영어어문교육, 14권 3호 2008년 가을

Factors Affecting Korean Learners' Perception of English Sibilants*

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

Lim, Jayeon & Seo, Misun. (2008). Factors affecting Korean learners' perception of English sibilants. *English Language & Literature Teaching, 14*(3), 201-219.

This study explores the perception of English sibilants by Korean learners. According to Schmidt (1996) and our pilot study, Koreans usually map all three English voiced sibilants (e.g., /z, ʒ, dʒ/) to one Korean sound (e.g., /ㅈ/) while perceiving voiceless counterparts (e.g., /s, ʃ, tʃ/) as several different native sounds (e.g., /ㅅ, ㅆ, ㅊ/). The study conducted an experiment of identifying English minimal pairs with sibilants to investigate the influence of such mapping relations between English and Korean sibilants and that of proficiency levels. In addition, positions of sibilants were included as another variable. The results showed that learners performed better in the identification task of English voiceless sibilants by showing higher accuracy and shorter reaction times than in distinguishing voiced ones. Learners' perception of sibilants was also influenced by positions of sibilants within words. Thus, the general pattern was that word-final sibilants caused less difficulty in the identification task than intervocalic or word-initial sibilants due to longer duration of word-final sibilants over word-initial or intervocalic ones. Learners' accuracy and reaction times did not improve according to their overall proficiency.

[English sibilants/L2/perception]

I. INTRODUCTION

When adult L2 learners perceive speech, they are often strongly influenced by

---

* An earlier version of this paper was presented at the 13th Pan-Pacific Applied Linguistic Conference held at University of Hawaii at Manoa, August 20-22, 2008.

** first author *** corespondent author

experience with the phonological system of their L1 (Abramson & Lisker, 1970; Best et al., 2001, among many others). However, the results of L1 influence are not always clear-cut. Differences between L1 and L2 cause difficulty among adult L2 learners. That is, when learners are faced with an L2 speech sound with no equivalent sound in the L1 sound system, they would experience more difficulty than when faced with an L2 speech sound that resembles an L1 sound system. Traditionally, this has been explained by L1 transfer (Odlin, 1989). Therefore, adult L2 learners will rely on their L1 system at all levels of language.

However, studies have shown that adult L2 learners do not uniformly manifest difficulty whenever they find differences between L1 and L2. Certain L2 sounds create no learning problems even when these sounds are first encountered by learners, whereas other sounds may pose difficulties. For example, Best, McRoberts and Sithercole (1988) have reported that adult L1 speakers of English showed no difficulty in discriminating Zulu click consonants with no previous exposure to such consonants. In a similar vein, Flege (1995) claims that learners are able to establish a new phonetic category in L2 when two sounds are distant.

If the differences between L1 and L2 system are not always clearly explained in terms of the sound differences between the two languages, we need to speculate other factors involved in creating the differences. Studies in L2 perception have pointed to many different phonological factors. For example, English /r/ and /l/ appear to be especially difficult to Japanese learners because these liquids do not map well onto any of their L1 category(s) (Aoyama et al., 2004; Best & Strange, 1992; MacKain et al., 1981, among others). The likelihood of hearing two English sounds in terms of one Japanese sound may cause the difficulty of Japanese learners of English. Likewise, English /r/ and /l/ give trouble to Korean learners of English since [r] and [l] are not contrastive sounds in Korean (Borden et al., 1983, among others).

Additionally, phonetic factors are also known to be one of the contributing factors in speech perception. For example, acoustic cues influenced by positions of a sound within a word may influence perception. For example, it is generally argued that onsets are more perceptible than codas since onsets occurring before a vowel have more robust acoustic cues than codas occurring after a vowel (Ohala, 1992; Wright, 1996, among others).

In addition to phonetic and phonological factors, researchers have also focused on the influence of 'learning'. For example, Flege (1995) argues that even adult learners can establish new phonetic categories in L2, as long as they can detect the differences between L1 and L2. He goes on to claim that basic speech learning mechanisms, which include the ability to establish long-term memory representations for speech sounds, stay undamaged throughout life. In support of Flege's claim, Flege, Takagi and Mann (1995) demonstrated that the acquisition of native-like proficiency of consonants is possible with increased L2

experience. Additionally, Bongaerts, van Summeren, Planken and Schils (1997) studied the highly advanced late learners and found that five of eleven learners were rated by native English judges to fall within the range of native speakers. The results have often been cited to support the claim that L2 perception can show progress.

A goal of this study is to investigate the perception of English sibilants /s, ʃ, tʃ, z, ʒ, dʒ/ by native speakers of Korean by running experiments. It is expected that the perception of English sibilants would create problems for Korean learners of English considering differences in the phonological systems of the two languages. Most of all, only voiceless fricatives and affricates /ㅅ, ㅆ, ㅈ, ㅉ, ㅊ/ (i.e. /s, s', tɕ, tɕ', tɕ^h/) exist in the Korean language with voiced counterparts occurring as allophones between voiced sounds. Thus, unlike in English, voicing is not contrastive in Korean fricatives and affricates. Instead, tenseness and aspiration are used as contrastive features in those sound groups. Since Korean has no contrastive voiced fricatives and affricates, Korean learners of English are particularly expected to undergo difficulties in distinguishing three English voiced sibilants. This prediction is also based on Schmidt (1996) and our pilot study with 29 Korean learners of English. According to Schmidt, word-initial English sibilants occurring in three vowel contexts of /i, a, u/ are labeled in Korean consonants by Korean learners as follows:

TABLE 1

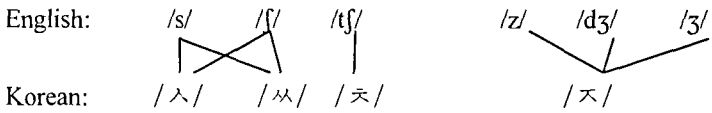
Percentage Labeling of Word-initial English Sibilants as Korean Consonants (adapted from Schmidt, 1996)

	English Sibilants				
Korean	/s/	/ʃ/	/tʃ/	/z/	/dʒ/
/ㅅ/	32	51			
/ㅆ/	67	48			
/ㅈ/			6	94	80
/ㅉ/			2	3	19
/ㅊ/			91		

As can be seen from Table 1, English voiced sibilants /z/ and /dʒ/ were predominantly labeled as the Korean consonant /ㅈ/ by Koreans while voiceless sibilants /s/ and /ʃ/ were labeled as /ㅅ/ or /ㅆ/ and /tʃ/ mostly as /ㅊ/.

In Schmidt, only word-initial consonants were considered and /ʒ/ was not included in the labeling experiment. Thus, to see how /ʒ/ is labeled and how English sibilants are labeled in different positions within words, we ran a pilot study of labeling all six English sibilants occurring in word-initial (#_V), intervocalic (V_V) and word-final (V_#) positions in three vowel contexts of /i, a, u/ by using recordings made by an English native speaker. The results of our pilot study were similar to those of Schmidt; additionally it was

found /ʒ/ was labeled as /ㅈ/ by Korean learners of English. Based on Schmidt and our pilot study, the mapping relations between English sibilants and Korean correspondents are represented as follows:



Since all three English voiced sibilants are mapped to the Korean consonant /ㅈ/, it is expected that Korean learners of English will experience difficulty in differentiating the three voiced sibilants. On the other hand, considering the mapping relations between English voiceless sibilants and Korean corresponding sounds, Korean learners are expected to face less difficulty in distinguishing them.

In addition to the influence of voicing and positions of English sibilants, this study also attempts to explore whether learners show different perception patterns according to their proficiency levels. That is, this study focuses on whether or not L2 perception would be influenced by various factors such as voicing, proficiency levels of learners, and positions of consonants to be perceived. In doing so, the study is set out to measure accuracy and reaction time of L2 learners by running a perception experiment. To this end, the study explores the following research questions:

(1) In L2 perception of English sibilants, would learners be influenced by voicing of sibilants in words?
(2) In L2 perception of English sibilants, would learners be influenced by their proficiency levels in English?
(3) In L2 perception of English sibilants, would learners be influenced by positions of sibilants in words?

II. METHOD OF RESEARCH

1. Participants

The perception experiment was run with 30 Korean learners of English. Among 30 Koreans, 10 (five males, five females) were a beginner group, 10 (four males, six females) were an intermediate group and 10 (three males, seven females) were an advanced group of learners. The beginner group of English learners was the ones who had TOEIC scores of 500's or below. The intermediate group of English learners was recruited from the ones

who got 700's TOEIC scores. Korean learners of English with 900's TOEIC scores were grouped as an advanced level. All learners were students at a certain University in Seoul, and they were paid for their participation.

2. Materials

For the perception experiment, 47 sets of minimal pairs with different sibilants in word initial, medial and final positions were selected. Eighteen pairs of them were minimal pairs with different voiced sibilants (e.g. *zealous – jealous, raise – rage, erasure – eraser*, etc.) and twenty-nine of them had different voiceless sibilants (*shell – cell, hushes – hutches, peace – peach*, etc.). A total of 91 words with sibilants were selected for our perception experiment. Table 2 illustrates the number of minimal pairs used in our experiment, classified by segment types and positions.

TABLE 2

Stimuli for the Perception Experiment: the Number of Minimal Pairs

	[z] – [dʒ]	[z] – [ʒ]	[s] – [ʃ]	[s] – [tʃ]	[ʃ] –[tʃ]
word-initial	5	0	5	4	5
Intervocalic	3	3	1	1	1
Word-final	5	2	3	3	6

Since no minimal pair was found for [dʒ] – [ʒ] from an online dictionary search, no words were included for that pair of segment type (For a full list of words used in our perception experiment, see Appendix A.).

Stimuli recordings were made by a female Canadian in a quiet room. The speaker has lived in Korea for 2.5 years. For the recordings, a head-mounted microphone (Audio-Technica M8531) and a TASCAM HD-P2 recorder were used. Recordings were made with the setting of 44100 Hz and 16 bits.

In addition to 91 target words containing sibilants, 39 words including other consonants such as stops or liquids were also recorded to use them as fillers in the perception experiment. For a recording, each word was put in a carrier sentence "please say _____". The speaker was asked to read each sentence at a natural speech rate three times. Sentences were randomized at each recording. Thus, sentences were presented to the speaker in different orders each time of a recording. The last recording was taken and used as the stimuli in the perception experiment.

3. Procedure

The task of the participants was to identify the member of a minimal pair they heard. For example, after listening to the sentence 'please say zest', listeners were presented with the two words 'zest' and 'jest' on a computer screen and they were asked to respond whether they heard the word 'zest' or 'jest' by clicking on a button. Participants were tested in front of a computer one at a time in a quiet room. The stimuli were presented to the participants through a headphone (SONY MDR-7505), using the *E-prime 1.0* program installed on a laptop computer. The interval between the stimuli was 100 ms. For the familiarization of the test, before starting the main session, a trial session was run with 5 words containing no sibilant. After the trial session, participants were asked if they had any questions about the procedure and proceeded with the main session. Accuracy rate and reaction time were recorded to use as an indicator of ease of perception in the analysis. Reaction time was measured from the onset of the sibilant of the target word in each stimulus. Participants were instructed to look at the computer screen during the experiment and to respond as accurately and quickly as possible. Participants heard 91 sentences including target words and 39 fillers once. The entire procedure took approximately 20 minutes. The stimuli were played in a randomized order to each listener and at a constant volume level.

4. Predictions

Korean learners of English at all levels are expected to show lower accuracy in identifying voiced sibilants as opposed to voiceless ones since voiced sibilants are mapped to one Korean consonant while voiceless sibilants are mapped to several different Korean consonants. Furthermore, it is predicted that it will take more time to distinguish voiced sibilants than voiceless ones under the assumption that similar sounds will be harder to discriminate. In addition, learners are expected to perform better and faster as their proficiency levels increase. In terms of positions of sibilants in English words, learners are expected to perform better in word-initial and intervocalic sibilants (i.e. onset sibilants) than in word-final sibilants (i.e. coda sibilants) since onsets are thought to be more perceptible than codas by containing more robust cues.

III. RESULTS

The results were analyzed in two parts: accuracy and reaction time. For the analyses, accuracy scores were calculated and reaction time was measured. Upon running a statistical analysis, following Duffield and White (1999), reaction times with above and

below two standard deviations of a particular participants' personal mean were corrected to the corresponding cut-off values. In the reaction time analysis, reaction times of only correct responses were included. For the statistical analysis, three-way ANOVAs were run with accuracy and reaction time as dependent variables respectively. In both analyses, proficiency levels of learners, voicing and positions of sibilants within words were treated as independent variables. Further, post-hoc analyses were run with proficiency levels and positions.

1. Accuracy

As illustrated in Table 3 below, there were significant main effects for voicing, proficiency levels and positions. In addition, there were significant interactions between voicing and proficiency, and also between voicing and positions. There were no significant interactions between proficiency and positions, or either among voicing, proficiency and positions[1].

TABLE 3

ANOVA Table for Accuracy

Source	SS	df	MS	F	p
Corrected Model	23.183	17	1.364	20.625	.000*
Intercept	1652.55	1	1652.55	24994.19	.000*
Voicing	13.958	1	13.958	211.104	.000*
Proficiency	1.833	2	.917	13.862	.000*
Position	2.022	2	1.011	15.288	.000*
Voicing * proficiency	.878	2	.439	6.639	.001*
Voicing * position	1.074	2	.0537	8.121	.000*
Proficiency * position	.117	4	.029	.444	.777
Voicing * proficiency * position	.381	4	.095	.1.441	.218
Error	175.343	2652	.066		
Total	2454	2670			

*p < .05

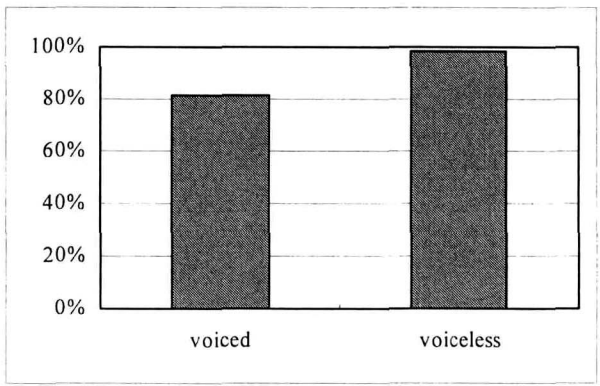
With regards to voicing, words with voiceless sibilants showed higher accuracy than those with voiced sibilants. As illustrated in Figure 1, the mean for voiceless sibilant words

[1] Although significant interactions were found between factors (voicing * proficiency and voicing * position), we will analyze main effects for voicing, proficiency and position since they carried meaningful information.

was 98.2% while the one for voiced was 81.8%. Interestingly, the standard deviation for the voiceless was 0.13 whereas the one for the voiced was much larger at 0.39.

FIGURE 1

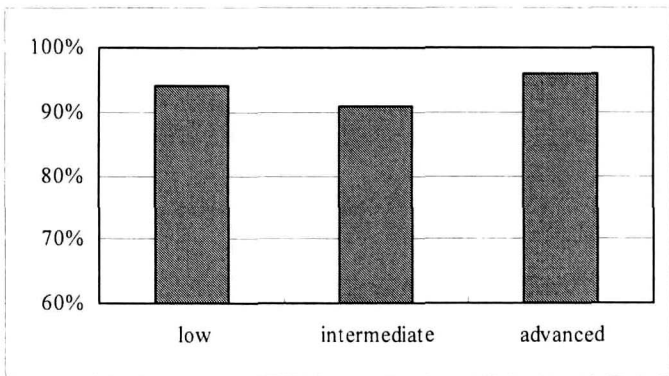
Accuracy Percentage for Voiced vs. Voiceless Sibilant Words



Learners across all levels exhibited fairly accurate results. However, there was a statistically significant main effect. A post-hoc analysis of Bonferroni showed significant differences between the following groups: low vs. intermediate ($p < .05$) and intermediate vs. advanced ($p < .05$). Unlike our prediction, there was no significant difference between low vs. advanced groups ($p = .196$). Figure 2 illustrates the finding. Advanced level learners (94.5%) exhibited the highest accuracy, followed by low (92.3%) and intermediate (89.0%) learners.

FIGURE 2

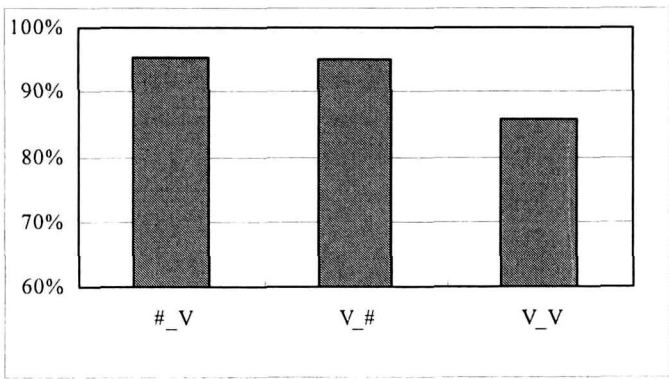
Accuracy Percentage by Proficiency Levels



Learners also exhibited a significant main effect for position of sibilants in words. A post-hoc analysis of Bonferroni showed significant differences between the following groups: word initial prevocalic (#_V) vs. intervocalic (V_V) positions ($p < .05$) and word final postvocalic (V_#) vs. V_V ($p < .05$) positions. There was no significant difference between #_V vs. #V_ ($p = 1.0$). Figure 3 shows learners showed a similarly high accuracy with both #V_(94.3%) and _V# (93.6%). V_V showed the lowest accuracy at 82.5%.

FIGURE 3

Accuracy Percentage by Positions of Sibilants in Words

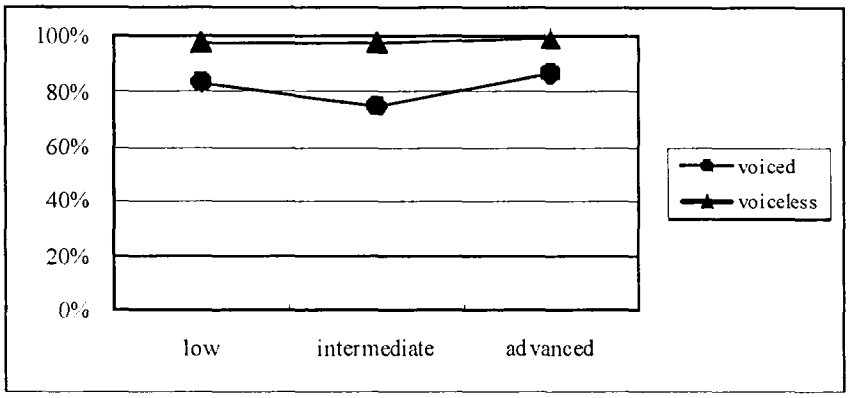


As mentioned above, there was a significant interaction of voicing and proficiency. The interaction is illustrated in Figure 4 below. As can be seen from the figure, learners were significantly better at identifying voiceless sibilants than voiced ones. With voiced, advanced level learners showed the highest accuracy (86.8%), followed by low (83.5%) and intermediate (75%) level learners. With voiceless, however, learners at all levels showed a similarly high accuracy (99.3%) for advanced and 97.6% for both intermediate and low).

Statistical analyses of ANOVAs revealed that a significant difference existed among different proficiency levels with voiced sibilant words ($F = 8.539$, $p < .05$). However, there was no significant difference among different proficiency levels with voiceless sibilant words ($F = 2.754$, $p = .064$). In case of voiced, a post-hoc analysis of Bonferroni test revealed that the significant differences existed between low and intermediate ($p < .05$) and between intermediate and advanced ($p < .05$). There was no statistical difference between low and advanced ($p = .815$).

FIGURE 4

Accuracy Interaction Plot for Voicing and Proficiency

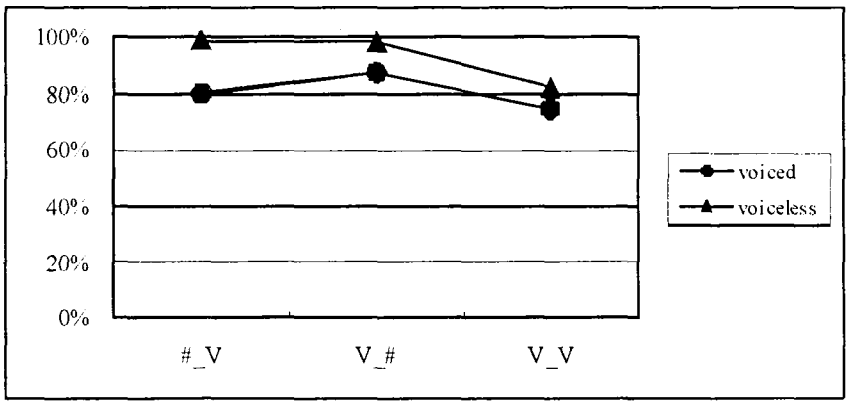


To test for differences between voiced and voiceless in each proficiency level, a series of of paired t-tests were performed. The results indicated significant differences existed between voiced and voiceless in all three proficiency levels ($t = -6.662$, $p < .05$ for low; $t = -8.881$, $p < .05$ for intermediate; $t = -6.469$, $p < .05$ for advanced).

Next significant interaction was observed with voicing and position. Again, the plot is provided below as Figure 5. Overall, learners showed different patterns with voiced vs. voiceless. First of all, as we can see from the figure, learners' accuracy with voiced was highest with V_# (87.6%), followed by #_V (80.3%) and V_V (75.0%). However, voiceless showed similar accuracy in #_V (99.0%) and V_# (98.2%). V_V position showed the lowest accuracy (82.5%).

FIGURE 5

Interaction Plot for Voicing and Position



As can be seen from the figure, in all positions, learners were significantly better at identifying voiceless sibilants than voiced ones ($t = -7.417$, $p < .05$ for #_V; $t = -5.813$, $p < .05$ for V_#; $t = -5.364$, $p < .05$ for V_V). Further analyses of ANOVAs on voicing in three different positions revealed significant differences within voiced ($F = 9.794$, $p < .05$) and within voiceless ($F = 6.328$, $p < .05$). A post-hoc analysis of Bonferroni on voiced sibilant words revealed significant differences existed between #_V and V_# ($p < .05$). No significant difference was found between #_V and V_V ($p = .265$). With voiceless, significant differences existed between #_V and V_V ($p < .05$) and between V_# and V_V ($p < .05$). No significant difference was observed between #_V and V_# ($p = .864$).

2. Reaction Time

As shown in Table 4 below, there were significant main effects for voicing, proficiency level and position. In addition, there was a significant interaction between voicing and position. Interactions between voicing and proficiency, proficiency and position and among voicing, proficiency and position were not significant.

TABLE 4

ANOVA Table for Reaction Time

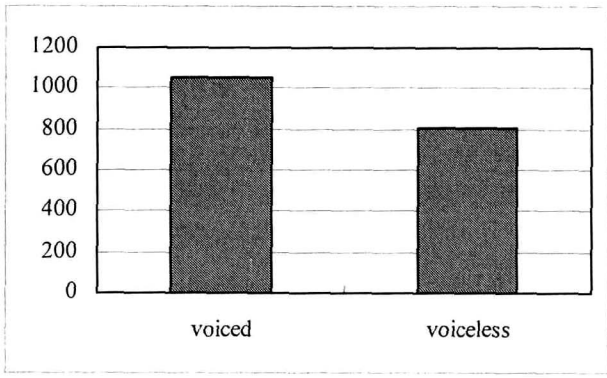
Source	SS	df	MS	F	p
Corrected Model	53959986	17	3174117	18.696	.000*
Intercept	1319594595	1	1319594595	9187.972	.000*
Voicing	15252848	1	15252848	89.843	.000*
Proficiency	1163550	2	581775.2	3.427	.033*
Position	23372352	2	11686176	68.835	.000*
Voicing * proficiency	254689.0	2	127344.5	.750	.472
Voicing * position	1957308	2	978654.1	5.765	.003*
Proficiency * position	547068.7	4	136767.2	.806	.521
Voicing * proficiency * position	509163.9	4	127291.0	.750	.558
Error	420856827	2436	169772.0		
Total	2333363958	2454			

*$p < .05$

With regards to voicing, words with voiceless sibilants yielded shorter reaction time than those with voiced sibilants. As illustrated in Figure 6, the mean reaction time for voiceless sibilant words was 804.8 ms, while the one for voiced was 1049.1 ms.

FIGURE 6

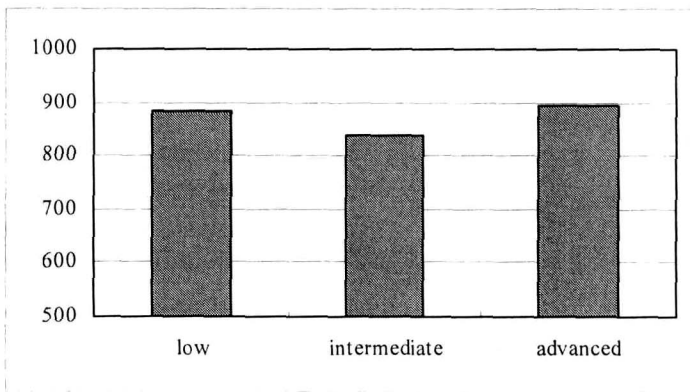
Mean Reaction Times of Voiced vs. Voiceless Sibilant words



As with proficiency level, there was a significant main effect. A post-hoc analysis of Bonferroni showed significant differences between the following groups: low vs. intermediate ($p < .05$) and intermediate vs. advanced ($p < .05$). However, unlike our prediction, and in par with the results of accuracy, there was no significant difference between low vs. advanced groups ($p = .582$). Figure 7 illustrates the finding. Advanced level learners exhibited the longest mean reaction time of 894.1 ms, followed by low level learners at 882.9 ms, and intermediate learners at 837.2 ms.

FIGURE 7

Mean Reaction Times by Proficiency Levels of Learners

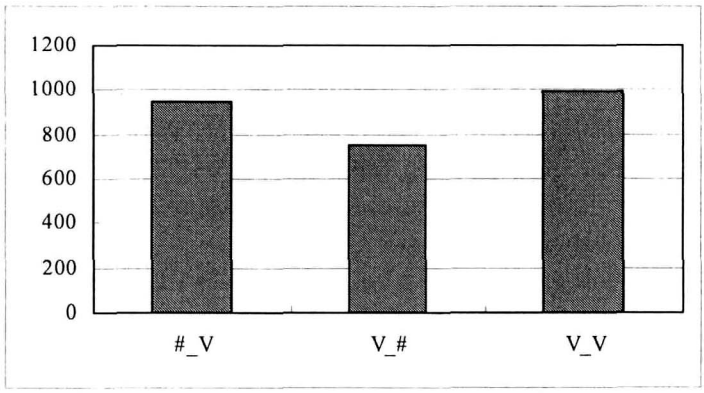


Learners also exhibited a significant main effect for position of sibilants in words. A post-hoc analysis of Bonferroni showed differences between the following groups: #_V vs. V_# ($p < .05$) and V_V vs. V_# ($p < .05$). There was no significant difference between #_V vs. V_V ($p = .077$). Figure 8 shows learners showed similarly long reaction times with

both #_V (951.2 ms) and V_V (994.4 ms). V_# showed the shortest reaction time at 754.4 ms.

FIGURE 8

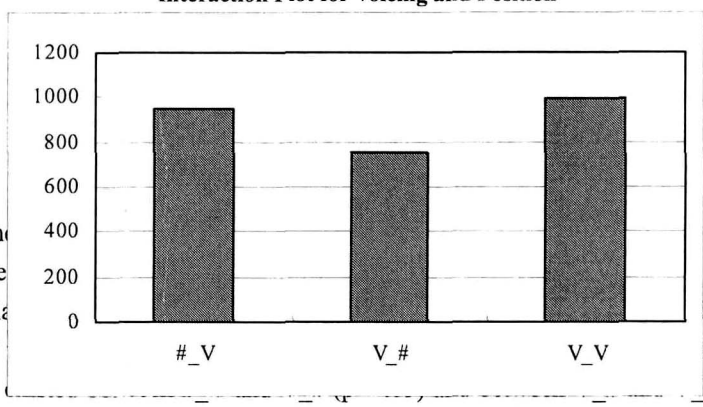
Mean Reaction Time by Position of Sibilants in Words



As mentioned above, there was a significant interaction between voicing and position. The interaction is illustrated in Figure 9 below.

FIGURE 9

Interaction Plot for Voicing and Position



For furth voiceless. It was reveale words with voiced sibil 575, $p < .05$). In case of significant differences V ($p < .05$). There was no statistical difference between #_V and V_V ($p = .068$). In case of voiceless, the same pattern was observed (i.e., #_V vs. V_#, $p < .05$; V_# vs. V_V, $p < .05$; #_V vs. V_V, $p = .222$).

To test for differences between voiced and voiceless in each position, a series of paired t-tests were performed. The results indicated significant differences between voiced and

voiceless within #_V ($t=5.586$, $p < .05$) and V_# ($t=4.825$, $p < .05$) respectively. However, no significant differences were observed within V_V ($t=1.887$, $p = .061$).

3. Summary of Results

In summary, voiceless sibilant words showed higher accuracy with shorter reaction time. Differences between proficiency levels were mainly observed within voiced sibilant words. With proficiency levels, low and advanced learners showed a similar pattern: higher accuracy with shorter reaction time. Intermediate learners showed the lowest accuracy with the longest reaction time. With positions, higher accuracy was observed with #_V and V_# as opposed to V_V. Shorter reaction time was correlated with V_#, followed by similarly longer reaction time with V_V and #_V.

IV. DISCUSSION

As with our prediction, Korean Learners of English showed different patterns in identifying voiced vs. voiceless sibilants in English. They showed higher accuracy and shorter reaction time with voiceless than with voiced. The results are in accordance with the mapping relations between English and Korean sibilants. Since all three English voiced sibilants /z, ʒ, dʒ/ are mapped to one Korean consonant /ㅈ/ according to Schmidt (1996) and our pilot study, learners in the study showed more difficulty (i.e. lower accuracy and longer reaction time) in the identification task. On the other hand, English voiceless sibilants, which are mapped to several different Korean consonants, did not pose much problem to learners. As mentioned in the results, learners' standard deviation for the voiced was higher by three times than that for the voiceless although the accuracy of the voiced was lower. This implies that there were much more individual differences in the responses for the voiced.

Next, learners across levels produced a fairly high accuracy. Still, there was a significant main effect of proficiency levels in both accuracy and reaction time analyses. This statistical difference was mainly due to the intermediate level with the lowest accuracy and the shortest reaction time. Thus, learners did not show a developmental progress in our study: low and advanced showed similarly high accuracy and similarly long reaction time. Although advanced learners did show progress when compared to intermediate, it is not likely to say that developmental progress across levels occurred. This is due to the fact that their reaction time is longer than that of intermediate learners. In L2 studies, it is normal to witness more advanced learners to show shorter reaction times in accordance with higher accuracy. However, in our study, shorter reaction times were not correlated to higher

accuracy. With the given data, it is not possible to say that L2 learning occurred, as Flege (1995) had claimed. This is in accordance with Piske, Mackay and Flege (2001) where the researchers supported the conclusion of Flege and Fletcher (1992) in that L2 learning proceeds at a faster rate in the initial learning period. It is suggested that most improvements in L2 pronunciation occur during the early stages of exposure to the L2. Thus, it is possible that learners in the present study, specifically when they are already past the early stages of being exposed to English, were not able to show conspicuous learning effects. Another reason may be due to the fact that learners were not exposed to focused instruction on the target structures. To substantiate the claim, a future study is needed with a carefully designed experiment providing a focused instructional input to L2 learners.

Finally, Korean learners' perception of English sibilants was influenced by positions of sibilants within words. Unlike our prediction that learners will perform better for onset sibilants (#_V and V_V) than coda sibilants (V_#), they identified voiceless sibilants in #_V and V_# more accurately than in V_V. In case of voiced sibilants, a similar pattern was found: sibilants in V_# and #_V were identified more accurately than those in V_V. Learners' reaction times were longer for both voiced and voiceless sibilants in V_V and #_V than in V_#. These results might be explained in terms of different durations of sibilants in the three positions. The acoustic analysis of perceptual stimuli used in our experiment showed that, regardless of voicing, sibilants' duration is the longest in V_# (268.7 ms for voiceless, 202.5 ms for voiced) followed by #_V (183.0 ms for voiceless, 137.6 ms for voiced) and V_V (130.0 ms for voiceless, 97.9 ms for voiced). Since sibilants in V_V have the shortest duration, they may have induced lower accuracy and longer reaction time. On the other hand, sibilants in V_# with the longest duration resulted in higher accuracy and shorter reaction time.

One more thing to note regarding the interaction of voicing and position is that there is no reaction time difference between voiced and voiceless sibilants in V_V as opposed to #_V and V_#. That is, voiceless sibilants showed shorter reaction times than voiced ones in #_V and V_#. According to the acoustic analysis of perceptual stimuli used in our experiment, voiceless sibilants have longer duration than voiced ones both in #_V and V_#. It is considered that such duration difference caused shorter reaction times of voiceless sibilants than voiced ones in the two positions. However, voiced and voiceless sibilants show similar reaction times in V_V although voiceless ones (130.0 ms) have longer duration than voiced ones (97.9 ms). We speculate this result to be due to duration of voiceless sibilants in V_V not being long enough to perceptually enhance the presence of voiceless sibilants.

The data provided from the study does not allow us to comprehensively explain the L2 perception in general. However, we are able to offer some implications for future research. We speculate that the effect of L1 is present in our study. To further substantiate our

speculation, we need a detailed assessment regarding perceived phonetic distance through L1-L2 mapping experiments. Along with this, we need to obtain the relevant identification data. Up to now, only a few studies offer both mapping and identification data (cf. Guion et al., 2000).

Next, we need to consider effects of word frequency. It is possible that learners might have been exposed to certain members in minimal pairs, because those particular members are simply used more frequently. If this is true, learners may be inclined to choose more frequent members over less frequent ones. Along with this, we also need to consider learners' semantic knowledge of target words. Although learners' overall proficiency in L2 was included to be a factor in the study, the results might have been influenced by learners' knowledge of the target words. Thus, for a future study, we need to include a semantic test that measures learners' knowledge on the words to be tested.

In addition, TOEIC scores were used in this study to divide learners into different proficiency levels. Though this standardized exam is currently one of the most widely used one in Korea, but because of its extensive employment, learners might have shown a 'test effect' in their scores: the scores might not have fully represented their genuine proficiency but instead reflected their familiarity with the exam.

V. CONCLUSION

This study examined various factors affecting L2 perception of English sibilants by adult Korean speakers. It was revealed that voicing is a prominent factor influencing the perceptual process. The learners in the study had difficulty in identifying voiced sibilant words. We speculated that different mapping relations in voiced vs. voiceless between Korean and English sibilants are correlated with the difficulty. In addition to voicing, positions of sibilants in words triggering different duration patterns affected L2 perception. Proficiency levels of learners were not a determining factor in L2 perception process, implying that there is no developmental progress in L2 perception of English sibilants in the study.

REFERENCES

Abramson, A., & Lisker, L. (1970). Discriminability along the voicing continuum: Cross language tests. *Proceedings of the 6th International Congress of Phonetic Sciences* (pp. 569-573). Prague: Academia.

Aoyama, K., Flege, J., Guion, S., Akahane-Yamada, R., & Yamada, T. (2004). Perceived

phonetic dissimilarity and L2 speech learning: The case of Japanese /r/ and English /l// and /r/. *Journal of Phonetics, 32*, 233-250.

Best, C., McRoberts, G., & Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *Journal of Acoustical Society of America, 109*, 775-794.

Best, C., & Strange. W. (1992). Effects of phonological and phonetic factors on cross-language perception of approximants. *Journal of Phonetics, 20*, 305-350.

Best, C., McRoberts, G., & Sithercole, N. (1988). The phonological bases of perceptual loss for non-native contrasts: maintenance of discrimination among Zulu clicks by English-speaking adults and infants. *Journal of Experimental Psychology, 14*, 345-360.

Borden, G., Gerber, A., & Milsark, G. (1983). Production and perception of the /r/-/l/ contrast in Korean adults learning English. *Language Learning, 33*, 499-526.

Bongaerts, T., van Summeren, C., Planken, B., & Schils, E. (1997). Age and ultimate attainment in the pronunciation of a foreign language. *Studies in Second Language Acquisition, 19*, 447-465.

Duffield, N., & White, L. (1999). Assessing L2 knowledge of Spanish clicitic placement: Converging methodologies. *Second Language Research, 15*, 133-160.

Flege, J. (1995). Second-language speech learning: Theory, findings and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 171-204). Timonium, MD: York Press.

Flege, J. E., & Fletcher, K. L. (1992). Talker and listener effects on the degree of perceived foreign accent. *Journal of the Acoustical Society of America, 91*(1), 370-389.

Flege, J., Takagi, N., & Mann, V. (1995). Japanese learners can learn to produce English /r/ and /l/ accurately. *Language and Speech, 38*, 25-55.

Guion, S., Flege, J., Akahane-Yamada, R., & Pruitt, J. C. (2000). An investigation of current models of second language speech perception: The case of Japanese adults' perception of English consonants. *Journal of Acoustical Society of America, 5*, 2711-2724.

McKain, K., Best, C., & Strange, W. (1981). Categorical perception of English /r/ and /l/ by Japanese bilinguals. *Applied Psycholinguistics, 2*, 369-390.

Odlin, T. (1989). *Language transfer: Cross-linguistic influence in language learning*. Cambridge: Cambridge University Press.

Ohala, J. (1992). Alternatives to the sonority hierarchy for explaining segmental sequential constraints. *Chicago Linguistics Society: Papers from the parasessions on the syllable* (pp. 319-338). Chicago: CLS.

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

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

Wright, R. (1996). *Consonant clusters and cue preservations in Tsou.* Unpublished doctoral dissertation. University of California at Los Angeles, Los Angeles.

APPENDIX A

Perceptual Stimuli: Words with voiced sibilants

	[ʒ]	[z]	[dʒ]
Word-initial		zealous	jealous
		zest	jest
		zam	jam
		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

Perceptual Stimuli: Words with voiceless sibilants

	[ʃ]	[s]	[tʃ]
Word-initial	shell	cell	
	shack	sack	
	ship	sip	
	shoot	suit	
		sane	chain
		sat	chat
		sin	chin
	sherry		cherry
	sheaf		chief
	sheer		cheer
	shore		chore
	shoe	sue	chew

Intervocalic	messing	meshing	
		sensory	century
	hushes		hutches
Word-final	furnish	furnace	
	mesh	mess	
	plush	plus	
		peace	peach
		piss	pitch
		stress	stretch
	wish		witch
	swish		switch
	mash		match
	crush		crutch
	wash		watch
	bush		butch

Examples in: English
Applicable Language: English
Applicable Levels: College

Jayeon Lim
Dept. of English Language & Literature
University of Seoul
90 Siripdae-gil, Dongdaemun-gu
Seoul, 130-743, Korea
Tel: (02) 2210-5692 /010-3203-1685
Email: limjy@uos.ac.kr

Misun Seo
Division of English Language & Literature
Hannam University
133 Ojeong-dong, Daedeok-gu
Daejeon, 306-791, Korea
Tel: (042) 629-8461/010-3078-7865
Email:misunseo@hnu.kr

Received in July, 2008
Reviewed in August, 2008
Revised version received in September, 2008