ/ in non-native speakers' incorrect production (df=1, F=8.278, p<.05). As shown in Figure 6, final /ʒ/ exhibited a significantly higher centroid frequency than /z/ (2870.9 Hz vs. 6229.3 Hz). However, in native speakers' production and non-native speakers' correct production, the opposite trend was attested although centroid frequency of final /z/ and that of /ʒ/ were not significantly different.

FIGURE 6

Average CF of [z] and [□] in Final

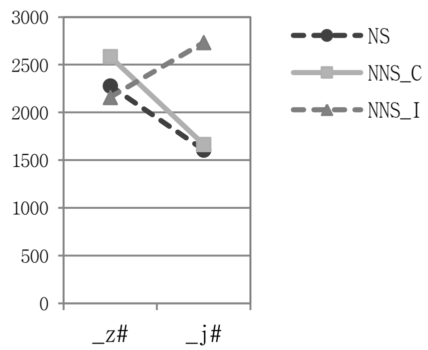


Note: j=□

The acoustic cue of SD was a differentiating cue with final /z/ and /□/ in only non-native speakers' correct production (df=1, F=18.405, $p<.05$), with higher /z/ at 2587.2 Hz than /□/ at 1666 Hz. As can be seen from Figure 7 below, despite the non-significant difference between the two sibilants in native speakers' production, a similar pattern was observed. On the other hand, in non-native speakers' incorrect production, an opposite pattern was observed.

FIGURE 7

Average SD of [z] and [□] in Final

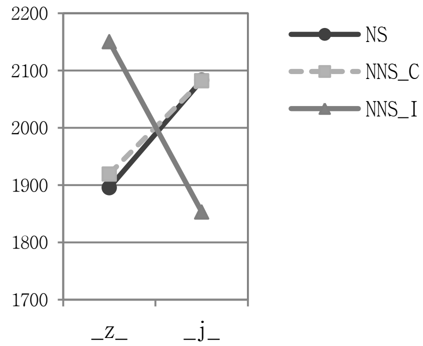


Note: j=□

With medial /z/ and /□/, onset F2 was a differentiating cue in native speakers' production (df=1, F=4.476, $p<.05$) and non-native speakers' incorrect production (df=1, F=6.912, $p<.05$). As illustrated in Figure 8, onset F2 of medial /z/ was significantly lower than that of /ʒ/ in native speakers' production (1895.4 Hz vs. 2083.2 Hz). On the other hand, the opposite pattern was attested in non-native speakers' incorrect production: /z/ at 2150.3 Hz vs. /ʒ/ at 1852.9 Hz. Non-native speakers' correct production showed a similar pattern with that of native speakers although the difference was not significant between the two sibilants.

FIGURE 8

Average Onset F2 of [z] and [ʒ] in Final

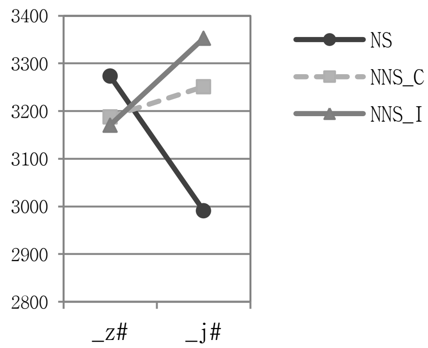


Note: j=ʒ

Offset F3 was a cue differentiating final /z/ from /ʒ/ in native speakers' production (df=1, F=6.634, p<.05). As can be seen from Figure 9, final /z/ exhibited a higher offset F3 at 3273.6 Hz than /ʒ/ at 2991.1 Hz. Non-native speakers' correct and incorrect production showed an opposite pattern: F3 was higher with /ʒ/ than with /z/.

FIGURE 9

Average Offset F3 of [z] and [ʒ] in Final



Note: j=ʒ

IV. DISCUSSION AND CONCLUSION

Acoustic analyses of /z/-/dʒ/ pairs show that SD and onset/offset F2 frequency are acoustic cues consistently used by native English speakers in differentiating /z/ from /dʒ/ across positions. As for SD, /z/ showed a larger average value than /dʒ/ in initial, medial and final positions. According to Li et al. (2009), this result indicates that the frication noise of /z/ has a more diffuse spectral shape than that of /dʒ/. Regarding onset/offset F2 frequency, /dʒ/ illustrated a higher average value than /z/ in all three positions. According to previous studies (Funatsu, 1995; Halle & Stevens, 1997), F2 frequency taken at the onset/offset of the vowel following/preceding a fricative indicates the length of the back cavity in the fricative and thus related to place of articulation of the fricative. During the production of /dʒ/, a long palatal channel and a shorter back cavity are formed due to the tongue-bunching gesture involved in the production of /dʒ/. Therefore, average F2 comes to be significantly higher for /dʒ/ than for /z/ (Jongman, Wayland & Wong, 2000; McGowan & Nittrouer, 1988; Nittrouer, 2002; Nittrouer, Studdert-Kennedy & McGowan, 1989). In non-native speakers' correct production, as in native English speakers' production, both SD and onset/offset F2 were acoustic cues differentiating /z/ from /dʒ/ in all the positions. In addition, patterns attested were similar to the ones found in native English speakers' production.

In addition to SD and onset/offset F2, there were additional acoustic cues that native English speakers used in differentiating /z/ from /dʒ/. Centroid frequency was a cue used in differentiating word-final /z/ from /dʒ/. The centroid of the fricative noise refers to the center of gravity of a defined part of the spectrum, each frequency being weighted according to its amplitude (Cho, Jun & Ladefoged, 2002; Forrest, Weismer, Milenkovic & Dougall, 1988; Jones & Munhall, 2003). Centroid frequency is negatively correlated with the length of the front resonating cavity and thus average centroid frequency around the midpoint of the frication noise comes to be higher for an alveolar fricative /z/ than for a palato-alveolar affricate /dʒ/ (Chang, Haynes, Yao & Rhodes, 2009; Nittrouer et al., 1989). In non-native speakers' correct production, as in native English speakers', average centroid frequency was significantly higher for /z/ than for /dʒ/ in final position.

Native English speakers used skewness, an acoustic parameter whose value is negatively correlated with the length of the front resonating cavity, in distinguishing /z/ from /dʒ/ in initial and medial positions. In both positions, average skewness was significantly higher for /z/ than for /dʒ/. In non-native speakers' correct production, skewness was a differentiating cue in initial and final positions. In initial position, like native English speakers, skewness was significantly higher for /z/ than for /dʒ/. However, unlike native English speakers' production where skewness was not significantly different for /z/ and /dʒ/ in final position, skewness was significantly higher for /dʒ/ than for /z/ in non-native speakers' correct production. In addition, skewness was not a differentiating cue in medial position within non-native speakers' correct production.

Onset/offset F3 frequency was used by native English speakers in distinguishing /z/ from /dʒ/ in initial and final positions. Onset/offset F3 frequency is an acoustic parameter, which is negatively correlated with the degree of a lip rounding. According to Ladefoged (2006), English palato-alveolar fricatives and affricates are produced with lip rounding which has the effect of lowering F3. In native English speakers' production, average F3 of /dʒ/ was significantly lower than that of /z/, indicating that native English speakers produced /dʒ/ accompanying lip rounding. In non-native speakers' correct production, /dʒ/ and /z/ did not show significantly different F3 values and it results from non-sufficient degree of lip rounding when producing /dʒ/.

As discussed, in non-native speakers' correct production, /z/ and /dʒ/ were acoustically differentiated in terms of acoustic parameters, some of which were also used by native English speakers. On the other hand, in non-native speakers' incorrect production, there was no single acoustic parameter used in differentiating /z/ from /dʒ/.

According to acoustic analyses of /z/-/ʒ/ pairs, there was no acoustic cue uniformly used by native English speakers across positions in differentiating the two fricatives. Onset F2 was the only differentiating cue in medial position, and offset F3 in final position. Onset F2 was significantly higher for /ʒ/ than for /z/ in medial position. As discussed above, the tongue-bunching gestures accompanied in pronouncing /ʒ/ results in a long palatal channel and a shorter back cavity which lead to a higher F2 for /ʒ/ than for /z/. In non-native speakers' correct production, /ʒ/ showed a higher average F2 value than /z/ in medial position although the difference was not significant. On the other hand, in non-native speakers' incorrect production where /ʒ/ and /z/ were distinguished by onset F2, the opposite pattern was observed. That is, a significantly higher average onset F2 value was attested for /z/ than for /ʒ/, indicating that places of articulation of the two fricatives are incorrectly targeted in non-native speakers' incorrect production.

Offset F3 was the only acoustic cue used by native English speakers in distinguishing /z/ from /ʒ/ in final position. Offset F3 was significantly lower for /ʒ/ than for /z/, indicating that native English speakers pronounced /ʒ/ with lip rounding. In non-native speakers' correct and incorrect production, offset F3 was not a differentiating acoustic cue, pointing to no lip rounding accompanied in the production of /ʒ/.

In non-native speakers' correct production, no acoustic cue differentiating /z/ from /ʒ/ was attested in medial position. SD was a differentiating cue in final position and it was significantly higher for /z/ than for /ʒ/. In native speakers' production, a similar pattern was attested although the difference was not significant.

In non-native speakers' incorrect production, onset F2 was a differentiating cue in medial position and centroid frequency in final position. With onset F2 of word-medial

/z/ and /□/, as discussed above, the pattern was opposite from the one observed in native English speakers. Centroid frequency of word-final /z/ was significantly lower than that of /□/, which is an opposite pattern from the one attested in native English speakers' production. As discussed, centroid frequency is negatively correlated with the length of the front resonating cavity. Therefore, it is expected that /z/ shows a higher value than /□/, which is the pattern attested in both native English speakers' production and non-native speakers' correct production although the difference is not significant. Thus, average onset F2 and centroid frequency values indicate that places of articulation of /z/ and /□/ were incorrectly targeted in medial and final positions in non-native speakers' incorrect production.

According to acoustic analyses conducted in the present study, non-native speakers' correct production and native English speakers' production illustrated more common acoustic parameters than non-native speakers' incorrect production and native English speakers' production in differentiating /z/ from /d□/ and /z/ from /□/. In addition, patterns of acoustic parameters were similar to the ones attested within native English speakers' production. In non-native speakers' incorrect production, either no differentiating acoustic parameters were found or their patterns were opposite from the ones found with native English speakers' production. This result shows that acoustic parameters which were found to be significant cues in distinguishing /z/ from /d□/ and /z/ from /□/ in native English speakers' production play major roles as perceptual cues to voiced sibilants when native English speakers make judgments L2 speakers' pronunciation.

Another finding of acoustic analyses is that there were more acoustic parameters differentiating /z/ from /d□/ than the ones differentiating /z/ from /□/ in native English speakers' production. In addition, consistent acoustic parameters (i.e. SD and onset/ offset F2) were involved in distinguishing /z/ from /d□/ across positions. On the other hand, differentiating acoustic parameter was inconsistently found with /z/-/□/ pairs. According to acoustic analyses conducted in the present study, Korean learners produced the /z/-/d□/ contrast more like native English speakers than the /z/-/□/ contrast. The existence of more acoustic cues and consistent cues present across positions with native English speakers' production may have facilitated Korean learners' learning of the /z/-/d□/ contrast. Additionally, there may have been the influence of L1 on such a result. According to Schmidt (1996) and Lim & Seo (2008), English voiced sibilants /z, □, d□/ are all mapped to Korean /ㅈ/, which is an affricate. When Korean learners pronounce the three voiced sibilants, they are likely to pronounce them as all affricates. Considering that English /d□/ is the only affricate among the three, Korean learners' /d□/ pronunciation would be judged as closer to the target phoneme than /z/ and /□/ which are fricatives. This is supported by acoustic analyses of the present study. According to the acoustic parameters such as centroid frequency, SD and skewness, which reflect the properties of a fricative noise in the present study, Korean learners are more likely to pronounce the ʒ-fricative noise correctly when it occurs as a subpart of an affricate /dʒ/ than when it occurs alone as a fricative /ʒ/.

In the present study, L2 production was analyzed in terms of both native speaker judgments and acoustic measurement. More consistency was observed between native speakers' production and Korean speakers' correct production than between native speakers' production and Korean speakers' incorrect production regarding the use of acoustic cues. The results of the acoustic analyses of L2 production confirms those of native speaker judgments and acoustic cues on which native speakers rely in judging L2 speech, illustrating that the acquisition of L2 contrasts can be facilitated when native speakers use more consistent acoustic cues across positions and also when they use more number of acoustic cues in differentiating members of the contrasts.

REFERENCES

Lim, J., & Seo, M. (2008). Factors affecting Korean learners' perception of English sibilants. *English Language and Literature Teaching, 14,* 201-219.

Lim, J., & Seo, M. (2011). Intelligibility and comprehensibility in the production of English sibilants by Korean learners. *Korean Journal of Applied Linguistics, 27*, 189-212.

Chang, C. B., Haynes, E. F., Yao, Y., & Rhodes, R. (2009). A tale of five fricatives: Consonantal contrast in heritage speakers of Mandarin. *University of Pennsylvania Working Papers in Linguistics, 15,* 36-43.

Cho, T., Jun, S.-A., & Ladefoged, P. (2002). Acoustic and aerodynamic correlates of Korean stops and fricatives. *Journal of Phonetics, 30*, 193-228.

Flege, J. E. (1981). The phonological basis of foreign accent: A hypothesis. *TESOL Quarterly, 15*, 443-455.

Flege, J. E. (1987). The production of "new" and "similar" phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics, 15*, 47-65.

Flege, J. E., & Hillenbrand, J. (1984). Limits on pronunciation accuracy in adult foreign language speech production. *Journal of the Acoustical Society of America, 76*, 708-721.

Flege, J. E., Bohn, O.-S., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics, 25*, 437-470.

Forrest, K., Weismer, G., Milenkovic, P., & Dougall, R. N. (1988). Statistical analysis of word-initial voiceless obstruents: Preliminary data. *Journal of Acoustical Society of America, 84*, 115-123.

Funatsu, S. (1995). Cross language study of perception of dental fricatives in Japanese and Russian. *Proceedings of the International Congress of Phonetic Sciences, 4,* 124-127.

Gass, S., & Varonis, E. (1984). The effect of familiarity on the comprehensibility of nonnative speech. *Language Learning, 34*, 65-89.

Halle, M., & Stevens, K. N. (1997). The postalveolar fricatives of Polish. In S. Kiritani, H. Hirose, & H. Fujisaki (Eds.), *Speech Production and Language*. (pp. 177-192). Berlin, New York: Mouton de Gruyter.

Hattori, K. (2009). *Perception and production of English /r/-/l/ by adult Japanese speakers.* Unpublished doctoral dissertation. University College London, London.

Joh, J.-S., & Lee, S.-S. (2001). Relationships between sound perception and production in L2 phonology acquisition. *Journal of the Applied Linguistics Association of Korea, 17*, 127-145.

Jones, J. A., & Munhall, K. G. (2003). Learning to produce speech with an altered vocal tract: The role of auditory feedback. *Journal of Acoustical Society of America*, *113*, 532-543.

Jongman, A., Wayland, R., & Wong, S. (2000). Acoustic characteristics of English fricatives. *Journal of Acoustical Society of America, 108,* 1252-1263.

Koo, H.-S. (1997). Umhyang chukjungkwa jikak pandane uihan hankukin yeongeoui unyul yeongu. (A study using acoustic measurements and perceptual judgment to identify prosodic characteristics of English as spoken by Koreans). *Speech Sciences, 2*, 95-108.

Ladefoged, P. (2006). *A course in phonetics*. 5th ed. Boston, MA: Thomson Wadsworth.

Leather, J. (1999). Second-language speech research: An introduction. *Language Learning*, *49,* 1-56.

Lee, B., & Guion, S. G. (2008). Effects of hyperarticulated clear speech in English stress production by Korean and Japanese bilinguals. *Studies in Phonetics, Phonology and Morphology*, *14*, 245-258.

Lee, B., Guion, S. G., & Harada, T. (2006). Acoustic analysis of the production of unstressed English vowels by early and late Korean and Japanese bilinguals. *Studies in Second Language Acquisition, 28*, 487-513.

Li, F., Edwards, J., & Beckman, M. (2009). Contrast and covert contrast: The phonetic development of voiceless sibilant fricatives in English and Japanese toddlers. *Journal of Phonetics, 37*, 111-124.

McGowan, R. & Nittrouer, S. (1988). Differences in fricative production between children and adults: Evidence from an acoustic analysis of /ʃ/ and /s/. *Journal of Acoustical Society of America, 83*, 229-236.

Munro, M. J., & Derwing, T. M. (1998). The effects of speaking rate on listener evaluations of native and foreign-accented speech. *Language Learning, 48*, 159-182.

Munro, M. J., Derwing, T. M., & Morton, S. L. (2006). The mutual intelligibility of L2 speech. *Studies in Second Language Acquisition, 28*, 111-131.

Nittrouer, S. (2002). Learning to perceive speech: How fricative perception changes, and how it stays the same. *Journal of Acoustical Society of America, 112*, 711-719.

Nittrouer, S., Studdert-Kennedy, M., & McGowan, R. S. (1989). The emergence of phonetic segments: Evidence from the spectral structure of fricative-vowel syllables spoken by children and adults. *Journal of Speech and Hearing Research, 32*, 120-132.

Piske, T., MacKay, I. R. A., & Flege, J. E. (2001). Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics, 29*, 191-215.

Prather, E., Hedrick, D., & Kern, D. (1975). Articulation development in children aged two to four years. *Journal of Speech and Hearing Disorders, 403*, 179-191.

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

Seo, M., & Lim, J. (2010). L2 production of non-contrastive sounds in L1. *English Language and Literature Teaching, 16*, 129-146.

Smit, A. B., Hand, L., Freilinger, J. J., Bernthal, J. E., & Bird, A. (1990). The Iowa articulation norms project and its Nebraska replication. *Journal of Speech and Hearing Disorders, 55*, 779-798.

Sung, E.-K. (2006). L2 sound perception and production by Korean adults and children. *Studies in Phonetics, Phonology and Morphology, 12*, 577-596.

Templin, M. (1957). *Certain language skills in children*. Minneapolis: University of Minnesota Press.

Zampini, M. L. (2008). L2 speech production research: Findings, issues and advances. In J. G. Hansen Edwards & M. L. Zampini (Eds.), *Phonology and second language acquisition* (pp. 219-249). John Benjamins Publishing Company: Amsterdam.

APPENDIX

Stimuli

	[ʒ]	[z]	[dʒ]
Word-initial		zealous	jealous
		zest	jest
		zam	zam
		Z	G
		zoo	Jew
Intervocalic		buzzing	budging
		gazing	gaging
		reason	region
	Seizure	Caesar	
	Composure	composer	
	Erasure	eraser	
Word-final	Beige	baize	
	Rouge	rues	
		chains	change
		raise	rage
		seize	siege
		frizz	fridge
		fuzz	fudge

Examples in: English
Applicable Languages: English
Applicable Levels: Secondary

Misun Seo
Division of English Language & Literature
Hannam University
133 Ojeong-dong, Daedeok-gu
Daejeon, 306-791, Korea
Tel: (042) 629-8461
Email: misunseo@hnu.kr

Jayeon Lim
Dept. of English Language & Literature
University of Seoul
90 Siripdae-gil, Dongdaemun-gu
Seoul, 130-743, Korea
Tel: (02) 2210-5692
Email: limjy@uos.ac.kr

Received in October, 2011
Reviewed in November, 2011
Revised version received in December, 2011