

TEACHING ENGLISH WORD-FINAL ALVEOLOPALATALS TO NATIVE
SPEAKERS OF KOREAN

By

SANG-HEE YEON

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2004

Copyright 2004

by

Sang-Hee Yeon

To my mom, Woon-Ja Lee

ACKNOWLEDGMENTS

I am very grateful to many people who helped me go through my years of doctoral studies. First and foremost, I am thankful to my mom, Woon-Ja Lee, who missed me and helped me the most from a far distant place, Korea. I can never imagine finishing my degree without her help and her sacrifice, as I progressed through the tiring work of studies. I do not know how to repay the debt to my mom.

Dr. Ratreë Wayland, my guide and chair of the committee, has led me throughout my entire 5 years of doctoral study. She always encouraged independent thinking and helped me whenever I had a stumbling block. Her comments always helped to make the dissertation better. She is also a role model for me. She always encourages students to think of a new project that we can develop. She shows what a true researcher does after the degree is given. Many thanks go to her for her asking, “Would it be interesting to study this?”

The members of my committee each contributed to my work greatly as well. Dr. Caroline Wiltshire, from the beginning of my study in the department, has always been supportive of any of my efforts, such as getting a job, and research funding. Of course she is full of great feedback on my work. I really appreciated her classes as well. In fact, the dissertation was partly developed from the class project with her in 2000. Many thanks go to her for working with me through my entire doctoral program.

Dr. Gillian Lord, with wit and suggestions, has been a great resource on second language acquisition. She shared my interests in this field, and gave many suggestions on

study, and on life. I am also grateful to Dr. Johnson, even though she joined later in the course of my dissertation, because she made a continuous effort to develop the dissertation.

I will never repay my debt to my brothers and their wives, Seung-Uk Yeon and Su-Kyoung Seo, and Je-Heon Yeon and Ji-Young Kim. They are truly an inspiration, and a source of energy to march on. The new addition to our family, Gavin Yeon, is the one who lights me up every day.

I thank all of my friends, especially my fellow “lab rat,” Mohammed Al-Khairi. He has been my resource, supporter, and friend ever since we shared a lab together. Without him, I do not know when I could have finished my dissertation. I am also thankful for HeeNam Park for her continuous support and humor. We shared many, many days of fun and laughter since Day One in Florida. I will miss her very much. I will miss the smiles of friends like Manjula Shinge, Alex Mouat and Andrea Dallars. Kyoung-Ok Pak’s support and encouragement will never be forgotten. Burdett and Jan Neal, host-family in Iowa, are always my emotional support. Finally, I greatly appreciated all the effort from LaTosha Csonka, who proofread the entire dissertation. Last but not the least, I am thankful to my Lord for everything.

TABLE OF CONTENTS

	<u>page</u>
ACKNOWLEDGMENTS	iv
LIST OF TABLES	x
LIST OF FIGURES	xii
ABSTRACT	xiii
CHAPTER	
1 INTRODUCTION	1
2 ISSUES IN SECOND LANGUAGE ACQUISITION (SLA).....	6
Factors Affecting the Varying Degrees of Foreign Accents	6
Age as a Factor in SLA	6
Physiological basis of the critical period hypothesis (CPH)	8
CPH: whys and why-nots	8
Length of Residence (LOR)	10
Language Use (First Language/Second Language Use)	12
First Language (L1) and Second Language (L2) Phonological Differences.....	14
Similarities between L1 and L2.....	14
3 RELATIONSHIP BETWEEN PERCEPTION AND PRODUCTION.....	18
Speech Perception Theories.....	18
Relation Between Perception and Production in L1 Acquisition	19
Perception and Production in L2	22
Second Language Speech Learning Model	28
Flege’s Speech Learning Model (SLM).....	28
Best’s Perceptual Assimilation Model (PAM).....	28
Define “New” and “Similar”	29
Effectiveness of Short-term Laboratory Perception Training	29

4	KOREAN AND ENGLISH OBSTRUENTS	33
	Differences between Korean and English Obstruents	33
	Stops	33
	Fricatives	36
	Affricates	37
	Errors Made by Korean Speakers of English	38
	Stops	38
	Fricatives	39
	Affricates	40
5	STUDY	42
	Hypotheses	42
	Methodology	44
	Perception Tests	44
	Pretests	44
	Posttest 1 and delayed posttest (posttest 2)	49
	Generalization tests	49
	Perception Training	51
	Stimuli	51
	Participants	51
	Procedure	51
	Data Analysis	52
	Comparison among English control (EC), Korean experimental (KE), and Korean control (KC)	52
	Pretest, posttest 1 and posttest 2	52
	Generalization tests I and II	53
	Production Test	53
	Pretest	53
	Posttest 1	56
	Posttest 2	56
	Judgment of production	57
	Acoustical Analysis	58
	Data Analysis	59
	KE vs. KC in Three Tests	59
	Individual Segments	59
	Acoustical Analysis of /dʒ/	60
	Correlation between Perception and Production	60
6	RESULTS: PERCEPTION TESTS	61
	Scores in Pretest	61
	Group and Time Comparison in Korean Experimental Group (KE) and Korean Control Group (KC)	62
	Generalization Tests, I and II (Subsets 4 and 5)	65
	Individual Subsets among KE and KC	67

Words Ending with an Alveolopalatal and an Alveolopalatal+i	69
Pretest Subsets in KE and KC (Lexical Status)	70
Subset 1	70
Subset 2 and 3	71
Generalization Test II (Subset 5)	71
Improvement in Final Alveolopalatal vs. Final Alveolopalatal+i Words in KE	74
Individual Differences	75
Training	76
Summary	79
7 RESULTS: PRODUCTION TESTS	81
Interrater Reliability	81
Group and Time Comparison in KE and KC	81
Individual Segments /ʃ, tʃ, dʒ/ in NK (KE and KC Combined)	84
Production Improvement in KE	86
Relation between Perception and Production	91
Correlation in NK (KE and KC combined) and Each Group	92
Individual Differences in Improvement in KE	92
Low Proficiency Group	93
Group with Low Proficiency in Production	94
Correlation between Perception and Production in Low Proficiency KE and KC (Combined and Separate)	95
Production Improvement in Low Proficiency KE and KC (Combined and Separate)	96
Individual Factors and Achievement in Perception and Production	100
Acoustical Analysis	104
Durational Differences in /dʒ/	104
Duration of /dʒ/ among correct tokens	105
Duration of vowel preceding /dʒ/ among correct tokens	105
Discrepancy between Judges and Acoustical Analysis	106
Correct vs. Incorrect Tokens in NK (KE and KC Combined)	109
Summary	111
8 DISCUSSIONS AND IMPLICATIONS	113
Effect of Training	114
Perception	114
Stop vs. alveolopalatal fricative/affricate	117
Individual differences in KE	120
Training	121
Production	123
Production of words ending with /i/	124
Errors in words ending with /i/	125
Individual differences in KE	126

Correlation between Perception and Production	126
Individual Differences and Other Factors.....	129
Acoustical Analysis	130
Different Models of Speech Perception.....	132
Educational Implications	136
Perception.....	136
Production.....	138
Research Implications.....	140

APPENDIX

A PERCEPTION PRETEST STIMULI	146
B BACKGROUND OF KOREAN PARTICIPANTS	148
C PERCEPTION GENERALIZATION TEST II (SUBSET 5) STIMULI	149
D PERCEPTION TRAINING STIMULI.....	150
E PRODUCTION STIMULI FOR WORDLIST GROUP	151
F PRODUCTION STIMULI FOR NAMING GROUP.....	152
LIST OF REFERENCES.....	153
BIOGRAPHICAL SKETCH	161

LIST OF TABLES

<u>Table</u>	<u>page</u>
4-1 Korean and English obstruents phonemic inventory.....	35
5-1 Stimuli for pretest.....	46
5-2 Stimuli for Generalization Test II (Subset 5).....	50
6-1 Score comparison in each subset among KE, KC and EC.....	62
6-2 Total score comparison between KE and KC.....	63
6-3 Generalization tests I and II.....	66
6-4 Mean scores in each subtest in KE.....	68
6-5 Mean scores in each subtest in KC.....	68
6-6 Comparison between words ending with alveolopalatal (P) vs. alveolopalatal+i (Pi).....	69
6-7 Non-words vs. words in generalization test II.....	72
6-8 Non-words ending with alveolopalatal+i (Pi) and alveolopalatal (P) in subset 5.....	73
6-9 Comparison between words ending with an alveolopalatal (P) and an alveolopalatal+i (Pi) in KE.....	75
6-10 Individual percentage of correctness and difference between the posttests and pretest.....	76
6-11 Raw scores for each training session.....	79
7-1 Mean percentage of correctness in the pretest and posttests 1 and 2.....	82
7-2 Mean percentage of correctness of words with final alveolopalatals in pretest and posttests 1 and 2.....	83
7-3 Mean percentage of correctness in words ending with /i/ for NE.....	83
7-4 Mean percentage of correctness in words ending with /ʃi/ and /tʃi/.....	89

7-5	Mean % of production accuracy and the differences of posttests and pretest in production of KE.....	90
7-6	Mean percentage of correctness in perception and production in KE.....	95
7-7	Mean percentage of correctness in low-proficiency KE	98
7-8	Mean duration (%) of /dʒ/ and preceding vowel in KE, KC and EC.....	105
7-9	Mean duration (%) between correctly and incorrectly produced /dʒ/ and its preceding vowel	109
7-10	Mean duration (%) comparison between CVdʒ, and CVdʒi words produced by EC.....	110
8-1	Percentage of correctness of participants 8 and 9 in KE.....	128
A-1	Perception pretest stimuli	146
B-1	Background of Korean participants.....	148
C-1	Perception generalization test II (Subset 5) stimuli	149
D-1	Perception training stimuli	150
E-1	Production stimuli for wordlist group	151
F-1	Production stimuli for naming group	152

LIST OF FIGURES

<u>Figure</u>	<u>page</u>
6-1 Score comparison between KE and KC at 3 testings	63
6-2 Mean score between non-words (N=20) and words (N=20) in the pretest subset 3 for NK	72
6-3 Non-words ending in the posttest 1 subset 5 (N=18 for each category)	73
6-4 Correlation between the pretest and a difference of posttest 1 and pretest for KE ($r=-0.88$, $r^2=0.78$)	77
6-5 The mean accuracy of training in weeks 1, 2, and 3 for KE	78
7-1 Mean percentage of correctness for final /ʃ, tʃ, dʒ/	85
7-2 Mean % of accuracy for words ending with /dʒ/ in pretest and posttests 1 and 2 in KE.....	88
7-3 Mean % of accuracy for words ending with /dʒ/ in pretest and posttests 1 and 2 in KC	88
7-4 Correlation between pretest and the difference of posttest 1 and pretest in KE ($r=-0.505$, $r^2=-0.26$)	91
7-5 Correlation between perception and production in the pretest in NK ($r=0.457$, $r^2=0.21$).....	93
7-6 Correlation between perception and production in pretest of low proficiency group ($r=0.457$, $r^2=0.028$)	97
7-7 Mean % of correctness comparison in the production of the low proficiency participants in KE.....	99
7-8 Correlation between posttest 1 production and LOR in KE ($r=0.56$, $r^2=0.31$)	101
7-9 <i>Change</i> mistakenly judged by EJ as <i>correct</i>	108
7-10 <i>Judge</i> mistakenly judged by EJ as <i>incorrect</i>	108

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

TEACHING ENGLISH WORD-FINAL ALVEOLOPALATALS TO NATIVE
SPEAKERS OF KOREAN

By

Sang-Hee Yeon

August, 2004

Chair: Ratreë Wayland
Major Department: Linguistics

Foreign accent and training in its reduction were studied. In the second language classroom, the source of a foreign accent is often disregarded, and mere articulation training is given to learners. Hence, this study discusses the source of accent, especially the possibility of perception errors being the source of production errors. For example, Korean speakers often do not produce English words with a final alveolopalatal and with an alveolopalatal+i distinctively (e.g., *fish* and *fishy*), and it is valuable to examine if they have perception errors, too. In addition, the effectiveness of intensive perception training was examined. The hypothesis is that the perception training helps both perception and production of English final alveolopalatal words. Perception and production tests were given to 15 adult Koreans (Korean experimental group), 12 adult Koreans (Korean control group), and 11 English speakers (English control group). Tests were given three times: Pretest, posttest, and delayed posttest (3 months after the posttest). Production tests were judged by another group of native speakers of English. After the pretest, 3 weeks of

perception training was given to the Korean experimental group. Training stimuli included minimally different words with a final alveolopalatal and with an alveolopalatal+i. Results showed that the Korean experimental group improved their perception of words with a final alveolopalatal after the training, and sustained the ability 3 months after. In the production of final alveolopalatal words, the experimental group did not improve it right after the training, but showed improvement three months after. Perception training seemed particularly beneficial for participants who started with a low level of proficiency.

CHAPTER 1 INTRODUCTION

Many second language (L2) speakers tend to have nonnative-like sound qualities when they speak in their L2. For example, Korean speakers of English tend to produce *fish* as *fishy*, *change* as *changy* and *church* as *churchy*. In other words, they produce words ending with an alveolopalatal with an extra vowel /i/. The source of this production error is not clear. It could be a simple articulation problem. However, it could also originate from a perception problem: Speakers might hear an extra vowel after a word-final alveolopalatal. Since the possibility of Korean speakers of English having a perceptual problem has not been explored yet, our study addressed the issue of the relationship between perception and production in L2 sounds. In addition, we trained a group of Korean speakers of English with intensive perception training, and tried to help them reduce production errors in word-final alveolopalatals. So far, short-term intensive perception training has resulted in improvement in perception and subsequently in production of L2 sounds (Bradlow et al., 1999), but our study is the first to incorporate perception training in teaching Korean speakers of English. By investigating the effect of perception training, we addressed the issue of the effect of perception training on foreign accent reduction. Finally, by examining the development of perception and possibly of production through the training, our study discussed the theories of speech perception and its relation to speech production, and in particular theories in L2 perception.

In our study, we trained adult L2 speakers and examined the possibility of improvement in perception and production. Adults are claimed to have more pronounced

accents than children, which relates to the issue of age in L2 acquisition. It seems that after a certain age, language acquisition becomes difficult. Children who come to the L2 community at an early age usually do not have any accents, but their parents do. However, varying degrees of foreign accents still exist in those who come to the L2 country at the similar age. According to Larsen-Freeman and Long (1991), the contributing factors include length of residence, motivation, aptitude, etc. In addition, when L2 speakers begin learning an L2, the differences between first language (L1) and L2 seem to affect L2 speakers' performance (Broselow et al., 1998). However, sometimes perceptual similarity makes it harder to acquire the L2. It is said that the more an L2 segment is perceptually close to that of L1, the harder it is to acquire (Flege, 1995). The difference or similarity between two languages could be one of the sources for errors in the production and perception of sounds in L2. Many researchers examined the production of L2 speech segments, but it is rare to examine perception and production of L2 sounds at the same time. Therefore, our study explored both domains, and examined the relationship between them.

Even though most adult learners seem to have foreign accents, it does not mean that they cannot improve. Studies show that training helps reduce foreign accents of adult speakers (Bradlow et al., 1997, 1999). Methods of pronunciation training are often discussed. However, it is debatable whether explicit teaching of pronunciation is effective, because the teaching approach has been shifted back and forth from audio-lingual to communicative teaching since the 1980s. However, the demand for pronunciation instruction is obvious for several reasons, for example, complaints from students of international teaching assistants (Morley, 1991), and longer processing time

to understand accented speech (Munro & Derwing, 1995). In addition, poor pronunciation is a socially undesirable phenomenon (Sheldon, 1985). Different types of pronunciation training are available, but traditional articulatory instruction is usually favored (such as the imitation of certain sounds several times), because teachers believe that the source of errors is incorrect articulation. We used an alternative method of pronunciation teaching, intensive perception training, which assumes that perception and production are related.

Intensive perception training methods often result in improvement in both domains. For example, perception training for discriminating English /ɹ/ and /l/ by Japanese speakers successfully resulted in improvement in perception and production (Bradlow et al., 1999). However, the results are not evaluated well enough to be recommended in actual classroom environments. Therefore, our study evaluates the effectiveness of perception instruction to Korean speakers of English. Particularly, Korean speakers of English often insert a vowel [i] after English alveolopalatal codas, for example pronouncing *catch* as *catchy*. By giving the perception training to Korean speakers, we explored the effect of training on perception. We also examined the effect of training on production of the final alveolopalatals, under the assumption that perception improvement leads to production improvement.

Perceptual sources of inserting extra vowels after final alveolopalatals by Korean speakers of English were also studied. It was expected that Koreans would identify words ending with an alveolopalatal as words ending with an alveolopalatal+i (e.g., *edge/edgy*). This is because Koreans tend to produce words ending with an alveolopalatal with an extra vowel [i]; therefore, they might be more familiar with words ending with an

alveolopalatal+i. We were also interested in the possibility that trainees would transfer the learned ability to perceive new stimuli, new words and new talkers that were not part of the training. If the training were successful, the trainees would have a stable perceptual model of words ending with an alveolopalatal; therefore they would perceive new stimuli with accuracy. We also discussed the long-term effect of perception training. It was hypothesized that Korean speakers who took the training would sustain the learned ability three months after the training, so that they would perceive and produce better than people who did not take the training. Finally, we examined the relationship between perception and production of an L2 sound. So far, the sequence of learning perception before production has not been agreed upon in L2 studies. In many L1 studies, this is agreed upon. We expected that a learner who performed poorly in the perception of word final alveolopalatals would also perform poorly in the production of word final alveolopalatals.

Scope of the research. Participants included 27 native Korean speakers of English (NK) who had lived in Gainesville, Florida. Their length of residence and age of arrival varied. However, all of them had come to the U.S. after puberty (in their 20s and 30s). The Koreans were divided into 2 groups: a Korean experimental group (KE, N=15) and a Korean control group (KC, N=12). Participants in KE received three weeks of intensive perception training of English alveolopalatals, /ʃ, tʃ, dʒ/. Participants in KC did not receive any training. Another control group consisted of native speakers of English (EC, N=11).

Our study involved perception and production tests. Each test was conducted 3 times: pretest, posttest 1, and delayed posttest (posttest 2). In the perception test

participants were asked to identify minimally different word pairs. Words ending with an alveolopalatal and with an alveolopalatal+i were included in the test. A production test included reading words from a wordlist, or naming words in English by reading flash cards. All Korean participants took both perception and production tests. The training lasted 3 weeks, and involved identifying words with final alveolopalatal and alveolopalatal+i (e.g., *itch/itchy*). Training stimuli were produced by three different talkers so that Korean participants were exposed to a varying range of acoustic quality of segments ("high variability training paradigm," Logan, Lively & Pisoni, 1991). Finally, to evaluate the production accuracy of Korean speakers of English, a panel of 6 native speakers of English participated.

CHAPTER 2 ISSUES IN SECOND LANGUAGE ACQUISITION (SLA)

It is widely accepted that even after years of exposure to a second language (L2), adult L2 speakers continue to produce it with varying degrees of foreign accents. Foreign accents are also difficult to eradicate, and it seems that in many cases adult L2 speakers are almost expected to have it. However, this does not mean that every adult L2 speaker has a strong accent. The degree of accent varies considerably from individual to individual. The question is what makes an L2 speaker's more or less heavy than others'. There are several factors that need to be considered. These include age, length of residence, first language (L1)/L2 use, etc. In addition, L1/L2 phonological differences contribute greatly to cross-language variations in foreign accents. The goals of our study do not include the effects of different factors in SLA, but it is worth recognizing different factors affecting achievement.

Factors Affecting the Varying Degrees of Foreign Accents

Age as a Factor in SLA

Age as a factor in SLA has been studied for a long time; however, there is no strong agreement on its role. Age in SLA usually refers to age of arrival (AOA), which is the first arrival time in a predominantly target language (TL) speaking country. Many researchers (Piske et al., 2001) claim that AOA is the most significant factor predicting the degree of foreign accents. Their claim is that 'younger is better.' The claim is based on the critical period hypothesis, first introduced by Lennenberg (1967). The hypothesis states that normal language acquisition is possible only if a person is exposed to a

language before a certain period of time. Others (Purcell & Suter, 1980) claim that age is not the most significant factor, and that even adults can achieve native-like proficiency. Studies on age sometimes produce different results. The source of these differences might lie in what aspect of the critical period a research study dealt with (rate of acquisition or ultimate attainment).

Regarding rate of acquisition, it seems that adolescents and adults have an advantage over younger children at the beginning stage. Snow and Hoefnagel-Hohle (1977) found that adolescent children were better at imitating Dutch words than younger children after 5 months of residence. They (1978) also reported that older children (8–10, 12–13 years old) performed better than younger children (3–5 years old). Other studies show that adults can achieve native-like proficiency. For example, Neufeld (1980) studied English speakers of French, and found that advanced speakers were judged as natives by native speakers of French. In addition, Loewenthal and Bull (1984) observed that even when chronological age was controlled, a group with early AOA did not always imitate Armenian sounds better than a group with late AOA. On the other hand, another study on Armenian sound imitation by English-speaking children shows that younger children imitated Armenian sounds and intonation better than older children (Tahta, Wood, & Loewenthal, 1981b).

Neufeld's study (1980) elicited several responses. For example, Long (1990) claims that the methodology to elicit the data had several flaws. Long explains that the native speaker judges who lived in English-French speaking communities, such as Montreal, had more tolerance and lower expectations of French from nonnatives. In addition, the participants were allowed to rehearse and repeat for the best performance.

With regard to ultimate attainment of native-like pronunciation, it seems that people who arrive in a TL community at an early age have an advantage over those who arrive as adults (Flege, Yeni-Komshian, & Liu, 1999; Oyama, 1976; Thompson, 1991). To support this, Flege, Munro and Mackay (1995b) carried out a study. In the production of several English consonants, Italian bilinguals whose AOA was earlier than 11 years generally performed better than those whose AOA was later than 21 years. This finding is supported by Piske, Mackay and Flege's 2001 study. The researchers proposed that even though other variables such as length of residence are parceled out, AOA remains the most critical factor in predicting degree of foreign accent.

Physiological basis of the critical period hypothesis (CPH)

The idea that 'younger is better' is derived from the CPH, and should be supported by physiological evidence. Related to that, the end point of "critical age" is often discussed. One of the physiological reasons is speculated to be the lateralization effect. The claim is the following: Brain function is lateralized in two hemispheres as a child is growing up, and when lateralization is complete, the hemispheres have separate functions. It is believed that after lateralization is complete, it is difficult to learn new things (e.g., L2), since brain plasticity has become limited. Another possible physiological basis would be myelination (Long, 1990). Different areas of the brain become myelinated as a child grows, and accumulation of myelin contributes to the maturation of behaviors. When myelination is complete, we believe that maturation of the brain ends, too.

CPH: whys and why-nots

Reduced brain plasticity and completion of myelination in the brain, which might indicate the end of critical age, usually coincides with puberty. However, the end point of

the critical age in SLA is controversial: Is there such a thing as an end point; and if so, when is it? For example, Krashen (1973) claims that lateralization is complete at the age of 5, not at puberty. Others claim that there are different critical ages for syntax, phonology, and other aspects (Seliger, 1978). For example, Flege, Yeni-Komshian and Liu (1999) suggest that age affects phonology more than morphosyntax. Finally, Walsh and Diller (1981) claim that the cells responsible for different linguistic aspects mature at different rates.

In addition, not everyone agrees with the CPH. For instance, Flege (1987) provides several arguments against the CPH. First of all, the CPH was devised to explain first language acquisition. It explains why children who suffer from aphasia can regain linguistic ability, but adults cannot. In addition, it seems that there is no abrupt discontinuity around puberty in accentedness. Rather, the increase is quite linear as AOA increases. Flege and Fletcher (1992) also claim that intersubject variability on foreign accents is greater in AOA of 17-39 years group than AOA of 3-15 years group. Tahta, Wood and Lowenthal (1981a), on the other hand, claim that the AOA of 7-11 years group has the most variation in the degree of accentedness, but not in earlier or later AOA groups.

Furthermore, studies show that adult speakers can attain native-like pronunciation. In Ioup and colleagues' 1994 study, two adult participants were rated as natives in the production and perception of Arabic. One of the participants, Julie, did not have any formal instruction of Arabic before she was exposed to a TL country. Obler (1989) also reports an exceptional speaker who learned several different languages after puberty and attained native-like proficiency. Finally, Bongaerts (1999) and Bongaerts et al. (1997,

2000) investigated L2 adult speakers of different L1 backgrounds, and reported that some speakers attained native-like pronunciation in sentence reading tasks and spontaneous conversation. It seems that there are some, but not many, exceptional speakers. However, Birdsong (1999) claims that almost 30 % of his participants in French speech tests reached native-like proficiency, and we cannot ignore these participants as outliers.

In conclusion, the age issue is not easily agreed upon (whether being old is good or bad for adult speakers of L2). It seems that many other variables exist (chronological age, LOR, motivation, etc). In addition, one reason for such different results might be the use of different test formats, speech segments, and L1 background.

Length of Residence (LOR)

Even though LOR seems to be one of the most heavily studied factors affecting SLA, results are not conclusive. Length of residence (LOR) is defined as the number of years spent in a country where L2 is a predominant language (Piske, Mackay & Flege, 2001). The LOR is closely related to AOA, but their effects on SLA seem different. For example, Flege and Fletcher (1992) claim that LOR and foreign accents have a negative correlation. Oyama (1976), on the other hand, claims that the degree of accuracy in L2 pronunciation is not correlated strongly with LOR, but it is with AOA. Finally, studying Italian bilinguals, Piske, Mackay, and Flege (2001) found that LOR and foreign accents had a negative relationship in a simple correlation analysis, but when AOA was parceled out, LOR did not show any significant correlation.

The reason that LOR was found as a non-significant factor is speculated as follows: First, after a certain age, the amount of input does not affect the L2 proficiency significantly, which supports the CPH. Before that age, LOR might affect the L2 significantly (Flege & Liu, 2001). Second, the amount of L1 use varies greatly among

bilinguals. As Piske, Mackay and Flege (2001) show, when L1 use was parceled out in a correlation analysis, LOR was no longer significantly correlated with the degree of accentedness. Third, the quantity of input might not be as important as the quality of input. For example, in Flege and Liu's study (2001), late Chinese bilinguals were divided into a group with short LOR (less than 3.8 years)/long LOR (more than 3.8 years), and a student/nonstudent group. Participants took several tests, including an English stop identification test and a listening test. The results showed that long LOR did not guarantee success, but when *student* and *nonstudent* variables were added in the analysis, the group with longer LOR performed significantly better than the group with shorter LOR. Furthermore, the student group as a whole performed better than the non-student group. Flege and Liu concluded that input quantity differed between school and nonschool settings, and might contribute to the differences between the student and nonstudent group.

Other studies claim that long LOR has a positive effect on L2 pronunciation. For example, Riney and Flege (1998) report that some Japanese students showed significant improvement in English /ɹ/ and /l/ perception and production after 4 years of college where English input was abundant. However, this result should be taken carefully, because these students did not reside in a predominantly English speaking country (although two of them who improved significantly stayed in the U.S. for a year). The factor here seems more to be L2 experience, rather than LOR.

Sometimes L2 experience is considered a factor, but no specific definition has been made. It could mean the number of years of formal education, the number of traveling experiences to a TL country, or LOR. However, researchers agree that, 'The more

experienced, the better.’ For example, Port and Mitleb (1983) report that Jordanians who lived in the U.S. produced /p/ more accurately, produced longer vowels before /b/ than /p/, and produced more flaps in intervocalic position than Jordanians in Jordan.

In conclusion, LOR does not seem to be a very significant factor in L2 pronunciation. Most studies cited above report that when AOA is parceled out, the LOR effect on foreign accents becomes nonsignificant. However, in the initial stage of residence, a significant LOR effect is evidenced. Flege and Fletcher (1992) suggest that their Spanish participants showed a LOR effect (LOR 0.7 vs. 14.3); in other words, the group with long LOR had less noticeable accents. However, in Flege’s earlier study (1988, cited in Flege & Fletcher, 1992) on Taiwanese participants (LOR 1.1 vs. of 5.1), the difference was not significant. In other words, in the Spanish bilingual study, the LOR gap between two groups of participants was great enough to observe a difference. The authors suspect that after a certain period, L2 speakers reach a plateau, and LOR does not affect L2 proficiency much, a process which is often referred to as ‘fossilization.’ Moyer (1999) concluded in the research on LOR that it has a positive correlation with perceived success on L2, and subsequently a person becomes “comfortable” in L2 after years of residence, which might contribute to the lack of further learning.

Language Use (First Language/Second Language Use)

It seems that L1 use and L2 use are correlated negatively. If people use their L1 more, they tend to use their L2 less and vice versa (Flege, Munro & Mackay, 1995a). Many studies (Tahta, Wood & Lowenthal, 1981a) show that L2 use at home, at work, or with friends has significant association with foreign accents: The more the L2 is used, the better is use of the L2. For example, Flege, Munro and Mackay (1995a) found that

language use at work, at home, or with friends was the second major factor in accentedness (after AOA), although the percentage of each L2 use was different between males and females. In other studies (Flege & Fletcher, 1992; Oyama, 1976), L2 use was not found to be a significant factor.

However, one important thing is usually ignored, namely that L2 use is often confounded with AOA: Early bilinguals use their L2 more than late bilinguals (Flege, Munro & Mackay, 1995a). As such, Thompson (1991) does not consider L2 use and formal education as factors: “the younger the age at arrival, the more likely the individual was to have had more education in English and to use English socially and at home, such as marrying an English speaker, having English-speaking children, and socializing with English-speaking friends” (p.193). However, Thompson implies that if L1 use and L1 fluency are maintained at a high level, it is unlikely for L2 speakers to attain native-like pronunciation. Finally, Piske, Mackay and Flege (2001) examined Italian bilinguals and noted that even if AOA was parceled out, L1 use remained as a significant factor.

Other questions still remain to be answered, such as why the increase of L2 use affects the degree of accentedness. One possible reason would be integrative motivation. Tahta, Wood and Lowenthal (1981a) found that using the L2 at home was the significant factor predicting the degree of foreign accents in the AOA 7-12 group. Using L2 at home indicates that L2 speakers want to be part of a target language community and be away from the L1 culture. Second, do L1 and L2 proficiency always have a negative correlation; in other words, it is not known whether high L1 use always means low L2 use. Thompson (1991) implies that they do, but in Piske, Mackay and Flege’s study (2001), L1 proficiency was not significantly correlated with foreign accents.

It is still inconclusive whether the use of L1/L2 could be a significant factor in the degree of accentedness, even when AOA is controlled. It is also not clear whether L1 and L2 proficiency are always negatively correlated.

First Language (L1) and Second Language (L2) Phonological Differences

L1 and L2 phonological differences are considered among the biggest influences on foreign accents, and they are less likely to be avoidable at the beginning stage of learning. A great number of studies in this aspect have been carried out for a long time.

Particularly, in the 1950s and 1960s, contrastive analysis of linguistic differences among languages was popular, and the analysis supported the contention that language problems arose if the L2 and the L1 had phonological differences (Larson-Freeman & Long, 1991).

One example of contrastive analysis would be that Spanish speakers of English tend to insert /e/ before initial /sk/ consonant clusters, since Spanish does not allow those (e.g., [skul] is pronounced as [eskul]). The L1 influence is often called ‘interference,’ implying a negative effect of the L1 on the L2. However, contrastive analysis is found to have several limitations. First, it does not predict all problems in SLA, and sometimes overpredicts problems. For example, Briere (1966, as cited in Major, 1987) found that when English speakers imitated French, Arabic, and Vietnamese, they had problems in /R/ or /B/; but not in /x/, which was not found in English either. It seems that other factors affect L2 pronunciation. Moreover, similarity between L1 and L2 also causes problems.

Similarities between L1 and L2

Similarities between L1 and L2 are often ignored in perception by L2 speakers, which causes foreign accents: New sounds are easier for L2 speakers to master. It seems that perceptually new sounds are at first difficult to learn, but eventually they are

produced correctly; however, perceptually similar sounds to L2 are not learned easily. Flege and Hillenbrand (1987b) observed that English speakers of French could not produce French /u/ correctly, but they produced /y/, which was not found in English, more correctly. English /u/ has a higher second formant than that of French, but English speakers failed to notice this subtle difference between English and French and did not produce /u/ accordingly.

As for problems in perception, Bradlow et al. (1997) observed that Japanese speakers had relatively greater difficulty in perceiving /l/ than /ɺ/. The authors suspected the reason for this is that Japanese has only one liquid, and it is more similar to English /l/ than /ɺ/. In addition, one of the allophones of the Japanese liquid is not a retroflex but a flap (Ingram et al., 1995).

Markedness and Developmental Factors. Not only differences and similarities between L1 and L2, but also other developmental factors affect second language phonology. One of these is markedness. It is considered to represent linguistic complexity or relative infrequency among languages (however, the definition of markedness is rather circular; see Ohala, 1990). For example, final position in a word is more marked than initial position, fricatives are more marked than stops, consonant clusters are more marked than singleton consonants, etc. Many times, markedness and linguistic differences between L1 and L2 work together in L2 phonology. Broselow (1983) notes that epenthesis patterns for initial consonant clusters differ depending on which dialects of Arabic a speaker uses. For example, when pronouncing *flow* in English, Iraqi Arabic speakers tended to produce it as [iflo], whereas Egyptian Arabic speakers produced it as [filo]. The author speculates that the reason for this is the influence of different syllable

structure of Iraqi and Egyptian Arabic: Iraqi dialect prefers VC syllables, and Egyptian dialect prefers CV syllables. In addition, Tarone (1980) examined Korean, Cantonese and Portuguese speakers of English and found that they modified syllable structures through deletion or insertion, depending on their L1 structure. Regardless of their L1, they preferred open syllables, which are universally unmarked.

Eckman (1977) also indicates that the issue of whether L2 speakers will have difficulties in pronouncing L2 words depends on markedness of the L1- L2 structure (Markedness Differential Hypothesis): if areas of the L2 are different and more marked than those of the L1, learning those areas will be difficult, but if L2 areas are different but less marked, learning will not be difficult. Larson-Freeman and Long (1991) provided examples of Eckman's hypothesis: English speakers would not have problems distinguishing /ʃ/-/ʒ/ in initial position in French, even though /ʒ/ is found only in medial and final position in English. Since the initial position is less marked than the medial and final positions, English speakers of French would not have problems producing /ʒ/ in the initial position. In addition, Major and Faudree (1996) examined Korean speakers of English and claimed that they had problems in producing obstruents in final position, even though Korean has obstruents in initial position. Lastly, Eckman and Iverson (1994) argue that the most difficult context in which L2 speakers produce coda consonants is before consonant onset (i.e., *C# #C*).

The L2 phonological system also affects L2 performance, and at some point, speakers produce a segment with an intermediate value between the L1 and L2. For example, Port and Mitleb (1983) claimed that experienced Jordanian speakers of English produced longer voice onset time (VOT) of /p/ in initial position than inexperienced

Jordanians, but not with the same value as that of native speakers. It seems that the more experienced learners are, the better they produce the L2. On the contrary, Gass (1984) noticed that an Italian speaker of English produced /b/ with longer VOT than native speakers did. The author suggests that L2 speakers sometimes “overshoot target language norms” (p. 70).

In summation, L1 influence on L2 is not avoidable, especially to adult L2 speakers, but it is not always negative, nor does it represent the only source of problems.

Depending on markedness, some L2 segments can be easier to learn. We can also observe L2 developmental effects. In addition, similarities between two languages do not always seem beneficial. L2 speakers tend to classify an acoustically different L2 sound according to their L1 phonemic category, which causes foreign accents. They, as Flege (1995) indicates, need to make the L1 and the L2 contrastive, and to do that an L2 phonetic category should be “deflected away” (p. 239) from an L1 category. Special instruction, in this case, helps L2 speakers detect the subtle differences between the L1 and the L2 and therefore change the L2 phonetic category. The subsequent question is whether improvement in perception is accompanied with improvement in production.

CHAPTER 3 RELATIONSHIP BETWEEN PERCEPTION AND PRODUCTION

The possible relationship between perception and production has been an important topic among researchers. From first language (L1) to second language (L2) studies, many have argued that perception and production are related (Flege, Munro, & Mackay, 1995b). Since adult L2 speakers already have their L1 established, their L2 acquisition seems much harder, and their production is generally deviant and nonnative-like. But is a L2 acquisition problem this simple? First of all, we do not know the source of the problems; we do not know if this is a perception or production problem, if any supporting evidence can be found in L1 acquisition studies, or if training helps to reduce L2 problems? These questions should be addressed to understand L2 production problems, and our study, by examining the perception training effect, will contribute to the field of second language speech perception and production.

Speech Perception Theories

There are three major hypotheses on how we perceive sounds. First, motor theorists argue that we perceive “intended” gestures from signals, and speech perception and production are linked with a specialized module, which converts acoustic signals into gestures automatically (Liberman & Mattingly, 1985). On the other hand, acoustic/auditory theorists claim that we perceive a gross shape of spectrum of acoustic cues that are invariant, regardless of context (Diehl, Kluender, & Walsh, 1990). Finally, direct-realists claim that we perceive “actual” articulatory gestures directly, without having any modules to link perception and production (Best, 1995).

There have been many studies to support each hypothesis. For example, categorical perception behavior is typical in human speech. However, it is also demonstrated that other mammals, such as chinchillas, can perceive sounds categorically (Kuhl & Miller, 1978). This implies that speech sounds have natural psychological boundaries, so that even chinchillas can detect acoustic differences. Other studies show that acoustic quality of a sound changes based on its surrounding segments, namely coarticulation phenomena. Even so, humans easily perceive a speech segment distinctively regardless of its surrounding segments, which implies that humans have a predisposition to perceive gestures, rather than acoustic cues (Liberman & Mattingly, 1985).

Relation Between Perception and Production in L1 Acquisition

In first language (L1) acquisition, it is obvious that children perceive sounds much earlier than they produce them. Many claim that infants can distinguish subtle acoustic differences. For example, one-month-old infants increased sucking rates when stimuli were changed from /pa/ to /ba/ (Eimas et al., 1971, as cited in Jusczyk, 1992). In addition, 6-month-old or younger infants can distinguish nonnative sounds (Werker & Tees, 1984a). However, the production of meaningful sounds by infants is not evidenced until after their first year of life.

During the first linguistic stages, children produce un-adult like words, a situation which is universal. When they do it, it is not because they perceive adult words incorrectly, but because their motor skills have not developed yet. For example, children have to learn to coordinate the movement of the tongue with other muscles simultaneously, like the jaw (Kent, 1992).

According to Kent (1992), production is developed and shaped by ambient language input, since children use innate systems to recognize linguistic codes as speech

sounds, store acquired sounds in memory and modify those according to an ambient language. Babbling of nonnative sounds ceases to exist, and children start to produce native sounds as they develop. After L1 acquisition is completed, these systems are claimed to become dormant (for the argument about this issue, see Birdsong, 1999).

Speech disorders in L1. Apart from normal language development, children with misarticulation problems (CM) shed light on the perception and production link.

Traditionally, it has been believed that CM have mere articulation problems, but now the idea that the problem has a perceptual origin, rather than an articulatory origin, has gained acceptance, although it remains controversial (Rvachew, 1994). For example, Locke (1980) reported that CM did not have any difficulty in distinguishing a misarticulated sound and a correct sound in several discrimination tests. On the other hand, Weiner (1967) proposes that auditory discrimination ability and articulatory proficiency have a strong correlation, if a child is younger than 8 or 9 years old and has sizeable articulation defects. In addition, Monninn and Huntington (1974) provide evidence that children who had problems in producing English /ɹ/ could not distinguish /ɹ/ and /w/. On the other hand, the children did not show any differences from normal children in perceiving other sounds.

Several studies on CM using perception training also support the contention that the problem lies in perception, rather than solely in production. Jamieson and Rvachew (1992) hypothesize that children with misarticulation problems do not have the same model of sounds as that of normal children. The authors found that CM had problems in both perception and production of English fricatives /s, ʃ, θ/. Consequently, Jamieson and Rvachew trained the children with identification tests using synthetic stimuli. Training

stimuli changed gradually from the most obvious contrast to the least obvious contrast (i.e., fading technique), varying major spectral amplitude peaks of the noise portion of the fricatives. After the training, most children who had participated improved in both perception and production. In addition, Rvachew (1994) trained CM with perception training as well as conventional production training (e.g., imitation of minimal pairs). Children who had problems in the production of /ʃ/ participated. They were divided into three training groups depending on the stimuli: the “shoe-Xoe” group, where X is a misarticulated version of /ʃ/; the “shoe-moo” group; and the “cat-Pete” group. In this study, natural stimuli were used. Results showed that children in “shoe-Xoe” group perceived and produced better in the posttest than in the pretest, but the children in other groups did not improve significantly. The author suggests that perception training gave CM an opportunity to establish a correct model for self-monitoring.

From these training studies, Rvachew and Jamieson (1995) suggest that L2 speakers and CM are alike: They both have production problems, which originate from perception problems, and their problems in perception occur either because “neither member of contrasting pair is not present in the underlying system (e.g., /θ-ð/ problems of French speakers),” “both members of the contrasting pair belong to a single category in the underlying system (e.g., /s-ʃ/ distinction problems of some CM),” or “both members of the contrasting pair exist as separate categories in the underlying system, but they are differentiated in terms of nonstandard cues, or non-standard values of a standard cue (e.g., English voiceless and voiced distinction by French speakers)” (p. 413-414).

The authors recommend perception training in order to resolve the discrepancy between a pre-existing system and a target system.

In conclusion, L1 acquisition studies support the idea that perception and production are related, and ‘perception precedes production.’ In addition, studies on children with misarticulation problems show that perception problems are the origin of some production problems.

In L1 studies it is well supported that acquisition of perception precedes that of production. The question remains as to why L2 speakers have foreign accents, since they already know how to produce using a perception system: we do not know if their L1 system prohibits them from learning an L2, or if they have perception difficulties.

Perception and Production in L2

When the L1 and the L2 have perceptually similar but different acoustic qualities of a speech segment, problems arise. Stop segments are well studied in this perspective, because every language has stop consonants. However, their acoustic properties are different from language to language. For example, Lisker and Abramson (1969) report that different languages have different voice onset time (VOT) values of stop segments. For example, Spanish and English both have two VOT values for stops as phonemic (i.e., voiced vs. voiceless). The Spanish median values of VOT range from –100 to +10 ms., but the English median values of VOT range from +10 to +75 ms. Due to this difference, some L2 speakers have difficulty in correctly perceiving L2 stops. Gass (1984) note that nonnatives’ perception of English stop sounds is continuous rather than categorical. In a VOT continuum of synthesized /b/ and /p/ stimuli identification (from –20 to +55 ms.), nonnative speakers of English, including L1 speakers of Italian, Portuguese, Farsi, Korean and Japanese, showed less monotonic and gradual changes of

identification functions. On the other hand, native English speakers (NE) showed more monotonic curves and abrupt changes of identification functions (i.e., categorical perception).

In addition, if L1 uses only one cue, but L2 uses multiple cues to perceive a certain sound, L2 speakers seem to have perception difficulties. In the study of final obstruents, Flege and Hillenbrand (1986) looked at the trading relation between final fricative duration and preceding vowel duration in English. Native speakers of English (NE) lengthen a vowel if a following final segment is short (i.e., voiced) and shorten a vowel if a following segment is long (i.e., voiceless). The authors questioned if L2 speakers could use this trading relation in a *peace-peas* continuum. When synthetic vowels and consonants were factorially varied, NE and French speakers of English could switch from /z/ to /s/, when a word had a shorter vowel and a longer friction part. However, Swedish and Finnish speakers of English could not: they only identified /z/ when a preceding vowel was long, regardless of consonant duration. The authors implied that this occurred because Swedish and Finnish did not use multiple cues to distinguish voiced and voiceless consonants, whereas English and French did. In addition, Flege (1984) conducted the same experiment with experienced and inexperienced Arabic speakers of English: the author asked the participants to identify final /s/-/z/ of English, controlling consonant and vowel duration. The experienced Arabic speakers and NE did not show any difference in identifying the two segments, but the inexperienced speakers were different in that they did not use consonant durations, but used only vowel duration to distinguish final /s/ and /z/. The speculated reason was that Arabic has phonemic vowel length, and inexperienced L2 speakers used that L1 perceptual strategy to identify /z/.

The other possible source of the problems would be that L2 speakers do not use appropriate acoustic cues as native speakers do. Japanese speakers of English are claimed to have problems in distinguishing English /ɹ/ and /l/ (Sheldon & Strange, 1982).

Comparing native speakers of Japanese, NE and Japanese-English bilinguals, Yamada (1995) varied the first three formants (F2 and F3 onset and transition frequencies, and F1 onset duration) of a synthetic /ɹ/ and /l/ series. It was found that native speakers of Japanese tended to focus more on the F2 than on the F1/F3, but English speakers predominantly focused on the F1/F3 to distinguish /ɹ/ and /l/. The bilinguals did not focus on the F2 as much as native speakers of Japanese did. In addition, native speakers of Japanese identified /w/ more than the NE and the bilinguals did. This is not because Japanese speakers do not possess auditory processing ability. Miyawaki et al. (1975, as cited in Flege & Wang, 1989) demonstrate that both Japanese speakers and NE discriminated the F3 part of the stimuli, when the F3 was presented in isolation. Flege and Wang speculate that the problem was not in auditory processing ability but in phonetic processing ability due to phonetic differences in liquids between English and Japanese.

In addition, in a study on the production of English final /t/ and /d/, Italian, Chinese and Spanish speakers of English all had difficulty with /d/ but not with /t/ (Flege, Munro, & MacKay, 1995b). The authors assume that this was caused by “ignoring certain properties of L2 phones which are phonetically relevant” (p. 22). For example, Italian has medial voicing contrast between /t/ and /d/. However, in Italian, closure voicing and duration are more important than preceding vowel duration. Preceding vowel duration, on the other hand, is a main cue in English to distinguish final /t/ and /d/. Italian speakers

failed to notice this, and consequently they produced final /d/ in a nonnative-like manner. Final /t/ was easier to produce than final /d/ because final /t/ is less marked than final /d/.

Finally, as for vowels, Bohn and Flege (1990) find that in a synthetic continuum of *beat-bit-bet-bat*, whose formants and duration of vowels were varied, German speakers of English could distinguish English /i/ and /I/ as NE did: NE and German speakers mainly used spectral cues to distinguish two vowels. It seems that German has almost the same perceptual cues for both vowels as English. However, when distinguishing English /ε/ and /æ/, German speakers used durational cues, whereas NE used spectral cues predominantly. This might be because German does not have /æ/ (Bohn & Flege, 1990). Another explanation is that whenever spectral cues are insufficient to distinguish vowels, nonnatives use durational cues, which are language independently preferred cues. The authors call it the “language independent perceptual principle” (p. 324).

Production and perception problems in L2. Many argue that when L2 speakers have production problems, namely foreign accents, they also have perception problems. For example, Rochet (1995) observed Portuguese and English speakers’ perception and production of French vowels, /u/, /i/ and /y/. In production, approximately 50 percent of the production of /y/ was identified correctly by native French speakers. When /y/ was produced incorrectly, Portuguese speakers tended to produce it more /i/-like, whereas NE produced more /u/-like vowels. In identification tests, Portuguese and English speakers behaved differently from native speakers of French. Native French speakers identified a stimulus as /y/ when the F2 frequency was within 1300-1900 Hz. On the other hand, Portuguese speakers identified it as /i/, and NE as /u/, which followed the same pattern as observed in their production.

In terms of vowels, Flege, Bohn and Jang (1997) examine the perception and production of the English vowels /i, I, ε, æ/ by L2 speakers. They had different L1 backgrounds (i.e., German, Spanish, Mandarin and Korean) and different degrees of L2 experience (i.e., experienced vs. inexperienced). Based on native-speakers' identification and acoustical analysis of the participants' production, the authors report that the production of some L2 speakers differed significantly from that of NE. In addition, in synthetic stimuli identification tests, where formant frequencies and duration were varied, L2 speakers did not use spectral cues as much as NE did to distinguish vowels. Many used durational cues more than spectral cues. In addition, depending on the L1, the pattern of production and perception varied. For example, Koreans used longer durational differences to produce and perceive /i-I/. It is of interest to note that some experienced speakers did not show much difference from inexperienced speakers in the production of English vowels. Finally, multiple regression analysis supports the contention that perception data accounted for a significant amount of variance in the production of both /i-I/ and /ε-æ/.

Furthermore, Flege and Efting (1987) discuss that in the perception of /da/-/ta/, English speakers needed 15 ms. more VOT than Spanish speakers to give a predominant /ta/ response. When production was examined, Spanish late bilinguals tended to produce English /t/ with shorter VOT in initial position. In addition, in the production of English vowels, inexperienced Spanish speakers of English tended to produce /I/ as an /i/-like vowel, and distinguished /i/ and /I/ using durational cues mainly, which are not the main cues used by NE (Flege, Bohn & Jang, 1997). Finally, Jamieson and Morosan (1986)

acknowledge that French speakers of English tended to produce /ð/ as /d/ and /θ/ as /t/, and also demonstrate that they had perception problems of these two English sounds.

With a limited number of studies on the relation between perception and production and many perception studies, it is implied that perception and production are related. Many studies also suggest that perception plays an important role in production, and production problems result not solely from motoric difficulties, but could result from perception problems: L2 speakers tend to perceive L2 phones using their respective L1 systems, and this might be one of the sources of foreign accents.

The idea that the loss of motoric ability is the main cause of production problems has not been supported heavily, first since it is very rare that L2 speakers have perception problems but not production problems (c.f., Sheldon & Strange, 1982), and second because it is hard to devise an experiment to support this idea (Gass, 1984). The study conducted by Sheldon and Strange (1982) was an exception. They found that Japanese speakers of English could produce /ɹ/ and /l/ distinctively, but failed to perceive them distinctively. However, this result should be taken carefully, because the Japanese speakers might have formal English education to direct them to use articulatory strategies such as “to produce /l/, put the tongue on the...” (p.265, Flege, 1991). In addition, as Yamada (1995) reports, Japanese speakers did not use the same acoustic cues to identify /ɹ/ and /l/ as NE did, and showed significant differences from NE in the identification tests.

With the above findings, Flege, Munro and Mackay (1995b) hypothesize that the correct perception of L2 sounds does not guarantee the correct production of them, but the correct perception is necessary for the correct production. It seems that perception

should come before production: “an L2 phone must be perceived in a fully native-like fashion if it is to be produced in a fully native-like fashion” (p.22). From this standpoint, Flege develops the Speech Learning Model in L2.

Second Language Speech Learning Model

Flege’s Speech Learning Model (SLM)

Since many studies show that perceptual similarities also cause problems (Flege & Efting, 1987), Flege (1995) hypothesizes that perceptually similar sounds in an L2 are more difficult to acquire than dissimilar sounds, and L2 speakers tend to merge similar L1 and L2 sounds together and identify them in the same way, which in turn causes foreign accents. In other words, L2 speakers tend to assimilate an acoustically different L2 sound to an equivalent L1 sound. For example, Korean speakers of English tend to produce English /æ/ as /ɛ/, because /æ/ sounds similar to Korean /ɛ/. This ‘equivalence classification’ hypothesis (which later became part of the SLM) is closely related to AOA: As AOA increases, L2 speakers have a harder time perceiving phonetic differences between L1 and L2 (Flege, 1995).

Best’s Perceptual Assimilation Model (PAM)

Additionally, Best (1995) proposes the ‘Perceptual Assimilation Model,’ and hypothesizes that when perceiving nonnative sounds, L2 speakers assimilate them to a L1 category. When nonnative sounds are in ‘single category assimilation’ (i.e., two nonnative sounds are categorized as one native sound, although they both are deviant from the native category) or in ‘category goodness differences’ (i.e., two nonnative sounds are categorized as one native sound, although one is closer to the native category than the other), discrimination of the two sounds is difficult. On the other hand, if two

nonnative sounds are not categorized based on any L1 category, discrimination of two sounds should be easy (e.g., click sounds in Zulu).

Define “New” and “Similar”

The question is how to define ‘different’ and ‘similar’ sounds and how much they are new/similar to equivalent L1 sounds. Many suggest different ways to do it, but there is little agreement. For example, Flege (1991) suggests two methods. First, a sound is new if it is not represented by any phonetic symbols in the L1 sound inventory. However, not all phoneticians agree with ‘phonetic symbols,’ and many times symbols are not detailed enough. For instance, /i/ and /I/ in English are sometimes represented as /i:/ and /i/. However, in reality they are different spectrally as well as durationally. If we use /i:/ and /i/, /i/ is not a new sound to some L2 speakers, because it is a short version of /i:/. The second method is to use acoustical analysis. However, it is also hard to decide what kind of acoustical analyses to use, since different analyses have different functions (Rochet, 1995). In terms of how closely a sound corresponds to a L1 category, many use a Likert scale to see the degree of similarities (e.g., Schmidt, 1996).

Effectiveness of Short-term Laboratory Perception Training

In order to ease the difficulties that L2 speakers have and to examine the relation between perception and production in L2, many researchers have tried to give short-term laboratory training to L2 speakers. A number of studies seem to have been mildly successful. For example, Werker and Tees (1984b) trained English speakers to perceive Hindi breathy voiced vs. voiceless aspirated stops (voicing contrast), and dental vs. retroflex stops (place contrast). After a short training, the participants could perceive voicing contrast, but not place contrast. In addition, the authors found that after one year

of learning Hindi, English speakers could perceive voicing contrast, but not place contrast.

Another example of the training of stop consonants in an L2 was provided by Pisoni and his colleagues' 1982 study. They trained English speakers to perceive prevoiced, voiceless unaspirated and voiceless aspirated stop distinctions (VOT -70 , 0 , $+70$ ms., respectively). After one hour of training, half of the participants reached 85 percent accuracy in identifying the three stops. In addition, McClaskey et al. (1983) trained English speakers to discriminate synthesized voiced and voiceless aspirated stops (VOT from -70 to $+70$ ms.). One group was trained with labial stops, and the other group with velar stops. It was found that after training, most of the participants could discriminate both labial and velar stops. Furthermore, they could discriminate alveolar stops, which was not part of the training. Finally, Rochet and Chen (1992, as cited in Rochet, 1995) trained Mandarin Chinese speakers to identify a synthetic /pu/-/bu/ distinction (starting from VOT -60 to $+90$ ms.) with a fading technique, which was devised by Jamieson and Morosan (1986). The speakers showed significant improvement in identifying /pu/ and /bu/ in the posttest. In addition, they could transfer the ability to stimuli that were not part of the training. However, they failed to show a significant difference in the perception of naturally produced voiced segments in the posttest, although they could perceive naturally produced voiceless segments significantly better than in the pretest.

Besides stop segments, fricative consonants are also frequently trained segments. Jamieson and Morosan (1986, 1989) conducted research on training French speakers of English to perceive /ð/ and /θ/. First, they used a fading technique (1986), where stimuli

were sequenced from the most acoustically distinct stimuli to the least distinct stimuli. The stimuli were synthesized to contain varying degrees of friction noise duration, from the longest gap between two segments to the shortest gap. The speakers who took the training showed improvement with synthetic and natural stimuli in both identification and discrimination posttests.

Another example of the use of synthetic stimuli is found in teaching the English /ɹ/ and /l/ contrast. Strange and Dittmann (1984) trained Japanese speakers using a synthetic /ɹak-lak/ continuum, varying F1, F2 and F3. Training was composed of AX discrimination tests. After the training, the participants showed significant improvement in synthetic stimuli identification and oddity discrimination in /ɹak-lak/ and /ɹejk-lejk/ continuums. However, they failed to identify natural stimuli of /ɹ-l/ minimal pairs. Two reasons the authors speculated are, first, that English /ɹ/ is intrinsically difficult and rare, and second, that the difference between /ɹ/ and /l/ is spectral, rather than durational.

However, Bradlow and her colleagues (1997, 1999) succeeded in training Japanese speakers to perceive English /ɹ/ and /l/, using multiple talkers, natural stimuli and identification tests. The training lasted 20-30 minutes in each session, 15 hours in total. After the training, the participants improved significantly both in perception and production, and the effect of training lasted three months after the training.

Other examples of perception training leading to the improvement of perception as well as production are studies by Jamieson and Rvachew (1992) and Rvachew (1994). They both trained children with misarticulation problems in English fricatives. English

fricatives /s, θ, ð/ are different in terms of the frequency of the spectral amplitude peak of noise portion. The differences between the two studies are, first, that Jamieson and Rvachew used synthetic stimuli, whereas Rvachew used natural stimuli. Second, the former investigators used perception training, but the latter used both perception and production training. Third, the duration of training was different (2 hours vs. 8 weeks, respectively). Both studies showed significant improvement both in the perception and the production of English fricatives.

In summary, short-term laboratory perception training seems helpful for improving perception of, and possibly production of, nonnative sounds. In addition, some training methods are better than others, and one of the more effective methods is giving natural stimuli intensively.

All in all, the positive aspect in the training studies is that L2 adult speakers can change their L2 sound category through training, which could also trigger more native-like production. If perception training is effective, it could be applied to train to perceive any kind of sounds distinctively in the L2. In addition, perception training gives hope to adult learners. They can improve their perception through training after critical period. In our study, the acquisition of English obstruents, especially alveolopalatal fricatives and affricates, by Korean speakers was examined. In order to do this, it is of importance to examine briefly the phonetic differences between Korean and English.

CHAPTER 4 KOREAN AND ENGLISH OBSTRUENTS

The present study focuses on the acquisition of English alveolopalatal sounds by Korean speakers. Since the pattern of production errors partly depends on L1 background, it is necessary to note the differences and similarities between Korean and English obstruents.

Differences between Korean and English Obstruents

Some might say that Korean and English obstruents sound similar, because their respective places of articulation seem rather similar. In reality, the nature of obstruents in the two languages is quite different. First of all, the number of obstruents is different. English has 6 stops (/p, b, t, d, k, g/), 9 fricatives (/f, v, θ, ð, s, z, ʃ, ʒ, h/) and 2 affricates (/tʃ, dʒ/) as phonemes. In contrast, Korean has 9 stops (/p^h(aspirated), p (lenis), p^ʰ(fortis), t^h, t, t^ʰ, k^h, k, k^ʰ/), 3 fricatives (/s, s^ʰ, h/) and 3 affricates (/tʃ^h, tʃ, tʃ^ʰ) (Kim, 1999). All obstruents are voiceless in Korean. In terms of a syllable structure, almost all obstruents are allowed in word initial and final position in English, but only lenis voiceless stops are allowed in final position in Korean; all obstruents are neutralized in final position into /p, t, k/, depending on their place of articulation (Kim & Jongman, 1996). Consonant clusters are not allowed in any position in Korean (Yoo, 1996).

Stops

As Table 4-1 shows, English and Korean voiceless stops /p, t, k/ share the same places of articulation, namely, bilabial, alveolar and velar. Even so, their acoustic

properties are different. In word or syllable final positions, for example, English stops are either released or nonreleased, i.e., whether the air stream is released after closure (Ladefoged, 2001). On the other hand, Korean stops are largely nonreleased (Kim, 1999). The definition of ‘releasing’ is debatable, but Kim (1999) defines it as “the removal of oral closure followed by a pulmonic egressive air stream flowing through the oral tract” (p. 362). In a linear predictive coding (LPC) spectrum, English released final stops have a sustained energy peak plateau. When unreleased, the stops have a sharp peak without a sustained energy plateau, or silence. On the other hand, Korean has a sharp peak in the word final without a plateau, which indicates the nonreleasing nature of final stops. In addition, the release after the final stop closure in Korean is followed by a velaric ingressive air stream, rather than a pulmonic egressive air stream.

In word initial position, the two languages differ in duration of voice onset time (VOT). Lisker and Abramson (1969) found that English has two phonemic values of VOT, with ranges from -155 to $+130$ ms. Voiced segments are produced with either lead VOT (prevoiced) or short lag VOT. Aspirated voiceless segments are produced with long lag VOT. Many times initial /b, d, g/ are produced with little voicing, and their VOT, being less than 10 ms., is close to that of unaspirated voiceless stops (Ladefoged, 2001). On the other hand, Korean has 3 phonemic VOT values, ranging from $+7$ to $+126$ ms. Lenis and fortis stops do not differ significantly in VOT values, but aspirated stops have significantly longer VOT than lenis and fortis stops (Lisker & Abramson, 1969). Interestingly, VOT range of Korean lenis and fortis stops falls within the English voiced stop VOT range. Therefore, those Korean stops can be heard as English voiced stops, and

Koreans possibly use their VOT value of the fortis or lenis stops when producing the English voiced stops.

Two studies provide the acoustic characteristics of Korean phonation. Cho, Jun and Ladefoged (2002) examine several acoustical correlates of Korean stops produced by Seoul and Cheju dialect speakers. The authors find that Korean stops are significantly different in terms of VOT, fortis stops being the shortest and aspirated stops being the longest. However, there is a dialectal difference: Seoul speakers use VOT more to identify 3 kinds of stops. Cheju speakers, on the other hand, use the fundamental frequency (F0) more: aspirated stops have a following vowel which has higher F0 than fortis stops, whose following vowel has higher F0 than that of lenis stops. In addition, the researchers find that the aerodynamic mechanism of vowels following the stops is different. Vowels following fortis stops are usually laryngealized due to the following abrupt adduction of the vocal folds. In contrast, vowels following aspirated stops are usually breathier due to the following less abrupt adduction of the vocal folds. Kagaya (1974) also examines Korean stops and reaches a similar conclusion.

Table 4-1. Korean and English obstruents phonemic inventory

		bilabial	labio-dental	inter-dental	alveolar	alveolo-palatal	velar	glottal
Stop	Korean	p ^h , p,			t ^h , t, t'		k ^h ,	
	English	p, b			t, d		k, g	
Fricative	Korean	s, s'						
	English		f, v	θ, ð,	s, z,	ʃ, ʒ,		
Affricate	Korean				tʃ ^h , tʃ,			h
	English				tʃ		tʃ, dʒ	h

Fricatives

Most of the English fricatives are not found in Korean. First of all, English has 9 fricatives, whereas Korean has only 3 fricatives. English fricatives /f, θ, s, ʃ/ have voiced counterparts, which Korean does not. English fricatives are distinct in terms of frequency, amplitude and duration of noise portions. For example, centroid frequency in the frication part of /s/ is higher than that of /ʃ/ (range: 3500-5000, 2500-3500 Hz, respectively).

Amplitude of the noise portion of the two is similar (63-64 dB). On the other hand, /f/ and /θ/ have diffused energy and exhibit lower amplitude than /s/ and /ʃ/ (Behrens & Blumstein, 1988). /s/ and /ʃ/ are also different in terms of the place of articulation, /s/ being apical or laminal alveolar and /ʃ/ being apical or laminal domed palatoalveolar (Ladefoged & Maddieson, 1996). Their articulatory setting is also different. /ʃ/ has lip rounding, which contributes to lowering the formant frequencies, and has a sublingual cavity, whereas /s/ is unrounded and does not have (or has in reduced size) a sublingual cavity (Johnson, 1997).

On the other hand, Korean has 3 fricatives /h/, /s/ and /s'/. Korean /s/ is similar to English /s/ in terms of centroid frequency of the frication part, but in Korean /s/ becomes an allophone [ʃ] before high front vowels (Yoo, 1996). However, in the studies of Cho, Jun and Ladefoged (2002) and Behrens and Blumstein (1988), in initial position Korean /s/ is found to be shorter than English /s/. Moreover, the duration of Korean /s' is closer to English /s/ (mean value of Korean /s/-105 ms., /s'/-150 ms., English /s/-174 ms.). Furthermore, Kim (1999, as cited in Joh & Lee, 2001) examines the loan word phonology

of English /s/ in Korean, and notes that it is more often represented as /sʰ/ before vowels than before consonants. The author speculates as to why the English /s/ is longer. It has also been observed that when Koreans borrow English words that start with /s/, Koreans perceive the durational differences, and transcribe English /s/ with Korean /sʰ/ orthography.

Korean /s/ and /sʰ/ are different in several ways. First, in initial position, /s/ has brief period of aspiration, following the frication part and before a vowel, whereas /sʰ/ does not (Kagaya, 1974). As a result, the duration of /s/ is longer than that of /sʰ/. However, if the aspiration portion is excluded, the duration of /sʰ/ is longer. In particular, /sʰ/ is significantly longer in intervocalic position, and /s/ is shorter, because of the disappearance of aspiration. /s/ has a lower centroid frequency than /sʰ/ (mean value 6200 vs. 6600 Hz, respectively), which implies that /sʰ/ has smaller front cavity than /s/. Finally, /s/ is followed by breathier vowels, whereas /sʰ/ is followed by laryngealized vowels (Cho, Jun & Ladefoged, 2002).

Affricates

English has two affricates, which are alveolopalatal /tʃ/ and /dʒ/, whereas Korean has 3 affricates, which are aspirated, lenis, and fortis (/tʃʰ/, /tʃ/, /tʃʰ/, respectively) voiceless alveolar. Korean and English affricates differ in several aspects. First of all, the place of articulation of Korean affricates seems different from that of English. According to Stevens (1993, as cited in Kim, 1999), if the highest spectral peak of the frication part of an affricate is higher than, or corresponding to, F4 of a following vowel the affricate is considered alveolar. If the highest peak is lower than, or corresponding to, F4 of a following vowel the affricate is alveolopalatal. In this vein, English affricates are

considered alveolopalatal, and Korean affricates are alveolar. All other Korean coronal obstruents have similar patterns in the highest spectral peaks of the frication part (Kim, 1999). Palatograms and linguograms also show that Korean affricates are alveolar, or denti-alveolar. This finding supports others' observation of Korean affricates being alveolar or dental (Schmidt & Meyer, 1995).

Errors Made by Korean Speakers of English

Due to the differences described above, Korean speakers of English (NK) make perception and production errors in English. Generally, L2 errors vary depending on the speakers' English proficiency, linguistic context and task types (Tarone, 1984).

Furthermore, most errors that are made in the early stage of learning English are similar to the Korean loan word representation.

Stops

English and Korean initial voiceless stops have a similar range of VOT, so it is rare that Korean speakers perceive English voiceless stops incorrectly. As for voiced stops, they are perceived as either lenis or fortis Korean voiceless stops. Schmidt (1996) conducted a study on NK's perception of English sounds. In this study, the author asked NK to label English sounds using Korean orthography. The finding is that voiced obstruents were perceived as either lenis or fortis Korean voiceless counterparts.

As for production, many have noticed that NK tend to devoice voiced stops in word-final position or epenthesize a vowel after final stops (Eckman & Iverson, 1994). Major and Faudree (1996) found that in wordlist reading, NK produced voiced stops as voiced only 68 percent of the time but produced voiceless stops as voiceless 98 percent of the time.

Many times it is witnessed that NK tend to insert a vowel (unround back high vowels, /ɯ/) after final stops (Eckman & Iverson, 1994). The reason for NK epenthesis after final stops is explained with syllable structure differences: Korean does not allow any stops in final position except voiceless stops (Yoo, 1996). However, many questions still remained unanswered. For example, it is debatable whether NK insert the vowel after all final stops. In addition, perception studies have not been conducted to examine why epenthesis is triggered.

Fricatives

Since English has 9 fricatives and Korean has 3, some English fricatives sound very foreign to Koreans. In Schmidt's 1996 study, when Koreans were asked to label English fricatives with Korean orthography, there was no clear one-to-one correspondence between English initial /f, θ, ð/ and Korean orthography. For example, /θ/ is perceived either /s/, /s'/, /t'/, or /p/. In addition, even though English /s/ and Korean /s/ sound similar, NK tended to perceive English /ʃ/ as Korean /s/, when /ʃ/ preceded high vowels, and to perceive /s/ as Korean /s'/ when /s/ preceded high vowels. Joh and Lee (2001) also noticed that Korean participants had difficulty in perceiving /s/ correctly in front of high vowels.

In the production of fricatives, the pattern of substitution is similar to that of perception. In initial position, NK tend to produce /θ/ as /s/ or /t/, and /s/ as /ʃ/ before high vowels (Joh & Lee, 2001). Furthermore, /z/ and /ʒ/ are all substituted by /dʒ/ (Schmidt & Meyer, 1995). Schmidt and Meyer (1995) also notice that all English palatal sounds are over-rounded: NK tend to round their lips more than necessary in producing English /ʃ/,

/tʃ/ and /ʒ/. This could be because of the transfer from loanword phonology: English loan words which start with /ʃ/ and /ʒ/ are usually represented with either a “CwV” or a “CV” structure, where C represents Korean orthography /s/ and /tʃ/. /w/ might be added after palatal fricatives when NK produce them because of “CwV” structure in Korean loanwords. English palatal consonants are also produced more front, influenced by Korean affricates, whose place of articulation is alveolar (Schmidt & Meyer, 1995). In final position, NK tend to insert vowels after fricatives, since no fricatives are allowed in Korean word final position. This phenomenon is also widely observed in loan word phonology: All fricatives are epenthesised by a Korean default vowel /ʉ/, but in case of /ʃ/, /i/ is epenthesised. A possible explanation of the vowel difference is that both /ʃ/ and /i/ are palatal (Yoo, 1996).

Affricates

In Schmidt’s study (1996), English initial /tʃ/ is perceived as Korean /tʃ^h/, and initial /dʒ/ is perceived as Korean /tʃ/ or /tʃ^ʔ/ by NK. In the production of English affricates in initial position, since NK perceived English /tʃ/ as /tʃ^h/, they produced /tʃ/ with more aspiration (Schmidt & Meyer, 1995). English /dʒ/ was correctly pronounced most of the time. However, Schmidt and Meyer (1995) indicate that sometimes the researcher they heard /tʃ^ʔ/ for /dʒ/, but /tʃ^ʔ/ was acceptable as /dʒ/. The researchers also found that NK produced both affricates more front and with more lip-rounding. In final

position, NK tended to produce English affricates with /i/ vowels (Major and Faudree, 1996).

In conclusion, since Korean and English obstruents might sound similar but not identical, Koreans make errors in the perception and production of English. They sometimes incorrectly produce and perceive English words with obstruents. For example, final palatals often are produced with an extra vowel [i]. Even very similar segments, such as English /s/, might be produced incorrectly. It seems that both perception and production errors exist and are closely linked.

CHAPTER 5 STUDY

The goal of our study is to examine whether Korean speakers of English had perception and production difficulties in word final alveolopalatals, whether they improved their perception and production ability after training, whether Koreans sustained the learned ability, and how much perception and production of final alveolopalatals were correlated.

Hypotheses

Seven hypotheses will be examined.

First, native speakers of Korean (NK) and native speakers of English (EC) will have significant differences in the perception and production of final alveolopalatals in English.

If NK have difficulty in perceiving and producing words ending with an alveolopalatal and with an alveolopalatal+i distinctively, then NK and EC will differ significantly in both perception and production.

Second, Koreans in an experimental group (KE), who receive perception training, will do better than Koreans in a control group (KC) in the perception of final alveolopalatals in English after the training.

The training involves the perception of final alveolopalatals. If KE learn how to distinguish words with a final alveolopalatal and with an alveolopalatal+i, they will improve their perception ability.

Third, KE will do better in generalization tests than KC after the training.

Generalization tests consist of new words with a final alveolopalatal sound and a final stop sound. If KE learn how to distinguish words with a final alveolopalatal and with an alveolopalatal+i, they will perceive and produce words with a final stop correctly in English, which should be easier than a final alveolopalatal fricative/affricate. In addition, KE will perform better with new words ending with an alveolopalatal consonant.

Fourth, 3 months after the training, KE will still perform better than KC in the perception of alveolopalatals in English.

If KE learned the ability to perceive the differences, they will sustain the ability to perceive final alveolopalatals correctly.

Fifth, Koreans in an experimental group (KE), who will receive perception training, will perform better than Koreans in a control group (KC) in the production of final alveolopalatals in English after the training.

KE will extend the benefits from perception training to the domain of production of final alveolopalatals.

Sixth, 3 months after the training, KE will do better than KC in the perception of alveolopalatals in English.

KE will produce final alveolopalatals more accurately than KC 3 months after the training pretest, possibly because improved perception ability will have extended to the production domain in KE.

Seventh, there will be a significant correlation between perception and production score in each test.

If perception and production have a strong correlation, participants who do well in the perception part will also do well in the production part of each test.

Methodology

Twenty-seven native speakers of Korean (NK) participated in both perception and production tests 3 times: first the pretest, then the posttest 1, and finally the delayed posttest (posttest 2). The perception tests were devised to examine whether NK perceived an extra vowel in English words that ended with an obstruent. The perception tests included discriminating English words and modified non-words ending with an obstruent and with a consonant+vowel produced by a native speaker of English. The production tests were designed to investigate if NK produced an extra vowel in words that ended with an obstruent. The production tests involved reading words from a wordlist or naming words by reading flash cards, which was intended to elicit more spontaneous responses. The production data were then presented to a panel of native speakers of English for identification. Training for the experimental group lasted 3 weeks. The training consisted of an identification task similar to the pretest. Training stimuli were produced by 3 different native speakers of English, so that Korean participants were exposed to a wide range of acoustic quality of segments under consideration, /ʃ, tʃ, dʒ/.

Perception Tests

Perception tests were designed to examine whether NK could distinguish English words or non-words which ended with either an obstruent or a vowel.

Pretests

Subset stimuli. Four subsets of stimuli were used in the pretest. Stimuli for the subsets were either English real words or possible English non-words, which had either a

C₁VC₂ or C₁V₁C₂V₂ syllable structure where C₂ was an English obstruent (/ʃ/, /tʃ/, /dʒ/, /s/, /f/, /θ/, /p/, /b/, /t/, /d/, /k/ or /g/). /ʒ/ was not included due to its rarity. All words in the perception tests were pronounced by a native speaker of English (Age: 22 years) in the University of Florida Linguistics program, who did not participate in any of the further recordings. Each subset, with the exception of subset 4, included 8 distractor words, 4 of which appeared initially, and the rest of which were inserted randomly in the test tokens. Initial distractor words were included for training participants to match aural and text stimuli.

Subset 1. Subset 1 involved identifying minimally different word pairs, which were all possible English non-words. This subset was composed of words ending with a fricative/affricate and with a vowel. In CVCV words, /i/ followed /tʃ/, /dʒ/ and /ʃ/. Otherwise, the fricative/affricate was followed by /ʊ/, which is a default vowel in Korean loan words. Fricatives with different places of articulation were used to examine the effect of different fricatives.

Subset 2. In subset 2, minimally different pairs of English real words, which ended with either C or Ci, were used. The final consonants were alveolopalatal fricatives or affricates.

Subset 3. Subset 3 involved minimally different word pairs, but this time one member of the pair was a real word and the other was not. English alveolopalatal fricatives and affricates in either real words or modified non-words were included. Non-words were slightly modified versions of real words. In other words, if a real word ended with a C, the modified version of the word ended with a Ci (e.g., *church* vs. *churchy*). If a

real word ended with a Ci, the modified version of the word ended with a C (e.g., *phonology* vs. *phonolog*). The total number of real words and modified words was equal.

Subset 4. Stimuli in this subset, words ending with stops, were not part of the training. NK also identified minimally different word pairs with final stops and possible English non-word pairs. These were produced with different releasing types: non-release, normal release and release followed by /u/. To examine the effect of release of air in final stops, the stimuli varied in terms of the preceding vowel (i.e., /I/ or /i/) or voicing status of final stops. One-syllable (i.e., C₁V₁) words, which had the same initial C₁V₁ as the C₁V₁C₂ stimuli, were inserted to examine whether NK distinguished words ending with a non-releasing stop and words without a final consonant. The order of all words in each set was randomized. The list of words that was used in the pretest is attached in Appendix A.

Table 5-1. Stimuli for pretest

	Stimuli
Subset 1 (non-words)	Affricates: 3 words * 2 final C (/tʃ,dʒ/) * 2 ending (C/Ci) =12 Fricatives: 3 words * 4 final C (/ʃ/, /s/, /f/ and /θ/) * 2 ending (C/Ci) = 24
Subset 2 (real words)	Affricates and fricatives: 6 words * 3 final C (/tʃ,dʒ,ʃ/) * 2 ending (C/Ci) + 1 repetition of 5 final C and 5 final Ci = 46
Subset 3 (real+non-words)	Affricates and fricatives: 3 words * 3 final C (/tʃ,dʒ,ʃ/) * 2 (words) * 2 (modified non-words) + 1 repetition of 3 final C and 3 final Ci = 42
Subset 4 (non-words)	Stops (C ₁ VC ₂ , C ₁ VC ₂ V): 6 C ₂ (/p,b,t,d,k,g/) * 2 preceding vowels (i/I) * 3 ending (release/ nonrelease/release with /u/) =36 C ₁ V: 3 words * 2 preceding vowels (i/I) =6

Talker recording. The recording of production was carried out individually in a sound attenuated room. A unidirectional head-mounted microphone (Shure SM 10A) and a SONY TCD D8-DAT recorder were used to capture stimuli. Then, data were transformed into WAV format using a Kay Lab CSL 4400 station and stored in an IBM

computer. The recording was redigitized at a sampling rate of 22.05 kHz and 16-bit quantization. All tokens were normalized for intensity with UAB software. Peak amplitude of all stimuli was normalized with 90 percent scale.

Participants. Participants consisted of 27 native Korean speakers of English (NK) who lived in Gainesville, FL. Their lengths of residence and ages of arrival varied. However, all of them had come to the U.S. after puberty, and they were all in their late 20s or early 30s (Mean=29.26, Range: 21-41 years). All of the participants had finished 6 years of English education in Korea before coming to the U.S. Both females and males participated (Female= 17, Male=10). Even though the participants all lived in the U.S., their L2 use was limited. Most participants self-reported that they used English less than 50 percent a day. The participants were divided into two groups: An experimental group and a control group.

Fifteen Koreans participated in the training as the Korean Experimental Group (KE). They received 3 weeks of intensive perception training of English alveolopalatals. The Korean control group (KC) of 12 Koreans did not receive any training. Another control group consisted of a group of native speakers of English (EC). EC were needed to establish the baseline of test difficulty. Eleven American English speakers, who were undergraduate students at the University of Florida, participated in the pretest (Mean age: 19.64). Both females and males participated (Female: 6, Male: 5). All members in EC did not have any sustained contact with Korean speakers or Korean language. KE and KC were paid for their participation, and EC received an extra credit for their contribution. The background information for KE and KC is attached in Appendix B.

Procedure. All perception tests were carried out individually. All directions to Korean participants were given in Korean by the researcher. EC were given directions in English. The participants were situated in a quiet room and provided with a headset (either Sennheiser HDC 451, or SONY MDR-V150). All stimuli were presented on the computer screen using UAB software, developed at the University of Alabama, Birmingham. The duration of each test was approximately 20-25 minutes. All directions and text stimuli were written in English orthography.

The participants were told to perform 4 subsets in the pretest. The order of presentation of subsets was the following: Subset 1, 4, 3 and 2. The participants did not take any breaks in between subsets. The raw score of correct identification was collected from each participant for further data analysis.

Subset 1 and 2--identification tests. In both subsets, participants were told that the task was to identify what they heard. In subset 1, they were told that that what they would hear was not English or Korean, and in subset 2, they were told that they would hear real English words. After listening to each stimulus, the participants were asked to select what they heard on the computer screen. The answer choices were written in English orthography. For example, when the participants heard /nuʃ/, the answer choices were *nush* and *nushi*.

Subset 3--correct/incorrect identification test. Here, the participants were told that the words that they would hear were either real English words or modified non-words, which were slightly different from real words. They had to decide whether the audio and text stimuli matched. The audio stimulus was presented first, and then the participants saw the visual stimulus on the screen. If they thought that they heard the

correct pronunciation of word shown on the screen, they would choose *correct*, and if not, *incorrect*. For example, if the text was *sash*, and /safi/ was pronounced, then the correct response was *incorrect*. The text stimuli were always real English words.

Subset 4--identification test. In the fourth subset, the participants were told that the words that they heard were not real English words. The participants were asked to identify what they heard among minimally different words that had either a C₁V₁, a C₁V₁C₂ or a C₁V₁C₂V₂ structure.

Posttest 1 and delayed posttest (posttest 2)

Stimuli. Stimuli for posttest 1 and 2 were the same as in the pretest.

Participants. All participants, except EC, who took the pretest came back and took posttest 1. In posttest 2, one participant from KE could not take it due to personal reasons.

Procedure. All NK came back for posttest 1. KE took the posttest 1 two or three days after their training had ended. KC took it approximately one month after the pretest. Posttest 2 was administered approximately 3 months after the posttest 1.

Generalization tests

Generalization test I stimuli. Subset 4 in the pretest was used as generalization test I in posttest 1 to examine whether the effect of training of final alveolopalatals extended to final stop identification.

Generalization test I participants. The same participants who took pretest took generalization test I.

Generalization test I procedure. The same procedure as in pretest subset 4 was used in the generalization test I.

Generalization test II stimuli. The stimuli in Generalization Test II were not part of the training. They consisted of new words produced by new speakerstalkers. New stimuli were not used in the training or in the pretest, although a small number of words from the training were included. New talkers did not participate in the recording of the pretest or the training stimuli. Two new talkers (one female and one male speaker) produced all stimuli in subset 5, and half of the stimuli from each talker were used to balance out the effect of each talker. The stimuli included words ending with an alveolopalatal and with an alveolopalatal+i. All words were either real words or modified non-words, which were slightly modified from real words: If a real word ended with a Ci, the modified version of the word ended with C, and vice versa. The format of the test was the same as subset 3 in the pretest, but there was no distractor word in this subset. The wordlist for Generalization Test II is attached in Appendix C.

Table 5-2. Stimuli for Generalization Test II (Subset 5)

Subset 5	$\int (11) + t\int (11) + d\int (12) * 2 \text{ (non-words/words)} + \text{repetition of 2 final C and 2 final Ci} = 72$
----------	--

Generalization test II participants. All participants who took posttest 1 and 2 took the generalization test. EC also took it when they took the pretest.

Generalization test II procedure. The procedure was the same as for subset 3 of the pretest, where the participants were asked to decide whether the text and aural stimuli were correctly matched. The researcher told the participants that this subset had the same format as subset 3. It was presented after subset 2 in posttest 1 and 2.

Perception Training

Stimuli

Talker recording. The stimuli for training were produced in the same manner as pretest stimuli.

Training. Training was composed of perception tasks with real English words. The words chosen were minimal pairs of words that ended with an alveolopalatal or an alveolopalatal+i. All words were searched for and found in the Merriam-Webster online dictionary. A total of 63 pairs of words were selected and used. The wordlist is attached in Appendix D. The stimuli were recorded by 3 native speakers of English who did not have any noticeable regional accents. One female and two male graduate students who were in the Linguistics program participated in the recording of training stimuli. The talkers' ages ranged from 25 to 28 years, and they all came from Florida. The word pairs were randomized in each session.

Participants

Fifteen Koreans in KE participated in a 3-week-long training session.

Procedure

Three 30-minute sessions were provided per week (however, the actual time of individual training varied each day). The total training time was approximately 4.5 hours. In Logan and colleagues' study (1991), there were 15 training sessions. Yet, the largest increment of performance was evidenced during the first ten sessions. Therefore, in our study it was expected that the participants would reach a maximum increase of improvement on the ninth or tenth session. In the training, the task was similar to pretest subset 2, except for the feedback part: KE were asked to identify whether a stimulus ended in C or Ci. Answer choices were written in English orthography. Feedback was

given in auditory and visual modalities: if their answer was incorrect, the participants heard a beep sound, and the stimulus was automatically repeated. At the same time, the correct text stimuli were blinked. The participants listened to stimuli from same talker for a week, and then proceeded to a different talker each week. All participants completed listening to all words in each session.

Data Analysis

Comparison among English control (EC), Korean experimental (KE), and Korean control (KC)

All data analyses were conducted using SPSS version 11. Raw scores of correctly identified stimuli were tabulated from each participant in each group (EC, KE and KC). For each test, the total number of correctly identified stimuli in each group was compared using a one-way analysis of variance (ANOVA). In addition, using a mixed design 2×3 ANOVA, scores of correctly identified stimuli were compared with *Time* (pretest, posttest 1 or 2) as a within subject variable, and *Group* (experimental and control) as a between subject variable. For KE and KC separately, a repeated measures one-way ANOVA, with *Time* as a within variable, was carried out for each subset as well as total scores to examine the improvement.

Pretest, posttest 1 and posttest 2

Among NK, the raw scores of correct identification in each subset of the pretest were calculated using a paired samples t-test (within subject variables: C vs. Ci) to examine the source of the incorrect perception by the Korean participants. Furthermore, among KE words ending with C and Ci were compared with *Time* as a within subject variable to examine the improvement.

Generalization tests I and II

NK (KE and KC combined), KE and KC were examined with a one-way ANOVA to explore the generalizability of the training in the generalization test I in pretest, posttest 1 and 2, and in generalization test II in posttest 1 and 2.

Production Test

In the production test, the aim was to examine whether Korean participants produced C and Ci words the same, whether KE improved their production, and whether there was a relation between perception and production. The production test was also conducted in the pretest, posttest 1 and 2. Both perception and production data were gathered at the same time, with production tests being administered first.

Pretest

Stimuli recording. The recording of production was carried out individually in a quiet room. A unidirectional head-mounted microphone (Shure SM 10A) and a SONY TCD D8-DAT recorder were used to capture production. Then, data was transformed into WAV format using a Kay Lab CSL 4400 machine and stored in an IBM computer. The recording was redigitized at a sampling rate of 22.05 kHz and 16-bit quantization. All tokens were normalized for intensity with UAB software. Peak amplitude was normalized with 50 percent of the scale.

Stimuli elicitation. Two types of production elicitation were used. One consisted of reading a wordlist, and the other consisted of the repetition of audio stimuli and naming words. Since the wordlist method could not elicit as many errors as the investigator had expected, a second type of elicitation was introduced. It seems that the participants became more spontaneous in the naming task. However, because the naming task could not elicit words with an alveolopalatal+i response (either because the Korean

translation sounded very odd or Korean equivalence consisted of very rarely used words), repetition of audio stimuli was added. Thirteen Korean participants (7 in KE and 6 in KC) and all EC used the wordlist method, and fourteen participants (8 in KE and 6 in KC) used the second method of elicitation.

Reading the wordlist (*wordlist* group). The wordlist included real English words ending with obstruents (i.e., alveolopalatal affricates/fricatives and stops) and with an alveolopalatal+i. Sixty-four words were produced in each test. Among these, 37 words ended with alveolopalatal fricatives or affricates ($\int=13$, $t\int=12$, $d\int=12$); 12 words ended with an alveolopalatal+i (3 for each alveolopalatal segment); and 20 words ended with stops. Stops in word final position were different in terms of the length of the preceding vowels and/or their voicing status. However, only 6 words from each NK participant were used in the further judgment: It seems the NK hardly inserted any vowels after stops in the data collection time. Those selected for the judgment were *beat*, *bead*, *pick*, *pig*, *rib* and *rip*. They were all minimal pairs and had either /i/ or an /I/ as the main vowel. All words used in the production elicitation were not used in the perception training. Sixteen words that were used in perception subtest 2 were included in the production stimuli. The list of words is attached in Appendix E. As for EC, they produced all words in the wordlist. However, because there were so many tokens for judgment, only a fraction of the tokens from EC were actually used in the judgment. Six tokens (*fish*, *fishy*, *such*, *itchy*, *page*, *judge*) from each of 5 of EC (participants 1, 2, 3, 5 and 10) were added to each Korean participant's tokens in the *Wordlist* group's judgment. Sex of EC matched to that of NK to make tokens sound as similar as possible.

Delayed repetition and naming (*naming group*). The second type of production tests consisted of the repetition of audio stimuli and naming words in English. Audio stimuli for repetition were recorded by one male native speaker of English who did not participate in any of the previous tests and the training. The stimuli were 9 words ending with an alveopalatal+i, preceded by 4 distractor words. For naming-words elicitation, English words that ended with either an alveopalatal or a stop, and that were easily translated into Korean words, were included. All words were generated from the previous wordlist (Appendix E), but several words were omitted, because some were hard to translate, and others were too rare to be named correctly. A total of 44 words were used in the repetition and naming-words task: Twenty-nine words ended with an alveopalatal, 9 words ended with an alveopalatal+i, and 6 words ended with a stop (*beat, bead, food, foot, pick, pig*). *Rip* and *rib* were not used in this group, because these words were not frequently used. Among 29 words ending with an alveopalatal, 7 final /ʃ/, eleven final /tʃ/, and ten final /dʒ/ words were used. The wordlist for this group is attached in Appendix F. In addition, 6 tokens (*ash, ashy, edge, judgy, watch, peachy*) from each of another 5 of EC (participants 4, 6, 7, 8, and 9) were inserted into each Korean participant's tokens in this group for judgment.

Participants. The same NK participants who took the perception tests took production tests. EC only took the pretest. For judgment, 6 native speakers of English who did not have any phonetic training participated as judges.

Procedure--wordlist group. The participants were asked to read a wordlist at a comfortable rate (i.e., a normal speaking rate) once. The wordlist was given before the reading so that the participants became familiar with the words and had a chance to ask if

they did not know the words. All words were produced in a carrier sentence, "Say ____ again." The order of the words was randomized for the judgment later. At the time of production, the investigator was present in the room to make sure that the participants produced as they were directed and to answer questions.

Procedure--naming group. First, the participants were asked to repeat what they heard after the native speaker's utterances from the computer. The audio stimuli were given either through audio speakers or headsets. In the recording, the native speaker said "I will repeat ____ to him." Right after that, the speaker followed with "What did I just say?". Then, the Korean participants repeated the first utterance. Four to five seconds were given for the repetition. After the repetition task, the participants were asked to name words in English, which were written in Korean orthography on note cards (i.e., translation). In case they could not think of words or produced inadequate words, English spelling was given to elicit appropriate tokens.

Posttest 1

Stimuli. The stimuli were the same as the pretest for both the *Wordlist* and the *Naming* group.

Participants. The NK participants were the same as those for the pretest. EC did not participate.

Procedure. The NK participants came back for the posttest 1 approximately one month after the pretest. All procedures were the same as that of the pretest.

Posttest 2

Stimuli. The stimuli were the same as in posttest 1 for both the *Wordlist* and the *Naming* group.

Participants. The NK participants were the same as those for the posttest 1. One participant in KE did not participate due to personal reasons. The data from one participant in KC were not included due to technical difficulties. EC did not participate.

Procedure. The NK participants came back for posttest 2 approximately 3 month after the posttest 1. All procedures were the same as that of the posttest 1.

Judgment of production

Stimuli. Stimuli for the judgment were from production of KE, KC and EC. All 3 tests of KE and KC and pretest of EC were prepared for the judgment. All stimuli of each participant were randomized.

Participants. Six native speakers of English participated as a panel of judges (EJ). Through a short interview, the judges were found not to have any sustained experience with Korean speakers (e.g., having a Korean roommate, or teaching experience in Korea) or fluent in any other languages. Their age ranged from 18 to 30 (Mean = 21.33 years). They did not have any hearing problems and did not have any noticeable regional accents. Most of them were born in Florida. One male and five females participated. For their contribution, 4 of the judges were paid, and two of them received extra credit for their classes. They were recruited from introductory linguistics classes at the University of Florida (LIN 2000, LIN 3010), but none were majoring in linguistics.

Procedure. EJ evaluated the production of KE, KC and EC. EJ were provided a headset in a quiet room. All tokens were presented on the computer screen using the UAB software.

All judges listened to all tokens of each of 6 to 7 NK participants each day for 4 consecutive days. EJ was given a paper-and-pencil based judgment task, so that they did not need to look at a computer screen and push buttons. The tokens were automatically

presented at approximately two-second intervals. The task was a forced-choice identification test: The judges were asked to choose between two minimally different words ending with a consonant or with a consonant+vowel, which were written in standard English orthography. The judges were encouraged to guess if they were uncertain about a stimulus, and to focus on the word final segment. If they heard the deviant form of a word ending with a consonant, the answer choice was “consonantE,” E being any kind of vowel. For example, if *bead* was heard, the answer choices were *bead* and *beadE*. If there was a real word equivalence ending with a consonant+vowel, the word was given (e.g., *fish/fishy*). If a token was entirely deviant from the answer choices, the judges were encouraged to write down possible spellings in a blank next to the answer choices.

Acoustical Analysis

The initial analysis of the judgment revealed that final /dʒ/ induced the most errors among 3 alveolopalatals. Hence, words ending with /dʒ/ for each participant were chosen for examination, first, if there was any vowel formant at the end of each word. This determines if there was epenthesis, and how strongly native speakers’ judgment and acoustic analysis were correlated. Second, the duration of final /dʒ/ and preceding vowels were measured to examine if there was any significant differences among words which were identified as *correct* and *incorrect*. Third, from 5 of the EC participants, 3 females and two males, words ending with /dʒ/ were chosen to examine if there was any significant difference in terms of duration between NK and EC. Praat version 4.1.22 was used to carry out acoustical measurement.

Definition of duration. /dʒ/ duration is defined as the duration from the end of a preceding vowel to the beginning of a following vowel. If there is no following vowel, the end of /dʒ/ is the endpoint of the frication of /dʒ/. The start of a preceding vowel is defined as the beginning of the F2, and the end of a vowel is defined as the beginning of the voicing of /dʒ/.

Data Analysis

For production data analysis, two kinds of analyses were carried out depending on *Group* (KE vs. KC vs. EC), and *Time* (pretest, posttest 1, posttest 2). Since two kinds of elicitation methods were used, and some tokens were missing due to technical failure, mean percentage of correctness was used rather than raw scores.

KE vs. KC in Three Tests

In order to examine whether KE did better than KC in their production, the mean percentage of correctly identified words was compared using a mixed design 2×3 ANOVA with *Group* (KE, KC) as a within subject variable and *Time* as a between subject variable (Pretest, posttest1 and posttest 2).

Individual Segments

For NK, in order to examine which segment was the most difficult to produce, a repeated measures ANOVA was conducted. The mean percentage of correctly identified /ʃ, tʃ, dʒ/ across 3 tests was compared. In addition, the mean percentage of correctly identified words ending with a /dʒ/, based on the initial analysis, was compared (pretest vs. posttest 1, pretest vs. posttest 2) using a paired-samples t-test. Finally, since many participants seemed to make mistakes in words ending with an alveolopalatal+i, the mean

percentage of correctly identified words ending with an alveolopalatal+i was also compared for pretest vs. posttest 1 and pretest vs. posttest 2.

Acoustical Analysis of /dʒ/

First, to examine how much discrepancy there was between the evidence in the spectrogram and EJ's judgment, the frequency was calculated. Second, using a paired samples t-test, the mismatched words were compared with matched words in terms of duration: a word which did not have any vowel trace but was judged as having a final /dʒi/ was compared with another version of the word which was judged correctly. The same procedure was used for words which had vowel trace but judges did not catch them. Third, using an independent samples t-test, correctly identified words were compared with those of EC in terms of duration of final /dʒ/ and its preceding vowel.

Correlation between Perception and Production

Using Pearson's' bivariate correlation analysis, the mean percentage of correctly identified tokens in perception and production was compared 3 different times. Raw scores in perception tests were converted into percentages for the analysis. First analysis was conducted with subset 1, 3 and 4 of perception test, and second analysis was conducted with 4 sets plus subset 5. This time, only posttest 1 and 2 were analyzed. In addition, participants who received lower than 80 percent of accuracy in production pretest were selected, and their perception and production were analyzed 3 different times.

CHAPTER 6 RESULTS: PERCEPTION TESTS

This chapter includes the results of the perception experiments for the pretest, posttest 1 and posttest 2, and reports the effect of training on the perception of English palatal codas. The analyses included comparison between KE, KC and EC, as well as KE and KC before and after the training in main subtests and generalization tests. In addition, words ending with an alveolopalatal and an alveolopalatal+i were compared to examine the possible source of extra vowel production after final alveolopalatals.

Scores in Pretest

Table 6-1 shows the score comparison among KE, KC and EC in the pretest before the training. Raw scores from each test were used in the perception test analyses. First, the total scores, with a maximum of 124, across 3 subsets (subsets 1, 2, and 3, which all included words ending with an alveolopalatal and an alveolopalatal+i) of the pretest among the English Control group (EC), Korean Experimental group (KE) and Korean Control group (KC), were analyzed to examine general perception ability for distinguishing words ending with an alveolopalatal and with an alveolopalatal+i. The mean scores in each group were 75.53, 75.17, and 119 in KE, KC and EC, respectively. EC made very few mistakes, and KE and KC performed almost identically in the pretest. A one-way ANOVA shows that there was a significant main difference among the 3 groups ($[F(2, 35)=26.57, p=0.00]$). A post-hoc Tukey Honestly Significant Difference test revealed that EC did significantly better than KE and KC ($p=0.000$). There was no significant difference between the two Korean groups in the pretest.

Table 6-1. Score comparison in each subset among KE, KC and EC

	subset 1 (N=36)	subset 2 (N=46)	subset 3 (N=42)	Total (N=124)
KE: Mean (SD)	22.60 (5.1)	29.20 (8.76)	25.80 (6.34)	75.53 (20.25)
KC: Mean (SD)	23.00 (6.02)	30.42 (8.89)	25.58 (6.17)	75.17 (19.1)
EC: Mean (SD)	34.09 (1.45)	45.55 (0.69)	39.36 (8.16)	119 (2.14)

Second, each subset was analyzed by a series of one-way ANOVA. In subsets 1, 2, and 3, the main effect of *Group* was significant ($[F(2, 33)=22.798, p=0.000]$, $[F(2, 35)=25.429, p=0.00]$, $[F(2, 35)=17.627, p=0.000]$, respectively). In subset 1, data from two of the KC were missing due to technical difficulties. The Tukey post-hoc test revealed that in all 3 tests, EC did better than KE and KC. KE and KC did not have any significant difference in performance.

Group and Time Comparison in Korean Experimental Group (KE) and Korean Control Group (KC)

In order to examine the effect of training on KE, both *Group* and *Time* were considered in this analysis. Table 6-2 shows the total score difference across subsets 1, 2, and 3 between KE and KC before and after the training. Before the training, KE and KC performed almost the same, but after the training, KE scored higher (Mean: 110.93, 84.5, respectively), and the same trend continued in posttest 2 (Mean: 110.36, 87.67, respectively). In addition, KE seemed to improve after the training. The following statistical analyses supported the assumptions. First, a mixed design 2×3 ANOVA was used with *Group* as between-group variables, and with *Time* (pretest, posttest 1, posttest 2) as within-group variables. The results showed that the main effect of *Time* and *Group*, and *Time/Group* interaction were all significant ($[F(2, 50)=17.095, p=0.000]$; $[F(1, 25)=4.704, p=0.04]$; $[F(2, 50)=4.85, p=0.000]$, respectively).

Second, a simple independent t-test, with *Group* as a variable, showed that, in posttest 1 and posttest 2, there was a significant difference between KE and KC. In both cases, KE did better than KC.

Table 6-2. Total score comparison between KE and KC

	KE: Mean (SD)	KC: Mean (SD)
Pretest	75.53 (20.25)	75.17 (19.1)
Posttest 1*	110.93 (9.55)	84.5 (20.9)
Posttest 2**	110.36 (10.25)	87.67 (19.46)

*[$t(25)=4.376, p=0.000$], ** [$t(24)=3.8, p=0.001$]

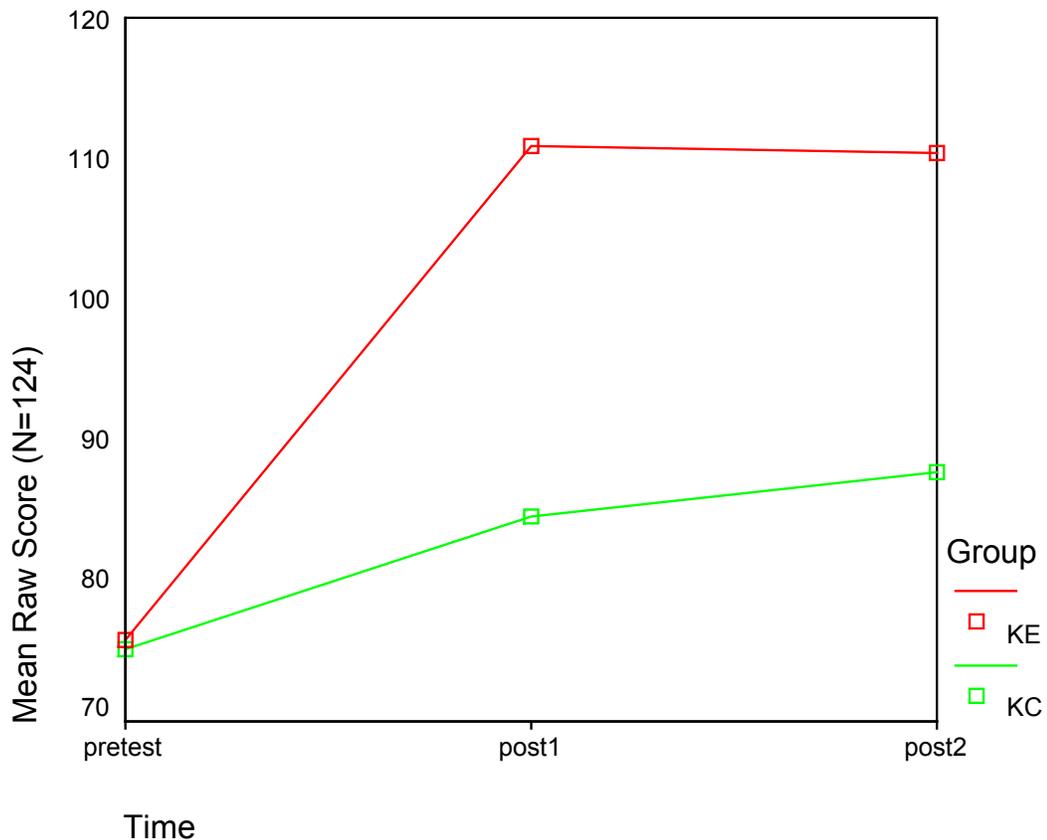


Figure 6-1. Score comparison between KE and KC at 3 testings

Third, a repeated measures ANOVA, with *Time* as a variable, was conducted for KE and showed a significant main effect ($[F(2, 28)=13.66, p=0.000]$). Pairwise comparison showed that the pretest and posttest 1, and the pretest and posttest 2, were significantly different ($p=0.000, p=0.007$, respectively). KE performed significantly

better in posttest 1 and posttest 2 than in the pretest. There was no significant difference in performance between posttest 1 and posttest 2.

Fourth, the same analysis was conducted for KC. A repeated measures ANOVA for KC showed significant main effect ($[F(2, 22)= 8.16, p=0.002]$). Interestingly, pairwise comparison showed that the pretest vs. posttest 1, the pretest vs. posttest 2, and posttest 1 vs. posttest 2 all showed a significant difference ($p=0.036, p=0.008, p=0.01$, respectively). KC performed best in posttest 2 and worst in the pretest.

Fifth, using an one-way ANOVA, when the total scores of posttest 1 (highest scores among 3 tests) from KE and KC were compared with EC, there was still a significant difference among the groups ($[F(2, 35)=22.066, p=0.000]$). However, the post-hoc Tukey test revealed that there was no significant difference between KE and EC, but there was a significant difference between KC and EC ($p=0.000$), and KE and KC ($p=0.000$). Similar results were obtained when subset 5, which included new alveolopalatal tokens, was included in the total score calculation ($[F(2, 35)=25.819, p=0.000]$): KE and EC did not show any significant difference, but KE vs. KC and EC vs. KC had significant differences ($p=0.000$ for both comparisons).

In conclusion, first, KE and KC perceived words with an alveolopalatal and an alveolopalatal+i worse than EC in the pretest. First language background should not be the cause of the differences, because meaning association to stimuli was not required. However, after the training, KE did not seem to have any differences from EC. The training seemed to help the participants reach native-like proficiency in perceiving English final alveolopalatals. KE performed better in both posttest 1 and posttest 2 than KC. In addition, KE improved from the pretest to posttest 1 and posttest 2. There was no

significant decrease in the ability to distinguish words ending with an alveolopalatal and with an alveolopalatal+i even 3 months after the training. Interestingly, KC also improved after the pretest, although their proficiency was poorer than KE at posttest 1 and posttest 2.

Generalization Tests, I and II (Subsets 4 and 5)

Table 6-3 shows the results of subsets 4 and 5. Subset 4 was the generalization test I for word final stops, which were not part of the training. The total number of items was 42. The mean scores for each group were 32.87 for KE, 33.42 for KC and 37.82 for EC, which seem to be very closely grouped. To investigate further, a series of statistical analyses were conducted. First, a one way ANOVA was conducted to examine group differences in the pretest. The main effect was significant ($[F(2,35)=3.317, p=0.048]$), although pairwise comparison showed no significance among the 3 groups. Only the comparison between EC and KC showed a near-significant difference ($p=0.052$). Second, to explore the effect of training on KE, a 2 (KE, KC) \times 3 (pretest, posttest 1, posttest 2) mixed design ANOVA analysis was conducted and showed that there was no main effect of *Group*, *Time* or their interaction.

Subset 5 was the generalization test section II for new words and new talkers with a final alveolopalatal and an alveolopalatal+i, which was added to posttest 1 and posttest 2. First, when the scores from EC in the pretest (since they took it only once) were compared with those of KE and KC in posttest 1, it seems as though KE and KC still did not achieve native-like proficiency (Mean: EC-71.18, KE-61.43, KC-44.67, respectively). Among the 3 groups, an one-way ANOVA showed that there was a significant effect of *Group* ($[F(2, 35)=28.361, p=0.000]$). Pairwise comparison showed that all 3 groups were significantly different (EC vs. KE $p=0.027$, EC vs. KC $p=0.000$, KE vs. KC $p=0.000$):

EC performed best and KC performed worst. Second, a 2 (KE, KC) \times 2 (posttest 1 and posttest 2) mixed design ANOVA showed that there was a significant main effect of *Group* ($[F(1,24) = 13.284, p = 0.001]$) and interaction between *Group* and *Time* ($[F(1, 24) = 9.032, p = 0.006]$). Second, independent t-tests, with *Group* as a within subject variable, revealed that KE did significantly better than KC in both posttest 1 and posttest 2 ($[t(26) = 4.408, p = 0.00]$, $[t(25) = 2.786, p = 0.01]$, respectively). Third, when a simple paired samples t-test, with *Time* as a variable, was conducted to examine a difference between posttest 1 and 2, there was no significant difference between posttest 1 and posttest 2 in KE, but there was in KC (KC mean in posttests 1, 2: 44.67, 49.08, respectively; $[t(10) = 2.573, p = 0.026]$). It seems that the significant interaction in a mixed design ANOVA analysis was produced due to the residue of significant difference in the *Time* variable in KC.

Table 6-3. Generalization tests I and II

	Pre:Mean (SD)	Post 1:Mean (SD)	Post 2:Mean (SD)
sub 4: KE	32.87 (5.53)	34.8 (6.06)	36.57 (3.88)
KC (N=42)	33.42 (4.72)	35.42 (5.05)	36 (3.71)
EC	37.82 (5.02)		
sub 5:KE		61.43 (7.28)	59.43 (7.86)
KC		44.67 (12.84)	49.08 (11.02)
EC (N=72)	71.18 (1.17)		

It is understandable that no significant difference was found in subset 2, since the participants almost reached the ceiling in the pretest. Koreans perceived the difference in stops with different releasing status, non-release, release and release with [ɰ]: They perceived non-release and release as the same, and the difference between those and release with [ɰ]. In addition, they were successful in discriminating words without any stops (CV) and with non-released stops (CVC).

However, in generalization test II, KE performed significantly worse than EC. It seems that the new tokens were a bit more difficult to perceive than old ones. Even so, KE did significantly better than KC in both posttest 1 and posttest 2. In other words, KE were able to use the ability to distinguish words ending with an alveolopalatal and with an alveolopalatal+i when the stimuli were not part of the training or the pretest. The reason that KC performed better in posttest 2 than in posttest 1 was assumed to be due to the test-retest effect. Otherwise, they simply improved the ability to perceive the difference just by being exposed to elements of the target tokens in the target language (TL) environment.

Individual Subsets among KE and KC

To examine the improvement in each subset in each group, paired samples t-tests were conducted, comparing the pretest vs. posttest 1, and the pretest vs. posttest 2. First, Table 6-4 shows the mean scores of each subset in KE. A repeated measures one-way ANOVA, with *Time* as a within subject variable, was carried out for separate subsets 1, 2, and 3. For separate subset scores, the main effect of *Time* was significant ($[F(2,26)=60.705, 20.244, 34.804, \text{ respectively, } p=0.000 \text{ in all 3 subsets}]$). Pairwise comparison revealed that all pretest scores were significantly worse than posttest 1 and posttest 2 scores ($p=0.000$). There was no significant difference in posttest 1 and posttest 2 in each subset score.

On the other hand, when KC were compared with *Time* as a within subject variable, they seemed to improve significantly in posttest 2, in comparison with posttest 1 (Table 6-5). A repeated measures one way ANOVA analysis revealed that in subset 4, KC showed a significant main effect of *Time* ($[F(2,22)=8.163, p=0.002]$). The difference between the pretest and posttest 2 was significant ($p=0.005$), but was not significant

between the pretest and posttest 1. In subset 3, although an ANOVA analysis did not show a significant main effect ($[F(2,22)=2.948, p=0.073]$), KC did better in the posttest 2 ($p=0.049$) in pairwise comparison than in the pretest.

Table 6-4. Mean scores in each subtest in KE

	Pretest	Posttest 1	Posttest 2
Subset 1 (36)	22.71	33.50	33.00
Subset 2 (46)	29.29	43.14	42.64
Subset 3 (42)	25.93	34.21	33.64

Note: number in the parentheses is the number of items in each subtest

In subset 1, *Time* was also shown to be significant ($[F(2, 18)=10.709, p=0.001]$): KC performed better in posttest 1 and posttest 2 than in the pretest ($p=0.034, 0.003$, respectively).

It is of interest that there was improvement after the pretest in KC, too. The reason could be the plain test-retest effect, even though posttest 1 and posttest 2 were carried out one month and 4 months after the pretest, respectively. The other possibility is that KC picked up the differences between words ending with an alveolopalatal and an alveolopalatal+i by being exposed to the target tokens in the TL community.

Table 6-5. Mean scores in each subtest in KC

	Pretest (SD)	Posttest 1 (SD)	Posttest 2 (SD)
Subset 1 (N=36)	23.00 (6.02)	26.40 (5.85)	29.30 (4.9)
Subset 2 (N=46)	30.42 (8.89)	31.75 (9.55)	36.33 (8.8)
Subset 3 (N=42)	25.58 (6.17)	26.25 (6.44)	27.92 (7.19)

In conclusion, after the training, participants in the experimental group improved their perception ability for final alveolopalatals in English and sustained it 3 months after the training. On the other hand, the control group improved more in posttest 2 than in posttest 1, possibly because of the mere exposure to the target tokens.

Words Ending with an Alveolopalatal and an Alveolopalatal+i

In order to find the source of mistakes in the production of final alveolopalatals, only alveolopalatal fricatives and affricates were examined among NK before and after the training. The total scores of words ending with an alveolopalatal and with an alveolopalatal+i across the 3 subsets (1, 2 and 3) were added up and compared across *Time* (pretest 1, posttest 1, posttest 2). Table 6-6 shows the mean scores across the 3 subsets. A paired samples t-test showed that, contrary to the hypothesis, words ending with an alveolopalatal were perceived more correctly than words ending with an alveolopalatal+i in the pretest and posttest 1 (pretest- $[t(26)=2.31, p=0.029]$; posttest 1- $[t(26)=4.16, p=0.000]$). However, in posttest 2 the result was reversed: NK perceived words ending with a vowel more correctly than words ending with a consonant ($[t(25)=2.715, p=0.012]$).

Table 6-6. Comparison between words ending with alveolopalatal (P) vs. alveolopalatal+i (Pi)

	Pretest: Mean (SD)	Posttest 1: Mean (SD)	Posttest 2: Mean (SD)
Pi	35.93 (11.87)	47.41 (11.54)	51.11 (12.61)
P	40.59 (8.99)	51.78 (9.39)	47.11 (14.41)

When KE alone were analyzed for the comparison of total scores of words ending with an alveolopalatal and with an alveolopalatal+i, the results became a bit different. In the pretest, there was no significant difference in the two types of stimuli. However, after the training, KE did perceive words ending with an alveolopalatal better than words ending with an alveolopalatal+i ($[t(14)=2.684, p=0.018]$). In posttest 2, however, the difference disappeared.

The results contradict the hypothesis that the perception of an extra vowel after a final alveolopalatal might be the reason for extra vowel production. NK actually

perceived words ending with an alveolopalatal more correctly than words ending with an alveolopalatal+i. However, the difference reversed about 4 months after the pretest. Two speculations could be proposed. First, words ending with an alveolopalatal+i were very novel to NK, and those words are not frequently used. Because NK simply were more familiar with words ending with an alveolopalatal, they had an easier time in perceiving words with an alveolopalatal. Second, lexical status might have affected the perception, meaning that there was a possibility that NK memorized real words case by case, but not non-words (i.e., modified non-words). Therefore, it was not possible to tap into the real phonetic model for words ending with an alveolopalatal and with an alveolopalatal+i. It seems that real words had been stored individually rather than categorically. Lexical status analysis was carried out subsequently.

Pretest Subsets in KE and KC (Lexical Status)

Only the pretest subsets were analyzed in detail to examine the source of extra vowel production before the training.

Subset 1

Subset 1 included non-words which ended with either a consonant or with a vowel. First, when words ending with an affricate and with an affricate+i were compared, words ending with an affricate+i were perceived more accurately ($[t(26)=4.233, p=0.000]$). Fricative endings ([s, f, θ, ʃ]) and fricative+vowel endings were also compared using a paired samples t-test, but there was no significant difference between the two types. When all words with an alveolopalatal (i.e., affricate and [ʃ]) and with an alveolopalatal+i were compared, vowel endings were easier to perceive than consonant endings ($[t(26)=4.233, p=0.000]$).

Subset 2 and 3

Subset 2 and 3 included words ending with alveolopalatal fricatives and affricates. To examine the effect of an alveolopalatal affricate alone, only words ending with an affricate and an affricate+i were selected and analyzed with a paired-samples t-test. Words ending with an affricate consistently were perceived better than words ending with an [i] (subset 2- $[t(26)=4.569, p=0.000]$, subset 3- $[t(26)=2.695, p=0.012]$). The same result was found when all words with an alveolopalatal (i.e., affricates and [ʃ]) and with an alveolopalatal+i were compared ($[t(26)=2.695, p=0.012]$), unlike the result in subset 1. Since subset 3 consisted of half non-words and half real words, the effect of lexical status was examined: As Figure 6-2 shows, real words were perceived significantly more correctly than non-words ($[t(26)=13.436, p=0.000]$). Further analysis was conducted in the non-word category. The non-words ending with a consonant and with a vowel did not have any significant difference; however, the significance might have not been revealed because of the small number of non-words (Number of tokens=20). Since the same format was used in subset 5, the same analysis was carried out.

Generalization Test II (Subset 5)

Words ending with an affricate and with an affricate+i were analyzed in posttest 1 and posttest 2. In posttest 1, affricate endings were easier to perceive ($[t(26)=2.894, p=0.008]$), and the same thing happened when an affricate and [ʃ] were combined ($[t(26)=3.204, p=0.004]$). In posttest 2, the difference disappeared when an affricate alone was compared, but when an affricate and [ʃ] were combined, the difference continued ($t(25)=2.077, p=0.048$): Alveolopalatal endings were easier to perceive correctly than

alveolopalatal+i endings. Since this subset also contained half non-words and half real words, the difference in lexical status was calculated. Table 6-7 shows the difference in

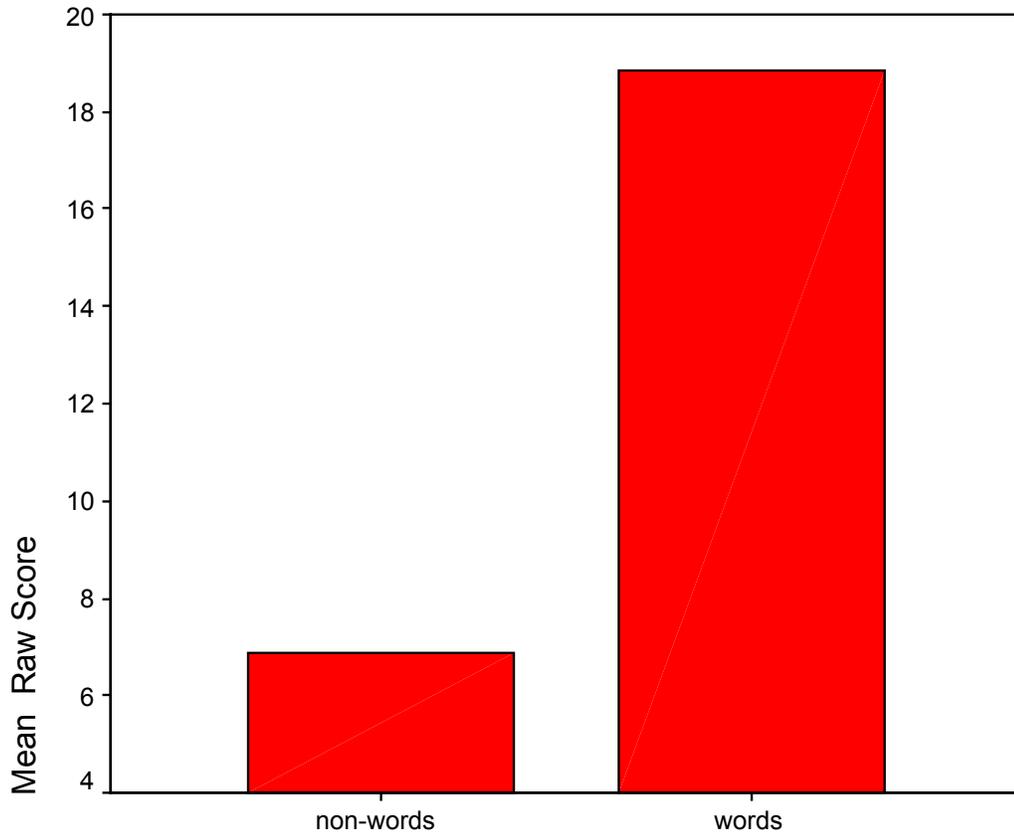


Figure 6-2. Mean score between non-words (N=20) and words (N=20) in the pretest subset 3 for NK

lexical status. A paired samples t-test revealed that real words were perceived significantly more correctly than non-words ($[t(25)=4.689, p=0.000]$). Further analysis was conducted in the non-word category. In this category (Number of tokens=36), alveolopalatal endings and alveolopalatal+i endings were compared, and the result was that non-words with alveolopalatal+i endings were perceived more correctly than words with alveolopalatal endings, as Figure 6-3 shows ($[t(25)=2.598, p=0.015]$).

Table 6-7. Non-words vs. words in generalization test II

Posttest 1 non-	Posttest 1	Posttest 2 non-	Posttest 2
-----------------	------------	-----------------	------------

	words	words	words	words
Mean	23.26	30.67	24.58	30.08
SD	10.75	3.4	7.99	3.53

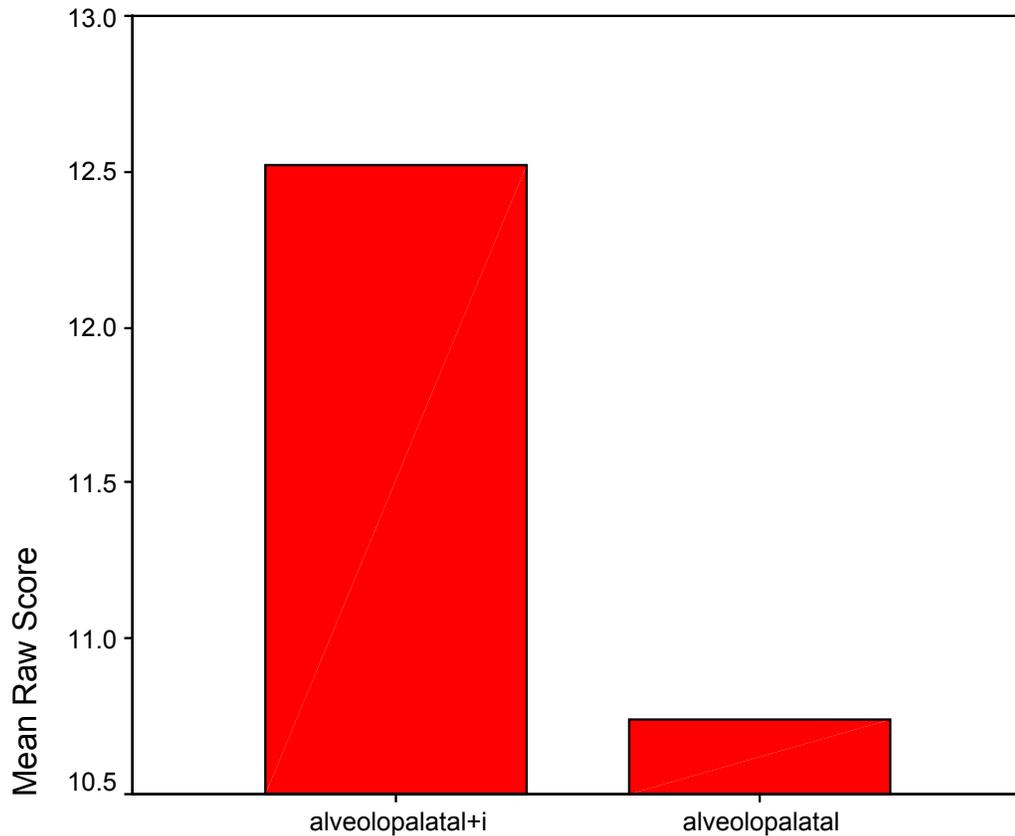


Figure 6-3. Non-words ending in the posttest 1 subset 5 (N=18 for each category)

The same result was produced in posttest 2 (Table 6-7): Real words were perceived more correctly than non-words ($[t(24)=4.474, p=0.000]$). In the non-word category, as Table 6-8 shows, non-words with alveopalatal+i endings were perceived more correctly ($[t(24)=3.819, p=0.001]$).

Table 6-8. Non-words ending with alveopalatal+i (Pi) and alveopalatal (P) in subset 5

	Posttest 1 Pi	Posttest 1 P	Posttest 2 Pi	Posttest 2 P
Mean	12.52	10.74	13.77	10.81
SD	6.72	4.36	5.4	3.25

In conclusion, when words ending with an alveopalatal and with an alveopalatal+i were compared, words ending with an alveopalatal were perceived

significantly better, but when lexical status was added in the analysis, the result changed. In a non-word category, NK consistently perceived words with a final vowel better than words with a final alveolopalatal. This could mean that NK perceived real words with a final alveolopalatal better, because NK learned those words case by case. When they heard a new word, they tended to perceive the words with a final vowel better than the words with a final consonant. This, in turn, might have affected the production of words with a final alveolopalatal. Finally, it is of interest to obtain the results of lexical status differences: Why did NK perceive words ending with an alveolopalatal more poorly than words ending with an alveolopalatal+i? It could be because of a certain aspect of alveolopalatal sounds that triggered extra vowel perception. The possible source would be sustained frication in word final position, which is not allowed in Korean phonology.

Improvement in Final Alveolopalatal vs. Final Alveolopalatal+i Words in KE

To examine the improvement in KE based on different types of words ending with an alveolopalatal and with a vowel, total scores across 3 subsets (1, 2, 3) were analyzed with regards to the *Time* variable, using a repeated measure one-way ANOVA. Table 6-9 shows the mean scores of each word type at 3 *Times*.

For words ending with an alveolopalatal, the analysis showed a significant main effect of *Time* ($[F(2, 26)=52.456, p=0.000]$). Pairwise comparison showed that the scores in all 3 tests were significantly different. When the pretest and posttest 1 and the pretest and posttest 2 were compared, KE improved the perception of words ending with an alveolopalatal ($p=0.000$ at both comparison). Similar results were found in words ending with vowels ($[F(2,26)=30.437, p=0.000]$): When the pretest and posttest 1, and the pretest and posttest 2, were compared, NK significantly improved words ending with [i] ($p=0.000$ for both comparison).

It seems that after the training, KE improved the perception of both words ending with an alveopalatal and with an alveopalatal+i. It is safe to say that the training helped the participants recognize the difference between those two different types of words. In addition, they sustained the ability for 3 months.

Table 6-9. Comparison between words ending with an alveopalatal (P) and an alveopalatal+i (Pi) in KE

	Pre P	Posttest 1 P	Posttest 2 P	Pre Pi	Posttest 1 Pi	Posttest 2 Pi
Mean	40.14	56.43	53.5	37.79	54.43	55.36
SD	9.26	6.02	6.17	12.52	4.15	6.07

Individual Differences

Although all KE completed nine sessions of training, the effect of training seemed varied. For example, Table 6-10 shows that participant 12 received the worst score among 15 participants in the pretest, but this participant improved the most in posttest 1. A similar result was shown in participants 1, 7, 8, and 15, whose perception abilities improved over 20 percent in the posttest 1 compared to the pretest. However, those who already received higher than 80 percent correct in the pretest did not improve as greatly as the others in posttest 1 (participants 6 and 11). This result implies that the training method helped learners with low proficiency more than learners with high proficiency.

Table 6-10 also shows the wide range of individual variation in the perception tests. To examine if the pretest scores could predict the effectiveness of the training, the perceptual improvement formula, "room for improvement," was calculated: posttest 1 scores minus pretest scores divided by 100 minus pretest scores (p. 2306, Bradlow et al., 1997). When the correlation between this formula and posttest 1 was calculated, a significant negative correlation was found ($r=-0.571$, $p=0.026$). This implies that the

pretest score and the rate of improvement were in the opposite direction: Individuals who started at a low level improved more than those who started at a high level.

Table 6-10. Individual percentage of correctness and difference between the posttests and pretest

	pre	post1	pre-post1	post2	post2-pre
1	50.81	96.77	45.96	96.75	45.94
2	57.26	91.94	34.68	90.24	32.98
3	60.48	94.35	33.87	97.56	37.08
4	62.9	83.87	20.97	83.74	20.84
5	73.39	95.16	21.77	95.12	21.73
6	98.39	98.39	0	99.19	0.8
7	46.77	83.06	36.29	86.99	40.22
8	56.45	88.71	32.26	89.43	32.98
9	58.87	90.32	31.45	—	—
10	40.32	66.94	26.62	65.85	25.53
11	83.87	95.16	11.29	95.12	11.25
12	34.68	92.74	58.06	93.5	58.82
13	64.52	88.71	24.19	86.99	22.47
14	71.77	86.29	14.52	88.62	16.85
15	53.23	89.52	36.29	86.99	33.76
Mean	60.91	89.46	28.55	89.72	22.83

This is confirmed by a correlation analysis of the pretest and a difference of the pretest and posttest 1 accuracy rates. When those two scores were compared with a Pearson's bivariate correlation, a strong negative correlation was found ($r=-0.88$, $p=0.000$). As Figure 6-4 shows, the lower one's score was in the pretest, the more this person gained from the training. However, the significant correlation disappeared in posttest 2.

Training

With respect to the training, as expected, KE reached the highest level of perceptual improvement in week 3, which was the final week. Figure 6-5 shows the mean score comparison for each week. When the correctness among 3 weeks was compared using a repeated measure ANOVA, a significant main effect was found ($[F(2, 28)=9.561$,

$p=0.001$]). There was a significant difference between week 1 and week 2 ($p=0.008$), and week 1 and week 3 ($p=0.000$). Although there was no significant improvement from week 2 to week 3, the raw scores in week 3 were better than those of week 2.

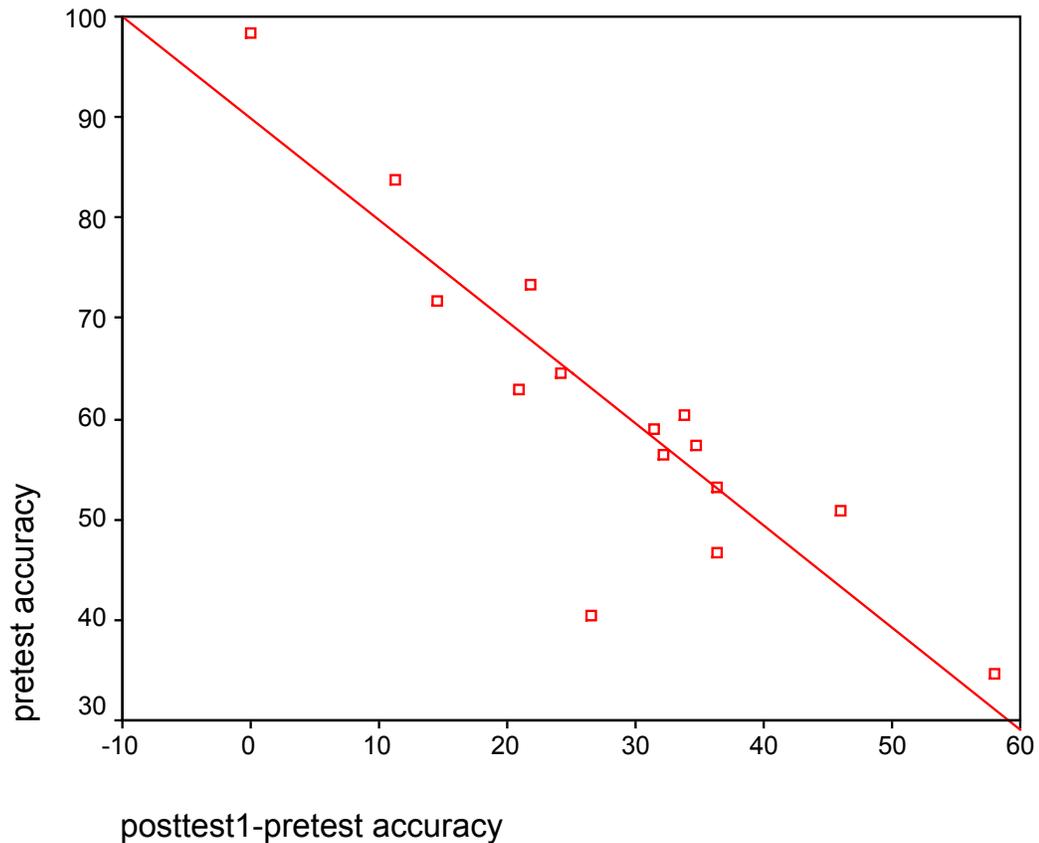


Figure 6-4. Correlation between the pretest and a difference of posttest 1 and pretest for KE ($r=-0.88$, $r^2=0.78$)

Since in this study one talker's tokens were used throughout a whole week, it was expected that the first day of exposure to a new talker would result in a decline of correctness. When the scores between the last day in week 1 and the first day in week 2, and the last day in week 2 and the first day in week 3 were compared with a repeated measures t-test, it was revealed that there was no significant difference. It seems that as the training progressed, the participants became well adjusted to a new talker after listening to the first 3 or 4 tokens.

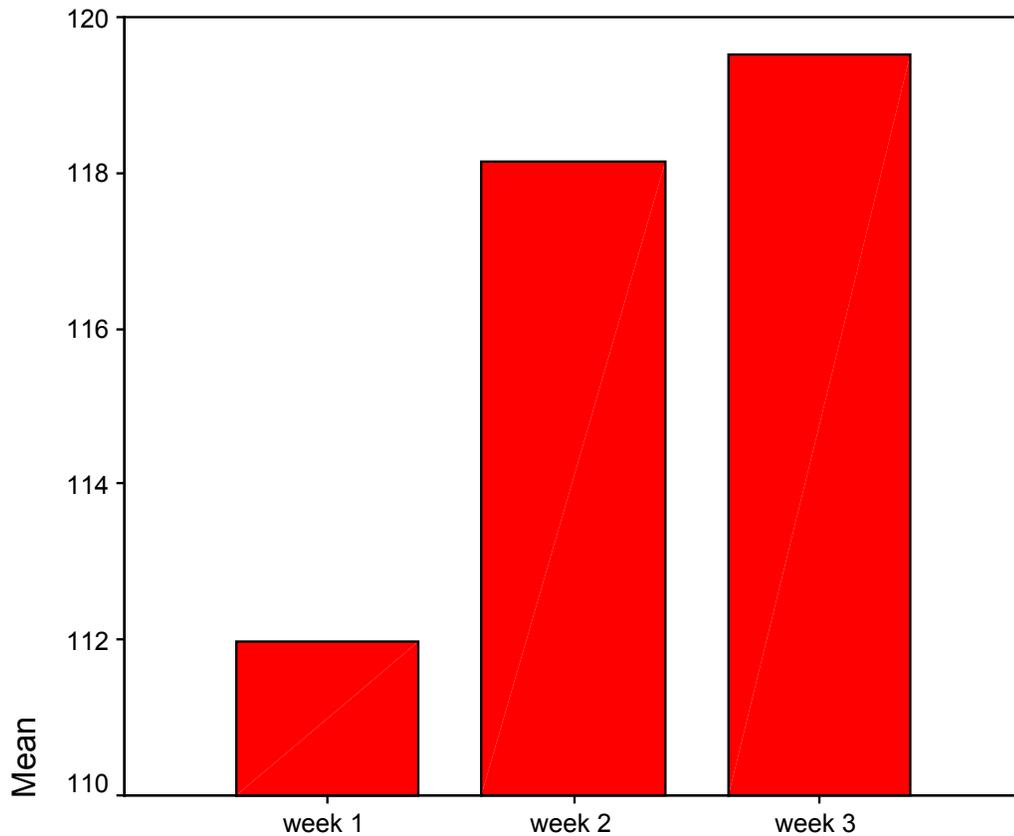


Figure 6-5. The mean accuracy of training in weeks 1, 2, and 3 for KE

Finally, Table 6-11 shows individual scores for each day of the training. It seems that there were participants who did not perform as well as the others. There were 3 participants (Participants 2, 4, 10) who struggled throughout the training. They never achieved scores higher than the group average. When their scores in the perception pretest were examined, two of them (participants 2 and 10) also received lower than the group average. However, all of them improved in posttest 1 and posttest 2, although most of them (except participant 2) received lower than group average scores in posttest 1 and posttest 2. It is suspected that they failed to attain their potential success due to lack of attention or motivation.

Table 6-11. Raw scores for each training session

	Week 1			Week 2			Week 3		
1	116	126	123	123	120	124	123	121	125
2	298	111	104	105	119	112	101	98	117
3	398	112	113	122	122	125	123	126	125
4	101	97	106	98	88	107	114	117	114
5	121	124	126	126	126	126	125	126	126
6	124	124	126	125	123	126	126	125	126
7	88	114	118	121	123	120	118	123	121
8	110	118	122	123	126	124	120	125	124
9	109	124	125	126	126	126	126	126	125
10	74	89	93	75	109	123	79	92	93
11	119	122	123	114	119	117	124	126	126
12	113	121	120	115	118	105	120	122	125
13	112	117	117	126	126	122	122	125	119
14	100	106	112	111	118	115	125	126	117
15	88	114	121	122	126	124	123	122	126

*Total tokens in one training session: 126

Using multiple talkers in the training seems to have been successful in two ways: first, the trainees performed better in posttest 1 and posttest 2 than in the pretest; and second, they perceived new talkers and new stimuli in the generalization test II better than KC. Although our study did not compare single talker and multiple-talker effect, the use of multiple talkers might have helped KE make each of two distinct categories robust.

Summary

In terms of perception, it seems that the training benefited KE. KE improved from the pretest to posttest 1 and sustained the ability to distinguish words ending with an alveolopalatal and an alveolopalatal+i 3 months after the training (posttest 2). In addition, KE and EC were not significantly different in posttest 1, which implies that KE almost attained native-like perception proficiency in distinguishing words ending with an alveolopalatal and with an alveolopalatal+i.

KE also seemed to transfer their ability to new tokens, including new words and new talkers. In generalization test II, KE did better than KC in both posttest 1 and posttest 2. In generalization test I, which included words ending with a stop and a stop+vowel, both KE and KC performed almost as well as EC for all 3 test times.

With respect to the cause of extra vowel perception, it seems that at least in the non-word category, KE made more mistakes in words ending with an alveolopalatal than words ending with a vowel. This indicates that certain acoustic properties of those sounds triggered extra vowel perception. However, in the real word category, NK made more mistakes in words ending with an alveolopalatal+i.

It seems that the training helped the participants who started at a low level of perception proficiency more. The “room for improvement” and the pretest percentage of correctness had a significant negative correlation, which means that participants who performed at a low level in the pretest gained more from the training. The same result was found when the pretest score and a difference of posttest 1 and the pretest were compared.

Finally, the intensive perception training seemed to help the trainees throughout the 3 weeks. Although there was no significant difference between week 2 and week 3, the mean scores of correctness in week 2 and in week 3 were significantly better than those of week 1. In addition, it seems that talker effect was not as strong as was expected. However, it is suspected that using several talkers helped the participants’ phonetic categories of final alveolopalatal and alveolopalatal+i become more robust.

CHAPTER 7 RESULTS: PRODUCTION TESTS

All analyses were carried out after converting raw scores into the mean percentage of correctness judged by the panel of American judges (EJ). Production tests consisted of either wordlist-reading or word-naming tasks to elicit words ending with an alveolopalatal and an alveolopalatal+i from Korean participants, as well as from the English control group. Almost all Korean participants took the pretest, posttest 1 and posttest 2, and their production was judged by EJ. EJ judged correctness of a token using a forced-choice identification test. The percentage of correctness in identification for each test and each group was compared.

Interrater Reliability

The reliability of the 6 judges (EJ) was calculated using a reliability intraclass correlation coefficient analysis. The sum of each participant's score from each judge was tallied and compared. The Cronbach's alpha was 0.9981 ($p=0.000$), which was a highly reliable interrater correlation.

Group and Time Comparison in KE and KC

Total scores from each Korean group were tabulated. The scores from EC were not included in this analysis for two reasons: one, not all tokens produced by EC were included in judgment due to the excessive number of items for judgment, and two, very few mistakes in their tokens were expected. When EJ judged the production of EC, there were only 8 out of 163 tokens judged incorrectly by at least one of the judges. Therefore, the error rate was 4.9 % (6 EC tokens were inserted in tokens of each NK, so that the

total number of EC tokens was $6 \times 27 = 163$). It is safe to say that EC produced English words correctly.

Table 7-1 shows the mean percentage of correctness in identification for KE and KC in the pretest, and this seemed to have improved in posttest 1 in KE (from 77.54 to 81.23 percent for KE; from 74.3 to 79.17 for KC). The mean percentage of both groups seemed to improve in posttest 2, too. When standard distribution was examined, we saw a wide range of variability in all 3 tests. In addition, when a mixed design 2×3 ANOVA, with *Group* (KE-Reading, KE-Naming, KC-Reading, KC-Naming), and *Time* (pretest, posttest 1, posttest 2) as variables, was performed, there was no significant main *Group* or *Time* effect, or interaction effects. This means that there was no statistically significant improvement in the production of English alveolopalatals from the pretest to posttests in either of the groups. Eta squares for each group were 0.2, 0.1 and 0.3 for the pretest, posttest 1 and posttest 2, respectively. This means that the *Group* variable only accounted for 20 percent of variability in production in the pretest, 10 percent in posttest 1 and 30 percent in posttest 2.

Table 7-1. Mean percentage of correctness in the pretest and posttests 1 and 2

Group	Pretest: Mean (SD)	Post1: Mean (SD)	Post2: Mean (SD)
KE-R (N=7)	79.49 (7.56)	82.18 (10.62)	86.07 (9.25)
KC-R (N=6)	80.72 (3.86)	81.34 (7.54)	87.13 (7.01)
KE-N (N=8)	74.53 (11.87)	81.67 (11.8)	79.94 (10.58)
KC-N (N=6)	67.82 (15.06)	73.97 (11.6)	72.94 (7.85)

R: Reading group, N: Naming group

Words ending with C and Ci improvement. Since one of the purposes of the production tests was to examine if Koreans made errors in words ending with a consonant, words ending with an alveolopalatal and with an alveolopalatal+i each were separated and analyzed. Table 7-2 shows the mean percentage of correctness of words

ending with an alveolopalatal. The mean percentage was compared among the pretest, posttest 1 and posttest 2 with a repeated measures one-way ANOVA for each of 4 groups.

The results showed that there were no significant differences in the 3 tests for all groups.

Table 7-2. Mean percentage of correctness of words with final alveolopalatals in pretest and posttests 1 and 2

	Group	Mean	SD
Pretest	KE-R	87.08	15.94
	KC-R	91.37	4.64
	KE-N	82.48	18.68
	KC-N	63.07	20.8
Posttest 1	KE-R	89.26	14.91
	KC-R	93.44	6.37
	KE-N	81.36	13.44
	KC-N	71.84	15.4
Posttest 2	KE-R	94.36	11.7
	KC-R	97.03	1.9
	KE-N	86.17	15.75
	KC-N	69.82	17.37

R: Reading group, N: Naming group

Interestingly, several mistakes were noticed in words ending with an alveolopalatal+i at the time of data collection. The participants did not produce final /i/, even though the researcher modeled some of the words prior to the recording. When those words were compared at the 3 different times in the KE groups, neither *reading* or *naming* groups showed significant differences at the 3 times (possibly because of the small number of participants in each group). However, the raw percentage of correctness improved after the pretest. Participants seemed to recognize the existence of final /i/ after the pretest.

Table 7-3. Mean percentage of correctness in words ending with /i/ for NE

	Pretest	Posttest 1	Posttest 2
KE-R: Mean (SD)	27.4 (22.35)	44.97 (38.04)	50.33 (34.98)
KE-N: Mean (SD)	34.71 (28.44)	55.65 (30.22)	44.28 (43.76)

R: Reading group, N: Naming group

To sum up, both KE and KC failed to show any significant improvement in the production of words ending with an alveolopalatal and with an alveolopalatal+i in posttests 1 and 2. Furthermore, they did not improve in the production of words ending with an alveolopalatal in posttest 1 and posttest 2. In addition, they failed to recognize the existence of words ending with an alveolopalatal+i in the pretest. However, KE noticed the difference between words ending with an alveolopalatal and with an alveolopalatal+i after the pretest, and consequently produced the words ending with an alveolopalatal+i more correctly.

Individual Segments /ʃ, tʃ, dʒ/ in NK (KE and KC Combined)

In order to examine which segment induced more errors in words ending with an alveolopalatal, the words ending with each segment were separated and tabulated. The mean percentage of accuracy for words ending with /ʃ, tʃ, dʒ/ from across 3 *Times* were analyzed separately with a repeated measures ANOVA. However, no significant difference was found: Production of all 3 segments did not seem to be changed from the pretest to posttest 1 and posttest 2.

When the mean percentage of accuracy for each segment across the 3 tests (pretest, posttest 1 and posttest 2 combined) was compared, words with final /dʒ/ seemed to be significantly more difficult for NK to produce correctly than words with final /ʃ/ and with final /tʃ/. Figure 7-1 shows the distribution of the mean percentage of correctness of final /ʃ/, final /tʃ/ and final /dʒ/ (88.21, 89.96, 72.31, respectively). A repeated measures ANOVA of words ending with /ʃ, tʃ, dʒ/ revealed a significant

difference ($[F(2, 52)= 13.952, p=0.000]$). Pairwise comparison showed that final /dʒ/ was harder than final /ʃ/ (Bonferroni adjusted $p=0.000$) and final /tʃ/ (Bonferroni adjusted $p=0.003$). There was no significant difference between final /ʃ/ and final /tʃ/.

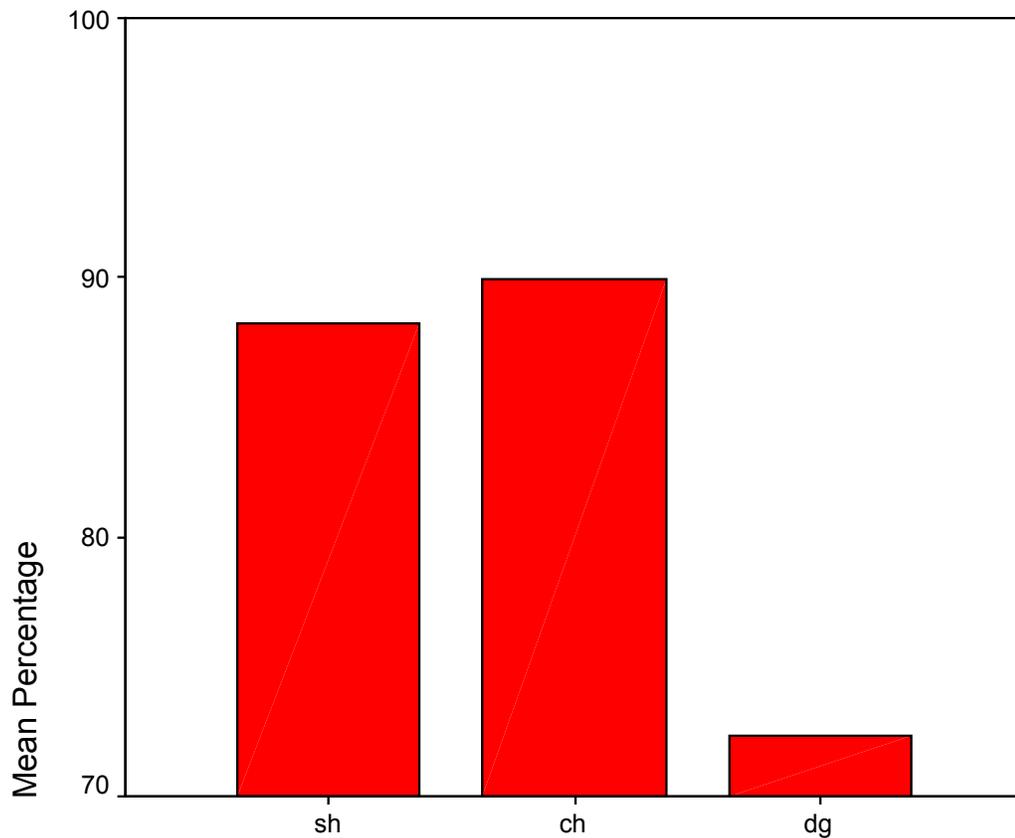


Figure 7-1. Mean percentage of correctness for final /ʃ, tʃ, dʒ/

As for the comparison of mean percentage of correct identification among words ending with /ʃi, tʃi, dʒi/, the total correctness across 3 *Times* was calculated. A repeated measures ANOVA showed that there was a significant main effect of the 3 segments ($[F(2, 52)=13.987, p=0.000]$). Pairwise comparison revealed that NK produced /ʃi/ and

/tʃi/ worse than /dʒi/ (/ʃi/ and /dʒi/: Bonferroni adjusted $p=0.009$; /tʃi/ and /dʒi/:

Bonferroni adjusted $p=0.000$). There was no significant difference between /ʃi/ and /tʃi/.

In summary, NK produced words ending with /dʒ/ less accurately than those with /ʃ/ and /tʃ/. With respect to the correctness of alveolopalatal+i words, voiceless segments induced more errors than voiced segments.

Words ending with a stop. Among each participant's tokens prepared for EJ, only 6 tokens of words ending with a stop were selected for the judgment due to the vast number of tokens: *beat, bead, pick, pig, rib* and *rip* for the *Wordlist* group and *beat, bead, food, foot, pick* and *pig* for the *Naming* group. Among 162 tokens (6 tokens for each of 27 participants) only 4 tokens were judged as mispronounced, and those were produced by 4 different speakers. It is safe to say that most participants produced words ending with a stop correctly. This was not an expected result: Koreans were said to insert a vowel after final stops. However, our result showed otherwise. This might be because the participants in the experiment were not true beginners: It is possible that only true beginners make such mistakes. The other possibilities are either that the tests failed to elicit true production ability of the participants, or they did not have any difficulty in producing those words.

Production Improvement in KE

First, a repeated measures ANOVA analysis (and refer to Table 7-1) showed that there was no significant main effect of *Time* in the mean percentage of production correctness in KE. Pairwise comparison with a Bonferroni adjustment did not show any significant difference among pairs. However, since we are interested in comparing

production before and after the training effect, simple paired samples t-tests were carried out subsequently, comparing the pretest vs. posttest 1 and the pretest vs. posttest 2. The results showed marginally significant improvement in posttest 2 ($t(2)=2.056, p=0.06$).

In addition, it was revealed that when KE and KC were combined and analyzed in terms of 3 different alveolopalatals, they made significantly more mistakes in words ending with /dʒ/ (Figure 7-1). Subsequently, only words ending with /dʒ/ were analyzed with *Time* as a within variable in two KE groups. When a repeated measures one-way ANOVA was performed, neither of the KE groups showed any significant difference. However, the mean percentage of correctness in posttest 2 was the greatest: KE did not seem to improve immediately after the training, but somehow the greatest improvement was shown (Figure 7-2) 3 months after the training. The difference between the pretest and posttest 2 was 7.46 and 11.16 for *Reading* group and *Naming* group, respectively. On the other hand, KC did not show any improvement in the mean percentage of correctness or in final /dʒ/ production.

With respect to words ending with /ʃi/ and /tʃi/, unlike the above results, posttest 1 showed significant improvement. Table 7-4 shows the mean percentage of correctness in words ending with /ʃi/ and /tʃi/. For both segments, scores gathered immediately after the training were the highest. A repeated measures ANOVA analysis for each segment was conducted, but the analyses showed no significant main effect of *Time*.

It seems that NK are more familiar with words ending with /dʒi/, which also implies that it was difficult to include a vowel after voiceless alveolopalatals. The participants were not used to the structure of /i/ following final /ʃ/ and /tʃ/. Specifically, in

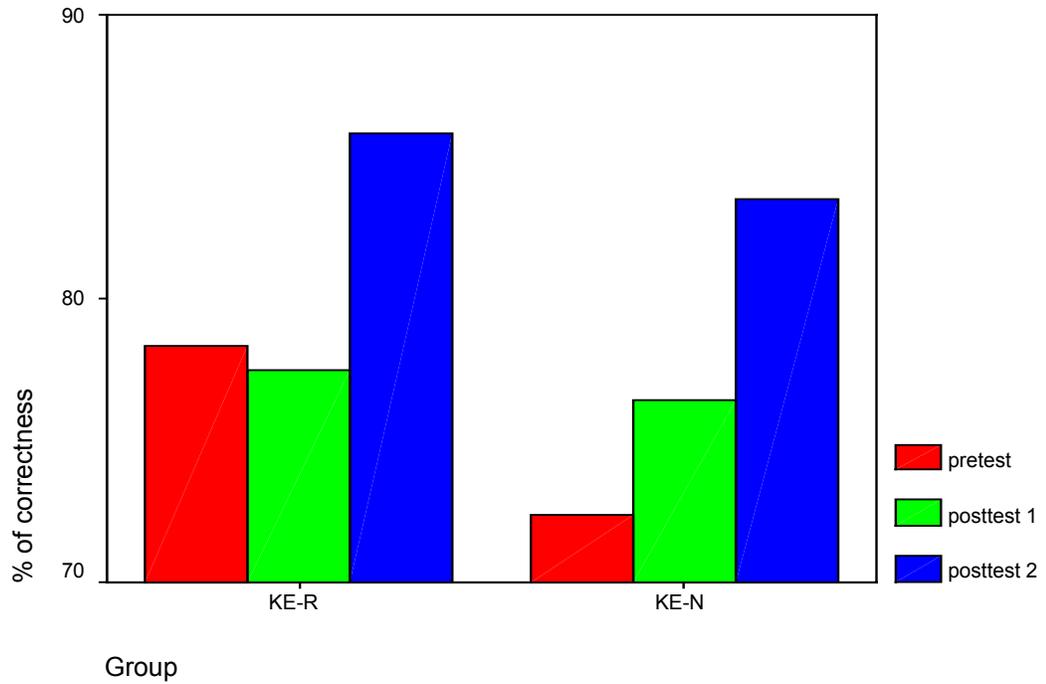


Figure 7-2. Mean % of accuracy for words ending with /dʒ/ in pretest and posttests 1 and 2 in KE

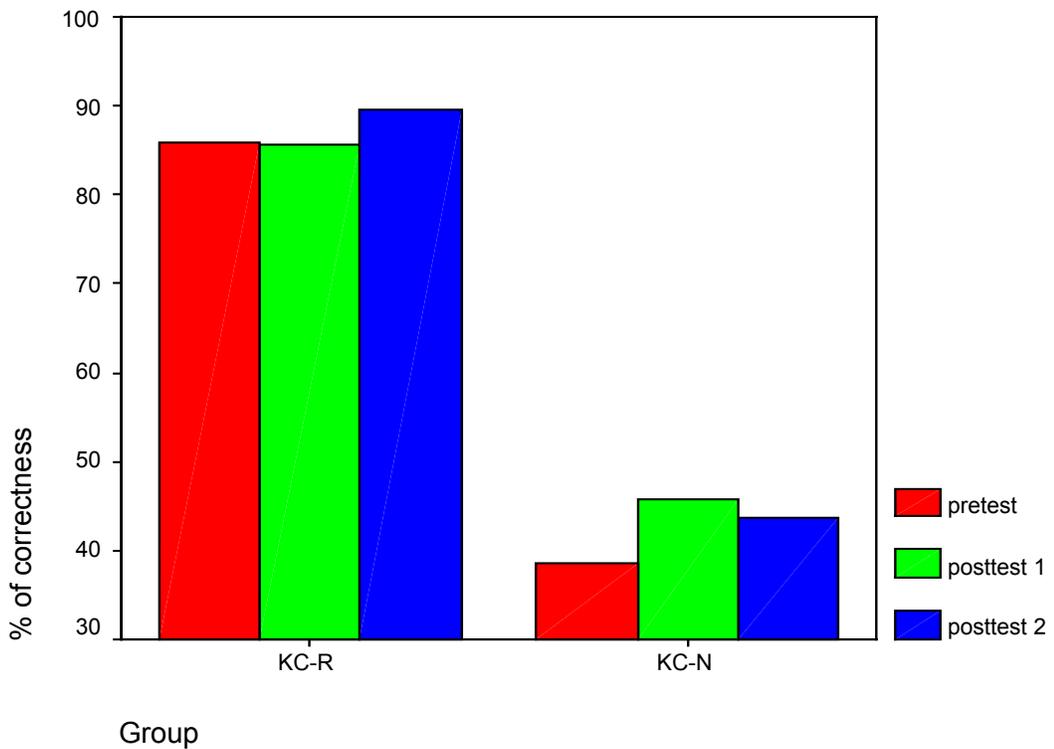


Figure 7-3. Mean % of accuracy for words ending with /dʒ/ in pretest and posttests 1 and 2 in KC

Table 7-4. Mean percentage of correctness in words ending with /ʃi/ and /tʃi/

	Pretestʃi	Post1ʃi	Post2ʃi	Pretest tʃi	Post1 tʃi	Post2 tʃi
KE-R:Mean	38.91 (44.07)	56.35 (36.91)	42.86 (44.06)	15.08 (18.89)	47.57 (42.43)	43.65 (52.74)
KE-N:Mean	35.42 (40.28)	61.9 (46.97)	53.17 (49.47)	20.83 (36.22)	39.28 (30.36)	34.13 (37.74)

R-Reading group, N-Naming group

the *Naming* group, most mistakes were made in the production of the word *catchy*. Out of a total of 35 tokens, only 6 tokens (17%) were correctly produced. When correctly and incorrectly produced *catchy* were compared, there was no significant difference in the preceding vowel and the consonant duration (vowel M=98.28, 109.42 ms.; /tʃ/ M=176.79, 193.63 ms., respectively). Therefore, when NK produced *catchy*, they did not shorten the preceding vowel or /tʃ/ but only inserted a vowel after /tʃ/, which probably happened in the case of other /ʃ/ and /tʃ/ words. In other words, NK might have not confidently produced final /ʃi/ and /tʃi/ words, whereas they did well with /dʒi/ words (See the Acoustical Analysis section). It is suspected the reason why final /dʒi/ was more correctly produced would be the higher frequency of words ending with /dʒi/, such as *technology*, in comparison with words ending with /ʃi/ and /tʃi/.

In summary, KE did not show any improvement in posttest 1, but showed near-significant improvement in posttest 2. In addition, the most troubling segment, final /dʒ/, was produced significantly better in posttest 2 than in the pretest. KE did not show a difference in the production of final /dʒ/ in posttest 1. As for final /ʃi/ and /tʃi/ production, which was harder than /dʒi/, improvement was shown immediately after the training, but

not 3 months after the training. Finally, the reason that final /ʃi/ and /tʃi/ were harder than final /dʒi/ might be its lack of frequency.

Individual differences among KE. Table 7-5 shows the individual percentage of correctness in the production of KE. The wide range of individual differences after the training in KE was also noticed in the production test. For example, participants 1, 6, 7 and 11 were almost at the same level of production accuracy in the pretest, but participants 6 and 11 improved in posttest 1, and participants 1 and 7 deteriorated. A similar pattern was noticed in posttest 2.

Table 7-5. Mean % of production accuracy and the differences of posttests and pretest in production of KE

participants	pre	pt1	pt1-pre	pt2	pt2-pre
1	84.29	76.8	-7.49	81.09	-3.2
2	65.6	67.67	2.07	72.67	7.07
3	72.22	74.67	2.45	97.33	25.11
4	81.41	85.29	3.88	81.37	-0.04
5	83.66	89.46	5.8	90.25	6.59
6	84.29	100	15.71	97.44	13.15
7	84.94	81.37	-3.57	82.37	-2.57
8	81.3	65.15	-16.15	76.74	-4.56
9	67.08	91.47	24.39	—	—
10	78.05	79.84	1.79	88.64	10.59
11	84.88	88.1	3.22	95.45	10.57
12	48.33	70.33	22	65.48	17.15
13	80.95	100	19.05	73.81	-7.14
14	75	85.77	10.77	86.74	11.74
15	80.68	72.73	-7.95	72.73	-7.95
Mean	76.85	81.91	5.06	83	5.47

As with the perception tests, the “room for improvement” was calculated (i.e., posttest-pretest/100-pretest) and compared with the pretest to examine if the pretest could predict the degree of improvement after the training. The results revealed that there was no significant relation ($r=-0.207$, $p=0.458$).

An alternative way of looking at the relation between the pretest and the posttests is to examine the relation between the pretest accuracy and the difference of the posttest 1 and pretest accuracy, and of the posttest 2 and pretest accuracy. The correlation analysis revealed that there was no strong correlation in the analysis between (posttest 1-pretest) and pretest ($r=-0.505$, $p=0.055$) and between (posttest 2-pretest) and pretest ($r=-0.495$, $p=0.072$). However, the correlation between the difference of posttest 1 and pretest and the pretest was suggestive (Figure 7-4), and this relation implies that individuals who performed well in the pretest did not gain as much as those who did poorly.

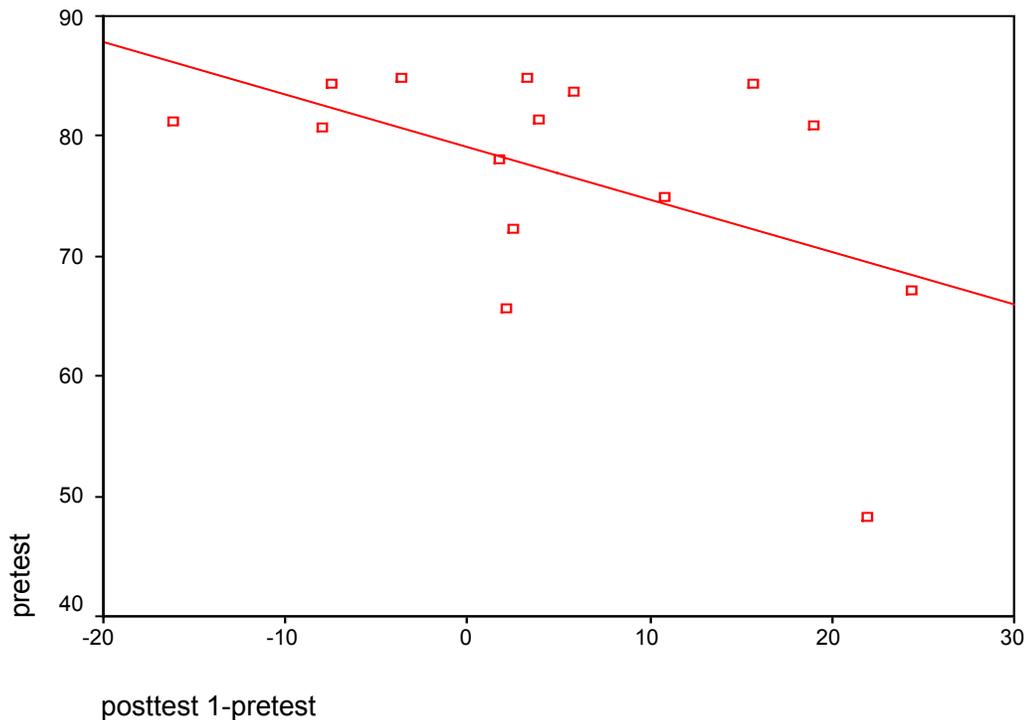


Figure 7-4. Correlation between pretest and the difference of posttest 1 and pretest in KE ($r=-0.505$, $r^2=-0.26$)

Relation between Perception and Production

The purpose of this analysis is to examine the relation between perception and production of words ending with an alveolopalatal and an alveolopalatal+i before and after the training, namely whether they were significantly correlated, and whether the

relationship changed after the training. In order to calculate the correlation between the two domains, the raw scores of the perception tests were converted into percentages.

Total scores across 3 subsets (subsets 1, 2, 3) in the perception tests were converted into mean percentage of correctness.

Correlation in NK (KE and KC combined) and Each Group

Using a Pearson's correlation analysis, with *Time* as an independent variable, the correlation coefficient between perception and production was calculated with NK's (KE and KC combined) mean percentage of correctness. In the pretest, the correlation coefficient reached a significant level ($p=0.017$, $r=0.457$), and in posttest 1 it reached a near-significant level ($p=0.06$, $r=0.366$). However, in posttest 2, the significant correlation disappeared ($r=0.352$). When subtest 5 (generalization test II) in the perception test was included in the analysis, similar results were produced: significant correlation was found in posttest 1 ($r=0.432$, $p=0.025$), but no significant correlation was found in posttest 2. Figure 7-4 and 7-5 showed the distribution of the relation and regression line.

When KE was examined alone, only the pretest showed a noticeable correlation with near-significance ($r=0.477$, $p=0.072$). It seems the linear relationship was disrupted after the training. However, when KC were examined alone, no significant correlation was evidenced in the pretest, posttest 1 and posttest 2.

Individual Differences in Improvement in KE

In order to examine the correlation between perception and production improvement in KE, a correlation analysis in the improvement in posttests 1 and 2 was calculated: The difference of posttest 1 and the pretest accuracy, and the difference of posttest 2 and the pretest accuracy, were compared between perception and production.

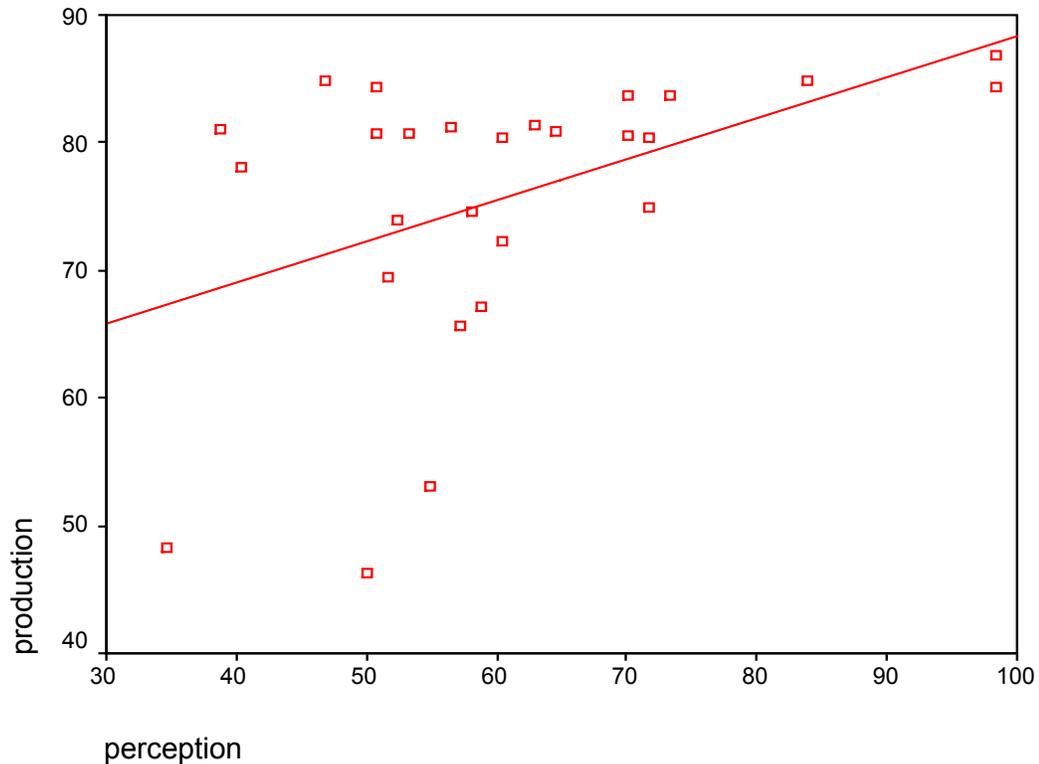


Figure 7-5. Correlation between perception and production in the pretest in NK ($r=0.457$, $r^2=0.21$)

The correlation at both times was revealed not to be strong ($r= -0.149$; $r=-0.057$, respectively). However, it is of interest that the correlation coefficient was negative, rather than positive. This means that some participants who improved in perception actually performed worse in production, which leads to the idea that the degree of improvement could vary as a function of proficiency. Subsequently, the Korean groups were divided according to their performance in the pretest.

Low Proficiency Group

To examine the correlation in more detail, NK were divided into high and low proficiency groups. Based on table 7-6 (for KE) and 7-7 (for KC), which show the individual mean percentage of correctness and group mean in perception and production, the participants were divided. For the high proficiency group, participants who scored

higher than 70 percent in perception (70 % was above the group average) and 80 percent in production (80 % was also above the group average) numbered 5 (KE-3, KC-2) in the pretest. Participants who received higher than 80 in both perception and production in posttest 1 (KE-8, KC-3) and in posttest 2 (KE-8, KC-3) numbered 11. As for the low proficiency group, the participants who scored lower than 70 percent in perception and 80 percent in production numbered 11 (KE-5, KC-6) in the pretest, and participants who received lower than 80 percent in both perception and production in posttest 1 numbered 6. There were also 6 people who were in this group in posttest 2. In particular, 5 in posttest 1 and all 6 in posttest 2 who scored lower than 80 percent were all KC. It is of interest that hardly anyone in KE was part of this group. This might imply that the relationship between perception and production ceased to exist after the training, especially in the low proficiency group. It seems that due to the training either perception or production or both improved more in the low proficiency group.

Group with Low Proficiency in Production

It seems that some participants already reached the ceiling when they started the experiment. The production results did not show any improvement after the training, but the effect could have leveled off due to highly proficient participants. It is possible that the method of training might have only been effective for a special group of people, considering that the number of participants who scored lower than 80 percent in both perception and production declined after the training in KE. Hence, only those who started from a low level in the production pretest were reexamined. The cutoff percentage was 80 in the production pretest: Fifteen participants who received less than 80 percent in correctness (80.68 % to be exact) by EJ were selected and compared. The rationale

behind the cutoff was the number of participants. Almost half of the participants (7 in KE, and 8 in KC) were in this group.

Table 7-6. Mean percentage of correctness in perception and production in KE

No.	Percept:Pre	Posttest 1	Posttest 2	Product:Pre	Posttest 1	Posttest 2
1	50.81	96.77	96.75	84.29	76.8	81.09
2	57.26	91.94	90.24	65.6	67.67	72.67
3	60.48	94.35	97.56	72.22	74.67	97.33
4	62.9	83.87	83.74	81.41	85.29	81.37
5	73.39	95.16	95.12	83.66	89.46	90.25
6	98.39	98.39	99.19	84.29	100	97.44
7	46.77	83.06	86.99	84.94	81.37	82.37
8	56.45	88.71	89.43	81.3	65.15	76.74
9	58.87	90.32	—	67.08	91.47	—
10	40.32	66.94	65.85	78.05	79.84	88.64
11	83.87	95.16	95.12	84.88	88.1	95.45
12	34.68	92.74	93.5	48.33	70.33	65.48
13	64.52	88.71	86.99	80.95	100	73.81
14	71.77	86.29	88.62	75	85.77	86.74
15	53.23	89.52	86.99	80.68	72.73	72.73
M	60.91	89.46	83.74	76.85	81.91	83.01

Correlation between Perception and Production in Low Proficiency KE and KC (Combined and Separate)

In the pretest of the low proficiency combined group, a significant correlation coefficient was produced between perception and production ($r=0.528$, $p=0.043$). Figure 7-6 shows the regression line and the scatterplot of the correlation. It seems that the correlation was stronger than that of entire NK ($r=0.457$). However, no significant correlation was found in posttests 1 and 2 between perception and production in this group. Again, it is speculated that training might have disrupted the relationship: Perception improvement was much greater than production improvement after the pretest. Another speculation would be small number of participants in this group. When KE and KC were separately analyzed in the correlation between perception and

production with *Time* as a variable, none of the *Groups* or *Times* reached significant correlation.

Table 7-7. Mean percentage of correctness in perception and production in KC

No.	Percept:Pre	Posttest 1	Posttest 2	Product:Pre	Posttest 1	Posttest 2
1	58.06	62.9	69.92	74.68	74.51	80.13
2	60.48	93.55	95.93	80.45	84.8	98.67
3	98.39	93.55	94.31	86.86	92.48	92.31
4	38.71	54.84	60.98	81.09	71.53	82.05
5	71.77	91.13	91.06	80.45	81.05	85.29
6	50.81	58.06	60.98	80.77	83.67	84.31
7	52.42	57.26	55.28	73.86	60.98	—
8	51.61	60.48	61.79	69.44	82.93	77.65
9	70.16	74.19	78.86	80.56	63.57	59.17
10	50	48.39	52.85	46.34	66.28	75.61
11	54.84	49.19	56.1	53.03	82.56	74.24
12	70.16	74.19	77.24	83.71	87.5	78.03
M	60.62	68.14	71.28	74.27	77.66	80.68

Production Improvement in Low Proficiency KE and KC (Combined and Separate)

To explore the possibility of improvement over time for low proficiency KE and KC combined group, a repeated measures ANOVA was performed, and showed significant difference ($[F(2, 24)=4.711, p=0.019]$). Pairwise comparison showed that there was no difference in production between the pretest and posttest 1. However, between the pretest and posttest 2 there was significant improvement ($p=0.021$). Furthermore, a repeated measures ANOVA, with *Time* as a within subject variable, revealed that perception ability also improved after the pretest ($[F(2, 26)=18.688, p=0.000]$). Pairwise comparison showed that posttest 2 results were the best, and the pretest was the worst (pretest vs. posttest1: $p=0.002$; pretest vs. posttest 2: $p=0.000$, posttest 1 vs. posttest 2: $p=0.049$).

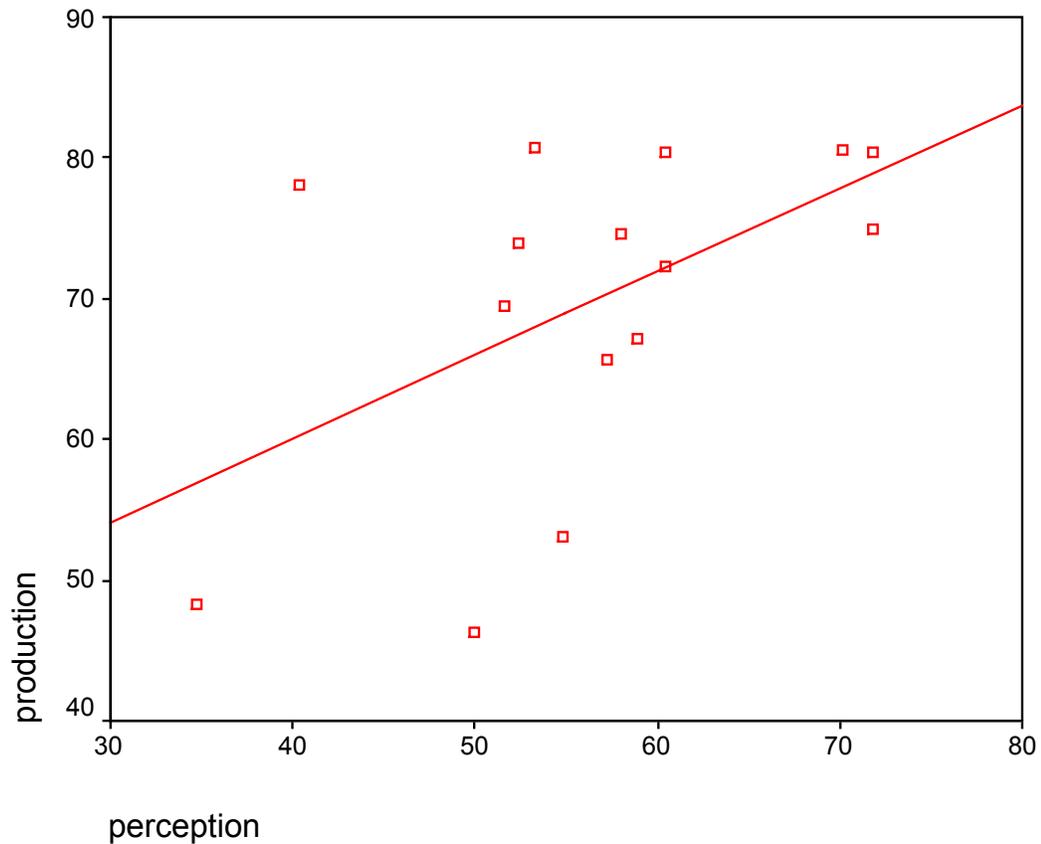


Figure 7-6. Correlation between perception and production in pretest of low proficiency group ($r=0.457$, $r^2=0.028$)

When KE in this group were examined alone ($N=7$, *Reading* 2, *Naming* 5), as shown in Table 7-6, production accuracy improved after the training. The differences of posttest 1 and the pretest were 2.26 % for the *Reading* group and 10.47 % for the *Naming* group. The differences of posttest 2 and the pretest were 16.09 % for the *Reading* group and 8.57 for the *Naming* group. Figure 7-6 also shows the production improvement in posttest 2. Many participants performed better in posttest 2 than in the pretest. A repeated measures ANOVA showed that there was no significant difference between the pretest vs. posttest 1 and posttest 2, but when the two KE groups were combined, there was a marginally significant difference between the pretest vs. posttest 2 ($p=0.065$). However,

there was no improvement in KC after the pretest. Hence, it is possible that the training helped this group more in production. When non-low proficiency members (N=8, *Reading* 5, *Naming* 1) in KE were chosen and compared, the increment after the training was less than that of low proficiency KE. The difference of posttest 1 and the pretest was -0.3 % for the *Reading* group and 9.13 % for the *Naming* group. The difference of posttest 2 and the pretest was 1.56 % for *Reading* group and 1.71 % for *Naming* group. When non-low proficiency KE *Reading* and *Naming* groups were combined and compared with a repeated measures ANOVA, there was no significant improvement in posttest 1 and posttest 2.

With respect to perception, low proficiency KE showed improvement in both posttests 1 and 2. A repeated measures ANOVA showed a significant main effect of *Time* ($[F(2, 12)=29.94, p=0.000]$), and at both times pairwise comparison was significant (pretest vs. posttest 1: $p=0.000$, pretest vs. posttest 2: $p=0.002$). The mean percentage of perception accuracy of KE is shown in Table 7-7. In KC, similar results were found ($[F(2, 14)=5.888, p=0.014]$, and although the pretest vs. posttest 1 was not significantly different, the pretest vs. posttest 2 was (pretest vs. posttest 2: $p=0.023$).

In conclusion, it seems that there was a stronger correlation between perception and production in the low proficiency group in the pretest than in the entire group. However, the relation somehow disappeared after the pretest. Perception training may have influenced the perception part more than the production part. In addition, the training

Table 7-7. Mean percentage of correctness in low-proficiency KE

	Pretest: Mean, (SD)	Posttest 1: Mean, (SD)	Posttest 2: Mean, (SD)
Perception	53.8, (12.6)	87.44, (9.4)	84.55, (12.2)
Production-Reading	68.91, (4.68)	71.17, (4.95)	85, (17.44)
Production-Naming	69.83, (13.06)	80.03, (8.82)	78.40 (11.16)

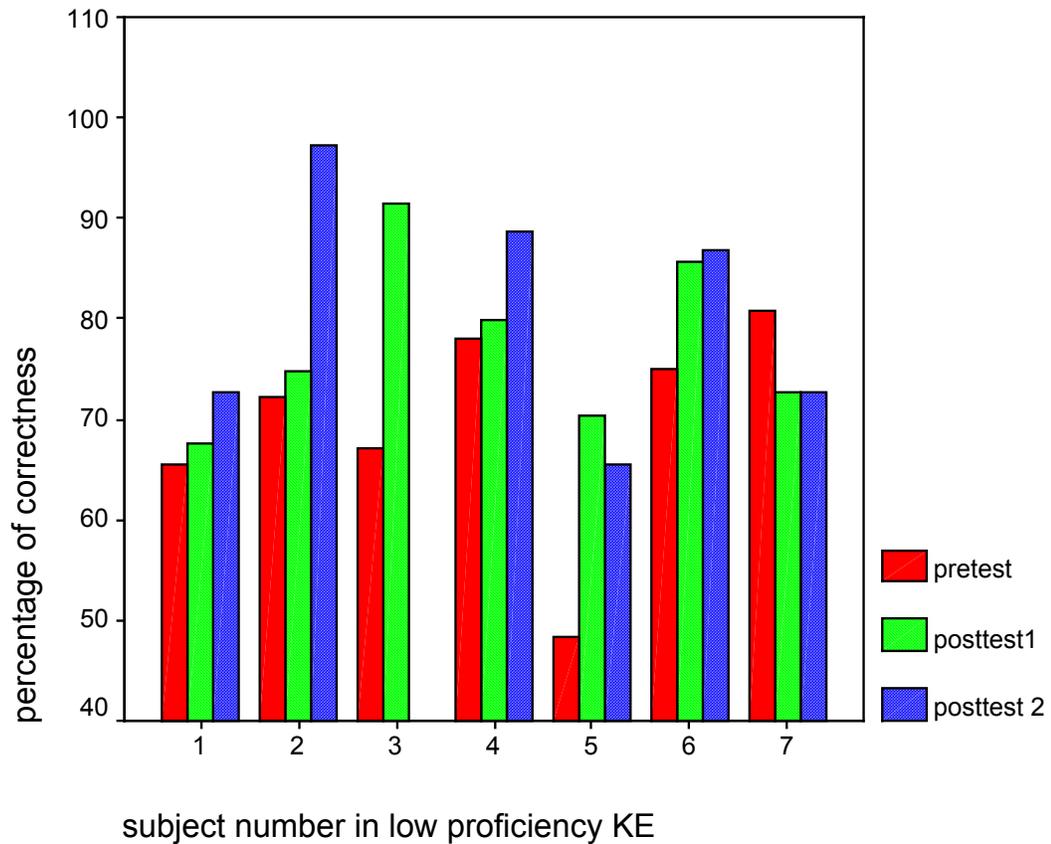


Figure 7-7. Mean % of correctness comparison in the production of the low proficiency participants in KE

seemed more helpful both in perception and production among people who began with a low level of proficiency in production. The participants in this group improved after the training. What is of more interest is that the low proficiency participants in KE actually produced words ending with an alveolopalatal better in posttest 2. Similar production results can be found in the entire KE group in posttest 2. Both results might imply that production takes more time to materialize; in other words, improvement in production lags behind that of perception, at least for those who were trained with perception tasks. This may be because perception training triggers the awareness of correct production of final alveolopalatals.

Individual Factors and Achievement in Perception and Production

Several individual factors can affect improvement: Motivation, learning style, L1/L2 use, L2 proficiency, etc. According to the background information obtained, which was filled out on the day of the pretest, the correlation between individual factors and posttest 1 accuracy in perception and production was calculated. The reason that the scores of posttest 1 were used is that the improvement of perception was the greatest in posttest 1. The factors considered were L2 use, length of residence (LOR), chronological age, TOEFL score, major and gender.

First, L2 use was found to be non-significant. Almost half of the participants indicated that they used English as less than 50 percent of their daily language use. The rest said that they used it less than 25 percent of the time.

Second, with respect to LOR, all participants differed greatly. Some had resided in the U.S. for more than 3 years, and others had LOR less than 3 months at the time of the pretest. No significant correlation was found between LOR and perception or production accuracy in posttest 1. Third, correlation between chronological age and perception/production accuracy produced the same results. Fourth, TOEFL scores and majors (humanity, science, and others) did not show any significant relationship with perception and production accuracy. Fifth, gender and proficiency level did not correlate strongly.

When KE alone were compared, only LOR showed a significant correlation with production accuracy ($r=0.56$, $p=0.03$): People who had lived in the U.S. longer performed better in posttest 1. However, this result should be taken with caution, because only 3 had lived in the U.S. for more than a year (Mean: 9.87 months, SD:11.49). As the scatterplot

in Figure 7-8 shows, the data were clustered around the area of 10 months. None of the factors showed a significant relation with perception accuracy.

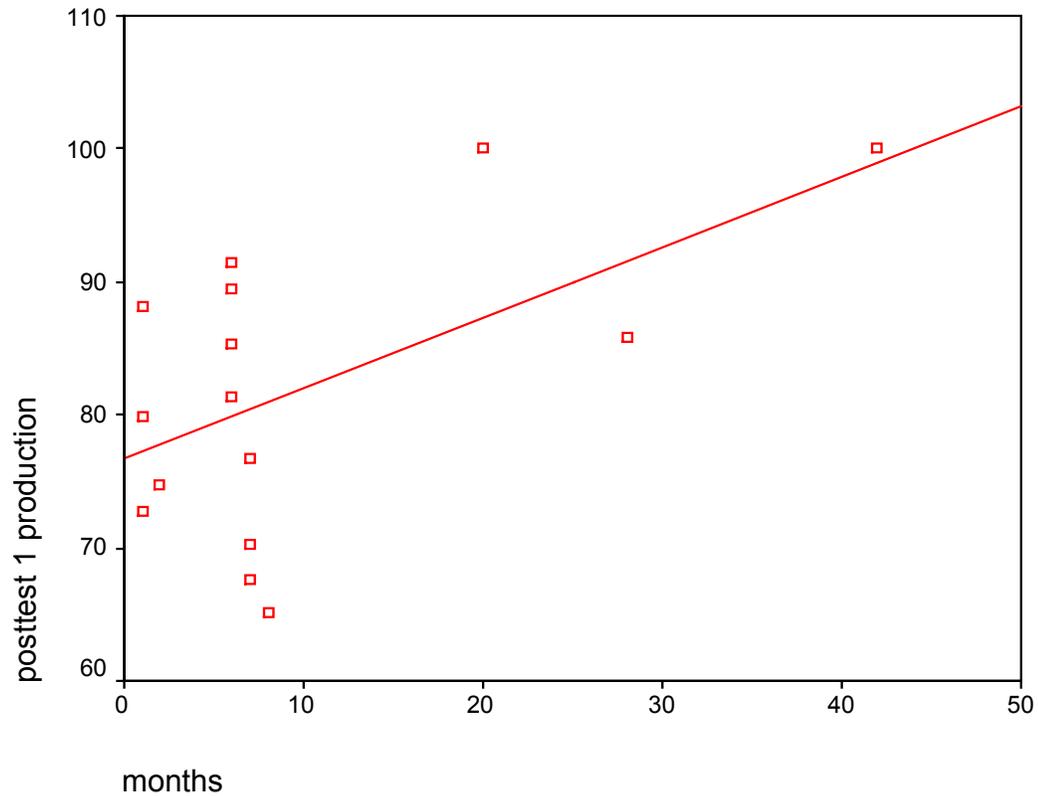


Figure 7-8. Correlation between posttest 1 production and LOR in KE ($r=0.56$, $r^2=0.31$)

Interview on speaking English. On the last day of training, the researcher asked some participants in KE questions regarding the experiment and their attitude towards speaking English in general. In addition, during the training, many expressed their ideas on learning and using English in general. All interviews were recorded upon the participants' agreement.

Several possible factors affecting the success in training were suggested, one of them being motivation. Although most of the participants in KE were motivated enough to participate in such a time-consuming study, some seemed more motivated than others.

For example, participant 3, who improved greatly in both perception and production, expressed a strong desire to learn English. He also made an effort to achieve his goal: He took Academic Spoken English classes, and once or twice a week he went to free English classes. He said that even though using English well was not required to be successful in his field of study, he wanted to speak it well to be more successful in the U.S. He was also a very outgoing person and said that he always tried to find a way to interact with native speakers of English.

A strong desire to learn was also expressed by participants 13 and 14, both of whom were in graduate school and improved greatly in posttest 1. They both believed that speaking English well would make them more successful in their respective careers.

Participant 11 showed improvement consistently throughout posttests 1 and 2 in both perception and production. Even though she improved greatly, she said that she was not 100 percent sure how to distinguish the two types of words. She recognized that it was difficult for Koreans to produce words ending with an alveolopalatal, and that it was difficult for her, too. She also expressed that there was a lack of chances to practice spoken English, even though improving English was very important to her.

On the other hand, participant 2 did not show such strong motivation toward improving her spoken English. Even though she took English classes 3 times a week during the data collection time, she did not seem to be really motivated. This may have been because she was not a student. She said that she made an effort to improve her English, but that she did not have any specific aim to achieve. She expressed a desire to speak English well because she lived in the target language community, but she did not seem to have many chances to communicate with native speakers outside of her English

class. Her initial production level was not high, and she did not improve much after the training.

The other possible factor would be anxiety. A high level of anxiety seems to hinder performance in L2. For example, participant 12 started from a very low level of proficiency. She expressed a general anxiety in speaking English, although she said that she performed well in other domains, such as reading. She seemed to have a severe level of anxiety in speaking, and she said that she even had a physical problem due to the anxiety. She showed a bit of anxiety in the actual production tests (she had a shaky voice), which might have led her to perform worse than she actually could have done. In all 3 production tests, her correctness was poorer than the group average, although her perception test scores were better than the group average after the training. Her level of motivation, from the impression that the researcher received based on our conversation, was not as great as the others.

In conclusion, even though these were not structured interviews, they revealed that the level of motivation was different among KE, and this might have contributed to the different level of success. Generally, students seemed to have a higher level of motivation than non-students, but this was not always true. In addition, some might say that students had more chances to interact with native speakers than non-students. However, it was also noticed that even though student participants had classes with native speakers, they said that they used English as less than 25 percent of their daily language use. Five of the KE participants were graduate students, and they said that they did not have many opportunities to interact with native speakers, which revealed that their situation was not much different from that of non-students.

In addition, most participants in KE took English conversation classes, whether they were free or not. This gave us the impression that their motivation level was higher than that of the people who did not take any classes, but the level of effort in improving English outside the classroom was different. It seems that whether they were students or not, people who actually made efforts to improve their English succeeded in improving it.

Acoustical Analysis

Acoustical analyses were conducted to explore the possibility of acoustical differences in duration between NK and EC, and of any discrepancy between acoustical characteristics and human judgment. The expectations were that there would be significant differences in the duration of words ending with an alveolopalatal, and that there would not be much discrepancy between acoustical characteristics and human judgment. The reason that words ending with /dʒ/ were chosen was due to their difficulty in production (See Individual Segments in the NK section of this chapter).

Durational Differences in /dʒ/

Acoustical analysis for the duration of final /dʒ/ and its preceding vowel in the 3 groups (KE, KC and EC) was performed. The preceding vowel duration was also measured, because during the perception training many KE participants indicated that preceding vowels were longer when there was no final /i/. This might have affected their production as well. Among the EC tokens, final /dʒ/ tokens of 5 randomly chosen participants, 2 males and 3 females, were analyzed. For the purpose of tallying, when there was 50 percent or greater agreement among EJ on the correctness of certain tokens, they were regarded as *correct/incorrect* tokens. All tokens with final /dʒ/ across 3 *Times*

from KE and KC were analyzed. Consonant and vowel durations were converted into a ratio to the whole word duration, due to the fact that speaking rate was not controlled at the time of data collection.

Duration of /dʒ/ among correct tokens

Correctness of tokens means that identification from EJ and analysis of the vowel traces matched (i.e., did not induce any discrepancy), and that at the same time the tokens were identified as *correct*. For example, when EJ identified a token as *age* (not *age+Vowel*), and this token did not have any vowel trace in the final position, then this token was regarded as *correct*. Table 7-8 summarizes the mean duration of final /dʒ/ and its preceding vowels among KE, KC and EC. It seems that EC produced a longer final /dʒ/. A one-way ANOVA was conducted to examine the mean durational difference in final /dʒ/ tokens among KE, KC and EC. The results showed there was no significant main effect among the 3 groups.

Duration of vowel preceding /dʒ/ among correct tokens

The mean duration of the vowel preceding /dʒ/ was measured and compared. As Table 7-8 shows, there was not much difference in the duration of preceding vowels. When the mean percentage of vowel duration from KE and KC and from EC was compared, no significant main effect was obtained among all 3 groups.

Table 7-8. Mean duration (%) of /dʒ/ and preceding vowel in KE, KC and EC

group	Preceding Vowel: Mean	SD	/dʒ/: Mean	SD
KE	36.03	5.79	35.64	6.29
KC	33.01	3.51	36.25	4.62
EC	35.51	2.21	40.50	2.70

In conclusion, KE and KC both produced final /dʒ/ and preceding vowels similarly to EC with regard to duration. Although EC produced longer final /dʒ/, it did not differ significantly from what KE and KC produced.

Discrepancy between Judges and Acoustical Analysis

Discrepancies between EJ and acoustical analysis mean that human judgment and the analysis of vowel traces did not match. Two possible cases were observed: The first case was when the judges said that they heard an incorrect version of words ending with /dʒ/ (i.e., a word ending with /dʒi/ was heard), but no vowel trace was shown in the spectrogram. Among the total number of 297 tokens judged *incorrect*, only 19 cases of this discrepancy were found (6.7 %). The second case was when the judges said that they heard the correct version of the words ending with /dʒ/, but a vowel trace was shown in the spectrogram. Among the total number of 594 words judged *correct*, only 8 discrepancies were found (1.35 %), which means that there was an almost absolute correlation between human judgment and the duration analysis (error rate=3.03%).

The discrepancy described above was in fact very rare, although it existed in almost every participant's tokens. Only 3 Korean participants failed to produce any discrepant tokens. The next step was to examine the cause of this gap between human judgment and the actual sounds produced.

Discrepant tokens. As discussed above, several tokens with a discrepancy between the acoustical analyses and the EJ' judgment were observed. Two kinds of discrepancy existed: Ones where the token had an /i/ trace but the judges did not catch it (case 1), and others which did not have an /i/ trace but where the judges said that there was (case 2).

Case 1: /i/ traces were there, but judges did not hear them. There were 8 instances of this case. Figure 7-9 shows an example of this case: *change* was judged as *correct*, but the spectrogram shows a vowel trace of /i/. This type of word was compared with the same words that were produced at different times and judged without discrepancy (N=12). For example, when *change* in the pretest was identified with discrepancy, but *change* in posttests 1 and 2 was identified as *correct* (without discrepancy), the duration of /dʒ/ and the preceding vowel of 3 versions of *change* were compared. In a paired samples t-test, the duration of /dʒ/ differed significantly between discrepant and non-discrepant tokens [$t(7)=2.879, p=0.024$]: Discrepant tokens had longer /dʒ/ than non-discrepant ones (mean percentage of /dʒ/ duration: discrepant-43.11 %, non-discrepant-30.14 %). However, with regard to preceding vowels, discrepant and non-discrepant tokens did not differ significantly (mean percentage of preceding vowel duration: discrepant-35.16 %, non-discrepant-40.33 %). It seems that if final /dʒ/ was long, EJ judged a word as *correct*, although there was a following vowel. However, it is possible that those 8 tokens that elicited discrepancies could have been sheer mistakes in human judgment, because of their small number.

Case 2: /i/ traces were not there, but the judges heard them. Nineteen instances of this case were found. Figure 7-10 shows an example of this case: *judge* was judged *incorrect*, but there was no /i/ at the end. This type of words was compared with the same words (N=12) that were produced at different times and identified without discrepancies, as described above. In a paired samples t-test, the duration of /dʒ/ was not significantly different between discrepant and non-discrepant tokens (Mean: discrepant-37.23 %, non-

discrepant-35.43 %). However, the duration of preceding vowel was different: Non-discrepant tokens had a longer preceding vowel than discrepant tokens ($t(11)=3.046$, $p=0.011$). Mean percentage of preceding vowel duration was 44.29 % and 25.71 %, respectively.

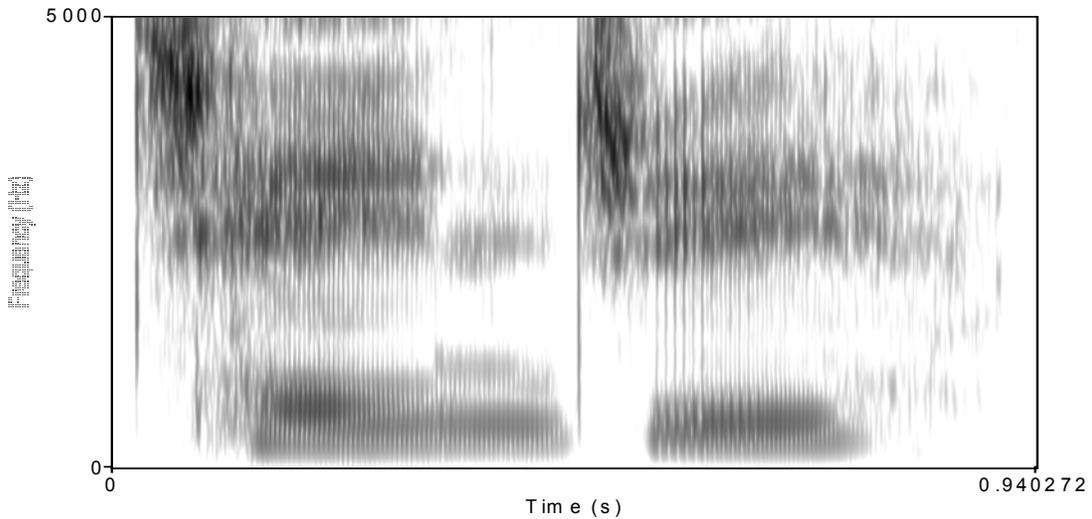


Figure 7-9. *Change* mistakenly judged by EJ as *correct*

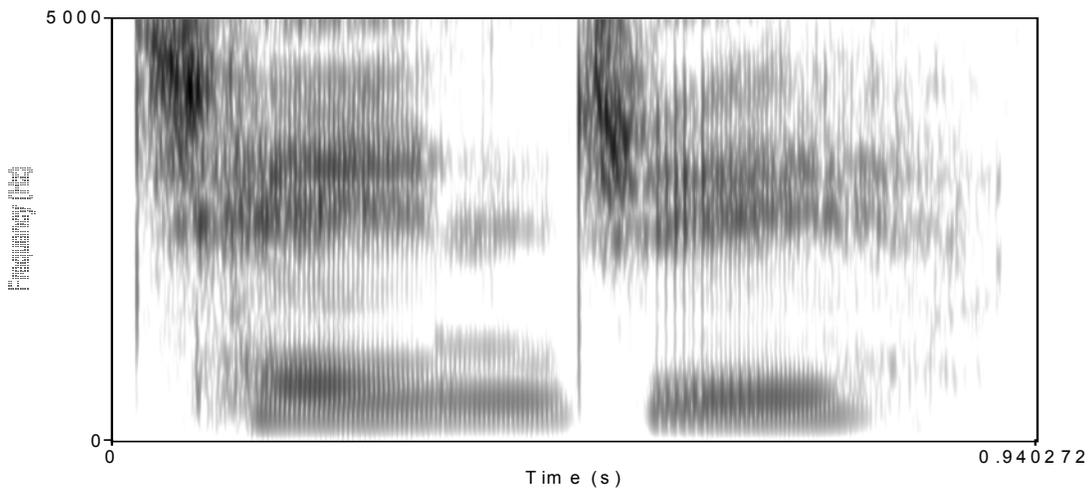


Figure 7-10. *Judge* mistakenly judged by EJ as *incorrect*

When the vowel before final /dʒ/ was short, EJ identified the word as a CVCi word, even though a token had /dʒ/ duration that was similar to that of CVC words and did not

have final /i/. This raises a question about the role of preceding vowels in words ending with an alveolopalatal and an alveolopalatal+i, namely whether native speakers also focus on preceding vowels to determine whether a word ends with an alveolopalatal and an alveolopalatal+i.

Correct vs. Incorrect Tokens in NK (KE and KC Combined)

The mean duration of correctly and incorrectly produced tokens with final /dʒ/ from NK was measured and compared. In this analysis, tokens that induced a discrepancy between EJ judgment and the acoustical analyses were excluded. Since 3 participants did not make any errors in words ending with /dʒ/, only 24 participants' tokens were analyzed. Table 7-9 shows the range of difference in duration of final /dʒ/ and its preceding vowels between correctly and incorrectly produced tokens. In the duration of final /dʒ/, a paired samples t-test revealed that correctly produced /dʒ/ was significantly longer than the incorrectly produced one ($[t(23)=5.337, p=0.000]$). As for the duration of preceding vowels, correctly produced tokens had significantly longer vowels than the incorrectly produced ones, too ($[t(23)=3.668, p=0.001]$).

Table 7-9. Mean duration (%) between correctly and incorrectly produced /dʒ/ and its preceding vowel

	Preceding Vowel: Mean (SD)	/dʒ/: Mean (SD)
Correct	35.07 (5.24)	35.19 (5.36)
Incorrect	27.34 (7.44)	25.97 (6.86)

However, it seems to be natural to have a shorter consonant and a preceding vowel when a word has a longer syllable, i.e., having final /i/. Six tokens that ended with /dʒ/ and /dʒi/ from each of the 5 members in EC were analyzed with respect to the mean

percentage of final alveolopalatal and its preceding vowel duration (i.e., *edge*, *judge*, *wedge*, *edgy*, *judgy* and *wedgy*). Table 7-10 shows the mean percentage of final /dʒ/ duration and its preceding vowel duration in two different types of words. A paired samples t-tests showed that /dʒ/ in CVdʒ was produced longer than those in CVdʒi ([$t(4)=6.593$, $p=0.003$]). In addition, preceding vowels before /dʒ/ in CVdʒ words were produced longer than those in CVdʒi words ([$t(4)=4.399$, $p=0.012$]).

Table 7-10. Mean duration (%) comparison between CVdʒ, and CVdʒi words produced by EC

	Preceding Vowel: Mean (SD)	Final /dʒ/: Mean (SD)
CVdʒi	22.28 (1.96)	28.91 (5.31)
CVdʒ	36.89 (6.14)	52.12 (3.19)

It seems that when native speakers of English judged words ending with an alveolopalatal and with an alveolopalatal+i, they used both the existence of final /i/ and its preceding vowels: when a preceding vowel was long, it also gave a clue that a word had a CVdʒ syllable, rather than a CVdʒi syllable. The preceding vowel duration could be a secondary cue, whenever the primary cue, final /i/, was unclear.

In conclusion, comparison of the production of words ending with /dʒ/ showed that Korean participants did not produce significantly shorter or longer final /dʒ/ and preceding vowels from EC. In addition, there was a discrepancy between human judgment and spectrogram analysis, such as in the case of a word that had a final /i/, but the judges did not detect it. It seems that when a word had a long preceding vowel, EJ tended to judge it as a CVdʒ word, even though a final /i/ trace existed. When correctly

produced and incorrectly produced tokens by NK were compared, correctly produced tokens had longer /dʒ/ and preceding vowels.

Summary

In production KE failed to improve after the perception training, but showed slight improvement in posttest 2 compared to the pretest. The same result was found in KE when words ending with /dʒ/ alone were analyzed.

Individual differences were also found in the KE: Some people improved and others declined in performance in the posttests. However, the correlation between the pretest and a difference between posttest 1 and the pretest suggests that the people who started with low proficiency benefited more.

With respect to a correlation between perception and production, a stronger correlation was found before the training, but not after. This might be due to the fact that the rate of improvement in the two domains was different. In addition, the group with low proficiency showed a stronger correlation in the pretest, but again the correlation disappeared after the pretest.

It seems that the different degree of success in posttest 1 might be related to LOR: The longer participants had lived in the U.S., the better they produced. Other factors might have been in effect as well. Through an interview, it was suggested that the level of motivation and the degree of effort in improving English were also related to the success of training.

Regarding the acoustical analysis of final /dʒ/, it seems that human judgment and vowel trace analysis had an almost absolute correlation (error rate=3.03 %). When correctly and incorrectly produced tokens were compared, correctly produced tokens had

longer final /dʒ/ and longer preceding vowels than incorrectly produced ones, and the same pattern was evidenced when words ending with /dʒ/ and with /dʒi/ produced by EC were compared. Considering the fact that EJ misidentified a certain word as *correct*, it is possible that the preceding vowel is a secondary cue in perceiving differences between words ending with an alveolopalatal and an alveolopalatal+i.

CHAPTER 8 DISCUSSIONS AND IMPLICATIONS

The results of the perception and production tests revealed several important aspects in the training of English final alveolopalatals for Korean speakers of English, the relationship between perception and production, individual differences and temporal properties of Korean participants' production of final /dʒ/.

First, the training was helpful in the perception of words ending with an alveolopalatal and with an alveolopalatal+i in the Korean experimental group (KE). They also sustained the ability 3 months after the training. Second, the training seemed to benefit individual members of KE differently. Participants who started with low level of proficiency improved more than the others. Third, the correlation between the perception and the production of words ending with an alveolopalatal and with an alveolopalatal+i was significant in the pretest, but the significant relation disappeared after the pretest. Fourth, the training did not affect the accuracy in producing words ending with an alveolopalatal immediately after the training in KE, but they seemed to improve 3 months after the training compared to the pretest. However, statistically significant improvement was not obtained in either posttest 1 or posttest 2. It is possible that perceptual accuracy is the precursor of production accuracy, but the latter takes more time. Finally, with respect to the acoustical analysis, KE and the Korean control group (KC) produced significantly shorter final /dʒ/ than EC, but shorter /dʒ/ did not seem to shift the identification decision of the English control group (EC). When correctly and incorrectly produced words

ending with /dʒ/ by KE and KC were compared, incorrectly produced words had a significantly shorter /dʒ/ and preceding vowel duration.

In this chapter, we will discuss the effect of training in both perception and production of word final alveolopalatals, the relation between perception and production, individual differences, and the durational properties of words ending with an alveolopalatal, which affected the judgment of EJ. In addition, the ways in which our results support or oppose theories in L2 perception and production will be explored. Finally, educational implications and research implications from our study will be discussed.

Effect of Training

Perception

The first question that we tried to address in our study was whether native speakers of Korean and English differed in their perception of final alveolopalatals in English. The results obtained suggested that KE and KC both were significantly different from EC in the pretest with respect to distinguishing words ending with an alveolopalatal and an alveolopalatal+i. This shows that L1 background affected the perception of these words. However, after perception training KE and EC did not seem to show a difference in distinguishing two types of words. Hence, it is possible that through training, nonnative speakers are able to improve significantly, at least in the perception domain.

The second and third questions that were addressed in the hypotheses were whether KE would do better in posttest 1 and in the generalization tests than KC. Both of the hypotheses were supported: The training was effective for KE in improving the ability to discriminate words ending with an alveolopalatal and with an alveolopalatal+i. KE did

significantly better after the training than before the training, and there was a significant difference between KE and KC in posttest 1. In addition, KE performed better than KC in generalization test II, where new words and new talkers were included. Finally, the next hypothesis, which addressed whether KE would sustain the learned ability and distinguish words ending with an alveolopalatal and with an alveolopalatal+i better than KC 3 months after the training, was also supported with a statistical significance: KE successfully sustained the ability to distinguish the two types of words 3 months after the training.

These findings showed the effectiveness of intensive perception training with multiple talkers. Additionally, in generalization test II, KE performed better than KC in both posttest 1 and posttest 2. KE used the learned ability to perceive new stimuli, which were not part of the training or the pretest. In generalization test I, where final stop segments were included, KE did not show any significant improvement after the training, because in the pretest KE and KC had already achieved a high level of accuracy.

The improvement is of importance in providing evidence in agreement with the idea that perception training was effective for the trainees to separate a single category of an alveolopalatal+i in the L1 into two distinctive categories of an alveolopalatal and an alveolopalatal+i in the L2: The 9 intensive sessions of perception training were found to be effective in establishing two distinctive phonetic categories in the L2. In addition, it is promising to examine the fact that KE used the learned ability when they listened to new stimuli, which implies that the two newly separated categories were stable enough to process new tokens.

The results in the perception tests were in an agreement with prior studies by Bradlow and her colleagues (1997, 1999). In their studies, intensive perception training helped Japanese speakers improve their perception of English /ɹ/ and /l/. They also used multiple talkers in the training, and it resulted in helping the participants adjust to new talkers and new stimuli successfully.

One of the findings in our study was that KC also seemed to improve from the pretest to posttest 1 and posttest 2 in the analysis of the total scores across 3 subsets (subsets 1, 2, 3). They also improved in generalization test II, which included new stimuli, from posttest 1 to posttest 2. This implies that they also gained the sensitivity for two different types of words after taking the pretest. The reasons that they also improved are suggested to be due to two factors. First, it might be due to the test-retest effect. However, considering the fact that the interval between the pretest and posttest 1 was one month, and the interval between the pretest and posttest 2 was 4 months, it does not seem to be a simple test-retest effect. A second possibility is that exposure to the target sounds in the L2 community during this time helped KC become more sensitive to the differences between words ending with an alveolopalatal and an alveolopalatal+i. It seems that by taking the pretest, awareness to differences between the two types of words was increased. Although KE and KC showed a significantly different level of success, both benefited from taking the tests.

The idea of raising awareness in learning an L2 has been discussed mostly in the acquisition of grammatical morphemes. For example, VanPatten (1996) claims that L2 learners have to ‘detect’ a specific bit of information in L2, which requires more attentional energy than just processing information. Tests such as the one used in our

study might have triggered the detection strategy of L2 speakers. Detected information is stored in their working memory and available to the speakers when they encounter such words in real life.

Hence, the pretest raised awareness of the differences in two different types of words, which in turn helped KC improve their perception. In addition, the participants performed better in the perception of two different types of words 4 months after the pretest.

Our results are also in agreement with the idea that there are certain speech segments in a language which are easier to detect than others. In Bradlow and colleagues' study (1997), a Japanese control group did not show any difference between the pretest and the posttest in the perception of English /ɹ/ and /l/, which were different in a spectral domain, rather than a temporal one. It seems that accidental learning seems to only be possible for certain speech segments, such as final alveolopalatals. Furthermore, their participants were not residing in the target language community, where nonnatives could be surrounded by rich input. In our study, all participants were exposed to the target language community, which may have helped them improve greatly.

Stop vs. alveolopalatal fricative/affricate

In generalization test I, where final stop segments were included, we did not observe any significant improvement from the pretest to posttest 1 or posttest 2 in either KE or KC. This might have been because most of the participants already established an accurate model of final stops in English. The initial hypothesis was that Korean speakers might perceive an extra vowel /ʊ/ after final stops, and hence by being trained with more difficult segments (i.e., alveolopalatal fricatives/affricates), trainees would transfer the

gained ability to the perception of easier segments (i.e., stops). However, at least in terms of perception, they already perceived the difference between words ending with a stop and with a stop+ ω successfully in the pretest. In addition, they distinguished C₁VC₂ words, with C₂ indicating a stop with non-release, and C₁V words separately (e.g., [zat] with non-release [t] and [za]).

Question remains, however, with regard to why Koreans performed well at perceiving final stops, but not fricatives/affricates. This might be due to the differences in the acoustic quality between stops and fricatives/affricates in final position. It is suspected that acoustic energy was more distinct when a stop becomes an onset of the following syllable than when a stop was in final position. On the other hand, it is assumed that there was no significant amplitude difference between final and initial fricative/affricates. In addition, the release of airflow in any final consonants was very foreign to Koreans. Even though Korean language has voiceless stops in word-final position, they are unreleased. The fact that there was sustained airflow in the final consonant gave the perceptual impression that a word has a certain segment following, possibly a vowel. Another hypothesis is that Koreans used an L1 strategy. Korean allows word-final voiceless stops. Hence, when Koreans had to distinguish between words ending with stop and with / ω /, they resorted to using the L1 strategy in distinguishing between final voiceless stop and final / ω /. On the other hand, final alveolopalatal fricatives/affricates are not present in the L1 phonology, so that Koreans could not use their L1 phonological strategy.

The accurate perceptual model of final stops, in turn, seemed to result in the accurate production of final stops in English. Among 162 tokens produced by NK (6

tokens for each of the 27 participants), only 4 tokens were judged to have been produced incorrectly. Regardless of whether final stops were voiced (although most of the errors were found in the production of *bead* in the *Naming* group), or whether a preceding vowel was tense, NK produced final stops without inserting a vowel at the end. Most tokens with final stops produced by NK were not included in the judgment, because from the beginning, NK did not seem to make any errors (at least to the researcher's ears).

This is somewhat unexpected, first in that many studies have noted that Koreans produce an extra vowel after a final stop (Browselow & Park, 1995), and second in that many Koreans seem to make such errors when they talk in casual conversation. For example, when the researcher first recorded participant 1's data, as soon as the recording was over her friend came in the room, and they started to talk in English. At that time, it was noticed that some words with a final stop were produced with /ʉ/. The reason why most participants produced final stops without an extra vowel might be that the recording was carried out in the laboratory setting. Some researchers (Tarone, 1983) suggest that in such a formal context, L2 speakers tend to focus on forms so that they produce more correct forms than they normally produce. Hence, even though the Korean participants might have produced an extra vowel in casual conversation, they knew that they should not do it in the testing environment. In addition, all 4 errors were produced in the naming task, which might have diverted more attention from forms. The second possibility is that only true beginners make such errors. From the beginning, most Koreans who participated in the experiment did not produce an extra vowel. Even though they often rated themselves as being in the beginning/intermediate level, they were not true beginners. Since all of the Korean participants had completed at least 6 years of

compulsory English education, and all of them had gone to college, where they took several English classes, none of them was a true beginner.

Yet, the fact that only two percent of the total tokens showed errors in words ending with a stop implies that the Korean participants produced them in a native-like manner. This was also true in their perception: They successfully distinguished words ending with a stop and with a stop+u. These results lead to the conclusion that the interlanguage of the participants moved beyond the limitations of Korean phonology, which does not allow any word-final voiced stops. Regardless of the voicing status of a stop and the tenseness of a preceding vowel, Koreans did not perceive or produce an extra vowel after a final stop consonant.

Individual differences in KE

Although all KE completed 9 sessions of training, the effect of training seemed varied: the training was more effective for those who had started with lower accuracy in the pretest. For example, participant 12 received the worst score in the pretest, but this participant improved the most in posttest 1. On the other hand, participant 11, who received higher scores than the group average in the pretest, gained only slightly in posttest 1. These results imply that the training method helped learners with low proficiency more than learners with high proficiency.

These results were not the same as those observed by Bradlow et al. (1997), in whose study the poorest performers did not improve as greatly as the better performers. It is suspected that inherent phonetic difficulties relating to the distinction between /ɹ/ and /l/ necessitate more time and effort to distinguish between these segments than between words ending with an alveolopalatal and an alveolopalatal+i

(temporal difference). In other words, words ending with an alveolopalatal and with /i/ require less time and effort to separate them, a process which is easy for L2 learners as long as they “detect” the difference. Another possible explanation for the observation that participants with a high level of proficiency did not improve as much as participants with a low level of proficiency is that they had already reached their ceiling, or their perceptual domain was ‘fossilized’ due to lack of motivation, self-content, etc. (Selinker, 1972).

The fact that the rate of improvement varied as a function of proficiency was also partly confirmed by the "room for improvement." A significant negative correlation was found between room for improvement and pretest accuracy ($r=-0.571$, $p=0.026$). This implies that those who started at a low level improved more than the ones who started in the high level. This correlation was almost the opposite to that observed in Bradlow’s study (1997). Their correlation coefficient was positive ($r=0.73$, $p=0.021$). In our study, however, most participants who scored highly in the pretest did not improve in posttest 1 as greatly as those who scored low. It seems that the training benefited less proficient participants more greatly. The same finding was obtained with the simple correlation between the pretest and a difference of the pretest and posttest 1 accuracy: The lower a person’s accuracy was in the pretest, the more she or he benefited from the training.

Training

With respect to the training, KE performed better in weeks 2 and 3 than week 1, and there was no significant improvement from week 2 to week 3. This result, again, was different from the study on training /ɹ/ and /l/ to Japanese speakers (Logan & Pruitt, 1995). The researchers claimed that their participants’ ability was the best around the

tenth session out of fifteen sessions. It is possible that the spectral differences between /ɹ/ and /l/ take longer to learn than durational ones, such as with the words used in our study. In addition, Logan and Pruitt used 5 talkers, instead of 3 talkers (as in our study), to give a wider range of variability of segments to the trainees.

The training method used in our study involved multiple talkers, (the so-called “high variability training paradigm”), so that it would include the various ranges of target sounds (Logan, Lively & Pisoni, 1991). The premise of this training paradigm is that some talkers are more intelligible than others. Since in our study one talker's tokens were used throughout a whole week, it was expected that the first day of a new talker would have resulted in a decline in correctness. However, as the training progressed, it seems that the participants became well adjusted to new talkers after listening to the first 3 or 4 tokens. This might have been because words ending with an alveolopalatal and with an alveolopalatal+i were easier to distinguish than words with /ɹ/ and /l/.

Even though there was no significant decline in perception accuracy due to different talkers, the use of several talkers in training leads to improvement in two ways: KE performed better in posttest 1 and posttest 2 than in the pretest. In addition, they perceived new talkers and new stimuli in generalization test II better than KC. It seems that training with highly variable stimuli helped KE establish the two categories more distinctly.

Although the mean scores did not show any difference in the intelligibility of different talkers, some participants expressed the opinion that one talker was more difficult to understand than the others. However, opinions were not consistent: Some

participants said that the week 2 talker was more difficult to understand, and others said that the week 3 talker was more difficult to understand.

Production

We first addressed the hypothesis on whether EC, KE and KC had a significant difference in the pretest production. Unfortunately, this was not suitably explored due to the vast number of tokens. However, considering the fact that English Judges (EJ) made mistakes very rarely in judging tokens from EC (error rate: 4.9 %), we can assume that EC performance was almost perfect. On the other hand, the mean percentage of accuracy in the entire Korean group (NK) in pretest production was 75.7 percent. Although not all tokens from EC were used in the judgment, it is conceivable to claim that NK and EC had a significant difference in their production of English words ending with an alveolopalatal.

The hypothesis about whether KE would improve in production after perception training was not strongly supported in posttest 1 and posttest 2, but the training effect in posttest 2 was better than that of posttest 1: Improvement in perception after training appeared to extend to improvement in production 3 months after the training. Comparing the pretest and posttest 1, we did not see any significant improvement in KE, although KE improved and performed better than KC in terms of the raw increment of accuracy. KE performed better in posttest 2 than in posttest 1 without further training. Furthermore, the difference of posttest 2 and the pretest was greater in KE than in KC. In addition, when words ending with /dʒ/, which induced the greatest number of mistakes in NK, were compared across 3 different times (pretest, posttest 1 and posttest 2), KE improved greatly in posttest 2 (although no statistical significance was found due to the small

number of participants in each group). With this result, it is suspected that improvement in production might lag behind that of perception.

The issue of the development of production lagging behind that of perception has been discussed before (e.g., Flege, Munro & Mackay, 1995b). This is clearly demonstrated in L1 acquisition: Perception always precedes production, since it takes time for an infant to master in maneuvering muscles to make correct sounds (Kent, 1992). However, it is not always true in L2 acquisition. For example, Flege, Munro and Mackay (1995b) note that correct perception is necessary for correct production, but it does not necessarily lead to correct production. This could be true in the perception and production of final alveolopalatals: KE improved in the perception more than they did in the production. The question remains as to why this happened. The first possibility is that age affects perception and production differently. In perception, the ability to learn a new segment is dormant until new attention is paid (Wood, 1996), and in production, the ability is lost due to the absence of stimuli ('Use it or lose it,' Bever, 1981, cited in Birdsong, 1999). Therefore, it takes more time for adults to recover the lost ability to produce a new segment.

Furthermore, it is possible that perception and production are controlled by two different domains, the first one by the cognitive domain and the second one by the muscular domain, and if the two are controlled by separate domains, the rate of development might be different.

Production of words ending with /i/

KE also improved their production of words ending with /i/ greatly (but not in a statistically significant way) after the pretest. It seems that KE were not quite sure how to produce these types of words in the pretest, even though the researcher modeled them for

the *Wordlist* group. Because in the pretest the participants did not know of the existence of such words, they did not produce final /i/ in words ending with an alveolopalatal+i. However, even in posttest 1, some participants in KE did not realize that they were exposed to these types of words in training and did not produce final /i/. After the recording of posttest 1, the researcher reminded them that they had heard those words before, and then they finally realized that they had been exposed to this type of words in the training. It is suspected that since Korean speakers had been told not to put any vowels after a final alveolopalatal in school, they unconsciously treated words ending with an alveolopalatal and with an alveolopalatal+i as the same in production. This was shown more clearly in words ending with /ʃi/ and /tʃi/.

Errors in words ending with /i/

Among the words ending with /i/, NK produced more errors in words ending with /ʃi/ and /tʃi/ than with /dʒi/. The opposite was observed with the result of words ending with an alveolopalatal: NK produced more errors in words ending with a /dʒ/.

Participants who produced an extra vowel in words ending with /dʒ/ produced /dʒi/ words better than /ʃi/ and /tʃi/ words. It seems that they were used to adding an extra vowel to words ending with /dʒ/, but were not familiar with adding a vowel to the words ending with /ʃ/ and /tʃ/. One possible reason that NK tended to add an extra vowel to voiced alveolopalatal affricates more easily is because of the higher frequency of words ending with /dʒi/, such as *energy*, *strategy*, *technology*, etc. According to Waring (2004),

only those 3 words were appeared in *2000 Most Frequently Used Words* in English. No words ending with /ʃi/ and /tʃi/ were appeared in this list.

Individual differences in KE

The wide range of individual differences after the training in KE was also noticed in the production test. Some participants started with a comparable level of accuracy in the pretest, and some improved more than others. In addition, some showed a decline in accuracy in posttest 1. The correlation between *room for improvement* and pretest accuracy revealed that there was no significant relation. However, when the correlation between the difference of posttest 1 and pretest accuracy and the pretest accuracy was compared, the correlation coefficient was suggestive ($r=-0.51, p=0.055$). This relation implies that after training those who performed poorly in the pretest gained more than those who did well.

Group with low proficiency. This speculation was partly supported by the analysis of the group with low proficiency. When participants in KE who received lower than 80 percent accuracy in the production pretest were isolated and analyzed in terms of *Time*, they seemed to show a great increase in production accuracy between the pretest and posttest 2. Those who received more than 80 percent accuracy in the pretest showed a smaller increase from the training. Those who were in the KC low proficiency group showed smaller amount of improvement, too. The finding obtained here suggests that those who started with low proficiency benefited more from the perception training.

Correlation between Perception and Production

The hypothesis about the correlation between perception and production of final alveolopalatals was not strongly supported, but it seems that the strength of the

relationship depends on *Time*. In NK (KE and KC combined), the correlation between perception and production was significant before the training ($r=0.457, p=0.017$), but the strong relationship was disrupted after training. This is understandable because KE performed better in perception in posttest 1 and posttest 2 than in the pretest, but did not perform better in production.

Compared with other studies using similar methods of teaching nonnative sounds, such as Bradlow et al. (1997, 1999), the present results were a bit different. In their 1997 study, most Japanese participants perceived and produced better in the posttests than in the pretest, so that the relation between perception and production was strong in all tests. Why is there such a difference between our study and their study? It could be because of some methodological differences between the two studies. First, participants (i.e., L1 background) were different, Japanese vs. Korean. Second, the segments in question were different. Theirs were English /ɹ/ and /l/, with which Japanese learners had, arguably, perception difficulties more than production difficulties. In our study, Koreans were assumed to have more production difficulties in words ending with an alveolopalatal. Third, Bradlow and her colleagues trained the participants for fifteen sessions, but in our study, only 9 sessions were given.

Rate of improvement in KE. A correlation analysis of improvement between perception and production in posttest 1 and posttest 2 revealed the relationship not to be strong. Yet, the weak relationship between the perception and production improvement might be because of the wide range of individual differences. For example, as Table 8-1 shows, in the perception tests participants 8 and 9 started with a similar level of proficiency in the pretest and improved in similar ways in posttest 1; however, in terms of

production, participant 9 improved greatly in posttest 1 (a difference of posttest 1 and pretest: 24.39 percent), whereas the accuracy of participant 8 declined (a difference of posttest 1 and pretest: -16.15 percent). Bradlow and colleagues' experiment (1997) also showed varying degrees of improvement in perception and production after training.

Table 8-1. Percentage of correctness of participants 8 and 9 in KE

	Perception				Production				
	pre	pt1	*pt1-pre	pt2	pre	pt1	*pt1-pre	pt2	*pt2-pre
8	56.45	88.71	32.26	89.43	81.3	65.15	-16.15	76.74	-4.56
9	58.87	90.32	31.45	51.52	67.08	91.47	24.39	—	

* indicates a difference of two tests in accuracy

We do not know exactly what caused the decline of production accuracy of participant 8; however, it is suspected that the rate of improvement in production can vary depending on individual factors. In posttest 2, the accuracy of 6 participants declined compared to the pretest, and that of 8 participants improved. It is possible that more training and/or time might have resulted in better production later on. Given the fact that the rate of improvement differed from person to person, it is suspected that there were people who could detect the subtle differences from the perception training and extend the knowledge to the production domain faster than others.

It is also conceivable to speculate that some participants focused on secondary cues to perceive and produce the two different kinds of words, ending with an alveopalatal and with an alveopalatal+i. During training, many expressed their opinion about the differences. For instance, one explained that words ending with an alveopalatal+i had a stressed and longer final vowel. Many said that words ending with an alveopalatal had a longer vowel, and words ending with /i/ had a shorter vowel. This might be true, but the most important part of the differences was that words ending with an alveopalatal+i had a vowel in final position. It is possible that they never caught this difference and used the

previous vowel length difference when producing. This perceptual strategy that many KE described was reflected in actual performance. In the acoustic analysis, when incorrectly produced words ending with /dʒ/ and correctly produced ones were compared, correctly produced ones had longer previous vowels.

Individual Differences and Other Factors

The subsequent question is why some KE could improve production only by taking perception training, but others could not. Earlier in this chapter, it was suggested that participants with low proficiency benefited more from training, but the correlation between improvement and pretest accuracy levels in this group was not sufficiently strong. The other possibilities could be that other factors play a role. Several factors were considered and compared with perception and production accuracy first in the entire NK group and then in KE alone. None of the factors examined, including age, length of residency (LOR), TOEFL score, L2 use, major, and gender, turned out to be a significant factor. When KE alone were analyzed, LOR seemed to be a significant factor. However, this should be taken with caution, because the data were not normally distributed: LOR of most participants was less than 10 months. What was surprising was that chronological age did not seem to have a strong correlation with any of the tests, which may be because the age range was very wide (Range: 20 to 35 years). This leads to the idea that something else might have affected the result.

Interview. Several possible factors, besides the factors listed above, affecting improvement were suggested through the interview, and one of them was motivation. Although most of the participants in KE were motivated enough to participate in such a time-consuming study, some seemed more motivated than others. Participants who

showed strong motivation in speaking more accurately improved more greatly than those who did not.

Participants who were motivated performed better in both the perception and production of words ending with an alveolopalatal. In addition, initially it seemed that students were more motivated than non-students; however, the most important factor was not whether participants were students, but whether they actually made efforts to improve their spoken English: People who actually made efforts to improve their English did improve.

Acoustical Analysis

Since there were significantly more errors observed in words ending with /dʒ/ than with /ʃ/ and /tʃ/, the duration of final /dʒ/ and its preceding vowel was measured and compared. In comparing the duration of /dʒ/ between correctly and incorrectly produced words by NK (KE and KC combined), it was found that incorrectly produced /dʒ/ was shorter than the correctly produced one. In addition, incorrectly produced /dʒ/ had shorter preceding vowels, too. This is exactly how native speakers of English produced words ending with /dʒi/.

If this is true, the opposite result might have obtained in words ending with /ʃi/ and /tʃi/. NK produced more errors in those words than words ending with /dʒi/. In other words, when NK made errors in words ending with /ʃi/ and /tʃi/, we expected that they would produce longer preceding vowels and final consonants, as they produced words

ending with /ʃ/ and /tʃ/ correctly. The investigation of the word *catchy*, which was produced mostly as *catch*, revealed otherwise: There was no significant durational difference between the correctly produced *catchy* and incorrectly produced ones. In addition, the impression that the researcher received from listening to final /ʃi/ and /tʃi/ words also leads us to suspect that NK did not produce those words in a native-like manner. Even when words were judged to be *correct*, many times participants were not quite sure how to add final /i/. Therefore, sometimes they lengthened the final vowel, or final /ʃ/ and /tʃ/.

From a phonological perspective, all 3 segments, /ʃ/, /tʃ/ and /dʒ/, should have triggered errors equally, because Korean does not allow any alveolopalatal consonants in word final position. However, /dʒ/ triggered more errors than /ʃ/ and /tʃ/. Conversely, when NK were asked to produce words ending with a vowel, they produced /dʒi/ more correctly than /ʃi/ and /tʃi/. This could be because of the higher frequency of words ending with /dʒi/ (e.g., *energy*).

Duration of final /dʒ/ and preceding vowel. The analysis of duration in correctly identified tokens with final /dʒ/ revealed that with regard to duration of final /dʒ/ and preceding vowels, there was no significant difference among KE, KC and EC.

Among words that did not have any vowel traces in the spectrogram but in which EJ thought they heard final vowels, the duration of the preceding vowels was found to be shorter than that of the correctly produced words. This leads us to the question of whether

native speakers of English also focused on preceding vowel length as one of the critical cues to identify words ending with an alveolopalatal and an alveolopalatal+i. Of course, the final vowel was the most critical cue, given the fact that there were very few discrepancies between the vowel trace analysis and EJ identification on words that were incorrectly produced by KE and KC. The shorter preceding vowels were also found in words with /dʒi/ produced by EC: They produced shorter /dʒ/ and a shorter preceding vowel when they produced words ending with an alveolopalatal+i (e.g., *edge* vs. *edgy*).

Different Models of Speech Perception

Three different models of speech perception make different predictions in the relation between perception and production. Motor theory predicts that when a perceptual category changes, production also changes, because they are mediated by a module. Hence, perception improvement consequently results in production improvement. In acoustic/auditory theory, the two domains are separate. Perception domain only monitors the production domain. Hence, more precise categories in the perception domain help more correct production through monitoring. Direct realists believe that perception and production are directly related, and hence that changes in perception result in changes in production without any intermediating module (Bradlow et al., 1997).

The present results are not in strong agreement with any of the theories. Since perception and production were not strongly correlated after the training in KE, the results might support auditory/acoustic theory: Production only improves through producing sounds. If production training had also been given, KE might have improved in production more greatly.

However, the improvement of production in posttest 2 shed some light on the relationship between perception and production: Changes in perceptual categories lead to changes in production, although it takes more time to establish two distinct categories in production. Perception and production might be related, but the rate of changes is different.

The question remains as to why it took a longer time to show improvement in production. When KE took posttest 2, there had been no sustained training at that time. Yet, they produced better in posttest 2 than in posttest 1, when KE had just finished the training. Suspected reasons are first that KE sustained the enhanced sensitivity to those words ending with an alveolopalatal and with an alveolopalatal+i after the training. They could not use the sensitivity right after training because of lack of practice. Even though many participants did not have many chances to interact with native speakers of English, KE might have used the learned ability in real life, monitored their production and improved it as time went by.

The second reason would be that production just simply took a longer time to be materialized. This may be because the muscular settings needed more time to be rearranged: Production changes involved changes in orchestrating different muscular activities. This assumption implies that perception and production might lie in two separate levels, high-level linguistic representation and low-level muscular manipulation.

The third reason would be that, immediately after the training, KE were confused between the two types of words. Examining the negative, although not strong, correlation between the rates of improvement in perception and production, it is suspected that there might have been a category reversal in production, meaning that KE sometimes produced

words ending with an alveolopalatal as words ending with an alveolopalatal+i and vice versa. However, being exposed to the target language, KE might have sorted out the difference and produced more distinctively in posttest 2.

Second language speech learning model. The relationship between perception and production has been discussed within the realm of L2 learning models of the Speech Learning Model (SLM). In the SLM, learning an L2 segment is harder when an L2 segment is similar to an L1 segment which is the nearest equivalence. Acquisition of the L2 segment is completed when the phonetic category of the L2 deflects away from that of the L1. For instance, in our study, a final alveolopalatal and an alveolopalatal+i were not in a separate category in the L2 phonetic domain before the training due to Korean phonetic category. After training, the trainees successfully established separate categories of the two types of segments, which was evidenced in the perception posttest 1 and posttest 2. Even though SLM does not explain the relationship between perception and production, it is assumed that if phonetic categories become separated, production will follow accordingly.

However, in our study, the results do not strongly support the SLM. When words ending with an alveolopalatal and with an alveolopalatal+i were compared in the pretest, the results were not consistent at first. When total scores from 3 subsets were tallied and compared, words ending with an alveolopalatal were perceived more correctly than words ending with an alveolopalatal+i. According to the model, words ending with an /i/ should have been perceived more correctly. However, further analysis revealed that in the non-word category, words ending with /i/ were perceived more correctly. In the non-word category, our results support the SLM.

It is suspected that real words ending with an alveolopalatal were lexicalized in NK. In other words, Koreans memorized those words one by one without categorizing them. Furthermore, Koreans were instructed not to insert /i/ after final alveolopalatals. Although they were instructed in that way so that they could distinguish words ending with an alveolopalatal and with an alveolopalatal+i when they could use lexical meaning, it is possible that in the higher level, two types of words were in one category. Hence, when NK were presented with two new words together, their L1 phonological system took over.

The effect of lexical familiarity on perceiving and producing a L2 sound is well observed in many studies, although the effect is not conclusive. For example, Flege and his colleagues (Flege, Takagi & Mann, 1996) found that Japanese speakers of English identified familiar words better than non-familiar words in the tasks of identifying English /ɹ/ and /l/. On the other hand, in another study (Flege, Frieda, Walley & Randazza, 1998), lexical status (frequency, familiarity, etc.) and VOT of English initial voiceless stop production did not correlate strongly. However, this difference could have occurred because one was a perception study, and the other one was a production study.

The reason that NK perceived real words better than non-words is speculated to be the following: First of all, as with children, L2 learners do not learn L2 sounds analytically, but rather holistically (Walley & Flege, 1999). Hence, identification of real words is faster and more correct than identification of non-words. In addition, lexical processes might be different in non-native sound processing, because there is an existing L1 system that L2 learners might use to access an L2 semantic system. Furthermore, lexical processing might have been triggered rather than phonetic processing, even

though meaning association was not required in the perception tests. In other words, real words automatically triggered meaning association, which might have helped the participants identify them more correctly.

The fact that there are such learners who improved greatly after training, for example participants 6 and 14, who were in their 30s when they took the training, is very promising in the L2 pronunciation teaching field. The fact is that there was only perception training available to the participants, but they seemed to extend the perception improvement to improvement in production. In addition, in the perception posttest 1, there was no significant difference in the perception of words ending with an alveolopalatal and with an alveolopalatal+i between KE and EC. In this regards, intensive perception training positively affected L2 learners' proficiency tremendously.

It seems to be true that the source of deviant speech of L2 speakers is the lack of attention. When attention is given to a specific segment through training, even adult learners can improve. To conclude, the training effect on both perception and production was greatly positive.

Educational Implications

Perception

The results from the perception training of English final alveolopalatals seem to offer hope to Korean speakers of English. First of all, after the training the Korean Experimental group (KE) improved their ability to perceive the difference between words ending with a final alveolopalatal and with an alveolopalatal+i. Second, KE sustained the ability 3 months after the training. Third, the effect of intensive training seems to override the critical period effect: In posttest 1, KE did not show any significant difference from EC in the perception of words ending with an alveolopalatal and an alveolopalatal+i.

Although the perception training was carried out in a laboratory setting, this does not mean that this method cannot be implemented in actual ESL classrooms. If a student is suspected of having a perception difficulty, it might be a good idea for a teacher to devise a session of intensive perception training. The training does not need to consist of 9 sessions as in our study; it might involve 2 or 3 sessions. However, they should be structured to give plenty of stimuli that have contrasting pairs with minimal differences.

In addition, using multiple talkers to train the Korean participants seemed helpful. Although in our study multiple-talker-training was not compared with single-talker-training, having multiple talkers in the training stimuli is suggested to be more helpful (e.g., Lively, Logan & Pisoni, 1993). Many participants noted that some talkers were easier to perceive than others, but opinions on who was easier or more difficult to understand were not in agreement. Exposed to a different degree of intelligibility of talkers, the participants started to normalize different acoustic qualities of words ending with an alveolopalatal and with an alveolopalatal+i, and make the categories robust. Although teachers may not be able to provide rich training stimuli such as the ones used in our study, it is important for a teacher to acknowledge the need for various stimuli for students.

It is also important to note that the Korean control group improved the perception of words ending with an alveolopalatal and with a vowel without the training. This could mean that simply by being exposed to a trouble spot in a specific way, learners might benefit. The stimuli could be enhanced by a certain method, such as contrasting minimal pairs. Learners themselves might be aware of the existence of certain problems and try to fix them without being explicitly taught. There should be ways to enhance this implicit

learning, but not much research has been conducted on this aspect of learning pronunciation.

Production

Unlike perception, production did not seem to improve immediately after the training. The training effect was shown more 3 months after the training. This does not mean that the perception training was not helpful for the production of word final alveolopalatals. Rather, it showed that the production improvement took longer to occur than the perception improvement. Teachers need to expect this gap in terms of rate of improvement and not to be frustrated by slow improvement in production.

It also seems that production did not improve as much as perception; this conclusion was reached by examining the correlation between the perception and the production of word final alveolopalatals after the training. It might be a little unrealistic to expect that by receiving perception training, learners improve perception and production at the same rate. This may be because perception and production are in two different domains. Production is a muscular activity, and muscular activities need time and practice to be improved. On the other hand, perception is more of a cognitive activity. Once something is understood, it is easier to improve it thereafter. Production requires practice. To raise the improvement rate of production, articulatory practice may be needed and may need to be accompanied with a perception activity.

It could also depend on how advanced a learner is. In our study, it seems that low level participants benefited more from the training than the high level participants in production. If learners are in the advanced level, articulatory practice might be needed to improve more. If learners are in the beginning level, perception practice might be more necessary in establishing a correct model so that they can practice on their own.

An intensive perception training method would work not only with word final alveolopalatal segments, but also with other segments as well. Giving perception training will help learners establish an accurate model of a speech segment, whether it is a consonant or a vowel. Then, learners can start to produce more accurately, although this might take more time.

The effect of training might depend on the kinds of segments that are targeted. Given the fact that the perception and production of final alveolopalatals in our study did not proceed hand in hand, it is suspected that the reason would be that words ending with an alveolopalatal and with an alveolopalatal+i demonstrated more of a durational difference. The studies of perception and production of English /ɹ/ and /l/ by Japanese speakers (Bradlow et al, 1997, 1999) give us more clues: Spectral differences in contrasting segments might trigger more linear improvement in perception and production. Hence, it would be possible to see perception and production improvement in certain vowel contrasts, which differ spectrally, more linearly if intensive perception training were given.

Of course, the success of any kind of L2 training depends on learners themselves. Because we saw the varying degrees of training effect in KE, it is apparent that other factors were at work. A suggested factor would be motivation. The level of motivation can be raised by teacher's effort. For example, teachers can devise a class to explain why it is important to improve pronunciation. Encouraging learners to listen to American pop music or watch movies may also be a good way to motivate learners.

Finally, learners should not be discouraged by their age limitation where L2 pronunciation is concerned. The critical period is not likely to be avoidable, but this does

not mean that adult L2 learners cannot attain a high level of proficiency. Many studies, including our study, have shown that training helps learners to overcome the age limitation. Although the ultimate attainment would not be the same as that of early bilinguals, training would help adult learners improve their pronunciation accuracy.

Research Implications

Since the participants in KE were volunteers and not really a random sample of Korean speakers of the English population, our study was not a truly controlled experimental study. However, in the SLA field, it is nearly impossible to achieve a random sample of a particular population. Even so, it would have been more beneficial to have more participants in order to have stronger reliability and validity. In particular, the production tests involved two different methods of elicitation, and this lowered the power of statistical analyses. Furthermore, the delayed posttest was conducted only 3 months after the training. Longitudinal studies, having a delayed posttest at least one year after the training, would be necessary to expect more generalizable sustainability effects of the training.

In addition, since most of the intensive perception training was performed on English consonants, it would be of interest to examine the effect of training on English vowels. For example, Korean speakers tend to produce English /i/ and /I/ in the same way, because the Korean language does not have /I/. Using the intensive perception training, some researchers might be able to answer questions such as: How long should training be for L2 learners to perceive and produce different vowels more correctly? Will there be more linear improvement between perception and production if speech segments are vowels? Will the use of multiple talkers be more beneficial than the use of a single talker? So far, there have been studies on the relationship between perception and

production of English vowels by Korean speakers of English (e.g., Ingram, Park & Mylne, 1997), but no training studies have been conducted yet.

Furthermore, it is of interest to explore the difference in real word and non-word perception. In the perception tests, Korean participants perceived real words ending with an alveolopalatal more accurately than words ending with alveolopalatal+i. In non-words, the pattern was reversed. It would be of interest to examine whether Koreans produce non-words better than real words.

Moreover, it would be important to explore possible factors affecting the different degrees of improvement. Tests on affective factors, such as motivation and anxiety (and aptitude), might reveal possible reasons to account for the varying degrees of success.

There are many other possibilities to expand the training studies. For example, a study might apply two different training methods (perception training vs. perception and articulatory training combined) and examine the difference in rate of achievement. Of course, time limitations and the methods of articulatory training should be considered. The combined technique might be more beneficial, but it might also take more time. In addition, there are several methods available for articulatory training: Computer-based imitation training, visual training such as showing movement of articulators through a sagittal section of the mouth, etc. All designs of the training, of course, depend on the learner's proficiency level. A well-controlled study, especially with control of time spent in both types of training, will also shed light on the relationship between perception and production.

Learning an L2 as an adult would be a tremendous challenge. Many learners complain about the poor improvement in their pronunciation even after living in a target

language community for several years. It is very rare that adult L2 speakers do not have foreign accents, so most adult learners envy the early bilinguals. However, appropriate training helps. Our study showed that training helped learners improve in perception and production. Learners and teachers alike need to recognize the possibilities of improvement, rather than being pessimistic about the age limitation.

APPENDIX A
PERCEPTION PRETEST STIMULI

Table A-1. Perception pretest stimuli

Subset 1		Subset 2	
bidge	tathe	geeb1	ti
bidgy	wof	geeb2	tik1
fos	wofe	geeb3	tik2
fose	zith	jee	tik3
hoch	zithe	jeeg1	vee
hochi		jeeg2	veeb1
huth		jeeg3	veeb2
huthe		lee	veeb3
laf		leeg1	zeed1
lafe		leeg2	zeed2
lage		leeg3	zeed3
lagy		mee	1-unreleased
lidge		meek1	2-released
lidgy		meek2	3-release
luch		meek3	with a vowel
luchi		meep1	
mich		meep2	
michi		meep3	
nes		peed1	
nese		peed2	
nish		peed3	
nishi		reet1	
nush		reet2	
nushi		reet3	
pesh		tee	
peshi		teet1	
taf		teet2	
tafe		teet3	
tas		theep1	
tase		theep2	
tath		theep3	

Table A-1. Continued

Subset 3		Subset 4	
badge	obuch	ash	peachy
badgie	obuch	ashy	ridge
bash	obuchi	ashy	ridge
bashi	punch	blush	ridgy
catch	punchi	blushy	sludge
catch	sash	catch	sludge
catchi	sashi	catchy	sludgy
effig	teach	dish	smudge
effigy	teachi	dishy	smudgy
elegy	wedge	edge	smutch
elleg	wedgy	edgy	smutchy
eulog		fish	smutchy
eulogy		fishy	wedge
eulogy		fishy	wedge
flush		glitch	wedgy
flushi		glitchy	
hibach		glitchy	
hibachi		hatch	
huge		hatchie	
hugie		itch	
hugie		itchy	
karach		itchy	
karachi		judge	
leash		judge	
leash		judgy	
leashi		mesh	
ledge		meshy	
ledgie		nebbish	
nash		nebbish	
nashy		nebbishy	
nashy		peach	

APPENDIX B
BACKGROUND OF KOREAN PARTICIPANTS

Table B-1. Background of Korean participants

Group	Sex	Age(year)	Major	TOEFL score	LOR (mon.)	Daily use of English (%)
KE	Female	30	Others	—	7	25
KE	Female	33	Humanity	—	7	25
KE	Male	28	Science	—	2	25
KE	Male	36	Science	580 & up	6	50
KE	Male	32	Science	580 & up	6	50
KC	Male	30	Science	580 & up	18	25
KE	Female	32	Science	580 & up	42	25
KE	Female	29	Others	580 & up	6	50
KC	Female	21	Science	—	1	25
KC	Female	30	Science	560–580	7	25
KC	Female	24	Humanity	—	1	75
KC	Male	24	Science	450–470	4	50
KC	Female	20	Others	—	1	50
KE	Female	20	Others	400–450	8	50
KE	Female	25	Others	—	6	25
KC	Male	30	Others	—	5	25
KC	Male	30	Science	480–500	8	50
KE	Female	36	Science	—	1	50
KE	Female	31	Humanity	—	1	25
KC	Female	28	Humanity	580 & up	8	50
KE	Female	28	Others	—	7	25
KC	Female	41	—	—	48	25
KC	Female	38	—	—	16	25
KE	Male	32	Science	560–580	20	25
KE	Male	31	Humanity	510–550	28	25
KC	Male	28	Humanity	560–580	24	25
KE	Female	23	Science	—	1	25

Major “Others” included fine art, architecture, etc.

APPENDIX C
PERCEPTION GENERALIZATION TEST II (SUBSET 5) STIMULI

Table C-1. Perception generalization test II (Subset 5) stimuli

bolsh	pouch	swash
bolshie	pouchy	swashy
bunge	pudg	thrush
bungee	pudgy	thrushy
clerge	rage	trench
clergy	ragy	trenchy
duch	ranch	usage
dutchy	ranchy	usagy
garnish	range	
garnishee	rangy	
hatch	regg	
hatchy	reggie	
hedge	ridge	
hedgy	ridgy	
image	shush	
image	shushy	
imagy	sketch	
imagy	sketchy	
irish	slash	
irishy	slashy	
mich	sluggish	
michie	sluggishy	
ouch	smooch	
ouchy	smoochy	
parish	squash	
parishy	squashy	
pinch	stodg	
pinchy	stodg	
plash	stodgy	
plashy	stodgy	
podg	such	
podgy	suchy	

APPENDIX D
PERCEPTION TRAINING STIMULI

Table D-1. Perception training stimuli

barge	dinge	plush	squelch
bargy	dingy	plushy	squelchy
beach	dodge	pouch	squish
beachy	dodgy	pouchy	squishy
beige	flash	push	starch
beigy	flashy	pushy	starchy
bitch	flesh	raunch	stench
bitchy	fleshy	raunchy	stenchy
blotch	french	rich	stretch
blotchy	frenchy	richie	stretchy
botch	grouch	rubbish	swish
botchy	grouchy	rubbishy	swishy
branch	grunge	rush	torch
branchy	grungy	rushy	torchy
brush	gush	scratch	trash
brushy	gushy	scratchy	trashy
bunch	kitsch	sketch	tush
bunchy	kitschy	sketchy	tushy
bush	ledge	slouch	twitch
bushy	ledgy	slouchy	twitchy
cage	lush	slush	varnish
cagey	lushy	slushy	varnishy
church	marsh	smooch	veg
churchy	marshy	smoochy	veggie
clash	mush	smutch	wash
clashy	mushy	smutchy	washy
crunch	orange	spinach	wish
crunchy	orangy	spinachy	wishy
cush	patch	splash	witch
cushy	patchy	splashy	witchy
dash	pitch	splotch	
dashy	pitchy	splotchy	

APPENDIX E
PRODUCTION STIMULI FOR WORDLIST GROUP

Table E-1. Production stimuli for wordlist group

Say pine/ again	Say ashy/ again	Say English/ again
Say rain/ again	Say large/ again	Say catch/ again
Say watch/ again	Say pick/ again	Say itchy/ again
Say language/ again	Say touch/ again	Say peach/ again
Say sausage/ again	Say bench/ again	Say edge/ again
Say judgy/ again	Say loop/ again	Say wedgy/ again
Say rib/ again	Say dishy/ again	Say judge/ again
Say rip/ again	Say fish/ again	Say ash/ again
Say catchy/ again	Say lube/ again	Say fishy/ again
Say bridge/ again	Say hood/ again	Say came/ again
Say polish/ again	Say finish/ again	Say lame/ again
Say food/ again	Say anguish/ again	Say same/ again
Say edgy/ again	Say suit/ again	
Say varnish/ again	Say foot/ again	
Say average/ again	Say fresh/ again	
Say page/ again	Say salish/ again	
Say church/ again	Say leak/ again	
Say much/ again	Say torch/ again	
Say wedge/ again	Say league/ again	
Say sit/ again	Say seat/ again	
Say Sid/ again	Say peachy/ again	
Say itch/ again	Say coach/ again	
Say beat/ again	Say change/ again	
Say bead/ again	Say inch/ again	
Say mash/ again	Say seed/ again	
Say such/ again	Say hash/ again	
Say age/ again	Say bit/ again	
Say pig/ again	Say bid/ again	
Say arrange/ again	Say British/ again	

*Direction: Read these words with a comfortable speaking rate. (/: pause one second)

APPENDIX F
 PRODUCTION STIMULI FOR NAMING GROUP

Table F-1. Production stimuli for naming group

Repetition of words stimuli	Naming of words stimuli	
Ashy	Age	Such
Catchy	Arrange	Torch
Dishy	Ash	Touch
Edgy	Average	Watch
Fishy	Bead	
Itchy	Beat	
Judgy	Bench	
Peachy	Bridge	
Wedgy	British	
	Change	
	Church	
	Coach	
	Dish	
	Edge	
	English	
	Finish	
	Fish	
	Food	
	Foot	
	Fresh	
	Inch	
	Itch	
	Judge	
	Language	
	Large	
	Much	
	Page	
	Peach	
	Pick	
	Pig	
	Polish	

LIST OF REFERENCES

- Behrens, S., & Blumstein, S. (1988). Acoustic characteristics of English voiceless fricatives: A descriptive analysis, Journal of Phonetics, 16, 295-298.
- Best, C. (1995). A direct-realist view of cross-language speech perception. In W. Strange (ed.), Speech Perception and Linguistic Experience: Issues in Cross-Language Research (pp. 171-204). Timonium, MD: York Press.
- Best, C., McRoberts, G., & Sithole, M. (1987). Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by speaking-speaking adults and infants. Haskins Laboratories Status Report on Speech Research, 91, 1-29.
- Birdsong, D. (1999). Introduction: Why and why-nots of the critical period hypothesis for second language acquisition. In D. Birdsong (ed.), Second Language Acquisition and the Critical Period Hypothesis, (pp.1-22). Mahwah, NJ: Lawrence Erlbaum Associates.
- Bohn, O.-S., & Flege, J. (1990). Interlingual identification and the role of foreign language experience in L2 vowel perception. Applied Psycholinguistics, 11(3), 303-328.
- Bongaerts, T. (1999). Ultimate attainment in L2 production: The case of very advanced late L2 learners. In D. Birdsong (ed.), Second Language Acquisition and the Critical Period Hypothesis (pp.133-159). Mahwah, NJ: Lawrence Erlbaum Associates.
- Bongaerts, T., Mennen, S., & Van Der Slik, F. (2000). Authenticity of pronunciation in naturalistic second language acquisition: The case of very advanced late learners of Dutch as second language. Studia Linguistica, 54(2), 298-308.
- 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
- Bradlow, A., Pisoni, D., Yamada, R., & Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ & /l/: IV. Some effects of perceptual learning on speech production. Journal of the Acoustical Society of America, 101, 2299-2310.
- Bradlow, A., Yamada, R., Pisoni, D., & Tohkura, Y. (1999). Training Japanese listeners to identify English /r/ & /l/: Long-term retention of learning in perception and production. Perception and Psychophysics, 61, 5, 977-985.

- Broselow, E. (1983). Non-obvious transfer: On predicting epenthesis errors. In S. Gass & L. Selinker (eds.) Language Transfer and Language Learning (pp.269-280). Rowley, MA: Newbury House.
- Broselow, E., Chen, S., & Wang, C. (1998). The Emergence of the Unmarked in Second Language Phonology. Studies in Second Language Acquisition, 20, 261-280.
- Cho, T., Jun, S.-A., & Ladefoged, P. (2002). Acoustic and aerodynamic correlates of Korean stops and fricatives, Journal of Phonetics, 30, 193-228.
- Diehl, R., Kluender, K., & Walsh, M. (1990). Some auditory bases of speech perception and production. In W. Ainsworth (Ed.), Advances in Speech Hearing and Language Processing, (Vol. 1, pp. 243-267). London, England: JAI Press.
- Eckman, F., & Iverson, G. (1994). Pronunciation difficulties in ESL: Coda consonants in English interlanguage. In M. Yavas (Ed.), First and Second Language Phonology (pp. 251-266). San Diego, CA: Singular Publishing Group.
- Flege, J. (1984). The effect of linguistic experience on Arabs' perception of the English /s/ vs. /z/ contrast. Folia Linguistica, 18(1-2), 117-138
- Flege, J. (1987). A critical period for learning to pronounce foreign languages? Applied Linguistics, 8(2), 162-177.
- Flege, J. (1991). Perception and production: The relevance of phonetic input to second language phonological learning. In T. Huebner & C. Ferguson (eds.), Crosscurrents in second language acquisition and linguistic theories (pp. 249-290). Philadelphia, PA: John Benjamins
- Flege, J., Bohn, O.-S., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. Journal of Phonetics, 25(4), 437-470.
- Flege, J., & Eefting, W. (1987). Production and perception of English stops by native Spanish speakers. Journal of Phonetics, 15(1), 67-83.
- Flege, J., & Fletcher, K. (1992). Talker and listener effects on degree of perceived foreign accent. Journal of the Acoustical Society of America, 91, 370-389.
- Flege, J., Frieda, E., Walley, A., & Randazza, L. (1998). Lexical Factors and Segmental Accuracy in Second Language Speech Production. Studies in Second Language Acquisition, 20(2), 155-187.
- Flege, J., & Hillenbrand, J. (1986). Differential use of temporal cues to the /s/-/z/ contrast by native and non-native speakers of English. Journal of the Acoustical Society of America, 79, 508-517.

- Flege, J., & Hillenbrand, J. (1987a). A differential effect of release bursts on the stop voicing judgments of native French and English listeners. Journal of Phonetics, 15(2), 203-208.
- Flege, J., & Hillenbrand, J. (1987b). Limits on phonetic accuracy in foreign language speech production. In G. Ioup & S. Weinberger (eds.), Interlanguage Phonology: Acquisition of a Second Language Sound System (pp.176-203). Cambridge, MA: Newbury House.
- Flege, J., & Liu, S. (2001). The effect of experience on adults' acquisition of a second language. Studies in Second Language Acquisition, 23(4), 527-552.
- Flege, J., Munro, M., & MacKay, I. (1995a). Factors affecting strength of perceived foreign accent in a second language. Journal of the Acoustical Society of America, 97, 3125-3134
- Flege, J., Munro, M., & MacKay, I. (1995b). Effects of age of second-language learning on the production of English consonants. Speech Communication, 16(1), 1-26.
- Flege, J., Takagi, N., & Mann, V. (1996). Lexical Familiarity and English-Language Experience Affect Japanese Adults' Perception of /r/ and /l/. The Journal of the Acoustical Society of America, 99(2), 1161-1173.
- Flege, J., Yeni-Komshian, G., & Liu, S. (1999). Age constraints on second-language acquisition. Journal of Memory and Language, 41(1), 78-104.
- Flege, J., & Wang, C. (1989). Native-language phonotactic constraints affect how well Chinese subjects perceive the word-final English /t/-/d/ contrast. Journal of Phonetics, 17, 299-315.
- Gass, S. (1984). Development of speech perception and speech production abilities in adult second language learners. Applied Psycholinguistics, 5, 51-74.
- Guion, S., Flege, J., & Loftin, J. (2000). The effect of L1 use on pronunciation in Quichua-Spanish bilinguals. Journal of Phonetics, 28(1), 27-42.
- Ingram, J., Park, S-G., & Mylne, T. (1997). Studies in cross-language speech perception. Asia Pacific Journal of Speech, Language and Hearing, 2, 1-23.
- Ioup, G., Boustagui, E., Tigi, M., & Moselle, M. (1994). Reexamining the critical period hypothesis: A case study of successful adult SLA in a naturalistic environment. Studies in Second Language Acquisition, 16, 73-98.
- Jamieson, D., & Morosan, D. (1986). Training non-native speech contrasts in adults: Acquisition of the English /ð/ - /θ/ contrast by francophones. Perception & Psychophysics, 40(4), 205-215.

- Jamieson, D., & Morosan, D. (1989). Training new, nonnative speech contrasts: A comparison of the prototype and perceptual fading techniques. Canadian Journal of Psychology, 43, 88-96.
- Jamieson, D., & Rvachew, S. (1992). Remediating speech production errors with sound identification training. Journal of Speech-Language Pathology and Audiology, 16(3), 201-210.
- Joh, J., & Lee, S. (2001). Relationships between sound perception and production in 12 phonology acquisition, Journal of the Applied Linguistics Association of Korea, 17, 127-146.
- Johnson, K. (1997). Acoustic and auditory phonetics. Madden, MA: Blackwell Publisher.
- Jusczyk, P. (1992). Developing phonological categories from the speech signal. In C. Ferguson, L. Menn, & C. Stoel-Gammon (eds.), Phonological development: Models, Research and implications (pp.17-64). Timonium, MD: York Press.
- Kagaya, R. (1974). A fiberoptic and acoustic study of the Korean stops, affricates and fricatives. Journal of Phonetics, 2, 161-180.
- Kent, R. (1992). The biology of phonological development. . In C. Ferguson, L. Menn, & C. Stoel-Gammon (eds.), Phonological development: Models, Research and implications (pp.65-90). Timonium, MD: York Press
- Kim, H. (1999). The place of articulation of Korean affricates revisited. Journal of East Asian Linguistics, 8, 313-347.
- Krashen, S. (1973). Lateralization, language learning and the critical period: Some new evidence. Language Learning, 23, 63-74.
- Kuhl, P. & Miller, J. (1978). Speech-Perception by Chinchilla - Identification Functions for Synthetic VOT Stimuli. Journal of the Acoustical Society of America, 63(3), 905-917
- Ladefoged, P. (2001). A course in Phonetics (4th ed.). Orlando, FL: Harcourt college.
- Ladefoged, P., & Maddieson, I. (1996). The Sounds of the World's Languages (pp. xxi+425). Oxford, England: Blackwell Publishers Ltd.
- Larsen-Freeman, D., & Long, M. (1991). An Introduction to Second Language Acquisition Research. New York: Longman.
- Liberman, A., & Mattingly, I. (1985). The motor theory of speech perception revisited, Cognition, 21, 1-36.
- Lisker, L., & Abramson, A. (1969). A cross-language study of voicing in initial stops: Acoustical measurements., Word, 20, 527-565.

- Llisterri, J. (1995). Relationships between Speech Production and Speech Perception in a Second Language. In K. Elenius & P. Branderud (Eds.), Proceedings of the XIIIth International Congress of Phonetic Sciences (Vol. 4, pp. 92-96). Stockholm, Sweden.
- Locke, J. (1980). The inference of speech perception in the phonologically disordered children. Part II: Some clinically novel procedure, their use, some finding. Journal of Speech and Hearing Disorders, 45, 445-468.
- Loewenthal, K., & Bull, D. (1984). Imitation of foreign sounds: what is the effect to of age. Language and Speech, 27, 95-98.
- Logan, J., & Pruitt, J. (1995). Methodological issues in training listeners to perceive non-native phonemes. In W. Strange (Ed.), Speech Perception and Linguistic Experience: Issues in Cross-Language Research (pp. 351-377). York: Timonium, MD.
- Long, M. (1990). Maturational constraints on language development. Studies in Second Language Acquisition, 12, 251-285.
- Mack, M. (1989). Consonant and Vowel Perception and Production: Early English-French Bilinguals and English Monolinguals. Perception and psychophysics, 46, 187-200.
- Major, R., & Faudree, M. (1996). Markedness universals and the acquisition of voicing contrasts by Korean speakers of English. Studies of Second Language Acquisition, 18, 69-90.
- McClaskey, C., Pisoni, D., & Carrell, T. (1983). Transfer of training of a new linguistic contrast in voicing. Perception and Psychophysics, 34(4), 323-330.
- Monninn, L., & Huntington, D. (1974). Relationship of articulatory defects to speech-sound identification. Journal of Speech and Hearing Research, 17, 352-366.
- Morley, J. (1991). The pronunciation component in teaching-English to speakers of other languages. TESOL Quarterly, 25(3), 481-520.
- Munro, M., & Derwing, T. (1995). Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. Language and Speech, 38(3), 289-306.
- Neufeld, G. (1980). On the adult's ability to acquire phonology. TESOL Quarterly, 14(3), 285-298.
- Obler, L. (1989). Exceptional second language learners. In S. Gass, C. Madden, D. Preston & L. Selinker (eds.), Variation in Second Language Acquisition: Psycholinguistic issues (pp.141-159). Philadelphia, PA: Multilingual Matters.

- Ohala, J. (1990). There is no interface between phonology and phonetics - a personal view. Journal of Phonetics, 18(2), 153-171.
- Oyama, S. (1976). Sensitive period for acquisition of a non-native phonological system. Journal of Psycholinguistic Research, 5(3), 261-283.
- Piske, T., MacKay, I., & Flege, J. (2001). Factors affecting degree of foreign accent in an L2: A review. Journal of Phonetics, 29(2), 191-215.
- Pisoni, D., Aslin, R., Perey, A., & Hennessy, B. (1982). Some effects of laboratory training on identification and discrimination of voicing contrasts in stop consonants. Journal of Experimental Psychology: Human Perception and Performance, 8, 297-314.
- Port, R., & Mitleb, F. (1983). Segmental features and implementation in acquisition of English by Arabic speakers. Journal of Phonetics, 11, 219-229.
- Purcell, E., & Suter, R. (1980). Predictors of pronunciation accuracy - a reexamination. Language Learning, 30(2), 271-287.
- Riney, T., & Flege, J. (1998). Changes over time in global foreign accent and liquid identifiability and accuracy. Studies in Second Language Acquisition, 20(2), 213-243.
- Rvachew, S. (1994). Speech perception training can facilitate sound production learning. Journal of Speech and Hearing Research, 37(2), 347-357.
- Rvachew, S., & Jamieson, D. (1995). Learning new speech contrasts: Evidence from adults learning a second language and children with speech disorders. In W. Strange (Ed.), Speech Perception and Linguistic Experience: Issues in Cross-Language Research (pp. 411-432). Timonium, MD: York Press, Inc.
- Sato, A. (1987). Limits on phonetic accuracy in foreign language speech production. In G. Ioup & S. Weinberger (eds.), Interlanguage Phonology: Acquisition of a Second Language Sound System (pp.176-203). Cambridge, MA: Newbury House.
- Schmidt, A. (1996). Cross-language identification of consonants part 1: Korean perception of English., Journal of the Acoustical Society of America, 99, 3201-3211.
- Schmidt, A., & Meyer, K. (1995). Traditional and phonological treatment for teaching English fricatives and affricates to Koreans, Journal of Speech and Hearing Research, 38, 828-838).
- Seliger, H. (1978). Implications of a multiple critical periods hypothesis for second language learning. In W. Ritchie (ed.), Second Language Acquisition Research (pp. 11-19). New York: Academic Press.

- Selinker, L. (1972). Interlanguage. IRAL, 10, 209-231.
- Sheldon, A. (1985). The relationship between production and perception of the [r]-[l] contrast in Korean adults learning English: A reply. Language Learning, 35(1), 107-113.
- Snow, C., & Hoefnagel-Hohle, A. (1977). Age differences and the pronunciation of foreign sounds. Language and Speech, 20, 357-365.
- Strange, W., & Dittmann, S. (1984). Effects of discrimination training on the perception of /r-l/ by Japanese adults learning English. Perception and Psychophysics, 36(2), 131-145.
- Tahta, S., Wood, M., & Loewenthal, K. (1981a). Age changes in the ability to replicate foreign pronunciation and intonation. Language and Speech, 24, 363-372.
- Tahta, S., Wood, M., & Loewenthal, K. (1981b). Foreign accents: factors relating to transfer of accent from the first language to a second language. Language and Speech, 24, 265-272.
- Tarone, E. (1980). Some influences on the syllable structure of interlanguage phonology. IRAL, 18, 139-152.
- Tarone, E. (1983). On the Variability of Interlanguage Systems. Applied Linguistics, 4(2), 142-164.
- Thompson, I. (1991). Foreign accents revisited: The English pronunciation of Russian immigrants. Language Learning, 41, 177-204.
- Ueyama, M., & Jun, S. (1996). Focus realization of Japanese English and Korean English intonation. University of California working papers in phonetics, 94, 110-125.
- VanPatten, B. (1996). Input processing in second language acquisition, Input Processing and Grammar Instruction (pp. 13-53). Norwood, NJ: Ablex.
- Walley, A., & Flege, J. (1999). Effect of lexical status on children's and adult's perception of native and non-native vowels. Journal of Phonetics, 27, 302-337.
- Walsh, T., & Diller, K. (1981). Neurolinguistic considerations on the optimum age for second language learning. In K. Diller (Ed.), Individual Differences and Universals in Language Learning Aptitude (pp. 3-29). Rowley, MA: Newbury House.
- Waring, R. (2004). The Word Frequency List. <http://www1.harenet.ne.jp/~waring/vocab/wordlists/vocfreq.html>. (date last accessed July 16, 2004)
- Weiner, P. (1967). Auditory discrimination and articulation. Journal of Speech and Hearing Disorders, 32, 19-38.

- Werker, J., & Logan, J. (1985). Cross-language evidence for three factors in speech perception. Perception and Psychophysics, 37, 35-44.
- Werker, J., & Tees, R. (1984a). Cross-language speech-perception - evidence for perceptual reorganization during the 1st year of life. Infant Behavior & Development, 7(1), 49-63.
- Werker, J., & Tees, R. (1984b). Phonemic and phonetic factors in adult cross-language speech-perception. Journal of the Acoustical Society of America, 75(6), 1866-1878.
- Wode, H. (1996). Speech perception and second language phonological acquisition. In P. Jordens & J. Lalleman (Eds.), Investigating Second Language Acquisition (pp. 321-353). New York: Mouton de Gruyter.
- Yamada, R. (1995). Age and acquisition of second language speech sounds: Perception of American English /r/ and l/ by native speakers of Japanese. In W. Strange (Ed.), Speech Perception And Linguistic Experience: Issues In Cross-Language Research, (pp. 305-320), Timonium, MD: York Press, Inc.
- Yoo, H. (1996). A constraints-based analysis of Korean loanwords. In Y. Kim (Ed.), Studies in Phonetics, Phonology and Morphology (Vol. 2, pp. 147-167). Seoul, Korea: The Phonology-Morphology Circle of Korea.

BIOGRAPHICAL SKETCH

Sang-Hee Yeon, the daughter of Woon-Ja Lee and Hung-Kyu Yeon, was born in Seoul, Korea. She went to HanKuk University of Foreign Studies and earned her B.A. in English Literature and Linguistics. She completed an M.A. in TESOL (Teaching English to Speakers of Other Languages) at the University of Northern Iowa in 1999. At the University of Florida, she started a doctoral study in linguistics in 1999. During her study, she taught introductory linguistics classes from 2000 to 2004. She was awarded a Gibson Dissertation Fellowship in Spring 2003.