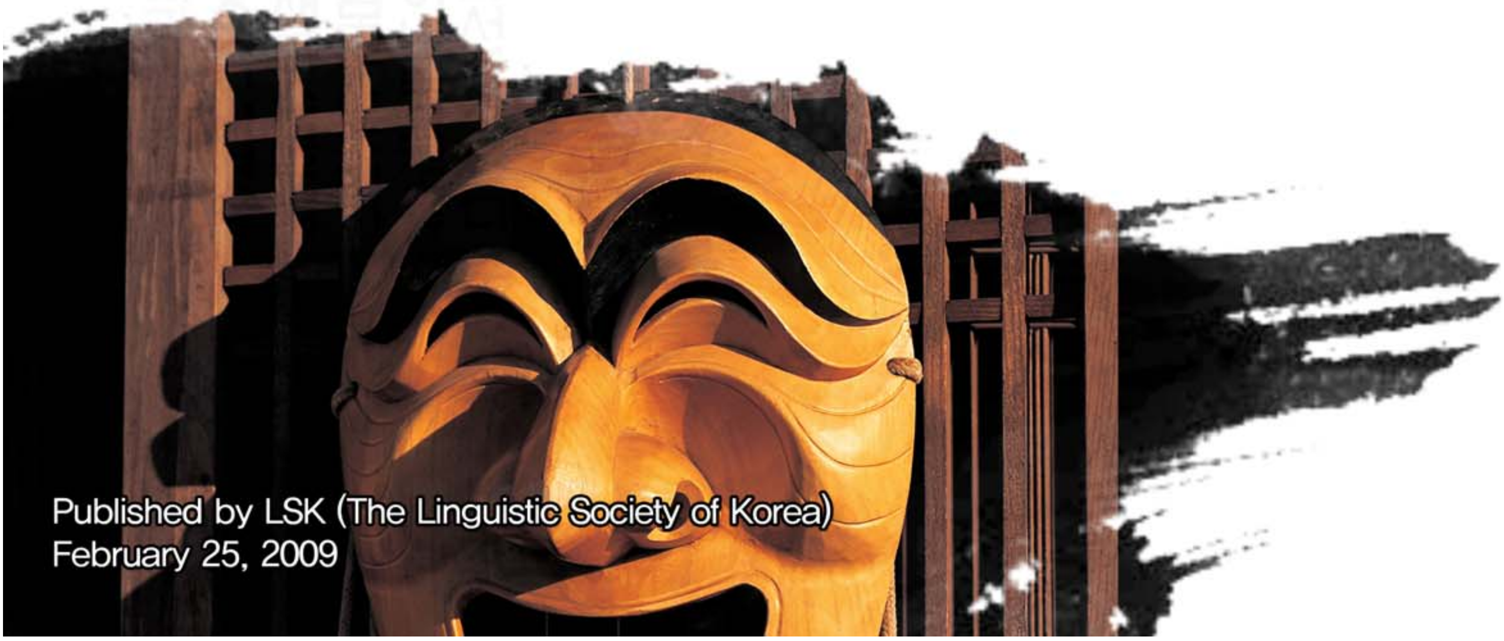
The Linguistic Society of Korea

CIL XVIII July 21~26, 2008 Seoul, Korea
www.cil18.org

Current Issues in Unity and Diversity of Languages

Collection of the papers selected from the CIL 18,
Held at Korea University in Seoul,

on July 21~26, 2008



Published by LSK (The Linguistic Society of Korea)
February 25, 2009

Current Issues in Unity and Diversity of Languages

Collection of the papers sellected from the CIL18,
Held at Korea University in Seoul

Published by LSK (The Linguistic Society of Korea)
February 25, 2009
Printed by Dongnam Publishing Co., Republic of Korea
ISBN 978-89-90696-71-7 98040

The Linguistic Society of Korea

Perception versus Production in Korean L2 Acquisition of English Sibilant Fricatives

Robert Allen Fox,[1] Ewa Jacewicz,[1] Fred R. Eckman,[2] Gregory K. Iverson,[3] Soyoung Lee[2]

[1]The Ohio State University, [2]University of Wisconsin-Milwaukee, [3]University of Maryland Center for Advanced Study of Language

Department of Foreign Languages and Linguistics
University of Wisconsin-Milwaukee
Milwaukee, WI 53201-0413
USA
eckman@uwm.edu

Abstract

An important question in the L2 acquisition of phonemic distinctions is whether learners have to perceive contrasts before they can successfully produce them. In this paper, data on both the perception and production of the English /s/-/š/ contrast were elicited from native speakers of Korean, a language in which [s] and [š] are allophones of the same phoneme. General principles of phonology predict that, in acquiring this contrast, Korean learners will suppress the application of the allophonic rule relating these sounds across a morpheme boundary only if they can prevent the rule from applying within morphemes. This study examines these predictions in both production and perception of the /s/-/š/ contrast as a function of learner's training.

Keywords: production, perception, interlanguage grammar, second language, fricatives, Korean.

1. Introduction

This paper reports data from a pilot project that is part of an ongoing study investigating the acquisition of phonemic contrasts in second language acquisition. The focus of the research is the second-language (L2) acquisition of a target-language (TL) contrast between two phonemes where the native language (NL) contains the sounds in question as allophones of the same phoneme. The task of the L2 learner in this case is to acquire the contrast by suppressing the application of the NL allophonic rule in the pronunciation of the TL words.

For the purposes of this study, we are interested in the acquisition of the contrast in two phonological contexts, in production as well as perception. The two general phonological contexts, or environments, are termed *morphologically basic* and *morphologically derived*, and refer to the morphological structure of representations that would trigger the application of the NL allophonic rule. Morphologically basic (henceforth, basic) environments are mono-morphemic words containing the appropriate segments for the application of the allophonic rule. For our purposes morphologically basic environments can be further subdivided into three phonological contexts: word-initial position before a vowel, word-medial position between vowels, and word-final position following a vowel. Morphologically derived environments (henceforth, derived) refer to words containing a representation to which the allophonic rule would apply, but the segments in question are separated by a morpheme boundary.

General principles of phonology predict that learners will be able to acquire a contrast in some positions before they can acquire the contrast in other positions. In basic environments, learners should acquire a contrast in word-initial and word-medial positions before they acquire the contrast in word-final position. This is because word-initial and word-medial positions are both syllable-onsets and word-final position is a syllable coda. Contrasts in onsets have been shown to be acquired before contrasts in codas.

General principles of phonology also make predictions about the acquisition of contrasts involving the learner having to split two NL allophones into separate TL phonemes by virtue of suppressing the application of the allophonic rule. The prediction is that learners will be able to suppress the application of the allophonic rule across a morpheme boundary only if they are also able to prevent the rule from applying within morphemes. In other words, a learner will acquire the contrast in derived environments only if that learner has acquired the contrast in basic environments. These general principles have always been formulated, and to date, have always been tested, on only production data. The purpose of this study is test whether these principles hold also for perception data.

The remainder of this paper is structured as follows. The next two sub-sections present the context for this study in terms of the connection between a learner's perception and production of a contrast, as well as the relationship between acquisition and phonological context. Section 2 outlines the methodology for eliciting the production and perception data, and Section 3 reports the results, which are, in turn, interpreted and discussed in the following section. The last section concludes the paper.

1.1. The production-perception relationship

One of the overarching questions in the L2 acquisition of phonemic distinctions is whether learners have to perceive contrasts before they can successfully produce them. Although it would seem to be intuitively clear that a learner must perceive any given contrast before being able to implement it in production, the literature in fact attests all four logical possibilities.

Two of these possibilities are straightforward and require little or no elaboration. There are numerous documented examples of learners who can neither perceive nor produce a TL contrast, as well as instances in which L2 learners can both perceive and produce TL contrasts (e.g., Bion, Escudero, Rauber and Baptista 2006, and others). The other two logical possibilities are not as straightforward, and therefore require discussion.

Flege's (1995) Speech Learning Model (SLM) is predicated on the notion of "equivalence classification", in which an L2 learner sets up categories for TL phonemes on the basis of that learner's perception of the segments in question. Flege (1999) examined correlations between perception and production in results from selected studies on L2 consonants and vowels (Flege, Bohn & Jang, 1997; Flege, MacKay and Meador, 1999), and found that L2 segmental production and perception are generally correlated, although factors such as age limits or methodological differences among the variables investigated by different researchers may weaken the correlation. Thus, the results from a number of studies support the claim that L2 learners perceive contrasts before they can successfully produce them.

The fourth logical possibility has also been documented. Sheldon and Strange (1982) reported that L2 learners' production of certain contrasts can exceed their ability to perceive that contrast. The authors tested native speakers of Japanese learning English on their ability to perceive and produce the distinction between /r/ and /l/, a contrast lacking in Japanese. Specifically, it was found that native speakers of English, when listening to recordings of the subjects' productions of minimal pairs containing /r/ and /l/, could successfully distinguish /r/ from /l/ better than the subjects could distinguish this contrast in their own productions.

Given the conclusion that all four logically-possible combinations of learning to perceive and produce a contrast have been attested, we turn our attention to the phonological environment in which the contrast is acquired.

1.2. The relationship between acquisition and phonological environment

Within the context of the above-referenced work on the relationship between the production and perception of phonemic distinctions by L2 learners, we investigated the ability of Korean L2 learners of English both to produce and perceive the contrast between /s/ and /š/ in derived as well as basic environments. This contrast was chosen because the phonemic status of these two segments is different in Korean than it is in English. In Korean [s] and [š] are in complementary distribution, with [š] occurring only before high front vowels, [s] elsewhere.

General principles of phonology predict that learners will be able to acquire a contrast in some positions before they can acquire the contrast in other positions. In basic environments, learners should acquire a contrast in word-initial and word-medial positions before they acquire the contrast in word-final position, because word-initial and word-medial positions are both syllable-onsets and word-final position is a syllable coda. Contrasts in onsets are acquired before contrasts in codas, in both first and second language acquisition (Eckman and Iverson 1994; Goldsmith 1990).

General principles of phonology also make predictions about the acquisition of contrasts involving the learner having to split two NL allophones into separate TL phonemes by virtue of suppressing the application of the allophonic rule. The prediction is that learners will be able to suppress the application of the allophonic rule across a morpheme boundary only if they are also able to prevent the rule from applying within morphemes (Eckman, Elreyes and Iverson 2003). In other words, a learner will acquire the contrast in derived environments only if that learner has acquired the contrast in basic environments.

To be more specific, at the beginning stages of acquisition the learner's IL grammar lacks the contrast altogether and the transferred NL rule will apply "across the board" in both basic and derived environments. This causes words such as *sip* and *ship* as well as *messing* and *meshing* to be pronounced homophonously. As the learner begins to acquire the contrast in question, the two segments become part of the IL lexicon, and the principles in question restrict the rule to apply only in derived contexts, causing the learner to distinguish the pronunciation of *sip* and *ship*, but to continue to pronounce *messing* and *meshing* homophonously. The last stage would be one in which the contrast has become acquired to the point where the rule is suppressed altogether.

As pointed out above, these general principles have always been formulated on, and to date, have always been tested on, only production data. The purpose of this study is to test whether these principles hold also for perception data.

2. Methods

2.1. Study participants

Ten native speakers of Korean served as participants of the study. All were students at the City University of Incheon, Korea, ranging in age from 19-25 years. All participants learned English in a formal school setting (for 8-10 years) and none of them had any direct informal experience with English, neither by means of a personal native-English-speaking tutor or travel to an English-speaking country. They were enrolled in various majors at the University but none of them included English, linguistics, or other language-related studies. The English teachers of the participants in their elementary, middle, and high school years were also native speakers of Korean. While in the college, however, they were enrolled in a conversation course taught by native English speakers.

2.2. Stimulus materials

Two sets of stimuli were used in the study, one pertaining to the production of the target sounds and one to their perception. These two sets were used twice to collect subjects' responses at two points in time: (1) as a pretest, at the beginning of the study before each subject entered the training phase, and (2) as a posttest, after the training had been completed.

2.2.1. Production stimuli

A set of 60 target words and 30 fillers were selected. All words were existing lexical items in English. Each target word contained /s/ or /š/ in three different positions in a morphologically basic word: initial (e.g., *sip/ship*), medial (e.g., *lesson/ocean*) and final (e.g., *pass/crash*) and in one additional position (medial, at the juncture with another morpheme) in morphologically derived words such as *passing/brushing* or *messy/bushy*. In the word-initial position, /s/ or /š/ occurred before a high front vowel which was also the case for the derived words, in which it was followed by the suffix -*ing* or –*y*. No high front vowel followed /s/ or /š/ in either the medial or final position in the basic word. A complete list of the target words used in the production task can be found in Appendix A.

2.2.2. Perception stimuli

Stimuli for the perceptual testing consisted of naturally produced single words recorded by a male native speaker of American English. The words were existing lexical items in English. Only minimal pairs were used in which /s/ or /š/ occurred in initial and final positions in basic words and in medial position followed by the suffix *-ing* or *-y*, e.g., *seep/sheep*, *plus/plush*, *classing/clashing*. The stimulus set consisted of 72 items (4 stimuli x 2 contrastive segments x 3 positions in a word x 3 repetitions). Appendix B lists all words used in the perception task.

2.3. Experimental procedures

Several custom programs were written in MATLAB for the purposes of the present study. First, a program controlling the recordings in the production task displayed on a computer screen a set of pictures, clues, and commands such as "Wait" or "Speak" designed to guide the subject and the experimenter in order to elicit the word in question. If a word could not be elicited through the picture display or written clues that followed, a verbal model was played and the subject repeated the word. The stimuli were presented in a random order. They were recorded directly into a hard disc drive at the sampling rate of 44.1 kHz. The subject spoke into a head-mounted microphone at a distance of 1 inch from the lips.

Another MATLAB program controlled the perception experiment. In this task, a single-interval two-alternative forced choice (2AFC) identification procedure was used with the two response choices, /s/ and /š/, displayed on the computer monitor. After hearing the stimulus word, the subject indicated with the press of a mouse button whether the word contained an /s/ or an /š/. The stimuli were presented in a random order over Sennheizer HD600 headphones at a comfortable listening level (~70 dB HL). Each subject was tested individually. First, the stimuli were presented in a no-masker condition so that no signal degradation was introduced to mask the signal. To make the perception task more demanding, the stimuli were then presented in multitalker babble and in masking white noise at two different levels of sound-to-babble (S/B) sound-to-noise (S/N) ratios: 0 dB, and -4 dB.

The testing took place in Incheon, Korea, in a quiet room near the University. All experiments were completed over two days. In the first day, each subject completed a production and a perception task to elicit the pre-training data. Next, the subject was trained on the production of the /s/-/š/ contrast. This was done through a series of training steps which were developed as a part of a training phase whose implementation was guided by a separate MATLAB program. Only nonsense words were used in the training phase. The training was administered to the subject in a way somewhat similar to the production task, using appropriate pictures and verbal models which were repeated and learned by the subject.

In the second day of testing, the same production and perception tasks were conducted to elicit subjects' production and perception responses to the /s/ - /š/ contrast. These tasks were administered as post-training tests to assess the effects of learning.

2.4. Data analysis

The production data were first transcribed by a native speaker of American English. All transcriptions were completed at The Ohio State University. The transcription process was fully automated using a custom MATLAB program that displayed the following choices on the computer monitor: "strong palatal," "weak palatal," "non-palatal" and "other, with comments." The data were transcribed in terms of degree of palatalization rather than categorically, i.e., as an instance of an /s/ or an /š/, because in some instances, it was difficult to determine for a native speaker of American English which fricative was produced. Degree of palatalization was easier to detect and therefore it was considered a more reliable measure of the difference between /s/ and /š/. After hearing each word, the transcriber selected an appropriate box on the computer screen to register her choice.

In analyzing the production results, we determined the percentage of correct productions for each fricative. The production of /s/ was considered correct if it was transcribed as "non-palatal" while the production of /š/ was considered correct if it was transcribed as either a "strong palatal" or a "weak palatal."

The perception results were analyzed in terms of overall percent correct score for both segments, i.e. an /s/ and an /š/. The total percent score provides information regarding the listener's ability to make a perceptual distinction between the two fricatives. As for the production data, we examined the perception responses for each position in the word.

3. Results

3.1. Results of production tests

Shown in Figure 1 are the overall mean percent correct scores for productions in the pre-training and post-training tests. The data are displayed as a function of morphological context (basic and derived) and position in a word (initial, medial and final).

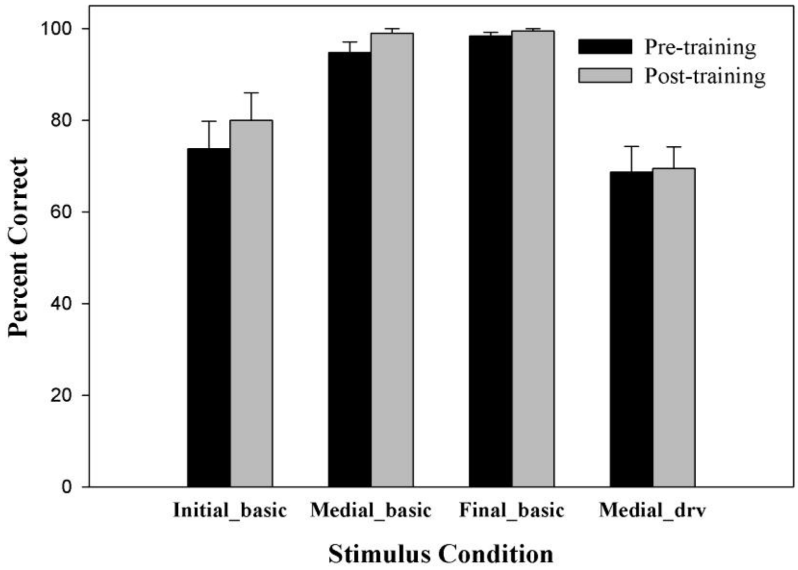


Figure 1. Mean correct productions of the /s/-/š/ contrast before and after training. Error bars indicate one standard error.

The production data were first analyzed using a two-way repeated measures ANOVA with the factors stimulus condition (initial_basic, medial_basic, final_basic and medial_derived) and test time (pre-training and post-training). The results showed a significant effect of stimulus condition ($F(3,27)=20.65$, $p<.001$, $\eta^2=.696$), indicating that the percentages of correct productions in the medial and final basic contexts were significantly higher than in either the initial basic or medial derived conditions. The main effect of training on speakers' production just failed to reach significance ($F(1,9)=4.88$, $p=0.54$, $\eta^2=.352$), primarily because there was very little improvement in either the final basic or medial derived positions. There was no significant stimulus condition by test time interaction.

Looking at only the basic contexts using a two-way repeated measures ANOVA with the factors position (initial, medial and final) and test time we find a significant effect of training ($F(1,9)=9.89$, $p=.012$, $\eta^2=.5.23$), with the greatest improvement found in the production of /s/ and /š/ in initial position. There was a significant main effect of position as well ($F(2,27)=14.18$, $p<.001$, $\eta^2=.612$). The percentage of correct productions was significantly lower in initial position than in either medial or final position.

Of particular interest in the current study is the effect of the occurrence of the fricative in a derived vs. basic word. A two-way repeated measures ANOVA with the factors morphological context (basic medial and derived) and test time of the production data in medial position only showed a significant effect of morphological context ($F(1,9)=52.51$, $p<.001$, $\eta^2=.854$) but no significant overall training effect or morphological context by training interaction. Matched-pairs t-tests showed that speakers did have a higher percentage of correct productions following training for the basic condition ($t(9)=2.44$, $p=.037$) but not for the derived condition ($t(9)=.192$, n.s.).

3.2. Results of perception tests

Each set of identification tests (in the no-masker, the babble, and the noise condition) was analyzed using the appropriate repeated-measures ANOVAs with the factors position and test time, with the additional factors babble-to-noise or signal-to-noise ratio in the later two conditions. Note that all tokens in which the target fricative occurred in medial position were derived words and the remaining tokens were basic words.

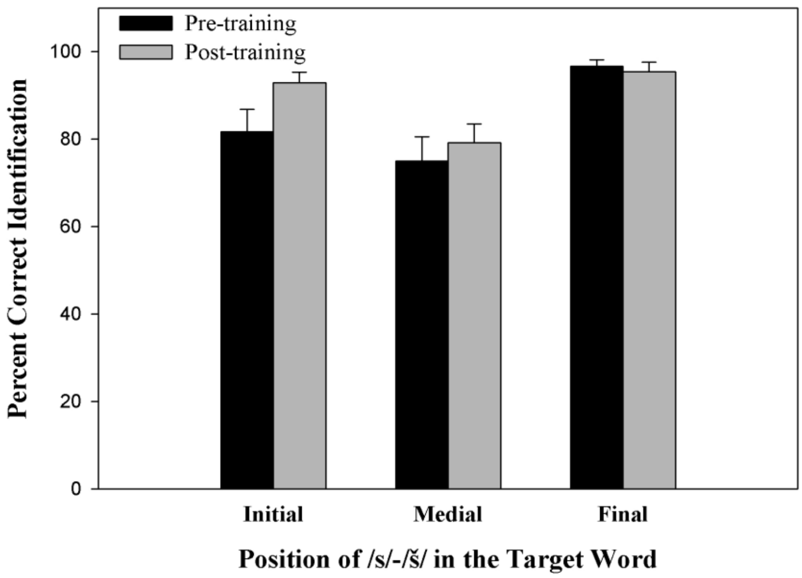


Figure 2. Mean correct identification of the /s/-/š/ contrast in no-masker condition before and after training. Error bars indicate one standard error.

3.2.1. Identification in no-masker condition

The no-masker identification task represents the perceptual condition that makes the least demands on the perceptual system and should show the best listener performance. Shown in Figure 2 are the mean percent correct identifications of the fricatives in three different positions both before and after training. There was a significant effect of position ($F(2,18)=17.9$, $p<.001$, $\eta^2=.666$), indicating that listeners had the lowest percentage of correct identifications in medial (derived) position (77%). There was also a main effect of test time ($F(1,9)=6.18$, $p=.035$, $\eta^2=.407$). Overall, subjects improved in their identification of /s/ and /š/ following training by an average of 5%. However, this must be evaluated in light of the significant test time by position interaction ($F(2,18)=4.73$, $p=.022$, $\eta^2=.345$). Matched-pairs t-tests showed that listeners improved significantly after training only in initial position ($t(9)=3.42$, $p=.008$). There was no significant improvement in final position ($t(9)=0.61$, n.s.), but this was likely because listeners were already close to ceiling performance (over 95% correct) before training. Again, as in the production tasks, the performance was poorer in the medial (derived) position compared to the positions in the basic environment.

3.2.2. Identification in babble masker condition

The target-in-babble condition represents a more challenging testing condition for the listener than does the no-masker condition as it adds a masker. In particular, since babbling is a combination of different talkers producing speech (or speech-like) sounds, babbling may introduce both energetic masking (i.e., "noise" at speech frequencies) and informational masking (linguistic content).

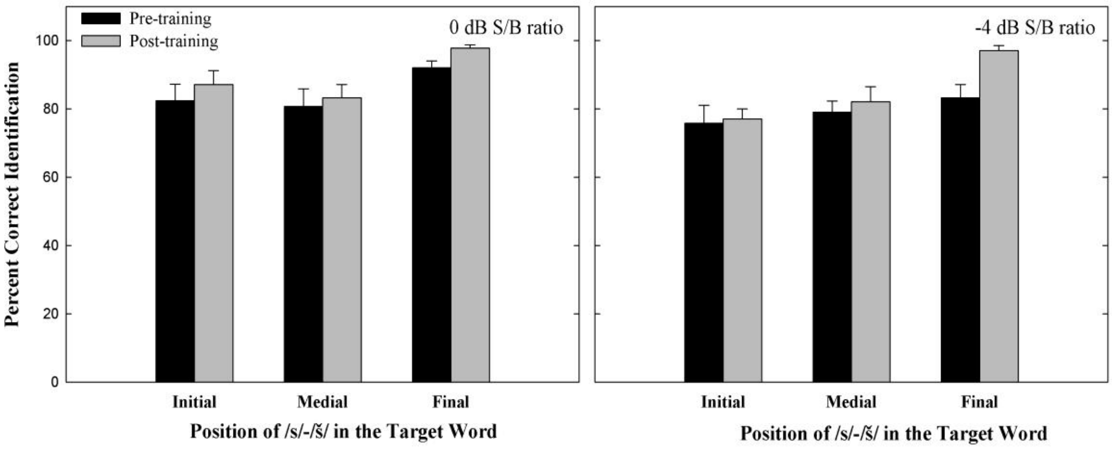


Figure 3. Mean correct identification of the /s/-/š/ contrast in target-in-babble condition before and after training. Error bars indicate one standard error.

Shown in Figure 3 are mean percent correct identifications of the fricatives in three different positions in the word both before and after training. The left panel presents the responses in the 0 dB S/B ratio condition and the right panel presents the responses in the -4 dB S/B ratio condition.

As might be expected due to the increase in masking energy, the percentages of correct identifications were, overall, significantly lower (by 5%) in the -4 dB S/B condition than in the 0 dB S/B condition ($F(1,18)=16.96$, $p=.003$, $\eta^2=.653$). There was also a significant main effect of position. The percentage of correct identifications of fricatives in the final position (92.6%) was significantly higher than the identification of fricatives in either initial (80.6%) or medial (derived) position (81.3%). The latter two positions did not differ significantly from one another.

There was an increase in the percentage of correct responses (5%) following training ($F(1,9)=6.00$, $p=.037$, $\eta^2=.400$). However, the effect of training was strongest for final position in the -4 dB S/B condition (14%), producing a significant training-by-position interaction ($F(2,18)=3.55$, $p=.05$, $\eta^2=.283$). None of the other interaction effects was statistically significant. These data continue to demonstrate that subjects had greater difficulty correctly producing and perceiving these fricatives in initial and medial (derived) positions. Conversely, their performance was significantly better when the fricative occurred in final position.

3.2.3. Identification in noise masker condition

The target-in-noise condition also presents a more challenging testing condition for the listener than does the no-masker condition. However, the noise masker is different acoustically from the babble masker. In the target-in-noise condition, the masker is Gaussian noise whose energy is spread across the spectrum (and not concentrated in the frequency range of human speech). The addition of "white noise" represents the introduction of energetic masking only (without any linguistic content whatsoever). The difference between the two maskers (babble and noise) almost certainly accounts for the different patterns we find in the perceptual responses.

Shown in Figure 4 are mean percent correct identifications of the fricatives in three different positions both before and after training. The left panel presents the responses in the 0 dB S/N and the right panel presents the responses in the -4 dB S/N ratio condition.

A three-way repeated-measures ANOVA of these data with the factors test time, position and S/N level revealed a significant effect of S/N level ($F(1,9)=15.68$, $p=.003$, $\eta^2=.635$) and training ($F(1,9)=10.01$, $p=0.01$, $\eta^2=.527$). Identification performance improved by 7% following training and was 5% better in the 0 dB than the -4 dB condition. However, for these data, the main effect of position was not significant.

Instead, there was a significant S/N level by position interaction ($F(2,18)=20.46$, $p<.001$, $\eta^2=.695$) produced by the fact that the identifications of the fricatives in medial (derived) position were significantly better than expected in the -4dB condition. In fact, the mean percent correct identification of the fricatives in medial position was actually higher in the -4 dB condition (83.4%) than in the -0 dB condition (81.0%). It is important to examine each of these S/N conditions separately in order to understand better this unexpected outcome.

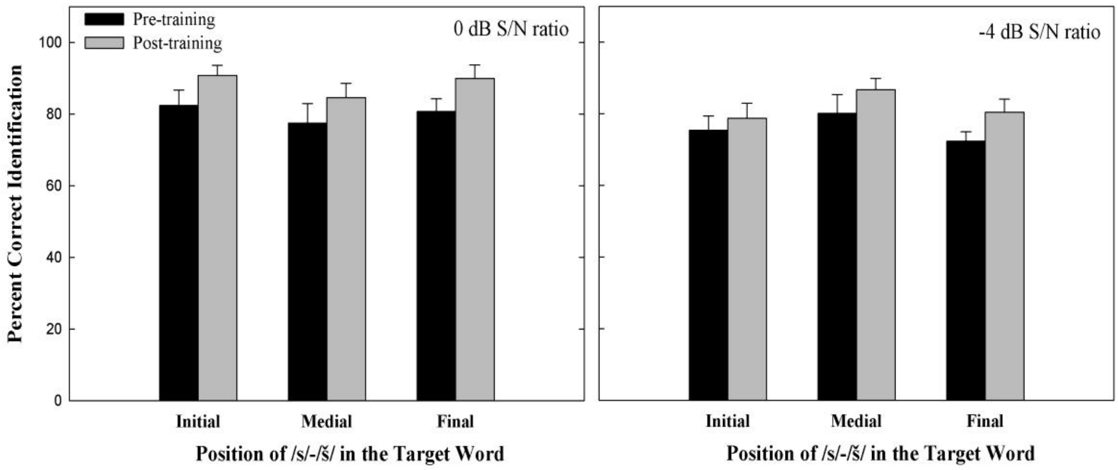


Figure 4. Mean correct identification of the /s/-/š/ contrast in target-in-noise condition before and after training. Error bars indicate one standard error.

The pattern of responses in the 0 dB S/N condition matches many of the patterns that we found in the production data and in the perception results for the no-masker and target-in-babble conditions. Namely, the mean percent correct was lower in the medial (derived) condition (81%) than in either initial (86%) or final position (85%). This was true for data collected both before and after training.

The pattern of responses in the -4 dB S/N condition does not match earlier patterns in an important way. Although, as expected, the mean percent correct identifications for initial and final positions drop (by 9%) for both positions in the -4 dB S/N condition compared to the 0 dB S/N condition, mean percent correct identifications actually improve by 2% in the -4 dB S/N condition for medial (derived) position. The anomalous behavior in the -4 dB S/N condition can be explained, we believe, by more closely examining the acoustic characteristics of the target stimulus (/s/ and /š/), the nature of the masker and the likely strategies used by the listeners in the identification tests.

Phonetically, both /s/ and /š/ are voiceless, sibilant fricatives. Acoustically, these two fricatives differ primarily in terms of their spectral mean (/s/ has more energy in the higher frequencies than does /š/). The babble masker will include spectral energies across the speech range (and the talkers in the background babble are producing the entire range of English speech sounds). The noise masker, on the other hand, is actually more like the fricatives themselves in that it represents acoustic noise. In the no-masker and babbling conditions we are assuming that listeners are making their decisions on the basis of a linguistic judgment (i.e. whether there is an /s/ or an /š/ in the word). The lower performance of listeners in the medial (derived) position (even after training) matches their difficulties as speakers in producing these fricatives in the same derived context.

Although listeners are likely making decisions on a linguistic basis in the 0 dB S/N condition, one can explain the anomalous pattern of identifications by suggesting that, in the -4 dB S/N condition, the present listeners have adopted a non-linguistic strategy in making their identification decisions. In particular, they may have been listening to the nature of the noise and when they heard a higher-pitched noise embedded in the noise masker, they responded with /s/ and when they heard a lower-pitched speech noise, they responded with /š/. Why would this occur for only the medial (derived) fricatives and not the initial and final fricatives? The answer is that the rms levels (a measure of mean amplitude) of the fricatives in the medial (derived) position are approximately 7 dB higher than the fricatives in either the initial or final position. Thus, the strategy will benefit listeners when the fricative is in medial position, but not when it is in initial or final position.

4. Discussion

This study examined the relationship between production and perception of English fricatives /s/ and /š/ by native speakers of Korean along linguistic variables (morphological context and position in a word) and as a function of subject’s training. The results indicate a close relationship between production and perception for the above factors in most of the testing conditions.

The production data indicated a positional asymmetry, which was manifested in significantly lower accuracy in initial position in a word for the basic environment and in medial position for the derived environment as compared to medial and final positions in the basic environment. Moreover, for the two positions in which the accuracy was lower, the improvement with training was found only in the initial basic position and not in the derived. This indicates that the English fricatives in initial position in a morphologically basic word are difficult to produce by a Korean speaker, which is comparable with a difficulty encountered in a morphologically derived context. However, while the pronunciation in initial position improved with training, the production in the derived context remained essentially the same. Conversely, fricatives in the medial and final positions in a basic word were produced with greater accuracy, which was reflected in high number of correct classifications by a transcriber who was a native speaker of American English. The question arises whether the asymmetry in production is matched in the perception results from the same subjects.

Although we can compare the production and perception data only for selected positions in a morphologically basic word due to the lack of the medial position in the present design, we still find a correspondence between the lower performance in initial position in production and poorer identification in this position in perception. Furthermore, both the production and perception were significantly better in the final position in a word. As for the derived context, the production and perception results show again a close match in that the accuracy in both tasks was significantly lower. Finally, as was found for the production data, the improvement with training in perception was most evident in the word-initial position.

It needs to be pointed out that this close production-perception relationship was found only when the perception stimuli were delivered in the no-masker condition, i.e., when no distortion was introduced to the speech signal. When the target stimuli were presented in a multitalker babble, we found some variation in this pattern. In particular, although the identification was again highest in final position in a basic word, the improvement with training was also significant in this position. Unlike for the no-masker condition, no improvement was found in initial position when target stimuli were presented in a babble masker. However, listeners continued to show lower accuracy in initial and medial (derived) positions, which is consistent with the pattern observed for the no-masker condition.

Using noise as a masker introduced yet another type of variance to the general pattern found thus far in perception. Most importantly, there was no significant effect of word position, indicating no difference in the accuracy of identification of the fricatives across all positions and morphological contexts. Another discrepancy was found in the -4 dB condition, where the percentage of correct identifications significantly increased in medial (derived) position. This anomalous effect can be explained on the basis of differences in processing of a fricative sound in background noise as opposed to background babble, as already discussed. Another discrepancy between the target-in-noise condition and the two other listening conditions lies in the significant improvement after training in all positions in a word, not only in selected positions (either initial or final). This improvement was observed in both 0 and -4 dB S/N conditions.

In summary, the present data support the existence of a production-perception link in the acquisition of segmental contrast in a second language. However, this relationship may not be manifested in all testing conditions and, especially in perception, listener behavior may depend on other than linguistic factors in processing the second language sound. We have observed a departure from the general pattern of a close match between production and perception when the fricatives were presented with background noise. This environmental factor "successfully" interfered with the linguistic processing of both fricatives so that the positional asymmetry found in other testing conditions was to a great extent lost in the target-in-noise condition. This result suggests that for some types of speech sounds, the acoustic content and distribution of spectral energy may contribute to increased confusion or to unexpected improvement in performance. This may happen when the listener is uncertain about the source of the noise, i.e., linguistic or environmental, when attending to a speech signal which itself contains a great portion of noise, such as fricatives and aspirated stops.

The lower accuracy in word-initial position, both in production and perception, was an unexpected finding in the present study. It has been widely accepted that word-initial position is the strongest and most salient, so that a consonant in this onset position tends to be acquired earlier than a consonant in the final position (or coda position in a CVC syllable structure). The syllable-initial position is often linked to early babbling, which is composed mostly of CV syllables (Locke, 1993). A consonant in the coda (and not in the onset) tends to be deleted or mispronounced in early child language (Smith & Stoel-Gammon, 1983; Vihman, 1996). Similarly, second language learners were shown to produce more errors in word final position than in word initial position (e.g., Flege & Davidian, 1984). The current results clearly depart from this established view.

However, a possible explanation of the present outcome may be linked to models of lexical access which emphasize the importance of word initial information for word recognition (e.g., Cole & Jakimik, 1980; Walley & Metsala, 1990, Vitevich, 2002). Accordingly, word initial information is essential for activating words in lexical processing, which stresses the importance of accurate production of initial consonants. It has been shown that errors in word initial position produced by second language speakers are more detrimental to word intelligibility for native speakers of a language than errors in word final position (Bent et al., 2007). The present second language listeners had reduced ability to make a distinction between minimally contrastive words when the contrast was in word-initial position, indicating that the contrast between the two fricatives may not have been yet fully acquired. Consequently, they might not have developed lexical representations of English words which contrast /s/ and /š/ word initially and their perceptual identification of this contrast was also poorer than in other positions in the morphologically basic word. This possibility must be tested in further experiments, however.

5. Conclusion

This study investigated the ability of Korean L2 learners of English both to produce and perceive the contrast between /s/ and /š/ in derived as well as basic environments. The results show a close relationship between production and perception for each position in the word. This was the case when listeners' perception was tested in no-masker listening condition. The perceptual responses did not always match the production pattern when the stimuli were presented in background babble or noise, which introduced considerable distortion to the signal. However, this outcome can be somewhat expected in the case of L2 learners.

Of particular interest here is the close production-perception relationship found for the contrast occurring across a morpheme boundary, i.e. in derived environments. This indicates that learners' perceptual response corresponds to their ability to produce the contrast in this difficult position. An unexpected finding of the study was that the current learners seemed to acquire the contrast in word-final position (in a syllable coda) before they acquired the contrast in the initial position (in a syllable onset). Although this matter warrants further investigation, it is noteworthy that this pattern was reflected again in both production and perception of the contrast, showing that these two systems are closely related.

References

Aoyama, K., J. E. Flege, S. Guion, R. Akahane-Yamada, & Y. Tsuneo. 2004. "Perceived phonetic dissimilarity and L2 speech learning: The case of Japanese /r/ and English /l/ and /r/". *Journal of Phonetics 23*, 233-250.

Bent, T., Bradlow, A. R. & B. L. Smith. 2007. Segmental errors in different word positions and their effects on intelligibility of non-native speech. In O.–S. Bohn & M. J. Munro (eds.), *Language experience in second language speech learning: In honor of James Emil Flege.* 331-347. Amsterdam & Philadelphia: John Benjamins.

Bion, R. A. H., R. Escudero, A. S. Rauber, & B. O. Baptista. 2006 Category formation and the role of spectral quality in the perception and production of English front vowels. *Proceedings of Interspeech 2006,* 1363-1366.

Cole, R. A., & J. Jakimik. 1980. How are syllables used to recognize words? *Journal of the Acoustical Society of America*, 67. 965-970.

Eckman, F., A. Elreyes, & G. Iverson. 2003. Some principles of second language phonology. *Second Language Research 19*, 169-208.

Eckman, F. & G. Iverson. 1994. Pronunciation difficulties in ESL: Coda consonants in English interlanguage. In M. Yavas (ed.) *First and second language phonology,* 251 – 265. San Diego: Singular Publishing Group.

Flege, J. E. 1995. Second language speech learning: Theory, findings and problems. In W. Strange (Ed.), *Speech production and linguistic experience: Issues in cross-language research* (pp. 233-277). Timonium, MD: York Press.

Flege, J. E. 1999. The relation between L2 production and perception. *ICPhS 99*, 1273-1276.

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

Flege, E. & R. Davidian. 1984. Transfer and developmental processes in adults FL Spanish productions. *Applied Psycholinguistics*, 5. 323-347.

Flege, J. E., I. McKay, & E. Meador. 1999. Native Italian speakers' production and perception of English vowels. *Journal of the Acoustical Society of America 106*, 2937-2987.

Goldsmith, J. 1990. *Autosegmental and metrical phonology.* Oxford and Cambridge, MA: Basil Blackwell.

Ingram, J., & S-G. Park. 1997. Cross-language vowel perception and production by Japanese and Korean learners of English. *Journal of Phonetics 25*, 343-370.

Kluge, C. D., A. S. Rauber, M. S. Reis, & R. A. H. Bion. 2007. The relationship between perception and production of English nasal codas by Brazilian learners of English. *Interspeech* 2007, 2297-2300.

Locke, J. L. 1993. *The child's path to spoken language*. Cambridge, MA: Harvard University Press.

Sheldon, A. & W. Strange. 1982, The acquisition of /r/ and /l/ by Japanese learners of English: Evidence that speech perception can precede speech perception. *Applied Psycholinguistics 3*, 243-261.

Smith, B. L. & C. Stoel-Gammon. 1983. A longitudinal study of the development of stop consonant production in normal and Down's Syndrome children. *Journal of Speech and Hearing Disorders,* 48. 114-119.

Vihman, M. M. 1996. *Phonological development: The origins of language in the child.* Cambridge, MA: Blackwell Publishers.

Vitevich, M. S. 2002. Influence of onset density on spoken-word recognition. *Journal of Experimental Psychology: Human Perception and Performance,* 28. 270-278.

Walley, A. C. & J. L. Metsala. 1990. The growth of lexical constraints on spoken word recognition. *Perception and Psychophysics,* 47. 267-280.

Appendixes

Appendix A: Words used in the production task

sink	fish	race	dressy
sick	wash	pass	passing
scissors	wish	price	pricey
seats	rush	kiss	pricing
see	crush	she	brushing
bicycle	brush	sheet	bushy
lesson	crash	ship	crashing
license	splash	sheep	fishing
message	polish	shoe	crushing
motorcycle	busing	ocean	splashing
bus	crossing	parachute	washing
cross	grassy	patient	wishing
dress	kissing	tissue	polishing
grass	messy	vacation	rushing
mess	racing	bush	

Appendix B: Words used in the perception task

ship
sip
shock
sock
sheep
seep
shed
said
clashing
classing
leashing
leasing
meshing
messing
mashing
massing
clash
class
leash
lease
mesh
mess
plush
plus