THE PERCEPTION AND PRODUCTION OF PALATAL CODAS BY KOREAN L2
LEARNERS OF ENGLISH

BY

AMANDA R. HUENSCH

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Linguistics
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2013

Urbana, Illinois

Doctoral Committee:

Associate Professor Tania Ionin, Chair
Assistant Professor Annie Tremblay, University of Kansas, Director of Research
Professor Wayne Dickerson
Associate Professor Chilin Shih
ABSTRACT

One of the central questions within the field of the acquisition of second language (L2) phonology is the role that speech perception plays in accurate speech production and whether, and if so, how, the speech perception and production systems are linked. Existing theories of L2 speech perception such as the Speech Learning Model (SLM) (Flege, 1991, 1995, 2003), the Native Language Magnet Model (NLM) (Kuhl & Iverson, 1995; Kuhl, 2000), and the Perceptual Assimilation Model (PAM) (Best, 1994, 1995; Best, McRoberts & Goodell, 2001), have made predictions about the acquisition of a second language phonological system, but are mostly concerned with the acquisition of L2 segments and segmental contrasts in relation to first language (L1) segments. Previous work indicates that syllable structure constraints can also play a role in speech perception (e.g., Dupoux, Kakehi, Hirose, Pallier, & Mehler, 1999; Kabak & Idsardi, 2007) and speech production (e.g., Abrahamsson, 2003; Hancin-Bhatt & Bhatt, 1997; Hancin-Bhatt, 2000).

This dissertation comprises three sets of experiments designed to investigate speech perception and production in relation to syllable structure constraints, as well as the mediating effect that perceptual training has on both perception and production, thereby shedding light on the relationship between L2 speech perception and L2 speech production. The experiments investigate the perception and production of existing and novel phonemes within an existing but restricted syllable structure, namely palatal codas in the English of native Korean speakers. Using an AXB perception task and a read-aloud task, Experiment 1 compares L2 perception and production accuracies of palatal codas. Experiment 2 uses a forced-choice word-identification task and a read-aloud task to investigate cues that may help L2 learners perceive palatal codas, and it corroborates results from the production task in Experiment 1 with a different set of
learners. Experiment 3 implements perceptual phonetic training on palatal codas using a pretest/post-test design, and it compares the effects of training on improvements in perception and production of palatal codas for familiar and novel words and talkers. A control group who completed a perceptual training on targets unrelated to the structures that are the focus of this dissertation is also included.

The results of Experiment 1 show that (1) the existence of a phoneme in the L1 does not necessarily facilitate its acquisition in an existing, but restricted syllable structure, (2) no direct relationship between learners’ perception and production accuracies emerges, and (3) learners at higher proficiencies show evidence of having been successful in the acquisition of palatal codas. Experiment 2 demonstrates that some learners are able to use native-like cues to perceive palatal codas, but do so only in certain tasks. Experiment 3 indicates that (1) learners who received perceptual phonetic training on palatal codas outperform those who did not in perception and production tasks, (2) perceptual phonetic training on palatal codas is successful in improving the perception and production accuracies of palatal codas, (3) learners are able to generalize learning from perceptual training not only to new words and new talkers, but also to new discourse contexts, and (4) similar to the findings in Experiment 1, improvements in perception are not always directly linked to improvements in production.

The finding that accurate perception of segments within an existing but restricted syllable structure can be difficult provides implications for L2 speech perception theories that syllable structure must be taken into consideration to fully understand acquisitional patterns. The finding that perceptual training improves production and allows for generalizability to new words and talkers in both perception and production provides implications for L2 speech learning theories that perception and production systems are linked. It also provides important pedagogical
implications for pronunciation classes and teachers in that supplying a variety of input for
learners is necessary. Because the perceptual training used in this research was designed to be
pedagogically feasible, it provides one promising means of supplementing out-of-class activities
in pronunciation classes. The finding that perceptual training can improve production accuracies
implies a connection between perception and production systems. However, the lack of a
consistent correlation between perception and production improvements adds to the growing
body of work in which questions the existence of a direct link between perception and
production systems.
ACKNOWLEDGEMENTS

There are many people I would like to thank for helping me in the completion of this research. First and foremost, I want to thank my director of research, Annie Tremblay. Her invaluable feedback, support, attention to detail and drive for excellence have been assets. I would also like to thank my chair, Tania Ionin and the other members of my committee, Wayne Dickerson and Chilin Shih. Their willingness to provide guidance, feedback, and support has truly impacted and strengthened my work.

I would also like to thank all who participated in my research over the years. I have made lasting friends in the Korean community in Champaign-Urbana. Without their volunteering their time and believing in my work, the research in this dissertation could not have been completed. In particular, I would like to thank Man-Ki and Jung Eun for their invaluable help in recruiting as well as for being great friends. I would also like to thank others who have helped me by allowing me to recruit from their classes, listening to practice talks, bouncing ideas around, and just generally being there for me: Patti, Laura, Lisa, Sam, Kayla, Sue, Rhi and Dustin; I couldn’t have done it with you.

I would also like to thank my undergraduate research assistants, Kate Tyndall and Hannah Greening, who made me excited about my work all over again by breathing a breath of fresh air into the research experience. I hope they learned even half of what I did while working with them.

My family has also been invaluable to me during this time. To my sister Anna, beyond being the best sister in the word, if I had to choose one person to work with, it would be her. She continually helped me to find the most efficient solution to my problem, and her willingness to
test out experiments or just lend an ear when I needed to discuss ideas saw no bounds. Without her, I truly would have been lost. Thanks buddy.

To my brother Michael who wrote me programs at a moment’s notice to help speed up data processing and save me time. Thank you also for introducing me to my rubber duck.

Of course, I also have to thank my parents for believing in me and supporting me through all my years of schooling. They were the first ones to instill a passion of learning and sense of inquisitiveness in me.

I would also like to thank my colleagues at the University of Illinois. To Ryan, the best office neighbor ever, always ready to lend an ear or answer an email. Thank you for your time. To my fellow graduate students in the program, you made the department a home for me: Karen, Erin, Veronica, Sue, and Matt.

To my friends on campus who made my many years in Champaign-Urbana full of unforgettable memories and who also participated in my experiments. Thank you Tom, Emma, Alex, Jeremy, Kate, Alex, Ryan, Duncan, Dom, Alex and Nicole. I love you all and miss you already. Abe Lincoln’s Pants forever!

And finally, to my best friend Luke, for all the hours he spent listening to me talk about my work, I feel like he could have written this dissertation by now.

This research was supported financially by a University of Illinois graduate college dissertation completion fellowship.
# TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION ........................................................................................................1

CHAPTER 2: BACKGROUND ........................................................................................................8

CHAPTER 3: EXPERIMENT 1: PERCEPTION AND PRODUCTION OF LEARNERS AT DIFFERENT PROFICIENCY LEVELS ................................................................. 39

CHAPTER 4: EXPERIMENT 2: NATIVE SPEAKER AND LEARNER SENSITIVITY TO STEM VOWEL AND FINAL VOWEL LENGTH ........................................................................ 79

CHAPTER 5: EXPERIMENT 3: PERCEPTUAL TRAINING AND ITS EFFECTS ON PERCEPTION AND PRODUCTION ......................................................................................... 104

CHAPTER 6: GENERAL DISCUSSION AND CONCLUSION ......................................................... 169

REFERENCES ............................................................................................................................... 182

APPENDIX A: CLOZE TEST ....................................................................................................... 194

APPENDIX B: LANGUAGE BACKGROUND QUESTIONNAIRE ............................................... 195

APPENDIX C: LIST OF STIMULI USED IN EXPERIMENT 1 .................................................... 197

APPENDIX D: EXPERIMENT INSTRUCTIONS ....................................................................... 198

APPENDIX E: EXPERIMENT 1 AXB PERCEPTION DATA RESULTS PRESENTED IN PERCENT ACCURACY .......................................................................................... 201

APPENDIX F: EXPERIMENT 3 STIMULI ................................................................................. 205
APPENDIX G: COMPLETE LIST OF STIMULI FROM DIALOG/PARAGRAPH PRODUCTION MEASURE IN EXPERIMENT 3 ................................................................. 206

APPENDIX H: EXAMPLE DIALOG FROM EXPERIMENT 3 PRODUCTION MEASURE ................................................................................................. 209

APPENDIX I: ANALYSIS OF DAILY TRAINING DATA ......................................................... 210

APPENDIX J: EXPERIMENTAL PERCEPTUAL TRAINING INSTRUCTIONS ............... 214

APPENDIX K: PERCEPTUAL TRAINING TIME COMPARISON BY PARTICIPANT ...... 216
CHAPTER 1
INTRODUCTION

The acquisition of native-like phonology appears to be one of the most difficult hurdles for late second language learners to overcome. In fact, some researchers conducting studies within the domain of the Critical Period Hypothesis have even suggested that exposure to a second language by six years of age is necessary to achieve native-like proficiency (e.g., Birdsong, 1992; Johnson & Newport, 1989; Long, 1990). It seems that complete acquisition of second language (L2) phonology is the exception rather than the rule when it comes to late learners. This is problematic for learners in that having an L2 accent not only negatively affects communication by causing problems with intelligibility, but also results in negative stereotypes and perceptions. Many factors have been argued to be responsible for these difficulties: cognitive constraints related to age of acquisition, limited amount of exposure to the target language, limited use of the target language, and influences of the native language, among others (for discussion, see Bohn & Munro, 2007; Strange, 1995).

L2 learners must acquire, among other things, the segmental contrasts of that language and how segments fit together in permissible combinations. It has been shown, for example, that English speakers have difficulty differentiating Hindi retroflexes and dental stops, contrasts that do not exist in English (Polka, 1991). Familiar first language (L1) segment contrasts become difficult for L2 learners if they are located in new positions in words (e.g., Broersma, 2005, 2010; Flege & Wang, 1989). Even advanced L2 learners evidence difficulties. For example, the /l/-/l/ contrast is notoriously problematic for some learners of English (e.g., Goto, 1971; Sheldon & Strange, 1982). In fact, learning an L2 in childhood does not guarantee native-like acquisition (e.g., Flege & MacKay, 2004). Nevertheless, we have evidence from training studies that
acquiring perceptual contrasts (e.g., Aoyama, Flege, Guion, Akahane-Yamada, & Yamada, 2004; Jamieson & Morosan, 1986; Lively, Logan, & Pisoni, 1993; Logan, Lively, & Pisoni, 1991) and improving production (e.g., Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997; Bradlow, Akahane-Yamada, Pisoni, & Tohkura, 1999) are possible.

One of the central questions within the field of the acquisition of L2 phonology is the role that speech perception plays in accurate speech production and whether, and if so, how, speech perception and production systems are linked. Existing theories of L2 speech perception such as the Speech Learning Model (SLM) (Flege, 1991, 1995, 2003), the Native Language Magnet Model (NLM) (Kuhl & Iverson, 1995; Kuhl, 2000), and the Perceptual Assimilation Model (PAM) (Best, 1994, 1995; Best, McRoberts & Goodell, 2001), have made predictions about the acquisition of a second language phonological system, but are mostly concerned with the acquisition of L2 segments and segmental contrasts in relation to L1 segments. We have indications from previous work that syllable structure constraints can also play a role in speech perception (e.g., Dupoux, Kakehi, Hirose, Pallier, & Mehler, 1999; Kabak & Idsardi, 2007) and speech production (e.g., Abrahamsson, 2003; Hancin-Bhatt & Bhatt, 1997; Hancin-Bhatt, 2000). In addition, while some of these theories of L2 speech perception make implicit predictions regarding the relationship between speech perception and production (SLM, PAM), that relationship is less clear in others (NLM).

The goal of this dissertation is to investigate speech perception and production in relation to syllable structure constraints, as well as the mediating effect that perceptual training has on both perception and production, thereby shedding light on the relationship between L2 speech perception and L2 speech production. The language group that is the focus of this dissertation is L2 learners of English who speak Korean as a first language (L1). Differences between Korean
and English provide an interesting test case for investigating questions concerning the relationship between perception and production and the role of syllable structure constraints. Korean and English, for example, both allow coda consonants, but they differ in the segments allowed in that position.

More specifically, English contains obstruents in the following categories: six plosives (p, b, t, d, k, g), nine fricatives (f, v, θ, s, z, j, ʒ, h), and two affricates (ʧ, ʤ), and allows for a variety of syllable types: (C)(C)(C)V(C)(C)(C)(C). Although there are some restrictions for which consonant can, for example, be the first C in a CCCV sequence, important for our discussion is that English allows palatals in coda position.¹ On the other hand, the Korean obstruents consist of nine stops (unaspirated/lenis /p t k/, aspirated /pʰ tʰ kʰ/, and fortis /p’ t’ k’/), three fricatives (/s s’ h/), and three affricates (/ʧʰ, ʧ’, ʧ/). Korean has a robust system of neutralization in codas. In coda position, stops neutralize to the lenis variety. Korean contains /ʃ/, but only as an allophone of /s/ before high vowels. With regard to syllable structure, Korean allows V, CV, and CVC syllables; however, only lenis stops, nasals /m n ɲ/, and the lateral /l/ are allowed in codas (Yeon, 2004). These similarities and differences provide an ideal test case for determining how syllable structure constraints play a role in speech perception and production.

The perception and production of Korean L2 learners of English has been studied with a variety of segments, including /ɹ/-/l/ (Borden, Gerber, & Milsark, 1983), word-final stops (Tsukada, Birdsong, Mack, Sung, Bialystok, & Flege, 2004), vowels (Ingram & Park, 1997; Tsukada, Birdsong, Bialystok, Mack, Sung, & Flege, 2005), and /s/-/ʃ/ (Fox, Jacewicz, Eckman, Iverson, & Lee, 2009), among others. There is also work that considers the coda consonants of Korean L2 learners of English, but it focuses on non-palatal obstruents (/p, b, f, v, θ, t, d, s, z/)

¹ Some dialects of English do not have /ʒ/ in final position whereas others do. These differences emerge in words like garage.
Less studied is the acquisition of the English palatals /ʃʧʤ/, especially in coda position. If we return to the descriptions of English and Korean provided above, important differences can be noted between English and Korean with regard to palatals: (1) Korean does not have the phoneme /ʤ/; (2) although Korean has the segment /ʃ/, it only surfaces as a context-dependent allophone; and (3) English syllable structure allows palatals in codas whereas Korean syllable structure does not.

Investigations of Korean adaptations of loanwords show that palatals in word-final codas are produced with an epenthetic [i] (e.g., Kim, 2009). In English production, Korean L2 learners of English sometimes produce an epenthetic vowel in coda or word-final position following English palatals (both fricatives and affricates, /ʃʧʤ/), resulting in productions such as language[i] instead of language (Schmidt & Meyer, 1995). Yet, we do not have systematic evidence regarding the magnitude (i.e., what percentage of final palatals have a vowel epenthesized after them) or cause (e.g., difficulties in perception, articulation, or a combination of the two) of this problem, nor do we know how these errors fit into the larger interlanguage (IL) system of Korean L2 learners of English (e.g., in what phonological contexts Korean learners of English can correctly produce palatals). This dissertation represents a first step in attempting to answer some of these questions. Note that it is not claimed that Korean speakers’ vowel epenthesis after final palatals is a major contributing factor to this group’s unintelligibility in English. Rather, motivation for investigating this issue comes from the desire to determine why this problem exists and why it is a typical feature of a Korean’s IL system. Although these errors might not significantly affect intelligibility, they do contribute to a Korean speaker’s noticeable foreign accent, and the IL system of Korean L2 learners of English with regard to
these consonants has not been adequately described or explained in the literature, thus warranting a systematic investigation.

It is clear from the basic inventories of Korean and English that several differences exist between the two languages that could account for the difficulties Korean learners of English demonstrate with respect to palatals. What is unclear is whether these difficulties originate at the level of perception (or representation) or production (articulation). Here, I presume that perception reflects, at least at some level, representations stored by learners. We know from previous research (e.g., Werker and Logan, 1985) that different tasks access representations at different levels (e.g., AXB tasks\textsuperscript{2} with short interstimulus intervals tap into representations at the acoustic level; AXB tasks with long interstimulus intervals tap into representations at the phonemic/phonological level; categorization tasks tap the phonemic/phonological representation because they involve lexical access). The tasks reported on in this dissertation tap into representations at the phonemic/phonological level. Thus, I attempt to determine whether learners can differentiate between CVC and CVCV at a higher level than the acoustic level.

Returning to a discussion of whether difficulties originate at the level of perception or production, it could be the case that Korean learners of English initially perceive a vowel following English palatals in coda position, thus storing that vowel as part of their lexical representations. Alternatively, Korean learners of English could perceive palatals in coda position accurately (i.e., without a vowel following them), but have difficulties with the articulation of these sounds. The purpose of this research is to investigate these issues in order to better understand the contributing effects of perception and production in the L2 acquisition of

\textsuperscript{2} An AXB task is one in which participants hear three stimuli in a row and decide whether the second stimulus (X) is the same as the first (A) or third (B). For example, if a participant heard the sequence lock – lock – rock, they should choose A because the second word (lock) was the same as the first word.
phonology, and ultimately design a perceptual training method that would be effective in improving these learners’ pronunciation.

Researching this phenomenon will allow us to test existing theories of L2 speech perception such as the Speech Learning Model and the Perceptual Assimilation Model, which attempt to explain why second language learners make certain mistakes and why some mistakes are more persistent than others. This particular phenomenon will allow us to extend these theories beyond the acquisition of segments in relation to other segments by including evidence related to syllable structure constraints. Ultimately, results from this dissertation contribute to a better understanding of both the role of speech perception in the acquisition of an L2 and the development of phonological IL systems.

Chapter 2 begins with a discussion of the existing models in speech perception that contribute to an understanding of L2 learners’ difficulties in the perception and production of L2 sounds. This discussion is followed by a review of research on the acquisition of L2 syllable structure from both a production and a perception perspective, on the relationship between the perception and production of L2 sounds, and on the use of perceptual phonetic training in testing that relationship. Next, I discuss a developmental sequence that has been proposed for the acquisition of codas, and outline the research questions and formulate specific predictions for the experiments reported upon in this dissertation. Chapter 3 presents results from Experiment 1, which compares the perception and production of palatal codas by Korean L2 learners of English. Chapter 4 presents results from Experiment 2, which investigates cues that may help L2 learners perceive palatal codas and corroborates results from the production task in Experiment 1 with a different set of learners. Chapter 5 presents results from Experiment 3, which utilizes
perceptual phonetic training to investigate how learning in one skill (e.g., perception) can influence performance in another (e.g., production), and Chapter 6 concludes the dissertation.
CHAPTER 2
BACKGROUND

2.1 Models of Speech Perception

In order to better understand inaccurate productions in second language pronunciation, it is also necessary to investigate speech perception and the possible links between perception and production systems. It could be the case that problems in production arise because words are initially inaccurately perceived and thus stored in a non-target-like manner by the learner. Here, I review three models that have dominated L2 research and discuss how they might contribute to our understanding of the phenomenon that is the focus of this dissertation. These models are Flege’s Speech Learning Model (Flege, 1995, 2003), Kuhl and Iverson’s Native Language Magnet Model (Kuhl & Iverson, 1995; Iverson & Kuhl, 1995, 1996; Kuhl, 2000), and Best’s Perceptual Assimilation Model (Best, 1995; Best & McRoberts, 2003).

The Speech Learning Model (SLM) is concerned with difficulties learners have in the ultimate attainment of L2 consonants and vowels. It is a psychoacoustic model—it takes as its primitives acoustic properties of the speech signal and investigates phonetic categories as opposed to, for example, articulatory gestures. It posits that when learning an L1, a child becomes attuned to the sound contrasts in that language and stores any language-specific aspects of those sounds in phonetic categories in long term memory. However, these categories are not fixed and can change over time. For example, if an adult acquires an L2, it will be represented in the same phonological space as the L1 and will thus affect the categories already residing there. The model states that learners might have trouble differentiating L2 sounds (either a new L2 contrast or an L1 sound and an L2 sound) for several reasons. It could be that the sound is similar enough to an existing sound that it is assimilated into an existing category. Alternatively, it could
be that the L1 phonology filters out important feature information, thus making the new sound difficult to differentiate. In either case, the SLM postulates that not having accurate perception will lead to problems in production. In terms of the relationship between perception and production, because it assumes a psychoacoustic view of speech perception, the SLM would not predict a direct relationship between perception and production, but rather an indirect one. Two important details to note about the SLM are that (1) it does not claim that all production errors have a basis in perception, and (2) it is primarily concerned with the acquisition of L2 segments and L2 segmental contrasts in relation to L1 segments and does not consider the potential effects of syllable structure constraints.3

Nevertheless, the SLM does potentially offer some predictions concerning the categorization of English fricatives and palatals in word-final position, in that it hypothesizes that L1 and L2 sounds are related to each other at a position-sensitive allophonic level rather than at an abstract phonemic level. Korean does not allow any fricatives or affricates in word-final positions; thus, there are no existing phonetic categories for these sounds in those positions. The SLM predicts difficulties in cases when a learner attempts to establish new phonetic categories in a position that has similar, existing phonetic categories in that same position. However, in the case of Korean, learners should be able to eventually establish new phonetic categories (for fricatives and affricates) in word-final positions that will not be affected by existing categories (because no affricates or fricatives are allowed in that position). In other words, we might expect that /sʃʧʤ/ would be eventually acquired by Korean learners of English, because they should be able to establish new phonetic categories in word-final position that are not affected by existing categories in that same position. Additionally, that establishing of categories should not

3 Flege qualifies that “motoric output constraints based on permissible syllable types in the L1 may cause Spanish speakers to pronounce the word ‘school’ as [eskul]” (Flege, 1995, p. 238).
be affected by whether those sounds exist in other positions in the language. Thus, the SLM would predict that native Korean speakers learning English could eventually establish new categories for fricatives and palatals in final position, and that during the process of acquisition, these L2 learners might have an equally easy or difficult time with the segments /sʃʧʤ/ because none are allowed in word-final position in Korean. This is not to say that the SLM denies that there can be differences in difficulty regarding the acquisition of segments. We know that even in L1 acquisition, some segments take longer to be fully acquired than others. If we consider the acquisition of /sʃʧʤ/ by child learners of English, /s/ is typically acquired by age 3 whereas /ʃʧʤ/ are typically acquired by age 4 (Sander, 1972; Smit, Hand,Frieilinger, Bernthal, & Bird, 1990). Therefore, it is possible that /s/ will show different patterns than /ʃʧʤ/. Despite this fact, it remains that both fricatives and affricates neutralize in coda position in Korean, and thus the establishment of a new phonetic category is required for all four of these segments. Therefore, we predict that there will be no difference in perception accuracy results regarding their acquisition in English.

The Native Language Magnet (NLM) model posits that when learning an L1, the acoustic space of a child is “warped” in such a way that there is “a change in perceived distances in the acoustic space underlying phonetic distinctions” (Kuhl & Iverson, 1995, p. 122). This results in a phenomenon known as the perceptual magnet effect. Initial evidence for this model comes from experiments with young children who have developed categorical perception by approximately six months of age. This model is somewhat neutral in identifying perceptual primitives; however, there is a strong bias toward auditory information as it operates within auditory-acoustic theories of perception (e.g., Diehl & Kluender, 1989; Stevens & Blumstein, 1981). For instance, they propose that “babies’ early speech representations are entirely auditory, but that they very soon
involve visual, kinesthetic, and motoric elements” (Pickett, 1999, p. 250). Important to understanding the NLM model and the perceptual magnet effect is the concept of a prototype. A prototype, as narrowly defined by Kuhl and Iverson, is a “good instance of a category” (p. 123). Research under this framework has demonstrated that prototypes function as perceptual magnets. In other words, the perceived distance between a prototype and another member of a category appears to be reduced while this is not the case for non-prototypes. Thus, when an adult learns an L2, L1 prototypes will distort the acoustic space, resulting in potential difficulties in perceiving new sounds accurately. If the L2 contains a sound that is similar to a prototype in the L1, then it might be attracted to that prototype and assimilated into the category of the L1 sound. Similar to the SLM, difficulties with perception may lead to inaccuracies in production. While the NLM does not make explicit claims about production, we might infer from its connection to auditory-acoustic speech perception theories that perception and production systems are not directly linked. Nevertheless, within these theories, internal acoustic representations monitor articulatory output; thus, when a speaker produces an utterance, it is monitored by the representations established from perception. In this way, perception and production systems can be ultimately linked, but require some intermediate step(s). Also like the SLM, it is important to note that the NLM model is concerned with segments. It is unclear, however, what predictions this model would make in relation to the Korean acquisition of consonantal segments in different syllable positions because position sensitivity is never addressed. The model does not make any predictions concerning having a sound similar to a prototype in the L1 but disallowed in certain syllable structures in the L2.

The Perceptual Assimilation Model (PAM) has its roots in Direct Realism (Fowler, 1986; Best, 1995) and focuses on cross-language research (rather than L2 research). Unlike the SLM
and the NLM model, the primitives in this model are articulatory gestures, not acoustic properties of the speech signal. The PAM posits that non-native speech sounds will be perceived in relation to their articulatory similarities to and differences from native speech sounds. Non-native segments can be (1) assimilated into a native category (as either a good, acceptable, or noticeably deviant exemplar), (2) perceived as an uncategorizable speech sound, or (3) not perceived as a speech sound (Best, 1995, pp. 194-195). Because it has been primarily concerned with cross-language research, the PAM does not have much to say with regard to how the phonological system might change with increasing proficiency in the L2. However, Best notes that within a Direct Realist framework, learning continues into adulthood so it is possible that categories could shift.

This model, although again concerned with the acquisition of L2 segments and L2 segmental contrasts in relation to L1 segments, provides interesting implications for the difficulties that Korean L1 learners have with word-final palatals. When initially confronted with palatal codas in English, learners could potentially assimilate those sounds into a native category (as either good, acceptable, or bad exemplars) or treat them as uncategorizable. Disregarding syllable structure and looking at the general acquisition of English /sʃʧʤ/, the PAM might predict that Korean speakers assimilate instances of /sʃ/, but not of /ʃʤ/, into native categories because they exist in Korean. Therefore, we might expect more perceptual difficulties with /ʃʤ/. However, since it does not take syllable structure into account, the PAM does not predict that Korean speakers would necessarily have more difficulty with these sounds in coda position than in other positions. In addition, because of its roots in Direct Realism and the fact that articulatory gestures are taken as primitives, we might predict a strong connection between perceptual and
production systems. Thus, if we were to find improved accuracies in the perceptual domain, those should strongly correlate with production.

To summarize, the models described above make the following predictions: Within the SLM, we might predict that Korean learners of English will have an equally easy or difficult time acquiring /sʃʃʧʤ/ in coda position because none of these segments are allowed in coda position in the L1; the NLM does not make clear predictions about the acquisition of these segments; and the PAM might predict that learners would have more difficulties with /ʃʤ/, irrespective of position-sensitive information, because those segments do not exist in the L1. We can also make predictions about how learning, or improvement, in one skill (e.g., perception) would affect the other (e.g., production) based on the posited links between perceptual and production systems in the different models. The PAM, with its roots in Direct Realism, posits linked systems that share representations and would thus predict that perception and production learning would be strongly correlated. The SLM, on the other hand, makes strong claims about perception leading production. Thus, although perceptual and production systems are eventually linked over time, they do not necessarily share representations. We might posit that perception and production learning will be correlated, but under the SLM, we would not predict that it would be the case that L2 learners can accurately produce sounds that they cannot perceive accurately. Finally, the NLM does not make claims between how perceptual and production systems relate, but if perception and production systems are linked indirectly, as they are in auditory-acoustic theories, we might predict dissociations or at least weaker correlations between learning in either skill. Overall, if we are to understand Korean L2 learners’ difficulties with the production of final palatals, we need to investigate not only the relationship between their
accuracies in both perception and production, but also how learning in one skill (i.e., perception) affects learning in the other (i.e., production).

One piece of the puzzle missing from the previous three models is a thorough understanding of how the syllable structure constraints of a language might lead to difficulties in the perception and production of L2 sounds. The case for Korean learners of English seems to point to difficulties that cannot be easily explained by comparing the L2 segments being acquired to L1 segments. With this in mind, let us now turn to research on L2 learners’ production of syllable structure.

2.2 Research on the L2 Acquisition of Syllable Structure with a Focus on Production

This section provides a brief review of the literature on L2 learners’ production of syllable structure. Many of the studies discussed below adopt an Optimality Theory (OT) framework to account for non-target-like L2 productions that do not appear to be explicable as a simple case of transfer from the L1. Although the present research does not adopt such a framework, some important research on the production of L2 syllables has been conducted in this area in the past few decades and thus should be acknowledged in light of the current discussion. Other work (e.g., Archibald, 1998; Broselow & Park, 1995) provides a similar account to the one below in terms of defining what an L2 phonological grammar consists of, but approaches it from a structural perspective and does not adopt the OT formalization.

Broselow and Park (1995) investigated the syllable structure of L2 learners within moraic theory. Theirs was an attempt to understand the IL grammar of Korean L2 learners of English with regard to syllable structure and to explain why they epenthesized vowels in some words (e.g., beat and cheap) but not in others (e.g., bit and tip). Because the consonants in coda position
in each of these sets of words are the same, it could not simply be the case that these particular consonants trigger vowel epenthesis generally. Broselow and Park claimed that the representations stored by learners are different for the two sets of words because of the vowels in each (e.g., long [bimoraic] vs. short [monomoraic]). Thus, a Korean L2 learner of English would hear a word like *beat*, set up a representation that has a bimoraic structure and ultimately perform vowel epenthesis because long vowels are not allowed in Korean. Nevertheless, what remains unanswered and potentially problematic from this work is that Broselow and Park presume that learners perceive and thus establish a representation with a bimoraic structure, despite the fact that this is an illicit structure in Korean. What needs to be established first is what these learners perceive from the input, as it would guide the formation of representations.

Broselow, Chen, and Wang (1998) investigated the production of coda consonants of Mandarin L2 learners of English. Mandarin and English were chosen for comparison because English allows a variety of segments in coda position, but Mandarin allows only glides and the nasals /n η/. Their goal was to use constraints within OT to explain the varying simplification strategies employed by learners (e.g., vowel epenthesis, coda deletion, coda devoicing) when producing words in the L2. One pattern they noted for Mandarin speakers is the preference for vowel epenthesis with monosyllabic words in comparison to disyllabic words. As an example, the target nonce word /vig/ was more likely to be produced as /vi.gə/ than /vik/ or /vi/, whereas a target nonce word /filig/ was more likely to produce as /fi.li/ or /fi.lik/ than /fi.li.gə/. The authors attributed these findings to the word binarity constraint, which states that words should consist of two syllables. What is relevant from this study for the current research is that syllable structure constraints predict that L2 learners will make errors, and that differences between the L1 and L2 can account for the types of errors learners make in the L2. Thus, differences between Korean
and English syllable structure could account for the preference for epenthesizing vowels after palatals in coda position.

Other studies that have looked at syllable structure within an OT framework include Hancin-Bhatt and Bhatt (1997) and Hancin-Bhatt (2000). Hancin-Bhatt and Bhatt (1997) investigated the productions of complex onsets and codas by Japanese and Spanish L2 learners of English. They argued that OT provided good predictions for not only the types of errors that L2 learners make, but also the simplification strategies that they adopt. One finding from their study is that learners made more production errors with complex codas than with complex onsets. They explained this finding by pointing out that complex onsets are more often allowed across languages than complex codas. In addition, they found that vowel epenthesization was more likely to occur in complex onsets than in complex codas, but that deletion was more likely to occur in codas. Their explanation was that onsets are a privileged position where sounds are more likely to be maintained in comparison with codas. Finally, Hancin-Bhatt (2000) looked at both simple and complex coda productions of Thai L2 learners of English. She found that simple codas were easier for learners to produce than complex ones, but unlike Hancin-Bhatt and Bhatt (1997), that substitution was the most common strategy for simple codas and that.

The above findings can be summarized as follows: Differences between L1 and L2 syllable structures can result in (1) errors that display a preference for disyllabic words, (2) a greater number of errors with complex codas than with complex onsets, and (3) a greater occurrence of vowel epenthesization in complex onsets than codas. Although the above studies provide explanations within an OT framework, other possible explanations exist, for example, those that consider working memory and the perceptibility of onsets as compared to codas. It could be the case that L2 learners’ working memory limitations influence the type of production
strategy they use, such that they delete or devoice codas rather than epenthesizing a vowel in longer words. Relative perceptibility of onsets as compared to codas might also provide insights into the above findings. There is evidence from work investigating both adults (e.g., Kochetov, 2004; Redford & Diehl, 1999) and children (e.g., Jusczyk, Goodman, & Bauman, 1999; Zamuner, 2006) that perception is more difficult in coda position than in onset position. Hancin-Bhatt and Bhatt (1997) discuss the possibility that learners may be unable to perceive codas because they are more difficult to hear, and thus less perceptible, than onsets, leading them to delete the sound. If we consider the Korean-English case that is the focus of this dissertation, anecdotally, learners appear to be epenthesizing a vowel rather than deleting the palatal consonant. However, perceptibility could also explain this apparent asymmetry if the palatal fricatives and affricates that pose difficulties to Korean L2 learners of English were considered to be more readily perceptible (e.g., Redford & Diehl, 1999) than the obstruents in the Hancin-Bhatt and Bhatt (1997) study.

In addition to these alternative explanations, the major shortcoming of the work within OT is that a majority of it is heavily focused on the productions of L2 learners and does not take their perceptions into consideration. OT hypothesizes an “input” or underlying representation that feeds into the language system, but it is unclear how we can establish what the learners perceive as input and what exact underlying representation they are storing. Hancin-Bhatt (2000) acknowledged this as a concern for any OT study on acquisition. Ultimately, OT as a framework cannot provide a satisfactory explanation of the difficulties learners encounter in both the production and the perception of syllable structure. With this in mind, let us now turn to a discussion of the L2 perception research addressing syllable structure constraints and the perceptual illusion effect.
2.3 L2 Acquisition of Phonotactics and Syllable Structure, and the Perceptual Illusion Effect

In addition to posing difficulties in the production of L2 sounds, phonotactics and syllable structure constraints from the native language also influence the learners’ perception of L2 sounds. Before reviewing some of the literature on this issue, it is helpful to differentiate between phonotactic constraints and syllable structure. I will refer to phonotactic constraints as those constraints on permissible sequences of sounds in a language (e.g., irrespective of where these sequences of sounds are located in relation to syllable boundaries), whereas I will consider syllable structure as the organization of a syllable and permissible segments in certain syllabic positions. The importance of this distinction will become clear during the discussion of Kabak and Idsardi (2007) later in this section.

Dupoux, Kakehi, Hirose, Pallier, and Mehler (1999) investigated the role of phonotactic constraints in cross-language perception. The language groups in their study included both Japanese and French speakers. Japanese, like Korean, has a limited syllable inventory, allowing V, VV, CVN, and CVQ sequences (where Q is a geminate). French on the other hand, is more similar to English in that it allows a wider variety of syllable structures. The experiment used nonwords such as *ebuzo* in which the medial vowel was removed in a step-wise fashion, resulting in experimental items on a continuum of *ebzo* to *ebuzo*. Listeners were asked to indicate whether or not they heard a [u] vowel in the middle of each word. Findings demonstrated that in the stimuli with the vowel completely removed, French learners indicated

---

4 The medial vowel *u* was spliced out at zero-crossings and included five conditions: (1) with little or no vowel, (2) containing the two most extreme pitch periods of the vowel, (3) containing four pitch periods, (4) six pitch periods, (5) eight pitch periods. Two other conditions included a recording of *ebzo* and a version of the word with a medial vowel other than *u*.
hearing a vowel only approximately 10% of the time. In contrast, Japanese learners reported hearing a vowel more than 70% of the time. Thus, this study demonstrated that Japanese speakers perceived an “illusory” vowel inside these consonant clusters.

Matthews and Brown (2004) also discussed the perceptual illusion effect but extended Dupoux et al.’s (1999) work to the context of L2 learners. In their study, they compared the performance of Japanese and Thai L2 learners of English on their perception of clusters, also using nonsense words. They included Thai learners because, although Thai speakers had been reported as hearing an illusory vowel in some prosodic environments (Imsri, 1999), Thai’s syllable structure constraints allow the cluster sequences they were testing (unlike Japanese). They found that whereas Thai speakers performed at ceiling, Japanese learners had significantly lower accuracy rates and perceived illusory vowels. They argued that in cases of perceptual illusion, the intake a learner receives actually exceeds the input because they perceive an illusory vowel where none exists in the acoustic signal. This finding has consequences for the early stages of phonological processing and lexical storage. Production inaccuracies were previously thought to be the result of L1 phonological processes, but if learners initially perceive words with illusory vowels (and perhaps store these words with the illusory vowel), then they might not begin the production process with a target-like representation. Nevertheless, it is unclear whether consonantal contact or syllable structure violation was causing the illusory vowel effect in Japanese speakers; therefore, Kabak and Idsardi (2007) expanded on the Dupoux et al. (1999) study by attempting to answer this question with Korean learners of English.

The Kabak and Idsardi (2007) study included Korean L2 learners of English and looked at word-medial clusters. They were able to differentiate contexts of consonantal contact restrictions, or those “that ban the co-occurrence of certain heterosyllabic consonants” (p. 23),
from syllable structure restrictions, which do not allow certain segments in coda position. They tested two different sequences of word-medial English consonant clusters of the type VC₁C₂V: In one type, the C₁ was licit in coda position, but the sequence of C₁C₂ produced a contact not allowed in the language. In the other type, the C₁ was not licit in coda position. Results showed that Korean learners of English had trouble distinguishing only those instances where the consonant in the coda position was illicit. Based on this finding, they claim that syllable structure restrictions, and not consonantal contact violations, influence the perception of an illusory vowel.

The findings of the above studies suggest that Korean L2 learners of English not only demonstrate difficulties in the production of palatal codas, but also have difficulties in perceiving them accurately, and that L2 learners’ difficulties in production may be related to perceptual difficulties with segments within syllables that violate L1 syllable structure constraints. In other words, Korean learners of English may hear an illusory vowel when palatal consonants are in word-final position, which may then lead to their production inaccuracies. There are, of course, orthographic considerations to take into account. It might be the case that these L2 learners initially perceive a final vowel but that learning the spelling of the word (which would provide evidence that there is no vowel) helps them know that there is no vowel. But even if the learner ‘knows’ that there is no vowel, this does not mean that the learner no longer perceives one in the input. It is also possible that the learner could use this orthographic knowledge as input to compensate for the perceptual illusion, but the likelihood that this would lead them to restructure their representations is probably low. We know, for example, that this is not the case for Japanese learners of English, who have difficulty with the /u/-/l/ contrast (see e.g., Goto, 1971; Sheldon & Strange, 1982; Bradlow, Pisoni, Akahane-Yamada & Tohkura, 1997; Logan, Lively & Pisoni, 1991; Lively, Logan & Pisoni, 1993; among others), and the number of minimal pairs
with those sounds is much higher than the number of minimal pairs with the palatal codas.
Nevertheless, while not a focus of this research, possible orthographic influences should be kept in mind when discussing this issue.

Having now discussed models of speech perception and learners’ difficulties with both perception and production of segments within certain syllable structures, we can turn to a discussion of the research that has investigated the relationship between perception and production.

2.4 Relationship between L2 Perception and Production

The link between L2 perception and production has been of growing interest to researchers in the past decades, and a wide variety of research has been conducted in order to better understand the relationship between the two and how that relationship relates to the acquisition of L2 phonology. One of the basic questions asked is whether accurate perception is a necessary precursor to accurate production and whether (at least some) production errors are caused by perception errors.

Unfortunately, despite a growing body of research, it appears that a definitive answer to these questions remains unreachable for several reasons. First of all, trends in the relationship between perception and production vary with regard to the phenomenon studied. For example, research investigating the perception and production of vowels seems to indicate a close relationship between these two abilities in learners. For example, Bohn and Flege (1990) investigated the perception and production of the vowels /ɛ/ and /æ/ by German L2 learners of English with differing L2 experience. Experience was operationalized as time spent in an L2-speaking country. One group had lived in the US for at least five years (m=7.5 years) and the
inexperienced group had recently arrived to the US (m=0.6 year). A group of monolingual American English speakers was included as a control. Participants completed both a production task, which consisted of reading minimal pairs in a carrier phrase, and a perception task, which was a word-identification task using synthesized minimal pairs (e.g., bet-bat) on a vowel continuum. Results showed that experienced learners were able to produce the vowels in a native-like way, as assessed by comparing spectral and durational measurements to those of native speakers, while the inexperienced learners’ vowels differed from native speakers’ productions both spectrally and in terms of duration. For perception, results showed that spectral cues were more important for native speakers than for experienced learners and in turn more important for experienced learners than for inexperienced learners. The reverse was true for duration cues; inexperienced learners relied most heavily on this cue. Thus, even though experienced learners produced the vowels in a native-like way, their perception was different from native speakers.

Bohn and Flege also compared individual perception and production behavior with regard to both spectral and duration cues. They found that native speakers uniformly relied on spectral cues in perception, but they showed a varying degree of magnitude of use of these cues in production. Inexperienced learners produced little to no spectral differences, but were variable in their use of it as a perceptual cue. Finally, experienced learners who relied greatly on that cue in perception had a greater magnitude of variation in production, while those that did not use that cue in perception showed little to no variation in production. Thus, Bohn and Flege concluded that spectral cues in perception and production are independent of each other for these learners. They found a similar lack of relationship between the perception and production of duration cues. However, Bohn and Flege did provide evidence that extended experience with the
language influenced the types of cues learners use to differentiate the vowel pairs /e/ and /æ/ and that in the acquisition of L2 vowels, experience seems to have more of an effect on production than on perception.

Rochet (1995) explored the perception and production of the high vowel /y/ by Canadian English and Brazilian Portuguese L2 learners of French. Using an imitation task, he demonstrated that Portuguese L2 learners of French more often produced /y/ as an /i/ or /i/-like vowel, whereas English speakers more often produced /y/ as a /u/ or /u/-like vowel. In the perception task, participants categorized synthetic stimuli that differed in terms of their F2 value, and those categorizations were compared to the perceptions of native French speakers. French speakers categorized the vowels in the F2 range of 1300-1900 Hz as being /y/ vowels. In comparison, English speakers more often categorized vowels in this range as /u/ and Portuguese speakers more often as /i/. Rochet thus claimed that the parallels between production and perception for the two groups of learners provided evidence that the “accented” speech of these learners might be perceptually motivated.

Thus, in terms of research on vowels, we find learners in all four categories: 1) accurate productions and perceptions, 2) inaccurate productions and perceptions or 3) accurate perceptions but inaccurate productions and 4) accurate productions but inaccurate perceptions. However, those in the last category are not as common. Research on the perception and production of vowels seems to point to a close relationship between these abilities in learners. By contrast, studies of the perception and production of consonants (although mostly restricted to stop and liquids) point to a less clear relationship.

A large body of this research has investigated the acquisition of the English /ɹ/-/l/ contrast. Goto (1971), who investigated the perception and production of Japanese L2 learners of
English, showed that accurate production of these liquids could precede accurate perception. Sheldon and Strange (1982) provided additional evidence for this finding in their extension of Goto’s work. Although the overall accuracies for participants in their study were higher than those in Goto’s, the same pattern of having higher production accuracies than perception ones was maintained.

Research investigating the link between perception and production with stops has shown similar results to those with liquids in that high accuracy in productions do not always indicate high accuracies in perception. Flege and Eefting (1987) compared the production and perceptions of the stop consonants /t/ and /d/ by Dutch L2 learners of English. They showed that Dutch learners of English were able to produce VOT differences for these sounds in Dutch and English, but that their perception of these differences in a continuum of synthetic stimuli did not change when the language setting of the experiment did.  

Thus, Flege and Eefting concluded that the relationship between perception and production for these sounds was not as clear as, for example, the learning of vocalic differences.

In a study that incorporated a perceptual training component, Bradlow, Pisoni, Akahane-Yamada, and Tohkura (1997) showed that the production of /ɹ/ and /l/ could be improved through perceptual training, although variations in the amount of improvement existed (this study is discussed in more detail in the section on phonetic training). Therefore, when we consider phenomena beyond vowels, we find more of a tendency for learners to fit into the category of those with accurate productions but inaccurate perceptions. At first glance, this might seem counterintuitive—how could learners produce something accurately when they cannot hear it? One possibility that has been suggested for /ɹ/ and /l/ in the literature is that learners receive

---

5 In order to control for ‘language context,’ Flege and Eefting divided the experiment into two ‘language sets’. When investigating English, all materials and instructions were in English; when investigating Dutch, all materials and instructions were in Dutch.
articulatory training and rely on the kinetic sensation that occurs when two articulators make contact (which does not occur in the production of vowels) to utter accurate productions. This would be aided by the fact that learners would know which consonant to produce based on orthography, which is relatively straightforward in English for /ɹ/ and /l/. Nevertheless, these explanations would not be equally as relevant for all consonants, depending on their regularity in orthography.

More recently, De Jong, Hao, and Park (2009) investigated the relationship between the perception and production of Korean L2 learners of English with regard to their acquisition of the consonants /p, b, t, d, f, v, θ, ð/. They argued that while perception and production systems are connected, the units of acquisition for perception and production are not the same: Acquisition in perception seems to involve features while acquisition in production seems to involve gestures and their coordination, at least for learners at some proficiency levels.

Overall, while many researchers have attempted to understand the link between perception and production in L2 learning, many questions remain unanswered. The work in this dissertation contributes to this discussion in important ways. First, it adds to the growing body of literature related to perception and production by investigating relatively understudied consonant sounds: fricatives and affricates. In addition, this dissertation goes beyond segmental information and takes syllable structure into account by investigating the acquisition of consonants in an existing, but restricted syllable structure in Korean (codas). Will we also find learners whose accuracies/inaccuracies in perception and production follow the patterns found for the consonants /ɹ/ and /l/ (i.e., who demonstrate more accurate production before perception), or will production accuracies be more closely tied to perception? Ultimately, in order to fully answer questions related to the relationship between perception and production, we must investigate how
learning in one skill (e.g., perception) affects learning in the other (e.g., production). One method for doing this involves perceptual phonetic training, which is the focus of the final experiment presented in this dissertation. The next section reviews some of the literature on this type of training component that will shape the design of the training put forth in this dissertation.

2.5 High-Variability Phonetic Training

The type of training upon which the perceptual training in this dissertation is based is called high-variability phonetic training (HVPT). It entails perceptually training learners with multiple words from multiple talkers (typically between 4-6 talkers, as opposed to one). The idea is that L2 learners who are exposed to this type of training are able to establish more robust categories and thus are able to generalize learning to new words and new talkers. In their widely cited studies, Logan, Lively and Pisoni (1991) and Lively, Logan and Pisoni (1993) put forth a highly effective perceptual training method that improved on previous methods in a number of important ways. Motivation for these studies stems from prior work failing to show generalizability of training to novel stimuli and novel talkers. The segments in question were /ɹ/ and /l/ and participants were Japanese L2 learners of English. The general organization of the method includes a pre-test/post-test design with three weeks of perceptual training in between. The pre-test, perceptual training, and post-test all used a forced-choice word-identification task rather than an AX task. In an AX task, listeners hear two words and are asked to determine whether the second word (X) is the same as the first word (A). In a forced-choice word-identification task, learners hear a stimulus (from a minimal pair), are visually presented with two possible choices, and asked to choose which of the two written words they heard. Logan et al. (1991) claimed the forced-choice word-identification task is preferable for at least two
reasons. First, the task forces learners to develop phonetic memory codes rather than simply relying on information in sensory memory, which learners can do if they perform an AX task. Second, the forced-choice word-identification task encourages learners to classify stimuli into categories. Another improvement was the use of natural stimuli produced by five talkers. Natural stimuli are preferable to synthetic stimuli because they contain more variation. Using more than one talker also provides increased variability, which allows learners to form more robust phonetic categories. The actual perceptual training in each study consisted of three weeks (15 days) of approximately 30-40 minute sessions. On each day of training, listeners would hear a total of 242 stimuli from one talker in the forced-choice word-identification task. If a response was correct, the next stimulus was presented. If a response was incorrect, a light on the answer box would become lit and the listeners would hear the stimulus again. Finally, both studies also included generalization tasks, which were similar to the pre- and post-tests but contained novel words produced both by a familiar talker and an unfamiliar one.

Results from Logan et al. showed that training learners with phonetically variable natural stimuli using a forced-choice word-identification task resulted in significant improvements by learners. In addition, unlike previous studies, learners were able to generalize to both novel words and novel talkers. Lively et al. (1993) extended these results by comparing learners trained with only one talker to those trained with five talkers. Not only did learners trained with one talker fail to generalize to novel words produced by a new talker, but they also did not do as well with novel words produced by a familiar talker. This result highlights the importance of being trained with multiple talkers for robust category formation to take place.

Before turning our attention to work that has investigated the effects of HVPT on production, let us first consider the underlying mechanisms that might account for why this type
of training results in generalizability. When considering the establishment of representations, two theories dominate the field: episodic-trace theories and abstractionist theories. Within an episodic-trace perspective (e.g., Goldinger, 1998; Goldinger & Azuma, 2004; Johnson, 1997; Jacoby, 1983), learners store exemplars in memory with detailed phonetic and non-linguistic information. Adopting this perspective would posit that receiving multiple tokens of input from multiple talkers would result in a greater number of exemplars stored in memory, thus resulting in more robust or stable representations. Nevertheless, within this theory, we might question why having a greater number of exemplars stored in memory would lead to abstraction, enabling learners to listeners to generalize to new words and new talkers. We might also conclude that if episodic-trace theories are correct, we would find better performance for trained words within the training experiment because these are the exemplars that would be stored. Alternatively, abstractionist theories (e.g., Bowers, 2000; Bowers & Michita, 1998; Norris, McQueen & Cutler, 2003) posit abstract prelexical representations that mediate recognition. Receiving varied input from multiple talkers would strengthen these abstract representations and allow for generalization from this input to new words and new talkers. While it is not the purpose of this research to investigate which, if either, of the two theories above is correct, we see that both provide potential explanations for why the HVPT might result in perceptual learning. We now turn our attention to the effects of HVPT on production.

In a continuation of Logan, Lively and Pisoni (1991) and Lively, Logan and Pisoni (1993), Bradlow et al. (1997) investigated the effects of perceptual training on the production accuracies of /ɹ/ and /l/ for Japanese L2 learners of English. A similar pre-test/post-test design with 3–4 weeks of perceptual training in between was employed to determine if training improved L2 learners’ perception and production of these sounds. A comparison of trained
learners (experimental group) and untrained learners (control group) indicated that the experimental group’s perception scores were significantly higher the control group’s perception scores. In order to measure production accuracies, participants’ productions from before and after training were presented to a group of native listeners (NLs) in two separate tasks. In the first, NLs heard both productions and were asked to rate on a seven-point scale “which version was a clearer and more intelligible version of the word presented on the screen” (p. 2303). In the second, NLs were presented with single instances of participants’ productions and chose which one they heard in a forced-choice word-identification task. Bradlow et al. then compared these production scores to the perception scores and examined whether perceptual training had a positive effect on production accuracy. They report, however, that individual differences in production gains were substantial and that no clear correlation could be seen between gains in perceptual accuracy and gains in production accuracy. In other words, a participant with high gains in perceptual accuracy did not necessarily have high gains in production accuracy, and a participant with lower gains in perceptual accuracy did not necessarily have lower gains in production accuracy. They concluded that while perception and production are clearly linked, the exact details of the relationship are unclear and perception accuracy is not a sufficient condition for production accuracy.

Let us return to our discussion of speech perception theories and their predictions on how learning in one skill would affect learning in the other. Recall that the PAM, with its roots in Direct Realism, posits linked systems that share representations. As such, it would predict that perception and production systems would have a direct relationship. Thus, even though the PAM is ultimately a perception theory, it would posit that perception and production learning would be strongly correlated. The SLM makes strong claims about perception leading production.
However, it still posits an indirect relationship between modalities, in that perceptual learning allows for a reorganization of the acoustic-auditory space that ultimately feeds the system used for both perception and production. Because it assumes a psychoacoustic view of speech perception, the SLM would not predict a direct relationship between perception and production, but rather an indirect one. Therefore, unlike the PAM, we would not necessarily expect to find that perception and production learning are strongly correlated. Finally, we hypothesized that the NLM might demonstrate weaker correlations between modalities because they are not directly linked within auditory-acoustic speech perception theories. While all three of these models can account for the findings of Bradlow et al., as the authors point out, none can account for the individual differences reported with regard to varying degrees of improvement in each skill. One explanation they offer for this is to suggest that the “specific motor commands necessary for improved /ɹ-/l/ production may be acquired at different rates for different subjects. This suggests that modification of an underlying perceptuomotor, phonetic representation is not sufficient on its own to result in corresponding modifications in speech production.” (p. 2308) With regard to the present research investigating the acquisition of palatal codas, results will allow us to determine whether the type of perceptual training that was beneficial for the perception of /ɹ-/l/ will show similar benefits for palatal segments in a restricted syllable structure. In addition, finding whether the perceptual training has an effect on the production of palatal codas will allow us to determine whether the results from Bradlow et al., which showed no direct link between gains in perception and gains in production, were unique to the acquisition of /ɹ-/l/ or whether they indicate a more general trend in the acquisition of L2 phonology.

The methods put forth in the Logan et al., Lively et al., and Bradlow et al. studies guide the design of the training component set forth in the final experiment presented in this
dissertation. It is not the goal of this dissertation to compare the benefits of high-variability phonetic training to, for example, low-variability phonetic training, but rather to use the methods found to be beneficial in high-variability phonetic training (e.g., using natural stimuli, training on multiple talkers) as a guide for designing perceptual phonetic training materials for palatal codas. Before stating the research questions and predictions that guide the research in this dissertation, I present a developmental path that learners might take in their acquisition of English codas.

2.6 Developmental Path in the Acquisition of Syllable Structure

In his 2003 study, Abrahamsson discussed a developmental path for the acquisition of codas by true beginner Chinese L2 learners of Swedish. In order to contextualize this study, it is helpful to know that Chinese learners in this study came from language backgrounds only allowing one or two segments in coda position (the velar nasal /ŋ/ and the dental nasal /n/). Swedish, on the other hand, appears to have no limit to the number of consonants allowed in coda position, although the practical cut-off seems to be somewhere around five segments (and the only consonants that are not allowed are /h/ and /ç/). Abrahamsson’s study focused on how Chinese learners of Swedish perform vowel epenthesis and consonant deletion as modifications of codas. Using data collected from interviews, he investigated whether Chinese learners would follow a particular developmental sequence, specifically coda deletion > vowel epenthesis > closed syllables (p. 341), and whether they would do so in a U-shaped curve. His principal motivation for positing a u-shaped development is the fact that, in early stages of development, learners would be saying grammatically or structurally simple (e.g., one-word) utterances and might avoid difficult consonant clusters, producing words with fewer codas. However, as L2 learners’ interlanguages become more complex and they focus greater attention to fluency, their
overall accuracy rates would decrease. Finally, at a more advanced stage of learning, they would again reach higher accuracy rates. His motivations for positing the developmental sequence stem from the recoverability principle. Abrahamsson explains that within a functional approach of phonology, learners simultaneously attempt to maximize intelligibility while minimizing articulatory effort. In order to maximize intelligibility, listeners may attempt to keep as much information in the uttered form of a word as possible. If we consider the two forms of coda simplification that are the focus of Abrahamsson’s study, we see that vowel epenthesis provides more information to a listener than coda deletion. This is because in the case of coda deletion, the surface form retains no information with regard to the deleted consonants, whereas with vowel epenthesis, segmental information about the coda consonant is available. Nevertheless, in order for a learner to perform vowel epenthesis, he or she must have the phonetic ability to do so. Thus, Abrahamsson predicts a greater overall proportion of epenthesis as proficiency increases. In light of the current research, this would mean that beginning learners of Korean would be more likely to delete palatal codas, but as their proficiency increases, they would be more likely to epenthesize a vowel after it, and finally, they would attain native-like production.

When attempting to incorporate these hypotheses into the Korean palatal problem, one thing to notice immediately is the seeming lack of consonant deletion as a coda simplification strategy. One possible explanation for why Korean learners of English might use vowel epenthesis and not coda deletion could be related to the fact that Korean does allow some consonants in coda position, but just not palatals. This would follow the predictions of Flege’s (1995) SLM. Another possibility is that Korean learners of English may epenthesize a vowel because they initially perceived a vowel in the input for palatal codas. Alternatively, it could be that the Korean L2 learners of English in this research have already passed the proficiency level
at which they would delete codas. Finding evidence for such development would be in line with Abrahamsson’s (2003) predicted development. The experiments reported on in this dissertation will shed further light on which of these accounts is correct.

2.7 Research Questions and Predictions

The overall goals of this dissertation are to investigate: (1) syllable structure and how it may filter perception in second language acquisition; (2) the relationship between perception and production in the acquisition of a second language phonological system; (3) the effects of perceptual phonetic training on the perception and production of palatal codas; and (4) the IL system of Korean learners of English with regard to their acquisition of palatal codas. Questions and predictions related to each of these issues are presented in the following subsections and are addressed in the remaining chapters of this dissertation.

2.7.1 Syllable Structure

We know from previous research that the perception of L2 segmentals can be influenced by the segmental categories that exist from the L1. Nevertheless, we do not know much about how the syllable structure constraints of an L1 filter perception. However, recall the perceptual illusion studies, which have shown that syllable structure, not consonantal contact, was at the root of perceptual illusions. Furthermore, if syllable structure constraints filter perception from the outset of acquisition of an L2, we do not know how they are eventually modified to allow the learner to learn the accurate and relevant information from the L2 phonology that has been filtered out. One goal of this dissertation is to answer these questions. Korean L2 learners of English provide an interesting test case because of the differences between English and Korean:
Both languages allow codas, but Korean has a restricted inventory of segments that can appear in coda position. Thus, we can ask the following questions: In a syllable structure that is restricted (codas), is perception equally affected for segments that exist in other syllable positions in the L1 as it is for those that do not exist?

Based on previous literature, we can make the following predictions. Within the PAM, we can predict that perceptions will be better with segments that exist in the L1 than with those that do not. For the questions posed in this research, this would mean that Korean L2 learners of English have more difficulty with /ʃ ʤ/ than with /s ʧ/ in coda position or, for that matter, elsewhere in the word (recall that the PAM does not consider position as an important factor in its predictions). On the other hand, the SLM states that segments are perceived as a function of their position in a word. Therefore, the SLM might predict that learners will have an equally easy or difficult time perceiving /s ʃ ʧ ʤ/ in coda position because none of these segments are allowed in coda position in the L1. Finally, we might predict that affricates will be more difficult than fricatives in the event that learners perceive them as two segments rather than one, but because previous literature has not addressed this issue, this remains an empirical question. It might also be the case that affricates are more difficult than fricatives because of the dual alveolar and palatal places of articulation of affricate palatals, which could possibly result in these segments being articulatorily and acoustically more complex. These predictions are summarized in Table 1. These questions will be addressed in Experiment 1, reported in Chapter 3.
Table 1: Predictions for the Perception of /sʃʧʤ/ in Codas by Korean L2 Learners of English

<table>
<thead>
<tr>
<th></th>
<th>SLM</th>
<th>PAM</th>
<th>Fricatives vs. Affricates</th>
</tr>
</thead>
<tbody>
<tr>
<td>/s/</td>
<td>easier</td>
<td>easier</td>
<td></td>
</tr>
<tr>
<td>/ʃ/</td>
<td>similarly difficult</td>
<td>more difficult</td>
<td></td>
</tr>
<tr>
<td>/ʧ/</td>
<td>easier</td>
<td>easier</td>
<td>more difficult</td>
</tr>
<tr>
<td>/ʤ/</td>
<td>more difficult</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

2.7.2 Relationship between Perception and Production

Previous research investigating the relationship between perception and production has yielded complicated results with regard to whether or not accurate perception precedes accurate production or vice versa. It has shown some indications of a link between perception and production systems, but has yet to provide evidence for a direct link between these two systems. While the type of segment (i.e., vowel vs. consonant) seems to influence results, little has been done in relation to the role of syllable structure. The second goal of this dissertation is to investigate the relationship between perception and production with regard to both syllable structure constraints and segments in the categories of palatals. I ask the following questions: Is there a direct relationship between perception and production accuracies? In other words, is there co-variation between accuracies in perception and production? And do improved accuracies in perception lead directly to improved accuracies in production?

Based on the literature related to the perception and production of liquids and stops, it is unclear whether we can predict that there will be co-variation between accuracies in perception and production for final palatals. Regardless of the pattern we find, we will not be able to draw
direct conclusions about the relationship between perception and production systems unless we attempt to change L2 learners’ perception or production. In other words, what we need is evidence that improvements in one skill (e.g., perception) can directly affect the other skill (e.g., production). This is the focus of the next subsection.

With regard to the final question above, which asks whether improved accuracies in perception lead directly to improved accuracies in production, we can make the following predictions. The PAM, with its roots in Direct Realism, posits linked systems that share representations and would thus predict that perception and production systems would have a direct relationship. Therefore, the PAM would posit that perception and production learning would be strongly correlated. Because it assumes a psychoacoustic view of speech perception, the SLM would not predict a direct relationship between perception and production, but rather an indirect one. Therefore, unlike the PAM, we would not necessarily expect to find that perception and production learning are strongly correlated.

2.7.3 Effects of Perceptual Phonetic Training on Perception and Production of Palatal Codas

The third focus of this dissertation is related to the effects of perceptual phonetic training on the perception and production of palatal codas. The goal is not only to determine whether perceptual training results in positive gains of both perception and production of palatal codas, but also to design materials that can be pedagogically viable and realistically implemented by teachers and/or used by students. In other words, the time commitment and implementation decisions should reflect practical classroom considerations. We also want to determine whether perceptual phonetic training will demonstrate similar results in terms of generalizability for these
structures as it has for other contrasts such as the /ɹ/-/l/ contrast in English. Thus, I ask the following questions: Can pedagogically viable perceptual phonetic training on palatal codas improve perception accuracies of palatal codas? What, if any, will be the effects of perceptual training on productions of palatal codas? Do learners’ improvements generalize to new words and new talkers?

Based on the previous literature reviewed in this dissertation, which focused on /ɹ/ and /l/, we might predict that perceptual phonetic training will improve both the perception and production of palatal codas and allow for generalizability. Nevertheless, as the segments investigated in this dissertation are being acquired in a syllable structure that is restricted in the L1 of learners, it is possible that we will find a different trend. Finally, as discussed above, the results related to whether improvements in perception lead to direct improvements in production will provide a better understanding of the relationship between those two systems.

### 2.7.4 The Developing Interlanguage System of Korean L2 Learners of English with Regard to Palatal Codas

Finally, it is necessary to have a more systematic understanding of the IL system of Korean learners of English in relation to palatal codas in order to know what contexts are most difficult for learners and should be a focus in a pronunciation classroom. Thus, I ask the following questions: In which contexts (words in isolation, words within a larger discourse, final singleton palatals, final palatal clusters, palatals before –ed morphemes, etc.) do learners have the most difficulty with palatals in production?
2.8 Summary

Overall, the questions posed in this dissertation fall into two larger categories: The first investigates the relationship of the perception and production of learners at different proficiency levels with regard to syllable structure constraints. The second examines perceptual phonetic training and how the perception and production of palatal codas might be changed as a result of that training. Chapters 3 and 4 report on findings from two experiments designed to answer questions related to the perception and production of learners at different proficiency levels with regard to syllable structure constraints. Chapter 5 reports on findings from a perceptual phonetic training experiment that focuses on whether, and if so, how, perceptual training can affect learners’ perception and production of palatal codas. Chapter 6 presents a general discussion and summary of the findings and concludes the dissertation.
CHAPTER 3
EXPERIMENT 1: PERCEPTION AND PRODUCTION OF LEARNERS AT DIFFERENT PROFICIENCY LEVELS

Experiment 1 investigates the effect of syllable structure constraints on the perception and production of Korean L2 learners of English. It compares the perception and production of palatals in coda position in isolated words by L2 learners at varying proficiency levels in order to gain preliminary information about that relationship for different levels of learners. The specific questions addressed in this chapter are:

1. In a syllable structure that is restricted in the L1 (codas), is perception equally affected for segments that exist in other syllable positions in the L1 as it is for those that do not exist?
2. Does the type of segment (fricative or affricate) influence perception?
3. Is there a direct relationship between perception and production accuracies? In other words, is there co-variation between accuracies in perception and production?
4. How does proficiency level play a role, if at all, in the above?

3.1 Participants

Eight native speakers (NSs) of English (5 men and 3 women) and 19 Korean L2 learners of English who were either enrolled at the University of Illinois or the Intensive English Institute participated in Experiment 1. The L2 learners were divided into two proficiency groups (8 high-proficiency, 2 men and 6 women; 11 mid-proficiency, 5 men and 6 women) based on their performance on a cloze test (Brown, 1980; see Appendix A). A cloze test was chosen as a global proficiency measure rather than a more specific test related to the learners’ oral language
proficiency in order to avoid circularity in the interpretation of the results: If L2 learners were grouped based on their oral language proficiency and then results indicated a significant difference between them in terms of their performance on the perception and production tests, then one could argue that the oral proficiency measurement was circular with the experiment. A cloze test is a sufficiently global measure of proficiency (for discussion, see Tremblay, 2011), and it has the advantage of not being circular with the object of this study: L2 learners’ perception and production of sounds. More specifically, proficiency groups were determined by performing a hierarchical cluster analysis on the cloze test scores of L2 learners using Ward’s Method to determine group (Kaufman & Rousseeuw, 1990). The same grouping outcome was found by performing a k-means cluster analysis (Tremblay, 2011) presupposing two groups.

Participants also completed a language background questionnaire (see Appendix B) to gather information about their age of first exposure to English, years of English instruction, years spent in an English immersion context, and so forth. Table 2 shows the participants’ cloze test scores, as well as the means and ranges for a subset of relevant language background information.
Table 2: Language Background Information, Experiment 1

<table>
<thead>
<tr>
<th></th>
<th>Cloze Test (out of 50)</th>
<th>Daily % Usage</th>
<th>1st Exposure to English (years)</th>
<th>Years in Immersion Context (years)</th>
<th>Years of Instruction (years)</th>
<th>Age</th>
</tr>
</thead>
<tbody>
<tr>
<td>NSs (n=8)</td>
<td>Mean 48</td>
<td>96</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td>SD 0.7</td>
<td>5.1</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>6.2</td>
</tr>
<tr>
<td></td>
<td>Range 47-49</td>
<td>90-100</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>20-40</td>
</tr>
<tr>
<td>High-Level (n=8)</td>
<td>Mean 42</td>
<td>56</td>
<td>7</td>
<td>7</td>
<td>12</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td>SD 3.2</td>
<td>31.5</td>
<td>4.5</td>
<td>4.3</td>
<td>7.8</td>
<td>3.4</td>
</tr>
<tr>
<td></td>
<td>Range 37-47</td>
<td>10-97</td>
<td>1-12</td>
<td>2.4-16</td>
<td>1-21</td>
<td>19-29</td>
</tr>
<tr>
<td>Mid-Level (n=11)</td>
<td>Mean 31</td>
<td>39</td>
<td>12</td>
<td>4</td>
<td>10</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td>SD 2.7</td>
<td>22.5</td>
<td>2.8</td>
<td>3.5</td>
<td>4.9</td>
<td>4.2</td>
</tr>
<tr>
<td></td>
<td>Range 25-35</td>
<td>10-80</td>
<td>5-15</td>
<td>0.75-13.5</td>
<td>6-22</td>
<td>27-40</td>
</tr>
</tbody>
</table>

3.2 Materials: Perception and Production Experiments

All participants completed an AXB perception experiment and a read-aloud production experiment. The stimuli to investigate the perception of coda consonants consisted of one- and two-syllable English words. Real words, as opposed to nonce words, were chosen in Experiment 1 in an effort to approximate the real language that these learners would perceive and produce. Real words also have the advantage of having more ecological validity. Thus, as this was the first
experiment conducted, it was decided to begin with real words.\(^6\) Stimuli generally conformed to
the following sequences: (1) \(C_1V_1C_2\) (e.g. *push*) and (2) \(C_1V_1C_2+/i/\) (e.g. *pushy*).\(^7\) \(C_2\) represents
one of four test sounds (/s ʃ ʧ ʤ/), which do not occur in coda position in Korean, as well as a
control sound (/n/), which is permitted in coda position in Korean (see Appendix C for a
complete list of stimuli). In choosing the four test consonants, the following were included: (a)
two sounds representing phonemes that exist in Korean (/s ʧ/) and two that do not (/ʃ ʤ/) in order
to determine if that had any effect on perception accuracies; and (b) two fricatives and two
affricates in order to determine if the type of consonant affected perception accuracies. The
condition of whether or not the phoneme exists in Korean was included because of the
predictions made by both the SLM and PAM. The SLM takes positional considerations into
effect, and because Korean allows none of these segments in coda position, they should be
equally easy or difficult for Korean learners of English. On the other hand, the PAM, which does
not take positional considerations into account, would predict that having the sound in the L1
would facilitate acquisition. The decision to test fricatives and affricates was made for two
reasons: First of all, it is not clear whether Korean learners of English would treat affricates as
one segment or a series of two segments. If it is the case that Korean learners of English treat
affricates as a series of two consonants rather than as one segment, we might predict more
difficulties because of the complex coda in which these would result. The fricative and affricate
conditions were also included for the practical reason of having a 2X2 experimental design. The
final consonant /n/ was included because Korean allows /n/ in final position; thus, it was
hypothesized that participants would have no difficulty in hearing the difference between word

\(^6\) However, because of their limited availability as well as the differences in frequency with which the different
words occur, nonce words were added to the stimuli in Experiment 3, described in Chapter 5.

\(^7\) Of the 60 experimental word pairs, 56 conformed to this pattern. The remaining four also included a pre-palatal
consonant /n l ɹ/. In Experiment 3, reported on in Chapter 5, all word pairs conformed to the CVC/CVC+/i/ pattern
and none had complex codas.
pairs like *fun-funny*. Each coda consonant condition included 12 items, for a total of 60 experimental items. The perception experiment also included 97 fillers that focused on sounds unrelated to the palatal coda test conditions, for a total of 157 items. The production task included all of the words from the coda experimental conditions—both those containing a vowel and those without (*n*=120).

The stimuli for the perception experiment were produced by three female native speakers of American English. Table 3 presents biographical information for the talkers who produced the stimuli. As we can see from Table 3, the three talkers are from the Inland North (Labov, Ash, Boberg, 2006) and had all been living in Illinois for at least 3.5 years at the time of recording. Stimuli were recorded in a sound-attenuated booth at the University of Illinois at Urbana-Champaign via a Marantz PMD570 solid state recorder using an AKG c520 head-mounted microphone at 44.1 kHz. After recording, stimuli were normalized to 65dB using Praat (Boersma & Weenink, 2010).8

<table>
<thead>
<tr>
<th>Talker</th>
<th>Age</th>
<th>Hometown</th>
<th>Length of Residence in IL (in years)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>29</td>
<td>Upstate NY</td>
<td>3.5</td>
</tr>
<tr>
<td>2</td>
<td>45</td>
<td>Northeast PA</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>28</td>
<td>Northern IL</td>
<td>28</td>
</tr>
</tbody>
</table>

Table 3: Biographical Information for the Three Female Talkers who Produced Stimuli

---

8 I would like to thank Chris Carignan for his help in developing and modifying the scripts that I used to automate processes for segmenting and normalizing data in Praat.
3.3 Procedure

Experiment 1 included four tasks: (1) a language background questionnaire; (2) the cloze test; (3) an AXB perception task; and (4) a read-aloud production task. The procedures for the perception and production tasks are described in the following subsections. Participants always completed the perception and production tasks in that order, but varied as to when they completed the cloze test and language background questionnaire.

3.3.1 Perception Experiment

Stimuli were presented using E-Prime\(^9\) following an AXB discrimination design. Participants heard three stimuli in a row and decided whether the second stimulus (X) is the same as the first (A) or third (B). For example, if a participant heard the sequence *push – push – pushy*, they should choose A because the second word (*push*) was the same as the first word. The interstimulus interval (ISI) was 1.5 seconds. A relatively long inter-stimulus interval was chosen so that participants could not simply rely on acoustic information, but would access categorical information from long-term memory (e.g., Pisoni, 1973; Werker & Tees, 1984). One of the benefits of choosing an AXB perception task is that word familiarity effects should not matter because AXB results are computed over both monosyllabic and disyllabic words. Instructions were given to participants explaining the AXB task, and then the participants began with 10 practice items, receiving feedback after each, in order to ensure that they understood the procedure (see Appendix D for complete instructions). With the practice items, the experiment contained a total of 167 items. The task took approximately 15-20 minutes to complete.

---

\(^9\) E-Prime is a computer software application for experiment design, data collection and analysis. For more information see Schneider, Eschman, and Zuccolotto (2001a, 2001b).
Four lists of stimuli were created, balancing conditions in across lists (i.e., whether X was A or B and whether A or B contained a word-final vowel) so that no participant heard the same word pair in more than one condition. In addition, stimuli from a given talker were always presented in the same position. In other words, A tokens were the first talker’s stimuli, X tokens the second talker’s stimuli, and B tokens the third talker’s stimuli. Test items were pseudo-randomized in E-Prime and responses were recorded as accurate or inaccurate.

3.3.2 Production Experiment

The production experiment took place after the perception experiment in order to avoid participants guessing the focus of the study (as it included only experimental items and no fillers). For the purposes of this experiment, all productions were coded as either accurate or inaccurate with respect to the final C/CV syllable by two trained English pronunciation teachers. In other words, productions of orthographically monosyllabic words were rated as accurate if they contained no epenthetic vowel after the coda, and productions of orthographically disyllabic words were rated as accurate if they contained a word-final vowel. Any other errors in pronunciation (e.g., substituting /p/ for /f/ in a word like fish) were ignored in the data analysis. A high inter-rater reliability coefficient was found between the two codings (r=.956, p<.001).

3.3.3 Data Analysis

Answers on the perception tests were scored as either accurate or inaccurate. Participants’ results on the perception task were then transformed into $d^\prime$ scores. Calculating $d^\prime$ scores is a method used within Signal Detection Theory to provide a measure of listeners’ sensitivity. The $d^\prime$ calculation is done by converting proportions of hits (H) (i.e., identifying A as A) and false
alarms (F) (i.e., identifying B as A) into z-scores under a normal distribution: \( d' = z(H) - z(F) \).

In Experiment 1, \( d' \) scores were calculated to control for the potential bias of selecting the first word (A) or the third word (B) following the hits and false alarms presented in Table 4 (Macmillan & Creelman, 1991).

### Table 4: Explanation of \( d' \) Scoring

<table>
<thead>
<tr>
<th></th>
<th>Hit: A = A</th>
<th>False Alarm: B = A</th>
</tr>
</thead>
<tbody>
<tr>
<td>Miss: B = A</td>
<td>Correct Rejection: B = B</td>
<td></td>
</tr>
</tbody>
</table>

Percent accuracy scores were also calculated from the AXB perception data. The percent accuracy results mirror the \( d' \) scores (see Appendix E).

### 3.4 Results

#### 3.4.1 Perception Experiment

First, results from the perception task are reported. Figure 1 shows the \( d' \) scores of the native, high, and mid groups by type of consonant (fricative or affricate). For these results, a \( d' \) score of 0 represents no sensitivity and a \( d' \) score of 3.50 represents perfect sensitivity.
A mixed-design repeated-measures ANOVA was performed on the $d'$ scores with type of consonant (fricative, affricate) and existence of the segment as a phoneme in Korean (yes, no) as within-subject variables, and with proficiency (native, high, mid) as between-subject variable. Type of consonant had a significant effect, $F(1,24)=7.89, p<.05$. As can be seen from the results by type, affricates appeared to be more difficult overall.

There was no effect of existence of phoneme ($F<1$) and no interaction between type of phoneme and existence of phoneme ($F<1$). Figure 2 shows the $d'$ scores of the native, high, and mid groups by existence of the segment as a phoneme in Korean (yes, no).
Proficiency had again a significant effect, $F(2,24)=7.60, p<.01$. Post-hoc Tukey tests showed a significant difference ($p<.05$) between the mid-proficiency group and the other two groups, but no difference between the native and high-proficiency groups ($p>0.1$). Hence, these results suggest that lower-level L2 learners had more difficulty perceiving palatal codas than higher-level L2 learners and native speakers.

Recall that the consonant /s/ was included in the design to test for differences between affricates and fricatives as well as to complete a 2X2 design. However, it was not hypothesized that learners would have difficulties perceiving these sounds based on the original observation that Korean L2 learners of English epenthesize a vowel after final palatals. It could be the case that the significant result found between affricates and fricatives above is driven by the fact that the affricate category contains two palatals, but the fricative category contains one palatal and one non-palatal. Because palatals and non-palatals are not balanced in this experiment, we will
not compare them directly. However, we can consider the individual segments. Figure 3 displays the $d'$ scores of all groups separated by consonant type.

![Figure 3: Perception accuracies separated by consonant for all proficiency levels reported as $d'$ scores](image)

When we conduct a mixed-design repeated-measures ANOVA with consonant type (/s, f, θ, dʒ/) as within-subject variable and proficiency (native, high, mid) as between-subject variable, we find significant main effects of consonant, $F(3,72)=2.97, p<.05$, and of group, $F(2,24)=7.60, p<.01$, but no interaction between consonant and group $F(6,72)=1.42, p<.221$. As mentioned above, affricates may have been more difficult to perceive than fricatives if, for example, the status of /s/ as a non-palatal somehow made perception easier. In order to test the fricative vs. affricate question, let us then compare only /ʃ/ and /θ/. In this way, we can directly ask the question of whether palatal affricates are more difficult than palatal fricatives without the
potentially confounding factor of /s/. We should, however, keep in mind that this comparison contains relatively few items in comparison to the previous analysis.

A mixed-design repeated-measures ANOVA was performed on the $d'$ scores with type of consonant (/ʃ ʧ/) as within-subject variable, and with proficiency (native, high, mid) as between-subject variable. There was a significant main effect of proficiency, $F(2,24)=4.58$, $p<.05$, but there was no effect of type of consonant, $F(1,24)=3.95$, $p<.058$, and no interaction between type of consonant type and proficiency $F(2,24)=1.03$, $p<.371$. Taken together, these results suggest that /s/ is behaving differently from the palatals and driving the finding that affricates are more difficult than fricatives. In other words, we do not find evidence for the existence of the phoneme in Korean affecting perception accuracies, nor do we find strong evidence that consonant type (fricative, affricate) affects perception accuracies.

### 3.4.2 Production Experiment

Here, the results from the production experiment are presented. Figure 4 shows the coda accuracy of all proficiency groups by type of consonant (fricative, affricate).
Figure 4: Production results of all groups for fricatives and affricates

A mixed-design repeated-measures ANOVA was performed on the accuracy rates with type of consonant (fricative, affricate) and existence of the segment as a phoneme in Korean (yes, no) as within-subjects variables, and proficiency (native, high, mid) as between-subject variable. Consonant type had a significant effect, $F(1,24)=17.68$, $p<.001$. As can be seen from the results by consonant type, affricates appeared to be more difficult overall.

Existence of the consonant in Korean also had a significant effect, $F(1,24)=11.68$, $p<.01$. Figure 5 shows the coda accuracy of all proficiency groups by existence of the segment as a phoneme in Korean (yes, no). As can be seen from the results, segments that do not exist as phonemes in Korean appeared to be more difficult to produce.
There were also interactions between type and existence, $F(1,24)=19.52, p<.001$, type and proficiency $F(2,24)=8.98, p<.001$, existence and proficiency, $F(2,24)=5.03, p<.05$, and type and existence and proficiency, $F(2,24)=20.22, p<.001$. Proficiency also had a significant effect, $F(2,24)=15.50, p<.001$. We can see from the data that the NS group is at ceiling. Given the three-way interaction, repeated-measures ANOVAs are conducted separately for each learner group with alpha levels adjusted to .025. A repeated-measures ANOVA was performed on the high proficiency group’s accuracy rates, with type of consonant (fricative, affricate) and existence of the segment as a phoneme in Korean (yes, no) as within-subjects variables. Consonant type did not have a significant effect, $F(1,7)=2.22, p<.180$, nor did existence, $F(1,7)=1.61, p<.245$. There was also no interaction between consonant type and existence ($F<1$). A similar repeated-measures ANOVA was also performed on the mid proficiency group’s accuracy rates, with type of consonant (fricative, affricate) and existence of the segment as a phoneme in Korean (yes, no) as within-subjects variables. Consonant type had a significant effect, $F(1,10)=24.78, p<.001$, and
existence had a significant main effect, $F(1,10)=17.11, p<.01$. There was also an interaction between consonant type and existence, $F(1,10)=47.65, p<.001$. Given this significant interaction, we tested for the effect of existence separately for affricates and fricatives in the mid-proficiency group’s data, with the alpha level being further adjusted to 0.0125. Paired-samples t-tests showed no significant difference between sounds that exist and sounds that do not exist for affricates, $t(10)=.697, p<.501$, but significant differences between sounds that exist and sounds that do not exist for fricatives, $t(10)=5.78, p<.001$. Thus, mid-proficiency learners performed significantly better on /s/ than /ʃ/, but they performed similarly on /ʃ/ and /ʤ/. 

As in the perception results, we will also consider accuracies separately for each consonant. Figure 6 displays the production accuracy rates of all proficiency groups separated by consonant type. Results are reported as the average percent correct productions for both monosyllabic words (e.g., push) and disyllabic words ending in [i] (e.g., pushy) combined, by coda type for each group.
When considering the results for the coda context, we see a trend: high-proficiency learners appear to be more accurate than mid-proficiency learners. A mixed-design repeated-measures ANOVA with consonant (/sʃʧʤ/) as within-subject variable and proficiency (native, high, mid) as between-subject variable shows main effects of consonant, $F(3,72)=14.63$, $p<.001$, and proficiency, $F(2,24)=15.50$, $p<.001$. There was also an interaction between consonant and proficiency, $F(6,72)=8.28$, $p<.001$. Again, the native speakers were at ceiling on all the conditions. Given the significant interaction, one-way ANOVAs with proficiency (native, high, mid) as between-subject variable were conducted separately for each consonant, with alpha levels adjusted to .0125. There were main effects of proficiency for the consonants /ʃ/, $F(2,26)=12.02$, $p<.001$, /ʧ/, $F(2,26)=21.15$, $p<.001$, and /ʤ/, $F(2,26)=12.53$, $p<.001$, but not /s/, $F(2,26)=3.13$, $p<.062$. Post-hoc Tukey tests showed a significant difference between the mid-proficiency group and the other two groups for the consonants /ʃʧʤ/ ($p<.01$), but not /s/, ($p>0.1$). These tests did not show significant differences between the native and high-proficiency groups for any consonant ($p>0.1$). In summary, native speakers and high-proficiency learners performed significantly better than mid-proficiency learners on all consonants but /s/, and there were no differences found between native speakers and high-proficiency learners for any consonant.

As with the results for perception, we might also consider an analysis that compares /ʃ/ and /ʧ/. When we conduct a mixed-design repeated-measures ANOVA with consonant type (/ʃʧ/) as within-subject variable and proficiency (native, high, mid) as between-subject variable, we no longer find a main effect of consonant type ($F<1$). The main effect of proficiency remains significant, $F(2,24)=17.61$, $p<.001$. As was the case for perception, this is further evidence that
the result identifying affricates as more difficult than fricatives is the product of /s/ behaving differently from palatals.

If we consider the mid-proficiency learners’ production error rates, we see that they are at approximately 30% for the palatal segments in coda position. However, if we separate their productions of monosyllabic words with coda consonants (e.g., fish) and their productions of disyllabic words ending in [i] (e.g., fishy), as shown in Figure 7, we see that the majority of errors are not errors of epenthesis, but rather omissions of the final [i] on disyllabic words like fishy.

![Figure 7: Mid-proficiency palatal production accuracies by word type](image)

As we can see from Figure 7, mid-proficiency learners perform quite well with the voiceless palatals /ʧ/ and /ʃ/ in coda position (having accuracies above 95%); however, their performance with /ʤ/ is only 71% accurate. While learners’ performance with this segment is at 66% in the disyllabic condition, for the other palatals, their production performance in the disyllabic condition is only at 45%, indicating that the L2 learners do not produce the vowel
when they should. It can also be noted that standard deviations for /dʒ/ in final position and all the palatals in the disyllabic position show relatively high variability among learners. A repeated-measures ANOVA was performed with two within-subject factors: word type (monosyllabic, disyllabic) and consonant (/ʃ dʒʧ/). This analysis revealed a main effect of word type,
\[ F(1,10)=14.30, \quad p<.01 \], and an interaction between word type and consonant, \[ F(2,20)=8.61, \quad p<.01 \]. Post-hoc Bonferroni comparisons were conducted with alpha levels adjusted to \( p<.017 \).
Paired-samples t-tests showed no significant difference between mono- and disyllabic words for /dʒ/, \( t(10)=.344, \quad p<.738 \), but significant differences between mono- and disyllabic words for /ʃ/, \( t(10)=4.98, \quad p<.001 \), and /ʧ/, \( t(10)=5.79, \quad p<.001 \). Thus, mid-proficiency learners performed significantly better with monosyllabic words for /ʃ,ʧ/, but not for /dʒ/.

The finding that learners demonstrate more errors in the disyllabic condition contrasts with the impressionistic observation in the classroom that learner’s epenthesize a vowel after final palatals. One potential explanation may be related to the existence of voiceless vowels in Korean. More specifically, [i] has been shown to be devoiced to varying degrees after voiceless consonants in Korean (e.g., Kim, Hirose, & Niimi, 1992; Jun & Beckman, 1993; Jun, Beckman, & Lee, 1998). The variability of the vowel devoicing ranges from voiced vowels, partially devoiced vowels, to completely devoiced vowels, and appears to be a phonetic feature caused by gestural overlap rather than the result of a phonological rule. Furthermore, while not much is known about this feature, it is the case that vowel devoicing can occur only when [i] is preceded by a voiceless consonant. This feature is relevant to the current discussion, because we are considering the existence of epenthetic [i] following both voiced and voiceless palatals in English.
Another possibility is that of overcorrection. It could be the case that learners are over-correcting in some instances and thus not producing final [i] vowels on disyllabic words like *pushy*. If Korean L2 learners of English are aware of their tendency to epenthesize a vowel after final palatals, they might attempt to avoid this error, even in cases where a final [i] is indicated in the orthography. In any case, further investigation into this issue is warranted and is presented in the following sections.

Mid-proficiency learners might be producing a voiceless vowel after final voiceless palatals whether or not there should be one, but this could not explain their production of an [i] after /ʤ/ because it is a voiced consonant, which they produce as voiced. Because of the syllable structures allowed in Korean and the existence of voiceless vowels, when a learner encounters voiceless palatals followed by a vowel in English (in a word like *pushy*), the orthography indicates there is a vowel, but they produce none, which suggests they are devoicing that vowel. In other words, it could be the case that learners produce words with voiceless palatals like *pushy* with a voiceless [i], which would be perceived by English listeners/raters as *push*. What remains unclear is whether, when learners successfully produce codas (in a word like *push*), they are in fact producing a vowel and devoicing it. In these cases, the current analysis of the production data would not necessarily reflect this because they were rated by two trained pronunciation teachers for accuracy and not subjected to an acoustic analysis. In the case of full vowel devoicing, an acoustic analysis would also not be informative if we simply considered what follows the palatal consonant. Nevertheless, we can perform a visual analysis of mono- and disyllabic word pairs for the presence/absence of a voiced vowel and compare their word lengths, which might provide insight into this question. The analysis was conducted on the

---

10 In fact, an analysis of disyllabic items scored as incorrect indeed demonstrated no periodicity in waveforms or indications of voicing in the spectrogram (as shown by pulses in Praat).
mono- and disyllabic /ʃ/ and /ʧ/ words produced by six randomly selected mid-proficiency learners. For these learners, we want to know whether words with a final voiceless palatal in the orthography (e.g., *push*) are produced acoustically differently from words with a voiceless palatal+[/i/] in the orthography (e.g., *pushy*). If L2 learners produce voiceless vowels after words like *push* and *pushy*, we would expect their acoustic realization of monosyllabic words to be similar to their acoustic realization of disyllabic words.

When comparing mono- and disyllabic word pairs for the presence/absence of a voiced vowel, four possible production patterns could be found: (A) mono- and disyllabic words could both appear as CVC; (B) monosyllabic words could appear as CVC and disyllabic words could appear as CVCV; (C) monosyllabic words could appear as CVCV and disyllabic words could appear as CVC; and (D) mono- and disyllabic words could both appear as CVCV.

For each learner, mono- and disyllabic words were visually compared and categorized into one of the four previous production patterns. Figures 8 and 9 are examples of a learner producing a word pair in Category A with no visible vowel following the palatal consonant in either word. The x-axis in all figures represents time (in seconds).
Figure 8: Spectrogram and waveform for the production of *rash* with no visible final vowel

Figure 9: Spectrogram and waveform for the production of *rashy* with no visible final vowel
Figures 10 and 11 are examples of a learner producing a word pair in Category B with no visible vowel following the palatal consonant in the monosyllabic word, but one present in the disyllabic word.

Figure 10: Spectrogram and waveform for the production of *ash* with no visible final vowel
Figure 11: Spectrogram and waveform for the production of *ASHY* with visible final vowel

Figures 12 and 13 are examples of a learner producing a word pair in Category C with a visible vowel following the palatal consonant in the monosyllabic word, but not in the disyllabic word.
Figure 12: Spectrogram and waveform for the production of *mush* with visible final vowel

Figure 13: Spectrogram and waveform for the production of *mushy* with no visible final vowel
Figures 14 and 15 are examples of a learner producing a word pair in Category D with a visible vowel following the palatal consonant in both words.

Figure 14: Spectrogram and waveform for the production of *twitch* with visible final vowel
Figure 15: Spectrogram and waveform for the production of *twitchy* with visible final vowel

Table 5 provides the percentage of word pairs produced in each Category for each of the six learners (24 pairs were analyzed for each learner) along with their overall error score from the production task.
As we can see in Table 5, a comparison of mono- and disyllabic words shows a trend for categories A and B, representing 53.5% and 41.7% of the pairs, respectively. Categories C and D only comprise 4.9% of the data (or seven pairs). Now that we see a trend for production patterns, let us consider what these categories could represent. Category A, in which both mono- and disyllabic words appear as CVC, might represent both mono- and disyllabic words being produced as CVC¥ (where ¥ represents a voiceless vowel) or monosyllabic words being produced as CVC and disyllabic words being produced as CVC¥. We do not expect that learners produce disyllabic words as CVC for several reasons. First, learners see the vowel in the orthography, thus they know that the vowel should be present. Second, producing the CVC¥ word does not violate L1 syllable structure constraints. Therefore, we would not predict that

### Table 5: Percentage of Productions by Category Type for Learners

<table>
<thead>
<tr>
<th>Learner</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>Task Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learner 7</td>
<td>58.3%</td>
<td>37.5%</td>
<td>0%</td>
<td>4.2%</td>
<td>66.7%</td>
</tr>
<tr>
<td>Learner 8</td>
<td>50.0%</td>
<td>33.3%</td>
<td>8.3%</td>
<td>8.3%</td>
<td>56.9%</td>
</tr>
<tr>
<td>Learner 18</td>
<td>45.8%</td>
<td>54.2%</td>
<td>0%</td>
<td>0%</td>
<td>77.8%</td>
</tr>
<tr>
<td>Learner 19</td>
<td>87.5%</td>
<td>12.5%</td>
<td>0%</td>
<td>0%</td>
<td>55.6%</td>
</tr>
<tr>
<td>Learner 23</td>
<td>41.7%</td>
<td>54.2%</td>
<td>0%</td>
<td>4.2%</td>
<td>76.4%</td>
</tr>
<tr>
<td>Learner 24</td>
<td>37.5%</td>
<td>58.3%</td>
<td>0%</td>
<td>4.2%</td>
<td>75.0%</td>
</tr>
</tbody>
</table>

| Mean (SD) | 53.5% (18.1%) | 41.7% (17.5%) | 1.4% (3.4%) | 3.5% (3.1%) | Production Task Score |

As we can see in Table 5, a comparison of mono- and disyllabic words shows a trend for categories A and B, representing 53.5% and 41.7% of the pairs, respectively. Categories C and D only comprise 4.9% of the data (or seven pairs). Now that we see a trend for production patterns, let us consider what these categories could represent. Category A, in which both mono- and disyllabic words appear as CVC, might represent both mono- and disyllabic words being produced as CVC¥ (where ¥ represents a voiceless vowel) or monosyllabic words being produced as CVC and disyllabic words being produced as CVC¥. We do not expect that learners produce disyllabic words as CVC for several reasons. First, learners see the vowel in the orthography, thus they know that the vowel should be present. Second, producing the CVC¥ word does not violate L1 syllable structure constraints. Therefore, we would not predict that
leaners would produce disyllabic words as CVC. Based on these possibilities, we would make different predictions in terms of word lengths: If mono- and disyllabic words are both produced as CVC, their word lengths should be similar. If monosyllabic words are produced as CVC and disyllabic words are produced as CVC, we would predict their word lengths would differ.

Category B, in which monosyllabic words appear as CVC and disyllabic words appear as CVC, might represent monosyllabic words being produced as CVC and disyllabic words being produced as CVC, or the native-like pattern in which monosyllabic words are produced as CVC and disyllabic words are produced as CVCV. Similarly to Category A, when we consider the possibilities, we would make different predictions in terms of word lengths: If monosyllabic words are produced as CVC and disyllabic words are produced as CVC, we predict that their word lengths would differ. If monosyllabic words are produced as CVC and disyllabic words are produced as CVC, we would predict their word lengths would be similar.

Table 6 summarizes the predictions for Categories A and B.
Table 6: Possible Learner Production Patterns of Mono- and Disyllabic Palatal Words and Categories for the Comparison Analysis

<table>
<thead>
<tr>
<th>Pattern</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monosyllabic</td>
<td>CVC</td>
<td>CVCY</td>
<td>CVCY</td>
<td>CVC</td>
</tr>
<tr>
<td>Disyllabic</td>
<td>CVCY</td>
<td>CVCY</td>
<td>CVCV</td>
<td>CVCV</td>
</tr>
<tr>
<td>Word length duration comparison</td>
<td>different</td>
<td>similar</td>
<td>similar</td>
<td>different</td>
</tr>
<tr>
<td>Category</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>B</td>
</tr>
</tbody>
</table>

A length analysis was conducted on Category A and B words to determine whether there are significant differences between the productions of these words. The number of items found in Categories C (n=2) and D (n=5) is not enough to conduct an analysis. Productions were segmented and labeled for the entire word length, and the average durations of the mono- and disyllabic words were compared.

First, we consider productions from Category A, in which both mono- and disyllabic words contained no voiced vowel. Figure 16 below presents the average word length durations of the Category A productions of mono- and disyllabic words for the learners. The y-axis represents duration in milliseconds (ms).
Figure 16: Average word length (in milliseconds) of mono- and disyllabic words in Category A

A paired-samples $t$-test showed significant differences between the word lengths of mono- and disyllabic words in Category A, $t(5)=3.11$, $p<.027$. Learners are producing mono- and disyllabic words with consistently different durations; thus, we have evidence supporting Pattern 1 from Table 6. Pattern 1 represents the possibility of native-like production in which words like *push* are produced as CVC and words like *pushy* are produced as CVCV.

Figure 17 below presents the average word length durations of the Category B productions of mono- and disyllabic words for the learners. The y-axis represents duration in milliseconds (ms).
A paired-samples $t$-test showed significant differences between the word lengths of mono- and disyllabic words, $t(5) = -3.63$, $p < .015$. Learners are producing mono- and disyllabic words with consistently different durations; thus, we have evidence supporting Pattern 4 from Table 6. Pattern 4 represents the possibility of native-like production in which words like *push* are produced as CVC and words like *pushy* are produced as CVCV.

By conducting a visual comparison of productions and comparing the durations of mono- and disyllabic words, we have evidence supporting the possibility that learners are producing voiceless vowels after orthographically disyllabic words. We do not, however, have evidence to support the possibility that learners are producing voiceless vowels after orthographically monosyllabic words (e.g., *push*).

Finally, it should be noted that although L2 learners are far more accurate at not producing an epenthetic vowel after voiceless palatals (as compared to /dʒ/), this could be related
to the fact that the target words were produced in a relatively easy context: in isolation. It might be the case that having these words in larger contexts will cause more difficulties for learners. Therefore, a sentential context is included in Experiment 3, presented in Chapter 5.

3.4.3 Comparing Perception and Production

Although data patterned similarly on the perception and production tasks, the relative difficulty level of each task is unknown. Therefore, the relationship between perception and production is examined via correlations. Comparing individuals’ perception and production scores to determine whether there is co-variation will determine whether there is a relationship between perception and production.

Table 7 shows the perception scores reported as $d'$ and production scores reported in percent accuracy for each learner. For these results, a $d'$ score of 0 represents no sensitivity and a $d'$ score of 3.76 represents perfect sensitivity. The participants’ proficiency level is listed in the column on the far right.
Table 7: All Learners’ Perception and Production Scores

<table>
<thead>
<tr>
<th>Perception Accuracy</th>
<th>Production Accuracy</th>
<th>Proficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.76</td>
<td>100%</td>
<td>High</td>
</tr>
<tr>
<td>3.76</td>
<td>100%</td>
<td>High</td>
</tr>
<tr>
<td>3.76</td>
<td>100%</td>
<td>High</td>
</tr>
<tr>
<td>3.76</td>
<td>100%</td>
<td>High</td>
</tr>
<tr>
<td>3.76</td>
<td>100%</td>
<td>High</td>
</tr>
<tr>
<td>3.76</td>
<td>100%</td>
<td>High</td>
</tr>
<tr>
<td>3.76</td>
<td>100%</td>
<td>High</td>
</tr>
<tr>
<td>2.04</td>
<td>75.8%</td>
<td>High</td>
</tr>
<tr>
<td>3.76</td>
<td>95.0%</td>
<td>High</td>
</tr>
<tr>
<td>3.25</td>
<td>66.7%</td>
<td>Mid</td>
</tr>
<tr>
<td>2.44</td>
<td>56.9%</td>
<td>Mid</td>
</tr>
<tr>
<td>3.25</td>
<td>77.8%</td>
<td>Mid</td>
</tr>
<tr>
<td>0.57</td>
<td>55.6%</td>
<td>Mid</td>
</tr>
<tr>
<td>1.93</td>
<td>76.4%</td>
<td>Mid</td>
</tr>
<tr>
<td>1.52</td>
<td>75.0%</td>
<td>Mid</td>
</tr>
<tr>
<td>2.22</td>
<td>52.8%</td>
<td>Mid</td>
</tr>
<tr>
<td>3.25</td>
<td>81.9%</td>
<td>Mid</td>
</tr>
<tr>
<td>1.74</td>
<td>81.9%</td>
<td>Mid</td>
</tr>
<tr>
<td>3.25</td>
<td>47.2%</td>
<td>Mid</td>
</tr>
<tr>
<td>3.76</td>
<td>95.8%</td>
<td>Mid</td>
</tr>
</tbody>
</table>
As can be seen in Table 7, all but one high-proficiency learner had perception scores at ceiling and one mid-proficiency learner had a perception score at ceiling. This finding suggests that some learners are able to eventually perceive palatal codas in English in a native-like way. Because it is difficult to examine the relationship between perception and production for participants who performed at ceiling on either the perception or the production task, all remaining analyses will focus on L2 learners whose perception and production scores were not at ceiling. Figure 18 plots learners’ perception and production scores for those who were not at ceiling.

![Figure 18: Scatterplot comparing learners’ perception in d’ and production accuracies in percent accuracy for palatal codas](image)

A correlation test was run to investigate the relationship between perception and production accuracies of learners not at ceiling. This test revealed no correlation between perception and production accuracy, \( r = .038, p < .912 \). Thus, it appears that for learners,
perception accuracy is not directly related to production accuracy. I return to this in the discussion section.

3.5 Discussion

We set out in this chapter to answer the following research questions. Here I consider each in turn.

1. In a syllable structure that is restricted in the L1 (codas), is perception equally affected for segments that exist in the L1 in other positions as it is for those that do not exist?
2. Does the type of segment (fricative vs. affricate) influence perception?
3. Is there a direct relationship between perception and production accuracies? In other words, is there co-variation between accuracies in perception and production?
4. How does proficiency level play a role, if at all, in the above?

Recall the predictions we made at the end of chapter two (see Table 1 in section 2.7.1).

3.5.1 In a syllable structure that is restricted in the L1 (codas), is perception equally affected for segments that exist in other positions in the L1 as it is for those that do not exist? Does the type of segment (fricative vs. affricate) influence perception?

In order to test whether existence of a phoneme in an L1 affected its perception in a restricted syllable structure in that L1 (codas), we compared the perceptions of /ʃ s/ and /dʒʃ/ in coda position. Experiment 1 provided preliminary evidence that for Korean learners of English, the existence of the phoneme in the L1 did not have an effect on perception. To test whether type of segment (fricative vs. affricate) affected perception in a restricted syllable structure in that L1
(codas), we compared the perceptions of /ʃ s/ and /dʒ ʧ/ in coda position. Initial results provided some evidence for affricates being more difficult to perceive than fricatives. However, it was noted that the categories of fricative and affricate were not balanced for the palatal/non-palatal distinction. When comparing the perceptions of /ʃ/ and /ʧ/, no significant differences in perception accuracy rates were found. We originally predicted that affricates might be more difficult than fricatives if they were perceived as two segments rather than one, or because of the dual alveolar and palatal places of articulation which could possibly result in these segments being articulatorily and acoustically more complex. The findings from Experiment 1 did not support these predictions.

High-proficiency learners performed relatively well on all segments tested in this experiment with regard to perception. In comparison, mid-proficiency learners demonstrated some difficulties with the perception of palatal segments, but not with the perception of /s/. If we return to our earlier predictions and consider the results of the mid-proficiency group, we can see that learners did not have an equally difficult time with all consonants in coda position, thus the SLM is not supported. Recall, however, that even child learners of English show different rates of acquisition for /s/ compared to palatals. Palatals are typically acquired by age 4, whereas /s/ is typically acquired by age 3. It could be the case that /s/ has acoustic and articulatory properties that make it easier to perceive and produce than palatals, but this is an issue that goes beyond the scope of this dissertation. Learners also did not follow the predictions of the PAM with sounds existing in their L1 being easier to perceive than sounds that are not. We did, however, find the interesting result that existence of the phoneme in the L1 was significant in production. If we compare Figures 4 and 6, which represent the perceptions and productions of participants for each consonant, we see that for perception, high-proficiency learners are at ceiling for all
consonants except /ʃ/. For production, we note that while /s/ is still at ceiling, /ʃ dʒʃ/ are not. Again it could be the case that /s/ behaving differently than palatals is driving this result. While it appears that high-proficiency learners have mastered perception of the segments /dʒʃ/, they still have difficulties in their productions.

Mid-proficiency learners demonstrated difficulties with both perception and production of palatals, indicating that L1 syllable structure constraints are playing a role in perception. While it appears to be the case that syllable structure constraints are playing a role in the perception of palatal segments, we can note that perception results for the palatals were relatively high overall: all but one learner scored between 80%-100%, and eight learners were at ceiling. Thus, it appears that the AXB perception task in Experiment 1 was quite easy for most learners. If we are able to increase the difficulty level of the task, we might have a clearer picture of how learners perceive these words. Therefore, Experiments 2 and 3, reported on in Chapters 4 and 5, respectively, adopt a more difficult task: a forced-choice word-identification task.

### 3.5.2 Is there a direct relationship between perception and production accuracies? In other words, is there co-variation between accuracies in perception and production?

The findings from Experiment 1 did not demonstrate a direct relationship between perception and production accuracies in that these accuracies did not co-vary. These findings are in line with much of the previous research investigating the relationship between perception and production, which has failed to find a direct link between these systems. Despite showing no clear correlation between perception and production accuracies, looking solely at the steady-state results of learners limits how much we can say about the link between the perception and production of palatal codas. If we want to have a better understanding of the relationship
between these two systems, we must employ perceptual training to determine what effects, if any, learning in the perceptual domain has on production. If we find that productions improve with perceptual training, we will have more evidence of the link between these systems. In addition, we will be able to determine whether perceptual training allows for generalizability in learning for palatal codas. These questions are addressed in Experiment 3 presented in Chapter 5.

3.5.3 How does proficiency level play a role, if at all, in the above?

With regard to the perception of palatal codas, results indicated that high-proficiency learners pattern with native speakers, and both groups perform significantly better than mid-proficiency learners. When we considered the production of final palatals, results indicated that high-proficiency learners performed significantly better on all palatals than the mid-proficiency group. These results suggest that high-proficiency learners in this experiment have acquired final palatals while mid-proficiency learners have not. In addition, mid-proficiency learners demonstrated significantly more errors in producing disyllabic words than monosyllabic words (see Figure 7), although this was only the case with /ʃ/ and /ʧ/ words and not /ʤ/ words. Because of this finding, a duration analysis was performed on the average word lengths of mono- and disyllabic words containing voiceless palatals for a subset of mid-proficiency learners. The results indicated support for hypothesis that mid-proficiency learners were producing voiceless vowels after palatal consonants in orthographically disyllabic words, but not those of orthographically monosyllabic words. These findings contribute to a better understanding of the developing interlanguage system of Korean L2 learners of English regarding palatal codas.
3.6 Impetus for Experiment 2

One final concern to address with Experiment 1 is the fact that overall accuracy rates were high. NSs and all but two high-proficiency learners performed at ceiling on the perception task, and accuracy rates for the mid-proficiency learners were relatively high. This raised the question as to whether natural tokens of words with full [i] vowels (e.g., pushy) were representative of the perceptual illusion Korean listeners might have when hearing word-final palatals. It could be the case that learners (as well as NSs) were using other information to guide their decisions on the perception task. We already know from the duration analysis above that the learners’ word lengths differed between mono- and disyllabic words. We might expect that NSs’ productions in the stimuli also contained differences (such as differences in stem vowel length, palatal consonant length, f0 patterns related to stress, etc.) that could have provided additional cues beyond the presence or absence of a final [i] vowel. These cues might have aided learners in making accurate perceptual decisions. We know, for example, that monosyllabic words have longer vowels than disyllabic words (Klatt, 1973; Lehiste, 1972). Thus, we can predict that the lengths of the stem vowels in the monosyllabic and disyllabic words of the talkers who produced the stimuli are not the same. Figure 19 presents the average stem vowel length of mono- and disyllabic words for each of the palatal condition words separated by talker. The y-axis represents the length of the stem vowel in milliseconds (ms).
Figure 19: Stem vowel length comparison between mono- and disyllabic words for the three talkers who produced stimuli

It appears that the average stem vowel length is longer for the monosyllabic words than the disyllabic ones. A repeated-measures ANOVA with word type (monosyllabic, disyllabic) as within-subject variable shows a main effect of word type, $F(1,2)=40.20$, $p<.024$. NSs are producing stem vowels in monosyllabic words with durations significantly longer than those of disyllabic words. Thus, participants could be using this cue, rather than (or in addition to) the following vowel, for determining whether words are the same or not. For this reason, Experiment 2 was designed to begin to investigate the possible confounding factor stem information may be contributing to perception.
CHAPTER 4
EXPERIMENT 2: NATIVE SPEAKER AND LEARNER SENSITIVITY TO STEM VOWEL AND FINAL VOWEL LENGTH

Experiment 1 provided preliminary evidence that L1 syllable structure constraints play a role in learners’ perceptions of codas. While we found that learners had more difficulty with palatal codas than with /s/, we also saw that the overall accuracy rates of the perception of final palatals were relatively high. There were also differences in the stimuli between the lengths of the stem vowels in mono- and disyllabic words like *fish* and *fishy*, potentially cueing the L2 learners to the correct answer in the AXB task. Therefore, Experiment 2 was designed to tease apart the relative contributions of stem vowel length and final vowel length. The specific research questions of Experiment 2 are as follows:

1. Do native speakers of English show sensitivity to the presence or absence of a word-final vowel in words like *fish*/*fishy* if the length of the stem vowel is controlled for?
2. Do Korean L2 learners of English show sensitivity to the presence or absence of a word-final vowel in words like *fish*/*fishy* if the length of the stem vowel is controlled for? Does proficiency affect results?

4.1 Participants

Twenty-four NSs (11 men and 13 women) and 15 L2 learners who did not participate in Experiment 1 participated in Experiment 2. As in Experiment 1, L2 learners were divided into two proficiency groups (8 high, 4 men and 4 women; 7 mid, 5 men and 2 women) based on their performance on a cloze test (Brown, 1980; see Appendix A). The same proficiency group categorization scores that were used as in Experiment 1 were used in Experiment 2 for
consistency. Participants also completed a language background questionnaire (see Appendix B) to gather information about their age of first exposure to English, years of English instruction, years spent in an English immersion context, and so forth. Table 8 shows the participants’ cloze test scores, as well as the means and ranges for a subset of relevant language background information.

### Table 8: Language Background Information, Experiment 2

<table>
<thead>
<tr>
<th></th>
<th>Cloze Test ( /50)</th>
<th>Daily % Usage</th>
<th>1st Exposure to English (years)</th>
<th>Years in Immersion Context</th>
<th>Years of Instruction</th>
<th>Age</th>
</tr>
</thead>
<tbody>
<tr>
<td>NSs (n=24)</td>
<td>Mean n/a 95 n/a n/a n/a n/a 21</td>
<td>SD n/a 5.3 n/a n/a n/a 3.1</td>
<td>Range n/a 80-100 n/a n/a n/a 18-32</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High Level (n=8)</td>
<td>Mean 40 66 7 5 13 21</td>
<td>SD 2.5 27.6 2.7 4.2 3.4 3.7</td>
<td>Range 37-44 35-95 4-11 0.17-15 8-17 19-29</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mid Level (n=7)</td>
<td>Mean 25 40 12 3 10 35</td>
<td>SD 6.0 24.0 3.2 2.0 7.1 7.4</td>
<td>Range 15-30 2-75 6-17 1-6 4-25 24-44</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### 4.2 Materials

All participants completed a forced-choice word-identification experiment and a vowel-detection experiment. Learners also completed a read-aloud production experiment (native
speakers did not complete the production experiment as all native speakers in Experiment 1 were at ceiling for production ratings and it was predicted the same would be true for this group).

Experimental stimuli included the experimental items from Experiment 1 of the form $C_1V_1C_2$ (e.g. push) and $C_1V_1C_2+/i/ \text{ (e.g. pushy)}$. The purpose of this experiment is to determine whether listeners have a sensitivity to the presence or absence of a word-final vowel when the stem vowel length is controlled for. Therefore, using Praat, the final [i] vowel of the CVCi words was condensed to 25% and 12.5% and subjected to the fade out function in Audacity. In order to control for the effects of differing acoustic information from the first syllable of these words, the condensed [i] vowel was also appended to the monosyllabic stem of each of the experimental items. This resulted in six conditions, presented in Table 9.

### Table 9: Conditions of Vowel Modification

<table>
<thead>
<tr>
<th>Vowel Manipulation</th>
<th>None</th>
<th>Condensed to 25%</th>
<th>Condensed to 12.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stem from Monosyllabic Word</td>
<td><em>push</em></td>
<td><em>push+[i]25</em></td>
<td><em>push+[i]12.5</em></td>
</tr>
<tr>
<td>Stem from Disyllabic Word</td>
<td><em>pushy</em></td>
<td><em>pushy25</em></td>
<td><em>pushy12.5</em></td>
</tr>
</tbody>
</table>

As in Experiment 1, each consonant condition (/n s /ʃ /dʒ/) included 12 items, for a total of 60 experimental items. Note that in the above conditions, for each CVC word that participants heard, they heard 5 CVCi words. This could possibly result in a bias to hear a vowel. However, several measures were taken to avoid this. First of all, in addition to the 60 coda consonant items, there were 95 fillers which focused on sounds unrelated to the coda consonant test conditions and 8 practice items, for a total of 163 items. This, in conjunction with the use of a forced-choice

---

11 These particular cut-off points were established on the basis of pilot experiments with NSs.
word-identification task designed to tap into representations at the phonemic/phonological level (described in the next section), allowed for optimal conditions to avoid this bias.

4.3 Procedure

4.3.1 Perception Tasks

Stimuli were presented using E-Prime. The first task was a forced-choice word-identification task. A forced-choice word-identification task was chosen over an AXB task for several reasons. First, because of the manipulations of the stems/vowel and the nature of wanting to know how listeners categorize these words, a forced-choice word-identification task was the more appropriate option. In addition, because of the high accuracies overall in Experiment 1, a more difficult task was desired. Identification tasks are typically more difficult than discriminations tasks. In each trial, participants heard one stimulus after which two words appeared on the computer screen. They were directed to press the button corresponding to the word they heard as quickly as possible (see Appendix D for the exact instructions that were provided in the experiment).

The second perception task was a vowel-detection task. It followed a design similar to that of Dupoux et al. (1999) and asked participants to simply answer yes or no to whether they heard a vowel at the end of the word. It was emphasized that the number of yes or no answers need not be balanced (see Appendix D for the exact instructions). Because the second task was quite explicit and could possibly draw participants’ attention to the focus of the study, it always followed the word-identification task. This task included only the coda context items (along with 10 practice items, for a total of 70 items).
For both experiments, six lists of stimuli were created, balancing conditions in six lists so that no participant heard the same word in more than one condition. In addition, participants did not hear the same word in the same condition in each of the two tasks (e.g., if a participant heard pushy25 in the forced-choice task, s/he would not hear that in the vowel-detection task). Test items were pseudo-randomized in E-Prime. For both tasks, responses were recorded as percentages of vowels perceived. In other words, a score of 75% would indicate that listeners perceived a vowel 75% of the time. This method of reporting results was chosen because we are investigating participants’ sensitivity to the presence or absence of a vowel. $d'$ prime scores were not computed because it was not possible to calculate them. In this experiment, manipulated vowels are appended to mono- and disyllabic stems for six possible conditions in a forced-choice word-identification task. Because of this, it is not possible to compute hits and false alarms. In addition, percent accuracy is not an appropriate method of reporting results because determining which response would be accurate for the manipulated words is not possible. Instead, reporting responses as percentages of vowel perceived allows us to investigate differing trends of when learners reported hearing a vowel and when they did not.

4.3.2 Production Task

The L2 learners who completed this series of perception tasks also completed a read-aloud production task as in Experiment 1. The procedure was the same as in Experiment 1. The production experiment took place after the perception experiment in order to avoid participants guessing the focus of the study. A read-aloud production task with words in isolation was chosen because it will allow for a clearer comparison to be drawn to the results from Experiment 1.
The production task was analyzed similarly to Experiment 1 for the coda context words in that all productions were coded as either accurate or inaccurate with respect to the final C/CV syllable. Productions were rated by one trained English pronunciation teacher and one naïve listener. A correlation analysis was run on 87% of the data, and showed a high inter-rater reliability coefficient between the two codings ($r=.904, p<.001$). Productions of the learners in this experiment are compared to those of NSs from Experiment 1.

4.4 Predictions

Based on the results from Experiment 1, it is predicted that NSs will be at ceiling for un-manipulated words in the perception tasks. What remains unknown is the relative contribution of stem and final vowel length in perceiving these words. If it is the case that the stem (i.e., from mono- vs. disyllabic words) influences perception more than final vowel length does, we might expect native speakers to report perceiving a disyllabic word more often when presented with a stem from a disyllabic word regardless of the length of the final vowel. Simultaneously, we might expect that native speakers will indicate perceiving a monosyllabic word more often when presented with a stem from a monosyllabic word regardless of the length of the final vowel. If the above predictions hold true, we would be able to conclude that the stem (i.e., from mono- vs. disyllabic words) drives perceptions to a greater extent than final vowel length does (25% vs. 12%).

Based on the result from Experiment 1 that showed significant differences between the high- and mid-proficiency learners for the perception and production of palatals in codas, we might predict that high-proficiency learners will perform like NSs and that mid-proficiency

---

12 Two participants (or 13%) were tested after the research assistant who rated this data’s appointment ended. Thus, only 87% of the data were rated by two raters.
learners will perform differently from both groups. Finally, results from the production task are expected to mirror results from Experiment 1 because it is the same task but with a different set of learners.

4.5 Results

4.5.1 Perception

We begin with a presentation of the NSs’ results. Because of their similarity, results from both the forced-choice word-identification task and the vowel-detection task are reported together. Figures 20 and 21 indicate the average percent vowel perceived in each of the tasks by NSs. Our goal is to determine whether stem (mono- vs. disyllabic) modulates perception when the final vowel is either 12% or 25%. The unmanipulated mono- and disyllabic words are included in the graphs for comparison.
Figure 20: Average percent vowel perceived in the forced-choice word-identification task for NSs

Figure 21: Average percent vowel perceived in the vowel-detection task for NSs
As we can see in Figures 20 and 21, NS participants reported hearing a vowel in the un-manipulated disyllabic condition 100% of the time, and only 2% and 3% in the monosyllabic cases, respectively. This result is expected because these participants are NSs and performed at ceiling on Experiment 1. We can also see that the stem (i.e., from mono- vs. disyllabic words) does affect whether the participant hears a vowel at the end of the word: Even when the word-final vowel was reduced to 12% of its original length in disyllabic words, NSs still reported hearing a vowel 83% and 86% of the time in, respectively, the word-identification and vowel-detection tasks. In comparison, when the vowel added to a monosyllabic word was reduced to 12%, NSs only reported hearing a vowel 20% and 26% of the time in, respectively, the word-identification and vowel-detection tasks. Because the acoustic information of the word-final vowel in each of these cases was exactly the same, we can conclude that the information from the stem (e.g., longer duration of the stem vowel in monosyllabic words as compared to disyllabic words, shorter duration of the second consonant in monosyllabic words as compared to disyllabic words, and other attributes such as f0 patterns) was driving perception decisions.

When considering what contributes to the perception of a disyllabic word over a monosyllabic word, we have at least two cues: the stem and the presence of a final vowel. As we decrease the final vowel from 25% to 12.5%, if the information from the stem did not matter, we might expect to see the perception of the vowel steadily decreasing; however, in the current data, this appears to happen less so with the disyllabic stem than the monosyllabic stem. Ultimately, if we consider the cases where there is 12.5% of the original vowel in the explicit vowel-detection task, we see that NSs only perceive a vowel 26% of the time with the monosyllabic stem. On the other hand, with the disyllabic stem with the same vowel, NSs perceive a vowel 86% of the time. Therefore, cues from the stem are strongly affecting this perception.
Next, the data from the high-proficiency learners are presented. Again, because of their similarity, results from both the forced-choice word-identification task and vowel-detection task are reported together. Figures 22 and 23 indicate the average percent vowel perceived in each of the tasks by the high-proficiency learners. Results from the NSs are included in the figures for comparison.

**Figure 22: Average percent vowel perceived in the forced-choice word-identification task for NSs and high-proficiency learners**
Figure 23: Average percent vowel perceived in the vowel-detection task for NSs and high-proficiency learners

As we can see in Figures 22 and 23, in the conditions that were not manipulated, high-proficiency learners reported hearing a vowel in disyllabic words 91% and 90% of the time in, respectively, the word-identification and vowel-detection tasks, and they reported hearing a vowel in monosyllabic words 3% and 17% of the time, respectively. We can also see that the stem does seem to affect whether the high-level learners hear a vowel at the end of the word. Even when the vowel was reduced to 12% of its original length in disyllabic words, high-proficiency learners still reported hearing a vowel 75% and 73% of the time, respectively. In comparison, when the vowel added to a monosyllabic word was reduced to 12%, these learners only reported hearing a vowel 23% and 33% of the time, respectively. Before performing statistical analyses, let us turn to the data from the mid-proficiency learners.
We begin with the results from the forced-choice word-identification task. Figure 24 indicates the average percent vowel perceived in the word-identification task by mid-proficiency learners.

Figure 24: Average percent vowel perceived in the forced-choice word-identification task for all groups

As we can see in Figure 24, in the conditions that were not manipulated, mid-proficiency learners reported hearing a vowel only 69% of the time in the disyllabic condition and 10% of the time in the monosyllabic condition. NSs and high-proficiency learners reported hearing a vowel 100% and 91% of the time in the disyllabic condition, respectively, and 2% and 3% of the time in the monosyllabic condition, respectively. It appears that mid-proficiency learners are reporting hearing a vowel more often than NSs and the high-proficiency group in the monosyllabic cases, but less often in the disyllabic cases.
It is also less clear for mid-proficiency learners whether the stem affects whether the participant hears a vowel at the end of the word. When the vowel was reduced to 12% of its original length in disyllabic words, mid-proficiency learners reported hearing a vowel only 50% of the time, compared to 83% and 75% for, respectively, NSs and high-proficiency learners. On the other hand, when the vowel added to a monosyllabic word was reduced to 12%, mid-proficiency learners reported hearing a vowel 19% of the time, which is similar to NSs and high-proficiency learners, who reported hearing a vowel 20% and 23% of the time, respectively.

A mixed-design repeated-measures ANOVA was performed on the word-identification data, with stem (monosyllabic, disyllabic) and vowel manipulation (12%, 25%) as within-subject variables and with proficiency (native, high, mid) as between-subject variable. Recall that the unmanipulated conditions are not included in this analysis, as its purpose is to examine the orthogonal effects of stem and of vowel length. The analysis of the forced-choice word-identification data revealed main effects of stem, $F(1,36)=103.76, p<.001$, vowel, $F(1,36)=62.53, p<.001$, and proficiency, $F(2,36)=5.49, p<.008$. There was also an interaction between stem and proficiency, $F(2,36)=3.98, p<.028$, and between stem and vowel, $F(1,36)=4.81, p<.035$, but no interaction between vowel and proficiency or between stem, vowel, and proficiency ($F<1$).

Given the stem-by-proficiency interaction, we test whether the effect of stem is significant for each proficiency group across vowels, with alpha levels adjusted to .017. Paired-samples $t$-tests showed a significant difference between mono- and disyllabic stems for the NS group, $t(23)=-11.86, p<.001$, the high-proficiency group, $t(7)=-4.72, p<.002$, and the mid-proficiency group, $t(6)=-14.78, p<.003$. Each group perceives a vowel significantly more for the disyllabic stem as opposed to the monosyllabic stem; however, the effect is much larger for some of the groups than for others. Thus, we can conclude that although all groups show some
sensitivity to the stem, the native speakers and high-proficiency learners show much more sensitivity than the mid-proficiency learners.

There was also a significant interaction between stem and vowel, with the effect of vowel being larger for words with monosyllabic stems than for words with disyllabic stems. However, because the critical point for this study is not to examine the effect of the word-final vowel length itself, but rather to determine whether the stem modulates the perception of the word-final vowel, pairwise comparisons that compare the effect of length of the final vowel were not conducted.

Before drawing more conclusions, let us consider the data from the vowel-detection task. Figure 25 indicates the average percent vowel perceived in the vowel-detection task by all groups.

![Figure 25: Average percent vowel perceived in the vowel-detection task for all groups](image)

As we can see in Figure 25, in the conditions that were not manipulated, mid-proficiency learners reported hearing a vowel only 57% of the time for disyllabic words and 29% of the time
for monosyllabic words. These results do not pattern with those of NSs and high-proficiency learners, who reported hearing a vowel 100% and 90% of the time in disyllabic words, respectively, and 3% and 17% of the time in monosyllabic words, respectively. Again, it appears that mid-proficiency learners are reporting hearing a vowel more often than NSs and the high-proficiency group in the monosyllabic cases (except in the edge25 case), but less often in the disyllabic cases.

It is also less clear for mid-proficiency learners whether stem affects whether participants hear a vowel at the end of the word. When the vowel was reduced to 12% of its original length in disyllabic words, mid-proficiency learners reported hearing a vowel only 17% of the time, whereas NSs and high-proficiency learners reported hearing a vowel 86% and 73% of the time. When the vowel added to a monosyllabic word was reduced to 12%, mid-proficiency learners reported hearing a vowel 40% of the time, while NSs and high-proficiency learners reported hearing a vowel 26% and 33% of the time.

A mixed-design repeated-measures ANOVA was also performed on the vowel-detection data, with stem (monosyllabic, disyllabic) and vowel (12%, 25%) as within-subject variables and with proficiency (native, high, mid) as between-subject variable. Recall that for the word-identification task, there were significant main effects for stem, vowel, and proficiency, as well as stem-by-proficiency and stem-by-vowel interactions. A mixed-design repeated-measures ANOVA was performed on the vowel-detection data, with stem (monosyllabic, disyllabic) and vowel manipulation (12%, 25%) as within-subject variables and with proficiency (native, high, mid) as between-subject variable. The analysis of the vowel-detection data revealed main effects of stem, $F(1,36)=29.98, p<.001$, vowel, $F(1,36)=10.22, p<.003$, and proficiency, $F(2,36)=16.70, p<.001$. There was also an interaction between stem and proficiency, $F(2,36)=12.42, p<.001$, \textit{etc.}
between vowel and proficiency, $F(2,36)=9.91, p<.001$, and between stem, vowel, and proficiency, $F(2,36)=6.16, p<.005$, but no interaction between stem and vowel ($F<1$).

Given the three way interaction, repeated-measures ANOVAs are conducted separately for each group with alpha levels adjusted to .017. A repeated-measures ANOVA was performed on the native speaker group’s vowel-detection data, with stem (monosyllabic, disyllabic) and vowel manipulation (12%, 25%) as within-subject variables. This analysis revealed main effects of stem, $F(1,23)=124.20, p<.001$, and vowel, $F(1,23)=44.23, p<.001$. There was also an interaction between stem and vowel, $F(1,23)=9.53, p<.005$. Thus, we can also test for the effect of stem separately for 12% and 25% vowels for the NS group. Paired-samples t-tests showed significant differences between mono- and disyllabic stems for 12% vowels, $t(23)=-12.34, p<.001$, and 25% vowels, $t(23)=-6.33, p<.001$. Thus, the NS group reported hearing a vowel significantly more often with the disyllabic stem in both vowel conditions.

A similar repeated-measures ANOVA was performed on the high proficiency group’s vowel-detection data, with stem (monosyllabic, disyllabic) and vowel manipulation (12%, 25%) as within-subject variables. Unlike the results for the NS group, the effect of stem did not reach significance, $F(1,7)=6.93, p<.034$, nor did the effect of vowel, $F(1,7)=9.21, p<.019$. There was also no interaction between stem and vowel ($F<1$). A similar repeated-measures ANOVA was also performed on the mid-proficiency group’s vowel-detection data, with stem (monosyllabic, disyllabic) and vowel manipulation (12%, 25%) as within-subject variables. Similar to the high-proficiency group, this analysis revealed no main effects of stem, $F(1,6)=1.62, p<.251$, or vowel, $F(1,6)=1.23, p<.310$, and no interaction between stem and vowel, $F(1,6)=5.02, p<.066$. 
If we look at the percent vowel perceived by the mid-proficiency group in the vowel-detection task, we see that it ranges between 17%-40% in all contexts except for disyllabic words that were not manipulated, where it is 57%.

In summary, all three proficiency groups show a main effect of stem in the word-identification task, but the learners rely on the stem to a lesser extent than native speakers. It is likely that the mid-proficiency group is driving the interaction. Only the NS group demonstrates a significant effect of stem in the vowel-identification task and the learner groups did not. These results indicate that learners are able to use stem cues similarly to NSs in some tasks, but not others. I return to this in the discussion section of this chapter.

4.5.2 Production

Now let us turn our attention to the production results from Experiment 2. Figure 26 shows the production accuracies of fricatives and affricates by high- and mid-proficiency learners (the NSs’ results are those from Experiment 1).
As we can see from Figure 26, high-proficiency learners performed better on each of the three palatal types compared to mid-proficiency learners. A mixed-design repeated-measures ANOVA was conducted on the participants’ production accuracies, with consonant (/ʃ ʤ ʧ/) as within-subject variable and proficiency (native, high, mid) as between-subject variable. As in Experiment 1, there were main effects of proficiency, $F(2,20)=18.58, p<.001$, and consonant, $F(3,60)=24.53, p<.001$, and an interaction between consonant and proficiency, $F(6,60)=9.47, p<.001$.

Given the significant interaction, one-way ANOVAs with proficiency (native, high, mid) as between-subject variable were conducted separately for each consonant, with alpha levels adjusted to .0125. There were main effects of proficiency for the consonants /ʃ/, $F(2,22)=16.50, p<.001$, /ʧ/, $F(2,22)=17.27, p<.001$, and /ʤ/, $F(2,22)=14.13, p<.001$, but not /s/, $F(2,22)=4.08, p<.033$. Post-hoc Tukey tests showed a significant difference between the mid-proficiency group and the native speaker group for the consonants /ʃ ʧ ʤ/ ($p<.001$), but not /s/ ($p<.053$). These tests
also showed a significant difference between the mid-proficiency group and the high-proficiency group for /ʃʧ/ \((p<.01)\), but not /dz/ \((p<0.15)\) or /s/ \((p<.053)\). These tests did not show significant differences between the native and high-proficiency groups for /ʃʧs/ \((p>0.1)\) or /dz/ \(p<.081\). In summary, there were no differences found between any of the groups in terms of performance on the production of /s/, and the learner groups also did not differ for /dz/.

As with the results from Experiment 1, we might also consider an analysis that compares /ʃ/ and /ʧ/. When we conduct a mixed-design repeated-measures ANOVA with consonant type (/ʃʧ/) as within-subject variable and proficiency (native, high, mid) as between-subject variable, we no longer find a main effect of consonant type \((F<1)\). The main effect of proficiency remains significant, \(F(2,20)=18.83, p<.001\). This is further evidence that /s/ is behaving differently from palatals.

If we look separately at mono- and disyllabic words (shown in Figure 27), which showed differences in Experiment 1, we again see that mid-proficiency learners are making more mistakes with the disyllabic words than the monosyllabic words. In other words, the more common error is to omit the final [i] in words like pushy rather than to epenthesize a vowels in words like push. Again, similar to the results in Experiment 1, it seems that /dz/ displays difficulties in both contexts for mid-proficiency learners.
Figure 27: Production accuracies separated by mono- and disyllabic words for mid-proficiency learners

A repeated-measures ANOVA was performed with two within-subject factors: word type (monosyllabic, disyllabic) and consonant (/ʃ ʤ ʧ/). Unlike the results from Experiment 1, there was no main effect of word type ($F<1$). Similar to the results from Experiment 1, there was no main effect of consonant ($F<1$), but there was an interaction between word type and consonant, $F(2,12)=7.46, p<.008$. Post-hoc Bonferroni comparisons were conducted with alpha levels adjusted to $p<.017$. Paired-samples $t$-tests showed no significant difference between mono- and disyllabic words for the consonants /ʤ/, $t(6)=-.180, p<.863$, /ʃ/, $t(6)=-1.64, p<.152$, and /ʧ/, $t(6)=-2.34, p<.058$. Thus, mid-proficiency learners are not patterning similarly to those in Experiment 1. Recall that this set of mid-proficiency learners included seven participants while the group in Experiment 1 was comprised of 11 participants. We do not see exactly the same patterns as in Experiment 1; this set of learners’ average production accuracies for monosyllabic words is slightly lower than those for learners in Experiment 1 (87%, 65%, 82% as compared to...
95%, 71%, 95%), and their standard deviations showed more variability. Nevertheless, as in Experiment 1, learners are still trending differently between mono- and disyllabic production accuracies with /ʃʧ/ words as compared to /ʤ/ words.

4.5.3 Comparing Perception and Production

Because of the design of Experiment 2, it is not possible to conduct an analysis comparing the perception and production of palatal codas. This is because in the perception task, learners only heard four palatal words that had not been manipulated in each consonant condition. Because of the very low number of items, an analysis is inappropriate.

4.6 Discussion

We set out in this chapter to answer the following research questions. Here I consider each in turn.

1. Do native speakers of English show sensitivity to the presence or absence of a word-final vowel in words like fish/fishy if the length of the stem vowel is controlled for?
2. Do L2 learners of English show sensitivity to the presence or absence of a word-final vowel in words like fish/fishy if the length of the stem vowel is controlled for? Does proficiency affect results?

4.6.1 Do native speakers of English show sensitivity to the presence or absence of a word-final vowel in words like fish/fishy if the length of the stem vowel is controlled for?

The results from Experiment 2 demonstrate that yes, native speakers of English show a sensitivity to the presence or absence of a word-final vowel in words like fish/fishy if the length
of the stem vowel is controlled for. More specifically, we have seen that regardless of the length of the manipulated word-final vowel in words containing the stem of a monosyllabic word (25% or 12.5%), native speakers still identified these words as monosyllabic in a majority of cases. On the other hand, when the vowel was shortened in disyllabic words, native speakers still reported hearing disyllabic words in a majority of cases. Thus, we can conclude that relative differences in stem vowel length (as well as all other differences present in the stem such as palatal consonant length, f0 patterns, etc.) were guiding these perceptions. The question is whether these differences also guide L2 learners’ perceptions, potentially leading them to judge the word accurately in spite of difficulties with the word-final coda.

4.6.2 Do L2 learners of English show sensitivity to the presence or absence of a word-final vowel in words like fish/fishy if the length of the stem vowel is controlled for? Does proficiency affect results?

Results from Experiment 2 also demonstrate the effects of the stem on learners’ perceptions of a following vowel. We have seen that high- and mid-proficiency learners behave like NSs in that the stem significantly affects how they perceive that word in the word-identification task. In contrast, mid-proficiency learners perform differently from NSs and high-proficiency learners in the word-identification task, showing a smaller effect of stem than the first two groups did. The results on the vowel-detection task indicated that high- and mid-proficiency learners are not able to use cues from the stem in a similar way to NSs. It could be the case that in the vowel-detection task, the learners are able to focus on the final vowel and disregard, or ‘tune-out’, the stem information. The NSs, for whom the stem cues are stronger, are not able to disregard stem information and thus continue to show effects in the vowel-
identification task. These results may shed some light on the results from Experiment 1 in that high-proficiency learners may have been able to use the cues from the stem to help guide their perception decisions rather than information relating to the existence of a final vowel, but that mid-proficiency learners may not have been able to do so.

4.7 Summary

The results from Experiment 2 provide additional support to the findings from Experiment 1. With regard to the production of palatals, we saw that, as in Experiment 1, high-proficiency learners performed significantly better than mid-proficiency learners. We also saw that mid-proficiency learners tended to make more mistakes with the disyllabic words than the monosyllabic words.

One possible concern with the design of Experiment 2, that was not a concern in Experiment 1, relates to potential word-familiarity effects that could have influenced perception results. If we consider that learners were completing a forced-choice word-identification task, if it were the case that a learner was more familiar with one word than another in the minimal pair, the learner could be biased to hear the more familiar word. Up to this point, stimuli in the experiments have included only real words. This was originally done for ecological validity purposes. However, as word familiarity might potentially affect results, it would be beneficial to include nonce words in perception and production tasks. Therefore, the stimuli in Experiment 3 are designed such that half of the words are real and half are nonce. This will allow us to compare results on real and nonce words to determine if word familiarity does have an effect. I return to this in the results section of Experiment 3 in Chapter 5.
4.8 Impetus for Experiment 3

From the results of Experiment 1, we have preliminary evidence that the perception and production systems are not directly linked, in that we did not find co-variation among perception and production accuracy scores. Nevertheless, in order to have a better understanding of the relationship between perception and production with regard to syllable structures constraints, we still need to see what effects perceptual training may have on each. Experiment 3, presented in the following chapter, aims to answer these questions by reporting on results from a perceptual phonetic training experiment.

Recall from our discussion of perceptual phonetic training that we can investigate whether learners generalize learning from perceptual training to new words and new talkers. We can determine whether they generalize that learning in both perception and production. We considered two theories that discussed underlying mechanisms that might account for generalizability following perceptual training: episodic-trace theories and abstractionist theories. While we concluded that both might provide explanations for why this type of perceptual training would results in generalization to new words and new talkers, we also suggested that episodic-trace theories might predict better performance on trained items, a point to which I will return in the discussion section of Chapter 5.

This type of perceptual training can also have pedagogical implications. We want to determine whether perceptual training can allow for generalization to new words and new talkers. If we find this is the case, then we have a potential argument for the incorporation of this type of training in pronunciation classrooms. However, if that is to be the case, then we will also want to consider ways to create a perceptual training paradigm that is pedagogically feasible as
well as determine whether learning can be extended to larger discourse contexts. With these considerations in mind, we now turn the Experiment 3 in Chapter 5.
CHAPTER 5
EXPERIMENT 3: PERCEPTUAL TRAINING AND ITS EFFECTS ON PERCEPTION AND PRODUCTION

The goal of Experiment 3 is to provide evidence to answer research questions related to: (i) the relationship between learners’ perceptions and productions of segments in an existing, but restricted syllable structure (palatals in coda position); (ii) the effects of pedagogically-viable perceptual phonetic training materials on perception and production accuracies, including whether or not learners generalize training to new words and new talkers for palatal codas in both perception and production measures; and (iii) the IL system of Korean learners of English with regard to palatal production.

Using a pretest/perceptual training/post-test design, I investigate whether perceptual phonetic training on palatal codas has an effect on perception and/or production accuracies, and whether improvements generalize to new words and new talkers. The research questions of Experiment 3 are as follows:

1. Can pedagogically viable perceptual phonetic training on palatal codas improve perception accuracies of palatal codas?
2. Does perceptual phonetic training on palatal codas allow generalization to new words and new talkers?
3. What will be the effects, if any, of perceptual phonetic training on palatal codas on productions of palatal codas?
4. What is the relationship, if any, between improvements on perceptions and productions of palatal codas?
5. In which contexts (words in isolation, words within a larger discourse, final singleton palatals, final palatal clusters, palatals before –ed morphemes, etc.) do learners have the most difficulty with palatals?

5.1 Participants

Participants included 24 adult, Korean L2 learners of English who did not participate in Experiments 1 or 2, randomly assigned to two groups. The experimental group (7 men and 5 women) received perceptual training on palatal codas, and the control group (7 men and 5 women) received perceptual training on tense/lax vowel pairs.13 This latter group completed both perception and production pretests, a training task unrelated to palatals to ensure a similar amount of time on task, as well as the post-tests. Five NSs also completed the pretests (2 males and 3 females).

All participants completed a language background questionnaire (identical to the one used in the research reported upon earlier in this dissertation, presented in Appendix B), and L2 learners completed a cloze test (identical to the one used earlier, presented in Appendix A). Table 10 shows the participants’ cloze test scores, as well as the means and ranges for a subset of relevant language background information.

---

13 In Bradlow et al., 1997’s study, the control group completed the pre/post-tests, but received no training of any sort, nor was it reported that they spent a similar amount of time on another task. In order to ensure improvements could not be contributed to time spent on task, the control group in the current study also completed a perceptual training, but that training focused on tense/lax vowel distinctions rather than on palatal codas.
Table 10: Language Background Information for Experiment 3

<table>
<thead>
<tr>
<th></th>
<th>Cloze Test (%)</th>
<th>Daily Usage (%)</th>
<th>1st Exposure to English (years)</th>
<th>Years in Immersion Context</th>
<th>Years of Instruction</th>
<th>Age</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>NSs (n=5)</strong></td>
<td>Mean n/a</td>
<td>97</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td>SD n/a</td>
<td>4.0</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>7.3</td>
</tr>
<tr>
<td></td>
<td>Range n/a</td>
<td>90-100</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>22-39</td>
</tr>
<tr>
<td><strong>Experimental Group (n=12)</strong></td>
<td>Mean 28</td>
<td>42</td>
<td>11</td>
<td>4.1</td>
<td>10</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>SD 7.4</td>
<td>27.3</td>
<td>4.3</td>
<td>3.5</td>
<td>5.7</td>
<td>8.7</td>
</tr>
<tr>
<td></td>
<td>Range 9-38</td>
<td>5-85</td>
<td>0-15</td>
<td>0-10</td>
<td>3-22</td>
<td>18-46</td>
</tr>
<tr>
<td><strong>Control Group (n=12)</strong></td>
<td>Mean 27</td>
<td>49</td>
<td>12</td>
<td>4.5</td>
<td>9</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>SD 10.1</td>
<td>35.3</td>
<td>2.7</td>
<td>4.0</td>
<td>3.7</td>
<td>8.4</td>
</tr>
<tr>
<td></td>
<td>Range 15-38</td>
<td>10-99</td>
<td>6-16</td>
<td>0.2-13.3</td>
<td>5-17</td>
<td>21-48</td>
</tr>
</tbody>
</table>

5.2 Materials

In the pre- and post-test phases, participants completed a forced-choice word-identification experiment and two read-aloud production experiments, one in which they read the words that had been heard in the forced-choice word-identification task and one in which they read dialog/paragraphs eliciting palatal codas in a wide variety of contexts including, but also extending beyond, those that are the focus of the training. In the perceptual training phase, participants completed a forced-choice word-identification task.

Experimental stimuli for the perception experiment and for the first production experiment were 48 minimal pairs of natural tokens, half of which were real words and half of
which were nonce words (see Appendix F for a complete list of stimuli). They included singleton 
palatals in coda position (e.g., real words: *push/pushy, dodge/dodgy, catch/catchy*; nonce words: 
*mish/misy, tudge/tudgy, tetch/tetchy*). The decision to use nonce words is based on the limited 
availability of these word pairs in English and to avoid potential word frequency effects present 
in Experiment 2. Each of the three conditions (ʃ, ʧ, ʤ) included eight minimal pairs used in the 
perceptual training as well as eight additional pairs used in the pretests/post-tests. Thus, only a 
subset of stimuli was presented in the training condition. This was done to determine whether 
learners can generalize improvement from the training to novel words.

For both the perception tests and the perceptual training, minimal pairs were presented in 
isolation as well as within the carrier sentences “He said X angrily” and “He said X frequently” 
to provide contexts in which the target word is followed by a vowel (*angrily*) as well as a 
consonant (*frequently*). The addition of testing more than just words in isolation was included to 
determine whether a larger, sentential context affects perception/production accuracy rates. 
Having both prevocalic and pre-consonantal conditions would allow us to test whether these 
contexts affect accuracy rates. The prevocalic condition might be easier for learners in that it 
would allow them to parse the palatals as the onset of the following syllable, thus not violating 
any syllable structure constraints of their L1. In the pre-consonantal condition, on the other 
hand, learners have no choice but to parse palatals as codas because /ʃʃ, ʤʤ, ʧʧ/ are not possible 
onset clusters in either Korean or English. The 48 stimuli were thus encountered three times: 
twice in context (*angrily/frequently*) and once in isolation. Following the design of Bradlow et al.

---

14 The talkers were explicitly instructed to read sentences such that they included the natural re-syllabification that 
occurs during English speech. If a talker inserted a pause between a target word and *angrily/frequently*, they were 
instructed to read the sentence again.
(1997), there were also 28 minimal pairs that contrasted other phonemes of English, both in isolation and within sentences.\(^{15}\)

All of the stimuli were recorded by six native speakers of English (three men and three women). Table 11 lists biographical information for the six talkers who produced the stimuli. As we can see from Table 13, two talkers are from the Inland North, three are from the Midland and one is from the South (Labov, Ash, Boberg, 2006). All talkers had been living in Illinois for at least 4.2 years at the time of recording.

**Table 11: Biographical Information for the Six Talkers who Produced Stimuli**

<table>
<thead>
<tr>
<th>Talker</th>
<th>Gender</th>
<th>Age</th>
<th>Hometown</th>
<th>Length of Residence in IL (years)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Talker 1</td>
<td>male</td>
<td>32</td>
<td>Central IL</td>
<td>32</td>
</tr>
<tr>
<td>Talker 2</td>
<td>male</td>
<td>33</td>
<td>Southern MI</td>
<td>8.3</td>
</tr>
<tr>
<td>Talker 3</td>
<td>male</td>
<td>25</td>
<td>Central IL</td>
<td>20</td>
</tr>
<tr>
<td>Talker 4</td>
<td>female</td>
<td>26</td>
<td>Central SC</td>
<td>4.2</td>
</tr>
<tr>
<td>Talker 5</td>
<td>female</td>
<td>27</td>
<td>East Central, NJ</td>
<td>4.3</td>
</tr>
<tr>
<td>Talker 6</td>
<td>female</td>
<td>31</td>
<td>Upstate NY</td>
<td>5.2</td>
</tr>
</tbody>
</table>

\(^{15}\) It was not the intent that these items would be used as fillers in the traditional sense of the word such that they would attempt to completely obscure the focus of the experiment from participants. In fact, previous research (Guion & Pederson, 2007) has shown that the addition of explicitly instructing participants to attend to phonetic cues significantly increases the benefits of perceptual training. While such instructions were not included in this experiment, it was also not the intent to hide the focus of the experiment or training from participants. These items were included to match the procedure of Bradlow et al. (1997) as closely as possible for purposes of comparing results.
Stimuli were recorded in a sound-attenuated booth at the University of Illinois at Urbana-Champaign via a Marantz PMD570 solid state recorder using either an Earthworks M30 standing microphone or an AKG c520 head-mounted microphone at 44.1 kHz. Stimuli were then segmented into individual files and normalized to 65dB using Praat. Recordings from two men and two women were used in both the pretests/post-tests as well as the training, while the recordings from the other man and woman were used only in the pretests/post-tests to determine whether learners can generalize improvement gained in the training to novel talkers. The two talkers not used for training were chosen at random.

The purpose of the second production test was to gain insight into the whole IL system of learners with regard to their production of palatals. It consisted of a dialog/paragraph reading task (with only real English words) containing palatals in a wide variety of contexts including, but also extending beyond, those that are the focus of the training. Conditions included singleton codas and their disyllabic counterparts (e.g., push/pushy), complex codas including /n/, /l/ or /ɹ/ before the palatal (e.g., pinch, perch, mulch), and each of these conditions before –ed morphemes (e.g., perched, dodged). A final condition of disyllabic words with stress on the first syllable (e.g., language, foolish) was also included. The context of the following sound (before a consonant, before a vowel, phrase-final) was also balanced to allow for an investigation of what effect, if any, context may have on the production of these palatals and for comparison purposes to the sentence contexts in the experiment. The conditions that matched the perception phase (i.e., singleton codas and their disyllabic counterparts) had a total of eight targets in each context (before a consonant, before a vowel, phrase-final), for a total of 72 items per consonant type (/ʃ /ʧ 16

16 –ed endings in environments that undergo consonant cluster simplification were not included. Consonant cluster simplification occurs when the ending is located between two consonants, but not when the second consonant is /w, h, j, ɹ/. For example, saved stamps undergoes consonant cluster simplification while kept out does not (Hahn & Dickerson, 1999).
\( \tilde{d}g \), or 216 words. Complex coda words included three consonants in the pre-palatal environment: /\(\tilde{u}\ n\ l/ in words like perch, pinch, and squelch. Where possible, conditions contained ten targets, although in some cases, real English words were limited (e.g., /\(\tilde{u}\j/\)). A complete list of stimuli as well as a count of each category and the contexts in which they appeared can be found in Appendix G. Appendix H provides an example of one of the dialogs.

5.3 Procedure

The procedure consisted of a pretest phase, a training phase, and a post-test phase conducted over approximately 10 days. The pretests and post-tests were administered individually in a quiet room. The perceptual training phase was completed online. For the perceptual training, participants were instructed to wear headphones and complete the tasks in a quiet environment. A more detailed description of each phase can be found in the following subsections.

5.3.1 Pretest Phase

The pretest phase consisted of both perception and production experiments. The perception tests included a forced-choice word-identification task of both words in isolation and in carrier phrases. A forced-choice word-identification task was chosen instead of an AXB task for two reasons. First, when we compare results from the AXB task in Experiment 1 and those from the forced-choice word-identification task in Experiment 2, we see that accuracy rates are higher for the AXB task. To avoid learners being at ceiling, the more difficult task was chosen. In addition, Bradlow et al. (1997) used a forced-choice word-identification task for their phonetic training study. Therefore, the forced-choice word-identification task was also chosen in order to
maintain similarity for comparison to those results. At the beginning of each trial, participants heard a word/sentence. Immediately after it was played, they saw the two words/sentences from each pair presented on the left and right side of the screen. They were instructed to choose the correct response as quickly as possible by pressing one of two marked keys on the keyboard (see Appendix D for the exact instructions). The ‘d’ and ‘l’ keys were marked with colorful tape. The ‘d’ key indicated a response of choosing the word on the left side of the screen, and the ‘l’ key indicated a response of choosing the word on the right side of the screen.

Each of the 96 experimental words (from the 48 minimal pairs) and 56 filler words (from the 28 minimal pairs) were presented once in isolation (in one block) and once in each carrier phrase context (in another block), along with eight practice items at the beginning of each task to familiarize participants with the procedure. The pre-test thus included a total of 472 trials. The isolated-word block lasted approximately 12-15 minutes and the carrier-phrase block lasted approximately 27-32 minutes. Because of the length of the carrier phrase block, participants were offered breaks at a third and two-thirds of the way through the experiment. Whether the correct word was on the left or right was counterbalanced across trials. The talker heard was also counterbalanced across trials, such that a learner did not hear both words from a minimal pair spoken by the same talker. Six lists were created to counterbalance across participants, such that all words from all talkers were heard. The order of block (whether a participant began with words in isolation or words in a carrier phrase) was counterbalanced across participants. Stimuli were presented using E-Prime, and participants wore either Beyerdynamic DT 770 or Sony MDR 7506 headphones and had control over the volume level via an Alesis iO2 USB interface.

The production pretests were completed after the perception pretest and were composed of two different tasks: a read-aloud task modeled on the perception task and a dialog/paragraph
reading task. Participants were balanced such that half completed the read-aloud task modeled on the perception task first and the other half completed the dialog/paragraph reading task first. Recordings were completed in a sound-attenuated booth at the University of Illinois at Urbana-Champaign via a Marantz PMD570 solid state recorder using an AKG c520 head-mounted microphone. The first set of participants was recorded at 44.1 kHz, but the settings were changed and the remaining participants were recorded at 48 kHz. However, recordings at 48 kHz were all converted to 44.1 kHz for the assessment phase. All participants’ pretest and post-test productions of target words were segmented into individual files and normalized to 65 dB using Praat.

In the read-aloud task modeled on the perception task, participants received a visual word/sentence prompt and read the word/sentence. All stimuli (real and nonce words in isolation and sentences) were combined, randomized, and presented using PowerPoint. Participants were instructed to read at a comfortable pace and to give their ‘best guesses’ for any unfamiliar words. Participants recorded all 456 tokens and were offered a break one third and two-thirds of the way through the list. The duration of the task was approximately 15-30 minutes, depending on the reading pace of participants and whether they took breaks.

For the dialog/paragraph reading task, participants received a print-out packet containing each dialog/paragraph on a separate page. Because of the large number of targets, the 14 dialogs were randomly divided into three sets, and participants were balanced as to whether they started with the first, second, or third set. Participants were instructed to read at a comfortable pace and to give their ‘best guesses’ for any unfamiliar words. Each set took approximately 10-12 minutes to read, for a total time on task of approximately 30-40 minutes.

17 In some cases, the participant did not record the word. From a total of 10,944 possible words (456 productions x 24 speakers), this occurred 16 times. In these cases, the items were omitted from the analysis.
5.3.2 Experimental Training Phase

The perceptual training phase for the experimental group consisted of eight, 20-minute, daily sessions of online training delivered via Paradigm Player (Perception Research Systems, 2007). An online delivery system for training was chosen for practical purposes. First, it was presumed that if participants were required to come to a lab daily, there would be a high percentage of participant attrition. Of course, the trade-off for having an online training system is that environmental context could not be controlled. Nevertheless, every measure was taken to ensure consistency across participants. For example, participants were instructed to complete the training in a quiet room and to wear headphones during their sessions. Online delivery was also chosen for pedagogical purposes. Presumably, if this type of training is to be incorporated into pronunciation classrooms, much of it would occur outside of classroom time. Therefore, an online training system would be a practical option.

The decision to have eight sessions was made for several reasons. First, listener performance has been shown to improve the most in the first ten training sessions, after which subsequent improvement is marginal (Logan & Pruitt, 1995). Nevertheless, these results mostly come from studies investigating /ɹ/ and /l/, which are particularly difficult for Japanese learners of English. Based on the results from Experiment 1, we know that learners perform fairly well with the perception of palatal contrasts. Therefore, it was predicted that they would need fewer training sessions to reach ceiling perception accuracy rates. The second reason for choosing eight, rather than, for example, ten sessions, was because a multiple of four was needed to balance the number of times participants heard each of the talkers. The final reason for choosing

---

18 Paradigm Player is a computer software application for experiment design, data collection and analysis.
eight training sessions is in line with the practical consideration of keeping materials pedagogically viable. The amount of time spent on /s/ and /l/ training sessions in Logan et al. (1991), Lively et al. (1993), and Bradlow et al. (1997) ranged from approximately 7.5-22 hours. A typical semester-long pronunciation course at the University of Illinois at Urbana-Champaign meets for approximately 40 hours over 16 weeks and covers a wide variety of topics. It was predicted that an 8-day training program lasting approximately four hours would not only feasibly fit into existing pronunciation classes, but also provide enough training to improve learners’ perceptions of these contrasts. The above considerations motivated choosing eight days for training. The training software allowed for daily tracking. An analysis of the changes in improvement during training is presented in Appendix I.

Each training session was comprised of a forced-choice word-identification task similar in procedure to the one used in the perception pretest/post-test, except feedback was provided and the words appeared before the sound file was played. Each training session was presented in two blocks: one including words in isolation and the other, words in carrier phrases. Participants always began with words in isolation and continued with words in carrier phrases. Each block consisted of (a) the set of 48 words (eight minimal pairs from each of the three conditions) in isolation from one talker along with 16 distractors, or (b) the set of 96 words in each of the carrier contexts from one talker along with 32 distractors. During each session day, learners heard stimuli from two different talkers (of the four who were randomly selected to be training stimuli). Blocks were counterbalanced such that over the course of the eight sessions, learners heard each word in isolation and in carrier phrases from each of the four talkers two times. The instructions given to participants as well as an example schedule are included in Appendix J.
The procedure for each trial was identical to the word-identification task of the perception pretest/post-test except that (a) during training participants received feedback as to whether or not they answered correctly, and (b) participants saw the words for 500 ms before the audio stimulus. For every response (whether correct or incorrect), participants heard the stimulus again during the feedback screen. During each training day, participants spent approximately 20 minutes on task for a total of approximately 160 minutes of perceptual training.

Participants were instructed to begin the perceptual training sessions as soon as possible, but no earlier than the day following the pretest. They were also instructed to complete the sessions in eight successive days, with a night’s sleep in between each session. For the purpose of learning, what is important is that participants wait at least one night before doing the next session; or, in other words, they should not complete two sessions in one day. This is because the brain consolidates information while asleep (see e.g., Walker & Stickgold, 2004; Stickgold, 2005; Marshall & Born, 2007). No participant completed two sessions in the same day. Nevertheless, because of participants’ schedules, sometimes there was more than one day between sessions. What is important for comparison purposes with the control group is to determine whether participants began and completed training in a comparable manner with control participants, and whether they spent approximately the same amount of time on task. After a description of the control training phase in the next subsection, a table is presented comparing completion habits of both groups.

19 McCandliss, Fiez, Protopapas, Conway, and McClelland (2002) investigated the success of perceptual training with and without feedback and demonstrated significant benefits when feedback is present.
5.3.3 Control Training Phase

The perceptual training phase for the control group consisted of eight daily sessions of online training delivered via Pierceive\textsuperscript{20} and had as its focus three tense/lax vowel pairs, \([\varepsilon \sim \varepsilon]\), \([i \sim i]\), and \([oo \sim o]\), presented in monosyllabic nonce pairs.\textsuperscript{21} Control group participants were randomly assigned to one of four training paradigms within which stimuli varied along three dimensions: talker, consonant context, and speech rate. Stimuli consisted of single-syllable, nonce minimal pairs in which ten consonants, /d, t, n, b, p, m, k, g, h, s/, were distributed between onsets and codas of three tense-lax vowel pairs \([\varepsilon \sim \varepsilon]\), \([i \sim i]\) and \([oo \sim o]\). Training on these nonce words always occurred in isolated words. Speech rate varied from ‘slow/careful’, to ‘normal/casual’, to ‘fast’. Talkers were eight NSs (four men and four women) of North American English.

Participants were instructed to spend approximately 20 minutes on task to mirror the time spent by the experimental group; however, unlike in the experimental perceptual training, these participants controlled the amount of time spent on task. Because of this, time on task varied across participants. Table 12 indicates the mean, standard deviation, and range of time spent on task by both the experimental and control groups, and it compares the pretest/post-test completion times (see Appendix K for a full list by participant). The results in Table 12 are reported presuming a night’s sleep in between each pretest/post-test and training day. Thus, a report of 0 days between pretest and training start indicates that a participant began training on the day following the pretest. A report of 1 day between pretest and training start, on the other hand, indicates that a participant began training with two night’s sleep after the pretest.

\textsuperscript{20} Liam Moran, a consultant for ATLAS Digital Media at the University of Illinois at Urbana-Champaign, created this software. I would like to thank him for allowing me to use it for this experiment.

\textsuperscript{21} I would like to thank Lisa Pierce at the University of Illinois at Urbana-Champaign, who created this training paradigm, for allowing me access to it for the purposes of using it as the control group training.
### Table 12: Perceptual Training Timing Comparison

<table>
<thead>
<tr>
<th></th>
<th>Days between pretest and training start</th>
<th>Days off during training</th>
<th>Days between training and post-test</th>
<th>Total time on training task (minutes)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Experimental Group</strong></td>
<td>Mean</td>
<td>2.50</td>
<td>0.58</td>
<td>0.50</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>2.35</td>
<td>1.00</td>
<td>0.80</td>
</tr>
<tr>
<td></td>
<td>Range</td>
<td>0-6</td>
<td>0-3</td>
<td>0-2</td>
</tr>
<tr>
<td><strong>Control Group</strong></td>
<td>Mean</td>
<td>2.33</td>
<td>1.25</td>
<td>0.25</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>3.87</td>
<td>1.48</td>
<td>0.62</td>
</tr>
<tr>
<td></td>
<td>Range</td>
<td>0-13</td>
<td>0-5</td>
<td>0-2</td>
</tr>
</tbody>
</table>

As we can see from Table 12, the experimental and control groups are similar in terms of (1) days between the pretest and start of training, (2) days off during training, and (3) the number of days between the end of training and the post-test. In contrast, the total time on task was quite different between groups, because the control group did not adhere to the instructions of spending 20 minutes/day on training. Ultimately, the control group spent less time completing perceptual training than the experimental group. I return to this issue in the discussion section of this chapter.

### 5.3.4 Post-test Phase

The post-test phase was identical to the pretest phase, including both the forced-choice word-identification tasks and the two production tasks. The perception tests were balanced.
across participants such that if a participant began with the words in isolation in the pretest, they began with the words in carrier phrases in the post-test. The production tasks were completed after the perception task and again were balanced such that if a participant began with the read-aloud task modeled on the perception task in the pretest, they began with the dialog/paragraph reading task in the post-test.

5.4 Data Analysis

Answers on the perception tests were scored as either accurate or inaccurate. As in Experiment 1, participants’ results on the perception task were then transformed into $d'$ scores. In Experiment 1, $d'$ scores were calculated to control for the potential bias of selecting the first word (A) or the third word (B) in the AXB task. In Experiment 3, the potential bias would be the likelihood of choosing the monosyllabic or disyllabic word. Thus, $d'$ scores were calculated following the hits and false alarms presented in Table 13 (Macmillan & Creelman, 1991).

Table 13: Explanation of $d'$ Scoring

<table>
<thead>
<tr>
<th>Hit: Mono = Mono</th>
<th>False Alarm: Di = Mono</th>
</tr>
</thead>
<tbody>
<tr>
<td>Miss: Di = Mono</td>
<td>Correct Rejection: Di = Di</td>
</tr>
</tbody>
</table>

The $d'$ score is calculated by subtracting the $z$ transformation of false alarms from the $z$ transformation of hits: $d' = z(H) - z(F)$.

There are several reasons for deciding to calculate $d'$ scores for the data in Experiment 3. First, $d'$ scores are more appropriate to answer the questions posed in Experiment 3 because we want to have a precise measure of the effect of training. This was not the case in Experiment 2, whose goal was to gather a sense of what causes difficulties for the L2 learners. In addition,
calculating $d'$ scores was not possible in Experiment 2. In Experiment 3, however, our ultimate goal is to isolate the effect of training. Calculating $d'$ scores allows us to remove potential biases and guesses. In the present experiment, a bias might mean having an initial tendency (during the pretest) to select disyllabic over monosyllabic words because learners have perception difficulty hearing palatal codas and hear epenthetic vowels. Then, if training results in making learners aware of this difficulty, training might lead learners to select more monosyllabic words in the post-test regardless of what they hear. Thus, training could lead to a bias in that monosyllabic words improve and disyllabic words do not. As $d'$ is a measure of sensitivity, it allows us to factor out the potentially confounding effect of bias. It is therefore the measure used for reporting the results of this experiment.

Production responses from the pretest and post-test read-aloud task modeled on the perception task were rated by a group of native listeners (NL) in two listening tasks: a paired-comparison task as well as a forced-choice word-identification task (described below). For the dialog/paragraph task, similar to Experiments 1 and 2, productions were coded as either accurate or inaccurate (with respect to the palatal) by a trained English pronunciation teacher. Twenty percent of the data (or five participants) were coded by a second trained English pronunciation teacher. Inter-rater reliability showed a high coefficient ($r=.917, p<.001$). In some cases, participants produced a consonant other than the target (e.g., /g/ instead of /ʤ/ in a word like *dodgy*). This happened to a varying degree with different participants. If a participant produced a consonant other than the palatal, then determining accuracy would be impossible, so these items were excluded from analysis. Of a total 8,856 items, 742 had to be excluded because either the pretest, the post-test, or both versions had an error in consonant.²²

²² Several common patterns of errors contributed to this high exclusion rate: (1) substituting /ŋ/ for /ʤ/ in low frequency words like *dingy, mangy, stingy*; (2) substituting /g/ for /ʤ/ in low frequency words like *clergy, bulgy,*
5.4.1 Paired-Comparison Task

Following Bradlow et al. (1997), a group of native-listeners (NL) performed a paired-comparison task with the learners’ pretest and post-test productions. For each trial, the target word was presented on a screen for 500 ms after which NLs heard both productions of the learner separated by 500 ms of silence. Listeners used a 7-point scale to judge which of the target words was ‘better,’ or, following Bradlow et al., which was a “clearer and more intelligible pronunciation of the word shown on the screen” (p. 2303). A response of ‘1’ indicated that the first version was better than the second, a response of ‘4’ indicated no noticeable differences between the two versions, and a response of ‘7’ indicated that the second version was better than the first. Listeners were instructed that they could use all seven points on the rating scale (see Appendix D for a complete list of instructions). This task was designed to determine if NLs judged the post-test productions of the experimental group as more native-like than pretest productions, but not those of the control group. If perceptual training had an effect on experimental-group learners’ productions, then NLs should identify these learners’ post-test productions as being more native-like more often than their pretest productions. Figure 28 shows what NLs saw on the computer screen for a test item.

fudgy, stodgy; and simplification of –ed endings in environments where simplification is not allowed in English (e.g., before a vowel as in matched only).
Listeners heard both the experimental (48 minimal pairs) and filler (28 minimal pairs) stimuli. Recall that stimuli were recorded in three contexts (isolated word, before a vowel, before a consonant), and that target words were segmented out of sentences. Because there are 152 words ([64 + 12] x 2) in each context, this resulted in a total of 456 trials for each listener. The words from each context (isolated word, before a vowel, before a consonant) were separated into three blocks. Each block began with five practice items to familiarize participants with the procedure and lasted approximately 10-12 minutes. Words were balanced such that in half of the cases, the pretest version preceded the post-test version, and in the other half, the post-test version preceded the pre-test version. Thus, for each L2 learner, two lists were created such that
in one version, the pre-test was presented first and in the other, the post-test was first. Each learner was assigned a minimum of two listeners.

Stimuli were presented via Paradigm Player. NLs either completed this task in the lab or online. Those who completed the task in the lab wore either Beyerdynamic DT 770 or Sony MDR 7506 headphones and had control over the volume level via an Alesis iO2 USB interface. NLs who completed the task online were instructed to wear headphones and complete the tasks in a quiet room.

In the data analysis, to facilitate the interpretation of the results, scores were converted from a scale of 1 to 7 to -3 to 3 such that a negative score indicated a preference for the pretest item and a positive score indicated a preference for a post-test item. Next, average scores were calculated across learners, taking into consideration that some learners had more than two NL raters. These averages are reported in the results section.

### 5.4.2 Forced-Choice Word-Identification Task

Following Bradlow et al., learners’ productions were also presented to NLs in a forced-choice word-identification task. While the paired-comparison task can tell us whether experimental group’s post-test productions were judged as more native-like than its pretest productions, it does not provide us information about whether NLs can more accurately identify learners’ post-test productions as the target word. Thus, for example, while a post-test version of *fishy* might have been judged as better than the pre-test version of *fishy*, it would remain unknown without further testing whether a NL would categorize either production as *fish* or *fishy*. 
In order to answer that question, NLs completed a forced-choice word-identification task similar to the one learners completed in the perception task. At the beginning of each trial, participants saw the two words from each pair presented on the left and right side of the screen for 500 ms. Next, a version of the word was played and participants were asked to choose the correct response as quickly as possible by pressing one of two marked keys on the keyboard (see Appendix D for complete instructions). The ‘d’ and ‘l’ keys were marked with colorful tape. The ‘d’ key indicated a response of choosing the word on the left side of the screen, and the ‘l’ key indicated a response of choosing the word on the right side of the screen.

Listeners heard both the experimental (48 minimal pairs) and filler (28 minimal pairs) stimuli. Recall that stimuli were recorded in three contexts (isolated word, before a vowel, before a consonant). Because there are 152 words ([64 + 12] x 2) in each context, this resulted in a total of 456 trials for each listener. The words from each context (isolated word, before a vowel, before a consonant) were separated into three blocks such that listeners completed three tasks. Each task began with six practice items to familiarize participants with the procedure. Stimuli were presented using E-Prime, and participants wore either Beyerdynamic DT 770 or Sony MDR 7506 headphones and had control over the volume level via an Alesis iO2 USB interface. Each task lasted approximately 10 minutes. Tasks were balanced such half the stimuli were pretest versions and half were post-test versions. Thus, for each L2 learner, two lists were created. Each L2 learner was assigned a minimum of two listeners.

5.4.3 Native Listener Participants

Native listeners were 97 (27 men and 70 women) native speakers of English who had learned only English between the ages of 0-5. Most of them were undergraduate students at the
University of Illinois at Urbana-Champaign. In addition to the production assessment tasks, each filled out a language background questionnaire (see Appendix B). A relevant subset of information from the background questionnaire is presented in Table 14.

| Table 14: Language Background Information for Experiment 3 Native Listeners |
|-------------------------------------------------|-----------------|--------|
| Native Listeners (n=97) | Daily % Usage | Age |
| Mean 98 | 24 |
| SD 4.9 | 8.6 |
| Range 70-100 | 18-61 |

Of the 97, 49 listeners completed both the forced-choice word-identification and paired-comparison tasks, but never with productions from the same learner. Of the 49 listeners who completed both tasks, 15 did so with approximately one week in between tasks. The other 34 listeners completed both tasks on the same day. The 49 listeners who completed both tasks were balanced for whether they started with the paired-comparison task or the forced-choice word-identification task. The remaining 48 listeners completed only one of the two tasks.

5.4.4 Predictions

Based on previous research implementing this type of perceptual phonetic training, it is predicted that participants in the experimental group will improve their perception of final palatals as compared to those in the control group. It is also predicted that their improvements will extend to novel words and novel talkers. In addition, it is predicted that those in the experimental group will improve their productions of final palatals in comparison to those in the
control group. The magnitude of improvement for both perception and production of learners in the experimental group will most likely vary individually, as it has in previous literature. If we return to our discussion of speech perception theories and the predictions they make regarding the relationship between perception and production, recall that the PAM, with its roots in Direct Realism, posits linked systems that share representations and would thus predict that perception and production systems would have a direct relationship and that perception and production learning would be strongly correlated. Because it assumes a psychoacoustic view of speech perception, the SLM would not predict a direct relationship between perception and production, but rather an indirect one. Therefore, unlike the PAM, according to the SLM, we would not necessarily expect to find that perception and production learning are strongly correlated. Based on the findings of Bradlow et al. (1997) and those from Experiment 1, we do not expect to find a direct relationship between improvements in perception and improvements in production.

Results from the dialog/paragraph reading task will provide a more comprehensive understanding of the current IL system of learners with regard to palatal codas. We will also be able to determine whether perceptual training/improvement on final singleton palatals will have an effect on the productions of these palatals in a wider variety of contexts. Based on previous research it is difficult to determine whether training will lead to improvements in these extended contexts; however, previous research with /r/ and /l/ has demonstrated that training with these segments in certain contexts can have benefits for the production of these sounds in other contexts.
5.5 Results

Results are presented in the subsections below, beginning with the perception tasks, followed by the production tasks and ending with a comparison of perception and production.

5.5.1 Perception Results

First, let us consider the results for improvements in perception in the isolated-word context. Figure 29 presents the pretest and post-test d’ scores that each group obtained on the perception task in the isolated word context. To determine whether the experimental group improved more than the control group, a mixed-design repeated-measures ANOVA was performed with test (pretest, post-test) as within-subject variable and group (experimental, control) as between-subject variable. There was a main effect of test, $F(1, 22)=22.81, p<.001$, but no effect of group ($F<1$). The interaction between test and group did not quite reach significance, $F(1,22)=2.44, p<.132$. Thus, in the isolated-word context, both the experimental and control groups improved. I return to this point in the discussion section.
Next, results from the experimental group in isolated words were analyzed to determine whether participants were able to generalize to new words and new talkers. Because the concept of new words and new talkers does not exist for the control group (they heard all words in the pretest/post-test and were not trained on these words, and their training consisted of a different set of voices), it is not possible to conduct test of generalizability for them.

In order to determine whether learners are able to generalize to new words and new talkers, we compare their pretest and post-test $d'$ scores on four categories of words: (1) new words spoken by new talkers, (2) new words spoken by talkers from the training, (3) words from the training spoken by new talkers, and (4) words from the training spoken by talkers from the training. If their improvements are the same for all four categories, then we can say that learners were able to generalize to new words and new talkers because they improved equally in all categories. On the other hand, if it is found, for example, that learners improved more on the

![Figure 29: Pretest and post-test perception scores by group for the isolated word context](image)
categories of words spoken by words or talkers used in the training, then we might conclude that learners were not able to generalize to new words or talkers.

Figure 30 shows the pretest and post-test $d'$ scores for palatal codas in isolated words separated by new and trained talkers and new and trained words for the experimental group.

**Figure 30: Word/Talker generalization in the perception task for palatal codas in isolated words for the experimental group**

A repeated-measures ANOVA was performed with test (pretest, post-test), word (novel, trained), and talker (novel, trained) as within-subject variables. There were main effects of test, $F(1, 11)=5.33, p<.041$, and word, $F(1,11)=15.06, p<.003$, but no main effect of speaker ($F<1$), nor interactions between word and speaker, $F(1,11)=1.57, p<.236$, speaker and test, $F(1,11)=2.45, p<.146$, word and test ($F<1$), or word and test and speaker ($F<1$). Despite the numerical tendency for the new word-new talker condition to receive lower $d'$ scores in the post-test, the results indicate that learners performed better on the post-test than on the pre-test across
words and talkers. Furthermore, they suggest that learners performed better on trained words than on new words across pre- and post-tests and across talkers, with the trained words being perhaps intrinsically easier than the new words. The lack of interaction between test and either word or talker indicates that learning was not modulated by either word or talker. Hence, we can conclude from these results that learning was generalized to new words and new talkers.

Next, we consider the results of the perception tasks for the sentence context. Figure 31 presents the pretest and post-test $d'$ perception scores of each group with regard to palatal codas in sentences. In order to determine whether there was a difference between the experimental and control group, a mixed-design repeated-measures ANOVA was performed with test (pretest, post-test) as within-subject variable and group (experimental, control) as between-subject variable. There was a main effect of test, $F(1, 22)=23.62, p<.001$, and an interaction between test and group, $F(1, 22)=9.30, p<.006$, but no main effect of group, $F(1,22)=2.57, p<.123$.

![Figure 31: Pretest and post-test perception scores by group for the sentence context](image-url)
Post-hoc Bonferroni comparisons were conducted with alpha levels adjusted to $p<.025$. A paired-samples $t$-test showed no significant difference between pretest and post-test for the control group, $t(11) = -1.50$, $p<.161$. There was, however, a significant difference between the pretest and post-test scores for the experimental group, $t(11) = -4.096$, $p<.001$. Thus, the experimental group showed significant improvement between the pretest and post-test for palatal codas in sentences, but the control group did not.

Thus, perceptual phonetic training on palatal codas had a significant effect on the experimental group, improving their perceptions of palatal codas in both isolated words and sentences. The control group, on the other hand, improved only in perception of palatal codas in isolated words. I return to this finding in the discussion section below.

Next, results from the experimental group were analyzed to determine whether participants were able to generalize to novel words and novel talkers. Recall that in order to determine whether learners are able to generalize to new words and new talkers, we compare their pretest and post-test $d'$ scores on four categories of words: (1) new words spoken by new talkers, (2) new words spoken by talkers from the training, (3) words from the training spoken by new talkers, and (4) words from the training spoken by talkers from the training. If their improvements are the same for all four categories, then we can say that learners were able to generalize to new words and new talkers because they improved equally in all categories.

Figure 32 shows the improvement results in the sentence context for the experimental group for palatal codas.
A repeated-measures ANOVA was performed with word (novel, trained), talker (novel, trained), and test (pretest, post-test) as within-subject variables. There was a main effect of test, $F(1, 11)=27.43$, $p<.001$, but no other main effects or interactions. Thus, learners improved equally between pretests/post-tests on all four categories of words. Therefore, we conclude that participants were able to generalize learning to new words and new talkers. Recall that in the isolated-word context, learners also demonstrated generalization to new words and new talkers, but that trained words may have been intrinsically easier than new words. Whatever advantage these words had in isolation, it did not persist in the sentence context.

The next question relates to the environment that follows the palatal coda words. Recall that the sentence items contained target words in the phrases ‘He said X angrily’ and ‘He said X frequently.’ Here, we test whether sentence context (before a vowel or before a consonant) had any effect on the experimental group’s perception improvements. Recall that we predicted the
prevocalic condition to be easier for learners in that it would allow them to parse the palatals as the onset of the following syllable, thus not violating any syllable structure constraints of their L1. Figure 33 displays experimental group’s $d'$ scores for the pretest and post-test in the prevocalic and pre-consonantal conditions.

![Figure 33: Pretest and post-test perception scores reported in $d'$ for the sentence context separated for whether the following word was angrily or frequently](image)

A repeated-measures ANOVA was performed on the pretest and post-test perception data with test (pretest, post-test) and sentence context (angrily, frequently) as within-subject variables. The analysis revealed main effects of test, $F(1,11)=23.16$, $p<.001$, sentence context, $F(1,11)=31.81$, $p<.001$, and an interaction between test and sentence context, $F(1,11)=13.42$, $p<.004$. Post-hoc Bonferroni comparisons were conducted with alpha levels adjusted to $p<.025$. Paired-samples $t$-tests showed significant differences between the pretest and post-test for the angrily context, $t(11)= -2.97$, $p<.013$, as well as for the frequently context, $t(11)= -5.25$, $p<.001$. 

132
Thus, the experimental group improved on perceptions of palatal codas in both the *frequently* and the *angrily* contexts.

One final question we can address from the perception results relates to whether word familiarity might be playing a role in the identification of these words. Recall that there were potential word familiarity effects in Experiment 2. That motivated the inclusion of both real and nonce words in the stimuli for the current experiment. Thus, we want to determine whether there were differences in pretest and post-test identifications based on whether words were real or nonce. If we find similar results for real vs. nonce words, then we can conclude that word familiarity does not play a crucial role in the perception of these pairs, because learners have no previous experience, and thus no familiarity, with the nonce words. Figure 34 presents the experimental group’s pretest and post-test *d’* scores in isolated words separated by real vs. nonce words.
Figure 34: Pretest and post-test perception scores reported for the experimental group for the isolated-word context separated by real vs. nonce words

A repeated-measures ANOVA with test (pretest, post-test) and lexical status (real, nonce) as within-subject variables showed a main effect of test, $F(1,11)=17.19$, $p<.05$, but no main effect of lexical status ($F<1$), and no interaction between test and lexical status, ($F<1$). Thus, learners are performing equally well in their perception of real and nonce words in the isolated-word context. Figure 35 presents the experimental group’s pretest and post-test $d'$ scores in the sentence context separated by real vs. nonce words.
A repeated-measures ANOVA with test (pretest, post-test) and lexical status (nonce, real) as within-subject variables showed a main effect of test, $F(1,11)=22.99, p<.001$, but no main effect of lexical status ($F<1$), and no interaction between test and lexical status, $F(1,11)=2.51, p<.141$. Similar to the perception results for real vs. nonce words in isolated words, in the sentence context, learners’ perceptions are improving equally for both real and nonce words. Thus, it does not appear that word familiarity was affecting perception results in Experiment 3.

In summary, for isolated words, both the experimental group and the control group performed significantly better on the perception post-test than on the perception pretest. The experimental group also showed evidence of generalizing perception learning to new words and new talkers in the isolated-word context, although the trained words appeared to have been intrinsically easier in the isolated-word context.
Unlike the perception results in isolated words, the experimental group, but not the control group, showed significant perception improvement between the pretest and post-test in the sentence context. These findings suggest a beneficial effect of perceptual phonetic training on codas for words in sentences. In addition, we found no correlation between perception pretest and perception improvement scores for the sentence context. Results of generalizability for the sentence context indicated that the experimental group was able to generalize learning to novel words and novel talkers, suggesting that the perceptual training was successful in allowing learners to establish more robust representations. An investigation of whether the context following the target word (angrily, frequently) affected perception indicated that participants in the experimental group showed improvements between the pretest and post-test in both the pre-consonantal and the prevocalic context, suggesting that context did not affect perceptions.

Finally, no effect of lexical status was found for the perception results of the experimental group in isolated words or sentences, indicating that word familiarity did not influence the results. We now turn our attention to the production results.

5.5.2 Production Results

The following subsections report on results from the two production tasks. I begin with results from the read-aloud task modeled on the perception task, followed by the dialog/paragraph reading task.

5.5.2.1 Paired-Comparison Task

As in the results for the perception tasks, data are first presented for words in isolation and then for words in sentence contexts. Recall that for this task, listeners were presented with
both the pretest and post-test productions of a learner and judged which was more native-like; thus, as listeners chose only one word, results cannot be separated by pretest and post-test. Also, recall that the scores from the paired-comparison task were converted such that a score above 0 indicated the post-test production was more native-like and a score below 0 indicated that the pre-test production was more native like. The scores were then averaged by participant for the control and experimental groups. Figure 36 displays the results for the production of palatal codas in isolated words.

![Figure 36: Paired-comparison results for the production of palatal codas in the isolated-word context by group](image)

As we can see in Figure 36, it appears that the average rating for the experimental group’s productions in isolated words was higher than that for the control group. To determine whether the experimental group’s post-test productions were rated more native-like than the control group’s productions, an independent-samples t-test was performed on the NL paired-comparison ratings for the experimental and control groups. The t-test showed a significant
difference between groups, $t(22)=2.76, p<.011$. Thus, more of the experimental group’s post-test productions of palatal codas in isolated words were rated more native-like in comparison to those of the control group.

Now let us consider the results for productions of palatal codas in the sentence context shown in Figure 37.

![Figure 37: Paired-comparison results for productions of palatal codas in the sentence context by group](image)

We again see that the average rating for the experimental group’s productions was higher than that for the control group. To determine whether the experimental group’s post-test productions were rated more native-like than the control group’s productions, an independent-samples $t$-test was performed on the NL paired comparison ratings for the experimental and control groups. The $t$-test showed a significant difference between groups, $t(22)=3.05, p<.006$. Thus, similarly to the results for productions in isolated words, more of the experimental group’s
post-test productions of palatal codas in the sentence context were rated more native-like in comparison to those of the control group.

In summary, the above results indicate that for both isolated words and sentences, more of the post-test productions of the experimental group were judged more native-like in comparison to those of the control group. These results indicate that the perceptual phonetic training on palatal codas helped learners improve their production of palatals, and improvements, as measured by the paired-comparison task, were minimal for those in the control group. We now turn our attention to the results of the word-identification task to determine whether NLs were able to categorize the experimental group’s post-test productions more accurately.

5.5.2.2 Forced-Choice Word-Identification Task

Data are first presented for the isolated words and then for words in sentences. Figure 38 displays the $d'$ scores of NL’s ratings of the pretest and post-test productions in the isolated-word context for both the experimental and control groups.
Figure 38: Production assessment results from the forced-choice task separated by pretest/post-test for the isolated-word context by group

A mixed-design repeated-measures ANOVA was performed with test (pretest, post-test) as within-subject variable and with group (experimental, control) as between-subject variable. There was a main effect of test, $F(1, 22) = 28.11, p < .001$, and an interaction between test and group, $F(1, 22) = 6.00, p < .023$, but no main effect of group, $F(1, 22) = 1.06, p < .314$. Post-hoc Bonferroni comparisons were conducted with alpha levels adjusted to $p < .025$. A paired-samples $t$-test showed no significant difference between pretest and post-test for the control group, $t(11) = -1.63, p < .131$. There was, however, a significant difference between the pretest and post-test scores for the experimental group, $t(11) = -4.67, p < .001$. Thus, NLs were able to accurately identify more post-test productions of the experimental group for palatal codas in isolated words. The same was not true for the control group.
Next, we consider the data from the productions of palatal codas in sentences. Figure 39 displays the $d'$ scores of NL’s ratings of pretest and post-test productions in the sentence context for both the experimental and control groups.

![Figure 39: Production assessment results from the forced-choice task separated by pretest/post-test for the sentence context by group](image)

A mixed-design repeated-measures ANOVA was performed with test (pretest, post-test) as within-subject variable and with group (experimental, control) as between-subject variable. Similar to the results for isolated words, there was a main effect of test, $F(1, 22)=15.07, p<.001$, and an interaction between test and group, $F(1,22)=5.91, p<.024$, but no main effect of group, $F(1, 22)=3.84, p<.063$. Post-hoc Bonferroni comparisons were conducted with alpha levels adjusted to $p<.025$. Again, a paired-samples $t$-test showed no significant difference between pretest and post-test for the control group, $t(11)= -2.55, p<.027$. There was, however, a significant difference between the pretest and post-test scores for the experimental group, $t(11)= 
-3.53, \( p < .005 \). Thus, NLs were able to accurately identify more post-test productions of the experimental group for palatal codas in sentences. The same was not true for the control group.

In summary, for palatal codas in both isolated words and sentences, NLs rated more of the experimental group’s post-test productions as more native-like, which was not the case for the control group. In addition, for both contexts, NLs were able to accurately identify more of the experimental group’s post-test productions, which was not the case for the control group. These findings suggest that the experimental group’s productions improved more than the control group’s productions, thus implying a beneficial effect of perceptual phonetic training on the production of on palatal codas.

5.5.2.3 Dialog/Paragraph Reading Task

The results of the dialog/paragraph reading task are meant to help answer questions related to production in larger discourse contexts as well as palatal codas beyond just the singleton coda context. Recall that this task included palatals in complex codas (e.g., perch, mulch, sponge), palatals in simple codas but in disyllable words like spinach, and palatals before -ed endings (but not those that undergo consonant cluster simplification). Results from this task will allow first for comparison to results in the read-aloud task modeled on the perception task. In addition, results from this task will allow for a more exploratory analysis of the larger interlanguage system of Korean L2 learners of English with regard to palatals. I first report the results from the dialog/paragraph production task that parallel the words in focus from the perceptual training (i.e., those with singleton word-final palatals, push, and singleton palatal + [i], pushy), as seen in Figure 40. Recall that unlike the results from the paired-comparison task
and the forced-choice word-identification task, these productions were assessed as either correct or incorrect based on the ratings of two trained English pronunciation teachers.

![Graph of pretest and post-test accuracies for palatal coda words separated by group.]

**Figure 40: Dialog/paragraph production task pretest and post-test accuracies for palatal coda words separated by group**

As we can see in Figure 40, it appears the experimental group improved their productions of palatal coda words, but the control group did not. A mixed-design repeated-measures ANOVA was performed on the accuracy scores with test (pretest, post-test) as within-subject variable and with group (experimental, control) as between-subject variable. Test had a significant effect, $F(1,22)=26.13, p<.001$, and there was an interaction between test and group, $F(1,22)=8.33, p<.01$. There was not a significant main effect for group ($F<1$). Post-hoc Bonferroni comparisons were conducted with alpha levels adjusted to $p<.025$. A paired-samples $t$-test showed no significant difference between pretest and post-test for the control group, $t(11)=-2.06, p<.064$. There was, however, a significant difference between the pretest and post-test
scores for the experimental group, \( t(11) = -4.75, p < .001 \). Thus, the experimental group showed significant improvement between the pretest and post-test on the production of palatal codas in the dialog/paragraph reading task, but the control group did not. These results suggest that perceptual training had more of a beneficial effect for the experimental group on palatal codas in larger discourse contexts.

Next, we turn our attention to the phonological contexts beyond those present in the perception tasks of Experiment 3. Figure 41 presents the experimental and control groups’ pretest and post-test results for the two word types unique to this production measure: disyllabic words ending in a palatal (e.g., spinach) and palatal words with –ed ending morphemes (e.g., pushed).

![Figure 41: Dialog/paragraph production task pretest and post-test accuracies for the disyllabic final-palatal and –ed ending words separated by group](image)

Figure 41: Dialog/paragraph production task pretest and post-test accuracies for the disyllabic final-palatal and –ed ending words separated by group
As we can see, learners are producing disyllabic palatal-final words relatively accurately overall, but there do not appear to be changes between the pretest and post-test for either groups. On the other hand, learners appear to be having more difficulty with palatal codas before –*ed* ending words.

A mixed-design repeated-measures ANOVA was performed on the accuracy scores with test (pretest, post-test) as within-subject variable, and with group (experimental, control) as between-subject variable for both the disyllabic palatal words and the –*ed* ending words. For the disyllabic palatal words, there were no effects of test or group and no interaction between test and group (*F*<1). Thus, neither group improved. Similarly, for –*ed* ending words there was no effect of test, *F*(1,22)=3.30, *p*<.083, or group (*F*<1) and no interaction between test and group (*F*<1). This suggests that learning did not generalize to these two phonological contexts that were not present in the perceptual training.

Next, we investigate whether the pre-palatal environment in the word (i.e., a vowel or the consonant /n l ɹ/ in words like push, pinch, squelch and perch) had an effect and thus whether simple vs. complex codas affected the production results. Figure 42 displays the pretest and post-test results for all four word types separated by pre-palatal environment for the experimental and control groups.
Figure 42: Dialog/paragraph production task pretest and post-test accuracies separated by pre-palatal environment

As we can see in Figure 42, pre-palatal environment does not seem to have an effect on improvement in that learners in the experimental group appear to be improving equally in both environments. A mixed-design repeated-measures ANOVA was performed on the production accuracy with test (pretest, post-test) and pre-palatal environment (vowel, consonant) as within-subject variables, and with group (experimental, control) as between-subject variable. Test had a significant effect, $F(1,22)=21.33, p<.001$, and there was an interaction between test and group, $F(1,22)=6.55, p<.05$. There was not a significant effect for group ($F<1$), nor was there a significant effect for pre-palatal environment, $F(1,22)=1.20, p<.285$, or any other interactions ($F<1$). These results suggest that pre-palatal environment did not affect production results, which indicates that improvements occurred similarly in simple and complex codas.

Next, we investigate whether environment (before a consonant, before a vowel, phrase-final) had an effect on production improvements for palatal codas. Recall that in the
dialog/paragraph production measure, the context following the target palatal word was balanced for whether the word that followed the target word began with a vowel, a consonant, or whether there was no word (making the target word phrase-final). Figure 43 displays the results for all four word types separated by environment for both groups of learners.

![Figure 43: Dialog/paragraph production task results pretest and post-test accuracies separated by whether the target word appeared before a consonant, before a vowel, or in phrase-final position](image)

As we can see in Figure 43, context did not appear to affect results in that learners in the experimental group seem to be improving equally in all three environments. A mixed-design repeated-measures ANOVA was performed on the accuracy scores with test (pretest, post-test) and environment (before a consonant, before a vowel, phrase-final) as within-subject variables, and with group (experimental, control) as between-subject variable. Test had a significant effect, $F(1,22)=21.80, p<.001$, and there was an interaction between test and group, $F(1,22)=6.44,$
There was not a significant effect for group ($F<1$), nor was there a significant effect for environment or any other interactions ($F<1$). These results suggest that environment did not affect production results, which is in line with the results from the read-aloud task modeled on the perception task, in which learners improved in both the pre-consonantal and prevocalic environments.

Results from the dialog/paragraph reading task provide preliminary indications that perceptual training has a beneficial effect on words in larger discourse contexts. The second goal of this task was to gain more information about the developing IL system of Korean L2 learners of English with regard to palatal codas. It was found that learning did not extend to disyllabic words final-palatal words (e.g., *spinach*), or to those with –ed ending morphemes. It was also found that pre-palatal environment did not appear to have an effect on production results, demonstrating that perceptual phonetic training on singleton palatal codas extended to production improvements on complex codas. Finally, similar to findings from the perception tasks, which demonstrated improvements in both the prevocalic and pre-consonantal positions, results from the dialog/paragraph task indicated no effect of following context (before a consonant, before a vowel, phrase-final). I return to this discussion in the final chapter of the dissertation.

5.5.3 Individual Variability and the Relationship between Perception and Production

In this subsection, we consider the relationship between changes in perception and changes in production for the experimental group with regard to palatal codas in both isolated words and sentences. Production scores represent the NL judgments from the forced-choice word-identification task or, in other words, NL accuracy in identifying learner productions.
Production scores from the word-identification task are a more appropriate measure than production scores from the paired-comparison task: Recall that the paired-comparison task was designed to provide indications of whether post-test productions were more native-like than pretest productions, but not necessarily to indicate word accuracy. It could be the case that a post-test production was rated as more accurate than a pretest production, but still not categorized accurately in the word-identification task. NL judgments from the word-identification task give us an indication of whether a learner went from inaccurate on the pretest to accurate on the post-test. They are thus the more appropriate measure for purposes of comparing changes in perception and production improvements.

Table 15 shows the pretest, post-test, and improvement scores (as $d'$ scores values) for each participant for palatal codas in isolated words. Improvement scores were calculated by subtracting pretest scores from post-test scores: $d'$ (improvement) = $d'$ (post) – $d'$ (pre). Columns on the left report perception results and columns on the right report production results.
Table 15: Pretest, Post-test and Improvement Scores (as $d'$ Scores) by Participant for Palatal Codas in Isolated Words

<table>
<thead>
<tr>
<th></th>
<th>Perception</th>
<th>Production</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pretest</td>
<td>Post-test</td>
<td>Improvement</td>
<td>Pretest</td>
<td>Post-test</td>
</tr>
<tr>
<td>5</td>
<td>0.48</td>
<td>2.56</td>
<td>2.08</td>
<td>-0.10</td>
<td>4.50</td>
</tr>
<tr>
<td>7</td>
<td>2.02</td>
<td>4.20</td>
<td>2.17</td>
<td>2.81</td>
<td>3.88</td>
</tr>
<tr>
<td>8</td>
<td>0.21</td>
<td>2.88</td>
<td>2.67</td>
<td>0.12</td>
<td>2.62</td>
</tr>
<tr>
<td>9</td>
<td>2.58</td>
<td>3.65</td>
<td>1.07</td>
<td>2.22</td>
<td>3.17</td>
</tr>
<tr>
<td>10</td>
<td>0.42</td>
<td>3.09</td>
<td>2.66</td>
<td>-0.14</td>
<td>3.63</td>
</tr>
<tr>
<td>11</td>
<td>0.27</td>
<td>0.62</td>
<td>0.34</td>
<td>-0.11</td>
<td>0.25</td>
</tr>
<tr>
<td>13</td>
<td>1.35</td>
<td>3.07</td>
<td>1.72</td>
<td>1.79</td>
<td>2.65</td>
</tr>
<tr>
<td>15</td>
<td>0.56</td>
<td>0.71</td>
<td>0.16</td>
<td>0.33</td>
<td>0.49</td>
</tr>
<tr>
<td>27</td>
<td>2.64</td>
<td>2.95</td>
<td>0.31</td>
<td>3.25</td>
<td>5.12</td>
</tr>
<tr>
<td>30</td>
<td>4.65</td>
<td>3.96</td>
<td>-0.70</td>
<td>2.56</td>
<td>4.17</td>
</tr>
<tr>
<td>31</td>
<td>1.94</td>
<td>4.20</td>
<td>2.26</td>
<td>1.33</td>
<td>4.17</td>
</tr>
<tr>
<td>32</td>
<td>2.56</td>
<td>3.50</td>
<td>0.94</td>
<td>2.33</td>
<td>3.73</td>
</tr>
<tr>
<td>Mean</td>
<td>1.64</td>
<td>2.95</td>
<td>1.31</td>
<td>1.37</td>
<td>3.20</td>
</tr>
</tbody>
</table>

As we can see in Table 15, all but one participant (P30) showed improvement in the perception of palatal codas in isolated words. However, note that for this participant, his pretest
score was the highest among the group. We also see considerable variation among participants in terms of both their pretest scores and their amount of improvement for perception.

When we consider production scores, we see that all learners improved on their production of palatal codas in isolated words. Similar to perception results, we also see considerable variation in the amount of improvement in production. For example, participants 10 and 11 had similar pretest scores (-0.14 and -0.11, respectively), but showed major differences in their amount of improvement (3.77 and 0.36, respectively). On the other hand, participants 9 and 32 showed similar pretest scores (2.22 and 2.33, respectively) and somewhat similar improvement scores (0.95 and 1.39, respectively).

Before investigating the relationship between perception and production, it is relevant to examine the relationship between pretest scores and improvements scores because of the individual variability found. Given the individual variability observed in Table 15, we might hypothesize that learners who obtained high pretest scores may not have been able to improve their perception and production performance as much as learners who obtained lower pretest scores did. In other words, we can examine whether there is a relationship between the initial perceptual and production abilities that learners had and how much they improved as a result of perceptual training. Thus, we would predict a negative correlation between pretest scores and the amount of improvement. Figure 44 plots learners’ perception pretest scores and perception improvement scores for palatal codas in isolated words.
A correlation analysis between the perception pretest scores and perception improvement scores for the isolated-word context did not quite reach significance ($r = -0.548, p < 0.065$). Thus, it was not the case that pretest scores modulated the amount of improvement for the perception of palatal codas in isolated words.

Similar to the perception analysis, we can also investigate whether there is a connection between pretest scores and amount of improvement for production. Figure 45 plots learners’ production pretest scores and production improvement scores for palatal codas in isolated words.
Figure 45: Scatterplot comparing learners’ production pretest and production improvement scores (reported in $d'$) for palatal codas in isolated words

Similar to the results for the perception, a correlation analysis between the production pretest scores and production improvement scores from the isolated-word context did not reach significance ($r = -.356, p < .256$). Thus, similar to perception results, it was not the case that pretest scores modulated the amount of improvement for the production of palatal codas in isolated words.

Next, we turn our attention to examining the relationship between perception and production. Figure 46 plots learners’ perception and production improvement scores for palatal codas in isolated words.
Figure 46: Scatterplot comparing learners’ perception and production improvement scores (reported in $d'$) for palatal codas in isolated words

A correlation analysis between the improvement scores on the perception task and the improvement scores on the word-identification measure (production) revealed a moderate coefficient ($r=.581, p < .05$). Thus, we see a moderate relationship between improvements in perception and improvements in production for palatal codas in isolated words. These results are unlike those from Experiment 1, where no correlation was found between perception and production accuracy scores.

Table 16 shows the pretest, post-test, and improvement scores (as $d'$ scores) for each participant for palatal codas in the sentence context. Columns on the left report perception results and columns on the right report production results.
Table 16: Pretest, Post-test and Improvement Scores (as $d'$ Scores) by Participant for Palatal Codas in Sentences

<table>
<thead>
<tr>
<th>Participant</th>
<th>Perception Pretest</th>
<th>Perception Post-test</th>
<th>Perception Improvement</th>
<th>Production Pretest</th>
<th>Production Post-test</th>
<th>Production Improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>0.21</td>
<td>0.84</td>
<td>0.63</td>
<td>0.11</td>
<td>3.47</td>
<td>3.36</td>
</tr>
<tr>
<td>7</td>
<td>0.16</td>
<td>2.11</td>
<td>1.95</td>
<td>2.52</td>
<td>3.50</td>
<td>0.98</td>
</tr>
<tr>
<td>8</td>
<td>0.24</td>
<td>1.53</td>
<td>1.29</td>
<td>0.30</td>
<td>3.53</td>
<td>3.23</td>
</tr>
<tr>
<td>9</td>
<td>0.37</td>
<td>1.66</td>
<td>1.30</td>
<td>1.69</td>
<td>3.59</td>
<td>1.90</td>
</tr>
<tr>
<td>10</td>
<td>0.48</td>
<td>1.19</td>
<td>0.71</td>
<td>0.71</td>
<td>4.36</td>
<td>3.65</td>
</tr>
<tr>
<td>11</td>
<td>0.06</td>
<td>0.06</td>
<td>0.01</td>
<td>0.20</td>
<td>0.57</td>
<td>0.37</td>
</tr>
<tr>
<td>13</td>
<td>0.72</td>
<td>2.13</td>
<td>1.41</td>
<td>1.90</td>
<td>2.90</td>
<td>1.00</td>
</tr>
<tr>
<td>15</td>
<td>-0.22</td>
<td>0.00</td>
<td>0.22</td>
<td>-0.14</td>
<td>-0.02</td>
<td>0.12</td>
</tr>
<tr>
<td>27</td>
<td>1.38</td>
<td>3.06</td>
<td>1.68</td>
<td>3.48</td>
<td>4.45</td>
<td>0.97</td>
</tr>
<tr>
<td>30</td>
<td>2.72</td>
<td>2.66</td>
<td>-0.06</td>
<td>2.63</td>
<td>2.87</td>
<td>0.23</td>
</tr>
<tr>
<td>31</td>
<td>1.29</td>
<td>2.12</td>
<td>0.83</td>
<td>2.31</td>
<td>3.95</td>
<td>1.64</td>
</tr>
<tr>
<td>32</td>
<td>2.33</td>
<td>3.48</td>
<td>1.15</td>
<td>3.50</td>
<td>2.93</td>
<td>-0.58</td>
</tr>
</tbody>
</table>

Similar to the results for perception of palatal codas in isolated words, all but one participant (the same as in isolated words, P30) showed improvement in the perception of palatal codas in sentences. Again, note for this participant that his pretest score was the highest among
the group. Also similar to the results for the perception of palatal codas in isolated words, we see considerable variation among participants in terms of both their pretest scores as well as their amount of improvement. When we consider production scores, we also see similar results to palatal codas in isolated words in that all but one participant (not the same participant as before, P32) showed improvements. We also see variability in the amount of improvement on production.

As with the isolated-word context, we can also investigate whether the amount of improvement is related to the pretest scores of participants for perception in sentences. Figure 47 plots learners’ perception pretest scores and perception improvement scores for palatal codas in sentences.

![Figure 47: Scatterplot comparing learners’ perception pretest and perception improvement scores (reported in $d'$) for palatal codas in sentences](image-url)
A correlation analysis between the perception pretest scores and perception improvement scores for the sentence context did not reach significance ($r = -.087, p < .788$). Thus, similar to the isolated-word context, it was not the case that pretest scores modulated improvement for perceptions of palatal codas in sentences.

As with the isolated-word context, we can also investigate whether the amount of improvement is related to the pretest scores of participants for production in sentences. Figure 48 plots learners’ production pretest scores and production improvement scores for palatal codas in sentences.

![Figure 48: Scatterplot comparing learners’ production pretest and production improvement scores (reported in $d'$) for palatal codas in sentences](image)

Figure 48: Scatterplot comparing learners’ production pretest and production improvement scores (reported in $d'$) for palatal codas in sentences

Similar to perception results, a correlation analysis between the production pretest scores and production improvement scores for the sentence context did not reach significance ($r = -.480,$
Thus, again, it was not the case that pretest scores modulated improvement for productions of palatal codas in sentences.

We again turn our attention to an examination of the relationship between perception and production. Figure 49 plots learners’ perception and production improvement scores for palatal codas in sentences.

![Figure 49: Scatterplot comparing learners’ perception and production improvement scores (reported in d') for palatal codas in the sentence context](image)

A correlation analysis between the improvement scores on the perception task and the improvement scores on the word-identification measure (production) did not reach significance ($r=.139, p<.666$). Thus, similar to the findings in Experiment 1 but unlike those for palatal codas in isolated words, we do not see a direct relationship between improvements in perception and improvements in production for palatal codas.
In summary, we found a moderate correlation between perception and production improvements for palatal codas in isolated words, but no correlation between perception and production improvements for palatal codas in sentences. We found no correlation between perception pretests and improvements in perception for either the isolated-word or sentence contexts, nor did we find any correlation between production pretests and improvements in production for either the isolated-word or sentence contexts. The findings from Experiment 3 are generally in line with findings from Experiment 1 and those from previous research that do not provide compelling evidence for a direct link between perception and production systems. Nevertheless, the fact that perceptual phonetic training led to improvements in production and the moderate correlation between perception and production improvements in the isolated-word context implies some sort of relationship. Ultimately, the null result of not finding a consistent, strong correlation between perception and production is difficult to interpret because there could be other explanations for this lack of a finding. Thus, while we do not have evidence against a direct link between perception and production, we do not have compelling evidence for it either.

5.6 Discussion

Experiment 3 was designed to answer the following questions. I consider each in turn in the following subsections.

1. Can pedagogically viable perceptual phonetic training on palatal codas improve perception accuracies of palatal codas?

2. Does perceptual phonetic training on palatal codas allow generalization to new words and new talkers?
3. What will be the effects, if any, of perceptual phonetic training on palatal codas on productions of palatal codas?

4. What is the relationship, if any, between improvements on perceptions and productions of palatal codas?

5. In which contexts (words in isolation, words within a larger discourse, final singleton palatals, final palatal clusters, palatals before –ed morphemes, etc.) do learners have the most difficulty with palatals?

5.6.1 Can pedagogically viable perceptual phonetic training on palatal codas improve perception accuracies of palatal codas?

The results from Experiment 3 indicated that the experimental group demonstrated improved perception accuracies of palatal codas in both isolated words and sentences. The results for the control group showed improvement between pretest and post-test perception scores for isolated words, but not in the sentence context. Thus, it appears that the perceptual phonetic training on palatal codas used in this experiment was beneficial for improving perceptions of palatal codas in the sentence context.

The fact that the control group improved in the isolated-word context requires further discussion. First of all, we should note that in previous training studies of this kind, the control group completed the perception and production pretests and post-tests, but did not complete a task in between the tests. In Experiment 3, the control group completed a comparable perceptual phonetic training task but focused on an unrelated target, tense/lax vowels. One possible explanation for their improvement in perceptions in the isolated-word context is that the training they received on tense/lax vowels was somehow beneficial for the palatal codas that were the
focus of this study. Recall from the design of the control group training that stimuli were presented in isolated words and were monosyllabic nonce words of the form CVC with the V representing vowels in the tense/lax pairs [æ~ɛ], [i~ɪ], and [o~ʊ], and with the C representing the following consonants /d, t, n, b, p, m, k g, h, s/. We can see from the tense/lax vowel pairs that /i/ was included. The training on this vowel might have led to improvements in the palatal words because the minimal pairs were palatal words that were differentiated by the presence or absence of a final [i] vowel. Thus, training and improvement on this vowel could have resulted in improved differentiations of push-pushy minimal pairs. We can also note that the stimuli in the control group training contained codas produced in a ‘careful’ speech rate style in which the final codas were released. Although there was no training on palatal codas, the fact that the stimuli for the control group training contained consonants in coda position that are not allowed in Korean may have led to improved perceptions of this syllable structure that is restricted in Korean for words presented in isolation. It is also interesting to note that the control demonstrated improvement in isolated words (but not sentences) despite having spent quite a bit less time on perceptual training (68 minutes vs. 160 minutes over the course of 8 days).

Another contributing factor to the control group improvement in perception in the isolated-word context might relate to the relatively long pretest/post-test phases in conjunction with the relatively transparent focus of the experiment. As described in the materials and procedure sections in this chapter, the pretests and post-tests contained a large number of items produced by six talkers. In addition, only 28 of 76 word pairs in the experiment focused on something other than palatals. The length of the perception and production pretests/post-tests was approximately 1.5-2 hours. In other words, the pretests/post-tests themselves may have acted as a form of training. When we consider the perceptual training results of the experimental group
(see Appendix I), we notice that improvements in the isolated-word context plateau relatively quickly. This occurs after two days of approximately 40 minutes of training in sentences and isolated words which amounts to having completed two isolated word blocks with a total of 96 stimuli and two sentence blocks with a total of 192 stimuli. Improvements in the sentence context, on the other hand, take longer to plateau (approximately 4-5 days). While it may be the case that the explanations just discussed can provide some insight into why the control group showed improvements in their perceptions in the isolated-word context, it remains the case that they did not demonstrate improvements in the sentence context. The experimental group, on the other hand, showed significant improvements between the pre- and post-tests in both the isolated-word and the sentence contexts.

When we considered sentence context, or whether the target word was followed by a vowel or consonant (*angrily*/frequently), we found that the experimental group showed improvements in perception for both the pre-consonantal and the prevocalic context. We also considered whether word familiarity might be playing a role in the perception of these word pairs. There was no difference between improvements for real and nonce words in either the isolated-word or the sentence context. Thus, word familiarity did not affect perception. Despite this finding, word familiarity has been shown to affect perceptions in other studies (e.g., Flege, Takagi & Mann, 1996). In their study, learners were better able to identify segments in familiar words than in unfamiliar ones. This allows for the possibility that learners are not necessarily attending to acoustic cues in these cases, but rather that top-down information influences decisions (MacKay, 1987). The lack of familiarity effects in Experiment 3 suggests that the strategies that L2 learners used were similar for both real and nonce words, and thus cannot be attributed to lexical factors.
5.6.2 Does perceptual phonetic training on palatal codas allow generalization to new words and new talkers?

Recall that we considered the underlying mechanisms that might account for why perceptual phonetic training might result in generalizability from both episodic-trace and abstractionist points of view. Within episodic-trace theories, learners store exemplars in memory with detailed phonetic and non-linguistic information. Receiving multiple tokens of input from multiple talkers would yield a greater number of exemplars stored in memory, thus resulting in more robust or stable representations. We also discussed the possibility that if episodic-trace theories are correct, we would find better performance for trained words for the experimental group because these are the exemplars that would be stored. On the other hand, abstractionist theories posit abstract prelexical representations that mediate recognition. Receiving varied input from multiple talkers would strengthen these abstract representations and allow for generalization from this input to new words and new talkers. The findings from Experiment 3 demonstrated that learners in the experimental group were able to generalize learning to both new talkers and new words in both the isolated words and sentence contexts. This is in line with previous work (e.g., Bradlow et al., 1997, 1999) that has demonstrated that perceptual phonetic training is beneficial to the establishing of new categories. These findings also indicate that the experimental group did not perform better on trained words. These findings provide further support for abstractionist theories of phonological processing.

The findings from the dialog/paragraph reading task indicated that learners in the experimental group were able to generalize their learning from training on singleton palatals to complex palatals in words like perch. It was not the case, however, that learners were able to
generalize to disyllabic palatal-final words or \textit{ed} ending morphemes. Thus, while we saw generalization to a larger discourse context and to complex palatals it was not the case that learning generalized to all palatals. I return to this discussion in Chapter 6.

5.6.3 What will be the effects, if any, of perceptual training on palatal codas on productions of palatal codas?

The results from Experiment 3 show that the experimental group’s productions improved significantly from the pretest to the post-test. The same was not true for the control group. The findings from the paired-comparison task indicated that more post-test productions of the experimental group were judged as more native-like in comparison to those of the control group. Nevertheless, averages for the control group were above zero, indicated a slight preference for post-test productions. This is not surprising, however, if we consider that the control group also received perceptual phonetic training on tense/lax vowels. Recall that this finding does not mean that NLs were able to better identify the control group’s post-test productions, but rather that they indicated some of their post-test productions were more native-like than their pretest productions. The phonetic training the control group received on vowels could have (and might even have been predicted to) make their post-test productions as more native-like. Nevertheless, the significant difference between the experimental and control group demonstrates that perceptual phonetic training on palatal codas contributed to significantly more native-like judgments from NLs for the experimental group.

The results from the forced-choice task indicated that NLs were able to more accurately identify the experimental group’s post-test productions of palatal codas words in both isolated words and sentences. The same was not true for the control group. These findings are in line with
those from the paired-comparison task, indicating that the experimental group’s productions of palatal codas improved significantly more than the control group’s productions. This work extends these findings with the inclusion of a dialog/paragraph reading task. As we have seen, learners in the experimental group demonstrated improvements on words that were the focus of the training, but they also demonstrated improvements with productions of palatal words with complex codas. They did not, however, demonstrate improvements with palatals before disyllabic word-final palatal words or –ed ending morphemes. Nevertheless, we do have evidence that the perceptual training extended to words in larger discourse contexts.

Taken together, the findings from Experiment 3 indicate more production improvement for the experimental group than the control group, demonstrating the beneficial effects of perceptual phonetic training on palatal codas on production. This finding implies that perception and production systems must be connected in some way because perceptual phonetic training on palatal codas affected productions. I return to this discussion after summarizing results that compared perception and production improvements.

5.6.4 What is the relationship, if any, between improvement on perceptions and productions of palatal codas?

To summarize the findings from Experiment 3 with regard to the relationship, if any, between improvement on perceptions and productions of palatal codas, we found that perceptual phonetic training on palatal codas resulted in improvements on productions of palatal codas. This finding suggests a connection between perception and production systems. We also found a moderate correlation between improvements in perception and improvements in production for the isolated-word context. We did not find a correlation between improvements in perception and
improvements in production in the sentence context. We also saw quite a bit of individual variability in the improvement scores of learners. We considered the possibility that pretests scores were modulating improvements. We predicted that if we found a correlation, it would be negative because lower pretest scores would allow for more improvements. Nevertheless, there were no correlations between pretest scores and improvement scores for either perception or production in the isolated-word context or the sentence context. Thus, we cannot account for the variability as being modulated by pretest scores. Similar to the findings of Bradlow et al., it is clear that learning in the perceptual domain is transferring to learning in the production domain, but that improvements in each domain are occurring at different rates for the different learners.

As in Bradlow et al. (1997), while we saw individual improvements in both perception and production for palatals, we did not find a consistent, clear relationship between these improvements. If we return to the predictions made by the PAM and SLM regarding the relationship between perception and production, recall that the PAM posits shared representations, which would predict that perception and production systems would have a direct relationship and that perception and production learning would be strongly correlated. The SLM, on the other hand, would not predict a direct relationship between perception and production, and we would not necessarily expect to find that perception and production learning are strongly correlated. The results from Experiment 3 (and Experiment 1) are generally in line with results from previous literature in that, counter to the predictions made by the PAM, they do not provide compelling evidence for a direct link between perception and production systems. The finding that there was a moderate correlation between perception and production improvement for the isolated-word context does not pose a problem for the predictions made by the SLM in that perception and production might co-vary even if they are indirectly linked. On the other hand,
not finding a consistent correlation between perception and production elsewhere is arguably problematic for the PAM because shared representations should result in a direct relationship. Despite finding a moderate correlation between improvements in perception and production in the isolated-word context, we did not find a correlation with words in sentences. We know that because the amount of improvement was not related to the pretest scores, the lack of relationship between perception and production cannot be attributed to the initial scores of participants. This then raises the question of what about the sentence context makes production not directly related to perception. One possibility relates to the complexity of the sentence context in relation to the isolated-word context in production: Learners have more to produce in a sentence as opposed to an isolated word. In addition to having more segments to produce, suprasegmentals features like rhythm and sentence-level intonation must be attended to. Nonetheless, although caution is advised in drawing conclusions from null effects, the present results do not provide compelling evidence for a direct link between perception and production systems. I return to this discussion in the final chapter of this dissertation.

5.6.5 In which contexts do learners have the most difficulty with palatals?

Results from the dialog/paragraph reading task indicated that pre-palatal environment (e.g., vowel or a consonant /n l ɹ/) did not appear to have an effect on production improvements of the experimental group. These findings suggest that perceptual phonetic training on singleton palatals extended to production improvements not only in singleton codas, but also in complex codas. The same was not true for palatal words with the –ed ending morphemes or disyllabic final-palatal words. In the case of disyllabic final-palatal words, the lack of improvement could have been related to the low number of these items on the task as well as the high pretest scores
of learners (90% for the control group and 91% for the experimental group as compared to, for example, 71% and 65% for the control and experimental groups for the –ed ending pretest scores). For the –ed ending words, the lack of improvement could have been related to the fact that these words contained inflectional morphology, which we know poses production problems of their own. For example, Lardiere (1998) investigated the tense morphology and pronominal case usage of a Chinese-speaking L2 learner of English using naturalistic production data from interviews. This study demonstrated that even after living in the US for 18 years, this learner who mastered pronominal case (supplying it 100% of the time in obligatory contexts) still only supplied past tense marking in 34% of obligatory contexts. Similar to the results from the perception task in which learners showed improvements in both the pre-consonantal and prevocalic sentence contexts, the context in which the palatal coda was heard (e.g., before a consonant, before a vowel, phrase-final) did not appear to have an effect on production improvements of the experimental group in the dialog/paragraph reading task.
CHAPTER 6
GENERAL DISCUSSION AND CONCLUSION

This dissertation set out to answer questions related to: (1) syllable structure and how it may filter perception in second language acquisition; (2) the relationship between perception and production in the acquisition of a second language phonological system; and (3) the effects of perceptual phonetic training on the perception and production of palatal codas. It accomplished this by investigating the acquisition of palatal codas of Korean L2 learners of English. The sections below provide a brief summary of the general findings of this work and provide implications for both future research and pronunciation pedagogy.

6.1 Syllable Structure

The findings of this dissertation provide evidence that syllable structure constraints can play a role in the acquisition of new segments. Perception of palatal codas was found to be difficult for mid-proficiency learners. Thus, we can conclude that L1 syllable structure constraints mediate and filter the perception of segments being acquired by learners. We also gained evidence that learners are able to overcome these differences and modify constraints in a native-like way in that we found high-proficiency learners who performed like native speakers.

At the end of Chapter 2, we discussed Abrahamsson (2003)’s developmental path for the acquisition of codas. Abrahamsson posits a U-shaped learning path, proposing a particular developmental sequence in the learning of codas, specifically coda deletion > vowel epenthesis > closed syllables. He also predicts a greater overall proportion of vowel epenthesis as proficiency increases in that within a functional approach of phonology, learners simultaneously attempt to maximize intelligibility while minimizing articulatory effort. In order to maximize intelligibility,
listeners may attempt to keep as much information in the uttered form of a word as possible. Nevertheless, there was no evidence from the current research that learners deleted palatal codas. This finding could be attributed to the proficiency levels of learners in the current study, which were relatively high. In fact, most learners reported not only having studied English for many years, but also having lived in an English-speaking environment for an extended period of time. Because of this, we do not have information regarding beginning learners, and therefore, what happens at the outset of acquisition. It could be that coda deletion is a strategy used at an earlier stage of acquisition.

In relation to production results, while it was the case that Korean L2 learners of English demonstrated a trend for having more difficulty with disyllabic words as compared to monosyllabic words, we did find learners that epenthesized a vowel after palatal codas. This finding could support the hypothesized developmental path put forward by Abrahamsson. However, an alternative exists to Abrahamsson’s explanation that learners are epenthesizing vowels because it allows them to maximize intelligibility while minimizing articulatory effort: It could be that learners epenthesize a vowel because they initially perceived a vowel in the input for palatal codas and thus stored it in their representations. The findings from this dissertation indicate that it is indeed the case that learners have difficulty perceiving palatal codas. Thus, similar to the findings of Dupoux et al. (1999) and Kabak and Idsardi (2007), we have evidence that learners perceive illusory vowels, at least some of the time, and thus may store these vowels in their representations. Despite this finding, we also have evidence that learners are able to overcome these perceptual illusions and ultimately perceive and produce palatal codas in a native-like way.
6.2 Relationship between Perception and Production and the Effects of Perceptual Training on the Perception and Production of Palatal Codas

The second goal of this dissertation was to investigate the relationship between perception and production with regard to both syllable structure constraints and segments in the categories of fricatives and affricates. Relatedly, the third focus of this dissertation was to investigate the effects of perceptual phonetic training on the perception and production of palatal codas. The goal was not only to determine whether perceptual training results in positive gains in both the perception and production of palatal codas for both familiar and new words and talkers, but also to design materials that could be pedagogically viable and realistically implemented by teachers and/or used by students.

Results from Experiment 1 showed no correlation between perception and production accuracies for learners. Results from Experiment 3 showed a moderate correlation between improvements on perception and production for the isolated-word context, but no correlation between improvements on perception and production in the sentence context. The absence of a direct link between perception and production in two of the three correlation analyses that were run on the data are in line with much of the previous research on the relationship between perception and production (Bohn & Flege, 1990; Goto, 1971; Sheldon & Strange, 1982; Flege & Eefting, 1987; De Jong, Hao, & Park, 2009; Bradlow et al., 1997). Nevertheless, we also saw from Experiment 3 that perceptual phonetic training on palatal codas was successful in improving both the perception and production of palatal codas. Learners were able to generalize improvements to both new words and new talkers in the contexts trained as well as in larger discourse contexts, as shown by improvements in a dialog/paragraph reading task. The results for palatals provide evidence for a close connection between perceptual and production systems.
However, the exact relationship between these systems is not clear. We find that perceptual training leads to improvements in production, which implies some relationship between representations. On the other hand, we did not find a correlation between perception and production accuracies in Experiment 1, nor did we find a correlation between perception and production improvements for palatal codas in sentences in Experiment 3, indicating that there is not always a direct relationship.

If we compare the experimental group’s results from the perceptual training study in this dissertation to those of Bradlow et al. (1997), we find many similarities. First, Bradlow et al. found variation in learners’ pretest perception accuracies, post-test perception accuracies, as well as the amount of improvement in perception accuracy. The same was found for the experimental group in this dissertation for both isolated words and sentences. Bradlow et al. also found variation with regard to the relative improvements on perception and production in their study, but a correlation between the two was not significant. A similar result was found in the current research for palatal codas in sentences; however, there was a moderate correlation for palatal codas in isolated words. Thus, in both Bradlow et al. and the current research, it is clear that learning in the perceptual domain is transferred to learning in the production domain, but improvements in each domain are occurring at different rates for the different learners.

Bradlow et al. explained asymmetries between improvements in perception and improvements in production in their results in several ways. First, they considered the possibility that some learners’ production improvements did not match those of participants with similar perception improvements because they needed more time to acquire the motor skills to produce the /ɹ-/l/ contrast. They also considered the possibility that some learners were attending to cues during training that would not aid in improving perception but that might aid in improving
production. Recall that Bradlow et al. focused on the acquisition of the /ɹ/-/l/ contrast in English. Thus, they argued that a learner could have focused on durational rather than spectral cues during perceptual training. In this case, they argue, durational cues might have been “sufficient to signal an /ɹ/-/l/ contrast in production but were ineffective for the perceptual identification of /ɹ/ and /l/ by native English speakers” in the word-identification task (p. 2307). In relation to cue weighting in L2 speech perception, we know from previous work (e.g., Iverson, Kuhl, Akahane-Yamada, Diesch, Tohkura, Kettermann, & Siebert, 2003; Mirman, Holt, & McClelland, 2004; Lotto, Sato, & Diehl, 2004; Iverson, Hazan, & Bannister, 2005; Holt & Lotto, 2006; Idemaru, Holt, & Seltman, 2012) that a variety of perceptual cues are used in speech perception for a given target, that these cues are weighted, and that the weighting of these cues is not equally applied in categorization tasks and can be changed. For example, Holt and Lotto (2006) demonstrated that learners exhibit biases for cue weighting that do not necessarily reflect the informativeness of cues, but rather experience from L1 patterns. Nevertheless, they also found that cue weights could be positively altered by manipulating the distribution of the input.

We also know that individual differences in cue weighting can affect a learner’s ability to successfully acquire new contrasts (Chandrasekaran, Sampath, & Wong, 2010). Thus, in order to provide the most beneficial training for palatal codas, we must first determine the relevant cues and their relative weight as well as attempt to employ methods to encourage re-weighting of these cues in L2 learners. In Experiment 2, we saw that high-proficiency learners patterned more like native speakers than did mid-proficiency learners in their perception of manipulated mono- and disyllabic palatal words. If we consider the gestures used to articulate palatals in different syllable positions, we can note that they differ not so much in their place of articulation, but rather in their articulatory timing. Thus, in order to acquire native-like perception, L2 learners
must attend to the cues that signal these timing differences (e.g., duration of the consonant). The findings from Experiment 2 that show sensitivity to the presence/absence of a final vowel being mediated by a mono- vs. disyllabic stem provide evidence that learners are attending to stem cues to compensate for their difficulty in perceiving palatal codas. Future research investigating the acquisition of palatal codas should attempt to better understand the cues that native speakers use to identify these segments. We can thus manipulate stimuli to enhance these cues and encourage their re-weighting to allow for more native-like perception (and production). It might also be beneficial to explicitly instruct learners to pay attention to the differences in the stem, thus focusing learners’ attention to relevant cues.

Let us now return to our discussion of the underlying mechanisms that might account for why perceptual phonetic training results in generalizability from both episodic-trace and abstractionist points of view. Recall that within episodic-trace theories, learners store exemplars in memory; thus, the more exemplars stored from different speakers, the more robust or stable the representations. We also discussed the possibility that if episodic-trace theories are correct, we would find better performance for trained words within the training experiment, because these are precisely the exemplars that would be stored. This is not in line with findings from Experiment 3, in which learners in the experimental group were shown to generalize across new words and new talkers in both the isolated-word and sentence contexts. On the other hand, abstractionist theories posit abstract representations that mediate recognition; varied input would strengthen these abstract representations and allow for generalization to new words and new talkers. The results reported in this dissertation are more in line with abstractionist theories, suggesting that learning extends beyond the particular exemplars that learners may have stored during the training.
We also discussed speech perception theories and their predictions on how learning in one skill (perception) would affect the other (production). The PAM has roots in Direct Realism and posits linked systems that share representations; thus, it would predict that perception and production learning would be directly related. The SLM predicts that perception will lead production, but it does not assume a direct link between perception and production and therefore would not predict that perception and production learning are strongly correlated. The results from Experiment 3 are generally in line with results from previous literature and Experiment 1 in that, counter to the predictions made by the PAM, they do not provide compelling evidence for a direct link between perception and production systems. The finding that there was a moderate correlation between perception and production improvement for the isolated-word context does not pose a problem for the predictions made by the SLM in that perception and production might co-vary even if they are indirectly linked. On the other hand, not finding a consistent correlation between perception and production elsewhere is arguably problematic for the PAM because shared representations should result in a direct relationship between the two.

Bradlow et al. attempt to explain the lack of correlation between perception and production improvements by suggesting that perhaps the specific motor commands for improved /ɹ/-/l/ production are acquired at different, individual rates, suggesting that perception and production are not mediated by exactly the same type of representations. It is also possible in the current research that improvements in perception and production skills are occurring at different rates. This implies that representations are not directly shared between perception and production systems. With regard to the present research, we see that perceptual training has a beneficial effect on the perception of palatal segments, as it does for /ɹ/-/l/. Thus, it appears that the results
from Bradlow et al. were not unique to the acquisition of /d/-/l/ and perhaps indicate a more general trend in the acquisition of L2 phonology.

One important consideration for the findings reported on in this dissertation relates to the proficiency level of learners who participated in the experiments. A majority of participants represented intermediate to advanced learners (with a few exceptions in the perceptual training study reported in Chapter 5). As previously mentioned, learners reported not only having studied English for many years, but also having lived in an immersion context for an extended period of time. Because of this, we do not have information regarding beginning learners, and therefore, what happens at the outset of acquisition. Recall that the SLM is concerned with ultimate attainment and therefore makes predictions about advanced learners, whereas PAM focuses on cross-language research and therefore makes predictions for learners with no prior linguistic experience with the target language. Thus, future work investigating the performance of lower-level learners might yield different results from those reported on in this dissertation and provide important insights for these speech perception models. If we are to fully understand the relationship between perception and production with regard to the structures investigated in this dissertation, it will be necessary to investigate what happens at the beginning stages of acquisition.

The research presented in this dissertation also extends previous work by adding a perceptual training beyond isolated words, and demonstrates that while gains are not as significant, improvements occur. However, we saw that while accuracies in isolated words plateaued in the mid 90% range, plateaus within sentences were at ~80%. This is an important finding for several reasons. First, we know that language learners ultimately need to improve perceptions in larger discourse contexts to be successful communicators. Second, for phenomena
that are relatively less difficult to acquire (final palatals vs. /ɹ/-/l/ contrast), we still find lower performance in contexts beyond those of isolated words. The finding that sentential contexts are more difficult than the isolated-word context is most likely the result of the sentence context being more complex. Learners have more to attend to in a sentence as opposed to an isolated word. In addition to having more words to perceive, suprasegmentals features like rhythm and sentence-level intonation must be attended to. This is also the case for production in the sentence context.

Another interesting question, not a focus of this dissertation, relates to potential effects or benefits that production training might have for learners. One explanation for finding learners with higher production accuracies than perception accuracies in the literature investigating /ɹ/-/l/ was that learners might have received articulatory training. Training research has focused on the effects of perceptual training, but what about the effects of production training? One methodological consideration for work attempting to answer these questions is that production training necessarily includes an aspect of perceptual training (i.e., learners not only hear what the instructors produce in the training itself, but when learners produce a sound/utterance, they can also perceive what they are producing). Some work has begun to investigate these issues (e.g., Baese-Berk, 2010).

A final question, yet unanswered, is whether learning from perceptual training extends to naturalistic speech contexts. Related to this question is whether learning persists in the long term. We have some indication that the effects of perceptual phonetic training persist over time based on the findings of Bradlow et al. (1999), who demonstrated that effects were still present three months after training. However, extension of learning to naturalistic contexts has been less well studied. Part of the reason for this relates to methodological considerations. It can be difficult, for
example, to elicit naturalistic data that include the target items under investigation. If we can elicit target items, it might be difficult to obtain enough items to conduct statistical analyses and to do so within a reasonable time frame as to not demand too much investment from our participants. Even if we are able to elicit enough target items, it is then difficult to control for the phonological environment in which they appear. If we consider the design for the read-aloud production experiment in Chapter 5, recall that we had an equal number of target words before a consonant and before a vowel. Additionally, the dialog/paragraph reading task included words balanced for context (before a consonant, before a vowel, phrase-final), and it included complex coda clusters (e.g., /ʃ/, /ʃ/, /ʃ/). Some of the words in the latter context are low-frequency (e.g., *marshy*) and are unlikely to be elicited (and perhaps it is unlikely that learners would even know these words). When eliciting naturalistic data, obtaining this type of consistency and balance is impossible. These are just a few methodological reasons why extending laboratory research to naturalistic contexts is difficult and has therefore not been done. Nevertheless, future research regarding the effects of perceptual phonetic training needs to investigate these issues to provide ecological validity to results. If perceptual training methods are to be incorporated into pronunciation pedagogy, then definitive evidence that learning extends to naturalistic speech must be shown.

6.3 Methodological and Pedagogical Implications

Findings from this dissertation also provide methodological implications for related future work. We saw from Experiment 3 that the control group improved between the pretest and post-test for the isolated-word (but not the sentence) context. Several explanations were suggested to account for this. First, while the tense/lax vowel training completed by the control
group was meant be unrelated to the palatal coda targets investigated in this dissertation, the focus on the vowel [i] as well as the fact that stimuli contained consonants in coda position may have caused the training to inadvertently have a beneficial effect on the palatal word pairs when they were heard in isolation. Second, the extensive pretests/post-tests (lasting a total of 1.5-2 hours) could have provided enough input to act as training for the control group. This provides justification for the inclusion of similar time-on-task control groups and not just those that complete the pre- and post-tests. Nevertheless, the type of task that the control group completes should be carefully considered. Additionally, we saw that when learners were able to control the amount of time spent on task (as was the case with the control group), they spent less than half the time required.

One issue with perceptual training (which is also a consideration when we consider its pedagogical feasibility) relates to the interest level of users. Participants in this research often commented on the ‘less-than-exciting’ or ’boring’ nature of the perceptual training tasks. Thus, for experimenters, motivating participants (or not giving them the control to determine time on task) is recommended. Some researchers (e.g., Wade & Holt, 2005; Lim & Holt, 2011) have attempted to overcome this issue by using methods other than forced-choice word-identification tasks with explicit feedback. In Lim and Holt’s (2011) work investigating the /ɹ/-/l/ contrast, a custom computer videogame was used that connected target sounds to certain characters in the game and required participants to use visual and aural information to correctly identify and interact with those characters to be successful in the game. In this way, they did not participate in overt categorization, nor did they receive explicit feedback. With only 2.5 hours of training, the videogame paradigm also showed perceptual improvements similar to those found using perceptual training paradigms similar to those of the present study that included training for
much longer periods of time. While potentially difficult to implement (e.g., because of the technical skills required to design and implement a computer videogame), creative methods like this are not only successful at inducing learning, but also might be motivating and provide an alternative to the practical concern of some ‘less-than-exciting’ varieties of perceptual phonetic training. The issue of motivation is also relevant for pronunciation teachers who would consider using this type of training to supplement courses. Fortunately, we have seen that even short amounts of training can have enhancing effects, at least in terms of palatal codas.

Considering pedagogical feasibility was of import in the design of the perceptual training paradigm used in Experiment 3. This was apparent in the use of online implementation software as well as in limiting the amount of time spent on task. Results from the dialog/paragraph reading task demonstrated that learners were not only able to improve on their productions of trained items (mono- and disyllabic words with singleton palatals such as push/pushy), but in some cases, even extended that learning to complex codas (e.g., perch). It was not the case, however, that learners were able to generalize to disyllabic palatal-final words or –ed ending morphemes. Thus, while we saw generalization to a larger discourse context and to complex palatals it was not the case that learning generalized to all palatals. One possible explanation for the lack of generalization to disyllabic palatal words could be methodological and related to the low number of items in the task (n=26). In addition, pretest scores for these items were quite high (~90% for each group), indicating that perhaps these types of words are less difficult for learners. Palatals with –ed endings, however, contained more items and pretest scores were ~70% for each group. These words, on the other hand, contained a morpheme boundary. Thus, we see that there are some limitations to the generalizability of learning that took place from the training. Despite these limitations, we did see learners generalizing to new words and new talkers
as well as larger discourse contexts, which provides support for the incorporation of this type of training into pronunciation classrooms.

The research presented in this dissertation contributes to a better understanding of second language acquisition, in general, and the acquisition of second language phonology, in particular. It sheds light on issues related to the roles of perception and production in second language acquisition as well as how syllable structure constraints relate to these issues. It also allows us a better understanding of the developing IL of an L2 learner with respect to the phenomena investigated. Finally, the focus on pedagogically feasible perceptual phonetic training contributes to identifying better practices in the pronunciation classroom.
REFERENCES


APPENDIX A: CLOZE TEST

DIRECTIONS
1. Read the passage quickly to get the general meaning.
2. Write only one word in each blank next to the item number. Contractions (example: don’t) and possessives (John’s bicycle) are one word.
3. Check your answers.
NOTE: Spelling will not count against you as long as the scorer can read the word.
EXAMPLE: The boy walked up the street. He stepped on a piece of ice.

He fell (1) down____ but he didn’t hurt himself.

MAN AND HIS PROGRESS

Man is the only living creature that can make and use tools. He is the most teachable of living beings, earning the name of Homo sapiens. (1) ________ ever restless brain has used the (2) ________ and the wisdom of his ancestors (3) ________ improve his way of life. Since (4) ________ is able to walk and run (5) ________ his feet, his hands have always (6) ________ free to carry and to use (7) ________ . Man’s hands have served him well (8) ________ his life on earth. His development, (9) ________ can be divided into three major (10) ________, is marked by several different ways (11) ________ life. Up to 10,000 years ago, (12) ________ human beings lived by hunting and (13) ________ . They also picked berries and fruits, (14) ________ dug for various edible roots. Most (15) ________ , the men were the hunters, and (16) ________ women acted as food gatherers. Since (17) ________ women were busy with the children, (18) ________ men handled the tools. In a (19) ________ hand, a dead branch became a (20) ________ to knock down fruit or to (21) ________ for tasty roots. Sometimes, an animal (22) ________ served as a club, and a (23) ________ piece of stone, fitting comfortably into (24) ________ hand, could be used to break (25) ________ or to throw at an animal. (26) ________ stone was chipped against another until (27) ________ had a sharp edge. The primitive (28) ________ who first thought of putting a (29) ________ stone at the end of a (30) ________ made a brilliant discovery: he (31) ________ joined two things to make a (32) ________ useful tool, the spear. Flint, found (33) ________ many rocks, became a common cutting (34) ________ in the Paleolithic period of man’s (35) ________. Since no wood or bone tools (36) ________ survived, we know of this man (37) ________ his stone implements, with which he (38) ________ kill animals, cut up the meat, (39) ________ scrape the skins, as well as (40) ________ pictures on the walls of the (41) ________ where he lived during the winter. (42) ________ the warmer seasons, man wandered on (43) ________ steppes of Europe without a fixed (44) ________, always foraging for food. Perhaps the (45) ________ carried nuts and berries in shells (46) ________ skins or even in light, woven (47) ________. Wherever they camped, the primitive people (48) ________ fires by striking flint for sparks (49) ________ using dried seeds, moss, and rotten (50) ________ for tinder. With fires that he kindled himself, man could keep wild animals away and could cook those that he killed, as well as provide warmth and light for himself.
APPENDIX B: LANGUAGE BACKGROUND QUESTIONNAIRE

A. General Information
1. Sex: □ F □ M
2. Age: 
3. Do you have vision or hearing problems? 
4. University year: 1 2 3 4 5 6 7 8 9
5. Major: 

B. Known Languages and Uses
1. Native language: ______________ Dialect: 
2. Mother’s native language: ______________ Dialect: 
3. Father’s native language: ______________ Dialect: 
4. Language(s) spoken at home during childhood: 
5. Language(s) spoken at home during the first five years of your life: 
6. Country of residence during the first five years of your life: 
7. Language(s) of instruction during elementary school (content courses): 
8. Country of residence from 6 to 11 years old: 
9. Language(s) of instruction during middle and high school (content courses): 
10. Country of residence from 12 to 17 years old: 
11. Other language(s) that you know and proficiency levels

<table>
<thead>
<tr>
<th>Language</th>
<th>Reading</th>
<th>Writing</th>
<th>Speaking</th>
<th>Listening</th>
</tr>
</thead>
<tbody>
<tr>
<td>□ Beginner</td>
<td>□ Beginner</td>
<td>□ Beginner</td>
<td>□ Beginner</td>
<td>□ Beginner</td>
</tr>
<tr>
<td>□ Intermediate</td>
<td>□ Intermediate</td>
<td>□ Intermediate</td>
<td>□ Intermediate</td>
<td>□ Intermediate</td>
</tr>
<tr>
<td>□ Advanced</td>
<td>□ Advanced</td>
<td>□ Advanced</td>
<td>□ Advanced</td>
<td>□ Advanced</td>
</tr>
<tr>
<td>□ Near-native</td>
<td>□ Near-native</td>
<td>□ Near-native</td>
<td>□ Near-native</td>
<td>□ Near-native</td>
</tr>
</tbody>
</table>

□ Beginner  □ Intermediate  □ Advanced  □ Near-native
12. Weekly use of English and other language(s)
   a. % weekly use of English: ______________________
   b. % weekly use of __________________________ (language): ______________________
   c. % weekly use of __________________________ (language): _______ (a-c = 100%)

13. In what language are you the most comfortable at this time? ______________________

C. Learning of English (learners only)
1. Age of first exposure to English: ______________________
2. Context of first exposure to English: □ At school □ Outside school □ Both
3. English instruction
   a. Number of years of English instruction that you have received: ______________________
   b. English dialect spoken by your teachers
      □ American □ Australian
      □ British □ Canadian
      □ Scottish □ Other: ____________
      □ Irish
   c. Were the majority of your English teachers native speakers of English? □ Yes □ No
4. a. Have you ever taken an English pronunciation class? □ Yes □ No
   b. If so and the class was at Illinois, which course did you take and when? ______________________
   c. If you took the class at another institution, please describe the course below and include the topics you studied?
5. Immersion(s) in an English-speaking environment □ N/A
   a. First immersion
      i. Age: ______________________
      ii. Place: ______________________
      iii. Context: ______________________
      iv. Duration: __________ year(s) __________ month(s) __________ week(s)
   b. Second immersion
      i. Age: ______________________
      ii. Place: ______________________
      iii. Context: ______________________
      iv. Duration: __________ year(s) __________ month(s) __________ week(s)
## APPENDIX C: LIST OF STIMULI USED IN EXPERIMENT 1

| /dʒ/  | 1. edge, edgy  | 7. fuss, fussy  |
|       | 2. pudge, pudgy | 8. class, classy  |
|       | 3. wedge, wedgy  | 9. hiss, hissy  |
|       | 4. marge, margie | 10. race, racy  |
|       | 5. dodge, dodgy  | 11. grass, grassy  |
|       | 6. smudge, smudgy | 12. grace, gracie  |
|       | 7. hedge, hedgy  |   |
|       | 8. sponge, spongy |   |
|       | 9. cage, cagey  |   |
|       | 10. fudge, fudgy |   |
|       | 11. bulge, bulgy |   |
|       | 12. sludge, sludgy |   |

| /n/  | 1. push, pushy  | 7. fun, funny  |
|      | 2. ash, ashy  | 8. bun, bunny  |
|      | 3. trash, trashy  | 9. fan, fanny  |
|      | 4. fish, fishy  | 10. ann, annie  |
|      | 5. mush, mushy  | 11. sun, sunny  |
|      | 6. wish, wishy  | 12. john, johnny  |
|      | 7. bush, bushy  |   |
|      | 8. flash, flashy  |   |
|      | 9. wash, washy  |   |
|      | 10. cush, cushy  |   |
|      | 11. rash, rashy  |   |
|      | 12. slush, slushy  |   |

| /ʃ/  | 1. catch, catchy  | 7. starch, starchy  |
|      | 2. itch, itchy  | 8. patch, patchy  |
|      | 3. sketch, sketchy  | 9. rich, Ritchie  |
|      | 4. kitsch, kitschy  | 10. glitch, glitchy  |
|      | 5. bitch, bitchy  | 11. peach, peachy  |
|      | 6. stretch, stretchy  | 12. twitch, twitchy  |

| /s/  | 1. mess, messy  |   |
|      | 2. grease, greasy  |   |
|      | 3. boss, bossy  |   |
|      | 4. dice, dicey  |   |
|      | 5. sass, sassy  |   |
|      | 6. lace, lacey  |   |
APPENDIX D: EXPERIMENT INSTRUCTIONS

The following instructions were presented to participants completing the AXB perception task in Experiment 1.

_This is an AXB task. In this task, you will hear English words. For each test item, you will hear three words. The three words will be uttered by three different speakers. Your task is to determine whether the SECOND word you heard (Word X) is identical to the FIRST word (Word A) or the THIRD word (Word B) you heard. Once the question “A or B?” appears on the screen, press A if Word X is identical to Word A, and press B if Word X is identical to Word B. For example, if you heard the three words “bat,” “bit,” and “bit,” you should press B because the SECOND word (Word X), “bit,” is identical to the THIRD word (Word B), “bit,” and not identical to the FIRST word (Word A), “bat.”_

The following instructions were presented to participants completing the forced-choice word-identification task in Experiment 2.

_You are about to listen to a series of English words. After hearing each word, two words will appear on the computer screen. Your task is to identify AS QUICKLY AS POSSIBLE the word that you heard. Press D to select the word presented on the LEFT, and press L to select the word presented on the RIGHT._

_For example, if you listened to the word "bit" and saw the words "bat" on the left and "bit" on the right, you should press L, because the word you heard "bit" which corresponded to the word presented on the right._

The following instructions were presented to participants completing the Yes/No perception task in Experiment 2.

_You are about to listen to a series of English words. After hearing each word, your task is to decide AS QUICKLY AS POSSIBLE whether or not you heard a vowel at the end of the word. You will be prompted by a screen asking Yes or No? If you do hear a vowel, press_
the marked key on the left side of the keyboard. If you do NOT hear a vowel, press the
marked key on the right side of the keyboard. The number of yes and no answers need
not be balanced.

For example, if you listened to the word "happy", you should mark the key on the left side
of the keyboard to indicate that you heard a vowel at the end of the word.

The following instructions were presented to participants completing the forced-choice word-
identification task in Experiment 3.

You are about to listen to a series of sentences. After hearing each sentence, two
sentences will appear on the computer screen. Your task is to identify AS QUICKLY AS
POSSIBLE the sentence that you heard. Press the marked key on the left to select the
sentence presented on the LEFT, and the marked key on the right to select the sentence
presented on the RIGHT.
For example, if you listened to the sentence "He said bit angrily" and saw the sentence
"He said bat angrily" on the left and "He said bit angrily" on the right, you should press
the marked button on the right, because what you heard corresponded to the sentence
presented on the right.

The following instructions were presented to participants completing the paired-comparison task
in Experiment 3.

In this experiment you will see a word displayed on the screen. Next, you will hear two
versions of that word. Then, you will compare those words using a rating scale and
clicking your selection with the mouse. You will hear both real and made-up English
words. You might hear the same word more than once. Here is an explanation of the
rating scale:

- 1 will indicate the first word was a clearer and more intelligible pronunciation of
the word on the screen.
- 7 will indicate the second word was a clearer and more intelligible pronunciation
of the word on the screen.
• 4 will indicate that there were no noticeable differences between the words.
You can use all seven rating boxes to indicate your comparisons of the words.

The following instructions were presented to native-listener participants completing the forced-choice word-identification task in Experiment 3.

You are about to listen to a series of words. After hearing each word, your task is to identify AS QUICKLY AS POSSIBLE the word you heard. Press the marked key on the left to select the word presented on the LEFT, and the marked key on the right to select the word presented on the RIGHT.
For example, if you listened to the word "bit" and saw the words "bat" on the left and "bit" on the right, you should press the marked button on the right.
You might hear the same word multiple times.
APPENDIX E: Experiment 1 AXB Perception Data Results Presented in Percent Accuracy

Here, results from the perception task are reported in percent accuracy. Figure 50 shows the coda accuracy of the native, high and mid groups by type of consonant (fricative or affricate).

![Graph showing coda accuracy by type for native, high, and mid groups.

Figure 50: Coda accuracy by type (fricative vs. affricate) for all proficiency levels

A mixed-design repeated-measures ANOVA was performed on the accuracy rates with type of consonant (fricative, affricate) and existence of the segment as a phoneme in Korean (yes, no) as within-subject variables, and with proficiency (native, high, mid) as between-subject variable. Type of consonant had a significant effect, $F(1,24)=7.79, p<.05$. As can be seen from the results by type, affricates appeared to be more difficult overall.

There was no effect for existence of phoneme ($F<1$) and no interaction between type of phoneme and existence of phoneme ($F<1$). Figure 51 shows the coda accuracy of the native, high and mid groups by existence of the segment as a phoneme in Korean (yes, no).
Proficiency also had a significant effect, $F(2,24)=6.35, p<.01$. Post-hoc Tukey tests showed a significant difference ($p<.05$) between the mid-proficiency group and the other two groups, but no difference between the native and high-proficiency groups ($p>0.1$). Hence, these results suggest that lower-level L2 learners had more difficulty overall perceiving palatal codas than higher-level L2 learners and native speakers.

Recall that the consonant /s/ was included in the design to test for differences between affricates and fricatives as well as to complete a 2X2 design. However, it was not hypothesized that learners would have difficulties perceiving these sounds based on the original observation of Korean L2 learners’ epenthesizing after final palatals. It could be the case that the significant result found between affricates and fricatives above is driven by the fact that the affricate category contains two palatals, but the fricative category contains one palatal and one non-palatal. Because palatals and non-palatals are not balanced in this experiment, we will not
compare them directly. However, we can consider individual segments. Figure 52 displays the accuracy rates of learners separated by consonant type.

![Bar chart showing perception accuracies for different consonants by proficiency level]

**Figure 52: Perception accuracies separated by consonant for all proficiency levels**

When we conduct a mixed-design repeated-measures ANOVA with consonant type (/sʃʧʤ/) as within-subject variable and proficiency (native, high, mid) as between-subject variable, we find a significant main effect of group, $F(1,24)=6.35$, $p<.01$. The main effect of consonant only approaches significance, $F(3,72)=2.72$, $p<.051$, and there is no interaction between consonant and group ($F<1$). If we consider the groups of fricatives and affricates, the balance of palatals is not equal. The finding that affricates were more difficult than fricatives could be affected if, for example, the status of /s/ as a non-palatal somehow made perception easier. In order to test the fricative vs. affricate question then, let us compare only /ʃ/ and /ʧ/. In this way, we can directly ask the question of whether palatal affricates are more difficult than palatal
fricatives without the potentially influencing factor of /s/. We should, however, keep in mind that this comparison contains relatively few items in comparison to the previous analysis.

A mixed-design repeated-measures ANOVA was performed on the accuracy rates with type of consonant (/ʃ ʧ/) as within-subject variable, and with proficiency (native, high, mid) as between-subject variable. There was a significant main effect of proficiency, $F(2,24)=3.74, p<.05$, but there was no effect of type of consonant, $F(1,24)=3.39, p<.078$, and no interaction between type of consonant type and proficiency ($F<1$). Taken together, these results suggest that /s/ is behaving differently from the palatals and driving the finding that affricates are more difficult than fricatives. In other words, we do not find evidence for the existence of the phoneme in Korean affecting perception accuracies nor do we find strong evidence that consonant type (fricative, affricate) affects perception accuracies.
APPENDIX F: EXPERIMENT 3 STIMULI

/dʒ/
1. edge/edgy
2. dodge/dodgy
3. smudge/smudgy
4. hedge/hedgy
5. cage/cagey
6. fudge/fudgy
7. sludge/sludgy
8. wedge/wedgy
9. tudge/tudgy
10. pidge/pidgy
11. codge/codgy
12. bedge/bedgy
13. modge/modgy
14. leidge/leidgy
15. sodge/sodgy
16. feidge/feidgy
40. touch/touchy
41. tetch/tetchy
42. putch/putchy
43. petch/petchy
44. boatch/boatchy
45. mutch/mutchy
46. letch/letchy
47. sotch/sotchy
48. fatch/fatchy

/ʃ/
17. push/pushy
18. ash/ashy
19. trash/trashy
20. fish/fishy
21. bush/bushy
22. flash/flashy
23. slush/ slushy
24. mush/mushy
25. teesh/teeshy
26. pash/pashy
27. cosh/coshy
28. bosh/boshy
29. mish/mishy
30. leish/leishy
31. seish/seishy
32. fush/fushy
33. catch/catchy
34. itch/itchy
35. sketch/sketchy
36. stretch/stretchy
37. peach/peachy
38. twitch/twitchy
39. patch/patchy

Fillers
49. fend/pend
50. chief/cheap
51. flake/fleck
52. cologne/clone
53. train/terrain
54. drive/derive
55. polite/plight
56. filet/flay
57. parade/prayed
58. beret/bray
59. sale/sell
60. miss/mist
61. pass/past
62. blow/below
63. blike/belike
64. pleam/paleam
65. fape/pape
66. heff/hepp
67. tabe/tebb
68. clate/calate
69. treem/tereem
70. prume/perume
71. prace/perace
72. mape/mepp
73. tiss/tissed
74. rass/rassed
75. froy/feroy
76. drate/derate
APPENDIX G: COMPLETE LIST OF STIMULI FROM DIALOG/PARAGRAPH PRODUCTION MEASURE IN EXPERIMENT 3

<table>
<thead>
<tr>
<th>f</th>
<th>Vpal</th>
<th>Vpal + i</th>
<th>Vpal + ed</th>
<th>Vpal 2 syll</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>brush</td>
<td>ashy</td>
<td>pushed</td>
<td>finish</td>
</tr>
<tr>
<td>F (8)</td>
<td>Before C (8)</td>
<td>Before V (8)</td>
<td>F (9)</td>
<td>Before C (7)</td>
</tr>
<tr>
<td>posh</td>
<td>cash</td>
<td>fish</td>
<td>squishy</td>
<td>mushy</td>
</tr>
<tr>
<td>posh</td>
<td>fish</td>
<td>rash</td>
<td>pushy</td>
<td>cushy</td>
</tr>
<tr>
<td>flesh</td>
<td>cash</td>
<td>rash</td>
<td>mushy</td>
<td>squishy</td>
</tr>
<tr>
<td>rash</td>
<td>posh</td>
<td>squash</td>
<td>ashy</td>
<td>ashy</td>
</tr>
<tr>
<td>hush</td>
<td>fresh</td>
<td>fresh</td>
<td>fishy</td>
<td>flash</td>
</tr>
<tr>
<td>leash</td>
<td>posh</td>
<td>lash</td>
<td>bushy</td>
<td>bushy</td>
</tr>
<tr>
<td>bush</td>
<td>dish</td>
<td>wish</td>
<td>cushy</td>
<td>mushy</td>
</tr>
<tr>
<td>fish</td>
<td>trash</td>
<td>crash</td>
<td>flashy</td>
<td>slushy</td>
</tr>
<tr>
<td>f</td>
<td>n/l/l pal</td>
<td>n/l/l pal + i</td>
<td>n/l/l pal + ed</td>
<td>n/l/l pal 2 syll</td>
</tr>
<tr>
<td></td>
<td>marsh</td>
<td>banshee</td>
<td>NONE</td>
<td>NONE</td>
</tr>
<tr>
<td>F (4)</td>
<td>Before C (4)</td>
<td>Before V (4)</td>
<td>F (2)</td>
<td>Before C (2)</td>
</tr>
<tr>
<td>Walsh</td>
<td>Welsh</td>
<td>kolsch</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Walsh</td>
<td>Walsh</td>
<td>Welsh</td>
<td></td>
<td></td>
</tr>
<tr>
<td>marsh</td>
<td>kirsch</td>
<td>harsh</td>
<td></td>
<td></td>
</tr>
<tr>
<td>harsh</td>
<td>marsh</td>
<td>marsh</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

206
<table>
<thead>
<tr>
<th>Vpal</th>
<th>Vpal + i</th>
<th>Vpal + ed</th>
<th>Vpal 2 syll</th>
</tr>
</thead>
<tbody>
<tr>
<td>bleach</td>
<td>catchy</td>
<td>matched</td>
<td>sandwich</td>
</tr>
<tr>
<td>match</td>
<td>beach</td>
<td>much</td>
<td>grouchy</td>
</tr>
<tr>
<td>itch</td>
<td>beach</td>
<td>much</td>
<td>touchy</td>
</tr>
<tr>
<td>touch</td>
<td>touch</td>
<td>much</td>
<td>watch</td>
</tr>
<tr>
<td>watch</td>
<td>watch</td>
<td>much</td>
<td>watch</td>
</tr>
<tr>
<td>much</td>
<td>much</td>
<td>which</td>
<td>itchy</td>
</tr>
</tbody>
</table>

| n/l/ɹ pal |
|-----------|---------|-----------|-------------|-----------|---------|
| inch | crunchy | pinched | research | research |
| cinch | cinch | French | F (7) | Before C (7) | Before V (7) | F (10) | Before C (2) | Before V (10) | F (1) | Before C (1) | Before V (1) | research | research | research |
| pinch | lurch  | inch  | lurch | French | inch | cinch | French | cinch | cinch | F (7) | Before C (7) | Before V (7) | F (10) | Before C (2) | Before V (10) | F (1) | Before C (1) | Before V (1) | research | research | research |
### d5

<table>
<thead>
<tr>
<th>Vpal</th>
<th>Vpal + i</th>
<th>Vpal + ed</th>
<th>Vpal 2 syll</th>
</tr>
</thead>
<tbody>
<tr>
<td>badge</td>
<td>dodgy</td>
<td>staged</td>
<td>cabbage</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>F (9)</th>
<th>Before C (8)</th>
<th>Before V (7)</th>
<th>F (8)</th>
<th>Before C (8)</th>
<th>Before V (8)</th>
<th>F (5)</th>
<th>Before C (1)</th>
<th>Before V (5)</th>
<th>F (2)</th>
<th>Before C (2)</th>
<th>Before V (1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hodge</td>
<td>edge</td>
<td>fudge</td>
<td>edgy</td>
<td>veggie</td>
<td>pudgy</td>
<td>staged</td>
<td>caged</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>grudge</td>
<td>judge</td>
<td>judge</td>
<td>cagey</td>
<td>dodgy</td>
<td>pudgy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>edge</td>
<td>pledge</td>
<td>pledge</td>
<td>puggy</td>
<td>fudy</td>
<td>edgy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pudge</td>
<td>page</td>
<td>page</td>
<td>edgy</td>
<td>veggie</td>
<td>dodgy</td>
<td>dodged</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>judge</td>
<td>lodge</td>
<td>lodge</td>
<td>edgy</td>
<td>edgy</td>
<td>stodgy</td>
<td>paged</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>age</td>
<td>edge</td>
<td>edge</td>
<td>veggie</td>
<td>pudy</td>
<td>fudy</td>
<td>staged</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bridge</td>
<td>stage</td>
<td>grudge</td>
<td>dodgy</td>
<td>wedgie</td>
<td>veggie</td>
<td>smudged</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### d5

<table>
<thead>
<tr>
<th>n/l/I pal</th>
<th>n/l/I pal + i</th>
<th>n/l/I pal + ed</th>
<th>n/l/I pal 2 syll</th>
</tr>
</thead>
<tbody>
<tr>
<td>change</td>
<td>dingy</td>
<td>ranged</td>
<td>orange</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>F (10)</th>
<th>Before C (10)</th>
<th>Before V (10)</th>
<th>F (7)</th>
<th>Before C (7)</th>
<th>Before V (7)</th>
<th>F (10)</th>
<th>Before C (4)</th>
<th>Before V (10)</th>
<th>F (2)</th>
<th>Before C (2)</th>
<th>Before V (2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>strange</td>
<td>binge</td>
<td>change</td>
<td>range</td>
<td></td>
<td></td>
<td>cringed</td>
<td>changed</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>lounge</td>
<td>singe</td>
<td>change</td>
<td>lounge</td>
<td>dingy</td>
<td></td>
<td>changed</td>
<td>lunged</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>twinge</td>
<td>change</td>
<td>cringe</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>spurge</td>
<td>George</td>
<td>change</td>
<td>manic</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Marge</td>
<td>change</td>
<td>charge</td>
<td>stinky</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>George</td>
<td>change</td>
<td>large</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>large</td>
<td>charge</td>
<td>change</td>
<td>charge</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

208
APPENDIX H: EXAMPLE DIALOG FROM EXPERIMENT 3 PRODUCTION MEASURE

9. What’s that smell?

Dad: Something smells fishy in here. This place is trashed!

Son: Come on, it’s hardly grungy at all!

Dad: It’s dingy and disgusting! There’s not one clean dish here! When was the last time you bleached this blanket? I can’t believe I even touched it! It’s so dirty. I’m sure these bedbugs are all gorged. And when’s the last time you cleaned this glass? It’s all smudged. The floor is even squishy if you step on it!

Son: I’ve been busy recently and haven’t had a chance to clean! I don’t think it’s stENCHY at all!

Dad: I’m going to make sure you tidy up. I’ll be keeping a close watch. What’s that mushy thing stuck to the wall? Is that a lighter? You completely scorched this desk!

Son: Oh, HUSH! It’s not that bad. Since you cringe at the sight of it, I’ll tidy up, but there’s no need to get into a rage about it.

Dad: Okay, just make sure you purge the room of whatever creatures infest it and try to get rid of that fishy odor. You should also change around your laundry so you have clean clothes. You’ll feel much better once everything’s washed. I’ll bring a sponge and some soap so you can get started.

Son: Gosh, you think I can’t do anything. I’ve got a sponge right here!

Dad: Listen, there’s no need to get touchy. This is a cleanup job you can’t dodge. I really don’t mean to judge. It’s just that you need to start keeping this place cleaner, especially at your age! Once this room is purged of its filth, you can have your Wii back. Until then, no video games!
APPENDIX I: ANALYSIS OF DAILY TRAINING DATA

Here data are presented regarding the daily tracking of the training sessions of the experimental group for palatal codas in isolated words and sentences. Recall that each day a learner completed two listening tasks: one with isolated words and one with words in carrier phrases. In each of the tasks, learners heard stimuli from a different talker (of the four who were randomly selected to be training stimuli). First, I begin with the isolated-word context. Figure 53 shows the daily average of percent accuracy for palatal codas.

![Figure 53: Palatal codas in isolated words](image)

As we can see, after the first two days, scores from participants level off. If we look separately at the performance on mono- vs. disyllabic words, we see a similar trend, as shown in Figure 54.
Figure 54: Palatal codas in isolated words separated by mono- and disyllabic words

Next, I present results from the sentence context. Figure 55 shows the daily average of percent accuracy for the final palatals in sentences.
It appears that for the palatal codas in the sentence context, it takes learners longer to plateau. Whereas in the isolated words learners plateaued at Day 3, in the sentence context, they plateaued at Day 5. We can also note that accuracy levels are not as high as in the isolated-word context. We can further separate data in mono- vs. disyllabic words, as shown in Figure 56. Again, as was the case in isolated words, we see a similar trend between mono- and disyllabic words.

**Figure 55: Palatal codas in sentences**
Overall, from the daily training data we can see a clear trend that palatal codas increase until they reach a plateau point. This is approximately Day 2 for the isolated words and Day 5 for the words in sentences. This potentially indicates that learners are discovering a pattern with the palatal codas that is allowing them to improve. In addition, it is interesting to note that the plateau range for the palatal codas in isolated words is around 95%, while for words in sentences it is around 80%.

**Figure 56: Palatal codas in sentences separated by mono- and disyllabic words**
APPENDIX J: EXPERIMENTAL PERCEPTUAL TRAINING INSTRUCTIONS

Your Participant Number: 3

Thank you for participating in our training study! Please read the following instructions carefully.

In order for you to participate in the training, you’ll need to download the Paradigm Player here. You must be using a PC (Paradigm Player is not supported on Macs).

Once you’ve downloaded the program, you should click on the Dropbox icon on the menu bar and log in to the experiment using the ID and password below.

After you log in to Dropbox, you will be asked to “allow” the Paradigm Player to access the Dropbox account. Click “Allow”.

Once you’ve logged in, you should see a list of 8 experiments that you will complete during the duration of your training.

You will complete 8 training session days. Each training day will include two sessions: the first with isolated words (approx. 7 mins) and the second with words in sentences (approx. 12 mins). You should always begin with the isolated words and continue with the sentences.
You should wait 1 day in between each training session day, but no longer. It is very important that you keep to the schedule. Your personalized training schedule is below. You **MUST** follow the schedule. Do not work ahead or do other tests that you’re not scheduled to take. If you do so, you will not be able to continue in the study now will you be able to participate in the pronunciation workshops. If you have any questions at any time, please email huensch@illinois.edu

Before you begin each training session, you should verify that the sound is functioning on your machine and at an appropriate volume. Once you begin the experiment, you will not be able to adjust settings on your machine without exiting the program.

In order to start your training, choose the appropriate experiment based on your training schedule by highlighting it with your mouse. Then, press the play button (green arrow) at the top of the Paradigm Player window. **NOTE: Each experiment will take between 2-8 minutes to download each time you begin.** After the experiment has downloaded, you will be asked to enter your subject name (your participant number, ‘P1’). Next, you will be asked to enter the session number (‘1’ or ‘2’ depending on whether it’s your first or second time completing that training session). If you have any questions at any time about your schedule, please do not hesitate to contact us (huensch@illinois.edu).

<table>
<thead>
<tr>
<th></th>
<th>Day 1</th>
<th>Day 2</th>
<th>Day 3</th>
<th>Day 4</th>
<th>Day 5</th>
<th>Day 6</th>
<th>Day 7</th>
<th>Day 8</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Isolated</strong></td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
</tr>
<tr>
<td></td>
<td>(Session 1)</td>
<td>(Session 1)</td>
<td>(Session 1)</td>
<td>(Session 1)</td>
<td>(Session 2)</td>
<td>(Session 2)</td>
<td>(Session 2)</td>
<td>(Session 2)</td>
</tr>
<tr>
<td><strong>Sentence</strong></td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>A</td>
<td>C</td>
<td>D</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td></td>
<td>(Session 1)</td>
<td>(Session 1)</td>
<td>(Session 1)</td>
<td>(Session 1)</td>
<td>(Session 2)</td>
<td>(Session 2)</td>
<td>(Session 2)</td>
<td>(Session 2)</td>
</tr>
</tbody>
</table>
APPENDIX K: PERCEPTUAL TRAINING TIME COMPARISON BY PARTICIPANT

Table 17: Perceptual Training Time Comparison by Participant

<table>
<thead>
<tr>
<th>Participant</th>
<th>Days between pretest and training start</th>
<th>Days off during training</th>
<th>Days between training and post-test</th>
<th>Total time on training task (minutes)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental Group</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>160.00</td>
</tr>
<tr>
<td>7</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>160.00</td>
</tr>
<tr>
<td>8</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>160.00</td>
</tr>
<tr>
<td>9</td>
<td>4</td>
<td>0</td>
<td>1</td>
<td>160.00</td>
</tr>
<tr>
<td>10</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>160.00</td>
</tr>
<tr>
<td>11</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>160.00</td>
</tr>
<tr>
<td>13</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>160.00</td>
</tr>
<tr>
<td>15</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>160.00</td>
</tr>
<tr>
<td>27</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>160.00</td>
</tr>
<tr>
<td>30</td>
<td>6</td>
<td>0</td>
<td>1</td>
<td>160.00</td>
</tr>
<tr>
<td>31</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>160.00</td>
</tr>
<tr>
<td>32</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>160.00</td>
</tr>
<tr>
<td>Average</td>
<td>2.50</td>
<td>0.58</td>
<td>0.50</td>
<td>160.00</td>
</tr>
<tr>
<td>SD</td>
<td>2.35</td>
<td>1.00</td>
<td>0.80</td>
<td>n/a</td>
</tr>
<tr>
<td>Range</td>
<td>0-6</td>
<td>0-3</td>
<td>0-2</td>
<td>n/a</td>
</tr>
<tr>
<td>Control Group</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>61.68</td>
</tr>
<tr>
<td>17</td>
<td>4</td>
<td>2</td>
<td>0</td>
<td>59.12</td>
</tr>
<tr>
<td>18</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>56.20</td>
</tr>
<tr>
<td>19</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>96.32</td>
</tr>
<tr>
<td>20</td>
<td>13</td>
<td>5</td>
<td>0</td>
<td>28.12</td>
</tr>
<tr>
<td>21</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>127.87</td>
</tr>
<tr>
<td>24</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>71.05</td>
</tr>
<tr>
<td>25</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>90.85</td>
</tr>
<tr>
<td>26</td>
<td>6</td>
<td>2</td>
<td>0</td>
<td>53.18</td>
</tr>
<tr>
<td>29</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>58.72</td>
</tr>
<tr>
<td>33</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>50.13</td>
</tr>
<tr>
<td>35</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>61.00</td>
</tr>
<tr>
<td>Average</td>
<td>2.33</td>
<td>1.25</td>
<td>0.25</td>
<td>67.85</td>
</tr>
<tr>
<td>SD</td>
<td>3.87</td>
<td>1.48</td>
<td>0.62</td>
<td>26.01</td>
</tr>
<tr>
<td>Range</td>
<td>0-13</td>
<td>0-5</td>
<td>0-2</td>
<td>28-127</td>
</tr>
</tbody>
</table>