The Effects of Two Methods on Training EFL University Students in Taiwan to Identify
Three Non-Native Phonemic Contrasts

Dissertation

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy
in the Graduate School of The Ohio State University

By

Yao-Feng Huang, M.A.

Graduate Program in Foreign and Second Language Education

The Ohio State University
2013

Dissertation Committee:

Leslie C. Moore, Advisor

Marjorie K. M. Chan

Laura Justice
Abstract

The present study investigated and compared the effects of two methods on training 71 EFL university students in Taiwan to identify three non-native phonemic contrasts. The study also investigated whether the perceptual training effect could be generalized to new tokens, new talkers, and untrained final stop contrasts as well as whether perceptual training facilitated production. Moreover, to understand the participants’ language background and language attitudes toward accents, a survey was conducted.

Eighty-eight EFL university students in Taiwan were randomly assigned to one of two experimental groups, identification (ID) (9 males, 20 females) or same/different (SD) (9 males, 20 females), or to a control group (10 males, 20 females). After attrition, the total number of participants who finished the pretest and posttest was 71, including the identification (ID) group (n = 24; 7 males, 17 females), the same/different (SD) group (n = 23; 9 males, 14 females), and the control group (n = 24; 8 males, 16 females). The experimental groups received different training in identifying the English /i/-/ɪ/, /ɛ/-/æ/, and the word-final /t/-/d/ contrasts fifty minutes per week for eight weeks in a computer lab. A pretest was administered before the training and a posttest and a generalization test were conducted after the training. In addition, a follow-up posttest was held three months after the completion of the training to assess retention.

The results indicate a significant positive training effect for the two training methods on identifying the /i/-/ɪ/ contrast. With regard to the /ɛ/-/æ/ contrast, only the ID training
showed a significant training effect. As for the word-final /t/-/d/ contrast, although both trained groups showed gains, the training effects of the two training methods were not significant. In contrast, the control group achieved no gains at all.

With respect to the superiority of the training methods, although the ID group outperformed the SD group in identifying all the three contrasts, the mean difference between the two trained groups was not statistically significant.

With regard to generalization, the effect of perceptual training was generalized to certain contrasts in new tokens, new talkers, and other final stop contrasts. As for retention of the training effect, the results of a follow-up posttest showed that the training effect appeared retained for certain contrasts in new tokens, new talkers, and other final stop contrasts three months after the completion of the training. With respect to the relationship between perception and production, the results showed that the perceptual training did not facilitate production.

The analyses of performance of participants with different L1 backgrounds showed that native speakers of Mandarin outperformed native speakers of Mandarin-Taiwanese on all three contrasts at posttest, while the mean difference was significant only in the /ɪ/-/ɪ/ contrast.

The results of the language attitudes survey indicating that an overwhelming 97% of the participants would like to have a native-like accent and 75% of the participants agreed that EFL teachers should have a native-like accent suggest that non-native EFL teachers should reduce their foreign accents for the sake of intelligibility in class and as a role model for their students. The results of the survey also show that English is regarded as symbolic capital in Taiwan.
With regard to difficulty, the duration manipulated /i/-/ɪ/ contrast posed the most difficulty for the participants to identify. In addition, the /ɛ/-/æ/ contrast proved to be the most resistant to differentiate due to a high degree of spectral overlap of the two vowels. With respect to production, the participants’ production performance appeared to be related to specific phonetic environments.

On the whole, the results show a consistent trend that the ID training method is superior to the SD training method although the mean difference between the two trained groups was not statistically significant. However, the large effect sizes of both trained groups indicate that either of the training methods is effective for perceptual training for ESL/EFL students.

Taken together, the findings of the present study provide support for the efficacy of high talker and duration variability in perceptual training. The finding that the trained groups did not improve their oral production also indicates the necessity of production training.
Dedication

Dedicated to my dear wife, Ya-Chun, and our two adorable boys, Henry and Ryan.
Acknowledgements

First of all, I’d like to thank Dr. Moore, Dr. Chan, and Dr. Justice for their professional feedback on my dissertation and support. My deepest gratitude goes to Dr. Moore, who also provided me with substantial assistance and advice.

Second, I’d like to thank Dr. Flege for his suggestions and detailed explanation of his previous study. I learned a lot from our correspondence. I’d also like to thank Dr. Chia-Ju Chen and Dr. Ming-Long Wu for their help on statistical analyses. In addition, I’d like to thank Dr. Yuang-Shan Chuang for his professional opinion.

Third, I’d like to thank the students who participated in this study.

Fourth, I’d like to thank Cambridge, Longman, and Merriam-Webster for the permission to use their dictionary sound files for this study. I’d also like to thank the native talkers who helped me record the stimuli. You are friends indeed!

Finally, I’d like thank my family for their support, especially my mother-in-law, who gave me a lot of encouragement but unfortunately passed away in Taiwan during my study at OSU.
Vita

1990…………………………………………..B.A. Foreign and Literature, Tunghai University
1995…………………………………………..M.A. TESOL, New York University

Publication

Fields of Study

Major Field: Foreign and Second Language Education
Table of Contents

Abstract ........................................................................................................................................... ii
Dedication ......................................................................................................................................... v
Acknowledgements ......................................................................................................................... vi
Vita ...................................................................................................................................................... vii
List of Tables ................................................................................................................................... xvi
List of Figures ................................................................................................................................... xix

Chapter 1: Introduction

1.1. Background .......................................................................................................................... 1
1.2. Statement of the Problem ....................................................................................................... 3
1.3. Gap statement ......................................................................................................................... 7
1.4. Purposes of the Study ............................................................................................................ 10
1.5. Research Questions ............................................................................................................... 11
1.6. Research Hypotheses ........................................................................................................... 12
1.7. Significance of the Study ....................................................................................................... 13
1.8. Assumptions of the Study ..................................................................................................... 14
1.9. Definition of Key Terms ........................................................................................................ 14
Chapter 2: Review of the Literature

2.1. Introduction.................................................................................................................. 16

2.2. Theoretical Underpinnings in the Acquisition of L2 Phonology..................... 17
   2.2.1. Speech Production Theories.................................................................................. 18
      2.2.1.1. Contrastive Analysis Hypothesis (CAH) ...................................................... 18
      2.2.1.2. Markedness Differential Hypothesis (MDH)............................................. 19
      2.2.1.3. Structural Conformity Hypothesis (SCH) ............................................... 19
      2.2.1.4. Ontogeny Phylogeny Model (OPM) ...................................................... 20
   2.2.2. Speech Perceptual Theories.................................................................................. 21
      2.2.2.1. Perceptual Assimilation Model (PAM) ...................................................... 21
      2.2.2.2. Native Language Magnet (NLM) Model.................................................. 22
      2.2.2.3. Speech Learning Model (SLM) ............................................................... 22
   2.3 Foreign Accents Stigma and Identity in SLA............................................................ 25
      2.3.1. NNES Teacher’ Accents and Issues of Identity.............................................. 27
      2.3.2. NNES Learners’ Accents and Issues of Identity.......................................... 32
   2.4. Factors in L2 Phonology Acquisition................................................................. 38
      2.4.1. Age effects....................................................................................................... 39
         2.4.1.1. Debate on the Critical Period................................................................. 39
         2.4.1.2. Studies Related to Age Effects............................................................... 40
      2.4.2. Cross-linguistic Factors.................................................................................... 46
      2.4.3. Social factors.................................................................................................... 50
         2.4.3.1. Length of Residence ............................................................................. 50
2.4.3.2. Exposure to the L2 ................................................................. 54

2.4.3.3. Social Network ................................................................. 58

2.4.3.4. Amount of L1 Use .......................................................... 60

2.4.3.5. Negative Ethnic Stereotyping ........................................... 63

2.5. Empirical Research on Chinese Speakers’ Perception and Production of English Phonetic Segments ................................................................. 66


2.7. Chapter Summary .................................................................. 95

Chapter 3: Methodology

3.1. Introduction ........................................................................ 100

3.2. The Sociolinguistic Background in Taiwan .............................. 101

3.3. Comparison of Phonologies in English, Mandarin, Taiwanese, and Hakka.... 102

3.3.1. English ............................................................................. 102

3.3.2. Mandarin ........................................................................... 104

3.3.3. Taiwanese ......................................................................... 105

3.3.4. Hakka ................................................................................ 105

3.4. Identification Training vs. Categorical Same/Different Training ....... 106

3.4.1. Identification Training ....................................................... 106

3.4.2. Categorical Same/Different Training .................................... 107

3.4.3. Debate on the Effectiveness of Identification Training vs. Categorical

Same/Different Training .............................................................. 107

3.5. Research Questions ................................................................ 108
3.6. Research Hypotheses ................................................................. 109
3.7. Research Design ................................................................. 110
3.8. Research Site ....................................................................... 111
3.9. Participants ........................................................................ 112
3.10. Stimuli ............................................................................... 114
3.11. Measurement ..................................................................... 119
      3.11.1. Perception ................................................................. 119
      3.11.2. Production ................................................................. 119
3.12. Procedures ....................................................................... 119
      3.12.1. Introduction and Consent ........................................ 119
      3.12.2. Pretest ..................................................................... 120
            3.12.2.1. Perception ......................................................... 120
            3.12.2.2. Production .......................................................... 121
      3.12.3. Training .................................................................. 121
            3.12.3.1. Identification Training .................................... 122
            3.12.3.2. Categorical Same/Different Training .............. 123
      3.12.4. Posttest ................................................................... 124
3.13. Data Collection ............................................................... 125
      3.13.1. Survey ................................................................... 125
      3.13.2. Pretest and Posttests ............................................... 125
3.14. Data Analysis ................................................................. 125
Chapter 4: Results

4.1. Research Question 1: Training Effects .............................................................. 129
  4.1.1. /i/-/ɪ/ Contrast ......................................................................................... 129
  4.1.2. /e/-/æ/ Contrast ....................................................................................... 131
  4.1.3. Word-Final /t/-/d/ Contrast ..................................................................... 133
    4.1.3.1. ID Group vs. Control Group ............................................................... 133
    4.1.3.2. SD Group vs. Control Group .............................................................. 135
    4.1.3.3. ID Group vs. SD Group ..................................................................... 137
  4.1.4. Overall Comparison .................................................................................. 139

4.2. Research Question 2: Comparison of Two Training Methods ................. 141

4.3. Research Question 3: Generalization to New Tokens and New Talkers ....... 142
  4.3.1. Generalization to New Tokens.................................................................. 142
  4.3.2. Generalization to New Talkers ................................................................ 144

4.4. Research Question 4: Generalization to Other Final Stops ....................... 146

4.5. Research Question 5: Retention of Training Effects ................................. 148
  4.5.1. ID Group Retention of Training Effects .................................................. 150
    4.5.1.1. ID Retention of Training Effects at Posttest ..................................... 150
    4.5.1.2. ID Retention of Training Effects on Generalization to New Tokens .......................................................... 152
    4.5.1.3. ID Retention of Training Effects on Generalization to New Talkers ........................................................................................................................................ 152
    4.5.1.4. ID Retention of Training Effects on Generalization to Other Final

Stop Contrasts

4.5.2. SD Group Retention of Training Effects

4.5.2.1. SD Retention of Training Effects at Posttest

4.5.2.2. SD Retention of Training Effects on Generalization to New Tokens

4.5.2.3. SD Retention of Training Effects on Generalization to New Talkers

4.5.2.4. SD Retention of Training Effects on Generalization to Other Final Stop Contrasts

4.6. Research Question 6: Relationship between Perceptual Training and Production

4.7. Research Question 7: Performance of Participants of Different L1 Backgrounds

4.8. Research Question 8: Participants’ Language Attitudes toward Accents

4.9. Participants’ Difficulties in Perception and Production

4.9.1. Perception Difficulties

4.9.2. Production Difficulties

4.10. Chapter Summary

Chapter 5: Discussion and Conclusion

5.1. Discussion of Major Findings

5.1.1. Training Effects and Comparison of Two Training Methods
5.1.2. Generalization to New Tokens, New Talkers, and Other Final Stops... 177
5.1.3. Retention of Training Effects...................................................... 178
5.1.4. Relationship between Perceptual Training and Production............. 180
5.1.5. Performance of Participants of Different L1 Backgrounds.............. 181
5.1.6. Difficulties in Perception and Production...................................... 183
  5.1.6.1. Difficulties in Perception...................................................... 183
  5.1.6.2. Difficulties in Production..................................................... 186
5.1.7. Participants’ Language Attitudes toward Accents.......................... 187
5.2. Pedagogical Implications.................................................................. 187
  5.2.1. Integrating English Pronunciation Instruction into Teacher Preparation
         Curriculum and In-Service Teacher Training.................................. 190
  5.2.2. Research Informed Pronunciation Instruction............................... 190
  5.2.3. Policy Making............................................................................... 191
5.3. Limitations and Future Research Directions...................................... 192
5.4. Conclusion ....................................................................................... 195

References.............................................................................................. 197

Appendix A: Language Background and Attitude Questionnaire.................. 215
Appendix B: Permission to Use Real-person Pronunciation Sound Files in Cambridge
  Advanced Learner’s Dictionary.......................................................... 217
Appendix C: Permission to Use Real-person Pronunciation Sound Files in Longman
  Dictionary of Contemporary English.................................................. 219
Appendix D: Permission to Use Real-person Pronunciation Sound Files in
Appendix E: Word List for NES Talkers ............................................................... 222
Appendix F: Training Stimuli ........................................................................... 227
Appendix G: Pre/Post Tests Stimuli ................................................................. 228
Appendix H: Generalization Test Stimuli ....................................................... 230
Appendix I: Pretest Instructions ..................................................................... 232
Appendix J: Production Word List .................................................................. 233
Appendix K: ID Group Training Instructions ................................................ 234
Appendix L: ID Group Training Feedback ..................................................... 235
Appendix M: SD Group Training Instructions ................................................ 239
Appendix N: SD Group Training Feedback .................................................... 240
Appendix O: Generalization Test Instructions .............................................. 242
List of Tables

Table 3.1 Background Information of Control, ID, and SD Groups.......................... 113
Table 4.1 ANCOVA Results and Descriptive Statistics for /i/-/ɪ/ Contrast by Group and Posttest Scores................................................................. 130
Table 4.2 Multiple Comparisons and Mean Differences in /i/-/ɪ/ Contrast by Group and Controlling for Pretest Scores........................................................... 130
Table 4.3 ANCOVA Results and Descriptive Statistics for /ɛ/-/æ/ Contrast by Group and Posttest Scores..................................................................................... 132
Table 4.4 Multiple Comparisons and Mean Differences in /ɛ/-/æ/ Contrast by Group and Controlling for Pretest Scores.............................................................. 132
Table 4.5 Johnson Neyman Technique Results for ID Group and Control Group........ 134
Table 4.6 Johnson Neyman Technique Interaction and Points of Significance for ID Group and Control Group................................................................. 134
Table 4.7 Johnson Neyman Technique Results for SD Group and Control Group....... 136
Table 4.8 Johnson Neyman Technique Interaction and Points of Significance for SD Group and Control Group................................................................. 136
Table 4.9 Johnson Neyman Technique Results for ID Group and SD Group.............. 138
Table 4.10 Johnson Neyman Technique Interaction and Points of Significance for ID Group and SD Group................................................................. 138
Table 4.11 ANCOVA Results and Descriptive Statistics for All Contrasts by Group and Posttest Scores................................................................. 140
Table 4.12 Multiple Comparisons and Mean Differences in All Contrasts by Group and Controlling for Pretest Scores........................................... 140
Table 4.13 Mean Percentage Correct Identification Scores from Posttest to Three-month Follow-up by ID Group................................................................. 149
Table 4.14 Mean Percentage Correct Identification Scores from Posttest to Three-month Follow-up by SD Group................................................................. 150
Table 4.15 ANCOVA Results and Descriptive Statistics for Production by Group and Posttest Scores........................................................................... 158
Table 4.16 Multiple Comparisons and Mean Differences in Production by Group and Controlling for Pretest Scores......................................................... 158
Table 4.17 ANCOVA Results and Descriptive Statistics for /i/-/ɪ/ Contrast by L1 and Posttest Scores........................................................................... 160
Table 4.18 Pairwise Comparison and Mean Difference in /i/-/ɪ/ Contrast by L1 and Controlling for Pretest Scores................................................................. 161
Table 4.19 ANCOVA Results and Descriptive Statistics for /ɛ/-/æ/ Contrast by L1 and Posttest Scores........................................................................... 162
Table 4.20 Pairwise Comparison and Mean Difference in /ɛ/-/æ/ Contrast by L1 and Controlling for Pretest Scores................................................................. 162
Table 4.21 ANCOVA Results and Descriptive Statistics for Word-Final /t/-/d/ Contrast by L1 and Posttest Scores........................................................................... 163
Table 4.22 Pairwise Comparison and Mean Difference in Word-Final /t/-/d/ Contrast by L1 and Controlling for Pretest Scores.................................................................163
List of Figures

Figure 3.1 The auditory qualities of /i/-/ɪ/ and /ɛ/-/æ/ contrasts of Standard American Newscaster ................................................................. 103

Figure 3.2 The Waveforms of the Words mat and mad ............................. 104

Figure 4.1 Relationship between Posttest Scores and Pretest Scores by ID and Control Groups on Word-Final /t/-/d/ Contrast ................................................................. 135

Figure 4.2 Relationship between Posttest Scores and Pretest Scores by SD and Control Groups on Word-Final /t/-/d/ Contrast ................................................................. 137

Figure 4.3 Relationship between Posttest Scores and Pretest Scores by ID and SD Groups on Word-Final /t/-/d/ Contrast ................................................................. 139

Figure 4.4 Mean Percentage Correct Identification Scores for New Tokens on the Generalization Test by the Trained and the Control Groups ...................... 143

Figure 4.5 Mean Percentage Correct Identification Scores for New Talkers on the Generalization Test by the Trained and the Control Groups ...................... 145

Figure 4.6 Mean Percentage Correct Identification Scores for /k/-/ɡ/ and /p/-/b/ Endings on the Generalization Test by the Trained and the Control Groups ...................... 147

Figure 4.7 Mean Percentage Correct Identification Scores for the 3 Contrasts on the Pretest and Posttest by All Participants ......................................................... 167

Figure 4.8 Mean Production Intelligibility Scores on the Pretest by All Participants .... 169

Figure 4.9 Mean Production Intelligibility Scores on the Posttest by All Participants .. 170
CHAPTER 1: INTRODUCTION

1.1. Background

English is taught in school as a foreign language in Taiwan, which is in the Expanding Circle (Kachru, 1992; Kachru & Nelson, 1996), where English is not used for everyday communication. The main foci of English language teaching (ELT) in Taiwan have long been on reading and partly on writing to cope with national exams. Thus, the grammar-translation approach has been widely used by English-as-a-foreign-language (EFL) teachers. Since English listening and speaking are excluded on national exams, these two skills have been neglected on purpose in the test-oriented teaching environment in Taiwan. Students in Taiwan overall treat English as a difficult subject and seldom speak English in English classes. Even EFL teachers from Taiwan enrolled in a pronunciation pedagogy course in a TESL program in the U.S felt stigmatized due to lack of fluency and foreign accent and felt that their identity as a legitimate speaker and pronunciation teacher of English was damaged (Golombek & Jordan, 2005).

Research has found that degree of perceived foreign accent in a second language (L2) is related to exposure to L2 (e.g., Flege, 1993; Bohn & Flege, 1996; Munro & Derwing, 2008) and the amount of first language (L1) use (e.g. Flege et al, 1997; Piske et al., 2001). A foreign accent usually results from low perceptual ability of the target language, especially the inability to differentiate non-native phonemic contrasts. Although some researchers asserted that we should respect L2 learners’ wish to retain their accents as a
marker of their desired identities (e.g., Ur, 1996; Jenkins, 1998, 2000; D. Liu, 1999), most researchers contended that L2 learners, especially non-native ESL/EFL teachers, should reduce their foreign accent for the sake of intelligibility and as a role model for their students (e.g., Lado, 1964; Medgyes, 1999; Arva & Medgyes, 2000; Demirezen, 2007). The norm of ELT and the issue of foreign accents are highly political and will remain in debate for a long time. According to the Speech Learning Model (Flege, 1995b), the degree of phonetic similarity determines whether L2 phonetic segments will be assimilated into existing L1 phonetic categories or whether separate L2 phonetic categories will be established. Therefore, it is important to train ESL/EFL students to identify non-native phonemic contrasts. The question is how EFL teachers can achieve the goal in a country where English is rarely used outside the classroom. Flege (2009) emphasized the importance of native speaker input and quality of input. He noted that usually the first English input EFL learners receive is the input of accented EFL teachers in their own countries, which suggests that EFL teachers should utilize native speaker input in perceptual and production training. Thus, one of the top priorities in an EFL setting is to provide an effective method of perceptual training utilizing recordings of various native speakers of English (in this case, Standard American English).

Although segmentals have been the focus of pronunciation teaching (Derwing, 2008), some researchers think segmentals overemphasized in pronunciation teaching (e.g., Brown, 1995). Recently the focus of pronunciation teaching has shifted to suprasegmental level (Celce-Murcia et al., 2010). However, as Derwing (2008) pointed out, the inability to produce certain segments is likely to lead to a communication breakdown. In the same vein, Saito (2007) contended, “It can be readily argued that one
must understand the segmental in order to be able to understand the suprasegmental fully. Considering communicative significance, phoneme awareness should be prioritized” (p. 20). He asserted that if speakers mispronounce the suprasegmental parts, listeners can still guess the content of the message based on the intelligible segmental parts. On the other hand, if speakers mispronounce certain segments, communication is likely to be hindered. In the same vein, the inability to identify segments in a second language will for sure hamper listening comprehension and result in miscommunication. Although English pronunciation instruction in Taiwan focuses on segments, judging by the low English listening and speaking proficiency of Taiwanese students, it is safe to say that the English listening and speaking training methods in Taiwan are not effective enough and are not informed by research. Therefore, there is an urgent need of effective training methods to help EFL learners in Taiwan.

1.2. Statement of the Problem

The acquisition of the second language phonetic system is the first step to achieve communicative competence. The lack of distinction between L2 segments, as Munro (2008) pointed out, is realized in both perception difficulty and a recognizable spoken accent, which might lead to a communication breakdown. As we can see in previous studies (e.g., Golombek & Jordan, 2005; Derwing & Rossiter, 2002), speaking with a foreign accent can also be detrimental to identity construction for non-native-English-speaking ESL/EFL teachers and students alike.

To most EFL students in Taiwan, the classroom is the only place to acquire English since they have few chances to listen to and speak English outside the classroom. With
little exposure to authentic aural English input and very few chances to practice English, most Taiwanese students seldom feel the need to communicate in English and thus have low proficiency in English, which is reflected in their proficiency test scores. On the 2010 Internet-Based TOEFL exam (TOEFL iBT), Taiwanese test takers’ mean score was 76 out of a maximum score of 120, ranked 18\textsuperscript{th} in Asia. On the listening section of the TOEFL iBT, Taiwanese test takers scored a mean score of 18 out of a maximum score of 30, which was lower than their scores in the reading, speaking, and writing sections, and lower than the that of test takers in developing nations in the Expanding Circle such as Bangladesh (20) and Bhutan (19). In comparison with South Korean TOEFL test takers, whose overall mean score was 81 with 20 on the listening section, Taiwanese TOEFL test takers’ average overall English proficiency and listening ability were inferior to those of their South Korean counterparts (Educational Testing Service, 2011). It is noteworthy that most Taiwanese TOEFL test takers are those who plan to study in an English speaking country and are considered highly proficient in English in Taiwan. Logically the average Taiwanese students should have much lower English proficiency. As a matter of fact, a shortage of high English proficient professionals is hurting Taiwan’s international competitiveness (Lin, 2005) since English has become the dominant global language.

As shown above, Taiwanese TOEFL test takers’ mean listening score was the lowest among the mean scores of the four skills. Therefore, how to improve Taiwanese students’ English listening ability has become a formidable task. The first step is of course to improve the perception of English phonemic contrasts, especially those that do not exist in Taiwanese students’ first languages and pose difficulty for Taiwanese EFL learners.

Listening skills have been less emphasized in Taiwan. In most universities, students
are required to take Listening and Speaking, a one-credit course, during their freshman year and have no extra English listening training during their four years in college. Moreover, as Derwing, Thomson, and Munro (2006) pointed out, many ESL teachers in Canada feel reluctant teaching pronunciation, usually due to a lack of training. Similarly, a lack of teacher training in pronunciation instruction has been a prevalent phenomenon in Taiwan. As a result, most EFL teachers in Taiwan can only rely on teacher’s manuals of textbooks on listening and speaking. Fortunately, recently more and more teachers’ colleges in Taiwan have incorporated English pronunciation pedagogy in the curriculum for would-be English teachers, although the course is listed as one of required electives (e.g., National Changhua University of Education; National Kaohsiung Normal University). Furthermore, although empirical research is crucial to inform pronunciation teaching, Derwing and Munro (2005) noted that pronunciation research has been marginalized in applied linguistics even with the recent resurgence of interest in the acquisition of L2 phonology, which often leaves teachers to rely on their own experiences and intuitions rather than empirical evidence.

Nearly all studies on perception and production of L2 segments have indicated that training learners to perceive and produce English segments can be effective. As many researchers have noted, learners can benefit from focused pronunciation instruction (e.g., Derwing et al., 2006; Couper, 2006; Derwing, 2008). However, the effect of perceptual training on production is not consistent among studies. Wang’s (2002) study suggests that effect of training on identifying non-native phonemic contrasts is not transferable to production. In contrast, Rochet (1995) and Bradlow et al. (1997) suggest that the effect of perceptual training can be transferred to production. Hence, the present study aimed to
explore further to determine if the effect of training on identifying non-native phonemic contrasts is transferable to production. Flege (2009) noted that accented EFL teachers usually reinforce their students’ foreign accents. Therefore, one of the first priorities of pronunciation teaching in Taiwan is to introduce effective teaching methods to train EFL teachers to identify English segments, especially those that do not exist in their first languages and would pose difficulty for them. As a result, they can incorporate the most effective training method into their teaching practices to help their students overcome the obstacles in perceiving English segments.

There are many factors that influence second language phonology acquisition. According to Ioup’s (2008) extensive review of the literature, the results of most studies have indicated that native-like L2 phonology is mostly found with very early L2 learners and the likelihood diminishes as the age increases. However, the finding that some of the late learners were found to perceive English vowels accurately in Flege and Mackay (2004) suggests that late learners can still establish new phonetic categories in an L2. As Flege (1999) argues, decline in L2 pronunciation accuracy with age is due to reinforced establishment of L1 phonetic categories instead of loss of ability to perceive and produce new phonetic categories in an L2, which can be supported by the fact that adult learners can benefit from focused pronunciation instruction (e.g., Derwing, Thomson, & Munro, 2006; Couper, 2006; Derwing, 2008). In other words, with effective training, adult EFL learners in Taiwan can still learn to perceive segments that do not exist in their languages.

Aside from age effects, L1 transfer is probably the most important factor that affects L2 phonology acquisition. According to Flege’s (1995b) Speech Learning Model, a learner’s L2 phonological acquisition is influenced by the interaction of his L1 and L2.
Birdsong and Molis’ (2001) study also shows that L1 influence may play an important role on L2 phonological attainment. In the same manner, the Ontogeny Phylogeny Model (Major, 2001) assumes that the role of L1 is much greater than universals for similar phenomena. The Contrastive Analysis Hypothesis (Lado, 1957) even attributes all L2 errors to L1 transfer although this argument has been disapproved by several current researchers. Although there are other factors that can also contribute to a foreign accent and failure to perceive L2 segments accurately, oftentimes cross-linguistic influence plays a major role (Odlin, 2003).

Taiwan is a multilingual nation. According to Huang (2000), Mandarin Chinese, the official language, is spoken by nearly 90% of the population. However, the majority of the population (73%) is bilingual or multilingual and speaks Taiwanese as a first language. Despite the fact that most people in Taiwan might speak first languages other than Mandarin, most researchers (e.g., Decamp, 1972; Flege, Munro, & Skelton, 1992; Paolillo, 1995; Rau, Chang, & Tarone, 2009) did not take the possibility of different L1 phonotactic constraints into account when they studied Taiwanese participants. Since speakers from different language backgrounds may have different L1 transfer (e.g., Flege, 1989; Flege & Wang, 1989; Flege et al, 1992; Flege & Liu, 2001), it is imperative to verify participants’ L1 backgrounds before examining cross-linguistic influence.

1.3. Gap Statement

Previous studies have shown that speakers of different Chinese languages have difficulty identifying English final stop consonants (e.g., Flege, 1989; Flege & Wang, 1989; Flege et al, 1992; Flege & Liu, 2001). In fact, the effect of stop voicing on duration
of preceding vowels is larger in English than in other languages in the world (Flege, 1993). Thus, native English speakers rely on such acoustic cues in identifying final stop contrasts. Moreover, in conversational English, the word-final /t/-/d/ words are often unreleased. Little research has been conducted on training Taiwanese EFL students to identify the word-final /t/-/d/ contrast (e.g., Flege, 1989). Since there is no final /t/ or /d/ in Mandarin and no final /d/ in Taiwanese or Hakka and most Chinese cannot discriminate and produce the /t/-/d/ contrast correctly (e.g., Flege & Wang, 1989; Flege, 1995a), it is worthwhile to find an effective method for training Taiwanese EFL students to identify the contrast to contribute to the literature.

The debate on which perceptual training method, identification (ID) or same/different (SD), is more effective is still not settled. The effectiveness of the identification training in Flege and Wang (1989) motivated Flege (1995a) to hypothesize that identification training method might be more effective than same/different training in perceiving the word-final English /t/-/d/ contrast. However, Flege found no significant difference between the two training methods. Furthermore, Flege’s study was conducted in an ESL setting instead of an EFL setting, which is more in need of effective methods for perceptual training. Since Flege’s (1995a) comparison of the two perceptual training methods, there have been only two perceptual training studies (Wayland & Li, 2005, 2008) in the literature that focused on comparing the effects of the two interventions on training native English speakers and native Mandarin speakers to discriminate the mid- vs. low-tone contrast in Thai. The results also showed that the two methods were equally effective. Since ELT in Taiwan is in need of effective methods to train English phonemic identification, it is worthwhile to replicate Flege’s study with a few modifications to
investigate which intervention is a better solution or whether they are equally effective in an EFL setting.

In the past decade, there have been several studies on perception and production of English vowel contrasts (e.g., Wang, 2002; Lee, 2009). Wang (2002) focused on training Mandarin and Cantonese speakers to identify the English /i/-/ɪ/, /u/-/ʊ/, and /ɛ/-/æ/ contrasts in an ESL setting in Canada. Lee (2009) used duration manipulation to train pre-service Korean EFL teachers to identify the English /i/-/ɪ/ and /u/-/ʊ/ contrasts. The results showed that the /i/-/ɪ/ contrast appeared to pose most difficulty for Chinese ESL learners and Korean EFL learners alike. Wang and Lee used both natural and unnatural synthesized stimuli to sensitize learners to spectral differences between the /i/-/ɪ/ contrast in addition to temporal cues in identifying the structurally similar English tense and lax vowel pair. The high talker variability in Wang (2002) appeared to successfully shift the Chinese participants’ attention on temporal cues to spectral cues. Since synthesized tokens are not produced by real people and sound unnatural and synthetic stimuli do not represent the full range of acoustic properties of specific phonetic categories (Flege, 1995a), natural tokens with talker and duration variability maximized were used in the present study to sensitize learners to spectral differences in the /i/-/ɪ/ contrast. It was intended to contribute to the literature to test if natural tokens with talker and duration variability maximized would be sufficient for effective perceptual training.

All of Flege’s (1995b) Speech Learning Model, Kuhl’s (1993) Native Language Magnet, and Best’s (1994) Perceptual Assimilation Model take into account the degree of similarity between L1 and L2 phonological systems as well as L2 learners’ experience, which determines whether L2 phonetic segments will be assimilated into existing L1 or
not. Since there are no high-front /ɨ/-/ı/ and mid vs. low-front /ɛ/-/æ/ contrasts in Mandarin, Taiwanese, and Hakka and they are highly functionally loaded (Brown, 1988) and most untrained Chinese ESL learners cannot perceive the English vowel duration difference preceding /t/-/d/ accurately (e.g., Flege & Wang, 1989; Flege, 1995a), it is imperative to investigate if perception of non-native phonemic contrasts differs with learners of different L1s. As mentioned above, few studies took the possibility of different L1 phonotactic constraints into account when Taiwanese participants were involved. Research on how native Mandarin-Taiwanese speakers and native Mandarin speakers process English syllables is scarce in the literature. In the present study, the training effects between native Mandarin-Taiwanese speakers and native Mandarin speakers were investigated and discussed. Therefore, the present study aimed to fill the gap in the literature regarding perception of three non-native phonemic contrasts among Taiwanese EFL learners of different L1 backgrounds in an EFL setting.

1.4. Purposes of the Study

The present study investigated and compared the effects of two methods on training 71 EFL university students in Taiwan to identify three non-native phonemic contrasts. The purposes of the study are: (1) to investigate if the perceptual training methods are effective in identifying three non-native phonemic contrasts, (2) to compare the effects of two perceptual training methods in identifying three non-native phonemic contrasts, (3) to investigate if effect of perceptual training on identifying three non-native phonemic contrasts be generalized to untrained tokens produced by familiar talkers and familiar tokens produced by new talkers, (4) to investigate if the perceptual training in identifying
the word-final /t/-/d/ contrast could be generalized to the other final stop contrasts, i.e., /k/-/g/ and /p/-/b/, (5) to investigate if participants can retain information pertinent to the specific stimuli on which they received feedback training three months after the completion of the training, (6) to investigate if the effect of perceptual training can be generalized to production of the three non-native phonemic contrasts, (7) to investigate if there is a relationship between participants’ identification performance and their first language backgrounds, and (8) to understand the language attitudes of participants toward speaking Standard English and EFL teachers’ accents. Accordingly, eight research questions were formulated as follows.

1.5. Research Questions

1. Are the perceptual training methods effective in identifying three non-native phonemic contrasts?

2. Which training method, ID or SD, will be more effective for training EFL students in Taiwan in identifying three non-native phonemic contrasts?

3. Can the effect of perceptual training on identifying three non-native phonemic contrasts be generalized to new tokens produced by familiar talkers and familiar tokens produced by new talkers?

4. Can the effect of perceptual training on identifying the word-final /t/-/d/ contrast be generalized to other final stop contrasts?

5. Can the effect of the training methods persist three months after the completion of the training?

6. Can the perceptual training in identifying three non-native phonemic contrasts
facilitate production of the non-native phonemic contrasts?

7. Will the performance of the participants in the trained groups vary due to different L1 backgrounds?

8. What are the language attitudes of the participants toward speaking Standard English and EFL teachers’ accents?

**1.6. Research Hypotheses**

Based on the research questions, seven research hypotheses were proposed as follows:

1. Judging by the results of previous studies that perceptual training in identifying English segments can be effective (e.g., Flege & Wang, 1989; Lively, Logan, & Pisoni, 1993, 1994; Flege, 1995a; Bradlow, Akahane-Yamada, Pisoni, & Tohkura, 1999; Wang, 2002; Lee, 2009), I predicted that the perceptual training would be effective.

2. Concurring with Flege (1995a), I predicted that identification training would be more effective than same/different training in that the ID group in Flege (1995a) felt that they benefited more from the training and enjoyed the training more than did the SD group.

3. Judging by the results of previous studies, I expected the perceptual training in identifying three non-native phonemic contrasts to be generalized to new tokens produced by familiar talkers and familiar tokens produced by new talkers.

4. I predicted that the effect of perceptual training on identifying the word-final /t/-/d/ contrast would be generalized to other final stop contrasts.
5. Judging by the results of previous studies that trainees retained the effect of perceptual training (e.g., Lively, Logan, & Pisoni, 1994; Flege, 1995a; Bradlow, Akahane-Yamada, Pisoni, & Tohkura, 1999; Wang, 2002), I predicted that the effects of the training methods would persist three months after the completion of the training.

6. I predicted that the perceptual training in identifying three non-native phonemic contrasts could facilitate production of the non-native phonemic contrasts.

7. I hypothesized that the performance of the participants in the trained groups would vary due to different L1 backgrounds, as in Flege (1989), Flege and Wang (1989), and Flege and Liu (2001).

1.7. Significance of the Study

The results of the present study may inform policy-makers and practitioners toward the integration of an effective perceptual training method into teacher training programs and English teaching process. After all, the majority of EFL teachers in Taiwan are not native speakers of English but they are usually the most influential model for their students to acquire English phonology. If non-native EFL teachers can be trained through an effective method to accurately perceive English segments that do not exist in Chinese and even improve their production, they can incorporate the same training method into their own teaching practices to help their students overcome the obstacles to succeed in perceiving and even producing English segments.
1.8. Assumptions of the Study

1. It was assumed that all the participants followed the instructions of the researcher during the time of the study.

2. It was assumed that all the participants received the same amount of English instruction from teachers other than the training.

3. It was assumed that the perception and production scores measured by the pretest and posttest were valid and reliable.

1.9. Definition of Key Terms

The following terms were operationalized to ensure clarity and consistency throughout the present study.

**Second language:** In the present study, the term second language refers to a second language or languages other than a first/native language.

**Functional load:** The simplest expression of the functional load of a phonemic contrast is defined by Brown (1988, p. 600) as “the number of minimal pairs that this contrast serves to distinguish.”

**Effects of the intervention:** In the present study, effects of the intervention were defined by the comparison of the experimental groups and the control group’s difference of posttest and pretest scores by statistical analyses.

**Talker variability:** In the present study, talker variability is defined by the use of stimuli spoken by different speakers (e.g., Lively, Logan, & Pisoni, 1993). The more speakers, the higher talker variability.

**Familiar talker:** In the present study, familiar talker is defined by the speaker of stimuli
used in perceptual training (e.g., Lively, Logan, & Pisoni, 1993).

**Familiar token:** In the present study, familiar token is defined by words used in perceptual training (e.g., Lively, Logan, & Pisoni, 1993).

**Talker effect:** In the present study, talker effect is defined by the proposition that participants are more able to identify phonetic segments in words spoken by those who produced the training stimuli than in words spoken by unfamiliar speakers (e.g., Lively, Logan, & Pisoni, 1993).
CHAPTER 2: REVIEW OF THE LITERATURE

2.1. Introduction

Empirical studies are crucial to inform pronunciation teaching. However, Derwing and Munro (2005) noted that pronunciation research has been marginalized in applied linguistics even with the recent resurgence of interest in the acquisition of L2 phonology, which often leaves teachers to rely on their own experiences and intuitions rather than empirical evidence. They called for more pronunciation research to help set pedagogical goals and priorities in the classroom. They also called for greater collaboration between researchers and teachers to provide more effective approaches to pronunciation teaching.

As mentioned in Chapter 1, accuracy in perceiving and pronouncing English segments is significantly correlated with degree of foreign accent in English. Although segmentals have been the focus of pronunciation teaching (Derwing, 2008), some researchers think segmentals are overemphasized in pronunciation teaching (e.g., Brown, 1995). Recently the focus of pronunciation teaching has shifted to suprasegmental level (Celce-Murcia et al., 2010). However, as Derwing (2008) pointed out, the inability to produce certain segments is likely to lead to a communication breakdown. In the same vein, Saito (2007) contended, “It can be readily argued that one must understand the segmental in order to be able to understand the suprasegmental fully. Considering communicative significance, phoneme awareness should be prioritized” (p. 20). He asserted that if speakers mispronounce the suprasegmental parts, listeners can still guess
the content of the message based on the intelligible segmental parts. On the other hand, if speakers mispronounce certain segments or listeners failed to perceive certain segments, communication is likely to be hindered. As the debate on teaching segmentals versus suprasegmentals stills goes on, some researchers asserted that we should address the issue by teaching both segmentals and suprasegmentals within a communicative framework (e.g., Celce-Murcia et al., 2010).

As Major (2001) noted, “segmentals have been the most thoroughly studied area in L2 phonology” (p. 14), studies on suprasegmentals are relatively fewer than those on segmentals. As the present study focused mainly on perceptual training of English phonetic segments, this review focused on studies on perception and production of English phonetic segments.

In the first section of this review, I survey major theories and models for the acquisition of second language phonology. Then factors involved in perception and production of L2 segments are discussed in the second section. In the third section, empirical studies related to Chinese speakers’ perception and production of segments are reviewed and discussed. Next, I turn to studies on L2 perceptual training and a few of them also involve production training. Then two training methods in identifying non-native L2 phonemic contrasts are compared. Finally, I review literature on sociolinguistic background in Taiwan and its demographic distributions of languages.

2.2. Theoretical Underpinnings in the Acquisition of L2 Phonology

There have been many approaches, theories and models for the acquisition of L2 phonology. In this section, I review several important theoretical underpinnings in L2
phonology acquisition.

2.2.1. Speech Production Theories

2.2.1.1. Contrastive Analysis Hypothesis (CAH)

When we learn a second language, we tend to transfer similar patterns we already acquired in our L1 to the L2, especially at the beginning stage, which often leads to a foreign accent. According to Major (2001), early work on Contrastive Analysis (CA) can be found in Weinreich’s (1953) work, which enumerated different types of L1 transfer (interference) such as sound substitution and phonotactic interference in language contact situations. Lado’s (1957) Contrastive Analysis Hypothesis (CAH) was a seminal work on CA, which assumed that by comparing systematically the student’s L1 and L2, we can predict and describe the problems that will cause difficulty in learning an L2. The difference between Weireich and Lado is that Weinreich focused on negative transfer from L1, while Lado addressed both positive and negative L1 transfer. CAH posits that all errors in L2 acquisition are due to L1 transfer, which offered new perspectives to applied linguistics and seemed to offer the key to second language acquisition. Since then CA has been extensively applied to language teaching without empirical investigation. However, a few years later several researchers (cited in Major, 2001) disagreed with Lado. In his seminal work on interlanguage and fossilization, Selinker (1972) pointed out that transfer is only one of the factors that result in an interlanguage. Several other researchers contended that not all errors in L2 acquisition are due to L1 transfer and that CA exponents do not take universal factors into account. Tarone (1987) noted the limitations of CA and cited many studies to show that CA cannot explain sociolinguistic variation. A case in point is that in Celce-Murcia’s (cited in Tarone, 1987) study the process of
avoidance could not be predicted by CA. CA was also criticized by Gass and Selinker (2008), who noted, “it is an oversimplification to think that comparing two languages is a straightforward comparison of structures… not all actually occurring errors were predicted; not all predicted errors occurred” (pp. 100-101). Nonetheless, they also pointed out that Lado’s CAH had inspired second language researchers to test the hypothesis in empirical studies. In view of all the criticisms, Eckman (1977) proposed the Markedness Differential Hypothesis to complement the CAH.

2.2.1.2. Markedness Differential Hypothesis (MDH)

The Markedness Differential Hypothesis (MDH) (Eckman, 1977) was formulated based on the construct of typological markedness. Eckman (2008) noted, “the MDH must be incorporated into the Contrastive Analysis Hypothesis (CAH) as a measure of relative difficulty in SLA” (p. 98). As mentioned earlier, the CAH does not take universal constraints into account. Eckman attempted to modify the CAH by taking phonological universals into consideration. The MDH posits that those aspects of the L2 that are different from the L1 will be difficult to acquire if they are also typologically more marked than the L1, while aspects of the L2 that are different but less typologically marked will be less difficult to acquire. Eckman (1977) proposed that the notion of “degree of difficulty” works correspondingly to the notion “typologically marked” (p. 320). However, he found that there were error patterns that the MDH could not predict and proposed another complementary hypothesis.

2.2.1.3. Structural Conformity Hypothesis (SCH)

Another hypothesis invoked the generalizations underlying markedness principles is the Structural Conformity Hypothesis (SCH) (Eckman, 1991). Eckman used the SCH to
complement his MDH to explain error patterns which do not arise out of the difference between L1 and L2 and thus cannot be explained by the MDH. The SCH holds, “The universal generalizations that hold for primary languages hold also for Interlanguages” (p. 24).

2.2.1.4. Ontogeny Phylogeny Model (OPM)

The Ontogeny Phylogeny Model (OPM) (Major, 2001) assumes that an L2 learner’s interlanguage (IL) is composed of “parts of the L1, parts of the L2, and universals” (p. 4) and as L2 proficiency increases, L1 transfer decreases while universals (U) increase and then decrease (Chronological Corollary of the OPM). Major asserted that universals in an interlanguage are “not already part of the L1 or L2 system” (p. 83). Major theorized the other three corollaries of the OPM as follows:

1. **Stylistic Corollary of the OPM.** IL varies stylistically in the following manner:
   
   As style becomes more formal, (a) L2 increases, (b) L1 decreases, and (c) U increases and then decreases.

2. **Similarity Corollary of the OPM.** In similar phenomena, IL develops chronologically in the following manner: (a) L2 increases slowly, (b) L1 decreases slowly, and (c) U increases slowly and then decreases slowly. Thus, the role of L1 is much greater than U, compared to less-similar phenomena.

3. **Markedness Corollary of the OPM.** In marked phenomena, IL develops chronologically in the following manner: (a) L2 increases slowly, (b) L1 decreases and then decreases slowly, and (c) U increases rapidly and then decreases slowly. Thus, except for the earliest stages, the role of U is much greater than L1, compared to less-marked phenomena. (p. 85, italics in original)
Although many researchers now have agreed that L1 transfer is only one of the factors that affect L2 production and that CA does not take universal factors into account, Odlin (2003) noted that by testing the CAH, researchers started to examine contrastive predictions and actual learner difficulties more closely. Moreover, studies afterwards still showed effects of L1 transfer and researchers such as Eckman have attempted to modify the CAH by taking other factors into account. On the other hand, Major asserted that universals in an interlanguage are independent of the L1 or L2 system and proposed his own model of speech production. Eckman (2008) suggested that the framework of Optimality Theory may explain interlanguage patterns that cannot be explained by L1 transfer or L2 input. As Major (2008) noted, even Optimality Theory framework acknowledges the importance of L1 transfer, which can be stated as L1 rankings.

2.2.2. Speech Perceptual Theories

2.2.2.1. Perceptual Assimilation Model (PAM)

Best’s (1994) Perceptual Assimilation Model (PAM) assumes that a phonologically mature listener will be unable to detect discrepancies between native and non-native sounds if she or he perceives the phones to be very similar to their native phoneme category. Under such circumstances, the non-native phones will be assimilated to the most similar native phoneme category. Best (1994) theorized that there are four patterns by which the two members of a given nonnative contrast could be perceptually assimilated to native phonemes:

1. The members of a nonnative contrast may be gesturally similar to two different native phonemes, thereby becoming assimilated to Two Categories (TC type).
2. The nonnative phones may both be assimilated equally well, or poorly, to a single native category, in which case they may be equally similar/discrepant to native exemplars of that Single Category (SC type).

3. Alternatively, the nonnative pair may both be assimilated to a single native category, yet one may be more similar than the other to the native phoneme, that is, the nonnative phones may show differences in Category Goodness (CG type).

4. Finally, the nonnative sounds may be too discrepant from the gestural properties of any native categories to be assimilated into any categories of the native phonology and should, therefore, be perceived as nonspeech sounds, that is, they are Nonassimilable (NA type).

(p. 191)

2.2.2. Native Language Magnet (NLM) Model

Kuhl, Williams, Lacerda, Stevens, and Lindblom (1992) found that exposure to a specific language in the first 6 months of life changes infants’ phonetic perception, which indicates that a language-specific pattern of phonetic perception is independent of the emergence of contrastive phonology and understanding of word meanings. This “magnet effect” (p. 607) of native language led to Kuhl’s (1993) Native Language Magnet theory. In Kuhl’s (1993) study, she cited her (1992) study to show that infants’ initial discrimination abilities, a boundary phenomenon, are language-general and independent of linguistic experience, but later modified by linguistic experience. Thus, she argued that early language input results in infants’ development of native-language phonetic categories and influences their perception of foreign language contrasts. The NLM holds that “perceptual magnets warp the acoustic space underlying phonetic distinctions by
shrinking the perceived distance between a magnet and its surrounding stimuli” (p. 268), which explains the assimilation of foreign language sounds. However, Kuhl (1992, 1993) could not be sure if the magnet effect is innate or if it emerges with exposure to a particular language. Expanding Native Language Magnet theory based on her recent research, Kuhl et al. (2008) proposed the concept of native language neural commitment (NLNC), which argues that the brain’s early coding of language affects our subsequent abilities. The concept was confirmed by studies using magnetoencephalography to show that the adult brain takes longer duration and larger area processing non-native speech sounds than processing native language sounds. The Native Language Magnet model, expanded (NLM-e) incorporated five new basic principles:

(i) Distributional patterns and infant-directed speech are agents of change.

(ii) Language exposure produces neural commitment that affects future learning.

(iii) Social interaction influences early language learning at the phonetic level.

(iv) The perception-production link is forged developmentally.

(v) Early speech perception predicts language growth.

(Kuhl et al., 2008, pp. 982-985)

2.2.2.3. Speech Learning Model (SLM)

Flege’s (1995b) Speech Learning Model (SLM) claims that (1) L1 and L2 phonetic subsystems cannot be fully separated in experienced L2 learners, (2) Discrepancies between L1 and L2 learner’s productions are due to perceived cross-language similarity, and (3) The degree of phonetic similarity determines whether L2 phonetic segments will be assimilated into existing L1 phonetic categories or whether separate L2 phonetic categories will be established. In other words, the SLM proposes that similar sounds will
be difficult to perceive and produce while different sounds will be easier to perceive and
produce.

Like the PAM and the NLM, the SLM proposes that the establishment of separate
L2 phonetic categories is influenced by existing L1 phonetic categories rather than loss of
ability to perceive and produce new phonetic categories in an L2. The SLM proposes that
adults still retain the ability to establish new phonetic categories in an L2. However,
phonetic category establishment for L2 speech sounds becomes less likely due to
reinforced L1 phonetic categories with increasing linguistic experience with L1.

All of the models above take into account the degree of similarity between L1 and
L2 phonology systems as well as L2 learners’ experience, which determines whether L2
phonetic segments will be assimilated into existing L1 or not. In addition, they all posit
that linguistic experience with L1 make later L2 learning difficult, especially in adulthood.
However, with respect to Flege’s (1995b) Speech Learning Model, I do not agree that
different sounds are always easier to perceive and produce. A case in point is the
interdental fricatives in English, which might be easy to perceive but has been proved
difficult to produce despite their salient difference in many languages (e.g., Derwing &
Rossiter, 2002; Derwing, 2003; Rau et al., 2009).

With respect to Kuhl’s Native Language Magnet (NLM) Model and the Native
Language Magnet model, expanded (NLM-e), the use of magnetoencephalography and
other devices to compare how the adult brain processes non-native speech sounds as
opposed to native language sounds is convincing with the scientific proof. However, I
wonder how the models could be applied to bilingual children’s language perception,
especially when there are similar sounds between the children’s two languages.
Investigating bilingual infants’ phonetic perception utilizing technology seems to be a good research direction in the future. Moreover, the concept of native language neural commitment (NLNC) argues that the brain’s early coding of language affects our subsequent abilities. The argument reminds me of some special cases of those feral children who have never been exposed to any language. Could they still acquire one or more languages since their neural tissue and circuitry have not reflected any properties of language input? Future research should be conducted to explore the possibility.

In comparison, Best’s (1994) Perceptual Assimilation Model appears to be the soundest perceptual theory among the three for it has detailed all four possibilities of assimilation. Although it was not initially intended for L2 acquisition, it can be extended to early stages of naturalistic L2 speech acquisition as in Guion, Flege, Akahane-Yamada, and Pruitt (2000). Before I turn to factors that might influence L2 phonology acquisition, I review studies on how foreign accent influences non-native English speaker’s identity construction and the issue of whether L2 learners should reduce foreign accents.

2.3. Foreign Accents Stigma and Identity in SLA

Regarding non-native accent, Gass and Selinker (2008) noted,

It is commonly accepted that the native language origin of a second language speaker is often identifiable by his or her accent. In fact, nonnative speaker pronunciation is often the source of humor, as in the case of comedians mimicking particular accent types, or in cartoon characters adopting nonnative accents. (p. 178).

Some of the jokes on foreign accent might be just for fun and mean no harm, while the
majority of them are usually associated with offensive linguistic stereotypes as described by Lippi-Green (1997).

Lippi-Green (1997) defined L2 accent as “the breakthrough of native language phonology into the target language” (p. 43) which is influenced by L1 with different degrees of success. She noted that foreign accent discrimination is related to racism and immigration such as discrimination against Mexicans. Matsuda (1991) pointed out, “The recent push for English-only laws, and the attack on bilingual education, may represent new outlets for racial anxiety now that many traditional outlets are denied. The angry insistence that “they” should speak English serves as a proxy for a whole range of fears displaced by the social opprobrium directed at explicit racism” (p. 1397).

Investigating accented non-native and native speakers’ perceptions of stigmatization and discrimination, challenges in communication, and feelings of social belonging in the U.S., Gluszek and Dovidio (2010) conducted two studies and found the non-native speakers perceived more problems in communication than did the native speakers with a regional accent but both groups did not differ in their perceptions of stigmatization. Moreover, regional accents were not related to foreignness, whereas the stronger the foreign accent, the less sense of belonging the non-native speakers felt. The study showed how the stigma and bias of foreign accents were detrimental to communication and the speakers’ social identity.

As Lippi-Green (1997) noted, “It is crucial to remember that it is not all foreign accents, but only accent linked to skin that isn’t white, or which signals a third-world homeland, that evokes such negative reactions” (pp. 238-239) and “Mainstream US English is a flimsy cover to hide behind in the face of serious intent to exclude on the
basis of race or ethnicity” (p. 242), in many cases it is racial discrimination that drives negative linguistic stereotyping and accent discrimination. On the other hand, as Hansen Edwards (2008) found in the literature, L2 learners may exercise agency to decide which accents to use as markers of gender, social, and ethnic identity. However, foreign accent usually does disservice to non-native speakers if they decide to keep their accents because what they say would be more difficult to understand. Lev-Ari and Keysar (2010) demonstrate that this “processing difficulty” (p. 1093) causes non-native speakers to sound less credible than a person with a native accent. They demonstrated that people believe non-native speakers are less credible simply because of their foreign accents instead of stereotypes of prejudice against foreigners. A case in point is that even though non-native speakers were merely reciting statements provided by a native speaker, they were still thought to be less credible because of their heavy foreign accents. The results showed that foreign accent may have an insidious impact on identity construction of non-native speakers. With the influx of immigrants into the U.S. and Canada in the last two decades, there are more and more ESL learners who face many identity difficulties after moving to a new country. On the other hand, in both ESL and EFL settings, many non-native-English-speaking (NNES) teachers and students also have L2 identity problems. I first turn to NNES teacher accent and identity.

2.3.1. NNES Teacher’ Accents and Issues of Identity

The standard language ideology institutionalized in the school system in the U.S. has marginalized not only students who speak with an accent but also teachers with foreign accents. Lippi-Green (1997) enumerated several examples of accent discrimination against teaching assistants and teachers with foreign accents. In one case,
403 residents filed a petition against reassignment of 2 bilingual teachers with foreign accents to first or second grade regular classrooms. The petitioners and even the mayor feared that children might pick up the teachers’ accents since they would provide language role models for their students. Lippi-Green dismissed the fear and argued that the fear had no foundation since the children had acquired their languages before they entered school.

Although Lippi-Green repudiated the petitioners’ fear in the case of the two bilingual teachers, it appeared that she did not take into account those immigrant students who already speak English with foreign accents. According to researchers who support a critical period or a sensitive period, children before puberty can acquire a native-like accent if they receive substantial input from native speakers of the target language. It suggests that the petitioners’ and the mayor’s concern was not groundless. I next turn to NNEST identity in ESL settings.

Investigating minority ESL teachers’ perception of their students’ ideal ESL teacher, Amin (1997, 1999) interviewed 5 minority female ESL teachers who had immigrated to Canada as adults and found that they felt disempowered by their students’ stereotype of an authentic ESL teacher. The results showed that the majority of the students preferred white teachers and questioned or even challenged the authority and authenticity of non-white teachers, which was detrimental to identity construction of the NNESTs.

It appears that white privilege is institutionalized within the TESOL profession (e.g., J. Liu, 1999b). However, Widdowson (1994) did not mention the dichotomy between Whiteness and non-Whiteness when discussing the association between native speakers and ownership of English. He contended, “this authority is claimed by those who possess the language by
primogeniture and due of birth, as Shakespeare puts it. In other words, the native speakers” (p. 379). Although Widdowson might have addressed racism in the example of coinage of new English words by Indians, he did not associate ownership of English with Whiteness. Therefore, it might have been in part the NNESTs’ race and mainly their accents that incurred discrimination. As Matsuda (1991) noted, “Any speech that is different from that constructed norm is called an accent” (p. 1361), which might explain why those ESL students preferred white teachers, who are usually native speakers of the target English variety.”

Addressing the “identity crisis” (Brown, 1994, cited in Demirezen, 2007) of NNESTs, Demirezen (2007) examines if NNESTs have identity problems and have the right to retain a foreign accent. He noted that some NNESTs intentionally retained a strong foreign accent in their speech to assert their national identity. As a result, their English was not intelligible and sounds unnatural to the students. Demirezen asserted that professional NNESTs should not retain a foreign accent as a foreign language right because the unintelligibility and incomprehensibility are harmful to students’ learning and a strong foreign accent is definitely detrimental to the efficacy of a professional identity.

It appears that Demirezen has a double standard when he asserted that professional NNESTs should not retain a foreign accent as a foreign language right but at the same time he concurred with Ur’s (1996) opinion in that some NNESTs’ wish to maintain a slight native language accent as a marker of their personal or ethnic identity should be respected. There has been a heated debate on whether non-native teachers of English should or have the right to retain their foreign accents. For example, Medgyes (1999) contended that the attitude of retaining a foreign accent should be of an amateur but not of a professional in the TESOL profession and is “inexcusable” (p. 192). He asserted that
a TESOL professional should pronounce English words “like striking a note on the keyboard of the imagination” (Wittgenstein, 1958, cited in Medgyes, 1999, p. 192), which means native-like pronunciation and intonation. Lado (1964) also asserted that one must be professionally qualified to perform professional duties and a TESOL professional identity requires the acquisition of specific foreign language knowledge and teaching skills. Moreover, Sifakis and Sougari’s (2005) survey on Greek English teachers’ attitudes toward pronunciation showed that native-like pronunciation is an indicator of EFL teachers’ authority as second language models in the classroom. However, D. Liu (1999) held the opposite opinion about pronunciation. He asserted that an excellent command of English includes fluency and the ability to use English idiomatically but not native-like pronunciation or intonation because it is not necessary in most EFL settings.

With respect to EFL or EIL (English as an international language) settings, Jenkins (1998, 2000) also suggested that we should respect students’ wish to retain their accents as a marker of their desired identities. It appears that the issue of foreign accent is highly contentious and political and will remain in debate for a long time.

Taking up poststructuralist view on identity, especially Norton’s (1995, 2000) view on identity as fluid, multiple, and a site of struggle within relations of power as well as Norton’s (2001) notion of *imagined communities*, Golombek and Jordan (2005) examined accent and social identity and explored how NNESTs could assert their legitimacy as an English user and a TESOL professional. The participants were 2 Taiwanese EFL teachers of English enrolled in a pronunciation pedagogy course in a TESL program in the U.S. The course adopted a critical approach to transform students’ identities as legitimate speakers and pronunciation teachers of English. The participants were asked to write
critical response to assigned readings and were interviewed after completion of the course. The analysis of the first participant’s narratives showed that although she did not fully accept NESTs’ language superiority, any oral deviation from the NS norm had been damaging to her identity as a legitimate speaker and pronunciation teachers of English. The readings in the course provided her with the language she could appropriate to contest the standard language ideology and imagine a new identity for herself as a legitimate teacher. On the other hand, she realized that her resistance to the myth of native speaker (NS) superiority and stress on intelligibility had to take students’ needs into account. The second participant was an experienced EFL teacher and had a great deal of self-confidence in her speaking ability but still felt stigmatized due to her lack of fluency and her accent. The critical pedagogy enabled her to contest the NS superiority myth and imagine an identity for herself as a multicompetent speaker and legitimate EFL teacher. However, the analysis also showed that even with their imagined “alternative identities” (p. 513), both of the participants were still ambivalent and were aware that their pedagogical goals had to serve students’ needs.

As in Brutt-Griffler and Samimy (1999), the critical pedagogy in this pronunciation course empowered the EFL teachers to contest the NEST myth and transformed their identities from deficient accented teachers to multicompetent speakers and legitimate EFL teachers. It suggests that such courses should be offered in more universities in the U.S. to raise TESOL professionals’ awareness and eliminate biases. As an EFL teacher in Taiwan, I agree with what the second participant stated about white privilege in ELT in Taiwan. With respect to foreign accents, I agree with the first participant in that accent is natural and learners of English do not need be ashamed of their accent. However, both of
the participants were aware that their pedagogical goals had to serve students’ needs. As Golombek and Jordan (2005) pointed out, “We as ESL teachers often experience our own contradictions—as teachers, we do not expect our students to have native-like pronunciation, but as foreign language students, we may seek out native-speaker models” (p. 517). However, do teachers’ expectations for students align with students’ goals? I turn next to NNES learners’ accent and identity construction in search of the answer.

2.3.2. NNES Learners’ Accents and Issues of Identity

Investigating whether there are mismatches between ESL students’ pronunciation needs and instruction they receive as well as how they cope with communication breakdowns, Derwing and Rossiter (2002) interviewed 100 adult immigrant full-time ESL students from 19 different L1 backgrounds in Canada. The results showed that when facing a communication breakdown, 56% of the participants reported paraphrase as a favorite strategy, followed by self-repetition (28%) and over a third of them felt that their foreign accents were the primary cause of any communication breakdowns. With respect to pronunciation needs and the instruction they received, with an emphasis on prosodic elements and the lack of teacher training in pronunciation instruction, the participants complained that they were either not receiving pronunciation instruction at all or not benefiting from suprasegmental instruction. Although 78% of the learners reported that their pronunciation was becoming more native-like, 90% of them stated that they still needed pronunciation instruction.

Although 90% of the participants wanted pronunciation instruction, only 8 out of the 100 participants had received explicit pronunciation instruction in the ESL classroom and there was a lack of teacher training in pronunciation instruction. The mismatches
imply that the TESOL programs in Canada should be held accountable for not preparing
pre-service teachers for teaching pronunciation and that there is a need for ESL programs
to incorporate pronunciation courses into the curriculum. Moreover, the finding that the
students complained that they did not benefit from a communicative approach focused on
prosodic elements exposes the fact that the students’ need for a focus on segmentals was
neglected. As Derwing (2008) pointed out, the inability to produce certain segments is
likely to lead to a communication breakdown. The findings that over a third of the
students felt that their foreign accents were the primary cause of communication
breakdowns and that they still desired pronunciation instruction (Derwing & Rossiter,
2002) suggest that the students’ needs should be heard and fulfilled.

Investigating ESL students’ own experiences on the issue of accent, Derwing (2003)
interviewed 100 adult immigrant full-time ESL students who were slightly different from
the participants in Derwing and Rossiter (2002). The results showed that although 60% of
the participants claimed that they were not discriminated against because of their accents,
53% felt they would be respected more if they spoke without a foreign accent. In addition,
95% of them would like to speak English like a native speaker and 97% believed or
strongly believed that it was important to have native-like pronunciation. Interestingly,
when asked whether they felt their own identities would be threatened if they spoke
English without a foreign accent, many claimed that their identities were tied to their L1
only.

It is worth noting that many of the participants claimed that their identities were
tied to their L1 only. As most identity studies indicated that there are L2 identities, it is
worth further research on whether the participants’ identities are tied to their L1 only. On
another hand, although Derwing suggested that ESL teachers focus pronunciation teaching on suprasegmentals and intelligibility, in Derwing and Rossiter (2002), the students claimed that they did not benefit from prosodic elements. Moreover, the finding that 97% of the participants believed or strongly believed that it was crucial to have native-like pronunciation suggests that those immigrants were eager to reduce their foreign accents to fit in the mainstream Canadian community. It suggests that pronunciation instruction should focus on both segmentals and suprasegmentals within a communicative framework, as suggested by Celce-Murcia, Brinton, and Goodwin (1996). After reviewing NNES learners’ accent and identity in ESL settings, I turn to learners’ accent and identity in EFL settings in the Expanding Circle.

In a modified matched-guise test, Dalton-Puffer, Kaltenboeck, and Smit (1997) investigated the attitudes of 132 Austrian university students of English toward 5 varieties of English speech: Received Pronunciation (RP), Estuary English, General American (GA), GA with an Austrian accent, and Estuary English with an Austrian accent. As expected, the results showed that more than two thirds of the participants preferred RP as it had been the traditional norm of ELT in Austria. RP was followed by GA and Estuary English, while the two Austrian-accented English speakers were negatively rated. On the whole, the GA speaker was rated more positively as a potential friend but more negatively in terms of professionalism, while the RP speaker was rated more positively professionally but more negatively as a potential friend. The most important finding was that personal experience was an important factor for those who chose GA as one’s model. The results showed that personal exposure in English-speaking countries led to more positively individualized, situation-specific attitudes rather than the stereotypes of the
EFL learners who had not had this experience.

One methodological flaw in the study might be the fact that the two Austrian speakers had “weak” Austrian accents (p. 117), especially the one who spoke American English. Most of the participants could not recognize her accent and as a result only 17% detected her Austrian accent (pp. 120-121). It appears that the researchers could have included heavily-accented and moderately-accented Austrian speakers as in McKenzi (2008). Moreover, since the majority of the participants were would-be EFL teachers, their attitudes could not represent those of the average students in Austria. I turn next to a study in Norway.

Employing a matched-guise test and a production test, Rindal (2010) investigated the attitudes of 23 Norwegian adolescent learners of English toward American and British varieties of English and how their attitudes influenced their accent choices and construction of L2 identity. Aside from the two tests, the participants were also asked to fill in a questionnaire and 5 of them were interviewed. In the production test, the participants were asked to read a wordlist with four phonological variables that distinguish RP and GA: postvocalic /r/, intervocalic /t/, /o/ as in goat, and /ɑ/ as in lot. The results showed that although the participants preferred RP as a model of pronunciation, more than two thirds of the participants’ production was American-like with various degree of use, which could be explained in part by L1 transfer. With respect to their attitudes, consistent with the findings in Dalton-Puffer et al. (1997), RP scores were significantly higher than GA in every aspect except popularity and RP was associated with prestige while GA was associated with informality. Although those who aimed for a GA accent or an RP accent shared the attitudes toward GA and RP, they
disagreed on whether the evaluations were positive or negative and consequently made opposing language choices. The accent choices indexed social meanings which were adopted by the users to construct their L2 identity. For instance, a GA accent indicated an identity against the authority, which is represented by an RP accent.

Contrary to the claim that identities are tied to their L1 only by the participants in Derwing (2003), this study showed that the participants adopted variations linked to indexical meanings to construct their L2 identities. This third wave variationist sociolinguistic study exemplified that identity is constructed and negotiated rather than projected and reflected and the combination of quantitative approach and qualitative approach increased the validity of the study. Now I turn to next study that surveyed EFL, English as an international language (EIL) and ESL students and teachers.

Investigating whether and how far students wanted to conform to native-speaker norms, Timmis (2002) surveyed 400 EFL, EIL and ESL students from 14 countries and 180 teachers from 45 countries. The results showed that 67% of the students preferred to speak English like a native speaker, while only 32% of NNESTs preferred a native accent. Nonetheless, the researcher interpreted the teachers’ choice as “the more realistic, rather than the more desirable outcome” (p. 243). While some students regarded NS pronunciation as a benchmark of achievement, a number of NNESTs considered it the benchmark of perfection. The results also showed that only 34% of the students from South Africa, Pakistan, and India preferred to have an NS accent, which suggests that this issue is context-sensitive. The researcher concluded that since there are students who desire to conform to NS norms, teachers should be aware of their students own pronunciation goals and honor their needs.
Compared to the high percentage (95%) of the participants who would like to speak English like a native speaker in Derwing (2003), only 67% of the students in this study preferred to have a native accent. The difference might be due to the fact that all the participants in the former study were in an English-speaking country, whereas some of the participants in the latter study were in EIL or EFL settings and they might exercise their agency to resist the NS norms.

In sum, the issues of foreign accent and NS norms are highly contentious and political. Some researchers asserted that ESL/EFL learners, especially NNESTs should reduce their foreign accents as much as possible (e.g., Lado, 1964; Medgyes, 1999; Arva & Medgyes, 2000; Demirezen, 2007) while others contended that native-like pronunciation or intonation is not necessary in EIL and EFL settings and that we should respect students’ wish to retain their accents as a marker of their desired identities (e.g., Ur, 1996; Jenkins, 1998, 2000; D. Liu, 1999). From the above studies we can see that speaking with a foreign accent could be detrimental to the speaker’s self-esteem and affect the speaker’s identity construction. However, some L2 users still showed resistance and decided to keep their foreign accents as a marker of ethnic group identity (Gatbonton et al., 2005) or use certain native accent to convey social meanings (Rindal, 2010), which should be honored by the teacher and taken into account when assessing students’ second language acquisition. The finding that 95% of the participants would like to speak English like a native speaker and 97% of them believed that it was crucial to have native-like pronunciation in Derwing (2003) suggests that immigrants are eager to reduce their foreign accents to fit in the mainstream community, which suggests that the pronunciation goal of ESL classes should be achieving a native-like accent. With respect
to EFL or EIL settings, as Jenkins (1998, 2000) suggested, we should respect students’ wish to retain their accents as a marker of their desired identities. Apparently, the issues of NS norms and foreign accents will remain in debate for a long time. Nonetheless, for perceptual training, the norm is usually American English or British English. For example, American English is the norm of ELT and English proficiency tests in Taiwan. Regardless of the possibility that students might speak English with a foreign accent, when teaching listening, it appears that EFL teachers in Taiwan need to expose students to native speaker input of the target language, as Flege (2009) suggested, to help them achieve native-like L2 perception of the target language community in addition to improving their test scores.

When it comes to acquisition of L2 phonology, there are various factors that play important roles. In the next section, I turn to factors that may influence L2 phonology acquisition.

2.4. Factors in L2 Phonology Acquisition

There are a variety of factors that influence the acquisition of L2 phonology. It appears that the onset age is crucial in acquiring native-like L2 phonology and cross-linguistic influence, i.e., L1 transfer, also plays an important role in constraining or contributing L2 phonology acquisition. Aside from age effects and L1 transfer, social factors have also been taken into account in the acquisition of L2 phonology. In this section, I review studies addressing these factors.
2.4.1. Age Effects

2.4.1.1. Debate on the Critical Period

According to Ioup’s (2008) extensive review of the literature, the results of most studies suggested that native-like L2 phonology is mostly found with very early L2 learners and the likelihood diminishes as the age increases. Thus it is unlikely for late learners to achieve native-like pronunciation. Although there are cases of successful late L2 learners, Munro (2008) called them “the exception rather than the rule” (p. 194) and attributes their success to either special language talent or high motivation.

Among the age effects the most discussed has been the Critical Period Hypothesis (CPH). Lenneberg (1967) hypothesized that as the hemispheric lateralization completes in our brain at puberty, our neural plasticity ends and our ability to acquire a language starts to decline. However, Ioup (2008) contended that Lenneberg’s claim is for L1 acquisition only and does not necessarily apply to L2 acquisition. Flege (1987) questioned the existence of a critical period for human speech learning by citing counter-evidence in several empirical studies. He contended that the CPH is an oversimplification of the speech learning process and that “an inappropriate acceptance of the CPH … has led to potentially erroneous conclusions regarding why children’s speech performance might differ from adults” (p. 167). Flege enumerated potential confounding factors such as L2 input and social factors.

It appears that longitudinal studies are required to investigate long-term effects of L2 phonology acquisition between children and adults, and further provide evidence or counter-evidence to support or dismiss the CPH. Whether the CPH can be applied to L2 phonology acquisition will be an ongoing debate. Now I turn to studies related to age
effects.

2.4.1.2. Studies Related to Age Effects

Most researchers agree that age of language learning affects phonetic perception. The question is, if there is a critical or a sensitive period, how early do age effects affect phonetic perception. By testing 64 six-month-old infants, 32 in the United States and 32 in Sweden, with both native and foreign language vowel sounds, Kuhl, Williams, Lacerda, Stevens, and Lindblom (1992) showed that exposure to a specific language in the first 6 months of life changed infants’ phonetic perception, which contradicts the suggestions of previous studies that language-specific phonetic perception occurs around 12 months of age. The results also indicated that a language-specific pattern of phonetic perception is independent of the emergence of contrastive phonology and understanding of word meanings. Kuhl et al. asserted that linguistic experience “shrinks the perceptual distance” (p. 608) around a native language prototype and causes the prototype to perceptually assimilate similar sounds. This “magnet effect” (p. 607) of native language led to Kuhl’s (1993) Native Language Magnet theory mentioned earlier. In the study we can see that linguistic experience resulted in language-specific phonetic perception at an early stage. Next, let’s look at a study investigating the effect of age on the L2 learners’ production of English vowels.

Investigating the effect of age on production of 11 English vowels by Italian immigrants who had immigrated to Canada at different ages ranging from 2 to 23, Munro, Flege, and MacKay (1996) found that foreign accent in all of 11 vowels was in positive correlation with age of arrival and that the late arrivers could not produce native-like pronunciation despite the fact that they had lived in Canada for an average of 32 years.
The results confirmed Munro’s previous finding that it is unlikely for most late arrivers to master English vowels. On the other hand, the results of the second experiment suggested that even the late arrivers were able to produce vowels that were intelligible to the raters most of the time, in spite of their accents. The findings suggested that many of the late arrivers may have established partially accurate perceptual representations for the vowels. It is worth noting that although the 10 raters are native speakers of General Canadian English, 9 of them were untrained and showed varying degrees of sensitivity to accentedness. Thus, the inter-rater reliability in this study is not unquestionable. After reviewing the acquisition of a new vowel category in English phonology, I turn to age effects on the acquisition of consonants in English phonology.

Examining production of voice onset time (VOT) values of the English phoneme /t/ by Spanish speakers who had acquired English either in childhood or as adults, Flege (1991) found that the early learners’ VOT values were native-like, while the late learners produced values that were between the monolingual Spanish and English values. Flege proposed that the mechanism of “equivalence classification” (p. 396) might have prevented the late learners from noticing the acoustic phonetic difference between unaspirated Spanish /t/ with short-lag VOT and English aspirated /t/ with long-lag VOT. The study was the first, according to Flege, to compare the performance of early and late L2 learners. The theory of “equivalence classification” (p. 396) later led Flege (1995b) to his Speech Learning Model.

In a subsequent study investigating the production of VOT values of the English consonants /p/ and /t/ as well as the production of interdental fricatives by 240 native Italian speakers who had immigrated to Canada between the ages of 3 and 21 years, Flege,
Munro, and MacKay (1996) found that age of L2 learning was the most crucial determinant of the participants’ accurate production of both the VOT values of voiceless stops and interdental fricatives, while the other factors such as motivation, social use, home use, and work use were significant in accounting for the variance of production. In addition, the later learners tended to substitute /t/ for /θ/ and /d/ for /ð/. The researchers also found that some late learners’ production was native-like. It is worth noting that although age of L2 learning (43%), language use factors (5%), and motivation (4%) accounted for 51% of the variance in the /θ/ /ð/scores, 49% of the variance was still unaccounted for. With respect to the variance in the VOT values of the voiceless stops, age of L2 learning (20%) and language use factors (10%) accounted for the variance, leaving an astonishing 70% of the variance unaccounted for. Due to the inability to explain the variance unaccounted for in the study, Feige et al. formulated six hypotheses for testing in future research. Next, I review a study investigating Japanese speakers’ production and perception of English /r/ and /l/.

Japanese speakers’ production and perception of English /r/ and /l/ is probably the most studied areas in the L2 acquisition of English phonology. Aoyama, Flege, Guion, Akahane-Yamada and Yamada (2004) investigated Japanese speakers’ perception of English /r/ and /l/ to test the Speech Learning Model (SLM) regarding the effect of perceived cross-linguistic phonetic dissimilarity on the learning of L2 speech sounds. For Japanese speakers, English /l/ is perceptually more similar to Japanese /ɾ/ than English /ɾ/. Thus, according to the SLM, English /l/ would be more difficult for Japanese speakers to acquire than English /ɾ/. The researchers examined the perception of /l/-/ɾ/, /ɾ/-/w/, /b/-/s/, /s/-/θ/, and /b/-/v/ contrasts and the production of /l/, /ɾ/, and /w/ by 16
Japanese children and 16 Japanese adults in two different experiments at two different times with an interval of 1.1 years. The results showed that the Japanese children’s perception of /r/ and /l/ significantly improved during the year. In contrast, the Japanese adults performed better than the children at the beginning but did not significantly improve a year later. With regard to production, the Japanese children showed greater improvement in the production of English /r/ than /l/. Again, although the Japanese adults performed better in the production of English /r/ than the children in the first experiment, they did not improve in the second experiment. The findings support Flege’s SLM. However, the Japanese children showed significant improvement in the production of English /w/, which contradicts SLM since English /w/ has a similar sound in Japanese.

Compared with the previous studies, Aoyama et al.’s longitudinal study allows us to see the long-term development of L2 learners’ acquisition of L2 phonology. Moreover, as the researchers suggested, the significant improvement in the production of English /w/ by the Japanese children, which contradicted SLM, requires investigation in further research. Aside from Italian, Spanish, and Japanese speakers’ perception and production of English segments, Flege and his colleagues also investigated age effects on the English segment acquisition of Korean speakers.

Flege, Yeni-Komshian, and Liu (1999) tested the Critical Period Hypothesis by studying the degree of foreign accents and knowledge of English morphosyntax of 240 Korean speakers who had come to the United States between the ages of 1 to 23 years. As in the previous studies, the findings showed that those who had arrived at younger ages tended to have more native-like pronunciation and age effects on the grammaticality judgment test scores were also significant. However, the findings also showed that there
was no sharp increase of foreign accent at an age of 12 or 15 years, the previous proposed ends of critical ages, which appeared to be counter evidence to the Critical Period Hypothesis. Although age effects appeared to be salient, Flege et al. found that social factors such as differences in U.S. education and amount of English use played a much greater role in their knowledge of English morphosyntax.

Flege et al. found that age effects are stronger in the domain of phonology than morphosyntax. Although the findings appeared convincing, the construct validity of the foreign accent test in Flege et al.’s study seemed problematic. First, each sentence was presented twice on a loudspeaker and then the participants had to repeat the sentence when they heard a tone. It suggests that aside from testing foreign accent, the test also tested listening comprehension and short-term memory. That is, those participants who did not understand the sentences or forgot part of the sentences would not be able to repeat the sentences. Second, the researchers do not specify who recorded the 21 sentences for the accent test. Suppose the sentences had been recorded by a native speaker of English, the participants would have a model to imitate the sounds and intonation after listening to each sentence twice. In my view, since it was an accent test, the participants could have read a list of the sentences instead of repeating after the recordings.

Replicating Johnson and Newport’s (1989) study, Birdsong and Molis (2001) investigated maturational constraints on 61 Spanish speakers with different ages of arrival and found evidence that was inconsistent with Johnson and Newport’s (1989) findings. Although L2 attainment was in negative correlation with age of learning, onset of age effects and age effects prior and after maturation differed from what Johnson and
Newport had found for Chinese and Korean speakers. The results also showed modest evidence of native-like attainment among late learners and aside from age effects, there were other variables such as age of first exposure to English, L1 influence, and amount of English use. In terms of native-likeness, the results rejected the Critical Period Hypothesis. Thus, the researchers questioned the generalizability of Johnson and Newport’s study and argued that L2 attainment may depend on L1 influence and amount of L2 use. I now turn to cross-linguistic factors in the next part.

Regarding the correlation of neurological maturation and language acquisition, DeKeyser and Larson-Hall (2005) claimed that to this date no researchers can prove that the development or decline of a certain neurological mechanism is related to incomplete language acquisition of adult learners and thus regard such explanation as “highly speculative” (p. 101). For example, Ohala (2008) cited Werker and Tees’ (1984) study to show that infants’ weakening in perceptual acuity is not due to any physical deterioration of the human auditory mechanism but rather is a function of how listeners adjust their phonetic categories to “tune out” (p. 24) those contrasts that are not relevant to the native language. However, the studies in Kuhl et al. (2008), which used magnetoencephalography and other devices to show that the adult brain takes longer duration and larger area processing non-native speech sounds than processing native language sounds, might provide evidence that initial language exposure causes physical changes in neural tissue and circuitry and thus enhances sensibility to one’s native language phonemes and reduces sensitivity to non-native phonemes. DeKeyser and Larson-Hall noted that even if the Critical Period Hypothesis is correct, the elementary immersion programs in Canada did not capitalize on children’s implicit learning skills
because of the focus on form and the limited time of learning. On the other hand, the Critical Period Hypothesis might be detrimental to the learning of those who have passed the period and are not aware of successful instances of adult learners. The researchers asserted that children learn implicitly while adults learn explicitly and that language programs should adapt to different learning mechanisms between children and adults.

2.4.2. Cross-linguistic Factors

The results in Birdsong and Molis (2001) showed that L1 influence may play an important role on L2 attainment. The Ontogeny Phylogeny Model (Major, 2001) also assumes the role of L1 is much greater than universals for similar phenomena. The Contrastive Analysis Hypothesis (Lado, 1957) even attributes all L2 errors to L1 transfer although this argument has been disapproved of by current researchers.

Weinreich (1953) studied various types of sound transfer before Contrastive Analysis and enumerated different types of transfer (interference) in language contact situations: (1) sound substitution, in which the learner substitutes the L2 sound with the closest L1 sound, (2) phonological processes, in which the learner uses the L1 allophonic variant that does not apply to the same linguistic environment in the L2, (3) underdifferentiation, in which there are contrasts in the L2 but not in the L1, (4) overdifferentiation, in which there are contrasts in the L1 but not in the L2, (5) reinterpretation of distinctions, in which the learner reinterprets secondary features of the L2 as primary or distinctive features, (6) phonotactic interference, in which the learner makes the syllable structure in the L2 the same as the L1, and (7) prosodic interference, in which the learner produces the L2 according to the L1 prosodic rules.

Aside from the types of sound transfer listed by Weinreich, there are loan phonology
phenomena (Broselow, 2000), which can be considered as culturally induced transfer or forced transfer since the learner usually pronounces loan words as if they were native words in their L1. For example, there are a lot of loan words from English in Japanese. Of course, there is not only negative transfer from L1. As Zampini (2008) noted, there is positive L1 transfer which can have a facilitative effect on L2 pronunciation, especially when both the L1 and L2 sound systems share the same features. Now I turn to studies addressing L1 influence on L2 phonology acquisition.

Investigating L1 linguistic influence on the word-final /t/-/d/ deletion of 20 adult Chinese learners of English from both China and Taiwan, Bayley (1996) found that final /t/ and /d/ were less likely to be deleted following a liquid than an obstruent or nasal and that the more careful the style, the less likely final /t/ and /d/ were deleted. The results for the phonological environment, including the preceding and following segments, and voicing agreement with the preceding segment, were similar to the patterns for native speakers of English. Unlike native speakers of English, the Chinese learners in this study were more likely to delete inflectional final /t/ and /d/ clusters than lexical ones. One thing interesting is that since Mandarin has no final consonant clusters and only two final consonants, Bailey hypothesized that triple clusters would be much more likely to be reduced than double clusters. However, the results of analyses did not confirm the hypothesis. After comparing the likelihood of final /t/ and /d/ deletion from monomorphemes and participles, Bayley asserted that in addition to the rules for past tense marking, learners must acquire a system of boundary constraints that inhibits morpheme deletion to achieve native-like performance.

It is worth noting that, as Bailey pointed out, some of the participants, especially
those from Taiwan, also spoke other dialects. Because of the high prestige of Mandarin in Taiwan, their self-reported status as native-speakers of Mandarin is not unquestionable. On the other hand, L1 transfer was not evident in this study. It appeared that universals were more evident in the study for the higher percentage of final /t/ and /d/ deletion on inflectional clusters rather than lexical clusters can be generalized to the Vietnamese learners of English in Wolfram and Hatfield’s (1984, cited in Bailey, 1996) study.

Examining whether there are word class effects on L2 pronunciation acquisition, Yeni-Komshian, Robbins, and Flege (2001) studied 192 adult Korean immigrants who had arrived in the U.S. between 6 to 23 years. The results indicated that they pronounced verbs more accurately than nouns. Moreover, the results of a grammaticality judgment test showed that they could detect incorrect formulations of verbs more precisely than nouns. The effects were more significant on those who had arrived in the U.S. later than 12 years old, which reflected the influence of the linguistic structure of Korean on learning English as an L2 since previous studies had showed that Korean children acquire verbs earlier than English children, who receive more nouns than verbs from their mothers and thus have a better control of nouns than verbs. In the next study, I turn to L1 influence on Japanese speakers’ perception of English consonants.

The single liquid consonant, an apico-alveolar tap /ɾ/, found in Japanese might be considered similar to both English /ɾ/ and /l/ and the two English consonants have been proved the most difficult for Japanese learners to acquire. Investigating adult Japanese speakers’ perception of English consonants, Guion, Flege, Akahane-Yamada, and Pruitt (2000) conducted two experiments with Japanese speakers with varying English learning experience. In Experiment 1, nine near-monolingual Japanese listeners were asked to
identify English and Japanese consonants in terms of a Japanese category and rate the consonants for goodness-of-fit to that Japanese category. The results showed that the Japanese listeners identified the Japanese consonants correctly nearly 90% of the time. Of the eight English consonants examined, five were consistently classified as instances of a single Japanese consonant category while the other three English consonants, /θ/, /r/, and /l/, were identified as two Japanese consonants. In Experiment 2, the same set of stimuli in Experiment 1 was used. Forty Japanese speakers with varying English-language experience were divided into low/mid/high-experienced groups and 10 native English speakers participated as a comparison group. The participants were asked to distinguish English-English, English-Japanese, and Japanese-Japanese consonant contrasts. The results showed that the perceived phonetic distance of L2 consonants from the closest L1 consonant predicted the discrimination of L2 sounds. Moreover, the role of experience in learning L2 sounds was also investigated and the results confirm evidence of learning in positive correlation with experience. The likely overlap in category assimilation between English /θ/ and /s/ suggested a need for a minor revision of the Perceptual Assimilation Model. On the other hand, the results indicated that the Speech Learning Model cannot be applied to relatively inexperienced L2 learners. The researchers suggested that Japanese speakers must learn to build a hardly distinguished pair of English sounds into two contrasting phoneme categories in order to acquire the sound structure of English.

Although the Perceptual Assimilation Model was not initially intended for L2 acquisition, it can be extended to early stages of naturalistic L2 speech acquisition to better predict L2 learners’ perception than the Speech Learning Model, as shown in this study. One potential threat to validity is that the researchers equated near-monolingual
Japanese speakers with those who began to study English at the age of 12 and learned
English mostly in the classroom. However, it is possible that those participants’ learning
had been efficient and learning experience had been accumulated substantially. It appears
that the researchers could have taken the Japanese participants’ English proficiency levels
into account by referring to their English grades at school.

According to Zampini’s (2008) extensive review, the majority of studies have shown
that L1 transfer plays an important role in the production of L2 syllables. Moreover, the
acquisition of L2 syllables may show effects of universal constraints on syllable structure,
especially longer and more complex onsets and codas. Aside from age effects and L1
transfer, social factors are variables that make the investigation of L2 phonology
acquisition complicated. I turn next to social variables in L2 phonology acquisition.

2.4.3. Social Factors

Hansen Edwards (2008) reviewed extensive research and concluded that both
linguistic factors and social factors affect variable L2 phonology production. In this
section, I focus on five social factors that affect L2 phonology acquisition. Among them,
length of residence is often conflated with age effects. However, age effects focus on the
onset age of L2 learning, while length of residence focuses on the length of living in an
L2-speaking country.

2.4.3.1. Length of Residence

Investigating the roles of length of residence (LOR) and input on L2 acquisition,
Flege and Liu (2001) administered three experiments with 60 adult Chinese speakers with
different LORs and one of the three language backgrounds: Mandarin, Shanghainese, or
Cantonese. Among the participants, half were students who had to use English and the
others were nonstudents with a full-time job and were not required to use English. They were asked to identify word-final English stop consonants, take a grammaticality judgment test, and then take a listening comprehension test. The results showed that the students with long LORs scored significantly higher than those with short LORs in all the experiments. However, no significant difference was found among the nonstudents with different LORs. On the other hand, the results showed that living in the United States for a long time alone did not significantly increase the participants’ knowledge of English morphosyntax. Therefore, the role of LOR was not clear. The researchers suggested that native speaker input may play a crucial role on L2 acquisition.

Although Flege and Liu claimed that substantial native speaker input may play a crucial role in the Chinese speakers’ English proficiency, they did not measure the amount of L2 input in the study. Thus, the thorny questions are (1) how do researchers measure the amount of L2 input? and (2) how do we define the unit of L2 input? By counting spoken L2 words by native speakers per day? What about other input such as TV shows and English spoken by L2 speakers? Moreover, despite the fact that the Chinese participants had 3 different first languages, the researchers did not take the possibility of different L1 phonotactic constraints into account.

Examining the production and perception of English vowels by Korean speakers, Tsukada, Birdsong, Bialystok, Mack, Sung, and Flege (2005) administered three experiments with 72 Korean speakers, half adults and half children, with different LORs in a one-year longitudinal study. The results of Experiment 1 showed that some pairs of contrastive English vowels were difficult for the Korean speakers to discriminate. In Experiment 2 and 3, 36 age-matched native speakers of English, half adults and half
children, also participated as comparison. The results of Experiment 2 showed that the Korean children could discriminate English vowels more accurately than the Korean adults but less accurately than native-English-speaking children. After a year, in Experiment 3, English words containing 6 vowels were elicited using a picture-naming task. The results revealed that the Korean children produced significantly larger between-vowel contrasts than the Korean adults but did not differ from native-English-speaking children. However, the Korean children with an LOR of 3 years discriminated four pairs of English vowels less accurately than the native-English-speaking children, which contradicts the traditional view of children as fast and perfect learners of L2 phonology. The results of the study showed significant age effects but did not support the Critical Period Hypothesis for the performance of the groups before and after the period did not differ significantly. The age effects were confounded by the fact that the Korean children were likely to have received more input from native speakers than the Korean adults since they were immersed in the L2 environment. With respect to LOR, LOR effects were not significant.

Examining the role of LOR on the degree of foreign accent, Flege, Birdsong, Bialystok, Mack, Sung, and Tsukada (2006) elicited English sentences from 108 Korean children and adults in the U.S. and Canada with different LORs at two different times with an interval of 1.2 years. The results showed that the Korean children spoke English with a foreign accent and neither the difference between the three-year and five-year LOR groups nor the difference between the two elicitations was significant. The researchers suggested that there was a strong L1 interference in both the Korean children and Korean adults. Moreover, the fact that the Korean children, who had arrived in an
English-speaking environment at a very young age and had lived in an English speaking community for a long time, still spoke English with a foreign accent appeared to contradict the Critical Period Hypothesis. Again, the researchers concluded that LOR effects are evident only if an immigrant receives a substantial amount of native speaker input.

Baker and Trofimovich (2006) conducted two experiments to investigate whether English learners’ age of arrival (AOA) and length of residence (LOR) in an English-speaking country determine the relationship between English perception and production. In Experiment 1, 40 Korean speakers and 10 native English speakers participated in vowel perception and production tasks in English. The results show that the relationship between perception and production performance depended on learners’ AOA instead of LOR. In Experiment 2, the same Korean speakers were asked to judge their own and others’ productions of English vowels. The results suggested that the learners with good production skills were equally good at other-perception and self-perception and the learners with poor production skills were equally poor at other-perception and self-perception. In contrast, among the four learners with intermediate production skills, half of them performed better at other-perception than self-perception and production, while the other half performed better at self-perception accuracy than their production accuracy and other-perception. The results suggested that self-perception is an important factor in determining the relationship between perception and production.

In sum, the role of LOR is either not significant (Flege & Liu, 2001; Tsukada et al., 2005; Baker & Trofimovich, 2006) or is evident only if an immigrant receives a
substantial amount of native speaker input (Flege et al., 2006). Now I turn to studies on
the role of L2 input in L2 phonology acquisition.

2.4.3.2. Exposure to the L2

In a production and perception study on two groups of adult German speakers
differing in English language experience, Bohn and Flege (1996) suggested that
experience has a greater impact on the production than on the perception of a new vowel
category. The German speakers were asked to produce /æ/, a sound not found in German,
and to label stimuli from a *bet* to *bat* continuum. The researchers found that with
extensive L2 linguistic experience, adult learners can produce and perceive a new vowel
category in a way similar to native speakers of the L2, which supported the Speech
Learning Model in that L2 sounds which are different from L1 sounds will be easier to
learn. It appeared to be counterevidence to the CPH. However, as the results showed,
even though the native English group showed variation in the production task and the
native-like segmental production of minimal pairs such as *bat* and *bet*, it did not
necessarily guarantee native-like production at the discourse level. In addition, although
the results indicated that production of L2 sounds was easier than perception of the same
sounds, the researchers also cited studies that had shown that production of L2 sounds
lagged behind perception (p. 68). Thus, the generalizability of this study remained
tentative.

As shown in the study, native-like production of L2 minimal pairs does not
necessarily guarantee native-like production at the discourse level. On the other hand,
although the researchers claimed that production of L2 sounds is easier than perception of
the same sounds, they also cited studies that showed counterevidence. It suggested that
the relationship between perception and production is not clear and merits further investigation.

In the previous study, Bohn and Flege (1996) claimed that production of L2 sounds is easier than perception of the same sounds. The next study made the opposite claim. Investigating whether L2 learners have to perceive L2 phonemic contrasts before production, Eckman, Iverson, Fox, Jacewicz, and Lee (2009) examined the production and perception of /s/ and /ʃ/ by four Japanese speakers and six Korean speakers. For the production task, the productions of 60 target words and 30 fillers were elicited from each participant and recorded under computer control. The researchers hypothesized that a contrast between /s/ and /ʃ/ in basic environments must happen before it occurs in derived environments. The results of the Korean participants confirmed the hypothesis but those of the Japanese participants did not. For the perception task, stimuli composed of single-word minimal pairs with either /s/ or /ʃ/ were presented on the computer screen for the participants to identify. The researchers hypothesized that a contrast between /s/ and /ʃ/ in derived environments without a contrast in basic environments is likely. The data of both Korean and Japanese participants confirmed the hypothesis for all four of the logical possibilities for the combination of the /s/-/ʃ/ contrast occurred in both basic and derived environments. The researchers suggested that the differences between the Korean and Japanese participants in the production task could be attributed to the fact that a contrast of /s/ and /ʃ/ exists in certain Japanese words.

Due to the fact that this study was a pilot study, the sample size was small. As Eckman et al. suggested, the findings in this study cannot be generalized. Moreover, if it is a fact that a contrast of /s/ and /ʃ/ already exists in certain Japanese words, it would
have been a methodological flaw that the researchers chose the contrast. The validity of the production task would be called into question. With respect to the interrelation between perception and production, Eckman et al. argued that perception of L2 phonemic contrasts should occur before production, which is opposite to Bohn and Flege’s (1996) claim.

Investigating if segmental acquisition remains possible through L2 experience, Munro and Derwing (2008) elicited production of 10 English vowels in CVC context by 44 immigrants who had recently arrived in Canada. The participants were divided into two groups, with the first groups consisting of 20 Mandarin speakers and the other group composed of 24 speakers of Slavic languages. All of them had extremely limited speaking and listening skills and had never experienced massive exposure to spoken English before arrival. Data were collected six times every two months over the course of a year. Each time the participants repeated a set of target CVC words and the repetition was recorded and evaluated by both trained and untrained listeners. The results revealed that both groups showed improvement in intelligibility over the year without focused pronunciation instruction. However, the greatest improvement occurred in the first six months and slower progress occurred after that. The improvement of both groups supported the Speech Learning Model in that the adult L2 learners could still acquire native-like segmental pronunciation with substantial exposure to L2. However, both groups’ performance on /ɪ/ after a year was still unsatisfactory. On the other hand, the possibility of L1 influences and possible effects of word frequency were not excluded.

The only potential threat to the validity of the study is inter-rater reliability due to the employment of untrained listeners. However, the researchers successfully neutralized
the threat by conducting an experiment on inter-judge agreement and found that the untrained listeners’ vowel intelligibility ratings corresponded to those of the expert judges. The finding that the Mandarin group’ performance on /ɪ/ after a year was still unsatisfactory could be explained by L1 influence since there is no /ɪ/ in Mandarin and by the fact that they received no focused pronunciation instruction.

The above studies on L2 experience suggest that adult L2 learners can still acquire native-like segmental pronunciation with substantial exposure to L2, but native-like production of L2 minimal pairs does not necessarily guarantee native-like production at the discourse level. Flege (2009) reemphasized the importance of native speaker input and quality of input. He noted that usually the first English input immigrants receive is the input of accented EFL teachers in their own countries and even after they arrive in an English speaking country, they are still surrounded by mostly foreign-accented non-native speakers of English, which reinforces their own foreign accents. He reiterated that length of residence in an L2-speaking country is likely to play a role in L2 phonology acquisition only for learners who receive a substantial amount of native-speaker input.

So far the frequency of either L2 input or L1 use in current research has been self-reported by the participants, which is unreliable and problematic. Flege (2009) suggested that to accurately measure L2 input or L1 use seems impossible due to practical and ethical limitations. The Experience Sampling Method, as Flege recommended, might be able to provide estimates of the participant’s language use. However, for young ESL learners, cell phone use might be banned in school settings. Nonetheless, a systematic method to measure participants’ amount of L2 input and L1 use is imperative to accurately examine the roles of these variables. On the other hand, as
Flege noted, quality of L2 input is crucial for acquisition of standard spoken English. Therefore, how do we set standards to measure the quality of input is another thorny issue.

2.4.3.3. Social Network

As Flege (2009) pointed out, an L2 learner’s social network might be decisive in his L2 acquisition in that the number of native speakers in one’s social network might reflect the amount of one’s exposure to the models of the target language and the opportunity for the learner to use the target language. In this part, I turn to studies on the role of social network in L2 phonology acquisition.

Investigating the effect of social network, Bayley (1996) divided his Chinese participants into one group with a mixed social network composed of both Chinese and Americans, and the other group with a primarily Chinese social network. He also examined the effect of language proficiency and found both extralinguistic factors significant. The results showed that the participants with a mixed social network were more likely to have final /t/ and /d/ deletion than those with a primarily Chinese social network, which may be explained by the possibility that they acquired more native-like patterns of /t/ and /d/ deletion from native speakers’ variation patterns. On the other hand, the lower proficiency participants were more likely to delete final /t/ and /d/ than higher proficiency participants, which might be due to an immature acquisition of past tense and consonant clusters as well as L1 transfer since there are no final consonant clusters and only two final consonants, /n/ and /ŋ/ in Mandarin. Furthermore, phonetic environments also affected the Chinese speakers’ final /t/ and /d/ deletion. For instance, final /t/ and /d/ were less likely to be deleted following a liquid than an obstruent or nasal. Moreover, the
Chinese speakers were less likely to delete final /t/ and /d/ in a careful style. One interesting finding is that the Chinese speakers were more likely to delete final /t/ and /d/ in past tense than in past participles, while native speakers tend to retain the word-final /t/-/d/ in past tense.

The finding that the Chinese speakers were more likely to delete final /t/ and /d/ in past tense than in past participles might be explained by the fact that there are far fewer usages of passive voice in Mandarin than in English and past participles are more marked than past tense to Chinese speakers, which might have resulted in the Chinese speakers’ retention of final /t/ and /d/ in past participles in the study.

Investigating the role of native speaker input in L2 learners’ English pronunciation and fluency development, Derwing, Thomson, and Munro (2006) examined the speech of 20 Mandarin and 20 Slavic adult immigrants in Canada in a 10-month study. The participants were full-time ESL learners with no special instruction focused on either pronunciation or fluency. For the speaking task, each participant was asked to narrate a story in English based on pictures at the beginning, 2 months later, and at the end of the 10-month study. Then they responded to a questionnaire regarding their informal exposure to English outside school. The results showed that both groups showed a little improvement in reducing their accents over the 10-month period. With respect to fluency, the Slavic group showed significant progress, whereas the Mandarin group showed no improvement although they had been judged more fluent than the Slavic group at the beginning of the study, which showed the possibility of a ceiling effect. On the other hand, the Slavic group reported significantly more contact with native and non-native English speakers, which might account for their better fluency performance.
As mentioned above, the fact that the amount of exposure to English was self-reported by the participants could be unreliable and problematic. Moreover, in the study, 20 native-English-speaking listeners were recruited to judge the production of the Mandarin group, while another 20 native-English-speaking listeners judged the production of the Slavic group. Although the researchers made sure that the inter-rater agreement was acceptable, it appears that the same group of raters could have been used for both groups of participants. Finally, it is worth noting that different from Flege’s (2009) emphasis on quality of native speaker input, in this study, exposure to English included radio and TV as well as contact with native and non-native English speakers.

The above studies suggested that contact with L2 speakers is in positive correlation with L2 learners’ acquisition of L2 phonology. In the next part, I turn to the role of L1 use in L2 phonology acquisition.

### 2.4.3.4. Amount of L1 use

To determine if degree of perceived foreign accent in an L2 depends on the amount of L1 use and to test the Critical Period Hypothesis, Flege, Frieda and Nozawa (1997) studied two groups of Italian speakers who had immigrated to Canada at different ages. Forty Italian speakers were asked to self-report use of Italian and produce five short English sentences. The results showed that both high and low users of Italian were rated as having a detectable foreign accent although they were early learners and had lived in Canada for an average of 34 years. However, the participants who spoke more Italian had a significantly heavier foreign accent than those who spoke less Italian. The results showed evidence against the Critical Period Hypothesis and Flege et al. contended that the amount of L1 use plays a more crucial role in degree of perceived foreign accent.
The results of the study showed that although the children began learning English at an average age of 5 years 8 months, which is before the earliest proposed cutoff age, 6 years old, of the critical period (Long, 1990), they still maintained a detectable accent. Therefore, judging by the evidence in this study, if there is a critical period, the cutoff age has to be at least younger than 6 years old.

Examining factors affecting degree of foreign accents in an L2, Piske, MacKay, and Flege (2001) conducted an experiment with 72 Italian speakers who had been born in Italy and had arrived in Canada at different ages. The participants were asked to self-report use of Italian and then repeat sentences as in Flege, Yeni-Komshian, and Liu (1999). The results were consistent with the findings in the previous study (Flege et al., 1997) in that the participants who spoke more Italian had a significantly heavier foreign accent than those who spoke less Italian and the L1 use effect existed in both early and late learners with late learners having a stronger foreign accent. Although the results showed that there was an L1 use effect in degree of perceived foreign accent, Piske et al. noted that the results could not be used against the existence of a critical or sensitive period in L2 phonology acquisition.

In the study, the participants were asked to repeat sentences as in Flege, Yeni-Komshian, and Liu (1999). However, different from the way of presentation in Flege et al., the sentences were presented in both written form and aural input. As in my critique of the procedure in Flege et al. (1999), the construct validity in this study is also problematic for the same two reasons. In my view, the participants could have been provided with the written sentences only instead of repeating after the recordings while reading the texts so that they would not have the opportunity to imitate the recordings.
The results of the above studies showed an L1 use effect in degree of perceived foreign accents. However, in a follow-up study on the relationship between L1 use and pronunciation ability with 72 Italian speakers, Flege, Schirru, and MacKay (2003) found that an L1 use effect was found only in the early learners. The results supported the Speech Learning Model in that the late learners assimilated English /e/ to a similar phonetic category in Italian while the early learners with low L1 use exaggerated /e/ in order to establish a new category and to differentiate it from the Italian /e/. Next I turn to a study on the amount of L1 use and perception of non-native segments in an L2.

Investigating whether adults who are early learners can perceive L2 vowels like native speakers, Flege and Mackay (2004) conducted four experiments with adult Italian speakers and found that those who had lived in Canada for only three months could not discriminate certain English phonetic contrasts and, as the SLM predicted, assimilated each contrast to a similar Italian phonetic category, while the early learners scored significantly higher than the late learners. Some of the late learners were also found to perceive English vowels accurately.

The most important finding in this study is that the early learners with high Italian use did not perform as well as native speakers of English, which supports Kuhl’s Native Language Magnet theory. However, the fact the early learners with low Italian use and the native English speakers did not differ significantly appear to contradict Kuhl’s theory. Since the early learners had immigrated to Canada between the ages of 2 and 13, according to Kuhl’s theory, their L1 magnet effect should have been in effect long before they immigrated to Canada.

In sum, the results of the above studies indicated L1 use effects in L2 phonology
acquisition, with the exception of the late learners in Flege, Schirru, and MacKay’s (2003). The findings that some of the late learners were found to perceive English vowels accurately in Flege and Mackay (2004) suggest that late learners can still establish new phonetic categories in an L2. As Flege (1999) contended, decline in L2 pronunciation accuracy with age is due to reinforced establishment of L1 phonetic categories instead of loss of ability to perceive and produce new phonetic categories in an L2. Evidently, amount of L1 use plays a decisive role in establishment of L1 phonetic system and assimilation of L2 phonetic contrasts. Next, I turn to a frequently cited study that focused on non-language factors.

2.4.3.5. Negative Ethnic Stereotyping

Investigating if the perceived poor quality of non-native-English-speaking international teaching assistants (NNES ITAs) was due to their limited English proficiency or students’ negative stereotyping, Rubin (1992) replicated Rubin and Smith (1990) and conducted three studies with 62 North American undergraduates. In Study 1, they were asked to listen to a short lecture recorded in Standard American English with the projection of a photograph of either a Caucasian or a Chinese woman and then complete a cloze test based on the speech. Next, they responded to a questionnaire on ethnicity and accent of the speaker and her teaching competence. As expected, the results showed that the speech accompanied by the Chinese photograph was perceived to be more oriental and with a more foreign accent than that accompanied by the Caucasian photograph. The results of listening comprehension indicated linguistic stereotyping. In other words, when the participants visually identified the instructor as Chinese, they perceived an accent that did not exist and assumed that they would have difficulty
understanding the instructor. In Study 2, the same participants were asked to listen to a short lecture with a strong Chinese accent, a moderate Chinese accent, or a Standard American English accent accompanied by the photograph of an Asian teaching assistant, a Caucasian teaching assistant, or no photograph at all and then responded to the same questionnaire used in Study 1. The results showed that those who had taken more courses from NNES ITAs had significantly higher comprehension scores, which was consistent with the finding in Rubin and Smith (1990). In Study 3, a training program facilitating positive intercultural contact was implemented to mitigate undergraduates’ attitudes toward NNES ITAs. Fifteen undergraduates observed two classes taught by NNES ITAs and recorded their feedback to each class based on a structured observation sheet. After each session, they met with the ITAs and discussed the teaching together. The results showed that the intervention was not effective. On the contrary, there was evidence indicating that some of the student mentors became more critical of the ITAs’ communication behaviors and only one quarter of the student mentors showed increased intercultural sensitivity.

The negative ethnic stereotyping found in this study is prevalent in the U.S., as described by Lippi-Green (1997). The significant different perception of the same accent by seeing the photographs of the Asian and the Caucasian by the North American undergraduates showed their bias against Asian but not white speakers of English, which affected their perception of accentedness and their willingness to understand the speaker. That is, even if Asian ITAs speak Standard American English, they will still suffer the stigmatization by North American undergraduates just because they are non-White. A potential limitation of the study might be that the undergraduates were in a southeastern
university. It is very likely that they had a southern accent which is different from the
accent of the Ohioan speaker who recorded the short lectures. It might explain why they
perceived an accent in the short lectures even when they saw the photograph of a
Caucasian female.

In sum, as mentioned above, aside from age effects and L1 transfer, social factors
are variables that make the investigation of L2 phonology acquisition complicated.
Moreover, it is likely that there are other confounding factors that might play a role in L2
phonology acquisition. The studies reviewed also showed the imbalanced focus on
speakers of certain languages only, with most of them focusing on Italian speakers,
followed by speakers of Korean, Japanese, Chinese, and Slavic languages. It suggests that
future research should include native speakers of other languages. In addition, all of the
studies reviewed so far have been conducted in ESL settings, which shows the scarcity of
research in EFL settings. Moreover, the construct validity appeared to be threatened in
Piske, MacKay, and Flege (2001) as well as in Flege, Yeni-Komshian, and Liu (1999) as
the participants were asked to repeat sentences after the recordings and thus had the
opportunity to imitate the recordings. Future research should adopt methodology that
ensures construct validity. Furthermore, the frequency of either L2 input or L1 use in
current research has been self-reported by the participants, which is unreliable and
problematic. Although Flege suggested that to accurately measure L2 input or L1 use
seems impossible due to practical and ethical limitations, a systematic method to measure
participants’ amount of L2 input and L1 use is needed to accurately examine the roles of
variables. Also, the inter-rater reliability in Munro, Flege, and MacKay (1996), Derwing,
Flege, Frieda, and Nozawa (1997), and Thomson and Munro (2006) is not unquestionable

with the use of heterogeneous raters, which suggests the necessity to present all productions to all raters. Finally, I concur with Flege (2009) in that quality of L2 input is crucial for acquisition of standard spoken English. On the other hand, how to measure the quality of input is another thorny issue.

In the next section, I turn to studies related to Chinese speakers’ perception and production of English segments.

2.5. Empirical Research on Chinese Speakers’ Perception and Production of English Phonetic Segments

Although the four language skills are interrelated and integrated, speaking skills and listening skills are more closely interrelated and mutually reinforcing. With respect to the interrelation between perception and production, most researchers suggested that perception of L2 phonemic contrasts should occur before production (e.g., Fledge, 1995b; Eckman et al., 2009, Celce-Murcia et al., 2010), which is opposite to Bohn and Flege’s (1996) claim that production of L2 sounds is easier than perception of the same sounds. Baker and Trofimovich (2006) noted that the majority of L2 speech learning theories is based on the assumption that perception and production are related and that accurate perception lays a foundation for accurate production. Before production of a second language, naturally the learner has to perceive and identify phonemes of the language. Accurate perception of the target phonemes is crucial for accurate pronunciation.

In this section, I turn to difficulties Chinese speakers might encounter in acquiring English segments. First of all, I briefly overview the results of studies related to Chinese learners reviewed in the previous section.
Identifying English stop consonants occurred in final position, the Chinese participants in Flege and Liu (2001) obtained higher scores for the stimuli containing a release burst than for the edited stimuli without release bursts, which can be explained by the fact that a release burst is easier to perceive. Investigating L1 linguistic influence on final /t/ /d/ deletion of 20 adult Chinese learners of English from both China and Taiwan, Bayley (1996) found that final /t/ /d/ were less likely to be deleted following a liquid than an obstruent or nasal and that the more careful the style, the less likely final /t/ /d/ were deleted. Unlike native speakers of English, the Chinese participants were more likely to delete inflectional final /t/ /d/ clusters than lexical ones, which might be explained by the fact that Mandarin is “an isolating language with practically no grammatical morphology” (Li & Thompson, 1978, p. 224). This lack of morphology in Mandarin could be easily transferred to L2 phonology, especially at discourse level. Eliciting production of 10 English vowels in CVC context, Munro and Derwing (2008) found that their adult Chinese participants’ performance on /ɪ/ after a year was still unsatisfactory, which might be explained by the Speech Learning Model in that since there is no /i/-/ɪ/ contrast in Mandarin, the participants assimilated /ɪ/ to the phonemic category in Mandarin, which is similar to English /i/. Investigating the role of native speaker input in L2 learners’ English pronunciation and fluency development, Derwing, Thomson, and Munro (2006) found the Chinese adult immigrants showed a little improvement in accentedness but showed no improvement in fluency over the 10-month period, the latter might be explained by a ceiling effect. Now I turn to more empirical research on Chinese speakers’ perception and production of English segments.

Contrary to Bayley’s (1996) finding that the Chinese participants were more likely
to delete inflectional final /t\ /d\ clusters than lexical ones, Hansen (2001) found that her three female Chinese participants were more likely to delete lexical final /t\ /d\ clusters than inflectional clusters during her six-month study. The analysis of two sets of data collected six months apart showed that both grammatical conditioning and both preceding and following linguistic environments affected the production of codas. The most difficult final consonants for the Chinese participants were the liquid /l/, voiced fricatives (e.g., /z/), and the voiceless interdental fricative /θ/. With regard to the favored production modification, for single final consonants, the participants favored articulation change such as place assimilation. For two-consonant final clusters, epenthesis was favored, while for three-consonant final clusters, deletion was favored. In terms of L1 transfer, equivalence classifications might have resulted in positive transfer in the participants’ production of English segments which have counterparts in Mandarin such as voiceless fricatives and stops. However, positive transfer was not found in their production of /n/ in words comprised of four or five syllables such as pronunciation and comprehension, in which deletion was favored. On the other hand, markedness constraints might have accounted for the low accuracy rates of the two- and three-consonant clusters, which do not exist in Mandarin.

Different from previous experimental studies, the speech data were collected from the participants’ spontaneous production in one-on-one interviews at a six-month interval, which allowed the researcher to examine the participants’ L2 acquisition over time with their naturalistic speech. On the other hand, the sample size in the study, composed of only three participants, was definitely too small. Moreover, insufficient third person verbs were found in the data. Thus the generalizability of the finding of the tendency to delete
lexical final /t/ /d/ than inflectional clusters should remain tentative. Finally, it appeared that it is not easy to identify whether a constraint is due to L1 transfer or markedness effects.

Since there are initial but no final stops in Mandarin (Li & Thompson, 1978), aside from deletion of both voiced and voiceless stops (e.g., Bayley, 1996), several studies have shown that Chinese speakers of English devoiced /b/, /d/ and, /g/, employed epenthesis, produced duration vowels preceding /p/ and /b/ with the same duration, did not actively enlarge the oral cavity to sustain voicing in /b/, and did not produce /p/ with a greater force of labial constriction than /b/ (Tarone, 1980; Eckman, 1981; Anderson, 1983, 1987; Flege & Davidian, 1985; Heyer 1986; Weinberger, 1987; Flege, McCutcheon & Smith, 1987; Flege et al., 1987; Flege, 1988; cited in Flege & Wang, 1989).

Proposing that non-native speakers of English will use particular acoustic cues when identifying non-native vowel contrasts even if the acoustic cues are not used to differentiate contrasts in their native language, Bohn (1995) asserted that the use of duration as a cue in vowel perception by non-native speakers of English cannot be explained solely by L1 transfer. Both Spanish and Mandarin do not use duration contrasts to differentiate segmental contrasts. However, the results of two experiments showed that Spanish participants predominantly, and Mandarin participants almost exclusively, relied on durational cues to differentiate the English /i/-/ɪ/ contrast, which could not be explained by L1 transfer. Moreover, most of the Mandarin participants and many of the Spanish participants did not use spectral cues to differentiate the non-native vowel contrast. To explain the findings, Bohn proposed the “Desensitization Hypothesis”, which holds that “whenever spectral differences are insufficient to differentiate vowel contrasts
because previous linguistic experience did not sensitize listeners to these spectral differences, duration differences will be used to differentiate the nonnative vowel contrast” (pp. 294-295).

As mentioned in Chapter 1, Taiwanese EFL students are usually taught to rely on durational cues to differentiate the English /i/-/ɪ/ contrast and thus neglect the spectral differences in a particular area of the vowel space. According to Bohn’s study, duration cues in vowel perception are easy to access regardless of listeners’ language-specific experience with them. The question is how to sensitize ESL/EFL learners to spectral differences in the /i/-/ɪ/ contrast in addition to temporal cues in identifying the structurally similar English tense and lax vowel contrast.

Examining 15 Beijing Mandarin speakers’ production and perception of English vowels, Wang (1997) found that the participants’ productions of the vowels that have Mandarin counterparts were native-like while their productions of the vowels that do not exist in Mandarin were not indistinguishable to native English listeners. Acoustic analysis showed that the participants’ productions of the vowels that do not have Mandarin counterparts deviated from the native English norms in terms of spectral values, which caused their low intelligibility scores. Especially, the Mandarin speakers’ /ɛ/ and /æ/ were mostly confused with each other. The results of the perception task of identifying five English front vowels /i/ /ɪ/ /e/ /æ/ /ɛ/ in bVt showed that the majority of misidentified /ɪ/ were identified as /i/ and they systematically substituted /i/ and /u/ for the /ɪ/ and /ʌ/ categories.

The finding that the majority of misidentified /ɪ/ were identified as /i/ in Wang was later confirmed in Munro and Derwing (2008) in that their adult Chinese participants’
performance on /i/ after a year was still unsatisfactory, which lends support to the Speech Learning Model in that since there is no /i/-/ɪ/ contrast in Mandarin, the participants assimilated /i/ to the phonemic category in Mandarin which is similar to English /ɪ/. Moreover, the finding that the participants’ productions of the vowels that do not have Mandarin counterparts deviated from the native English norms in terms of spectral values in acoustic analysis confirms Bohn’s (1995) finding that native English speakers rely predominantly on spectral cues rather than temporal cues to identify the contrasts as opposed to Mandarin speakers, who rely almost exclusively on temporal cues. On the other hand, the /u/-/ʊ/ contrast trained in the study are considered unimportant by Brown (1988). He contended that since there are only four /u/-/ʊ/ minimal pairs in modern English (pool/pull, fool/full, who’d/hood, suit/soot), misunderstanding is highly unlikely to occur for these contrasts. However, perception and production of the contrast is worth training in terms foreign accent reduction.

The results of Wang’s (1997) study appeared to support Bohn’s (1995) Desensitization Hypothesis, which proposed that L2 learners will resort to duration differences to differentiate non-native vowel contrasts when there are insufficient spectral differences in the L1 to sensitize L2 learners to spectral differences. However, in a later study, Wang’s (2006) findings did not fully confirm Bohn’s theory. Wang argued that Bohn’s Desensitization Hypothesis is based on only one pair of English vowels contrast (/i/-/ɪ/) and her pervious study with Munro (Wang & Munro, 1999) found that although the native Mandarin speakers relied predominantly on duration cues for the /i/-/ɪ/ contrast, they did not relied predominantly on duration cues for the /u/-/ʊ/ contrast, which did not confirm Bohn’s hypothesis. Wang decided to test Bohn’s hypothesis in a perception test
in which synthesized vowel continua manipulating vowel spectral and temporal cues were used to investigate 17 native Mandarin speakers’ and 6 native English speakers’ perception patterns in perceiving the /i/-/ɪ/, /u/-/ʊ/, and /ɛ/-/æ/ contrasts. The results showed that the native English speakers relied predominantly on spectral cues when identifying all of the three contrasts. Although the Mandarin speakers also showed sensitivity to spectral cues, it was not close to that of native English listeners for any of the three target vowel contrasts. In addition, the Mandarin speakers showed little sensitivity to the duration cues for the /u/-/ʊ/ and /ɛ/-/æ/ contrasts, which was more like the pattern observed in the native English speakers. Finally, the Mandarin speakers relied significantly more on duration cues only for the /i/-/ɪ/ contrast. The results showed that the Mandarin speakers applied different strategies in identifying the three vowel contrasts, which did not fully support Bohn’s hypothesis.

The finding that the majority of the Mandarin speakers relied on neither spectral nor duration cues systematically to distinguish the /u/-/ʊ/ and /ɛ/-/æ/ contrasts suggested that they had not established category distinctions for the contrasts, which warrants perceptual training in identifying these contrasts. Moreover, the Mandarin speakers’ predominant reliance on duration cues for the /i/-/ɪ/ contrast also showed the need to train Mandarin speakers to rely on spectral cues as native English speakers do.

Investigating L1 phonotactic constraints, Flege, Munro, and Skelton (1992) examined perception and production of the word-final English /t/-/d/ contrast by 20 Chinese learners of English and found that even experienced learners could not produce the English vowel duration difference preceding /t/-/d/ accurately and only one Chinese (inexperienced) learner identified the English word-final /t/-/d/ contrast. One the other
hand, the larger number of experienced than inexperienced L2 learners who mastered the English voicing effect in the word-final /t/-/d/ was consistent with the hypothesis that the phonetic contrast between English /t/-/d/ is learnable.

Taiwanese permits /p/, /t/, /k/ but not /b/, /d/, /g/ in the word-final position (Cheng, 1968), while there are only two final consonants, /n/ and /ŋ/, in Mandarin (Li & Thompson, 1978). Therefore, it is important to verify the first language of the participants to investigate phonotactic constraints of different native languages. Although a native Mandarin research assistant verified that all of the 10 participants from Taiwan spoke Mandarin as their native language, judging by the demographic distribution in Taiwan, it appears that it would have been more appropriate to ask the Taiwanese participants about their first language. On the other hand, the failure of producing the difference of vowel durations preceding final /t/ /d/ by most of the participants suggested that this feature should be incorporated into pronunciation teaching to make learners aware of the difference.

Wang (1995, cited in Broselow et al., 1998) found epenthesis, deletion, and devoicing of English stop consonants in coda position by Mandarin speakers. Broselow, Chen, and Wang (1998) discussed the patterns within the frame of McCarthy and Prince’s (1993, 1995) as well as Prince and Smolensky’s (1993) Optimality Theory (OT) with the Markedness Differential Hypothesis (Eckman, 1977). OT chooses the best surface constraint according to input, whereby a function called GEN (p. 267) generates a set of possible forms for surface pronunciation. Which form a speaker chooses depends on which constraint is stronger in the speaker’s language. OT posits that “a bisyllabic word is optimal because every syllable can be parsed into a binary foot” (p. 271). The
researchers used this universal constraint of word binarity to account for a set of related constraints that combine to favor bisyllabic words. For instance, since there are no stops in coda position in Mandarin, Chinese learners of English tend to use either epenthesis or deletion to move the stop out of coda position. On the other hand, due to preference for bisyllabic words, epenthesis should be preferred over deletion in monosyllabic input forms of the final consonant because epenthesis will transform a monosyllabic form to a bisyllabic form, whereas deletion should be preferred over epenthesis in bisyllabic input forms of the final consonant because deletion will move the stop out of coda position but still maintain a bisyllabic form. In other words, an input bisyllabic form is most likely to resist the epenthesis for it will transform a bisyllabic form to a trisyllabic form, which is against the word binarity constraint. The theory is supported by the findings in Wang (1995). With respect to devoicing of English stops in coda position by Mandarin speakers, Broselow et al. noted that voiceless codas are universally less marked than voiced codas, which accounts for the Mandarin speakers’ devoicing of English stops in coda position in Wang (1995). According to OT, the constraint of no voiced obstruent coda in Mandarin is not visible because Mandarin forms are subject to the more general constraint of no obstruent coda, a higher ranking constraint. Broselow et al. argue that the sudden visibility of such constraint in second language acquisition may be caused by extensive input of target language or from the development of an interlanguage grammar that is different from the L1 grammar in its constraint rankings.

As an experienced EFL teacher in Taiwan, I also found instances of epenthesis in both my students and even in some of my colleagues when they produced words such as *net*, *gap*, and *hit*. According to Optimality Theory, since there are no obstruents in coda.
position in Mandarin and because of the universal constraint of word binarity, it makes sense that those speakers used epenthesis to make themselves more comprehensible when speaking Mandarin mixed with English words. In the remainder of this section, I turn to studies that focus on Taiwanese learners of English.

Exemplifying hypercorrection, Decamp (1972) noted that Mandarin speakers in Taiwan generalize the rule of the retroflection of certain lexically-marked morphemes of Mandarin to any noun. It appeared that Decamp assumed that Mandarin speakers in Taiwan can produce /r/ without difficulty. However, Decamp did not specify the first languages of the speakers in his study.

Employing a contrastive approach, Paolillo (1995) noted that Mandarin has distinct /r/ and /l/ phonemes. However, they contrast only in initial position but not in medial or final positions. In contrast, English has distinct /r/ and /l/ phonemes in all three positions. Investigating the distinction between /r/ and /l/ in five positions among Mandarin speakers, Paolillo recorded 20 Taiwanese graduate native Mandarin speaking students in a U.S. university. The five positions were (1) postconsonantal (C_), (2) word-initial (#_), (3) intervocalic (V_V), (4) syllabic ([r]and [l]), and (5) postvocalic (_C/#). Paolillo found four different degrees of neutralization (e.g., both class and crass are pronounced the same as crass) in the same position among different participants, ranging from complete neutralization to no neutralization at all. He also found more instances of subset neutralization where r-type phones were shared (82%) than those where l-type phones are shared (18%). Thus, he concluded, “Mandarin has r-like phones in all the relevant positions, which we expect to transfer and to be older than the l-type phones which have to be learned in most environments” (p. 281). However, there was an overwhelming
tendency for deletion of both postvocalic /r/ and /l/ due to low phonetic saliency. In the study, postvocalic /r/ and /l/ were the most difficult position for Mandarin, Japanese, Korean, and Indonesian speakers due to lack of a phonemic contrast and were the most marked. Therefore, postvocalic /r/ and /l/ were the most difficult for these speakers to learn based on Eckman’s (1977) Markedness Differential Hypothesis. Paolillo proposed to add one important qualification to Eckman’s hypothesis by arguing, “learners may begin acquiring the more frequent contrasts earlier, even if they are more marked crosslinguistically” (p. 289) and called this different sense of markedness “language-specific markedness,” which refers to “the degree of functional load that a contrast has in a particular environment in a particular language, and consequently to how early or late a learner may begin learning that contrast” (p. 288). In contrast, Eckman’s crosslinguistic markedness refers to the inherent difficulty of a contrast in a particular environment and dictates the relative order in which a learner may complete learning, in this case, the /l/-/r/ contrast in different environments. However, although learners may begin acquisition of more frequent, and more “marked” contrasts based on Paolillo’s language-specific markedness, Paolillo contended that the process of acquisition will not be complete until the crosslinguistically less marked contrasts are fully acquired.

Like Decamp (1972), Paolillo (1995) assumed that the Taiwanese participants in his study were native speakers of Mandarin and did not take their first languages into consideration. The overwhelming tendency deletion of postvocalic /r/ in Paolillo’s study contradicted Decamp’s assumption that Mandarin speakers can produce /r/ without difficulty, which might in part be due to the possibility that some of the participants are not native speakers of Mandarin. Even if they are, they might choose to drop the final /r/
in Mandarin (Cheng, 1973) or do not use the final /r/ at all, as in Taiwan Mandarin (Kubler, 1985). Thus, whether Mandarin speakers can pronounce postvocalic /r/ easily still remains a question and which variety of Mandarin is used by speakers should be taken into account. Moreover, Paolillo did not investigate the difference between word medial postvocalic /r/ and word final postvocalic /r/ among the Taiwanese students, which is still a gap in the literature.

Analyzing errors for the /r/ sound in the initial position (e.g., room), the medial position (e.g., front), and the final position (e.g., their) among Taiwanese senior high school students, Lee (2006) found the final retroflection used in Mandarin Chinese does not guarantee the correct use of /r/ in the word-final position within an English word and some of the participants adopted deletion of the final /r/, which is consistent with the results in Paolillo’s (1995) study. Although Lee studied postvocalic /r/ in the word-final position, she did not investigate postvocalic /r/ in the word-medial position. In addition, as mentioned above, the final /r/ is rarely seen in Taiwan Mandarin (Kubler, 1985).

All of the researchers above assumed that Mandarin speakers can produce the final retroflection in Mandarin. However, Cheng (1973) noted long time ago that many Mandarin speakers drop the final /r/ in Mandarin, which may affect their production of final /r/ in English. Moreover, the Mandarin spoken in Taiwan differs considerably from that spoken in mainland China in syntax, phonology, and lexicon due to language contact with Taiwanese (Kubler, 1985). Taiwanese is spoken by approximately 80% of the population in Taiwan (Cheng, 1985) and Taiwanese speakers have difficulty producing retroflex sounds because there is no retroflex sound in Taiwanese (Kubler, 1985). Due to the language policy in Taiwan, Taiwanese has been degraded and stigmatized but now
revitalized. Nowadays, most people in Taiwan are bilingual in both Taiwanese and Mandarin. Today, nearly all news broadcasters in Taiwan drop final /r/ when speaking Mandarin. Therefore, it cannot be assumed that Mandarin speakers in Taiwan can produce final /r/ without difficulty. Conversely, it is even safe to assume that most Taiwanese do not use final /r/ when speaking Mandarin. The interference of Taiwanese on the pronunciation of English was noted by Ing (1988), who indicated that native speakers of Taiwanese substitute /h/ for /f/ in *force*, /s/ for /θ/ in *thick*, and /l/ for /ð/ in *that*, which is not exactly the same as native speakers of Mandarin. Since Chinese speakers from different language backgrounds might have different L1 transfer, it is imperative to investigate L1 backgrounds of Chinese participants.

The English interdental fricatives /θ/ and /ð/ are often identified by learners of English as the most difficult segments (e.g., Derwing & Rossiter, 2002; Derwing, 2003) and previous studies have shown that Chinese speakers from different backgrounds substitute different variants for English /θ/: Cantonese-speaking children in Canada substitute /s/ for initial /θ/ and /z/ for final /s/ (Wang & Geva, 2003). In contrast, Cantonese speakers in Hong Kong substitute /f/ for /θ/, whereas Mandarin speakers in China and Taiwan substitute /s/ for /θ/ (Weinberger, 1994, cited in Rau, Chang, & Tarone, 2009). Peust (1996, cited in Rau et al., 2009) also confirmed the findings in Weinberger with Cantonese speakers in Hong Kong and Mandarin speakers in Taiwan, whereas Malaysia/Singapore Chinese substitute /t/ for /θ/.

Employing a quantitative variationist framework, Rau, Chang, and Tarone (2009) examined production of the English voiceless interdental fricative /θ/ by Mandarin speakers, 11 from China at a U.S. university and 16 undergraduates at a university in
Taiwan. They were asked to evaluate the acceptability of four possible substitutes for /θ/ and then perform four oral production tasks. As expected, the results showed that the participants from both Taiwan and China rated /s/ as the most acceptable substitute for /θ/.

In their actual production, all the participants from China substituted /s/ for /θ/ (99%), with /t/ and /ʃ/ accounting for only 1% of the substitutes. In contrast, although most of the Taiwanese participants also substituted /s/ for /θ/ (86%), there was a wider range of variation of substitutes for /θ/, including /s/, /z/, /ʃ/, /ɬ/, /d/, /t/. The results also indicated that immediate phonetic environment and speech style accounted for accurate production of /θ/. Moreover, those who could produce /θ/ with higher accuracy employed monitoring strategies, whereas those who had lower production accuracy depended on phonetic salience strategies. The findings support the markedness principle. For instance, the onset or coda with /r/ (#thr or rth#) inhibited the accurate production of /θ/ because they are more marked.

As we can see, Chinese speakers from different language backgrounds substitute different variants for English /θ/. Therefore, it is imperative to verify participants’ first language to understand their linguistic variation. Although Rau et al. claimed that all of the Taiwanese participants shared the same L1, Mandarin, according to their language background data (p. 589), three of the participants reported that they were bilingual in both Mandarin and Taiwanese. Judging by the reported high percentage of native speakers of Taiwanese in Taiwan (Cheng, 1985), it is unlikely that all of the 16 Taiwanese participants were native speakers of Mandarin and spoke standard Taiwan Mandarin. Although in the case of /θ/ native speakers of Taiwanese would also substitute /s/ for /θ/ (Ing, 1988), there is no guarantee that native speakers of Taiwanese will not show
different variations in other cases. In addition, the wider range of variation of substitutes for /θ/ among the Taiwanese participants might be explained by the possibility that some of them might speak other languages as a first language. This again shows the importance of verifying participants’ first language. Finally, the English interdental fricatives are often identified by learners of English as the most difficult segments (e.g., Derwing, 2003). However, it has been argued that they are not crucial sounds because they carry a low functional load and the inability to produce them is less likely to cause a communication breakdown (Brown, 1991; Catford, 1987, cited in Derwing, 2003). In fact, the English interdental fricatives are excluded in the Lingua Franca Core proposed by Jenkins (2002, 2007) in which English is used as a lingua franca among non-native speakers of English from different L1 backgrounds. However, whether the exclusion of English interdental fricatives would obstruct successful communication among non-native English speakers has not been widely tested yet.

The results of all the studies reviewed in this section are invaluable when it comes to teaching Chinese speakers English pronunciation. Hansen’s (2001) method of collecting speech data from participants’ spontaneous production in one-on-one interviews at a long interval allows us to examine participants’ L2 acquisition over time with their naturalistic speech. However, it will take huge amount of time and efforts if the researcher wishes to collect data from a large sample. Moreover, despite the fact that the Chinese participants might speak different first languages, the researchers (Decamp, 1972; Flege, Munro, & Skelton, 1992; Paolillo, 1995; Flege & Liu, 2001; Rau, Chang, & Tarone, 2009) did not take the possibility of different L1 phonotactic constraints into account. One exception is Bayley (1996), who noted the high prestige of Mandarin in Taiwan and questioned the
participants’ self-reported status as native-speakers of Mandarin. Since Chinese speakers of different language backgrounds might have different L1 transfer, it is imperative to verify their L1 before examining their perception and production of English segments. On another hand, the variationist framework employed by Rau, Chang, and Tarone (2009) appears to be a superior framework to examine the production of learners of English. Finally, the finding that the majority of the Mandarin speakers relied on neither spectral nor duration cues systematically to distinguish /u/-/ʊ/ and /ɛ/-/æ/ contrasts in Wang (2006) warrants perceptual training in identifying these contrasts. Moreover, Mandarin speakers’ predominant reliance on duration cues for the /i/-/ɪ/ contrast, as found in Wang (1997, 2006) and Bohn (1995), also shows the need to train Mandarin speakers to rely on spectral cues as native English speakers do.

2.6. Empirical Research on English Segments Identification Training

Investigating whether L1 phonotactic constraints influence how well Chinese with different L1s perceive the word-final English /t/-/d/ contrast and whether sensitivity to the contrast can be improved by a small amount of feedback training, Flege and Wang (1989) examined identification of English word-final /t/-/d/ contrast by 27 adult Chinese speakers with different length of residence and grouped by their L1s: Mandarin, Shanghainese, or Cantonese. The /t/-/d/ minimal pairs were beat, bead, bet and bed but closure voicing and release burst cues had been removed by editing. The participants were trained by drawing their attention to acoustic cues such as preceding vowel duration and providing feedback in multiple contrasts. The performance of the three Chinese groups was compared before, during, and after the feedback training. The Cantonese
participants were expected to perform best because their L1 permits unreleased /p/ /t/ /k/ in word-final position, whereas the Mandarin participants were expected to perform most poorly because Mandarin does not permit word-final obstruents. The Shanghainese participants, whose L1 permits a final glottal stop, were expected to perform between the other two groups. As expected, the results showed that the Cantonese participants were significantly more sensitive to the English /t/-/d/ contrast than the other participants. The participants in all of the three groups showed a significant increase in sensitivity as a result of the training in that they became better at making use of acoustic cues such as preceding vowel duration and quality. The results showed that native-language phonotactic constraints influenced how the participants with different L1s processed syllables. An unexpected finding is that three variables related to amount of L2 experience (length of residence in the U.S., years of instruction in English, percentage of daily use of English) were found to be significant predictors of the identification scores.

The effectiveness of the training program suggests that by drawing learners’ attention to acoustic cues such as preceding vowel duration and quality and providing feedback in sufficient minimal pairs, it is possible for adult L2 learners to achieve native-like sensitivity in identifying the /t/-/d/ contrast. On the other hand, although the analysis supported the researchers’ belief that the difference between the Cantonese and Mandarin groups was due to L1 syllable structure differences instead of L2 experience, the facts that the Cantonese participants had lived in the U.S., had received English instruction longer than the other groups, and reported higher daily use of English still posed potential threats to the validity of the study. It appears that such a Cantonese sample could have been avoided because of all the confounding factors.
Flege (1989) also conducted a similar study with the inclusion of 5 native speakers of Taiwanese. Since Taiwanese also permits unreleased /p/ /t/ /k/ in word-final position, the Taiwanese speakers were expected to perform better than the Mandarin speakers. The results showed that the training effect was indeed influenced by L1 background: the southern Chinese participants had the greatest increase in correct identification (21%), followed by the Taiwanese speakers (11%), and the Mandarin speakers had the least increase (5%). The results supported the hypothesis in Flege and Wang (1989) that native-language phonotactic constraints influence how the participants with different L1s processed syllables.

Although the training effect was indeed influenced by L1 background, the sample size of the Taiwanese speakers in the study ($n = 5$) is relatively small. Furthermore, the majority of population in Taiwan is bilingual in both Mandarin and Taiwanese. In other words, they are native speakers of Mandarin and Taiwanese. Research on how native Mandarin-Taiwanese speakers process English syllables is still scarce in the literature.

The effectiveness of the identification training in Flege and Wang (1989) motivated Flege (1995a) to hypothesize that identification training might be superior to same/different training. Twenty adult Mandarin speakers were randomly assigned to an identification group or a same/different group to receive 7 sessions (approximately 15 minutes each session) of feedback training on identifying final the /t/-/d/ contrast during 3 weeks. The stimuli, beat, bead, bet and bed with closure voicing and release burst cues removed, were the same as in Flege and Wang (1989). As the finding in Flege and Wang (1989), the participants frequently misidentified the word-final /t/-/d/ tokens in the edited stimuli. Pretests were administered before each training session and two posttests were
held after the completion of training, with the first one held two weeks after the training and the follow-up posttest held two months after the training. The first pretest showed that the two groups had the same score. Significant increases in the percentage of correct identification of the word-final /t/-/d/ contrast were obtained for both groups while scores of the identification group after the first training session and all subsequent training sessions were higher than those of the same/different group. However, the gain difference between the groups was not significant and the same/different group had better but not significant retention than the identification group in the follow-up posttest. Results also showed that the effects of training were generalized to words that were not used in the training.

Although the gain difference between the groups was not significant, the identification group did obtain scores higher than the same/different group throughout the training and the participants in the former group enjoyed their training more and were more willing to receive additional training than those in the latter group. It is very likely that the gain difference might have been significant had the training time been extended. In addition, lack of a control group did not eliminate the threat to internal validity for the increases in the percentage of correct identification for both groups might have been due to education in an English environment instead of the perceptual training. Moreover, Flege focused on the bVt-bVd contrast and did not incorporate word-final /t/-/d/ stimuli with word initials other than /b/ in the training. Neither did he investigate the generalizability of the perceptual training effect on identifying the word-final /t/-/d/ contrast to other final stop contrasts, i.e., /p/-/b/ and /k/-/g/. Furthermore, Flege’s study was conducted in an ESL setting instead of an EFL setting, which is more in need of
effective methods for perceptual training. Finally, it appears that Flege could have investigated whether the effects of perceptual training can be transferred to production by including a production task.

Japanese speakers’ production and perception of English /r/ and /l/ is probably the most studied areas in the L2 acquisition of English phonology because it is a difficult non-native phonemic contrast. Lively, Logan, and Pisoni (1993) conducted two experiments on training Japanese listeners to identify the English /r/-/l/ contrast. In Experiment 1, six Japanese speakers who had arrived in the U.S. for months prior to the experiment were trained in an identification task with English words containing the /r/-/l/ contrast in initial singleton, initial consonant clusters, and intervocalic positions recorded by five native talkers for three weeks. The results showed a significant identification improvement from the pretest to the posttest and the effect of training was generalized to stimuli with new phonetic environments and new talkers. In Experiment Two, another six Japanese speakers were trained by the same procedure with more stimuli in five phonetic environments recorded by only one native talker for the same amount of time as in Experiment 1. The results showed that although the participants made improvement after the training, they did not generalize the effect of training to either new tokens or new talkers.

The most obvious validity threat of this study is that there was no random assignment. Six Japanese speakers were recruited for Experiment One and another six Japanese speakers were recruited for Experiment Two. It can be argued that the participants could have been recruited at the same time and randomly assigned to one of the two experimental groups. Another validity threat is that more stimuli and more
phonetic environments were used in Experiment Two. Thus the two experimental groups were trained under different conditions. It appears that the researchers could have held the other variables constant (i.e. using the same number of stimuli and the same phonetic environments) and focused solely on talker variability. Nonetheless, the results of the study suggest that high talker variability plays a crucial role in establishment of new phonemic categories as well as elimination of talker effects.

Lively et al. (1994) replicated the previous study and focused on long-term retention of the effect of perceptual training. 19 Japanese speakers were recruited to receive perceptual training and another 23 Japanese speakers were recruited as a control group. After the pretest, the experimental group was trained in a two-alternative identification task on a computer with English words containing the /r/-/l/ contrast in five phonetic environments recorded by five native talkers. A posttest was administered to both groups after completion of the training. Two follow-up posttests were administered to the experimental group three and six months after the completion of the training. Two generalization tests were also administered to the experimental group in the posttest and two follow-up posttests. The results showed that, consistent with previous findings, the high-variability training procedure proved to be effective in improving the Japanese speakers’ perception of the non-native phonemic contrast and the effect of training was retained half a year after the training.

Again, the most obvious validity threat of this study is the lack of random assignment. It appears that the participants could have been recruited at the same time and randomly assigned to the experimental group or the control group. Another validity threat is that fewer participants from the experimental group returned for the two
follow-up posttests, with 16 out of 19 in the first follow-up posttest and only eight out of 19 in the second follow-up posttest. If the majority of the participants could not return for the second follow-up posttest, the second follow-up test could have been cancelled. Moreover, it could have been overwhelming or even boring for the experimental group to take the same test four times (one pretest and three posttests) as well as to take the same generalization test three times, which might explain why only eight out of 19 participants from the experimental group showed up for the second follow-up posttest. Nonetheless, the high-variability training procedure appeared to be effective in establishment of non-native phonemic categories and long-term retention.

Bradlow, Akahane-Yamada, Pisoni, and Tohkura (1999) continued investigating the long-term retention of training improvement in both perception and production of the English /r/-/l/ contrast. Eleven monolingual Japanese speakers attended 45 sessions for a total of 15 to 22.5 hours of training within a month. The high variability perceptual training was composed of 68 minimal pairs of the /r/ and /l/ contrast in five phonetic environments. During each session, a /r/-/l/ minimal pair appeared on a computer screen in written form and the stimulus was played. Then the participants were instructed to identify which word was presented. If the response was correct, a chime would be heard and a buzzer would be heard if the response was incorrect. Aside from identification test, the participants were also asked to produce 55 English /r/-/l/ minimal pairs to assess their improvement of intelligibility. A pretest and a posttest were administered. A follow-up posttest was administered three months after the training to assess their retention. The results showed that the participants made significant improvement in both perception and production of the contrast. The results of the three-month follow-up posttest showed that
the participants still retained their long-term improvements in both perception and production.

It appeared that the training was effective and by receiving feedback from the computer, the participants could learn at their own paces. The successful application of technology in L2 perceptual training suggested that computer software can contribute to minimal pair identification effectively. On the other hand, although perception might contribute to production, it appears that technology has made little progress in facilitating production. After all, speech production requires feedback much more complicated than just providing the right answer.

Cenoz and Lecumberri (1999) examined the effect of training on the discrimination of English vowels by Basque speakers and Spanish speakers. The participants were 109 university students in Spain. They were asked to complete questionnaires and vowel discrimination tests, which included eleven vowels and eight diphthongs. A total of 14 hours was devoted to aural discrimination training, which included listening and identification of English simple vowels, diphthongs, and consonants as well as some stress and rhythm exercises. The results showed that the training significantly improved the participants’ ability on perception of English vowels. In addition, the results indicated the desire to acquire a native accent was a significant predictor of improvement in phonetic discrimination.

Although the results showed that the training was effective, we cannot see the details of how the identification training was implemented. The researchers could have provided more details for replication. On the other hand, the finding that the desire to acquire a native accent was a significant predictor of improvement in phonetic discrimination
suggests that motivation, or investment (Norton, 1995, 2000), plays an important role in L2 acquisition.

Epenthesis and deletion have been found in Chinese learners of English in Wang (1995), Bayley (1996), and Hansen (2001). However, the researchers did not provide any solution to the problems. Couper (2006) attempted to offer a method to tackle the problems of 21 mostly Asian immigrant ESL students in New Zealand. First, the treatment group was given a diagnostic test to determine the focus of instruction, which turned out to be epenthesis and deletion. Then the treatment group received 12 30-minute sessions during two weeks to overcome their problems. The researcher believed that many learners were not aware of their pronunciation problems and could not distinguish the difference between their pronunciation and that of native speakers. Therefore, the instruction involved explicit explanation, modeling, finding patterns, practice, and real-time feedback to raise their awareness of the problems. The students were also asked to record themselves and compare their speech with the model. An immediate posttest and a follow-up posttest three months later were administered. The results showed that the treatment group significantly reduced error rate in the posttest and the error rate rose slightly in the follow-up post-test three months later. In contrast, the control group achieved no gains at all.

One limitation the researcher noted was that the practice activities were not communicative. Therefore, even though the treatment group dramatically reduced their errors on the segmental level, there is no guarantee for their improvement at the discourse level. Another limitation is that the researcher found that suprasegmental errors such as sentence stress and rhythm were too difficult to quantify, so he did not focus on
instruction in this area. However, it is a fact that most of the learners in the study had
difficulty on the suprasegmental level. Future research should explore effective
instruction in this area.

Saito (2007) also emphasized explicit pronunciation instruction and raising learner
awareness in his study. He noted that although a lot of research has been conducted on L2
speech acquisition, little attention has been given to the development of effective
strategies to address the problems. In his experimental study involving six Japanese
learners of English, four participants in the experimental group were given one-hour
explicit instruction through computer-generated visual feedback on English /æ/, while the
other two participants in the control group received no feedback at all. A pretest was
administered to all the participants and a posttest was administered to the treatment group
only immediately after the training. A delayed posttest was administered to all the
participants one week later. The results showed that the treatment group became more
aware of their speech and improved their pronunciation significantly, whereas the control
group made no improvement.

Although the training lasted only one hour, the explicit feedback appeared effective.
However, the sample size ($n = 6$) in this study was too small and jeopardized
generalizability and there was no signs of random assignment. With regard to technology
in assisting L2 pronunciation acquisition, although Celce-Murcia (2001) recommended
the use of Computer Assisted Language Learning (CALL) software to provide students
with corrective feedback, the accuracy of speech-recognition software is still unreliable
and even native speakers might get a low score due to different voice qualities. However,
with advances in technology, speech recognition software will definitely be much more
accurate and will play an important role in assisting L2 speech learning.

Investigating the effects of perceptual training on the perception and production, Wang (2002) trained 13 Mandarin and three Cantonese speakers in identifying the English /i/-/ɪ/, /u/-/ʊ/, and /ɛ/-/æ/ contrasts in a laboratory setting in Canada. In the perception test, the participants were asked to identify the three contrasts in both synthesized and naturally produced minimal pairs. In the production task, the participants were asked to read a list of words containing the target contrasts. A pretest was held as a baseline to measure the effects of the training. Identification tasks with immediate feedback were used for the training. Synthesized tokens from the endpoints of continua were used at the beginning, followed by spectrally less extreme tokens. Then the participants were exposed to highly variable natural tokens produced by multiple native English speakers. The results showed that the training shifted the participants’ attention on temporal cues to spectral cues for all the three contrasts and their scores increased significantly from the pretest to the posttest. Moreover, the effect of perceptual learning was generalized to new talkers and the effect persisted three months after the training. However, the effect of perceptual training on production was not significant in terms of the participants’ vowel duration differences within each contrast.

It appears that the high talker variability in this study was effective in shifting the Chinese participants’ attention on temporal cues to spectral cues. In my view, highly variable natural tokens are superior to synthesized tokens simply because they are naturally produced by native speakers instead of manipulated unnatural tokens. In addition, synthetic stimuli do not represent the full range of acoustic properties of specific phonetic categories (Flege, 1995a). As long as talker variability is maximized,
synthesized tokens are unnecessary. With respect to the effect of perceptual training on production, further research needs to explore whether the effect of perceptual training is not transferable to production.

So far we have not seen any study that focuses on training teachers to teach pronunciation. As mentioned above, research on training teachers to teach pronunciation is relatively rare and most ESL/EFL teachers lack training in teaching pronunciation. In response to a call for inclusion of a pronunciation component in ESL/EFL teacher preparation programs, the University of Kansas designed a seminar for teaching second language pronunciation for graduate students in both TESOL and foreign language education programs (Gonzalez-Bueno, 2001). The pilot course took into account the crucial role that speech perception plays and included components such as raising learners’ awareness in identifying and producing phonemic contrasts in context, employing communicative drills for perception and production, and application of certain sounds in contextualized activities.

As we can see, the training focused on teaching both segmental and suprasegmental features in contextualized situations. It also involved raising learners’ awareness about their speech by providing visual stimuli such as diagrams and spectrographic images. Moreover, it incorporated communicative drills to encourage perceptual and production practice first and then extended to free production in communicative activities at the discourse level. It appears that the pilot course should be a model for other TESOL and second language education programs that wish to include a pronunciation instruction course for their students. Next I turn to another teacher training study in an EFL setting.

Lee (2009) used duration manipulation to train pre-service Korean EFL teachers to
identify The English /i/-/ɪ/ and /u/-/ʊ/ contrasts in a quasi-experimental study. One native talker was asked to record 4 English high vowel stimuli, heed, hid, who’d, hood. Then vowel duration of each word was manipulated using Praat to create five different durations. Another three native talkers were asked to record new stimuli in two different vowel durations for later training. For example, the word feel was recorded with both a normal vowel duration as in dictation and a shorter vowel duration as in fast speech, whereas its contrast, fill, was recorded with a longer vowel duration as in emphatic speech as well as with a normal duration. The purpose of this manipulation of vowel durations was to sensitize learners to spectral differences in the contrasts in identifying the structurally similar English tense and lax vowel pairs because the results of a pilot study showed that the Korean participants mainly relied on duration as a cue to discriminate tense from lax vowels in English. The first experimental group (n = 49) received 60-minute instruction and took one 40-minute individual training and the second experimental group (n = 37) received 30-minute instruction only, while the control group (n = 25) received no instruction at all. During the perceptual training, the experimental groups were exposed to pair stimuli of the four words with different durations and were asked to identify the correct pair of words. For the production training, the experimental groups watched instructional video clips about how to pronounce the target words and then practiced pronouncing the words. The perceptual training also exposed the experimental groups to new words and new talkers. The results showed that both experimental groups’ posttest scores were significantly higher than that of the control group. However, there was no significant difference between the two experimental groups. Moreover, the overall production test scores were not significantly improved after
the formal pronunciation instruction. In terms of generalization, both experimental groups’ posttest perception mean scores were significantly higher than that of the control group. However, the experimental groups did not generalize their newly acquired knowledge into new words more than the control group in terms of production.

One of the strengths of the study is that the innovative idea of using two different natural vowel durations for perceptual training has been rarely seen in other related studies. It appears that if duration and talker variability are maximized, it would be redundant to use unnatural synthesized tokens. On the other hand, the /u/-/ʊ/ contrast trained in this study has been argued that they carry a low functional load and are considered unimportant by Brown (1988). Since there are only four /u/-/ʊ/ minimal pairs in modern English, misunderstanding is very unlikely to occur for these contrasts. Therefore, the /u/-/ʊ/ contrast is not the first priority in perceptual training. Moreover, the finding that there was no significant difference between the two experimental groups might be due to that fact that the training time (100 minutes vs. 30 minutes) was relatively short as opposed to other perceptual training studies. It is very likely that the difference between the two groups could have been significant if both groups had received a longer period of training. Furthermore, the little improvement in production might be also related to the short amount of training time. One fatal weakness in Lee’s study is that the minimal pairs were presented to the participants at the same time in the production list at pretest and posttest. Although there were only four words for production, the words could have been arranged in a way that no minimal pair would be presented at the same time.

All the studies reviewed in this section suggested that training learners to perceive
and produce English segments can be effective. As many researchers noted, learners can benefit from focused pronunciation instruction (e.g., Derwing, Thomson, & Munro, 2006; Couper, 2006; Derwing, 2008). However, as Derwing, Thomson, and Munro (2006) pointed out, many ESL teachers feel reluctant teaching pronunciation, usually due to a lack of training. Derwing and Rossiter (2002) conducted interviews with 100 immigrant ESL students in Canada and found that only eight of them had received explicit pronunciation instruction in the ESL classroom. It suggests that TESOL programs should be held accountable for preparing would-be ESL/EFL teachers for pronunciation instruction. The pilot course in Gonzalez-Bueno (2001) appears to be a good model for other teacher preparation programs. On the other hand, ESL/EFL teachers should be informed by research to improve their pronunciation instruction. The prerequisite is that research on training teachers to teach pronunciation should be encouraged by major English teaching and learning journals.

Since learners can benefit from focused pronunciation instruction, it is imperative to find the most effective method for training ESL/EFL learners to identify English segments, especially non-native phonemic contrasts. In the next section, I briefly introduce the sociolinguistic background in Taiwan.

2.7. Chapter Summary

According to Major’s (2001) extensive review of literature on the acquisition of L2 phonology, the majority of studies have indicated that the younger one acquires L2 phonology, the more native-like the outcome. However, there is still no consensus on whether there is a critical period and what the cutoff age is. We can also see
counterevidence of the Critical Period Hypothesis in this review and there are other confounding variables involved in acquisition of L2 phonology. As Odlin (2003) noted, “the interaction of age, social variation, and language transfer differs from one linguistic subsystem to the next” (p. 471). He argued that unless researchers can prove that there is a critical period that has real effects on the acquisition process, explanations of any age-related differences will have to involve multiple causes. Despite Flege’s repeated emphasis on input, DeKeyser and Larson-Hall (2005) contended that quantity and quality of input and other social variables play a very limited role when age effects are removed, while age effects still play a crucial role when the other variables are removed. Although Flege and his colleagues’ studies showed no signs of discontinuities before and after proposed cutoff age of the critical period, DeKeyser and Larson-Hall argued that there are also studies that showed discontinuities. It appears that the existence of a critical period will remain an ongoing debate. On the other hand, it appears that no researchers would deny the existence of a “sensitive period” and would rule out a maturational account of age-related changes in L2 phonology acquisition. Celce-Murcia, Brinton, and Goodwin (1996) cited Diamond (1988) to counter Lenneberg’s and Scovel’s theories of neurological maturation as “erroneous” (p. 16). However, acknowledging the age effect in their survey of the literature, Celce-Murcia et al. suggested that an intelligible pronunciation rather than a native-like pronunciation should be “a more realistic pedagogical goal” (p. 29). Derwing (2003) and Derwing and Munro (2005) concurred with Celce-Murcia et al. and contended that mutual intelligibility should be the primary concern in pronunciation teaching.

There is no doubt that L1 transfer plays an important role in the production of L2
speech. Moreover, the acquisition of L2 syllables may show effects of universal constraints on syllable structure, especially longer and more complex onsets and codas. Aside from age effects and L1 transfer, social factors are variables that make the investigation of L2 phonology acquisition complicated. The studies reviewed in this section show the imbalanced focus on speakers of certain languages. Therefore, future research should include native speakers of other languages. Moreover, the inter-rater reliability in some studies is questionable with the use of heterogeneous raters. Future research should use homogeneous and trained raters to ensure inter-rater reliability.

The findings in research can provide focuses for teaching English pronunciation to speakers of specific languages. Hansen’s (2001) method of collecting speech data from participants’ spontaneous production in one-on-one interviews at a long interval allows us to examine participants’ L2 acquisition over time with their naturalistic speech. Moreover, despite the fact that participants might speak different first languages, many researchers did not take the possibility of different L1 phonotactic constraints into account. Since speakers from different language backgrounds might have different L1 transfer, it is necessary to verify their L1 background before examining their production of English. Finally, the variationist framework employed by Rau et al. (2009) appears to be a superior framework to examine the production of learners of English.

As Derwing and Munro (2005) noted, pronunciation research has been marginalized in applied linguistics. Major English teaching journals should encourage pronunciation research, especially on effective methods to pronunciation instruction, to provide empirical evidence to inform ESL/EFL teachers instead of leaving them on their own. Finally, the lack of well-trained teachers to teach pronunciation suggests that TESOL
programs should take the responsibility for preparing would-be ESL/EFL teachers for teaching pronunciation.

Apparently, the issues of NS norms and foreign accents are sure to be ongoing debates. Nonetheless, for perceptual training, the norm is usually American English or British English. Regardless of the possibility that students might speak English with a foreign accent, when training perception, it is evident that EFL teachers need to expose students to native speaker input of the target language, as Flege suggested, to help them achieve native-like L2 perception of the target language community. Furthermore, Cenoz and Lecumberri (1999) found that the desire to acquire a native accent was a significant predictor of the improvement in phonetic discrimination, suggesting that motivation, or investment (Norton, 1995, 2000), plays an important role in L2 acquisition. The finding also suggests that motivating students to acquire a native accent as a goal to help with their perception of the target language is part of ESL/EFL teacher’s responsibility. In view of this, the present study used a spy game scenario suggested by Flege (personal communication, April 13, 2011) in the hope that participants in the trained groups might accomplish more in the game-like training.

Despite the fact that most people in Taiwan might speak first languages other than Mandarin, most researchers did not take the possibility of different L1 phonotactic constraints into account when they studied Taiwanese participants. Since speakers from different language backgrounds might have different L1 transfer, it is necessary to verify their L1 before examining cross-linguistic influence. Research on how native speakers of both Mandarin and Taiwanese process English syllables is scarce in the literature.

Bohn (1995) found that his Mandarin speakers relied almost exclusively on
durational cues to differentiate the /i/-/ɪ/ contrast, which cannot be explained by L1 transfer. Moreover, most Mandarin speakers do not use spectral cues to differentiate the the /i/-/ɪ/ contrast, as indicated in Wang (1997; 2006) and Bohn (1995). The results of the talker variability experiments in Lively et al. (1993) suggest that students should be trained with high talker variability. The high talker variability in Wang (2002) appeared to successfully shift the Chinese participants’ attention on temporal cues to spectral cues. Moreover, the high variability perceptual training in Bradlow et al. (1999) in various phonetic environments appeared to be effective in training Japanese speakers in both perception and production of the /r/-/l/ contrast. Furthermore, the innovative idea of using two different natural vowel durations for perceptual training in Lee (2009) suggests that if duration and talker variability are maximized, it is unnecessary to use unnatural synthesized tokens. In addition, synthetic stimuli do not represent the full range of acoustic properties of specific phonetic categories (Flege, 1995a).

With respect to the effect of perceptual training on production, Wang’s (2002) study suggests that effect of training on identifying non-native phonemic contrasts is not transferable to production. However, more studies (e.g., Rochet, 1995; Bradlow et al., 1997) suggested that the effect of perceptual training can be transferred to production. Therefore, it is worthwhile to explore further if the effect of training on perception can be generalized to production. Finally, whether the effect of training methods can persist after the completion of training should be a main concern when evaluating the effectiveness of training methods (e.g., Lively et al., 1994; Flege, 1995a; Bradlow et al., 1999).

In the next chapter, the methodology deployed in the present study to address the issues found in the literature is described and explained in detail.
CHAPTER 3: METHODOLOGY

3.1. Introduction

This chapter begins with a description of how the replication of Flege’s (1995a) study was modified. Then a brief introduction of the sociolinguistic background in Taiwan is presented. Next, I compare the trained segments in the present study in English, Mandarin, Taiwanese, and Hakka. Then I introduce the two perceptual training methods, identification method and categorical same/different method, in the present study. In the remainder of the chapter, I describe and explain the methodology deployed in the present study.

The present study replicated Flege’s (1995a) study with several modifications to investigate which intervention, ID or SD, is a better solution or whether they are equally effective. The modifications include: (1) the inclusion of a control group to eliminate the threat of maturation to internal validity, (2) the number of sample size in each group was increased from 10 to more than 20, (3) the perceptual training time was extended from approximately 105 minutes (7 sessions, approximately 15 minutes per session) to approximately 400 minutes (8 sessions, approximately 50 minutes per session), (4) the generalizability of perceptual training effects on identifying the word-final /t/-/d/ contrast to other final stop contrasts was investigated, (5) two extra non-native phonemic contrasts,
/i/-/ɪ/ and /ɛ/-/æ/, were included in the perceptual training, (6) word-final /t/-/d/ minimal pairs with word initials other than /b/ were incorporated in the perceptual training, (7) training stimuli in multiple phonetic environments were used, (8) manipulation of the durations of /i/-/ɪ/ stimuli was employed to maximize duration variability, (9) a production task was included to investigate whether the effects of perceptual training can be transferred to production, (10) the present study was conducted in an EFL setting instead of an ESL setting.

In the next section, I present a brief introduction of the sociolinguistic background in Taiwan.

### 3.2. The Sociolinguistic Background in Taiwan

Taiwan is a multilingual nation. According to Huang (2000), Mandarin, the official language, is spoken by nearly 90% of the population. However, the majority of the population (73%) is bilingual or multilingual and speaks Taiwanese as a first language. Given the high prestige of Mandarin in Taiwan, many native-Taiwanese-speaking parents have begun speaking Mandarin as home language. However, their Taiwanese-accented Mandarin, or Taiwanese Mandarin (Liao, 2000), is often joked about and certainly influences their children’s acquisition of Mandarin phonology (Peng, 1993). Most people in Taiwan speak Taiwanese Mandarin instead of Taiwan Mandarin (Cheng, 1985) or Guoyu 國語, standard Mandarin in Taiwan. As mentioned earlier, despite the fact that most people in Taiwan might speak first languages other than Mandarin, most researchers (e.g., Decamp, 1972; Flege, Munro, & Skelton, 1992; Paolillo, 1995; Lee, 2006; Rau, Chang, & Tarone, 2009) did not take the possibility of different L1 phonotactic
constraints into account when they studied Taiwanese participants. Since speakers from
different language backgrounds might have different L1 transfer (e.g., Flege, 1989; Flege
& Wang, 1989; Flege et al., 1992), it is imperative to verify their L1s before examining
cross-linguistic influence.

Three major languages are spoken in Taiwan: Mandarin, Taiwanese, and Hakka. In
the next section, the phonological systems of the three languages are compared to that of
English in relation to the trained segments in the present study.

3.3. Comparison of Phonologies in English, Mandarin, Taiwanese, and Hakka

In this section, the segments of English, Mandarin, Taiwanese, and Hakka are
compared in relation to the trained segments in the present study.

3.3.1. English

Due to the fact that American English is the norm of English language teaching and
testing in Taiwan, only American English segments are introduced here. According to
Ladefoged and Johnson (2011), General American English contains 11 monophthongs (/i/,
/ɪ/, /æ/, /ə/, /ʌ/, /ɔ/, /u/, /ʊ/, /ɔ/, and /ə/) and 5 diphthongs (/e/, /o/, /aɪ/, /aʊ/, and /aʊ/).
The consonants consist of 6 stop consonants (/p/, /b/, /t/, /d/, /k/, and /g/), 3 nasals (/m/,
/ŋ/, and /ŋ/), 8 fricatives (/f/, /v/, /θ/, /ð/, /s/, /z/, /ʃ/, and /ʒ/), 2 affricates (/tʃ/ and /dʒ/), 4
approximants (/w/, /r/, /j/, and /l/), and /h/, which Ladefoged and Johnson argue is the
voiceless counterpart of the surrounding sounds (p. 69).

As mentioned in Chapter 1, there are no high-front /i/-/ɪ/ and mid vs. low-front
/ɛ/-/æ/ contrasts in Mandarin, Taiwanese, and Hakka and they are highly functionally
loaded (Brown, 1988). Previous studies have shown that most untrained Chinese ESL
learners had difficulty perceiving these non-native phonemic contrasts (e.g., Bohn, 1995; Wang, 1997). The tongue positions of high-front /i/-/ɪ/ and mid vs. low-front /ɛ/-/æ/ contrasts are presented in Figure 3.1.

![Figure 3.1](image)

*Figure 3.1* The auditory qualities of /i/-/ɪ/ and /ɛ/-/æ/ contrasts of Standard American Newscaster (Adapted from Ladefoged & Johnson, 2011, p. 90)

Since final stops are commonly unreleased in American English, the major difference between final stop minimal pairs is in the vowel length. The effect of stop voicing on duration of preceding vowels is larger in English than in other languages in the world (Flege, 1993). A vowel preceding voiceless stops (/p/, /t/, and /k/) is much
shorter than it is when preceding voiced stops (/b/, /d/, and /g/) (Ladefoged & Johnson, 2011, p. 59). Figure 2.1 presents the waveforms of the words *mat* and *mad*.

![Waveforms of mat and mad](image)

*Figure 2.1 The Waveforms of the Words mat and mad (Ladefoged & Johnson, 2011, p. 60)*

### 3.3.2. Mandarin

A language with prosodic and phonetic structures very different from those in English, Mandarin is the official language of both mainland China and Taiwan. According to Li and Thompson (1978), a number of Chinese dialects are mutually unintelligible due to phonological and lexical differences. Among all Chinese dialects, Mandarin has the simplest tone system with only four phonemic (lexical) tones plus a neutral tone. Unlike English, most Chinese dialects forbid consonant clusters and as mentioned earlier, there are only two final consonants, /n/ and /ŋ/, in Mandarin (Cheng, 1973; Li & Thompson, 1978). Due to language contact with Taiwanese, the final /r/ in
Peiking Mandarin is rarely seen in Taiwan Mandarin (Kubler, 1985). Mandarin permits no obstruents in the word-final position (Cheng, 1973). Therefore, there is no word-final /t/-/d/ contrast in Mandarin.

According to Duanmu (2007), there are eight consonants (/θ/, /ð/, /dz/, /ʒ/, /ʃ/, /v/, /w/, and /j/) and five vowels (/i/, /æ/, /ɔ/, /ʌ/, and /ɔɪ/) in English that do not have equivalents in Mandarin. Since there is /i/ but no /ɪ/ in Mandarin, there is no /i/-/ɪ/ contrast in Mandarin. In addition, tenseness is not a feature to distinguish vowels in Mandarin.

Although Xu (1980, cited in Duanmu, 2007) claimed that /æ/ is a variant of the low vowel in Mandarin and Lin (2007) also contended that /æ/ may occur in Mandarin in specific contexts (p. 64), most linguists agreed that there is no /æ/ in Mandarin (Duanmu, 2007, p. 39). Therefore, the present study assumed that there is no /ɛ/-/æ/ contrast in Mandarin.

3.3.3. Taiwanese

According to Klöter (2005), there is /i/ and /ɛ/ but no /ɪ/ or /æ/ in Taiwanese. Therefore, there is no /i/-/ɪ/ or /ɛ/-/æ/ contrast in Taiwanese. There are six consonant endings (/p/, /t/, /k/, /m/, /n/, and /ŋ/) in Taiwanese. That is, in terms of final stops, Taiwanese has /p/, /t/, and /k/ but no /b/, /d/, or /ɡ/ in the word-final position (Cheng, 1968; Hsu, 1998; Klöter, 2005). Therefore, although there is final /t/ in Taiwanese, there is no word-final /t/-/d/ contrast in Taiwanese.

3.3.4. Hakka

Hakka is spoken by over 50 million people worldwide and there are many Hakka dialects based on region (Zhang, 1996, cited in Opper, 2010). This section focuses on
core phonological properties of Sixiàn Hakka dialect. According to Opper (2010), the dialect is spoken by the largest group of Hakka speakers in Taiwan and is spoken in the region where the research site is located (Pingtung County).

There are lots of similarities between Hakka and Taiwanese. There is also /i/ and /ɛ/ but no /ɪ/ or /æ/ in Sixiàn Hakka. Therefore, there is no /i/-/ɪ/ or /ɛ/-/æ/ contrast in Sixiàn Hakka.

According to Lu (2005, cited in Opper, 2010), there are six consonant endings (/p/, /t/, /k/, /m/, /n/, and /ŋ/) in Sixiàn Hakka. Like Taiwanese, in terms of final stops, Hakka has /p/, /t/, and /k/ but no /b/, /d/, or /g/ in the word-final position. Therefore, there is no word-final /t/-/d/ contrast in the dialect.

In the next section, I introduce the two perceptual training methods, identification method and categorical same/different method, in the present study.

3.4. Identification Training vs. Categorical Same/Different Training

The procedures of the two training methods in the present study, identification training and categorical same/different training, are exemplified in this section.

3.4.1. Identification Training

In identification (ID) training, the trainee listens to one stimulus of a minimal pair per trial and is instructed to identify the stimulus in terms of two categories (e.g., /i/ or /i/). Feedback regarding the acoustic cues of the segment (e.g., tense or lax) or the vowel preceding a specific segment (e.g., vowels before /t/ are pronounced shorter than they are before /d/) is given immediately after the trainee answers the trial. By learning the
phonetic properties from feedback, the trainee is expected to master the phonetic properties associated with each category.

Logan, Lively, and Pisoni (1991) noted that identification training forces the trainee to develop and use phonetic memory codes in short-term memory and eventually establish two categories of the trained contrast.

3.4.2. Categorical Same/Different Training

In categorical same/different (SD) training, the trainee listens to two stimuli per trial. Same and different trials are presented randomly. As in Flege (1995a), two stimuli presented in a given trial are produced by different talkers. Thus, the trainee’s task is to listen to the trial and decide whether the two stimuli are two different realizations of a single phonetic category instead of whether one stimulus was presented twice. Feedback (e.g., same or different) is given immediately after the trainee answers the trial.

Logan, Lively, and Pisoni (1991) noted that categorical same/different training relies on information in sensory memory and enhances the perception of fine within-category acoustic differences.

3.4.3. Debate on the Effectiveness of Identification Training vs. Categorical Same/Different Training

Logan et al. (1991) and Lively et al. (1994) are in favor of identification training in that the immediate feedback on phonetic properties of stimuli may promote formation of a new phonetic category in L2 in long term memory. Jamieson and Morosan (1986, 1989) are also in favor of identification training because same/different training may fail to sensitize the trainee to pay attention to the core properties that define the L2 categories by
focusing on intraphonemic sensitivity. As a result, the trainee may fail to establish a new phonetic category.

On the other hand, Polka (1992) contended that the trainee in identification training may attend to any phonetic properties, even ones not used by native speakers, to differentiate two non-native categories and still respond correctly. Polka is in favor of categorical same/different training because the procedure requires the trainee to maintain perceptual constancy to successfully identify multiple variants of each category. As a result, the trainee may establish distinctive categories in L2.

### 3.5. Research Questions

As mentioned earlier in Section 1.5, eight research questions were formulated as follows.

1. Are the perceptual training methods effective in identifying three non-native phonemic contrasts?
2. Which training method, ID or SD, will be more effective for training EFL students in Taiwan in identifying three non-native phonemic contrasts?
3. Can the effect of perceptual training on identifying three non-native phonemic contrasts be generalized to new tokens produced by familiar talkers and familiar tokens produced by new talkers?
4. Can the effect of perceptual training on identifying the word-final /t/-/d/ contrast be generalized to other final stop contrasts?
5. Can the effect of the training methods persist three months after the completion of the training?
6. Can the perceptual training in identifying three non-native phonemic contrasts facilitate production of the non-native phonemic contrasts?

7. Will the performance of the participants in the trained groups vary due to different L1 backgrounds?

8. What are the language attitudes of the participants toward speaking Standard English and EFL teachers’ accents?

3.6. Research Hypotheses

As mentioned earlier in Section 1.6, seven research hypotheses were proposed based on the research questions as follows:

1. Judging by the results of previous studies that perceptual training in identifying English segments can be effective (e.g., Flege & Wang, 1989; Lively, Logan, & Pisoni, 1993, 1994; Flege, 1995a; Bradlow, Akahane-Yamada, Pisoni, & Tohkura, 1999; Wang, 2002; Lee, 2009), I predicted that the perceptual training would be effective.

2. Concurring with Flege (1995a), I predicted that identification training would be more effective than same/different training in that the ID group in Flege (1995a) felt that they benefited more from the training and enjoyed the training more than did the SD group.

3. Judging by the results of previous studies, I expected the perceptual training in identifying three non-native phonemic contrasts to be generalized to new tokens produced by familiar talkers and familiar tokens produced by new talkers.

4. I predicted that the effect of perceptual training on identifying the word-final /t/-/d/ contrast would be generalized to other final stop contrasts.
5. Judging by the results of previous studies that trainees retained the effect of perceptual training (e.g., Lively, Logan, & Pisoni, 1994; Flege, 1995a; Bradlow, Akahane-Yamada, Pisoni, & Tohkura, 1999; Wang, 2002), I predicted that the effects of the training methods would persist three months after the completion of the training.

6. I predicted that the perceptual training in identifying three non-native phonemic contrasts could facilitate production of the non-native phonemic contrasts.

7. I hypothesized that the performance of the participants in the trained groups would vary due to different L1 backgrounds, as in Flege (1989), Flege and Wang (1989), and Flege and Liu (2001).

### 3.7. Research Design

An experimental study is the most ideal to compare the effects of the two perceptual training methods. Pretest-posttest designs allow us to compare and measure the effectiveness of interventions. Therefore, a pretest-posttest true experimental study was designed to address the research questions. To address the research questions, the present study employed a pretest-posttest design in a true experimental study. There were two experimental groups that received interventions and a control group that participated only in the pretest and the first posttest. Any gain by the control group can be assumed to be due to familiarity with the test tokens, and not to the effect of training. The participants were randomly assigned to one of two experimental groups or to a control group. Therefore, the independent variable of the present study is treatment (3 levels) and the dependent variables are the scores of perception tests and the scores of production tests.

In addition, to understand the language attitudes of the participants toward speaking
Standard English and EFL teachers’ accents, a language background and attitude survey (Appendix A) was administered to all the participants before the pretest.

There are four phases in the present study: a pretest phase, a perceptual training phase, a posttest phase, and a three-month follow-up posttest phase. In the pretest phase, both perception and production data from the trained and control groups were collected as a baseline. During the perceptual training phase, participants in both of the trained groups received 8 sessions of perceptual training. In the posttest phase, both the trained and the control groups repeated the pretest tasks. In addition, a generalization test was included to assess whether the effect of perceptual learning was transferred to new words, new talkers, and untrained final stops. During the follow-up posttest three months after the training, only the two trained groups repeated the posttest without the oral production task.

### 3.8. Research Site

The research site was a department of applied foreign languages in a university located in southern Taiwan, where the researcher was a full-time lecturer. After the researcher had obtained permission from the chair of the department, the application of the present study was reviewed and approved by the Behavioral and Social Sciences Institutional Review Board (IRB) at The Ohio State University (Protocol number 2011B0257).

The participants were informed of the procedures of the study in a written form and signed a consent form before the training.
3.9. Participants

Eighty-eight English major students, 37 males and 51 females, from a population of 231, were recruited to participate in the present study and were randomly assigned to one of the three groups based on their gender. During the study, 17 participants dropped out of the study. As a result, the total number of participants who finished the pretest and posttest was 71 (24 males, 47 females), including the identification (ID) group \( n = 24 \); 7 males, 17 females), the same/different (SD) group \( n = 23 \); 9 males, 14 females), and the control group \( n = 24 \); 8 males, 16 females). The participants’ age ranged from 18 to 22. Although the participants in the present study had been admitted to the university according to their entrance examination scores, their levels of English proficiency were not homogenous, which warranted a pretest after randomization to provide a baseline. Moreover, the research site is a Tier 4 (ranked in the bottom 25% among all the universities) university in Taiwan, which means most of the participants’ English proficiency was at the high-beginner level. In other words, most of the participants were expected to be unable to identify the three non-native phonemic contrasts correctly in the present study. The proportions of main home languages are Mandarin (60.6%), Taiwanese (38%), and Hakka (1.4%). Since the research site is located in southern Taiwan and Mandarin is the official language and medium of instruction, most of the participants are Mandarin-Taiwanese bilinguals (69%). The rest of the participants are native speakers of Mandarin (12.7%), Mandarin-Hakka (9.9%), Taiwanese (7%), and Hakka (1.4%). Table 3.1 presents participants’ background information in each group.
Table 3.1
Background Information of Control, ID, and SD Groups

<table>
<thead>
<tr>
<th></th>
<th>CON</th>
<th>ID</th>
<th>SD</th>
<th>Total</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>16</td>
<td>17</td>
<td>14</td>
<td>47</td>
<td>66.2</td>
</tr>
<tr>
<td>Male</td>
<td>8</td>
<td>7</td>
<td>9</td>
<td>24</td>
<td>33.8</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>4</td>
<td>5.6</td>
</tr>
<tr>
<td>21</td>
<td>4</td>
<td>2</td>
<td>4</td>
<td>10</td>
<td>14.1</td>
</tr>
<tr>
<td>20</td>
<td>10</td>
<td>9</td>
<td>7</td>
<td>26</td>
<td>36.6</td>
</tr>
<tr>
<td>19</td>
<td>6</td>
<td>9</td>
<td>9</td>
<td>24</td>
<td>33.8</td>
</tr>
<tr>
<td>18</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>7</td>
<td>9.9</td>
</tr>
<tr>
<td>L1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mandarin-Taiwanese</td>
<td>17</td>
<td>18</td>
<td>14</td>
<td>49</td>
<td>69.0</td>
</tr>
<tr>
<td>Mandarin-Hakka</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>7</td>
<td>9.9</td>
</tr>
<tr>
<td>Mandarin</td>
<td>1</td>
<td>5</td>
<td>3</td>
<td>9</td>
<td>12.7</td>
</tr>
<tr>
<td>Taiwanese</td>
<td>2</td>
<td>0</td>
<td>3</td>
<td>5</td>
<td>7.0</td>
</tr>
<tr>
<td>Hakka</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1.4</td>
</tr>
<tr>
<td>Main home language</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mandarin</td>
<td>12</td>
<td>19</td>
<td>12</td>
<td>43</td>
<td>60.6</td>
</tr>
<tr>
<td>Taiwanese</td>
<td>12</td>
<td>5</td>
<td>10</td>
<td>27</td>
<td>38.0</td>
</tr>
<tr>
<td>Hakka</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1.4</td>
</tr>
<tr>
<td>Mean years of English learning</td>
<td>10.5</td>
<td>11.5</td>
<td>10.9</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The sample size in the present study was calculated based on suggestions by research methodologists. Cohen (1988) contended that the minimum number of participants per variable should to be 14. According to Mertens (2009), the recommended number of participants per variable is 15. Hair et al. (1998) argued that 15 to 20 participants per independent variable are appropriate with regard to the generalizability of research results.
The present study was primarily designed with three independent variables (two main and one rival), and thus, it required at least 45 participants. Therefore, the number of participants (N = 71) in the present study was considered to have met all the above recommendations.

3.10. Stimuli

The results of the talker variability experiments in Lively et al. (1993) suggest that students should be trained with high talker variability. The high talker variability in Wang (2002) appeared to successfully shift the Chinese participants’ attention on temporal cues to spectral cues. Moreover, the high variability perceptual training in Bradlow et al. (1999) in various phonetic environments appeared to be effective in training Japanese speakers in both perception and production of the /r/-/l/ contrast. Furthermore, the innovative idea of using two different natural vowel durations for perceptual training in Lee (2009) suggests that if duration and talker variability are maximized, it is unnecessary to use unnatural synthesized tokens. In addition, synthetic stimuli do not represent the full range of acoustic properties of specific phonetic categories (Flege, 1995a) and the ultimate goal of perceptual training is to enable learners to perceive the target contrasts in real speech. Therefore, in the present study, only natural tokens in various phonetic environments were used for training with talker and duration variability maximized.

Hi-front English vowels /i/-/ɪ/, mid vs. low-front English vowels /ɛ/-/æ/, and the word-final /t/-/d/ contrasts were chosen for training because the first two contrasts don’t exist in Chinese dialects and pose serious problems for Chinese speakers as indicated by previous studies (e.g., Bohn, 1995; Wang, 1997). In addition, the first two contrasts are
among the most functionally loaded phonemes (Brown, 1988), so the inability to discriminate or produce them is likely to cause a communication breakdown. Similar to the Korean EFL learners in Lee (2009) and the Mandarin speakers in Wang (1997, 2006) and Bohn (1995), most Taiwanese EFL learners were expected to rely almost exclusively on durational cues to differentiate English the /i/-/ɪ/ contrast and have difficulty perceiving and producing the contrast. In comparison, the /u/-/ʊ/ contrast trained in Wang (1997, 2002) and Lee (2009) are considered unimportant by Brown since there are only four /u/-/ʊ/ minimal pairs in modern English (pool/pull, fool/full, who’d/hood, suit/soot) thus misunderstanding is very unlikely to occur for these contrasts. Therefore, the /u/-/ʊ/ contrast was excluded for the perceptual training in the present study.

With respect to the word-final /t/-/d/ contrast, similar to the results in previous studies, it was expected that most Taiwanese EFL learners do not know that the durations of the same vowel preceding word-final /t/-/d/ are different. We can see that even experienced Chinese EFL learners might not produce the English vowel duration difference preceding /t/-/d/ accurately (e.g., Flege, et al., 1992). Given the fact that in conversational English the word-final /t/-/d/ are often unreleased (Flege, 1995a), it is imperative to train EFL students to identify the contrast with edited stimuli in which final stops are removed.

The researcher obtained permission to use real-person pronunciation sound files from the following electronic dictionaries for the training: Cambridge Advanced Learner’s Dictionary (Appendix B), Longman Dictionary of Contemporary English (Appendix C), Merriam-Webster 11th Collegiate Dictionary (Appendix D), and m-w.com (the pronunciations are the same as those in Merriam-Webster 11th Collegiate Dictionary).
Considering the first two dictionaries have both British and American pronunciations, only the American pronunciation sound files were used in the present study since American English is the norm of ELT and proficiency tests in Taiwan. The stimuli from the dictionaries were recorded from one computer to another computer at 44100 Hz (mono channel) using Audacity 1.2.6 on Windows 7 Home Premium (32 bit) operating system.

Aside from the electronic dictionaries mentioned above, four native speakers of Midwestern American English in Ohio recorded tokens for the training and the pre- and post-tests. Two of the talkers’ tokens were used only for the posttests to test generalization of the training effect to new talkers and retention. As in Lee (2009), the four native talkers were asked to record all the /i/-/ɪ/ tokens with two different vowel durations. For example, the word beach was recorded with both a normal vowel duration as in dictation and a shorter vowel duration as in fast speech, whereas its contrast, bitch, was recorded with a longer vowel duration as in emphatic speech as well as with a normal duration (Appendix E). The purpose of this manipulation of vowel durations was to use natural stimuli instead of synthesized stimuli to sensitize learners to spectral differences in the /i/-/ɪ/ contrast instead of temporal cues in identifying the structurally similar English tense and lax vowel pair. The process doubled the /i/-/ɪ/ tokens recorded by the native talkers (4 Talkers × 76 Words × 2 Vowel durations), which made the total number of /i/-/ɪ/ tokens to 836 (including 3 Dictionary Talkers × 76 Words). As for the /ɛ/-/æ/ contrast, there were 350 tokens (7 Talkers × 50 Words) in total.

To edit the final /t/-/d/ endings as in Flege (1995a), all the final /t/-/d/ endings in the present study were edited with Audacity 1.2.6 so that the closure voicing, release burst,
and duration of the stop closure interval were removed. The remaining acoustic cues such as vowel duration and vowel quality were expected to provide sufficient information for native speakers to identify the word-final /t/-/d/ tokens correctly as in Flege (1995a). This process doubled the number of the word-final /t/-/d/ tokens (7 Talkers × 20 Words × 2 Editing conditions) and made the number of the word-final /t/-/d/ tokens to 280. The interstimulus interval (ISI) between the stimuli of each same/different pair for SD training was fixed at 250 ms. According to Flege (1995a), short ISIs encourage participants to rely on sensory memory.

Aside from the high variability of takers, the target segments in the stimuli were also in multiple phonetic environments. The majority of the /i/-/ɪ/ contrast words were monosyllabic CVC words with different initials and codas (e.g., bead-bid, feel-fill), with 2 pairs of VC words (eat-it, eel-ill), a pair of CCVC words (sleep-slip), and a pair of CVCC words (feast-fist). There was only one pair of /i/-/ɪ/ disyllabic contrast (beaten-bitten). Similarly, most of the /ɛ/-/æ/ contrasts were also monosyllabic CVC words with different initials and codas (e.g., bet-bat, guess-gas). The rest were 2 pairs of CVCC words (send-sand, spend-spanned) and 3 pairs of disyllabic contrasts (kettle-cattle, peddle-paddle, rebel-rabble). With respect to the word-final /t/-/d/ contrast, all of them were monosyllabic CVC words with various initials (/b/, /h/, /m/, /n/, /s/). As for other final stop contrasts (/k/-/g/, /p/-/b/) for the generalization test, all of them were monosyllabic CVC words with various initials (/b/, /d/, /h/, /k/, /l/, /p/, /t/, /s/, /t/) except a pair of CCVC words (snack-snag). Some of the tokens were used for both the training and the pre- and post-tests while some of the tokens were used only for the pre- and post-tests.
Participants in the training groups were exposed to some of the tokens above in the training (Appendix F), including two vowel durations of 8 minimal pairs for the /i/-/ɪ/ contrast (5 Talkers × 16 Words × 2 Durations), 16 minimal pairs for the /ɛ/-/æ/ contrast (5 Talkers × 32 Words), and 8 edited minimal pairs for the word-final /t/-/d/ contrast (5 Talkers × 16 Words) from the three electronic dictionaries and tokens recorded by three native talkers to maximize talker variability. It was intended to encourage participants to focus on general rather than token-specific properties of the stimuli, and thus to eliminate talker effects (Lively et al, 1994; Flege, 1995a).

The tokens for the pretest and posttest (Appendix G) were edited using Audacity 1.2.6. Tokens recorded by 2 male native talkers were presented only in the generalization test (Appendix H) to assess generalization to new talkers. In addition, tokens produced by familiar talkers not used in the training were used to test generalization of the training effect to new tokens and other final stop endings. The stimuli used in the pretest and posttest were all used for training in both experimental groups. Thus, the face validity and content validity (Carmines & Zeller, 1991) of the measurement were validated for the instrument measured what should be measured and reflected the specific intended domain of content to draw conclusions.

All the subtests were administered to two experienced native ESL/EFL teachers of American English to assess the intelligibility of the trials and to eliminate potential problematic trials. It turned out that two edited trials in the word-final /t/-/d/ contrast were ambiguous and they were replaced by tokens produced by the other speakers. In addition, there were four instances that the researcher had checked the wrong correct answer in the computer program and the answers were corrected. Thus, the instrument validity was
further established by expert review (Carmines & Zeller, 1991).

3.11. Measurement

3.11.1. Perception

The effect of training on perception was assessed through the pretest and posttest. After each test the participants’ scores were calculated by the computer system by group. Then each group’s scores were filed using SPSS.

3.11.2. Production

The possible effect of perceptual training on production was assessed through an intelligibility test. Three experienced native-English-speaking ESL/EFL teachers, all of whom teach at university level, were the judges for the productions of the participants in the pretest and posttest. To avoid the threat of low inter-rater reliability in Munro et al. (1996), Derwing et al. (1997), and Thomson and Munro (2006), the raters rated all of the production data. The production data from both the pretest and posttest were numbered and given to the native judges in a random order so that the judges would not expect progress of the training and thus give the posttest productions higher scores. They rated each token produced by the participants from “totally unintelligible” (1 point) to “native-like” (10 points) in relation to the token the participants were supposed to produce.

3.12. Procedures

3.12.1. Introduction and Consent

A brief introduction of the study was given by the researcher. Next, consent forms
were distributed and signed by the participants. Then the participants were guided to complete a language background and attitude questionnaire (Appendix A) before the pretest began.

3.12.2. Pretest

3.12.2.1. Perception

Both the experimental groups and the control group were given a pretest (Appendix I) on the computer in a language lab to provide a baseline. The format of the perception test is a two alternative forced-choice test. For the identification task, the participants were instructed to play a trial and click on one of the two choices on the computer screen after hearing each trial. The participants were asked to respond to each trial and to guess even if they were uncertain about the answer. The trials were presented over headsets at a comfortable level in a language lab. The reaction time for each trial was not limited.

There was no feedback for the pretest.

To familiarize the participants with the trials and rate of presentations, 6 examples (2 for each contrast) were presented to the participants without feedback in a short practice session in a computer lab. Then a pretest with 3 subtests for each contrast (30 randomly arranged trials in each subtest) was administered to the participants over headsets at a comfortable level.

The pretest also served as a screening procedure to eliminate those participants who might show native-like perceptual ability and thus would not benefit from the subsequent perceptual training. This would avoid ceiling effects that appeared in Wang (2002) and Derwing et al. (2006). In addition, Flege emphasized that “trimming” should also be conducted to manifest equal performance before any training is administered in order to
avoid experimenter bias effects (personal communication, January 7, 2011). The results of the pretest indicated that there were no potential ceiling effects and there was no significant difference among the three groups.

3.12.2. Production

With respect to the effect of perceptual training on production, Wang’s (2002) study suggests that effect of training on identifying non-native phonemic contrasts is not transferable to production. However, more studies (e.g., Rochet, 1995; Bradlow et al., 1997) suggested that the effect of perceptual training can be transferred to production. Therefore, the present study aimed to explore further if the effect of training on identifying three non-native phonemic contrasts could be generalized to production of these contrasts.

Participants in all groups were given a word list (Appendix J) with the Kenyon and Knott Phonetic Alphabet (K.K.), which is similar to the International Phonetic Alphabet (IPA) and are widely used in ELT in Taiwan. All the words were selected from words used in the training sessions and were randomly arranged so that two words in a minimal pair were not presented back to back. The participants were told that they had five minutes to practice pronouncing the words and then their productions were recorded on a computer hard drive via their microphone in a language lab. They were asked to read the word in each question twice in clear and loud speech. The production data were converted to MP3 files using Audacity1.2.6 for the convenience of storage and presentation to the native raters.

3.12.3. Training

The control group received no training at all and participated only in the pretest and
the first posttest. The experimental groups chose from four time slots to receive 50-minute individual training on the computer weekly for a total of 8 weeks during three months. To make the training more game-like, the researcher adopted a “naturalistic approach” rather than mere phonetic training, suggested by Flege (personal communication, April 13, 2011) in the hope that participants in the trained groups might accomplish more in a spy game scenario. The scenario was as follows:

You are a special agent and will work as a native speaker of American English. For the first step, you are going to train to identify 3 non-native phonemic contrasts for 8 weeks. If you fail to identify the contrasts after the training, you will reveal your identity and get arrested. Therefore, work hard when you train. Good luck!

3.12.3.1. Identification Training

Participants in the ID training group were told that they would be trained to identify three contrasts in eight training sessions on the computer and then read the training instructions (Appendix K). They listened to one stimulus per trial and were instructed to pay attention to the acoustic cues of the vowel and identify the stimulus in terms of two categories (e.g., /i/ or /ɪ/). The trials were presented over headsets at a comfortable level in a language lab and the participants were instructed to play a trial and click on one of the two choices on the computer screen after hearing each trial. After they clicked on the button on the left to confirm the answer, they checked the feedback that corresponded to the number of the question in a feedback handout. For the /i/-/ɪ/ contrast, they learned from the feedback that /i/ is a tense vowel and /ɪ/ is a lax vowel; for the /æ/-/ɛ/ contrast, they learned that for /æ/, the tongue is lower and the duration is longer than /ɛ/; for the word-final /t/-/d/ contrast, they learned that the vowel before /t/ is pronounced shorter
than that before /d/ (Appendix L).

There were 180 trials and 180 stimuli for each session (60 trials for each contrast; one stimulus per trial), while the number of trials for the SD group contained 360 stimuli (180 trials x 2 stimuli). Two stimuli per trial were presented to the participants in the SD training, whereas only one stimulus per trial was presented in the ID training. To balance out the trials and feedback of both training methods, feedback of 180 trials was given to participants in the ID group and in Question 61 in each contrast, participants in the ID group were instructed to listen to 60 stimuli without feedback. Thus, participants in both groups listened to the same number of stimuli during training and received the same amount of information regarding the accuracy of their answers.

3.12.3.2. Categorical Same/Different Training

Participants in the SD training group were told that they would be trained to identify three contrasts in eight training sessions and would listen to two stimuli per trial. Same and different trials were presented randomly. They were asked to read the training instructions (Appendix M). As in Flege (1995a), two stimuli presented in a given trial were produced by different talkers (i.e. from different electronic dictionaries or talkers). Thus, the participants’ task was to decide whether the two stimuli were two different realizations of a single phonetic category instead of whether one stimulus had been presented twice. The interstimulus interval (ISI) between the tokens of each pair was fixed at 250 ms. They listened to two stimuli per trial and were instructed to decide whether the two stimuli were the same. The trials were presented over headsets at a comfortable level in a language lab and the participants were asked to click on one of the two choices (Same or Different) on the computer screen after hearing each trial. After
they clicked on the button on the left to confirm the answer, they checked the answer (Same or Different) in the handout with no feedback on the acoustic cues of the stimuli. Aside from the answer (Same or Different), participants in the SD group received phonetic symbols of the stimuli in each question. For example, Same: /ɪ/, Different: /i/ vs. /ɪ/, Same: /ɪ/, and Different: /i/ vs. /i/ (Appendix N). This was intended to prevent that participants might decide on the opposite segment and still answer the question correctly. In total, there were 180 trials and 360 stimuli for each training session (60 trials for each contrast; two stimuli per trial).

3.12.4. Posttest

After completion of the training sessions, a posttest was administered to the participants a week later. The trials and the procedures for the posttest were identical to those of the pretest except that the order of the trials was randomized.

A generalization test was also included in the first posttest. The test consisted of 6 subsets with a total of 180 trials to assess the effect of the perceptual training on new tokens and other final stop endings produced by old talkers as well as familiar tokens produced by new talkers. All participants were asked to read the instructions (Appendix O), which are similar to the pretest instructions, and then answer the questions.

Whether the effect of training methods can persist after the completion of training should be a main concern when evaluating the effectiveness of training methods (e.g., Lively et al., 1994; Flege, 1995a; Bradlow et al., 1999). Therefore, in the present study, a follow-up posttest to assess retention was administered to the experimental groups three months after the first posttest. Again, the trials and the procedures were identical to those of the first posttest with the order of trials randomized.
3.13. Data Collection

3.13.1. Survey

A language background and attitude questionnaire (Appendix A) was completed by the original sample ($N = 88$) before the pretest to understand their language backgrounds and language attitudes toward speaking Standard English and EFL teachers’ accents.

3.13.2. Pretest and Posttests

After attrition, 71 participants ($n = 24$ in the control group, $n = 24$ in the ID group, and $n = 23$ in the SD group) finished the posttest. The pretest and posttest perception scores were calculated for data analysis. In addition, scores from native judges’ evaluation of the participants’ productions in the pre- and post-tests were calculated for data analysis.

Forty-one participants (ID group $n = 21$, SD group $n = 20$) finished the follow-up posttest three months after the first posttest. The data were used to test retention.

3.14. Data Analysis

In the present study, Statistical Packages for the Social Sciences (SPSS) was utilized to analyze the quantitative data. To determine significance throughout the study, the significance level was set at .05. Pretest and posttest data were statistically analyzed to test for the effects of the training methods.

Most researchers in previous studies used repeated measures analysis of variance (ANOVA) with pretest-posttest data to investigate the effect of training (e.g., Flege,
1995a; Wang, 2002; Lee, 2009). However, Dimitrov and Rumrill (2003) asserted that “the results provided by repeated measures ANOVA for pretest-posttest data can be misleading. Specifically, the F test for the treatment main effect (which is of primary interest) is very conservative because the pretest scores are not affected by the treatment” (p. 162). They contended that although one-way ANOVA is a better alternative, analysis of covariance (ANCOVA) is the best choice for pretest-posttest designs for it results in a more powerful test and allows for modifications leading to appropriate analysis when some assumptions do not hold. Moreover, ANCOVA leads to larger F-ratio and smaller error variance than ANOVA, which means more power to detect group differences. Therefore, ANCOVA was used for the pretest and posttest data. However, when the assumption of homogeneity of regression slopes is violated, the Johnson-Neyman technique (John & Neyman, 1936) was conducted to provide regions of significance. Kowalski, Schneiderman, and Willis (1994) regarded the technique as a generalization of the analysis of covariance (ANCOVA) because it does not make the assumption that the regression coefficients for the regression of X on the covariates are equal in the groups being compared. Cohen’s effect size value was also calculated for the overall scores in the pretest and the posttest on an online effect size calculator (Becker, 2012)

A series of 3 (group) x 2 (pre/post) analysis of covariance (ANCOVA) were conducted to evaluate the effects of perceptual training methods on the participants’ perception test scores. The two independent variables were training methods (control, ID, SD) as a between-group variable and Time (pretest and posttest) as a within-group variable, with the posttest scores as the dependent variable and the pretest scores as a covariate.
A series of one-way ANOVA were conducted to evaluate if the effect of training could be generalized to new tokens, new talkers, and other final stops contrasts.

In addition, to assess retention, a series of paired-samples t-tests were conducted to examine if the score differences between the posttest and the follow-up posttest were significant. Then a series of paired-samples t-tests were conducted to examine if the score differences between the pretest (drop-outs deleted) and the follow-up posttest were significant.

With respect to production, a series of 3 (group) x 2 (pre/post) analysis of covariance (ANCOVA) were conducted to evaluate the effects of perceptual training methods on the participants’ oral production scores. The two independent variables were training methods (control, ID, SD) as a between-group variable and Time (pretest and posttest) as a within-group variable, with the posttest oral production scores as the dependent variable and the pretest oral production scores as a covariate.

To investigate L1 influence, a one-way analysis of covariance (ANCOVA) was conducted at a significance level of .05. The independent variable, L1, included two levels: Mandarin and Mandarin-Taiwanese. The dependent variable was the participants’ posttest production scores and the covariate was the participants’ pretest production scores. A significant main effect of L1 was explored through Bonferroni post-hoc pair-wise comparisons.

In the next chapter, statistical results of the present study answering the eight research questions are presented.
CHAPTER 4: RESULTS

This chapter reports the statistical results of the present study answering the eight research questions. Section 4.1 reports results of the pre- and post-test and the effects of the perception training, Section 4.2 compares the effects of the two training methods, Section 4.3 reports results of the generalization test to show if the effect of training was generalized to new tokens and now talkers, Section 4.4 reports results of the generalization test to show if the effect of training was generalized to other final stop contrasts, Section 4.5 reports results of the follow-up posttest and compares the retention of training effects of the training methods, Section 4.6 reports results of the production task at pretest and posttest to show if the perceptual training facilitated production, Section 4.7 reports if performance of the participants in the trained groups varied due to different L1 backgrounds, and Section 4.8 shows the language attitudes of the participants toward speaking Standard English and Taiwanese EFL teachers’ accents. Each section includes statistical analyses, the conclusion of hypothesis testing, and a summary of findings. Furthermore, the participants’ difficulty in perception and production is presented in Section 4.9. To determine significance throughout the study, the significance level was set at .05. Finally, Section 4.10 summarizes findings in this chapter.
4.1. Research Question 1: Training Effects

Research question 1: Are the perceptual training methods effective in identifying three non-native phonemic contrasts?

Judging by the results of previous studies that perceptual training in identifying English segments can be effective (e.g., Flege & Wang, 1989; Lively, Logan, & Pisoni, 1993,1994; Flege,1995a; Bradlow, Akahane-Yamada, Pisoni, & Tohkura,1999; Wang, 2002; Lee, 2009), I predicted that the perceptual training would be effective. Therefore, the null hypothesis predicted that there would be no significant differences among the three groups’ performance at posttest. The pretest consisted of three subtests, the /i/-/ɪ/ contrast (Part 1), the /ɛ/-/æ/ contrast (Part 2), and the word-final /t/-/d/ contrast (Part 3). Therefore, the results of the pre- and post-test are presented in three parts following the order of the subtests. In Section 4.1.4, the overall mean scores of the three contrast from pretest to posttest were compared.

4.1.1. /i/-/ɪ/ Contrast

In order to evaluate the statistical significance of the difference in the mean scores, a one-way analysis of covariance (ANCOVA) was conducted at a significance level of .05. The independent variable, Group, includes three levels: CON, ID, and SD. The dependent variable is the participants’ posttest scores and the covariate is the participants’ pretest scores. The results are shown in Table 4.1 and Table 4.2.
Table 4.1
**ANCOVA Results and Descriptive Statistics for /i/-/ɪ/ Contrast by Group and Posttest Scores**

<table>
<thead>
<tr>
<th>Group</th>
<th>Posttest Score</th>
<th>Observed Mean</th>
<th>Adjusted Mean</th>
<th>SD</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>CON</td>
<td>51.17</td>
<td>50.60</td>
<td>11.24</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>ID</td>
<td>67.04</td>
<td>67.93</td>
<td>15.58</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>SD</td>
<td>64.43</td>
<td>64.11</td>
<td>13.24</td>
<td>23</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Posttest</td>
<td>3476.65</td>
<td>1</td>
<td>3476.65</td>
<td>26.27*</td>
</tr>
<tr>
<td>Group</td>
<td>3948.44</td>
<td>2</td>
<td>1974.22</td>
<td>14.92*</td>
</tr>
<tr>
<td>Error</td>
<td>8867.29</td>
<td>67</td>
<td>132.35</td>
<td></td>
</tr>
</tbody>
</table>

*Note. R² = .44, Adj. R² = .41, adjustments based on pretest scores mean of 51.82. Homogeneity of regression tested and not significant: F = 3.04, p = .054>.05. * p < .05

Table 4.2
**Multiple Comparisons and Mean Differences in /i/-/ɪ/ Contrast by Group and Controlling for Pretest Scores**

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Mean Difference</th>
<th>s.e.</th>
<th>Bonferroni Adjusted 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>CON vs. ID</td>
<td>-17.33*</td>
<td>3.33</td>
<td>-25.51, -9.14</td>
</tr>
<tr>
<td>CON vs. SD</td>
<td>-13.51*</td>
<td>3.36</td>
<td>-21.76, -5.27</td>
</tr>
<tr>
<td>ID vs. SD</td>
<td>3.82</td>
<td>3.37</td>
<td>-4.45, 12.08</td>
</tr>
</tbody>
</table>

*Note. Comparisons based upon ANCOVA adjusted means controlling for pretest scores mean of 51.82. CON = Control Group; ID = Identification Group; SD = Same/Different Group * p < .05, where p-values are adjusted using the Bonferroni method.

One-way ANCOVA results showed that there were statistically significant differences among groups at posttest, indicating a significant main effect for training (Table 4.1.). Multiple comparisons revealed significant differences between the control...
group and each of the two training groups. However, the mean difference between the ID group and the SD group was not statistically significant (Table 4.2.). Both the observed and adjusted means showed that participants in the ID group performed best, followed by participants in the SD group, with participants in the control group performing worst.

The results showed that the three groups had significant difference in their performance at posttest, but the difference between the two experimental groups was not significant. Thus, the null hypothesis predicting that there would be no significant differences among groups’ performance at posttest is rejected. Both the ID and SD groups outperformed the control group in identifying the /i/-/ɪ/ contrast, indicating that both training methods had a significant effect on performance in identifying the /i/-/ɪ/ contrast.

4.1.2. /ɛ/-/æ/ Contrast

A one-way analysis of covariance (ANCOVA) was conducted at a significance level of .05. The independent variable, Group, includes three levels: CON, ID, and SD. The dependent variable is the participants’ posttest scores and the covariate is the participants’ pretest scores. The results are shown in Table 4.3 and Table 4.4.
Table 4.3
**ANCOVA Results and Descriptive Statistics for /ɛ/-/æ/ Contrast by Group and Posttest Scores**

<table>
<thead>
<tr>
<th>Group</th>
<th>Observed Mean</th>
<th>Adjusted Mean</th>
<th>SD</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>CON</td>
<td>60.54</td>
<td>61.62</td>
<td>15.39</td>
<td>24</td>
</tr>
<tr>
<td>ID</td>
<td>71.88</td>
<td>71.11</td>
<td>12.97</td>
<td>24</td>
</tr>
<tr>
<td>SD</td>
<td>69.04</td>
<td>68.72</td>
<td>11.53</td>
<td>23</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Posttest</td>
<td>4460.01</td>
<td>1</td>
<td>4460.01</td>
<td>38.40*</td>
</tr>
<tr>
<td>Group</td>
<td>1156.81</td>
<td>2</td>
<td>578.41</td>
<td>4.98*</td>
</tr>
<tr>
<td>Error</td>
<td>7781.53</td>
<td>67</td>
<td>116.14</td>
<td></td>
</tr>
</tbody>
</table>

*Note.* $R^2 = .44$, Adj. $R^2 = .42$, adjustments based on pretest scores mean of 62.52. Homogeneity of regression tested and not significant: $F = 2.90$, $p = .062>.05$.

One-way ANCOVA results showed that there were statistically significant differences among groups at posttest, indicating a significant main effect for training (Table 4.3). Multiple comparisons revealed significant differences between the control
group and the ID group. However, the mean difference between the ID group and the SD group and that between the control group and the SD group were not statistically significant (Table 4.4). Both the observed and adjusted means showed that participants in the ID group performed best, followed by participants in the SD group, with participants in the control group performing worst.

The results suggest that the null hypothesis predicting that there would be no significant differences among groups’ performance at posttest is rejected. Although both the ID and SD groups outperformed the control group, the mean difference between the ID group and the SD group and that between the control group and the SD group are not statistically significant. Thus, only the ID training had a significant effect on performance in identifying the /ɛ/-/æ/ contrast.

4.1.3. Word-Final /t/-/d/ Contrast

Levene’s Test of Equality of Error Variances showed that the assumption of homogeneity of regression slopes was violated [F(2, 68) = 3.24, p = .045 < .05]. Therefore, the Johnson-Neyman technique (John & Neyman, 1936) was conducted to provide regions of significance. Johnson-Neyman technique software (Wu, 2007) was used for a series of analyses. Since the software can only analyze two groups that have the same number of participants, the mean score of the SD group was added as a case so that there was equal number of participants in each group and then the data were analyzed pair by pair.

4.1.3.1. ID Group vs. Control Group

The results indicated a statistically significant difference in pretest word-final /t/-/d/ scores when the scores were 40.30 or smaller and 83.02 or greater (Table 4.5, Table 4.6).
With 95% confidence, it was concluded that when a participant’s pretest score was less than 40.30, the control group training (no training) was more effective. On the other hand, when a participant’s pretest score was greater than 83.02, the ID training was more effective. When a participant’s pretest score was between 40.30 and 83.02, there was no statistically significant difference between the two training methods.

Table 4.5
Johnson Neyman Technique Results for ID Group and Control Group

<table>
<thead>
<tr>
<th></th>
<th>SSw(x)</th>
<th>SSw(y)</th>
<th>CPwj</th>
<th>df</th>
<th>ss&quot;w(y)</th>
<th>df</th>
<th>bwj</th>
<th>awj</th>
</tr>
</thead>
<tbody>
<tr>
<td>ID</td>
<td>5383.83</td>
<td>3769.62</td>
<td>2402.25</td>
<td>23</td>
<td>2697.75</td>
<td>22</td>
<td>.45</td>
<td>44.31</td>
</tr>
<tr>
<td>CON</td>
<td>1125.33</td>
<td>1730.96</td>
<td>460.83</td>
<td>23</td>
<td>1542.24</td>
<td>22</td>
<td>.41</td>
<td>30.80</td>
</tr>
<tr>
<td></td>
<td>6509.17</td>
<td>5500.58</td>
<td>2863.08</td>
<td>46</td>
<td>6509.17</td>
<td>5500.58</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note.* bwj = slope. awj = intercept

Table 4.6
Johnson Neyman Technique Interaction and Points of Significance for ID Group and Control Group

<table>
<thead>
<tr>
<th></th>
<th>Xo</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>XD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-368.32</td>
<td>-.46</td>
<td>28.61</td>
<td>-1552.25</td>
<td>40.30</td>
</tr>
</tbody>
</table>

*Note.* Xo = intersection of two regression lines. XD = points of significance

Since there were no participants who scored higher than 83.02, it was concluded that the ID training was not effective in identifying the word-final /t/-/d/ contrast. The relationship between posttest scores and pretest scores by the two groups on is shown in
4.1.3.2. SD Group vs. Control Group

The results indicated a statistically significant difference in pretest word-final /t/-/d/ scores when the scores were 52.48 or smaller and 86.06 or greater (Table 4.7, Table 4.8). With 95% confidence, it was concluded that when a participant’s pretest score was less than 52.48, the control group training (no training) was more effective. On the other hand, when a participant’s pretest score was greater than 86.06, the SD training was more effective. When a participant’s pretest score was between 52.48 and 86.06, there was no statistically significant difference between the two training methods.
Table 4.7
Johnson Neyman Technique Results for SD Group and Control Group

<table>
<thead>
<tr>
<th></th>
<th>SSw(x)</th>
<th>SSw(y)</th>
<th>CPwj</th>
<th>df</th>
<th>ss&quot;w(y)</th>
<th>df</th>
<th>bwj</th>
<th>awj</th>
</tr>
</thead>
<tbody>
<tr>
<td>SD</td>
<td>1258.96</td>
<td>4521.48</td>
<td>1103.61</td>
<td>23</td>
<td>3554.05</td>
<td>22</td>
<td>.88</td>
<td>15.93</td>
</tr>
<tr>
<td>CON</td>
<td>1125.33</td>
<td>1730.96</td>
<td>460.83</td>
<td>23</td>
<td>1542.24</td>
<td>22</td>
<td>.41</td>
<td>30.80</td>
</tr>
<tr>
<td></td>
<td>2384.29</td>
<td>6252.44</td>
<td>1564.44</td>
<td>46</td>
<td>5096.29</td>
<td>44</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note.* bwj = slope. awj = intercept

Table 4.8
Johnson Neyman Technique Interaction and Points of Significance for SD Group and Control Group

<p>| | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Xo</td>
<td>31.83</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>-.66</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>45.57</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>-2971.40</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>XD</td>
<td>52.48</td>
<td>86.06</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note.* Xo = intersection of two regression lines. XD = points of significance

Since there were no participants who scored higher than 86.06, it was concluded that the SD training was not effective in identifying the word-final /t/-/d/ contrast. The relationship between posttest scores and pretest scores by the two groups is shown in Figure 4.2.
4.1.3.3. ID Group vs. SD Group

The results indicated a statistically significant difference in pretest word-final /t/-/d/ scores when the scores were 40.23 or smaller and 80.04 or greater (Table 4.9, Table 4.10). With 95% confidence, it was concluded that when a participant’s pretest score was less than 40.23, the ID training was more effective. On the other hand, when a participant’s pretest score was greater than 80.04, the SD training was more effective. When a participant’s pretest score was between 40.23 and 80.04, there was no statistically significant difference between the two training methods.
Table 4.9
Johnson Neyman Technique Results for ID Group and SD Group

<table>
<thead>
<tr>
<th></th>
<th>SSw(x)</th>
<th>SSw(y)</th>
<th>CPwj</th>
<th>df</th>
<th>ss&quot;w(y)</th>
<th>df</th>
<th>bwj</th>
<th>awj</th>
</tr>
</thead>
<tbody>
<tr>
<td>ID</td>
<td>5383.83</td>
<td>3769.62</td>
<td>2402.25</td>
<td>23</td>
<td>2697.75</td>
<td>22</td>
<td>.45</td>
<td>44.31</td>
</tr>
<tr>
<td>SD</td>
<td>1258.96</td>
<td>4521.48</td>
<td>1103.61</td>
<td>23</td>
<td>3554.05</td>
<td>22</td>
<td>.88</td>
<td>15.93</td>
</tr>
<tr>
<td></td>
<td>6642.79</td>
<td>8291.10</td>
<td>3505.86</td>
<td>46</td>
<td>6251.80</td>
<td>44</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note.* bwj = slope. awj = intercept

Table 4.10
Johnson Neyman Technique Interaction and Points of Significance for ID Group and SD Group

<table>
<thead>
<tr>
<th></th>
<th>Xo</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>XD</th>
</tr>
</thead>
<tbody>
<tr>
<td>ID</td>
<td>65.94</td>
<td>-.38</td>
<td>21.06</td>
<td>-1200.95</td>
<td>40.23</td>
</tr>
<tr>
<td>SD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>80.04</td>
</tr>
</tbody>
</table>

*Note.* Xo = intersection of two regression lines. XD = points of significance

The relationship between posttest scores and pretest scores by the two groups is shown in Figure 4.3.
4.1.4. Overall Comparison

In order to evaluate the statistical significance of the difference in the overall mean scores, a one-way analysis of covariance (ANCOVA) was conducted at a significance level of .05. The independent variable, Group, includes three levels: CON, ID, and SD. The dependent variable is the participants’ overall posttest scores and the covariate is the participants’ overall pretest scores. The results are shown in Table 4.11 and Table 4.12.
Table 4.11
* ANCOVA Results and Descriptive Statistics for All Contrasts by Group and Posttest Scores *

<table>
<thead>
<tr>
<th>Group</th>
<th>Posttest Score</th>
<th>Observed Mean</th>
<th>Adjusted Mean</th>
<th>SD</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>CON</td>
<td>55.81</td>
<td>55.55</td>
<td>9.38</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>ID</td>
<td>69.76</td>
<td>70.16</td>
<td>11.10</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>SD</td>
<td>67.03</td>
<td>66.89</td>
<td>9.96</td>
<td>23</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Posttest</td>
<td>3126.82</td>
<td>1</td>
<td>3126.82</td>
<td>53.47*</td>
</tr>
<tr>
<td>Group</td>
<td>2810.75</td>
<td>2</td>
<td>1405.37</td>
<td>24.03*</td>
</tr>
<tr>
<td>Error</td>
<td>3917.92</td>
<td>67</td>
<td>58.48</td>
<td></td>
</tr>
</tbody>
</table>

*Note. R^2 = .60, Adj. R^2 = .58, adjustments based on pretest scores mean of 57.92. Homogeneity of regression tested and not significant: F = .41, p = .664>.05.*

* p < .05

Table 4.12
* Multiple Comparisons and Mean Differences in All Contrasts by Group and Controlling for Pretest Scores *

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Mean Difference</th>
<th>s.e.</th>
<th>Bonferroni Adjusted 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>CON vs. ID</td>
<td>-14.61*</td>
<td>2.21</td>
<td>-20.03, -9.18</td>
</tr>
<tr>
<td>CON vs. SD</td>
<td>-11.34*</td>
<td>2.23</td>
<td>-16.82, -5.86</td>
</tr>
<tr>
<td>ID vs. SD</td>
<td>3.27</td>
<td>2.23</td>
<td>-2.21, 8.75</td>
</tr>
</tbody>
</table>

*Note. Comparisons based upon ANCOVA adjusted means controlling for pretest scores mean of 57.92. CON = Control Group; ID = Identification Group; SD = Same/Different Group

* p < .05, where p-values are adjusted using the Bonferroni method.

One-way ANCOVA results showed that there were statistically significant differences among groups at posttest, indicating a significant main effect for training (Table 4.11). Multiple comparisons revealed significant differences between the control group and each of the two training groups. However, the mean difference between the ID group and the SD group was not statistically significant (Table 4.12). Both the observed
and adjusted means showed that participants in the ID group performed best, followed by participants in the SD group, with participants in the control group performing worst.

The results showed that the three groups had significant difference in their performance at posttest, but the difference between the two experimental groups was not significant. Thus, the null hypothesis predicting that there would be no significant differences among groups’ performance at posttest is rejected. Both the ID and SD groups outperformed the control group in identifying the three contrasts, indicating that both training methods had a significant effect on performance in identifying the three contrasts.

A factorial two-way ANOVA was conducted. The ANOVA revealed a main effect of group, $F(2, 65) = 10.108, p = .000$. In contrast, it failed to reveal a main effect of gender, $F(1, 65) = .559, MSe = 102.76, p = .458, \alpha = .05$ and showed no main effect of an interaction of group * gender $F(2, 65) = 1.467, MSe = 102.76, p = .238, \alpha = .05$.

Further, Cohen’s effect size value ($d = 1.36$) of the ID group compared to the control group suggested a high practical significance. In comparison, Cohen’s effect size value ($d = 1.16$) of the SD group compared to the control group also suggested a high practical significance. The large effect sizes of both trained groups indicate that it is worthwhile to receive either of the training methods.

4.2. Research Question 2: Comparison of Two Training Methods

Research question 2: Which training method, ID or SD, will be more effective for training EFL students in Taiwan in identifying three non-native phonemic contrasts?

Concurring with Flege (1995a), I predicted that identification training would be
more effective than same/different training. Therefore, the null hypothesis predicted that there would be no significant differences among the trained groups’ performance at posttest.

Although the results in the last section showed that the ID group outperformed the SD group in identifying all the three contrasts [4% higher on the /i/-/i/ contrast, 2% higher on the /e/-/æ/ contrast, 3% higher on the word-final /t/-/d/ contrast, and 3% on the overall scores], the differences were not statistically significant (Table 4.2, Table 4.4, Table 4.10, and Table 4.12). Therefore, the null hypothesis predicting that there would be no significant differences among the trained groups’ performance at posttest is accepted.

4.3. Research Question 3: Generalization to New Tokens and New Talkers
Research question 3: Can the effect of perceptual training on identifying three non-native phonemic contrasts be generalized to new tokens produced by familiar talkers and familiar tokens produced by new talkers?

Judging by the results of previous studies, I expected the perceptual training in identifying the three non-native phonemic contrasts to be generalized to new tokens produced by familiar talkers and familiar tokens produced by new talkers. Therefore, the null hypothesis is that there would be no significant differences among groups’ performance on identifying new tokens produced by familiar talkers and familiar tokens produced by new talkers. The results of generalization to new tokens will be presented first, followed by the results of generalization to new talkers.

4.3.1. Generalization to New Tokens
The mean percentage correct identification scores for new tokens on the
generalization test by the trained and the control groups are presented in Figure 4.4. A series of one-way ANOVA using a significance level of .05 were conducted to evaluate if the effect of perceptual training could be generalized to new tokens.

![Figure 4.4](image)

*Figure 4.4 Mean Percentage Correct Identification Scores for New Tokens on the Generalization Test by the Trained and the Control Groups*

The results indicated a significant main effect for training as the differences among groups were significant on all three contrasts: [F(2, 68) = 10.116, p = .000] for the /i/-/ɪ/ contrast, [F(2, 68) = 5.876, p = .004] for the /ε/-/æ/ contrast, and [F(2, 68) = 4.313, p = .017] for the word-final /t/-/d/ contrast. Post hoc analyses using Bonferroni at a significance level of .05 indicated significant differences between the control group and each of the two trained groups for the /i/-/ɪ/ contrast and the /ε/-/æ/ contrast. However, for the word-final /t/-/d/ contrast, the mean difference between the control group and the SD
group was not statistically significant (p = .141).

Again, the results showed that the mean difference between the ID group and the SD group was not statistically significant (p = .858 for the /i/-/ɪ/ contrast, p = 1.000 for the /ɛ/-/æ/ contrast, p = 1.000 for the word-final /t/-/d/ contrast). The results showed that participants in the ID group performed best on all three contrasts, followed by participants in the SD group, with participants in the control group performing worst. Thus, the null hypothesis predicting that there would be no significant differences among groups’ performance on identifying new tokens produced by familiar talkers is rejected. The results showed that the effect of perceptual training on identifying the three non-native phonemic contrasts was generalized to new tokens produced by familiar talkers, except for the SD group in the word-final /t/-/d/ contrast.

4.3.2. Generalization to New Talkers

The mean percentage correct identification scores for new talkers on the generalization test by the trained and the control groups are presented in Figure 4.5. A series of one-way ANOVA using a significance level of .05 were conducted to evaluate if the effect of perceptual training could be generalized to new talkers.
The results indicated a significant main effect for training for the differences among groups were significant on all three contrasts: $[F(2,68) = 9.714, p = .000]$ for the /i/-/ɪ/ contrast, $[F(2, 68) = 5.284, p = .007]$ for the /ε/-/æ/ contrast, and $[F(2, 68) = 8.724, p = .000]$ for the word-final /t/-/d/ contrast. Post hoc analyses using Bonferroni at a significance level of .05 indicated significant differences between the control group and each of the two trained groups for the /i/-/ɪ/ contrast and word-final /t/-/d/ contrast. However, for the /ε/-/æ/ contrast, the mean difference between the control group and the SD group was not statistically significant ($p = .105$).

Again, the results showed that the mean difference between the ID group and the SD group was not statistically significant ($p = 1.000$ for the /i/-/ɪ/ contrast, $p = .971$ for the /ε/-/æ/ contrast, $p = .884$ for the word-final /t/-/d/ contrast). Participants in the ID group performed best on the /ε/-/æ/ and the word-final /t/-/d/ contrasts, and participants in the
SD group performed best on the /i/-/ɪ/ contrast, while participants in the control group performed worst in all three contrasts. Thus, the null hypothesis predicting that there would be no significant differences among groups’ performance on identifying familiar tokens produced by new talkers is rejected. The results showed that the effect of perceptual training on identifying the three non-native phonemic contrasts was generalized to familiar tokens produced by new talkers, except for the SD group in the /ɛ/-/æ/ contrast.

4.4. Research Question 4

Research question 4: Can the effect of perceptual training on identifying the word-final /t/-/d/ contrast be generalized to other final stop contrasts?

I predicted that the effect of perceptual training on identifying the word-final /t/-/d/ contrast would be generalized to other final stop contrasts, i.e., the /k/-/g/ and the /p/-/b/ endings. Thus, the null hypothesis is that there would be no significant differences among groups’ performance on identifying tokens with the /k/-/g/ or the /p/-/b/ endings.

The mean percentage correct identification scores for the /k/-/g/ and the /p/-/b/ endings on the generalization test by the trained and the control groups are presented in Figure 4.6.
A series of one-way ANOVA using a significance level of .05 were conducted to evaluate if the effect of perceptual training could be generalized to new talkers. The results indicated a significant main effect for training as the differences among groups were significant on the /k/-/g/ endings, \[F(2, 68) = 4.246, p = .018\], but not on the /p/-/b/ endings, \[F(2, 68) = .882, p = .419\]. Post hoc analyses using Bonferroni at a significance level of .05 indicated a significant difference between the control group and the ID group \((p = .029)\) but no significant difference between the control group and the SD group \((p = .066)\) for the /k/-/g/ endings. With respect to the /p/-/b/ endings, the mean difference between the control group and the ID group \((p = .572)\) and that between the control group and the SD group \((p = 1.000)\) were not statistically significant.

The results also showed that the mean difference between the ID group and the SD
group was not statistically significant (p = 1.000 for both the /k/-/g/ and the /p/-/b/ endings). Participants in the ID group performed best on both contrasts, followed by participants in the SD group, with participants in the control group performing worst. Thus, the null hypothesis predicting that there would be no significant differences among groups’ performance on identifying other final stop contrasts is rejected for the /k/-/g/ endings but accepted for the /p/-/b/ endings. The results showed that the effect of perceptual training on identifying the /t/-/d/ endings was generalized to the /k/-/g/ endings for the ID group only but not generalized to the /p/-/b/ endings for both trained groups.

4.5. Research Question 5: Retention of Training Effects

Research question 5: Can the effect of the training methods persist three months after the completion of the training?

Judging by the results of previous studies that trainees retained the effect of perceptual training (e.g., Lively, Logan, & Pisoni, 1994; Flege, 1995a; Bradlow et al., 1999; Wang, 2002), I predicted that the effects of the training methods would persist for more than three months after the completion of the training. The results in Section 5.1 showed that although some of the gain differences between each of the trained groups and the control group were not significant, both trained groups did obtain scores higher than the control group in every subtest. If the effect of perceptual training was retained, there would be no significant difference between the posttest scores and the follow-up posttest scores. Thus, the null hypothesis is that there would be a negative significant difference between the posttest scores and the follow-up posttest scores as well as between the two generalization tests in each group.
To assess retention, a series of paired-samples t-tests were conducted to examine if the score differences between the posttest and the follow-up posttest were significant. For the trained contrasts, if there was a significant decrease from the posttest to the follow-up posttest, a paired-samples t-test was conducted to examine if the score differences between the pretest (with drop-outs deleted) and the follow-up posttest were significant. Both groups’ mean percentage correct identification scores at posttest and three-month follow-up are presented in Table 4.13 and Table 4.14 for comparison.

Table 4.13
Mean Percentage Correct Identification Scores from Posttest to Three-month Follow-up by ID Group (n = 21)

<table>
<thead>
<tr>
<th></th>
<th>Posttest</th>
<th>3-month Follow-up</th>
<th>Mean Difference</th>
<th>SD</th>
<th>df</th>
<th>s.e.</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>67.90</td>
<td>63.19</td>
<td>4.71</td>
<td>9.08</td>
<td>20</td>
<td>1.98</td>
</tr>
<tr>
<td>P2</td>
<td>71.86</td>
<td>67.48</td>
<td>4.38</td>
<td>9.40</td>
<td>20</td>
<td>2.05</td>
</tr>
<tr>
<td>P3</td>
<td>70.90</td>
<td>67.00</td>
<td>3.90</td>
<td>11.60</td>
<td>20</td>
<td>2.53</td>
</tr>
<tr>
<td>NTOP1</td>
<td>71.43</td>
<td>66.43</td>
<td>5.00</td>
<td>8.12</td>
<td>20</td>
<td>1.77</td>
</tr>
<tr>
<td>NTOP2</td>
<td>71.86</td>
<td>69.57</td>
<td>2.29</td>
<td>7.73</td>
<td>20</td>
<td>1.69</td>
</tr>
<tr>
<td>NTOP3</td>
<td>70.43</td>
<td>67.33</td>
<td>3.10</td>
<td>7.42</td>
<td>20</td>
<td>1.62</td>
</tr>
<tr>
<td>NTAP1</td>
<td>65.57</td>
<td>64.14</td>
<td>1.43</td>
<td>11.22</td>
<td>20</td>
<td>2.45</td>
</tr>
<tr>
<td>NTAP2</td>
<td>70.38</td>
<td>69.52</td>
<td>.86</td>
<td>6.99</td>
<td>20</td>
<td>1.52</td>
</tr>
<tr>
<td>NTAP3</td>
<td>70.86</td>
<td>67.76</td>
<td>3.10</td>
<td>7.89</td>
<td>20</td>
<td>1.72</td>
</tr>
<tr>
<td>/k/-/g/</td>
<td>65.52</td>
<td>64.48</td>
<td>1.05</td>
<td>5.90</td>
<td>20</td>
<td>1.29</td>
</tr>
<tr>
<td>/p/-/b/</td>
<td>64.76</td>
<td>64.71</td>
<td>.05</td>
<td>6.42</td>
<td>20</td>
<td>1.40</td>
</tr>
</tbody>
</table>

Table 4.14
Mean Percentage Correct Identification Scores from Posttest to Three-month Follow-up by SD Group (n = 20)

<table>
<thead>
<tr>
<th></th>
<th>Posttest</th>
<th>3-month Follow-up</th>
<th>Mean Difference</th>
<th>SD</th>
<th>df</th>
<th>s.e.</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>65.30</td>
<td>63.85</td>
<td>1.45</td>
<td>11.12</td>
<td>19</td>
<td>2.48</td>
</tr>
<tr>
<td>P2</td>
<td>69.56</td>
<td>68.40</td>
<td>1.15</td>
<td>8.68</td>
<td>19</td>
<td>1.94</td>
</tr>
<tr>
<td>P3</td>
<td>66.90</td>
<td>65.30</td>
<td>1.60</td>
<td>12.10</td>
<td>19</td>
<td>2.71</td>
</tr>
<tr>
<td>NTOP1</td>
<td>66.85</td>
<td>64.35</td>
<td>2.50</td>
<td>12.12</td>
<td>19</td>
<td>2.71</td>
</tr>
<tr>
<td>NTOP2</td>
<td>70.25</td>
<td>67.40</td>
<td>2.85</td>
<td>11.27</td>
<td>19</td>
<td>2.52</td>
</tr>
<tr>
<td>NTOP3</td>
<td>64.80</td>
<td>67.10</td>
<td>-2.30</td>
<td>11.93</td>
<td>19</td>
<td>2.67</td>
</tr>
<tr>
<td>NTAP1</td>
<td>65.50</td>
<td>64.65</td>
<td>.85</td>
<td>12.86</td>
<td>19</td>
<td>2.88</td>
</tr>
<tr>
<td>NTAP2</td>
<td>67.00</td>
<td>65.00</td>
<td>2.00</td>
<td>8.39</td>
<td>19</td>
<td>1.88</td>
</tr>
<tr>
<td>NTAP3</td>
<td>67.40</td>
<td>67.80</td>
<td>-.40</td>
<td>12.63</td>
<td>19</td>
<td>2.83</td>
</tr>
<tr>
<td>/k/-/g/</td>
<td>63.25</td>
<td>63.00</td>
<td>.25</td>
<td>11.11</td>
<td>19</td>
<td>2.48</td>
</tr>
<tr>
<td>/p/-/b/</td>
<td>62.90</td>
<td>64.30</td>
<td>-1.40</td>
<td>12.52</td>
<td>19</td>
<td>2.80</td>
</tr>
</tbody>
</table>


The results of analyses are presented by group as follows.

4.5.1. ID Group Retention of Training Effects

4.5.1.1. ID Retention of Training Effects at Posttest

First, a paired-samples t-test indicated that there was a significant difference in the scores between the posttest (M=67.90, SD=15.35) and the follow-up posttest (M=63.19, SD=13.86) for the /i/-/ɪ/ contrast; t(20)=2.380, p = .027. However, a paired-samples t-test indicated that there was a significant difference in the scores between the pretest (M=51.67, SD=10.81) and the follow-up posttest (M=63.19, SD=13.86) for the /i/-/ɪ/ contrast; t(20)= 3.388, p = .003. The results showed that although there was a significant decrease from the posttest to the follow-up posttest, the increase from the pretest to the follow-up posttest was still significant, indicating that the ID training effect on the /i/-/ɪ/ contrast was retained three months after the completion of the training. Thus, although
the null hypothesis predicting that there would be a negative significant difference
between the posttest scores and the follow-up posttest scores is accepted, the training
effect the /i/-/ɪ/ contrast was still retained when the follow-up scores was compared to the
pretest scores.

Second, a paired-samples t-test indicated that there was a significant difference in
the scores between posttest (M=71.86, SD=13.53) and the follow-up posttest (M=67.48,
SD=12.77) for the /ɛ/-/æ/ contrast; t(20)=2.14, p = .045. Next, a paired-samples t-test
showed that there was no significant difference in the scores between the pretest
(M=63.86, SD=11.59) and the follow-up posttest (M=63.19, SD=13.86) for the /ɛ/-/æ/
contrast; t(20)= 1.823, p = .083. The results indicated that there was a significant
decrease from the posttest to the follow-up posttest and the mean difference between the
pretest and the follow-up posttest was not significant, indicating that the ID training
effect for the /ɛ/-/æ/ contrast was not retained three months after the completion of the
training. Thus, the null hypothesis predicting that there would be a negative significant
difference between the posttest scores and the follow-up posttest scores for the /ɛ/-/æ/
contrast is accepted.

Third, a paired-samples t-test indicated that there was no significant difference in the
scores between the posttest (M=70.90, SD=12.84) and the follow-up posttest (M=67.00,
SD=12.46) for the word-final /t/-/d/ contrast; t(20)=2.380, p = .139. Next, a
paired-samples t-test indicated that there was a significant difference in the scores
between the pretest (M=57.10, SD=15.58) and the follow-up posttest (M=63.19,
SD=13.86) for the word-final /t/-/d/ contrast; t(20)= 3.051, p = .006. The results showed
that the ID training effect for the word-final /t/-/d/ contrast was retained three months
after the completion of the training. Thus, the null hypothesis predicting that there would
be a negative significant difference between the posttest scores and the follow-up posttest
scores is rejected.

4.5.1.2. ID Retention of Training Effects on Generalization to New Tokens

First, a paired-samples t-test indicated that there was a significant negative
difference in the scores at posttest (M=71.43, SD=12.41) and the follow-up posttest
(M=66.43, SD=14.33) for the /i/-/ɪ/ contrast; t(20)=2.823, p = .011.

Second, a paired-samples t-test indicated that there was no significant difference in
the scores at posttest (M=71.86, SD=10.26) and the follow-up posttest (M=69.57,
SD=11.73) for the /ɛ/-/æ/ contrast; t(20)=1.35, p = .191.

Third, a paired-samples t-test indicated that there was no significant difference in the
scores at posttest (M=70.90, SD=12.84) and the follow-up posttest (M=67.00, SD=12.46)
for the word-final /t/-/d/ contrast; t(20)=1.91, p = .070.

The results above showed that the ID training effect generalized to new tokens
appeared to have been retained three months after the completion of the training for the
/ɛ/-/æ/ contrast and the word-final /t/-/d/ contrast, but not for the /i/-/ɪ/ contrast. Thus, the
null hypothesis predicting that there would be a negative significant difference between
the posttest scores and the follow-up posttest scores appears rejected for the /ɛ/-/æ/
contrast and the word-final /t/-/d/ contrast, but appears accepted for the /i/-/ɪ/ contrast.

4.5.1.3. ID Retention of Training Effects on Generalization to New Talkers

First, a paired-samples t-test indicated that there was no significant difference in the
scores between the posttest (M=65.57, SD=13.38) and the follow-up posttest (M=64.14,
SD=13.78) for the /i/-/ɪ/ contrast; t(20)=.584, p = .566.
Second, a paired-samples t-test indicated that there was no significant difference in the scores at posttest (M=70.38, SD=10.68) and the follow-up posttest (M=69.52, SD=9.44) for the /ε/-/æ/ contrast; t(20)=.562, p = .580.

Third, a paired-samples t-test indicated that there was no significant difference in the scores at posttest (M=70.56, SD=13.58) and the follow-up posttest (M=67.76, SD=12.71) for the word-final /t/-/d/ contrast; t(20)=1.80, p = .087.

The results above showed that the ID training effect generalized to new talkers appeared to have been retained three months after the completion of the training for all three contrasts. Thus, the null hypothesis predicting that there would be a negative significant difference between the posttest scores and the follow-up posttest scores appears rejected for all three contrasts.

4.5.1.4. ID Retention of Training Effects on Generalization to Other Final Stop Contrasts

The results in Section 5.4 showed that the ID training effect on identifying the /t/-/d/ endings was generalized to the /k/-/g/ endings but not to the /p/-/b/ endings. Therefore, only the retention of the /k/-/g/ endings was investigated here.

A paired-samples t-test indicated that there was no significant difference in the scores at posttest (M=65.52, SD=10.16) and the follow-up posttest (M=64.48, SD=9.39) for the word-final /t/-/d/ contrast; t(20)=.426, p = .426.

The results above showed that the ID training effect generalized to the /k/-/g/ endings appeared to have been retained three months after the completion of the training. Thus, the null hypothesis predicting that there would be a negative significant difference between the posttest scores and the follow-up posttest scores appears rejected for the
Next, the analyses of the retention of SD training effect are presented.

4.5.2. SD Group Retention of Training Effects

4.5.2.1. SD Retention of Training Effects at Posttest

First, a paired-samples t-test indicated that there was no significant difference in the scores at posttest (M=65.30, SD=13.93) and the follow-up posttest (M=63.85, SD=12.19) for the /i/-/ɪ/ contrast; t(19)=2.380, p = .567. Thus, the null hypothesis predicting that there would be a negative significant difference between the posttest scores and the follow-up posttest scores appears rejected. The result indicated that the SD training effect for the /i/-/ɪ/ contrast appeared to have been retained three months after the completion of the training.

Second, a paired-samples t-test indicated that there was a significant difference in the scores at posttest (M=71.86, SD=13.53) and the follow-up posttest (M=67.48, SD=12.77) for the /ε/-/æ/ contrast; t(19)=2.14, p = .045. Next, a paired-samples t-test indicated that there was a significant difference in the scores between the pretest (M=62.05, SD=8.79) and the follow-up posttest (M=63.19, SD=13.86) for the /ε/-/æ/ contrast; t(19)= 2.224, p = .038. The results showed that although there was a negative significant decrease from the posttest to the follow-up posttest, the increase from the pretest to the follow-up posttest was still significant, indicating that the SD training effect for the /ε/-/æ/ contrast was retained three months after the completion of the training. Thus, although the null hypothesis predicting that there would be a negative significant difference between the posttest scores and the follow-up posttest scores is accepted, the training effect was still retained when the follow-up scores was compared to the pretest scores.
Third, a paired-samples t-test indicated that there was no significant difference in the scores at posttest (M=70.90, SD=12.84) and the follow-up posttest (M=67.00, SD=12.46) for the word-final /t/-/d/ contrast; t(19)=2.380, p = .139. However, the SD training effect was not significant from the pretest to the posttest. Therefore, a paired-samples t-test was conducted to investigate if there was a significant difference between the pretest and the follow-up test scores. The result showed that there was no significant difference in the scores between the pretest (M=63.86, SD=11.59) and the follow-up posttest (M=63.19, SD=13.86) for the word-final /t/-/d/ contrast; t(19)= 1.422, p = .171, indicating that the SD training effect, if any, for the word-final /t/-/d/ contrast was not retained three months after the completion of the training. Thus, although the null hypothesis predicting that there would be a negative significant difference between the posttest scores and the follow-up posttest scores is accepted, there was no training effect to be retained. A further test showed there was no significant difference in the scores between the pretest and the follow-up posttest for the word-final /t/-/d/ contrast, indicating that there was no training effect to be retained.

4.5.2.2. SD Retention of Training Effects on Generalization to New Tokens

The results in Section 4.3.1 showed that the effect of SD perceptual training was not generalized to the word-final /t/-/d/ contrast in the subtest of new tokens. Therefore, only the retention of the /i/-/ɪ/ and the /ε/-/æ/ contrasts of new tokens was investigated here.

First, a paired-samples t-test indicated that there was no significant difference in the scores at posttest (M=66.85, SD=16.09) and the follow-up posttest (M=64.35, SD=14.62) for the /i/-/ɪ/ contrast; t(19)=.922, p = .368.

Second, a paired-samples t-test indicated that there was no significant difference in
the scores at posttest (M=71.86, SD=13.53) and the follow-up posttest (M=67.48, SD=12.77) for the /ɛ/-/æ/ contrast; t(19)=1.13, p = .272.

The results above showed that SD training effect on both the /i/-/ɪ/ and the /ɛ/-/æ/ contrasts of new tokens appeared to have been retained three months after the completion of the training. Thus, the null hypothesis predicting that there would be a negative significant difference between the posttest scores and the follow-up posttest scores appears rejected.

4.5.2.3. SD Retention of Training Effects on Generalization to New Talkers

The results in Section 4.3.2 showed that the effect of SD training was not generalized to the /ɛ/-/æ/ contrast. Therefore, only the retention of the /i/-/ɪ/ and the word-final /t/-/d/ contrasts of new talkers was investigated here.

First, a paired-samples t-test indicated that there was no significant difference in the scores at posttest (M=65.50, SD=13.57) and the follow-up posttest (M=64.65, SD=15.86) for the /i/-/ɪ/ contrast; t(19)=.296, p = .771.

Second, a paired-samples t-test indicated that there was no significant difference in the scores at posttest (M=67.40, SD=15.71) and the follow-up posttest (M=67.80, SD=15.40) for the word-final /t/-/d/ contrast; t(19)=-.142, p = .889.

The results above showed that the SD training effect on both the /i/-/ɪ/ and the word-final /t/-/d/ contrasts of new talkers appeared to have been retained three months after the completion of the training. Thus, the null hypothesis predicting that there would be a negative significant difference between the posttest scores and the follow-up posttest scores appears rejected for both the /i/-/ɪ/ and the word-final /t/-/d/ contrasts of new talkers.
4.5.2.4. SD Retention of Training Effects on Generalization to Other Final Stop Contrasts

The results in Section 4.4 showed that the effect of SD perceptual training on identifying the /t/-/d/ endings could not be generalized to the /k/-/g/ or the /p/-/b/ endings. Therefore, retention could not be investigated for both contrasts.

In sum, for both trained groups, most of the training effects at posttest appeared retained three months after the completion of the training.

4.6. Research Question 6: Relationship between Perceptual Training and Production

Research question 6: Can the perceptual training in identifying three non-native phonemic contrasts facilitate production of the non-native phonemic contrasts?

I predicted that the perceptual training in identifying the three non-native phonemic contrasts could facilitate production of the non-native phonemic contrasts. Thus, the null hypothesis is that there would be no significant differences among the three groups’ production scores at posttest.

A one-way analysis of covariance (ANCOVA) was conducted at a significance level of .05. The independent variable, Group, includes three levels: CON, ID, and SD. The dependent variable is the participants’ posttest production scores and the covariate is the participants’ pretest production scores. The results are shown in Table 4.15 and Table 4.16.
Table 4.15
ANCOVA Results and Descriptive Statistics for Production by Group and Posttest Scores

<table>
<thead>
<tr>
<th>Group</th>
<th>Observed Mean</th>
<th>Adjusted Mean</th>
<th>SD</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>CON</td>
<td>82.43</td>
<td>84.74</td>
<td>11.81</td>
<td>24</td>
</tr>
<tr>
<td>ID</td>
<td>84.19</td>
<td>84.76</td>
<td>15.47</td>
<td>24</td>
</tr>
<tr>
<td>SD</td>
<td>82.71</td>
<td>79.71</td>
<td>11.47</td>
<td>23</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Posttest</td>
<td>9996.86</td>
<td>1</td>
<td>9996.86</td>
<td>416.14*</td>
</tr>
<tr>
<td>Group</td>
<td>382.94</td>
<td>2</td>
<td>191.47</td>
<td>7.97*</td>
</tr>
<tr>
<td>Error</td>
<td>1609.52</td>
<td>67</td>
<td>24.02</td>
<td></td>
</tr>
</tbody>
</table>

Note. $R^2 = .862$, Adj. $R^2 = .856$, adjustments based on pretest scores mean of 82.39.
Homogeneity of regression tested and not significant: $F = .623$, $p = .540 > .05$.
* $p < .05$

Table 4.16
Multiple Comparisons and Mean Differences in Production by Group and Controlling for Pretest Scores

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Mean Difference</th>
<th>s.e.</th>
<th>Bonferroni Adjusted 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>CON vs. ID</td>
<td>-.014</td>
<td>1.42</td>
<td>-3.50, 3.47</td>
</tr>
<tr>
<td>CON vs. SD</td>
<td>5.03*</td>
<td>1.45</td>
<td>1.46, 8.60</td>
</tr>
<tr>
<td>ID vs. SD</td>
<td>5.04*</td>
<td>1.44</td>
<td>1.51, 8.58</td>
</tr>
</tbody>
</table>

Note. Comparisons based upon ANCOVA adjusted means controlling for pretest scores mean of 82.39. CON = Control Group; ID = Identification Group; SD = Same/Different Group
* $p < .05$, where p-values are adjusted using the Bonferroni method.

One-way ANCOVA results showed that there was a statistically significant difference among groups at posttest (Table 4.15). Multiple comparisons revealed significant differences between the control group and the SD group, and between the ID
group and the SD group. Also, the mean difference between the ID group and the control group was not statistically significant (Table 4.16). Both the observed and adjusted means showed that participants in the ID group performed best, followed by participants in the control group, with participants in the SD group performing worst. Therefore, the null hypothesis predicting that there would be no significant differences among groups’ production scores at posttest is rejected. The results showed that both the ID group and the control group had significantly higher production scores than the SD group and the mean difference between the ID group and the control group was not statistically significant. It is noteworthy that the control group, receiving no perceptual training, had significantly higher production scores than the SD group and performed almost as well as the ID group. Also, the SD group, in spite of training, showed a 3% decrease from the pretest to the posttest, indicating that the perceptual training did not facilitate production.

4.7. Research Question 7: Performance of Participants of Different L1 Backgrounds

Research question 7: Will the performance of the participants in the trained groups vary due to different L1 backgrounds?

I hypothesized that the performance of the participants in the trained groups would vary due to different L1 backgrounds, with native speakers of Taiwanese performing better than native speakers of Mandarin, as in the results of Flege (1989), Flege and Wang (1989), and Flege and Liu (2001). Thus, the null hypothesis predicted that there would be no significant differences among the performance of trained participants with different L1s at posttest.

Since the total number of native speakers of Mandarin-Hakka (n = 4) and Taiwanese
(n = 3) are under 5, only the performances of native speakers of Mandarin (n = 8) and native speakers of Mandarin-Taiwanese (n = 32) were compared. The results are presented in three parts following the order of the subtests.

4.7.1. /i/-/ɪ/ Contrast

A one-way analysis of covariance (ANCOVA) was conducted at a significance level of .05. The independent variable, L1, includes two levels: Mandarin and Mandarin-Taiwanese. The dependent variable is the participants’ posttest /i/-/ɪ/ subtest scores and the covariate is the participants’ pretest /i/-/ɪ/ subtest scores. The results are presented in Table 4.17 and Table 4.18.

<table>
<thead>
<tr>
<th>L1</th>
<th>Posttest Score</th>
<th>Observed Mean</th>
<th>Adjusted Mean</th>
<th>SD</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td></td>
<td>73.63</td>
<td>74.14</td>
<td>16.47</td>
<td>8</td>
</tr>
<tr>
<td>M-T</td>
<td></td>
<td>64.47</td>
<td>64.34</td>
<td>13.26</td>
<td>32</td>
</tr>
<tr>
<td>Source</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Posttest</td>
<td></td>
<td>2010.33</td>
<td>1</td>
<td>2010.33</td>
<td>13.93*</td>
</tr>
<tr>
<td>L1</td>
<td></td>
<td>613.62</td>
<td>1</td>
<td>613.62</td>
<td>4.25*</td>
</tr>
<tr>
<td>Error</td>
<td></td>
<td>5339.51</td>
<td>37</td>
<td>144.31</td>
<td></td>
</tr>
</tbody>
</table>

Note: R² = .323, Adj. R² = .286, adjustments based on pretest scores mean of 52.05.
M = Mandarin; M-T = Mandarin-Taiwanese
* p < .05
Table 4.18
Pairwise Comparison and Mean Difference in /i/-/ɪ/ Contrast by L1 and Controlling for Pretest Scores

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Mean Difference</th>
<th>s.e.</th>
<th>Bonferroni Adjusted 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>M vs. M-T</td>
<td>9.80*</td>
<td>4.75</td>
<td>.170, 19.43</td>
</tr>
</tbody>
</table>

* p < .05, where p-values are adjusted using the Bonferroni method.

Note. Comparisons based upon ANCOVA adjusted means controlling for pretest scores mean of 52.05. M = Mandarin; M-T = Mandarin-Taiwanese

One-way ANCOVA results showed that there were was a statistically significant difference between the two L1 groups at posttest (Table 4.17). Pairwise comparisons revealed a significant difference between the Mandarin group and Mandarin-Taiwanese group (Table 4.18). Both the observed and adjusted means showed that native speakers of Mandarin performed significantly better than native speakers of Mandarin-Taiwanese. Therefore, the null hypothesis predicting that there would be no significant differences among the performance of trained participants with different L1s at posttest is rejected for the /i/-/ɪ/ contrast. The results showed that in the trained groups, native speakers of Mandarin performed significantly better than native speakers of Mandarin-Taiwanese on identifying the /i/-/ɪ/ contrast at posttest.

4.7.2. /ɛ/-/æ/ Contrast

A one-way analysis of covariance (ANCOVA) was conducted at a significance level of .05. The independent variable, L1, includes two levels: Mandarin and Mandarin-Taiwanese. The dependent variable is the participants’ posttest /ɛ/-/æ/ subtest scores and the covariate is the participants’ pretest /ɛ/-/æ/ subtest scores. The results are presented in Table 4.19 and Table 4.20.
Table 4.19

ANOVA Results and Descriptive Statistics for /ɛ/-/æ/ Contrast by L1 and Posttest Scores

<table>
<thead>
<tr>
<th>L1</th>
<th>Posttest Score</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Observed Mean</td>
<td>Adjusted Mean</td>
<td>SD</td>
<td>n</td>
</tr>
<tr>
<td>M</td>
<td>71.63</td>
<td>74.07</td>
<td>16.41</td>
<td>8</td>
</tr>
<tr>
<td>M-T</td>
<td>70.00</td>
<td>69.39</td>
<td>11.83</td>
<td>32</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Posttest</td>
<td>3057.64</td>
<td>1</td>
<td>3057.64</td>
<td>35.75*</td>
</tr>
<tr>
<td>L1</td>
<td>137.40</td>
<td>1</td>
<td>137.40</td>
<td>1.61</td>
</tr>
<tr>
<td>Error</td>
<td>3164.24</td>
<td>37</td>
<td>85.52</td>
<td></td>
</tr>
</tbody>
</table>

Note. R² = .493, Adj. R² = .465, adjustments based on pretest scores mean of 62.90. Homogeneity of regression tested and not significant: F = .122, p = .729>.05. M = Mandarin; M-T = Mandarin-Taiwanese

* p < .05

Table 4.20

Pairwise Comparison and Mean Difference in /ɛ/-/æ/ Contrast by L1 and Controlling for Pretest Scores

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Mean Difference</th>
<th>s.e.</th>
<th>Bonferroni Adjusted 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>M vs. M-T</td>
<td>4.678</td>
<td>3.69</td>
<td>-2.80, 12.16</td>
</tr>
</tbody>
</table>

Note. Comparisons based upon ANCOVA adjusted means controlling for pretest scores mean of 62.90. M = Mandarin; M-T = Mandarin-Taiwanese

One-way ANCOVA results showed that there were no statistically significant difference between the two L1 groups at posttest (Table 4.19). Pairwise comparisons revealed no significant differences between the Mandarin group and the Mandarin-Taiwanese group (Table 4.20). Both the observed and adjusted means showed that native speakers of Mandarin performed better than native speakers of Mandarin-Taiwanese but the difference was not significant. Thus, the null hypothesis predicting that there would be no significant differences among the performance of
trained participants with different L1s at posttest is accepted for the /e/-/æ/ contrast.

4.7.3. **Word-Final /t/-/d/ Contrast**

A one-way analysis of covariance (ANCOVA) was conducted at a significance level of .05. The independent variable, L1, includes two levels: Mandarin and Mandarin-Taiwanese. The dependent variable is the participants’ posttest word-final /t/-/d/ subtest scores and the covariate is the participants’ pretest word-final /t/-/d/ subtest scores. The results are presented in Table 4.21 and Table 4.22.

Table 4.21
**ANCOVA Results and Descriptive Statistics for Word-Final /t/-/d/ Contrast by L1 and Posttest Scores**

<table>
<thead>
<tr>
<th>L1</th>
<th>Observed Mean</th>
<th>Adjusted Mean</th>
<th>SD</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>73.75</td>
<td>72.99</td>
<td>15.03</td>
<td>8</td>
</tr>
<tr>
<td>M-T</td>
<td>68.75</td>
<td>68.94</td>
<td>13.76</td>
<td>32</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Posttest</td>
<td>1784.85</td>
<td>1</td>
<td>1784.85</td>
<td>11.66*</td>
</tr>
<tr>
<td>L1</td>
<td>104.91</td>
<td>1</td>
<td>104.91</td>
<td>.69</td>
</tr>
<tr>
<td>Error</td>
<td>5664.66</td>
<td>37</td>
<td>153.10</td>
<td></td>
</tr>
</tbody>
</table>

*Note. R² = .256, Adj. R² = .215, adjustments based on pretest scores mean of 58.20. Homogeneity of regression tested and not significant: F = .205, p = .654>.05. M = Mandarin; M-T = Mandarin-Taiwanese

* p < .05

| Table 4.22
Pairwise Comparison and Mean Difference in Word-Final /t/-/d/ Contrast by L1 and Controlling for Pretest Scores |
<table>
<thead>
<tr>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Comparison</td>
</tr>
<tr>
<td>------------</td>
</tr>
<tr>
<td>M vs. M-T</td>
</tr>
</tbody>
</table>

*Note. Comparisons based upon ANCOVA adjusted means controlling for pretest scores mean of 58.20. M = Mandarin; M-T = Mandarin-Taiwanese
One-way ANCOVA results showed that there were no statistically significant difference between the two L1 groups at posttest (Table 4.21). Pairwise comparisons revealed no significant differences between the Mandarin group and the Mandarin-Taiwanese group (Table 4.22). Both the observed and adjusted means showed that native speakers of Mandarin performed better than native speakers of Mandarin-Taiwanese but the difference was not significant. Thus, the null hypothesis predicting that there would be no significant differences among the performance of trained participants with different L1s at posttest is accepted for the word-final /t/-/d/ contrast.

The results of this section showed that both L1 groups made progress at posttest and the native speakers of Mandarin performed significantly better than the native speakers of Mandarin-Taiwanese on identifying the /i/-/ɪ/ contrast at posttest, while the difference between the two L1 groups was not significant in identifying the /ɛ/-/æ/ and the word-final /t/-/d/ contrasts.

4.8. Research Question 8: Participants’ Language Attitudes toward Accents

Research question 8: What are the language attitudes of the participants toward speaking Standard English and EFL teachers’ accents?

A language background and attitude questionnaire (Appendix A) was completed by the original sample (N = 88) before the pretest to understand their language backgrounds and language attitudes. This section reports the results of the survey.
In response to the question, “Would you like to speak English like a native speaker if you could? Why or why not?” an overwhelming 97% of the participants (n = 85) would like to speak English like a native speaker. Their typical arguments include the following:

- Speaking English like a native speaker is cool.
- It’s easier to get a good job.
- Native speakers can understand me more easily.
- An English major’s ultimate goal is to speak English like a native speaker.

With respect to the remaining three participants who wouldn’t like to have a native English accent. One of them did not give reasons and the other two arguments are as follows:

- Taiwanese should retain their Taiwanese accent as a marker of ethnic identity.
- As long as I can express myself in English, I don’t need to have a native accent.

In response to the question, “Do you think it is important for EFL teachers to have a native-like accent? Why or why not?” the majority of the participants (75%) agreed that EFL teachers should have a native-like accent. Their typical arguments are as follows:

- I don’t want to pick up my teacher’s Taiwanese accent.
- A language teacher should be a role model for language learning.
- Teachers’ native-like accents help my pronunciation and listening comprehension.
- Teachers’ mispronunciation hinders understanding in class.
- I can understand Taiwanese accented English but I can’t understand American accented English.

With regard to the 25% of the participants (n = 22) who did not agree that EFL
teachers should have a native-like accent, their typical arguments are as follows:

- Taiwanese should retain their Taiwanese accent as a marker of ethnic identity.
- A teacher with a Taiwanese accent sounds approachable.
- A native accent can be difficult to understand.
- A native accent is not necessary as long as teachers can pronounce words correctly.
- As long as students can understand the teacher, a native accent is not necessary.

The results of the participants’ language attitudes toward accents showed that almost 100% of the participants would like to have a native-like accent, while 75% of them thought that it is imperative for EFL teachers to have a native-like accent. On the other hand, a quarter of the participants did not think that it is necessary for EFL teachers to have a native-like accent.

An independent-samples t-test was conducted to compare pretest mean scores of participants with different language attitudes regarding teachers’ accents. The results show that there was not a significant difference between the scores for participants who agreed that EFL teachers should have a native-like accent (M=57.68, SD=7.33) and for participants who did not agree that EFL teachers should have a native-like accent (M=59.92, SD=7.33); t(86)=-1.24, p =.993, although the means of the latter is higher than the former.

4.9. Participants’ Difficulties in Perception and Production

In this section, the mean percentage correct identification scores by contrast and the mean production scores by token at pretest and posttest were analyzed to evaluate the
relative difficulty of the three contrasts for the participants to identify as well as the relative difficulty of the tokens for the participants to produce.

4.9.1. Perception Difficulties

The mean percentage correct identification scores at pretest and posttest for the three contrasts are presented in Figure 4.7.

![Figure 4.7 Mean Percentage Correct Identification Scores for the 3 Contrasts on the Pretest and Posttest by All Participants (N = 71)](image)

A series of one-way ANOVA using a significance level of .05 were conducted to evaluate which contrast posed the most difficulty and which contrast posed the least difficulty for the participants. The results of the pretest indicated a significant difference among the means: [F(2,210) = 19.073, p = .000]. Post hoc analyses using Bonferroni at a significance level of .05 showed significant differences between /i/-/ɪ/ and /ɛ/-/æ/ (p = .000), and between /i/-/ɪ/ and the word-final /t/-/d/ (p = .000). However, the mean
difference between /ɛ/-/æ/ and the word-final /t/-/d/ was not statistically significant (p = .247). The /i/-/ɪ/ contrast appeared to be the most difficult, followed by the word-final /t/-/d/ contrast and the /ɛ/-/æ/ contrast.

With respect to the posttest, the results indicated a significant difference among the means of the three contrasts: [F(2,210) = 3.499, p = .032]. Post hoc analyses using Bonferroni at a significance level of .05 indicated a significant difference between /i/-/ɪ/ and /ɛ/-/æ/ (p = .027), but not between /i/-/ɪ/ and the word-final /t/-/d/ (p = .373) or between /ɛ/-/æ/ and the word-final /t/-/d/ (p = .832). The /i/-/ɪ/ contrast was still the most difficult for the participants, followed by the word-final /t/-/d/ contrast and the /ɛ/-/æ/ contrast.

The results showed that the /i/-/ɪ/ contrast posed the most difficulty for the participants to identify throughout the study, followed by the word-final /t/-/d/ contrast and the /ɛ/-/æ/ contrast.

4.9.2. Production Difficulties

The pretest intelligibility test results by token are presented in Figure 4.8 to display which token posed the most difficulty and which token posed the least difficulty for the participants.
Figure 4.8 Mean Production Intelligibility Scores on the Pretest by All Participants ($N = 71$)

Figure 4.8 indicates that the highest mean score was observed in *need* and the lowest mean score was observed in *dip*, while its contrast, *deep*, was relatively easy for the participants. Likewise, *beat* had a higher mean score than *bit*. However, *ship* had a higher mean score than *sheep*, and *peek* and *pick* had about the same mean score. As for the */ɛ/-*/æ/* contrast, *pen*, *guess*, and *bed* had higher mean scores than their contrasts, *pan*, *gas*, and *bad*. With respect to the word-final */t/-*/d/* contrast, *need* had a higher observed mean score than *neat*, but *bat* had a higher observed mean score than *bad*, while *bed* and *bet* had a slight mean difference.

On the whole, */i/* is relatively easier than */t/* and */ɛ/* is relatively easier than */æ/* for the participants to produce. As for the word-final */t/-*/d/* contrast, there seemed no
difference.

The posttest intelligibility test results by token are also presented in Figure 4.9 to display which token posed the most difficulty and which token posed the least difficulty for the participants.

*Figure 4.9* Mean Production Intelligibility Scores on the Posttest by All Participants ($N = 71$)

Figure 4.9 indicates that the highest mean score was still observed in *need* and the lowest mean score was still observed in *dip*, while its contrast, *deep*, was relatively easy for the participants. As in the pretest, *beat* had a higher mean score than *bit*, and *ship* still had a higher mean score than *sheep*. *Peek* and *pick* still had about the same mean score.
As for the /ɛ/-/æ/ contrast, *pen, guess,* and *bed* still had higher mean scores than their contrasts, *pan, gas,* and *bad.* With respect to the word-final /t/-/d/ contrast, *need* had a higher observed mean score than *neat,* but *bat* had a higher observed mean score than *bad,* while *bed* and *bet* had a slight mean difference.

On the whole, like the pretest, /i/ is still relatively easier than /ɪ/ and /ɛ/ is relatively easier than /æ/ for the participants to pronounce. As for the word-final /t/-/d/ contrast, there was no difference. All in all, the trainee’s overall mean intelligibility scores were below par, with none of the token mean scores above 5 at posttest.

### 4.10. Chapter Summary

From the pretest to the posttest, the ID group showed a 17% increase on the /i/-/ɪ/ contrast, an 8% increase on the /ɛ/-/æ/ contrast, and a 12% increase on the word-final /t/-/d/ contrast; the SD group showed a 12% increase on the /i/-/ɪ/ contrast, a 6% increase on the /ɛ/-/æ/ contrast, and a 9% increase on the word-final /t/-/d/ contrast. With respect to the control group, the results showed slight decreases on all three of the contrasts, with a 2% decrease on the /i/-/ɪ/ contrast, a 1% decrease on the /ɛ/-/æ/ contrast, and a 5% decrease on the word-final /t/-/d/ contrast. The results showed a significant positive training effect for both training methods on identifying the /i/-/ɪ/ contrast, while for the /ɛ/-/æ/ contrast, only the ID training showed a significant training effect. With respect to the word-final /t/-/d/ contrast, although both trained groups showed gains, the training effects of the two training methods were not significant.

Although the results showed that the ID group outperformed the SD group in identifying all the three contrasts (4% higher on the /i/-/ɪ/ contrast, 2% higher on the
The mean difference between the two trained groups was not statistically significant. The large effect sizes of both trained groups indicate that it is worthwhile to receive either of the training methods.

With regard to generalization, the results showed that the effect of perceptual training on identifying the three non-native phonemic contrasts was generalized to: (1) new tokens produced by familiar talkers, except for the SD group in the word-final /t/-/d/ contrast, (2) familiar tokens produced by new talkers, except for the SD group in the /ɛ/-/æ/ contrast, and (3) the /k/-/g/ endings for the ID group only but not to the /p/-/b/ endings for both trained groups.

In terms of retention, for the ID group, the training effect appeared retained for the /i/-/ɪ/ contrast and the word-final /t/-/d/ contrast three months after the completion of the training. With respect to generalization to new tokens, the training effect appeared retained for the /ɛ/-/æ/ contrast and the word-final /t/-/d/ contrast. With regard to generalization to new talkers, the training effect appeared retained for all of the three contrasts. As for generalization to other final stop contrasts, the training effect appeared retained for the /k/-/ɡ/ endings.

For the SD group, the training effect appeared retained for the /i/-/ɪ/ contrast and the /ɛ/-/æ/ contrast three months after the completion of the training. With respect to generalization to new tokens, the training effect appeared retained for the /i/-/ɪ/ and the /ɛ/-/æ/ contrasts. With regard to generalization to new talkers, the training effect appeared retained for the /i/-/ɪ/ and the word-final /t/-/d/ contrasts.

With respect to the effect of perception training on production, both the ID group and the control group, with no perceptual training, had significantly higher production
scores than the SD group and the mean difference between the ID group and the control group was not statistically significant. Note that the SD group, in spite of training, showed a 3% decrease from the pretest to the posttest, indicating that the perceptual training did not facilitate production.

As for the perceptual performance of the trainees based on their L1 grounds, native speakers of Mandarin outperformed native speakers of Mandarin-Taiwanese on all three contrasts at posttest, while the mean difference was significant only in the /i/-/ɪ/ contrast, but not in the /ɛ/-/æ/ and the word-final /t/-/d/ contrasts.

A survey of the participants’ language attitudes showed that an overwhelming 97% of the participants would like to speak English like a native speaker and the majority of the participants (75%) agreed that EFL teachers should have a native-like accent, while a quarter of the participants did not agree that EFL teachers should have a native-like accent.

All in all, the /i/-/ɪ/ contrast posed the most difficulty for the participants to identify throughout the study, followed by the word-final /t/-/d/ contrast and the /ɛ/-/æ/ contrast. However, both the trained groups showed the least increase on the /ɛ/-/æ/ contrast from pretest to posttest (8% for the ID group and 6% for the SD group). When it came to production, /i/ is relatively easier than /ɪ/ and /ɛ/ is relatively easier than /æ/ for the participants to produce. However, ship had a higher mean score than sheep. As for the word-final /t/-/d/ contrast, there appeared to be no difference.

In sum, the ID group and the SD group showed gains on the three contrasts while the control group showed slight decreases on all of the three contrasts from the pretest to the posttest. Although the ID group outperformed the SD group in identifying all the
three contrasts, the mean difference between the two trained groups was not statistically significant. The effect of the ID training was generalized to seven out of the eight contrasts in the generalization test, while the effect of the SD training was generalized to only half of the eight contrasts in the generalization test. For both trained groups, most of the training effects at posttest appeared retained three months after the completion of the training. With the relationship of perception and productions, the perceptual training did not facilitate production of the trained contrasts in the present study. Although the trained native speakers of Mandarin outperformed the trained native speakers of Mandarin-Taiwanese on all three contrasts at posttest, the difference was significant only in the /i/-/ɨ/ contrast. The /i/-/ɨ/ contrast posed the most difficulty for the participants to identify, and /ɨ/ appeared to be the most difficult for the participants to produce. An overwhelming 97% of the participants would like to have a native-like accent and the majority of the participants (75%) agreed that EFL teachers should have a native-like accent, while a quarter of the participants did not agree that EFL teachers should have a native-like accent.

Discussion of major findings in the present study and pedagogical implications are presented in the next chapter.
CHAPTER 5: DISCUSSION AND CONCLUSION

In this chapter, major findings in the present study are discussed in relation to the research questions and previous studies. Next, pedagogical implications are provided for ESL/EFL teachers, curriculum designers, and policy makers. Finally, limitations of the present study are discussed and directions for future research are proposed.

5.1. Discussion of Major Findings

5.1.1. Training Effects and Comparison of Two Training Methods

From the pretest to the posttest, the ID group and the SD group showed gains on the three contrasts while the control group showed slight decreases on all of the three contrasts. The results showed a significant positive training effect for both training methods on identifying the /i/-/ɪ/ contrast, while for the /ɛ/-/æ/ contrast, only the ID training showed a significant training effect. With respect to the word-final /t/-/d/ contrast, although both trained groups showed gains, the training effects of the two training methods were not significant.

The effectiveness of both training methods on the /i/-/ɪ/ contrast and that of the ID training on the /ɛ/-/æ/ contrast support previous empirical studies on the effectiveness of perceptual training on identifying non-native phonemic contrasts (e.g., Flege & Wang, 1989; Lively et al., 1993,1994; Flege,1995a; Bradlow et al.,1999; Wang, 2002; Lee, 2009).
For the /ɛ/-/æ/ contrast, the finding that the ID group showed an 8% increase and showed a significant training effect while the SD group showed a 6% increase but showed no significant training effect indicates that the SD group may have been on the verge of showing a significant training effect, in view of the finding that the control group showed a 1% decrease on the contrast.

For the word-final /t/-/d/ contrast, the finding that the ID group showed a 12% increase and the SD group showed a 9% increase while the control group showed a 5% decrease suggests that both training methods were effective although there was no statistically significant difference. Furthermore, the finding that only the ID group showed a significant training effect in the /ɛ/-/æ/ contrast suggests that participants in the ID group may have benefited from learning acoustic cues of the non-native phonemic contrast in the training.

With respect to which training methods is more effective, although the ID group outperformed the SD group in identifying all the three contrasts, the mean difference between the two trained groups was not statistically significant. The finding is consistent with the findings of previous studies comparing these two methods (Flege, 1995a; Wayland & Li, 2005, 2008). However, after the follow-up posttest, the researcher announced that the ID group had learned acoustic cues of the contrasts in the training and some participants in the SD group expressed great interest in learning the acoustic cues. In addition, many of the participants in the ID group mentioned that they had never learned the acoustic differences in the contrasts before and have benefited from the cues. The reaction of the participants is consistent with the finding in Flege (1995a).

Moreover, there are three participants in the ID group and two participants in the SD
group who had a correct identification score of 90% or higher on the /ɪ/-/i/ contrast, two participants in the SD group who obtained a score of 90% or higher on the /ɛ/-/æ/ contrast, and two participants in the ID group and three participants in the SD group who scored 90% or higher on the word-final /t/-/d/ contrast. The finding not only supports the argument that late learners can still establish new phonetic categories in an L2, as found in previous studies (e.g., Flege et al., 1996; Birdsong & Molis, 2001; Flege & Mackay, 2004) but also shows counterevidence of the Critical Period Hypothesis. In addition, the high correct percentage of identification of some participants in the SD group provides evidence that these participants may have maintained perceptual constancy to successfully identify the contrasts and have established distinctive categories for the contrasts, as suggested by Polka (1992).

The overall findings in this section indicate that natural tokens with talker and duration variability maximized were effective for perceptual training. Moreover, the findings suggest that with effective training, EFL university students in Taiwan may learn to perceive segments that do not exist in Chinese. The large effect sizes of both trained groups indicate that it is worthwhile to receive either of the training methods.

5.1.2. Generalization to New Tokens, New Talkers, and Other Final Stops

The results showed that the effect of perceptual training on identifying the three non-native phonemic contrasts was generalized to: (1) new tokens produced by familiar talkers, except for the SD group for the word-final /t/-/d/ contrast, (2) familiar tokens produced by new talkers, except for the SD group for the /ɛ/-/æ/ contrast, and (3) the /k/-/g/ endings for the ID group only but not to the /p/-/b/ endings for both trained groups.

The results showed that the effect of the ID training was generalized to seven out of
the eight contrasts in the generalization test, including new tokens and new talkers, with the exception of the word-final /p/-/b/ contrast, while the effect of the SD training was generalized to only half of the eight contrasts in the generalization test. The finding suggests that high talker variability may play a crucial role in establishment of new phonemic categories as well as elimination of talker effects, as found in previous studies (e.g., Logan et al., 1991; Lively et al., 1993, 1994; Flege, 1995a; Wang, 2002; Lee, 2009).

It is noteworthy that although the ID training effect was not significant on the word-final /t/-/d/ contrast in the posttest, it was significant on the word-final /t/-/d/ contrast in the generalization test, indicating once more that that participants in the ID group may have benefited from learning acoustic cues of the non-native phonemic contrast. That is all the more reason why most researchers are in favor of the ID training rather than the SD training (e.g., Jamieson & Morosan, 1986, 1989; Logan et al., 1991; Lively et al., 1994; Flege, 1995a).

As for the finding that the training effect was not generalized to the /p/-/b/ endings for both trained groups, it is possible that most of the six tokens (nap, nab, pup, pub, rip, and rib) used in the generalization test are unfamiliar to the participants, as opposed to the relatively easier tokens used in the /k/-/g/ endings (back, bag, buck, bug, lock, log, pick, pig, tack, and tag).

5.1.3. Retention of Training Effects

For the ID group, most of the training effects appeared retained but the training effect on the /ɛ/-/æ/ contrast in the posttest and the training effect generalized to new tokens on the /i/-/i/ contrast appeared not retained three months after the completion of
the training. The fading training effect on the /ɛ-/æ/ contrast may be explained by the spectral overlapping of /ɛ/ and /æ/ (Hillenbrand & Clark, 2000), which will be discussed fully in Section 5.1.6. As for the loss of the training effect generalized to new tokens on the /i-/ɪ/ contrast, it is possible that the new tokens were less familiar to the participants as compared to familiar tokens used for the training. Thus, a “token effect” might have been observed in this case. As for the SD group, most of the training effects at posttest appeared also retained three months after the completion of the training.

The finding that most of the training effects at posttest appeared retained three months after the completion of the training is consistent with previous findings in that once a new phonemic category is established, it can be retained for a long period of time (e.g., Lively et al., 1994; Flege, 1995a; Bradlow et al., 1999; Wang, 2002).

The present study provides evidence that participants in the ID group may have benefited from learning acoustic cues of the non-native phonemic contrasts in the training. On the other hand, the high correct percentage of identification of some participants in the SD group also shows evidence that the participants may have maintained perceptual constancy to successfully identify the contrasts and have established distinctive categories for the contrasts. In other words, the ID training may have sensitized the trainee to pay attention to the core properties that define the L2 categories and the SD training may have enabled the trainee to maintain perceptual constancy to successfully identify multiple variants of each category. Both training may have contributed to the establishment of distinctive categories in L2.

Compared with previous studies that involved perceptual training (e.g., Lively et al., 1994; Flege, 1995a; Bradlow et al., 1999; Wang, 2002), the perceptual training in the
present study is relatively less effective, which might be explained by the fact that the participants in most of the previous studies are ESL students while the participants in the present study are EFL students. ESL students would feel the need more than EFL students to have training in perception and production to prevent a communication breakdown in their daily life (e.g., Derwing, 2008; Saito, 2007). Another explanation may be that some of the participants in the present study are low-motivated students, which is shown by the high attrition rate and that some of the trainees did not make much progress throughout the training. The participants may have been interested in the spy game scenario because it was more game-like than mere training; however, the novelty seemed to wear off over time for some trainees did not make progress till the end of training.

All in all, the overall pattern of results suggests that either of the training methods can be effective in training ESL/EFL learners to identify non-native phonemic contrasts although the ID training appeared more effective. The results also indicate that natural tokens with talker and duration variability maximized are sufficient for effective perceptual training. Moreover, the success of individual computer training shows a promising future for the application of a computer-assisted program for perceptual training in identifying L2 segments. It can also be part of an English remedial program to allow low-achieving students to learn at their own pace, which may be helpful, especially in a Tier 4 university.

5.1.4. Relationship between Perceptual Training and Production

The finding that the control group had significantly higher production scores than the SD group and performed almost as well as the ID group at posttest suggests that the percentage increase in the control group and the SD group from the pretest to the posttest
may have been due to maturation, i.e., the English education they received besides the
training during the period of the study may have contributed to the gains in production
scores. It also suggests that the perceptual training did not facilitate production of the
trained tokens, which is consistent with the findings in Wang (2002) and Lee (2009) but
inconsistent with the findings in Rochet (1995), Bradlow et al. (1997), and Bradlow et al.
(1999). The results suggest that perceptual training does not facilitate production of
trained contrasts and that additional production training is necessary.

5.1.5. Performance of Participants of Different L1 Backgrounds

Although the trained native speakers of Mandarin outperformed the trained native
speakers of Mandarin-Taiwanese on all three contrasts at posttest, the difference was
significant only in the /i/-/ɪ/ contrast, but not in /ɛ/-/æ/ and the word-final /t/-/d/ contrasts.

Theoretically, Mandarin-Taiwanese bilinguals should be able to identify more
phonemes than monolinguals of Mandarin. In this case, there is no word-final /t/ or /d/ in
Mandarin while there is word-final /t/ in Taiwanese. However, unlike the finding in Flege
(1989), the Mandarin-Taiwanese bilinguals in the present study did not perform better
than the monolingual Mandarin speakers in identifying the word-final /t/-/d/ contrast. As
a matter of fact, the result suggests the opposite, which disconfirms the hypothesis in
Flege (1989) and Flege and Wang (1989) that native-language phonotactic constraints
influence how people with different L1s processed syllables. The findings of the present
study show counterevidence to previous studies (e.g., Flege, 1989; Flege & Wang, 1989;
Flege et al, 1992; Flege & Liu, 2001) in which speakers of languages with word-final /t/
outperformed Mandarin speakers in identifying the word-final /t/-/d/ contrast.

The better performance of the native speakers of Mandarin may be explained by the
findings that out of the 32 Mandarin-Taiwanese bilinguals in the trained groups, 19 (58%) of them used Mandarin as main home language, that the majority (79%) used Mandarin as main language in school, and that 19 (58%) of them claimed that they could speak standard Taiwan Mandarin. The overwhelming domination of Mandarin may have resulted in language loss of the participants’ Taiwanese. In other words, it is highly likely that their Taiwanese proficiency is far below their Mandarin proficiency and thus signs of L1 transfer from Taiwanese may not have shown.

Another possible explanation of the better performance of the native speakers of Mandarin might have been due the high prestige of Mandarin in Taiwan (Bayley, 1996). A higher prestige of a language is often associated with a higher socioeconomic status, as found in Labov’s (1966) study of r-fullness in three department stores in New York City. In other words, the native speakers of Mandarin in the present study may have a higher socioeconomic status than their Mandarin-Taiwanese counterparts and their parents might have higher expectations for them, which might have motivated them to train harder and perform better.

A caveat is that due to the small sample size (8 native speakers of Mandarin and 32 native speakers of Mandarin-Taiwanese), the results may be biased.

Furthermore, the finding that most of the participants in the present study are Mandarin-Taiwanese bilinguals (69%) and only a small proportion are native speakers of Mandarin (12.7%) contradicts with findings in previous studies that all Taiwanese participants are native speakers of Mandarin (e.g., Decamp, 1972; Flege, Munro, & Skelton, 1992; Paolillo, 1995; Rau et al., 2009). Since speakers from different language backgrounds may have different L1 transfer (e.g., Flege, 1989; Flege & Wang, 1989;
Flege et al, 1992; Flege & Liu, 2001), future research should verify participants’ L1 backgrounds and take the possibility of different L1 phonotactic constraints into account when Taiwanese participants are involved.

Finally, the proportion of main home language of Mandarin-Hakka (1.4%, \( n = 1 \)) and Hakka (1.4%, \( n = 1 \)) showed serious language attrition of Hakka in Taiwan, especially the largest group of Hakka speakers in Taiwan inhabits the county where the research site is located. The finding indicates that the Hakka language is dwindling and that the Hakka people are facing language loss.

5.1.6. Difficulties in Perception and Production

5.1.6.1. Difficulties in Perception

In terms of perception, the lowest mean score was observed in the \(/i/-/ɪ/\) contrast at both pretest and posttest. It appears that the manipulation of durations of \(/i/-/ɪ/\) contrast tokens increased difficulty for the participants to identify the contrast. It is possible that the participants may have relied on temporal cues that they had used to rely on in identifying the \(/i/-/ɪ/\) contrast as suggested in previous studies (e.g., Bohn, 1995; Wang, 1997, 2006; Lee, 2009). In other words, many of them may have identified the shorter \(/i/\) as \(/ɪ/\) and longer \(/ɪ/\) as \(/i/\).

However, after the training, the most percentage increase for both trained groups was observed in the \(/i/-/ɪ/\) contrast, with a 17% increase for the ID group and a 12% increase for the SD group and both groups performed significantly better than the control group, indicating the maximized duration variability of the \(/i/-/ɪ/\) contrast has successfully sensitized the trained participants to spectral differences in the \(/i/-/ɪ/\) contrast instead of temporal cues they had used to rely on in identifying the structurally similar English tense.
and lax vowel pair.

Previous studies have shown that the /i/-/ɪ/ contrast is minimally affected by duration to native speakers of English (Bohn, 1995; Wang, 1997; Wang & Munro, 1999; Hillenbrand & Clark, 2000). Crystal and House (1988) found that although /i/ averages about 41% longer than /ɪ/, shortening /i/ or lengthening /ɪ/ show very little effect on vowel identity. The present study shows evidence that once some of the trainees were sensitized to spectral differences in the /i/-/ɪ/ contrast, they began to rely on spectral cue as native speakers do.

The finding that the participants’ assimilation of /ɪ/, a non-native segment, into /i/, which exists in Mandarin, Taiwanese, and Hakka, lends support to not only the Speech Learning Model (Flege, 1995b) but also Best’s (1994) Perceptual Assimilation Model as well as Kuhl’s (1993) Native Language Magnet theory.

With respect to the /ɛ/-/æ/ contrast, although the highest mean score was observed in the /ɛ/-/æ/ contrast at both pretest and posttest, the least percentage increase for both trained groups was also observed in the /ɛ/-/æ/ contrast, indicating that both trained groups made the least progress on identification of the /ɛ/-/æ/ contrast. Therefore, the /ɛ/-/æ/ contrast proved to be the most resistant to differentiate.

The evident difficulty in identifying the /ɛ/-/æ/ contrast may be due to a high degree of spectral overlap of the two vowels. Crystal and House (1988) found that although /æ/ averages about 18% longer than /ɛ/, while /i/ averages about 41% longer than /ɪ/, it is the /ɛ/-/æ/ contrast that is most affected by duration. Hillenbrand and Clark (2000) also noted that /ɛ/ and /æ/ have a high degree of spectral overlap, so native speakers predominantly rely on duration to differentiate the contrast. The finding that there was a significant ID
training effect indicates that participants in the ID group may have benefited from learning acoustic cues of the /ɛ/-/æ/ contrast from the training.

In Wang (2002), both the trained and control groups identified the /i/-/ɪ/ contrast more accurately than the /ɛ/-/æ/ contrast although the trainees spent more time on the /ɛ/-/æ/ contrast. The findings of the present study show just the opposite. It is possible that the manipulation of durations of /i/-/ɪ/ contrast tokens may have increased difficulty for the participants to identify the contrast and resulted in lower correct identification scores.

The effect of stop voicing on duration of preceding vowels is larger in English than in other languages in the world (Flege, 1993). Thus, native English speakers rely on such acoustic cues in identifying final stop contrasts. Although both trained groups showed gains on the word-final /t/-/d/ contrast, the training effects of both training methods were not significant. Also, the finding that the training effects were not generalized to the word-final /p/-/b/ contrast suggests that it is difficult for Chinese learners to perceive different durations of the same vowel preceding final stops. I suggest using spectrograms to help students understand different durations of the same vowel preceding final stops in English.

Furthermore, the difficulty in identifying the word-final /t/-/d/ contrast might also be exacerbated by the mixed presentation of the minimal pairs in multiple phonetic environments by multiple talkers. It appears that future training on the word-final /t/-/d/ contrast should focus on single environment (e.g., bVt and bVd, as in Flege, 1995a) to sensitize the trainees to the different vowel durations preceding the word-final /t/-/d/. The finding is inconsistent with the effective high variability perceptual training in Bradlow et
al. (1999) with tokens in various phonetic environments

5.1.6.2. Difficulties in Production

With respect to production, /i/ is relatively easier than /u/ and /e/ is relatively easier than /æ/ for the participants to pronounce. As for the word-final /t/-/d/ contrast, no evidence could indicate which is easier.

Although there is /i/ in Mandarin, Taiwanese, and Hakka and the highest mean production score was observed in need at both pretest and posttest, the most difficult tokens for the participants also include /i/ in sheep (2nd) and deep (5th), suggesting that the participants’ production performance may be related to specific phonetic environments. In other words, the onset /n/ may have facilitated the accurate production of /i/ and the onset /ʃ/ may have inhibited the accurate production of /i/, even though /i/ exists in the participants’ L1s. The finding lends support to the Markedness Differential Hypothesis (Eckman, 1977) in that there is no /ʃ/ followed by /i/ in the participants’ L1s, so it is more difficult to acquire. The finding is also similar to the findings in Rau et al. (2009) in production of the English voiceless interdental fricative /θ/ by Mandarin speakers in that /θ/ in certain phonetic environments were more difficult because they are more marked.

The participants’ performance on the word-final English /t/-/d/ contrast is similar to Flege, Munro, and Skelton (1992) in that even the experienced learners could not produce the English vowel duration difference preceding /t/-/d/ accurately.

Finally, the finding that none of the token mean scores was above five (out of 10 points) at posttest not only reflects the participants’ low English proficiency but also suggests that production training is imperative.
5.1.7. Participants’ Language Attitudes toward Accents

Although the survey was conducted in an EFL setting, the finding that an overwhelming 97% of the participants would like to speak English like a native speaker is consistent with Derwing (2003) in that 95% of the ESL learners would like to have a native accent. In addition, the response that “An English major’s ultimate goal is to speak English like a native speaker” is consistent with Derwing (2003) in that 97% of the ESL learners believed or strongly believed that it was important to have native-like pronunciation. The high percentage (97%) of the EFL students who wanted a native accent in the present study is much higher than the percentage (67%) of the 400 EFL, EIL and ESL students from 14 countries and 180 teachers from 45 countries who preferred to speak English like a native speaker in Timmis (2002). Furthermore, the response that “Speaking English like a native speaker is cool” might indicate that having a native accent might be related to identity construction in English as found in previous studies (e.g., Golombek & Jordan, 2005; Derwing & Rossiter, 2002). Moreover, the response that “It’s easier to get a good job” suggests that English has such symbolic capital (Bourdieu, 1977) in Taiwan. As mentioned in Chapter 1, as there is a shortage of high English proficient professionals in Taiwan (Lin, 2005), high English proficiency is usually equated with a high income and faster promotion. Finally, the response that “Native speakers can understand me more easily” indicates the participants’ worries about a communication breakdown due to foreign accents, which is similar to the response of 100 adult immigrant full-time ESL students in Derwing and Rossiter (2002) in that over a third of them felt that their foreign accents were the primary cause of any
communication breakdowns. It also supports the argument of Derwing (2008) and Saito (2007) that the inability to produce certain segments is likely to lead to a communication breakdown. The finding suggests the participants’ desire for a native-like accent to make them more intelligible to native speakers.

With respect to the remaining three participants who wouldn’t like to have a native English accent, the response that, “Taiwanese should retain their Taiwanese accent as a marker of ethnic identity” and the response that, “As long as I can express myself in English, I don’t need to have a native accent” concur with the argument of some researchers that we should respect L2 learners’ wish to retain their accents as a marker of their desired identities (e.g., Ur, 1996; Jenkins, 1998, 2000; D. Liu, 1999). The first response also shows that a heavy accent might be preferred in terms of social attractiveness, as indicated in McKenzie (2008).

In response to the second question, “Do you think it is important for EFL teachers to have a native-like accent?” the majority of the participants (75%) agreed that EFL teachers should have a native-like accent. Two of the responses, that “I don’t want to pick up my teacher’s Taiwanese accent” and that “A language teacher should be a role model for language learning” indicate that accented EFL teachers usually reinforce their students’ foreign accents, as noted by Flege (2009), suggesting not only that EFL teachers should utilize native speaker input in perceptual and production training but also that non-native ESL/EFL teachers should reduce their foreign accents for the sake of intelligibility and as a role model for their students (e.g., Lado, 1964; Medgyes, 1999; Arva & Medgyes, 2000; Demirezen, 2007). In addition, the response that “Teachers’ native-like accents help my pronunciation and listening comprehension” expressed
students’ needs in terms of test performance, as confessed by both of the Taiwanese EFL teachers of English in Golombek and Jordan (2005) that EFL teachers should be aware that their pedagogical goals had to serve students’ needs. Especially English proficiency has become a prerequisite for applying to schools and applying for jobs in Taiwan. Moreover, the responses that “Teachers’ mispronunciation hinders understanding in class” and that “I can understand Taiwanese accented English but I can’t understand American accented English” suggest that that EFL teachers’ foreign accents not only hinder understanding in class but also impede understanding of the target language in real life.

With regard to the 25% of the participants (n = 22) who did not agree that EFL teachers should have a native-like accent, the response that “Taiwanese should retain their Taiwanese accent as a marker of ethnic identity” concurs with some researchers’ contention that native-like pronunciation or intonation is not necessary in EIL and EFL settings and that we should respect people’s wish to retain their accents as a marker of their desired identities (e.g., Ur, 1996; Jenkins, 1998, 2000; D. Liu, 1999). It also suggests that the participants had higher expectations for themselves than for their EFL teachers. In the present study, the finding that a quarter of the participants did not think EFL teachers should have a native-like accent but 97% of them wanted to have a native-like accent concurs with Golombek and Jordan’s (2005) observation, however with the roles of teacher and student exchanged. Furthermore, the response that, “A native accent can be difficult to understand” suggests that the participants had long been exposed to EFL teachers’ foreign accents, as noted by Flege (2009).

In sum, the findings that an overwhelming 97% of the participants would like to have a native-like accent and 75% of the participants agreed that EFL teachers should
have a native-like accent suggest that non-native EFL teachers should reduce their foreign accents for the sake of intelligibility in class and as a role model for their students. It appears that speaking English like a native speaker should be not only an English major’s ultimate goal but also a non-native EFL teacher’s ultimate goal. In addition, the participants’ desire for a native-like accent reflects their perceived need for successful communication with native speakers. It might also be related to identity construction in an L2. Furthermore, the finding that some participants equated a native-like accent with a good job suggests that English is regarded as symbolic capital in Taiwan. The most interesting finding is that a quarter of the participants thought that EFL teachers should retain their Taiwanese accent as a marker of ethnic identity while 97% of all the participants wanted to have a native-like accent, suggesting that the participants had higher expectations for themselves than for their EFL teachers.

5.2. Pedagogical Implications

The results of the present study provide pedagogical implications for curriculum developers, ESL/EFL teachers, and policy makers.

5.2.1. Integration of English Pronunciation Pedagogy into Teacher Preparation

Curriculum and In-Service Teacher Training

As mentioned earlier, lack of pronunciation pedagogy training leaves most EFL teachers in Taiwan to rely on teacher’s manuals of textbooks, their own experiences, and intuitions. Fortunately, recently more and more teachers’ colleges in Taiwan have incorporated English pronunciation pedagogy in the curriculum for would-be English teachers, although the course is listed as one of the required electives (e.g., National
Changhua University of Education; National Kaohsiung Normal University). As English pronunciation is more and more emphasized in Taiwan, pronunciation pedagogy should be listed as a required course in the EFL teacher preparation curriculum. In addition, English pronunciation pedagogy workshops should be offered for in-service EFL teachers.

5.2.2. Research Informed Pronunciation Instruction

As Derwing and Munro (2005) noted, empirical studies are crucial to inform pronunciation teaching. ESL/EFL teachers should be informed by empirical evidence in pronunciation related research instead of relying merely on their own experiences and intuitions. For instance, they can understand students’ perceptual and pronunciation difficulties with different L1 backgrounds (e.g., Chinese s EFL learners’ assimilation of /u/ to /i/); learn about the importance of authentic input and the quality of input (e.g., Flege, 2009); learn how to utilize high talker variability to train students’ perception of non-native phonemic contrasts (e.g., Lively et al., 1993); and learn how to use explicit explanation, modeling, practice, and real-time feedback to raise students’ awareness of their problems (e.g., Couper, 2006).

In the meantime, more pronunciation research should be encouraged by major journals to help set pedagogical goals and priorities in the classroom. Researchers and teachers should collaborate to provide more effective approaches to pronunciation teaching, as noted by Derwing and Munro (2005).

5.2.3. Policy Making

As mentioned in Chapter 1, pronunciation has been less emphasized in EFL teaching in Taiwan due to the fact that most entrance examinations in Taiwan do not test English
listening comprehension. Policy makers in Taiwan should change the existent policy to foster a positive “test washback” (Shohamy, 2001, p. 101). A case in point is that the Ministry of Education in Israel introduced a new English oral test which was administered to all students graduating from high school. The washback effect was positive because the goal of the emphasis on the teaching of oral language was achieved and the test affected the content and methods of teaching. The same washback effect can happen to English listening. Therefore, policy makers should utilize the power of tests to emphasize the focus on English listening.

As a matter of fact, the Test of English Listening Comprehension was administered for the first time in Taiwan in late 2012 and the test score is scheduled to be officially listed as a requirement for applications to all universities in Taiwan starting 2015 (CEEC, 2012). Once the English listening comprehension has become a requirement, teachers and students alike will begin to emphasize listening and more courses in English listening comprehension may be added. As a result, teachers will teach to the test and students will benefit from it.

Furthermore, the Ministry of Education in Taiwan is planning on lowering the age of English education in elementary schools from Grade 3 to Grade 1 within a decade (Lin, 2011). Right now first graders in Taipei City and Hsinchu City have started learning English and the communicative approach is being used. This appears to support the Critical Period Hypothesis. It is hoped that the early start will improve the nation’s overall English proficiency and raise its international competitiveness.
5.3. Limitations and Future Research Directions

The limitations of the present study are discussed to make recommendations for future research.

First, there was no random selection from the population. All the participants volunteered to participate in the present study. Therefore, the sample cannot represent the whole population. No generalizations should be made beyond the sample. As mentioned in Chapter 3, the research site is a Tier 4 university and many of the participants had low English proficiency and even low motivation. Some of the drop-outs did not even attend school, let alone training sessions. There was a total of 17 drop-outs from pretest to posttest in the study, which comprised 19% of the original sample, and there was a second round of attrition (n = 6) from posttest to the three-month follow-up posttest due to the fact that some participants transferred to another school. However, there was no significant difference between the mean scores of the drop-outs and the remaining sample. It can be inferred that the sample may have higher English proficiency and higher motivation than the rest of the whole population because they were willing to participate in a study that might improve their English listening ability. Nonetheless, future replication of the study should recruit those who are willing to make a commitment to training sessions in order to reduce attrition. As we can see in Cenoz and Lecumberri (1999), the desire to acquire a native accent can be a significant predictor of the improvement in phonetic discrimination. Motivation, or investment (Norton, 1995, 2000), plays an important role in L2 acquisition. Thus, how to make perceptual training more interesting to increase student motivation has become a formidable task.

Second, the sample size was still too small, so the results may be biased. Especially,
the number of native speakers of Taiwanese or Hakka in the trained groups was under five. Even the number of native speakers of Mandarin in the trained groups was only eight. Therefore, no generalizations should be made from the present study in terms of whether perceptual training effect is influenced by L1 background. A further larger scale study should be conducted in order to make generalizations.

Third, during the training, the computer program randomized the order of the trials, making providing the answer and feedback before the next trial impossible. Thus, each trial began with a question number and the trainees had to check the feedback that corresponds to the number of the question in a separate handout after they answered each question during the training. Although the trainees were told not to peek at the answer, it is likely that some of the participants might have peeked at the answer before answering the question. For example, one participant in the ID group had 100% correct identification on the /ɛ/-/æ/ contrast in the training but had only 63% correct identification on the same contrast at posttest. After communication and modifications, now the researcher is able to choose to display trials in order, making providing the answer and feedback before the next trial possible.

Finally, given the respective advantages of the two training methods in the present study, a cross-training method may be beneficial for ESL/EFL students to take advantage of the effectiveness of each training method. That is, students may receive the ID training to learn core acoustic cues of a non-native phonemic contrast spoken by multiple talkers and then receive the SD training to achieve perceptual constancy to successfully identify multiple variants of distinctive categories in L2. Future research may compare the effectiveness of a cross-training method with the ID and SD training methods.
5.4. Conclusion

The present study investigated and compared the effects of two methods on training EFL university students in Taiwan to identify three non-native phonemic contrasts. The results indicate that both the trained groups showed gains on the three contrasts while the control group showed slight decreases on all of the three contrasts from the pretest to the posttest. Although the ID group outperformed the SD group in identifying all the three contrasts, the mean difference between the two trained groups was not statistically significant. The effect of the ID training was generalized to seven out of the eight contrasts in the generalization test, while the effect of the SD training was generalized to only half of the eight contrasts in the generalization test. For both trained groups, most of the training effects at posttest appeared retained three months after the completion of the training. The results suggest that the perceptual training did not facilitate production of the trained contrasts. Moreover, the trained native speakers of Mandarin outperformed the trained native speakers of Mandarin-Taiwanese on all three contrasts at posttest. The duration manipulated /i/-/ɪ/ contrast posed the most difficulty in perception throughout the study while the /ɛ/-/æ/ contrast proved to be the most resistant to differentiate due to a high degree of spectral overlap of the two vowels. With respect to production, the participants’ production performance may be related to specific phonetic environments. The findings that an overwhelming 97% of the participants would like to have a native-like accent and 75% of the participants agreed that EFL teachers should have a native-like accent suggest that non-native EFL teachers should reduce their foreign accents for the sake of intelligibility in class and as a role model for their students. The results of the survey also show that English is regarded as symbolic capital in Taiwan.
On the whole, the results show a consistent trend that the ID training method is superior to the SD training method although the mean difference between the two trained groups was not statistically significant. However, the large effect sizes of both trained groups indicate that either of the training methods is effective for perceptual training for ESL/EFL students.

Taken together, the findings of the present study provide support for the efficacy of high talker and duration variability in perceptual training. The finding that the trained groups did not improve their oral production also indicates the necessity of production training.

It is hoped that the results of the present study may inform policy-makers and practitioners toward the integration of either of the two training methods into ESL/EFL teacher training programs and English teaching process to help ESL/EFL learners accurately identify non-native English phonemic contrasts.
References


http://www.uccs.edu/~lbecker/


201


Internet-based and Paper-based Tests. Retrieved from


Journal of the Sociology of Language, 143, 139-149.


http://etds.ncl.edu.tw/theabs/service/ft_result.jsp?FT=Y&id=095NKNU5238017


Liu, J. (1999b). From Their Own Perspectives: The Impact of Non-Native ESL Professionals on Their Students. In G. Braine (Ed.), *Non-Native Educators in English Language Teaching* (pp. 159-176). Mahwah, NJ: Lawrence Erlbaum.


Integrating Diversity with Quantitative, Qualitative, and Mixed Methods.


Wang, X. (2002) *Training Mandarin and Cantonese speakers to identify English vowel*
Simon Fraser University.


Appendix A: Language Background and Attitude Questionnaire

1. Name: ____________________
2. Student #: ____________________
3. Gender: ____________________
4. Age: ____________________
5. Is your hearing normal? ____________________
6. Your first language(s) is (are) (may choose more than one):
   □ Mandarin □ Taiwanese □ Hakka
   □ Mandarin-Taiwanese bilingual □ Mandarin-Hakka bilingual
   □ Others: ___________________________________
7. The language(s) your parents speak: (List in the order from high to low frequency of use)
   Mother: __________________
   Father: __________________
8. What is the percentage of the language(s) you use at home? (e.g. 70% Mandarin and 30% Hakka, or 100% Taiwanese)
   ___________________________________
9. What is the percentage of the language(s) you speak with your friends in school?
   ___________________________________
10. The language you prefer to use: __________________
11. Can you speak Standard Mandarin in Taiwan (e.g., pronounce ㄓㄔㄕㄗㄘㄌㄈ without a problem at all)? __________________
12. Do you think that your pronunciation in Mandarin has been influenced by your first language(s) other than Mandarin? __________________
13. At what age did you begin to learn English? __________________
14. Will you receive English instruction in how to discriminate English minimal pairs in other places (e.g., school, cram school)? __________________
15. Would you like to speak English like a native speaker if you could? Why or why not?

_________________________________________________________________

16. Do you think it is important for EFL teachers to have a native-like accent? Why or why not?

_________________________________________________________________
Appendix B: Permission to Use Real-person Pronunciation Sound Files in Cambridge Advanced Learner’s Dictionary
3 March 2011

Yao-Feng Huang
631 Tuscarawas Ct
Columbus, OH 43210

Cambridge University Press Reference
ISBN: 978 0 521 712675
Title: Cambridge Advanced Learner’s Dictionary, Third edition, CD-ROM
Selection: up to 1000 audio pronunciations of entry items (headwords and runons) from
the CD-ROM

Research Use
Content described above may be used in research for a doctoral thesis exploring the
instruction of pronunciation in English, and may be edited and played to research subjects
to elicit their responses.

Dear Yao-Feng Huang:

Thank you for your request. Subject to the conditions below, Cambridge University Press
grants you permission to use the material referenced above.

Conditions of Use
- All materials may be used solely for the research described above.
- The audio files will not be sold, published, distributed or shared with any
  individual or organization without the written permission of Cambridge
  University Press.
- Yao-Feng Huang will provide Cambridge University Press with a copy of the
  research thesis when it is completed.
- Yao-Feng Huang will delete or destroy any copies of the audio files when the
  research is completed.

Sincerely,
Paul Heacock
Publishing Manager

Agreed: 3/15/2011
Yao-Feng Huang

218
Appendix C: Permission to Use Real-person Pronunciation Sound Files in Longman Dictionary of Contemporary English

RE: I’m wondering if I could use the real-person pronunciation on your CD-ROM
Gilbert, Alison [Alison.Gilbert@pearson.com]
Dear Yao-Feng Huang,
Thank you for your enquiry below that has been forwarded to me from the Permissions Department. We are happy to grant you a once only permission, to use the real-person pronunciation on the CD-ROM component of the Dictionary of Contemporary English title, for your research purposes. This permission is granted only on the basis outlined in your email below, that the material is not used for any commercial gain and that you credit the source of the audio pronunciation in your research document.
With best wishes
Alison Gilbert
IP & Versioning Contracts Coordinator
ELT Division
Pearson
Edinburgh Gate
Harlow
CM20 2JE
United Kingdom
D: +00 (44) 1279 623100
Pearson
Always Learning
Learn more at www.pearson.com
Appendix D: Permission to Use Real-person Pronunciation Sound Files in Merriam-Webster 11th Collegiate Dictionary
LIMITED NON-COMMERCIAL RESEARCH LICENSE AGREEMENT

IN CONSIDERATION OF THE MUTUAL PROMISES HEREIN, THE UNDERSIGNED AGREE AS FOLLOWS:

I. LICENSE. Merriam-Webster, Incorporated (“Merriam”) hereby grants to the party named below (“Licensee”) and Licensee hereby accepts a non-exclusive, non-transferable license to use the Database and the Reference Content solely for non-profit, non-commercial research purposes in the Research Project, subject to the following provisions and to the attached Schedules A and B and Terms and Conditions (collectively, the “Agreement”).

II. TERM. The term of this Agreement and the license granted herein is one (1) year from the date set forth below, unless sooner terminated as provided herein. Upon expiration or termination of this Agreement, this Agreement and the license granted herein shall terminate automatically as of such expiration or termination date.

III. LICENSE FEE. In consideration of the License granted herein, Merriam waives the License Fee for the term of this Agreement.

IV. DEFINITIONS.
“Database” means each compilation of Merriam’s proprietary lexical and/or reference materials described in Schedule A attached hereto.

“Reference Content” means a set of entries and any lexical, linguistic or reference information associated with the entries in a Database.

“Research Project” means Licensee’s specific research project described in Schedule B attached.

DATE: August 1, 2011.

MERRIAM-WEBSTEBER,
INCORPORATED
47 Federal Street, P.O. Box 281
Springfield, MA 01102

By: __________________________
James W. Withgott
Vice President & Associate Publisher

LICENSEE:
YAO-FENG HUANG
631 Tuscana Way
Columbus, OH 43219

By: __________________________
Print Name: YAO-FENG.HUANG
Title: OSU doctoral student/Lecturer at Taien University, Taiwan
Appendix E: Word List for NES Talkers

**Instructions:** Please record the following words in a normal speed. Please make sure that every segment, including final consonants, is pronounced in a clear speech.

**Sound File 1: /i/-/ɪ/ contrast**

1. beach, bitch
2. bead, bid
3. beat, bit
4. beaten, bitten
5. cheap, chip
6. deed, did
7. deep, dip
8. each, itch
9. ease, is
10. eat, it
11. deem, dim
12. eel, ill
13. feel, fill
14. feet, fit
15. feast, fist
16. green, grin
17. heap, hip
18. heat, hit
19. heed, hid
20. heel, hill
21. lead, lid
22. leap, lip
23. leave, live (v.)
24. meat, mitt
25. neat, knit
26. peace, piss
27. peach, pitch
28. peek, pick
29. peel, pill
30. read, rid
31. reach, rich
32. seat, sit
33. seek, sick
34. sheep, ship
35. sheet, shit
36. sleep, slip

**Sound File 2: /ɛ/-/æ/ contrast**

1. bet, bat
2. bed, bad
3. beg, bag
4. dead, dad
5. fed, fad
6. guess, gas
7. head, had
8. heck, hack
9. kettle, cattle
10. led, lad
11. leg, lag
12. mess, mass
13. met, mat
14. med, mat
15. net, gnat
16. peck, pack
17. peddle, paddle
18. pen, pan
19. pet, pat
20. rebel, rabble
21. said, sad
22. send, sand
23. set, sat
24. spend, spanned
25. ten, tan

**Sound File 3: Word-Final /t/-/d/ contrast**

1. bat, bad
2. bet, bed
3. beat, bead
4. bit, bid
5. hat, had
6. heat, heed
7. mat, mad
8. neat, need
9. not, nod

10. sat, sad

**Sound File 4: Final /k/-/g/ contrasts**

1. back, bag
2. buck, bug
3. dick, dig
4. hawk, hog
5. knock, nog

6. lock, log
7. pick, pig
8. rack, rag
9. snack, snag
10. tack, tag

**Sound File 5: Final /p/-/b/ contrasts**

1. bop, bob
2. lope, lobe
3. nap, nab
4. nip, nib
5. pup, pub
6. rip, rib

**Sound File 6: Shorter /i/**

**Instructions:** Please record all the following words with much shorter vowel durations as in fast speech.

1. (short) beach
2. (short) bead
3. (short) beat
4. (short) beaten
5. (short) cheap
6. (short) deed
7. (short) deep
8. (short) each
9. (short) ease
10. (short) eat
11. (short) deem
12. (short) eel
13. (short) feel 26. (short) peace
14. (short) feet 27. (short) peach
15. (short) feast 28. (short) peek
16. (short) green 29. (short) peel
17. (short) heap 30. (short) read
18. (short) heat 31. (short) reach
19. (short) heed 32. (short) seat
20. (short) heel 33. (short) seek
21. (short) lead 34. (short) sheep
22. (short) leap 35. (short) sheet
23. (short) leave 36. (short) sleep
24. (short) meat 37. (short) team
25. (short) neat 38. (short) wheel

**Sound File 7: Longer /ʌ/**

**Instructions:** Please record all the following words with much longer vowel durations as in emphatic speech.

1. (long) bitch 7. (long) dip
2. (long) bid 8. (long) itch
3. (long) bit 9. (long) is
4. (long) bitten 10. (long) it
5. (long) chip 11. (long) dim
6. (long) did 12. (long) ill
| 13. (long) fill  | 36. (long) slip |
| 14. (long) fit   | 37. (long) Tim  |
| 15. (long) fist  | 38. (long) will |
| 16. (long) grin  |                |
| 17. (long) hip   |                |
| 18. (long) hit   |                |
| 19. (long) hid   |                |
| 20. (long) hill  |                |
| 21. (long) lid   |                |
| 22. (long) lip   |                |
| 23. (long) live (v.) |          |
| 24. (long) mitt  |                |
| 25. (long) knit  |                |
| 26. (long) piss  |                |
| 27. (long) pitch |                |
| 28. (long) pick  |                |
| 29. (long) pill  |                |
| 30. (long) rid   |                |
| 31. (long) rich  |                |
| 32. (long) sit   |                |
| 33. (long) sick  |                |
| 34. (long) ship  |                |
| 35. (long) shit  |                |
Appendix F: Training Stimuli

<table>
<thead>
<tr>
<th>Sessions</th>
<th>/i/-/ɪ/ contrast</th>
<th>/ɛ/-/æ/ contrast</th>
<th>Word-Final /t/-/d/ contrast</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>beat-bit</td>
<td>bet-bat</td>
<td>edited bet-bed</td>
</tr>
<tr>
<td></td>
<td>deep-dip</td>
<td>dead-dad</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>short beat-long bit</td>
<td>guess-gas</td>
<td>edited bat-bad</td>
</tr>
<tr>
<td></td>
<td>short deep-long dip</td>
<td>head-had</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>neat-knit</td>
<td>met-mat</td>
<td>edited mat- mad</td>
</tr>
<tr>
<td></td>
<td>sheep-ship</td>
<td>med-mad</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>short neat-long knit</td>
<td>led-lad</td>
<td>edited neat- need</td>
</tr>
<tr>
<td></td>
<td>short sheep-long ship</td>
<td>leg-lag</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>heat-hit</td>
<td>heck-hack</td>
<td>edited heat-heed</td>
</tr>
<tr>
<td></td>
<td>peek-pick</td>
<td>kettle-cattle</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>short heat-long hit</td>
<td>fed-fad</td>
<td>edited beat-bead</td>
</tr>
<tr>
<td></td>
<td>short peek-long pick</td>
<td>net-gnat</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>heel-hill</td>
<td>pedal-paddle</td>
<td>edited bit-bid</td>
</tr>
<tr>
<td></td>
<td>wheel-will</td>
<td>rebel-rabble</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>short heel-long hill</td>
<td>mess-mass</td>
<td>edited sat-sad</td>
</tr>
<tr>
<td></td>
<td>short wheel-long will</td>
<td>pen-pan</td>
<td></td>
</tr>
</tbody>
</table>
Appendix G: Pre/Post Tests Stimuli

Part 1 /i/-/ɪ/ Contrast

1. beat (normal) 16. will (long)
2. bit (normal) 17. beat (short)
3. deep (normal) 18. bit (long)
4. dip (normal) 19. deep (short)
5. heat (normal) 20. dip (long)
6. hit (normal) 21. heat (short)
7. heel (normal) 22. hit (long)
8. hill (normal) 23. heel (short)
9. neat (normal) 24. hill (long)
10. knit (normal) 25. neat (short)
11. peek (normal) 26. knit (long)
12. pick (normal) 27. peek (short)
13. sheep (short) 28. pick (long)
14. ship (long) 29. sheep (short)
15. wheel (short) 30. ship (long)

Note: Normal refers to a normal speed. Short refers to a short duration for /i/, while long refers to a long duration for /ɪ/.

Part 2 /ɛ/-/æ/ Contrast

1. bet 6. fad
2. bat 7. guess
3. dead 8. gas
4. dad 9. head
5. fed 10. had
Part 3 Final /t/-/d/ Contrast

1. bat 16. nod
2. bad 17. bat
3. bet 18. bad
4. bed 19. bet
5. beat 20. bed
6. bead 21. beat
7. bit 22. bead
8. bid 23. bit
9. heat 24. bid
10. heed 25. heat
11. mat 26. heed
12. mad 27. mat
13. neat 28. mad
14. need 29. neat
15. not 30. need

Note: All final /t/’s and /d/’s are removed.
Appendix H: Generalization Test Stimuli

New Tokens Part 1/ɪ~/ˈi/ Contrast

1. cheap (normal) 16. fill (long)
2. chip (normal) 17. cheap (short)
3. each (normal) 18. chip (long)
4. itch (normal) 19. each (short)
5. sleep (normal) 20. itch (long)
6. slip (normal) 21. sleep (short)
7. heel, (normal) 22. slip (long)
8. hill (normal) 23. heel (short)
9. seat (normal) 24. hill (long)
10. sit (normal) 25. seat (short)
11. seek (normal) 26. sit (long)
12. sick (normal) 27. seek (short)
13. sheet (short) 28. sick (long)
14. shit (long) 29. sheet (short)
15. feel (short) 30. shit (long)

Note: Normal refers to a normal speed. Short refers to a short duration for /i/, while long refers to a long duration for /ɪ/. The above tokens were produced by familiar talkers in the training.

New Tokens Part 2 /ɛ~/æ/ Contrast

1. bed 7. pet
2. bad 8. pat
3. beg 9. said
4. bag 10. sad
5. peck 11. ten
6. pack 12. tan
13. set
14. sat

Note: The above tokens were produced by familiar talkers in the training.

New Tokens Part 3 Word-Final /t/-/d/ Contrast

1. bit
2. bid
3. hat
4. had

Note: The above tokens were produced by familiar talkers in the training.

New Talkers
The question tokens in the pretest and posttest were reproduced by two male talkers whose productions were not used for the training.

Final /k/-/g/ Contrast

1. back
2. bag
3. buck
4. bug
5. lock
6. log
7. pick
8. pig
9. tack
10. tag

Note: The above tokens were produced by familiar talkers in the training.

Final /p/-/b/ Contrast

1. nap
2. nab
3. pup
4. pub
5. rip
6. rib

Note: The above tokens were produced by familiar talkers in the training.
Appendix I: Pretest Instructions

**Part 1 /i/-/ɪ/ Contrast**
Directions: In this part of the test, you will click on  to play one word with either /i/ or /ɪ/ sound in each question. After you hear the word, click on one of the two choices and click on the button on the left to confirm your answer. It will take a second or two to send your answer to the server. Then the next question will appear on the top of the screen. Continue answering the questions till the end of the test.

**Part 2 /ɛ/-/æ/ Contrast**
Directions: In this part of the test, you will click on  to play one word with either /ɛ/ or /æ/ sound in each question. After you hear the word, click on one of the two choices and click on the button on the left to confirm your answer. It will take a second or two to send your answer to the server. Then the next question will appear on the top of the screen. Continue answering the questions till the end of the test.

**Part 3 /t/-/d/ Endings**
Directions: In this part of the test, you will click on  to play one word with either “t” or “d” ending in each question. The final “t” or “d” cannot be heard, which is common in conversational English. After you hear the word, click on one of the two choices and click on the button on the left to confirm your answer. It will take a second or two to send your answer to the server. Then the next question will appear on the top of the screen. Continue answering the questions till the end of the test.
Appendix J: Production Word List

**Instructions:** Read the word in each question **twice** in clear and loud speech. Pause for a second after each word.

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>bit /bɪt/</td>
</tr>
<tr>
<td>2.</td>
<td>dead /dɛd/</td>
</tr>
<tr>
<td>3.</td>
<td>sheep /ʃɪp/</td>
</tr>
<tr>
<td>4.</td>
<td>pick /pɪk/</td>
</tr>
<tr>
<td>5.</td>
<td>bet /bɛt/</td>
</tr>
<tr>
<td>6.</td>
<td>bad /bæd/</td>
</tr>
<tr>
<td>7.</td>
<td>deep /dɪp/</td>
</tr>
<tr>
<td>8.</td>
<td>gas /ɡæs/</td>
</tr>
<tr>
<td>9.</td>
<td>need /nɪd/</td>
</tr>
<tr>
<td>10.</td>
<td>beat /bɪt/</td>
</tr>
<tr>
<td>11.</td>
<td>neat /nɪt/</td>
</tr>
<tr>
<td>12.</td>
<td>peek /pɪk/</td>
</tr>
<tr>
<td>13.</td>
<td>pen /pɛn/</td>
</tr>
<tr>
<td>14.</td>
<td>bed /bɛd/</td>
</tr>
<tr>
<td>15.</td>
<td>guess /ɡɛs/</td>
</tr>
<tr>
<td>16.</td>
<td>pan /pæn/</td>
</tr>
<tr>
<td>17.</td>
<td>ship /ʃɪp/</td>
</tr>
<tr>
<td>18.</td>
<td>dad /dæd/</td>
</tr>
<tr>
<td>19.</td>
<td>dip /dɪp/</td>
</tr>
<tr>
<td>20.</td>
<td>bat /bæt/</td>
</tr>
</tbody>
</table>
Appendix K: ID Group Training Instructions

Part 1 /i/-/ɪ/ contrast
Directions: In this part of the test, you will click on 🎧 to play one word with either /i/ or /ɪ/ sound in each question. After you hear the word, click on one of the two choices and click on the button on the left to confirm your answer. It will take a second or two to send your answer to the server. At this time, check the feedback that corresponds to the number of the question in your handout. The next question will appear on the top of the screen. Continue answering the questions till the end of the test.

Note: The last numbered question (#61) contains 60 words spoken by different speakers. Listen to all the words carefully (with no feedback) and then click on “I’ve listened to all the words” to continue.

Part 2 /ɛ/-/æ/ contrast
Directions: In this part of the test, you will click on 🎧 to play one word with either /ɛ/ or /æ/ sound in each question. After you hear the word, click on one of the two choices and click on the button on the left to confirm your answer. It will take a second or two to send your answer to the server. At this time, check the feedback that corresponds to the number of the question in your handout. The next question will appear on the top of the screen. Continue answering the questions till the end of the test.

Note: The last numbered question (#61) contains 60 words spoken by different speakers. Listen to all the words carefully (with no feedback) and then click on “I’ve listened to all the words” to continue.

Part 3 /t/-/d/ endings
Directions: In this part of the test, you will click on 🎧 to play one word with either “t” or “d” ending in each question. The final “t” or “d” cannot be heard, which is common in conversational English. After you hear the word, click on one of the two choices and click on the button on the left to confirm your answer. It will take a second or two to send your answer to the server. At this time, check the feedback that corresponds to the number of the question in your handout. The next question will appear on the top of the screen. Continue answering the questions till the end of the test.

Note: The last numbered question (#61) contains 60 words spoken by different speakers. Listen to all the words carefully (with no feedback) and then click on “I’ve listened to all the words” to continue.
Appendix L: ID Group Training Feedback

Week 1 ID Training Feedback

Part 1 beat-bit, deep-dip
1. /ɪ/: lax
2. /ɪ/: lax
3. /i/: tense
4. /ɪ/: lax
5. /i/: tense
6. /ɪ/: lax
7. /i/: tense
8. /i/: tense
9. /ɪ/: lax
10. /ɪ/: lax
11. /ɪ/: lax
12. /i/: tense
13. /i/: tense
14. /i/: tense
15. /ɪ/: lax
16. /ɪ/: lax
17. /i/: tense
18. /ɪ/: lax
19. /ɪ/: lax
20. /i/: tense
21. /ɪ/: lax
22. /i/: tense
23. /i/: tense
24. /i/: tense
25. /i/: tense
26. /ɪ/: lax
27. /ɪ/: lax
28. /ɪ/: lax
29. /i/: tense
30. /i/: tense
31. /i/: tense
32. /ɪ/: lax
33. /ɪ/: lax
34. /i/: lax
35. /i/: lax
36. /i/: tense
37. /i/: tense
38. /i/: tense
39. /i/: lax
40. /ɪ/: lax
41. /i/: tense
42. /ɪ/: lax
43. /i/: tense
44. /i/: tense
45. /i/: lax
46. /i/: lax
47. /ɪ/: lax
48. /i/: tense
49. /ɪ/: tense
50. /ɪ/: lax
51. /i/: tense
52. /i/: tense
53. /ɪ/: lax
54. /i/: tense
55. /i/: tense
56. /i/: lax
57. /ɪ/: lax
58. /ɪ/: lax
59. /i/: tense
60. /i/: tense
61. Listen to the 60 words

Part 2 bet-bat, dead-dad
1. /ɛ/: the tongue is higher and the duration is shorter than /æ/
2. /ɛ/: the tongue is higher and the duration is shorter than /æ/
3. /æ/: the tongue is lower and the duration is longer than /ɛ/
4. /ɛ/: the tongue is higher and the duration is shorter than /æ/
5. /ɛ/: the tongue is higher and the duration is shorter than /æ/
6. /æ/: the tongue is lower and the duration is longer than /ɛ/
7. /æ/: the tongue is lower and the duration is longer than /ɛ/
8. /ɛ/: the tongue is lower and the duration is longer than /ɛ/
9. /ɛ/: the tongue is higher and the duration is shorter than /æ/
10. /ɛ/: the tongue is higher and the duration is shorter than /æ/
11. /æ/: the tongue is lower and the duration is longer than /ɛ/
12. /ɛ/: the tongue is higher and the duration is shorter than /æ/
13. /æ/: the tongue is lower and the duration is longer than /ɛ/
14. /æ/: the tongue is lower and the duration is longer than /ɛ/
15. /ɛ/: the tongue is higher and the duration is shorter than /æ/
16. /ɛ/: the tongue is higher and the duration is shorter than /æ/
17. /æ/: the tongue is higher and the duration is shorter than /ɛ/
18. /æ/: the tongue is lower and the duration is longer than /ɛ/
19. /æ/: the tongue is lower and the duration is longer than /ɛ/
20. /ɛ/: the tongue is higher and the duration is shorter than /æ/
21. /æ/: the tongue is lower and the duration is longer than /ɛ/
22. /æ/: the tongue is lower and the duration is longer than /ɛ/
23. /æ/: the tongue is lower and the duration is longer than /ɛ/
24. /æ/: the tongue is lower and the duration is longer than /ɛ/
25. /æ/: the tongue is lower and the duration is longer than /ɛ/
26. /æ/: the tongue is lower and the duration is longer than /ɛ/
27. /æ/: the tongue is lower and the duration is longer than /ɛ/
28. /æ/: the tongue is lower and the duration is longer than /ɛ/
29. /æ/: the tongue is lower and the duration is longer than /ɛ/
30. /æ/: the tongue is lower and the duration is longer than /ɛ/
31. /æ/: the tongue is lower and the duration is longer than /ɛ/
32. /æ/: the tongue is lower and the duration is longer than /ɛ/
33. /æ/: the tongue is lower and the duration is longer than /ɛ/
34. /æ/: the tongue is lower and the duration is longer than /ɛ/
35. /æ/: the tongue is lower and the duration is longer than /ɛ/
36. /æ/: the tongue is lower and the duration is longer than /ɛ/
37. /æ/: the tongue is lower and the duration is longer than /ɛ/
38. /æ/: the tongue is lower and the duration is longer than /ɛ/
39. /æ/: the tongue is lower and the duration is longer than /ɛ/
40. /æ/: the tongue is lower and the duration is longer than /ɛ/
41. /æ/: the tongue is lower and the duration is longer than /ɛ/
42. /æ/: the tongue is lower and the duration is longer than /ɛ/
43. /æ/: the tongue is lower and the duration is longer than /ɛ/
44. /æ/: the tongue is lower and the duration is longer than /ɛ/
45. /æ/: the tongue is lower and the duration is longer than /ɛ/
46. /æ/: the tongue is lower and the duration is longer than /ɛ/
47. /æ/: the tongue is lower and the duration is longer than /ɛ/
48. /æ/: the tongue is higher and the duration is shorter than /æ/

236
49. /ɛ/: the tongue is higher and the duration is shorter than /æ/
50. /æ/: the tongue is lower and the duration is longer than /ɛ/
51. /æ/: the tongue is lower and the duration is longer than /ɛ/
52. /æ/: the tongue is lower and the duration is longer than /ɛ/
53. /æ/: the tongue is lower and the duration is longer than /ɛ/
54. /ɛ/: the tongue is higher and the duration is shorter than /æ/
55. /ɛ/: the tongue is higher and the duration is shorter than /æ/
56. /æ/: the tongue is lower and the duration is longer than /ɛ/
57. /ɛ/: the tongue is higher and the duration is shorter than /æ/
58. /ɛ/: the tongue is higher and the duration is shorter than /æ/
59. /æ/: the tongue is lower and the duration is longer than /ɛ/
60. /æ/: the tongue is lower and the duration is longer than /ɛ/
61. Listen to the 60 words

Part 3 bet-bed
1. t ending, the vowel before /t/ is pronounced shorter
2. t ending, the vowel before /t/ is pronounced shorter
3. d ending, the vowel before /d/ is pronounced longer
4. t ending, the vowel before /t/ is pronounced shorter
5. d ending, the vowel before /d/ is pronounced longer
6. t ending, the vowel before /t/ is pronounced shorter
7. t ending, the vowel before /t/ is pronounced shorter
8. t ending, the vowel before /t/ is pronounced shorter
9. d ending, the vowel before /d/ is pronounced longer
10. d ending, the vowel before /d/ is pronounced longer
11. t ending, the vowel before /t/ is pronounced shorter
12. d ending, the vowel before /d/ is pronounced longer
13. t ending, the vowel before /t/ is pronounced shorter
14. t ending, the vowel before /t/ is pronounced shorter
15. d ending, the vowel before /d/ is pronounced longer
16. d ending, the vowel before /d/ is pronounced longer
17. d ending, the vowel before /d/ is pronounced longer
18. t ending, the vowel before /t/ is pronounced shorter
19. t ending, the vowel before /t/ is pronounced shorter
20. d ending, the vowel before /d/ is pronounced longer
21. d ending, the vowel before /d/ is pronounced longer
22. t ending, the vowel before /t/ is pronounced shorter
23. t ending, the vowel before /t/ is pronounced shorter
24. t ending, the vowel before /t/ is pronounced shorter
25. t ending, the vowel before /t/ is pronounced shorter
26. d ending, the vowel before /d/ is pronounced longer
27. d ending, the vowel before /d/ is pronounced longer
28. d ending, the vowel before /d/ is pronounced longer
29. t ending, the vowel before /t/ is pronounced shorter
30. t ending, the vowel before /t/ is pronounced shorter
31. d ending, the vowel before /d/ is pronounced longer
32. t ending, the vowel before /t/ is pronounced shorter
33. d ending, the vowel before /d/ is pronounced longer
34. d ending, the vowel before /d/ is pronounced longer
35. t ending, the vowel before /t/ is pronounced shorter
36. d ending, the vowel before /d/ is pronounced longer
37. d ending, the vowel before /d/ is pronounced longer
38. t ending, the vowel before /t/ is pronounced shorter
39. t ending, the vowel before /t/ is pronounced shorter
40. t ending, the vowel before /t/ is pronounced shorter
41. d ending, the vowel before /d/ is pronounced longer
42. d ending, the vowel before /d/ is pronounced longer
43. t ending, the vowel before /t/ is pronounced shorter
44. d ending, the vowel before /d/ is pronounced longer
45. d ending, the vowel before /d/ is pronounced longer
46. t ending, the vowel before /t/ is pronounced shorter
47. t ending, the vowel before /t/ is pronounced shorter
48. t ending, the vowel before /t/ is pronounced shorter
49. d ending, the vowel before /d/ is pronounced longer
50. d ending, the vowel before /d/ is pronounced longer
51. t ending, the vowel before /t/ is pronounced shorter
52. d ending, the vowel before /d/ is pronounced longer
53. t ending, the vowel before /t/ is pronounced shorter
54. t ending, the vowel before /t/ is pronounced shorter
55. d ending, the vowel before /d/ is pronounced longer
56. d ending, the vowel before /d/ is pronounced longer
57. d ending, the vowel before /d/ is pronounced longer
58. t ending, the vowel before /t/ is pronounced shorter
59. t ending, the vowel before /t/ is pronounced shorter
60. d ending, the vowel before /d/ is pronounced longer
61. Listen to the 60 words
Appendix M: SD Group Training Instructions

**Part 1 /i/-/ɪ/ contrast**
Directions: In this part of the test, you will click on 🎧 to play two words spoken by two speakers with either /i/ or /ɪ/ sound in each question. After you hear the words, if you think they are saying the same word, click on the answer marked “same.” Otherwise click on the answer marked “different.” Then click on the button on the left to confirm your answer. It will take a second or two to send your answer to the server. At this time, check the feedback that corresponds to the number of the question in your handout. The next question will appear on the top of the screen. Continue answering the questions till the end of the test.

**Part 2 /ɛ/-/æ/ contrast**
Directions: In this part of the test, you will click on 🎧 to play two words spoken by two speakers with either /ɛ/ or /æ/ sound in each question. After you hear the words, if you think they are saying the same word, click on the answer marked “same.” Otherwise click on the answer marked “different.” Then click on the button on the left to confirm your answer. It will take a second or two to send your answer to the server. At this time, check the feedback that corresponds to the number of the question in your handout. The next question will appear on the top of the screen. Continue answering the questions till the end of the test.

**Part 3 /t/-/d/ endings**
Directions: In this part of the test, you will click on 🎧 to play two words by two speakers with either “t” or “d” ending in each question. The final “t” or “d” cannot be heard, which is common in conversational English. After you hear the words, if you think they are saying the same word, click on the answer marked “same.” Otherwise click on the answer marked “different.” Then click on the button on the left to confirm your answer. It will take a second or two to send your answer to the server. At this time, check the feedback that corresponds to the number of the question in your handout. The next question will appear on the top of the screen. Continue answering the questions till the end of the test.
Appendix N: SD Group Training Feedback

Week 1 SD Training Feedback

**Part 1 beat-bit, deep-dip**

1. Same: /i/  
2. Different: /i/ vs. /ɪ/  
3. Different: /i/ vs. /ɪ/  
4. Same: /i/  
5. Same: /i/  
6. Different: /i/ vs. /ɪ/  
7. Same: /i/  
8. Same: /i/  
9. Different: /i/ vs. /ɪ/  
10. Different: /i/ vs. /ɪ/  
11. Different: /i/ vs. /ɪ/  
12. Same: /i/  
13. Different: /i/ vs. /ɪ/  
14. Same: /i/  
15. Same: /i/  
16. Different: /i/ vs. /ɪ/  
17. Same: /i/  
18. Different: /i/ vs. /ɪ/  
19. Different: /i/ vs. /ɪ/  
20. Same: /i/  
21. Same: /i/  
22. Different: /i/ vs. /ɪ/  
23. Different: /i/ vs. /ɪ/  
24. Same: /i/  
25. Same: /i/  
26. Different: /i/ vs. /ɪ/  
27. Same: /i/  
28. Same: /i/  
29. Different: /i/ vs. /ɪ/  
30. Different: /i/ vs. /ɪ/  
31. Different: /i/ vs. /ɪ/  
32. Same: /i/  
33. Different: /i/ vs. /ɪ/  
34. Same: /i/  
35. Same: /i/  
36. Different: /i/ vs. /ɪ/  
37. Same: /i/  
38. Same: /ɪ/  
39. Different: /i/ vs. /ɪ/  
40. Different: /i/ vs. /ɪ/  
41. Same: /i/  
42. Different: /i/ vs. /ɪ/  
43. Different: /i/ vs. /ɪ/  
44. Same: /i/  
45. Same: /i/  
46. Different: /i/ vs. /ɪ/  
47. Same: /i/  
48. Same: /i/  
49. Different: /i/ vs. /ɪ/  
50. Different: /i/ vs. /ɪ/  
51. Different: /i/ vs. /ɪ/  
52. Same: /i/  
53. Different: /i/ vs. /ɪ/  
54. Same: /i/  
55. Same: /i/  
56. Different: /i/ vs. /ɪ/  
57. Same: /i/  
58. Different: /i/ vs. /ɪ/  
59. Same: /i/  
60. Different: /i/ vs. /ɪ/

**Part 2 bet-bat, dead-dad**

1. Same: /e/  
2. Different: /æ/ vs. /e/  
3. Different: /æ/ vs. /æ/  
4. Same: /æ/  
5. Same: /æ/  
6. Different: /æ/ vs. /æ/  
7. Same: /æ/  
8. Same: /æ/  
9. Different: /e/ vs. /æ/  
10. Different: /æ/ vs. /e/  
11. Different: /æ/ vs. /æ/  
12. Same: /æ/  
13. Different: /æ/ vs. /æ/  
14. Same: /e/  
15. Same: /æ/  
16. Different: /æ/ vs. /æ/  
17. Same: /e/  
18. Different: /æ/ vs. /æ/  
19. Different: /æ/ vs. /æ/  
20. Same: /æ/  
21. Same: /æ/  
22. Different: /æ/ vs. /æ/  
23. Different: /æ/ vs. /æ/  
24. Same: /æ/  
25. Same: /æ/  
26. Different: /æ/ vs. /æ/  
27. Same: /æ/  
28. Same: /æ/  
29. Different: /æ/ vs. /æ/  
30. Different: /æ/ vs. /æ/  
31. Different: /æ/ vs. /æ/  
32. Same: /æ/  
33. Different: /æ/ vs. /æ/  
34. Same: /æ/  
35. Same: /æ/  
36. Different: /æ/ vs. /æ/  
37. Same: /æ/  
38. Same: /æ/  
39. Different: /æ/ vs. /æ/  
40. Different: /æ/ vs. /æ/  
41. Same: /æ/  
42. Different: /æ/ vs. /æ/  
43. Different: /æ/ vs. /æ/  
44. Same: /æ/  
45. Same: /æ/  
46. Different: /æ/ vs. /æ/  
47. Same: /æ/  
48. Same: /æ/  
49. Different: /æ/ vs. /æ/  
50. Different: /æ/ vs. /æ/
Different: /ɛ/ vs. /æ /
Same: /ɛ/
Different: /æ/ vs. /ɛ/
Same: /æ/
Different: /æ/ vs. /ɛ/
Same: /ɛ/
Different: /ɛ/ vs. /æ/
Same: /æ/
Different: /æ/ vs. /ɛ/
Same: /ɛ/
Different: /ɛ/ vs. /æ/
Same: /æ/
Different: /æ/ vs. /ɛ/
Same: /ɛ/
Different: /ɛ/ vs. /æ/
Same: /æ/
Different: /æ/ vs. /ɛ/
Same: /ɛ/
Different: /ɛ/ vs. /æ/
Same: /æ/
Different: /æ/ vs. /ɛ/
Same: /ɛ/
Different: /ɛ/ vs. /æ/
Same: /æ/
Different: /æ/ vs. /ɛ/
Same: /ɛ/

Part 3 bet-bed
1. Same: t ending
2. Different: d vs. t
3. Different: d vs. t
4. Same: t ending
5. Same: d ending
6. Different: t vs. d
7. Same: t ending
8. Same: d ending
9. Different: d vs. t
10. Different: t vs. d
11. Different: d vs. t
12. Same: t ending
13. Different: t vs. d
14. Same: d ending
15. Same: t ending
16. Different: d vs. t
17. Same: d ending
18. Different: t vs. d
19. Different: d vs. t
20. Same: t ending
21. Same: d ending
22. Different: t vs. d
23. Different: d vs. t
24. Same: t ending
25. Same: d ending
26. Different: t vs. d
27. Same: t ending
28. Same: d ending
29. Different: d vs. t
30. Different: t vs. d
31. Different: d vs. t
32. Same: t ending
33. Different: t vs. d
34. Same: d ending
35. Same: t ending
36. Different: d vs. t
37. Same: d ending
38. Same: t ending
39. Different: t vs. d
40. Different: d vs. t
41. Same: d ending
42. Different: t vs. d
43. Different: d vs. t
44. Same: t ending
45. Same: d ending
46. Different: t vs. d
47. Same: t ending
48. Same: d ending
49. Different: d vs. t
50. Different: t vs. d
51. Different: d vs. t
52. Same: t ending
53. Different: t vs. d
54. Same: d ending
55. Different: d vs. t
56. Different: t vs. d
57. Same: t ending
58. Different: d vs. t
59. Same: d ending
60. Different: t vs. d
Appendix O: Generalization Test Instructions

**New Tokens P1 /i/-/ɪ/ Contrast**
Directions: In this part of the test, you will click on \( \text{[Play]} \) to play one word with either /i/ or /ɪ/ sound in each question. After you hear the word, click on one of the two choices and click on the button on the left to confirm your answer. It will take a second or two to send your answer to the server. Then the next question will appear on the top of the screen. Continue answering the questions till the end of the test.

**New Tokens P2 /ɛ/-/æ/ Contrast**
Directions: In this part of the test, you will click on \( \text{[Play]} \) to play one word with either /ɛ/ or /æ/ sound in each question. After you hear the word, click on one of the two choices and click on the button on the left to confirm your answer. It will take a second or two to send your answer to the server. Then the next question will appear on the top of the screen. Continue answering the questions till the end of the test.

**New Tokens P3 /t/-/d/ Ending**
Directions: In this part of the test, you will click on \( \text{[Play]} \) to play one word with either “t” or “d” ending in each question. The final “t” or “d” cannot be heard, which is common in conversational English. After you hear the word, click on one of the two choices and click on the button on the left to confirm your answer. It will take a second or two to send your answer to the server. Then the next question will appear on the top of the screen. Continue answering the questions till the end of the test.

**New Talkers P1 /i/-/ɪ/ Contrast**
Directions: In this part of the test, you will click on \( \text{[Play]} \) to play one word with either /i/ or /ɪ/ sound in each question. After you hear the word, click on one of the two choices and click on the button on the left to confirm your answer. It will take a second or two to send your answer to the server. Then the next question will appear on the top of the screen. Continue answering the questions till the end of the test.

**New Talkers P2 /ɛ/-/æ/ Contrast**
Directions: In this part of the test, you will click on \( \text{[Play]} \) to play one word with either /ɛ/ or /æ/ sound in each question. After you hear the word, click on one of the two choices and click on the button on the left to confirm your answer. It will take a second or two to send your answer to the server. Then the next question will appear on the top of the screen. Continue answering the questions till the end of the test.

**New Talkers P3 /t/-/d/ Endings**
Directions: In this part of the test, you will click on \( \text{[Play]} \) to play one word with either “t” or “d” ending in each question. The final “t” or “d” cannot be heard, which is common in conversational English. After you hear the word, click on one of the two choices and
click on the button on the left to confirm your answer. It will take a second or two to send your answer to the server. Then the next question will appear on the top of the screen. Continue answering the questions till the end of the test.

/k/-/g/ Endings
Directions: In this part of the test, you will click on to play one word with either “k” or “g” ending in each question. The final “k” or “g” cannot be heard, which is common in conversational English. After you hear the word, click on one of the two choices and click on the button on the left to confirm your answer. It will take a second or two to send your answer to the server. Then the next question will appear on the top of the screen. Continue answering the questions till the end of the test.

/p/-/b/ Endings
Directions: In this part of the test, you will click on to play one word with either “p” or “b” ending in each question. The final “p” or “b” cannot be heard, which is common in conversational English. After you hear the word, click on one of the two choices and click on the button on the left to confirm your answer. It will take a second or two to send your answer to the server. Then the next question will appear on the top of the screen. Continue answering the questions till the end of the test.