When Phonological Systems Collide:
The Role of the Lexicon in L2 Phonetic Learning

by

Farrah Neumann

B.A., Florida State University, 2013

Submitted to the Graduate Faculty of the

Dietrich School of Arts and Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2021

UNIVERSITY OF PITTSBURGH

DIETRICH SCHOOL OF ARTS AND SCIENCES

This dissertation was presented

by

Farrah Neumann

It was defended on

March 25, 2021

and approved by

Matthew Kanwit, Associate Professor, Department of Linguistics

Marta Ortega-Llebaria, Associate Professor, Department of Linguistics

Seth Wiener, Associate Professor, Department of Modern Languages, Carnegie Mellon University

Dissertation Advisor: Melinda Fricke, Assistant Professor, Department of Linguistics

When Phonological Systems Collide:
The Role of the Lexicon in L2 Phonetic Learning

Farrah Neumann, PhD

University of Pittsburgh, 2021

The acquisition of a sound system is an integral component of second language (L2) communication, yet it is one of the most difficult skills to teach and is therefore largely ignored in L2 classrooms (Derwing, 2010). In laboratory settings, phonetic training studies have typically examined syllables, rather than words, with no referential meaning. Support for this decontextualization of the stimuli and the subsequent distributional learning that takes place has come from findings that word learning impedes phonetic learning, especially when involving minimal pairs (Feldman, Griffiths, Goldwater, & Morgan, 2013; Hayes-Harb & Masuda, 2008). This advantage for distributional learning over minimal pair learning, however, has not been demonstrated with regard to generalizing to untrained, analogous contexts.

To investigate the role of the lexicon in L2 phonetic learning, the experiment in this dissertation trained participants on an artificial language with a different VOT category boundary than that of their first language, English. The experiment featured a between-subjects design in which participants were exposed to one of two training conditions. In the explicit, minimal pair-based condition, participants learned fine-grained VOT differences through voiced-voiceless minimal pairs (*bilsu*/*pilsu*) that illustrated a VOT category boundary of around 0 ms, providing explicit evidence for the phones' contrasting phonological statuses. The implicit condition served to determine whether distributional learning could take place without explicit information from minimal pairs. The implicit condition's lexicon contained no minimal pairs to illustrate a direct meaningful relationship between voiced/voiceless pairs of phones (*binsu/pilsu*). Lexical learning

was measured by accuracy scores from the training phases, which took place across five different days. Participants' approximation to the new category boundary was measured by a discrimination task.

Most of the predictors examined showed similar effects in both groups, but the explicit condition group alone improved in perceiving prevoicing, providing support for the minimal pair (Maye & Gerken, 2000) and noticing (Schmidt, 2012) hypotheses. Both groups successfully generalized to analogous contexts, even outperforming the contexts on which they were trained. This may be attributable to overtraining and novelty effects. The data suggest that, when invoking the lexicon, minimal pairs can help, rather than hinder, L2 phonetic learning.

Table of Contents

Acknowledgements ..... **xi**

1.0 Background **1**

1.1 Phonetic Learning **3**

1.1.1 Generalizing from Learned to Analogous POAs **14**

1.1.2 Models of L2 Speech Sound Learning **16**

1.2 The Relationship between L2 Perception and Production **20**

1.3 Distributional and Minimal Pair Learning **22**

1.3.1 Distributional Learning **23**

1.3.2 Minimal Pair Learning **26**

1.3.3 The Role of the Lexicon in Phonetic Learning **27**

1.3.3.1 Sleep Consolidation **32**

1.4 Implicit and Explicit Learning **33**

1.5 The Role of Attention **42**

1.6 Research Questions **45**

1.6.1 Gap in the Literature **45**

1.6.2 Research Questions **46**

1.6.3 The Current Study **46**

1.6.4 Predictions **47**

2.0 Methodology **51**

2.1 Numana **52**

2.1.1 Lexical Items of Numana **54**

2.1.2 Stimuli Construction 58

2.1.3 Acoustic Properties of Stimuli 59

2.2 Participants 59

2.3 Experiment Procedure 60

3.0 Results 68

3.1 Accuracy 68

3.2 Discrimination 73

3.2.1 Absolute Change in AX 74

3.2.2 Direction of Change of AX 76

3.2.3 Category Boundary and Voicing 78

3.2.4 Category Boundary and Ambiguity 81

3.2.5 Speaker 83

3.2.6 Item 85

3.2.7 POA 86

3.2.8 Speaker and Item 89

3.2.9 POA and Item 92

3.2.10 Speaker & POA 94

4.0 Discussion and Conclusions 98

4.1 Lexical Learning 99

4.1.1 The Role of Feedback 105

4.2 Phonetic Learning 106

4.2.1 The Role of Attention in Distributional vs Minimal Pair Learning 107

4.2.2 Models of L2 Speech Perception 111

4.3 Generalization and Analogy 113

4.4 The Role of the Lexicon in L2 Phonetic Learning 117

4.5 Pedagogical Implications 120

4.6 Limitations and Future Directions 123

4.7 Conclusion 125

Appendix A Appendix A NOUN Objects 126

Bibliography 128

List of Tables

Table 1 Summary of Studies that Have Employed HVPT 9

Table 2 Numana Consonant Inventory 52

Table 3 Numana Vowel Inventory 52

Table 4 VOT Continuum of English and Numana 53

Table 5 Explicit Condition Items 56

Table 6 Implicit Condition Items 57

Table 7 Summary of Speakers 59

Table 8 Summary of Participant Language History 60

Table 9 Procedure Summary 67

Table 10 Mean (SD) Accuracy for Training B 2AFC Task 68

Table 11 Mean (SD) Accuracy for Training C 2AFC Task 69

Table 12 Difference Scores for Training B and C TACF Tasks 70

Table 13 Relative Improvement Scores (SD) for Training B and C 2AFC Tasks 72

List of Figures

Figure 1 Finite State Automoton of Numana Phonology 55

Figure 2 Example of Training Task 62

Figure 3 Feedback for 'Correct' Responses 65

Figure 4 Feedback for 'Incorrect' Responses 65

Figure 5 d' Scores for the Absolute Difference between A and X 76

Figure 6 d' Scores for the Direction of Change between VOT of A and X Stimuli 77

Figure 7 d' Scores for Boundary / Voicing Categorization 80

Figure 8 d' Scores for Boundary / Ambiguity Categorization 83

Figure 9 d' Scores for New versus Familiar Speaker 84

Figure 10 d' Scores for Trained versus Untrained Items 85

Figure 11 d' Scores for POA by Group 87

Figure 12 d' Scores for POA with Groups Combined 88

Figure 13 d' Scores for Item and Speaker Faceted by Group 90

Figure 14 d' Scores for Item and Speaker with Groups Collapsed 91

Figure 15 d' Scores for Item and POA Faceted by Group 93

Figure 16 d' Scores for Speaker and POA Faceted by Group 95

Figure 17 d' Scores for Speaker and POA with Groups Collapsed 96

Figure 18 Training B 2AFC Task Critical vs Distractor Items 100

Figure 19 Training C 2AFC Task Critical vs Distractor Items 102

Figure 20 d' Scores for 'Within Voiced' Items 108

Acknowledgements

In August of 2015, I showed up at Pittsburgh International Airport with my allotted one checked bag, one carry-on, and one personal item, containing all of my earthly possessions. I had never visited Pittsburgh, much less the University of Pittsburgh, and I did not know a single person in this city. It is amazing to think about how the people I have met over the last five and a half years have completely shaped my world. Without any of these people, I would not be where I am today.

I would like to start by thanking my friends who have become my family. Your love and support have revived me more times than I can count. You remind me that I am much more than just my work. I hope you know how much I care for you. A simple 'thank you' could never be enough to convey how much gratitude I hold for you. Although many of my closet friends were made in this PhD program, these are friendships that will last a lifetime. For this I am far luckier than I could have ever hoped to be.

Thank you to everyone in the Linguistics Lab who helped with piloting and running my experiment. Your help, support, feedback, and company on many late nights is what got me through. Thank you to my research assistants, Kinan Moukamal, Gillian Reed, Bret Morgan, Sarah Kerman, and Xiaotong Liang. Thank you to the "Numaniacs" who recorded stimuli for this experiment, specifically Juan Berrios, Melinda Fricke, Matthew Hadodo, Jevon Heath, Virginia Teran, Silvia Pisabarro-Sarrió, Angela Swain, and Anthony Verardi.

I would also like to thank the entire Pitt Linguistics Department. The comradery within our department is one to be rivaled. I am so glad to have had such amazing peers who motivated me to be my best and always supported me at my worst. Thank you, grad lings! And to our wonderful

faculty and staff - each professor has contributed to my understanding of the field and my survival of the PhD program in their own, unique way. Thank you to all of you who left your doors cracked and welcomed me with a smile, whether for help with linguistic theory, recording stimuli, or personal obstacles. Thank you, especially, to Na-Rae Han for helping me to wrangle Python. If not for you, I would probably still be manually counting or coding something or other.

This dissertation was made possible thanks to the financial support of the Language Learning Dissertation Grant, funding from the University of Pittsburgh Department of Linguistics, and the Dean's Tuition Scholarship. Thank you for believing in my project and making it all possible! And of course, thank you to our administrative assistant, Allison Thompson, who is the genius behind the whole department's operation.

I have been very fortunate to have a really wonderful group of committee members. Without their eye for potential and their in-depth understanding of their respective fields, this project would never have taken the form that it ultimately did. Marta Ortega-Llebaria, thank you for giving me my foundations in phonetics, and for helping me to grow as a graduate student and researcher. Seth Wiener, thank you for being supportive in absolutely every way; I only wish our paths had crossed earlier! Matt Kanwit, thank you for making me into a researcher and a critical thinker and for teaching me the ins and outs of publishing. You have gone above and beyond for me on so many occasions and I am truly, truly grateful to have learned from you. Your encouraging words after every conference presentation gave me the confidence to keep pushing when I felt ready to be done.

The final member of my committee, and the one to whom I owe the most gratitude, is my wonderful advisor, Melinda Fricke. You have taught me so much about speech perception, but you have taught me at least as much about how to be a good mentor and a kind human. In my own

teaching, I think of you every day and do my best to be as compassionate and human with my students as you have been with me. Thank you for believing in me when I did not. I know with certainly that I would not have made it across the finish line without your unwavering support.

Finally, I am forever grateful to my parents, who have been my greatest cheerleaders. They may not know exactly what I've been doing over the last six years, but you would never know it from their unwavering support and enthusiasm. I hope this dissertation makes them proud.

1.0 Background

A great number of factors present challenges to adult learners acquiring a second language (L2). For example, age, motivation, access to education, and a natural aptitude for language learning all play a role. This study focuses on L2 phonology, and in particular, the phonetic and lexical input that inform it. L2 speech sound learning has received relatively little attention in classroom settings (c.f. Derwing & Munro, 2005; Lord, 2005; Nagle, 2017; Pollock, 2020; Saito, 2018; Schmidt, 2018; Schoonmaker-Gates, 2017; Zampini, 1998) but has seen increased consideration in laboratory training studies. These laboratory studies have shed light on the effectiveness of such training under a number of different conditions, demonstrating that providing high levels of variability – in terms of number of speakers, items, and phonetic environments to name a few – leads to increased gains and improved accuracy in both perception and production (Bradlow, Pisoni, Yamada, & Tohkura, 1997; Iverson, Hazen, & Bannister, 2005; Logan, Lively, & Pisoni, 1991). These gains are not particular to the speakers, items, and environments on which participants are trained, however; these improvements have been shown to generalize to analogous contexts, pointing to the robust nature of the training (Bradlow et al., 1997; Lively, Logan, & Pisoni, 1993; McClaskey, Pisoni, & Carrell 1983; Thomson, 2011, 2012)

In addition to these strides in high variability phonetic training (HVPT), studies have also begun to shed some light on the relationship between phonetics and the L2 lexicon (Curtin, Goad, & Pater, 1998; Hayes-Harb, 2007; Hayes-Harb & Barrios, 2019; Hayes-Harb & Masuda, 2008; Maye & Gerken, 2000; Pater, 2003). In particular, understanding which lexical items to include (or whether to include any at all) in training paradigms in order to best facilitate phonetic learning will inform the particulars of how these laboratory studies might eventually come to be applied in

other settings, such as the classroom. In attempts to understand the role of the lexicon in L2 phonetic learning, researchers have explored and compared the efficacy of minimal pair training and distributional learning training (Feldman, Griffiths, Goldwater, & Morgan, 2013a; Feldman, Myers, White, Griffith, & Morgan, 2013b). Minimal pair learning centers around learners acquiring phonetic contrasts by virtue of being exposed to stimuli containing lexical items differing in a single sound, whereas distributional learning accounts for learning without necessarily incorporating such minimal pairs, which may be unnecessary and at times even disadvantageous. Although distributional learning appears to yield more favorable results than minimal pair learning in terms of allowing learners to create separate phonological categories (Thiessen, 2007, 2011; Feldman et al., 2013a, 2013b), the verdict remains out on which paradigm results in better generalization to novel contexts. Although some studies have examined this question, their contrasting findings mean that the field has not yet reached a consensus on the matter (Maye & Gerken, 2001; McClaskey et al., 1983).

Given this, the goal of this dissertation is to first examine whether L2 phonetic variables (namely voice onset time (VOT)) are more effectively acquired via minimal pair-based or distributional training. Additionally, this dissertation examines which of the two types of training are more effective tools for training participants to accurately perceive a phoneme with a slight phonetic difference in place of articulation (POA) between the first language (L1) and the L2. Specifically, /t/ is alveolar in English but is dental in the training language, Numana. Finally, this project aims to contribute to the discussion on whether minimal pair or distributional learning better facilitate generalization to an untrained POA in an L2 after being trained on an analogous contrast in two other POAs.

To provide an overview on these topics, this literature review begins by examining the phonetic elements to be acquired. Subsequently, I explore studies that have examined L2 speech sound training before turning to examine two models of L2 speech perception that help to account for their results. Because a great number of the training studies have focused on production instead of, or in addition to, perception, which is the primary focus of the current study, in this review of the literature I also discuss the relationship between the two modalities. In the following sections, I explore distributional learning and minimal pair learning, the role of the lexicon in phonetic learning, and learners' abilities to generalize from learned to analogous elements in an L2. Next, implicit and explicit learning and the role of attention are explored before finally turning to the research questions that motivate the current study.

1.1 Phonetic Learning

This overview will begin by examining an acoustic element that differs between many languages, including English and Spanish – voice onset time (VOT). VOT is the interval between the release of occlusion of a stop consonant and the onset of vocal fold vibration (Lisker & Abramson, 1964). Examining VOT provides a viewpoint of the malleability of the phonetic representation of the learner. This malleability can be seen in changes to the production or perception of VOT due to exposure to input with differing VOT values. Notably, much of the literature on the acquisition of L2 speech perception has focused on cross-linguistic differences in VOT. For example, the VOT category boundary of Spanish differs from that of English. Whereas English voiceless stops are produced with 30 ms or more of aspiration, Spanish voiceless stops have approximately 0 ms (Mora, 2008). Spanish voiced stops are produced with 40 ms or more of

prevoicing, or negative VOT, whereas English (phonologically) voiced stops are produced with a VOT of approximately 0 (Lisker & Abramson, 1964). Because many languages, such as Spanish, divide the VOT continuum differently than English, how and when learners acquire L2 VOT perception and production is of great interest because it provides a window onto phonetic malleability of the L2 sound system.

This is not to say that VOT is the only voicing cue to vary cross-linguistically for stop consonants. Although VOT is a primary cue used in Spanish, closure duration (Martínez-Celdrán, 1993; Simonet, 2012) and fundamental frequency are also important secondary voicing cues (Dmitrieva, Llanos, Shultz, & Francis, 2015). Fundamental frequency (f0) is a primary voicing cue in languages such as Korean, and L2 learners must become aware of differences in the individual cues or the ordering of their importance in the L2. Under some circumstances, these adjustments to L2 voicing cues can be made with relatively little exposure. Schertz, Cho, Lotto, and Warner (2015) examined voicing contrast cues in English by native English speakers and L1 Korean learners of English. Since the two languages use different cues to discriminate obstruent voicing, the authors manipulated VOT and f0 cues, such that in some cases the primary and secondary cues provided contradictory evidence for voicing. The Korean listeners attended primarily to f0, a primary cue in Korean, whereas English listeners attended primarily to VOT. Over the course of the experiment, both groups decreased their reliance on f0 due to its contradiction with VOT. This shift occurred even without explicit feedback and in under 100 exposures to the non-canonical stimuli. Although this moved the Korean listeners away from the English group in that they attended less to VOT, these findings contribute to our understanding of the rapid perceptual adaptability that has been demonstrated in both the L1 and L2 following laboratory training (Escudero, Benders, & Wanrooij, 2011; Iverson, Hazan, & Bannister, 2005;

Kondaurova & Francis, 2010; Norris, McQueen, & Cutler, 2003). The current study examines changes in the perception of a singular voicing cue – VOT – but it is important to keep in mind that additional voicing cues can be essential to discrimination, especially of exemplars with VOTs that fall within ambiguous areas of the VOT continuum.

Studies on longitudinal classroom learning or immersion have examined changes in the perception of phonetic variables such as VOT. For example, Zampini (1998) examined the perception the VOT of /p/ and /b/ by advanced L2 classroom learners of Spanish over the course of a semester. Although some participants approximated native speaker (NS) VOT norms by the end of the semester, there was considerable variation between participants. In another longitudinal study, Nagle (2017) followed L2 Spanish learners from the second semester until the fourth semester of instruction and tested their perception of VOT over five sessions. Only Spanish bilabial stops /p/ and /b/ were tested, but the perception of both became more native-like across sessions, despite receiving no intervention beyond their regular language classroom input. Like the learners in Zampini's study, however, Nagle's learners also exhibited high levels of individual variability.

In a cross-sectional study on the perception of VOT by L2 learners of Spanish, Pollock (2020) found that learners could successfully categorize voiced and voiceless pairs by the fourth semester of Spanish study, with more native-like perception following with additional semesters of study.

Typically, however, researchers have examined VOT via *production* rather than perception in classroom and immersion studies (Lord, 2005; Nagle, 2017; Saito, 2018 for classroom learning; Casillas, 2016; Jacobs, Fricke, & Kroll, 2016; Lord, 2010 for immersion). These production studies have found that learners generally do make improvements, but they can vary by program and individual and take several weeks to appear. Despite being distinct modalities, changes in

production are thought to reflect perceptual learning, with productions stemming from the mental representations that also include perception (Bradlow et al., 1997; Casillas, 2016; Flege, 1993). I return to this relationship between perception and production in a later section of this literature review.

The input found in either classroom instruction or immersion is generally naturalistic in that learners are exposed to natural, non-synthesized speech in a contextualized environment. However, because the segments being examined contain cues that participants do not attend to in the L1, such as prevoicing, the naturalistic input is not always targeted enough to raise listeners' awareness of those cues. An initial attempt to resolve this was with targeted, low variability training in a laboratory setting. Low-variability training involves training learners on a single contrast using a single item, minimal pair, or minimal triplet. It also typically involves synthetic speech, which is then modified along a continuum of some dimension such as VOT. Pisoni, Aslin, Percy, and Hennessy (1982) trained L1 English participants on a three-way voicing contrast using a discrimination task, and in doing so, were able to induce categorical perception. Importantly, however, the researchers tested only the three training items [ba], [pa], and [p^ha], and did not test for generalization to new items, speakers, or contexts. This is important because following low variability training, listeners tend to improve on the contrast or item for which they received training, but have difficulty generalizing to new items, speakers, and tasks (see Bradlow, 2008 for an overview). Thus, without testing generalization to these novel contexts, the true utility of the training remains unclear.

An additional difficulty with low-variability training paradigms is their frequent employment of synthetic stimuli. Strange and Dittmann (1984) trained L1 Japanese learners of English on the /r/-/l/ contrast using synthetic stimuli "rock" and "lock". Although listeners

experienced a “slow and effortful” improvement in their discrimination of these items, and in some cases were able to transfer these gains to additional task types and some of the 32 additional items they were tested on, they were unable to transfer their knowledge to non-synthetic stimuli or other phonetic environments (p. 132).

These problems can be largely ameliorated by high variability phonetic training (HVPT) due to its introduction of non-synthetic stimuli and greater variability in the training through additional speakers, items, or phonetic environments in the training phase (Bradlow et al., 1997; Iverson et al., 2005; Logan et al., 1991). This variability allows HVPT to target general rather than stimulus-specific learning. HVPT originated from Logan et al. (1991), who used the same word list as Strange and Dittmann (1984) but had better results, which they attributed to training participants on five different speakers using natural speech, as well as several different phonetic environments. This provided some of the first evidence for the benefit of incorporating high variability and natural stimuli in laboratory training paradigms.

The training phase of HVPT typically mirrors the testing phrase, but additionally includes feedback, which directs participants’ attention towards the relevant components of the input (Logan et al., 1991). Common tasks include either identification (e.g., Logan et al., 1991, 1993) or discrimination (e.g., Aliaga-García & Mora, 2009). For identification, the participant hears an aural stimulus and must select between two items (typically graphemes) on the screen. With discrimination tasks, participants hear two or more stimuli and are asked to decide whether they are the same or different.

HVPT is thought to be effective because, in exposing learners to multiple exemplars, learners must learn to disregard irrelevant acoustic differences and instead attend to the meaningful acoustic cues that are consistent across items, speakers, and phonetic contexts. Lively et al. (1993)

replicated the study done by Logan et al. (1991) but included an additional variable of number of speakers to whom participants were exposed. Participants trained on five different talkers were significantly more able to generalize to new talkers in the posttest as compared to participants trained on only one talker. By training learners to attend to acoustic cues that remain consistent across a variety of speakers, items, or phonetic contexts, HVPT allows learners to extract the relevant acoustic information and apply it to more accurately generalize to novel items, speakers, and contexts. Another key benefit of the HVPT paradigm is that it induces changes in perception remarkably quickly. Alves and Luchini (2017) successfully trained listeners on L2 English VOT using HVPT with only three 30-minute training sessions. In fact, the researchers found that not only did perceptual training improve participants' identification, but it also increased the production of VOT values towards those of NSs.

An additional benefit of HVPT over other paradigms is its longer lasting results. Several studies have shown that improvements can last for up to six months following training (Lively et al., 1994; Thomson, 2012). Beyond only perceptual improvements, Bradlow et al. (1999) showed that native Japanese participants also retained improvements in the production of the English /l/-/r/ contrast three months after completing training.

Although the HVPT paradigm induces rapid and long-lasting results in training L2 contrasts, the degree of success depends largely on the type of contrast trained. Aliaga-García and Mora (2009) trained Spanish-Catalan bilinguals on English VOT contrasts, /p/ - /b/ and /t/ - /d/ and two non-native vowel contrasts, /i:/ - /I/ and /æ/ - /ʌ/ using HVPT. Training consisted of both perception and production training tasks and included six two-hour sessions across six weeks. As in a number of studies that have trained only perception (e.g., Bradlow et al., 1997, 1999; Lambacher et al., 2005; Alves & Luchini, 2017), participants improved both their perception and

their production of VOT as a result of training on only one of these modalities. Participants in Aliaga-García and Mora's (2009) study did not improve significantly in either modality for the vowel contrasts, however. This finding appears to be consistent across modalities, since perception training has been shown to lead to productions that are more accurate for obstruents than for sonorants or vowels (for an overview, see Sakai & Moorman, 2018). A great number of studies have employed HVPT, and several seminal studies are summarized in Table 1. These studies have been selected in order to illustrate the development of HVPT over time, specifically with regard to the variables that have been found to make the paradigm particularly effective. Although these studies are of European training languages and examine segments, it is important to note that there is a large body of existing research on non-European languages as well as suprasegmental features. These studies are beyond the scope of the current investigation, however, and for that reason are not included here.

Table 1 Summary of Studies that Have Employed HVPT

Study	Segments Trained	Methodology	Items	Findings
Logan et al. (1991)	/l, r/	Identification training with feedback over 15 sessions. Trained on 5 talkers and tested on 1 new one. Trained on minimal	Real English minimal pairs that did not require lexical access.	Better identification accuracy with trained speakers. Phonetic environment also impacted results. Small but significant improvement in identification accuracy

		pairs using natural speech and several different phonetic environments.		after training. Provides evidence for high variability and natural stimuli being superior to low variability and synthetic stimuli, especially for generalization.
Lively et al. (1993)	/l, r/	Replication of Logan et al. (1991) but included two different training groups – one trained on one talker, and another trained on five.	Real English minimal pairs that did not require lexical access.	Successful generalization to new talkers only for group trained on different talkers.
Lively et al. (1994)	/l, r/	Replication of Logan et al. (1991) with greater number of participants and a longitudinal follow up.	Real English minimal pairs that did not require lexical access.	Improved identification of segments following training. Identification performance decreased slightly after three months. After six months, scores

				were still higher than at pretest.
Bradlow et al. (1997)	/l, r/	Perception training using minimal pair identification task. Forty-five sessions over three to four weeks.	Real English minimal pairs that did not require lexical access.	Perception and production (especially of /r/) improved as a result of perception training, including generalizing to new words.
Bradlow, Akahane-Yamada, Pisoni, and Tohkura (1999)	/l, r/	Replication of Bradlow et al. (1997) to examine long-term retention of trained contrast.	Real English minimal pairs that did not require lexical access.	Training led to improved perception and production, and results were maintained three months after training.
Lambacher, Martens, Kakehi, Marasinghe, and Molholt (2005)	/æ, ɑ, ʌ, ɔ, ɜˑ/	Six sessions of identification training over six weeks.	Closed English syllables without minimal pairs or lexical access.	Perception and production improved on trained segments.

Aliaga-García & Mora (2009)	/p, b, t, d, iː, ɪ, æ, ʌ/	Six sessions training both perception and production. Perception training tasks included identification, discrimination using minimal pairs, and phonetic transcription.	Real English minimal pairs that did not require lexical access.	Perceptual shift towards longer VOT durations following training and improved discrimination of vowel pairs.
Thomson (2012)	/i, ɪ, e, ɛ, æ, ɑ, ʌ, o, ʊ, u/	Trained participants to associate each vowel category with an image. Eight training sessions over three weeks. Training consisted of 20 tokens per vowel and two talkers.	Open English syllables without minimal pairs or lexical access, but with a visual referent for each vowel category.	Improved identification of vowels in the trained phonetic context. Transfer to one novel phonetic context but failed to transfer to a third. Results retained at delayed posttest one month after training.
Thomson and	/i, ɪ, e, ɛ, æ, ɑ, ʌ, o, ʊ, u/	Participants trained by having them select the IPA	Trained a group on open syllables and a	Training at the syllable level was more effective than the word level.

Derwing (2016)		symbol that corresponded to the vowel in the auditory input. Received auditory/visual feedback. Forty training sessions over a month.	group on real words to see if invoking lexicon detracts from phonetic learning. Information about minimal pair status was not provided.	Researchers concluded it was because participants were able to attend to the phonetic detail without needing to split attention between phonetics and the lexicon.
Alves & Luchini (2017)	/p, t, k/	Three sessions of training using an identification task and six different talkers and three items. Compared a standard training group to a group that was told explicitly to focus on the acoustics.	Real English minimal pairs that did not require lexical access.	Both groups improved in identification, but explicit group made greater gains in production transfer.

1.1.1 Generalizing from Learned to Analogous POAs

A core benefit of HVPT is that participants are trained on a small set of talkers and items and yet demonstrate improvement on additional talkers and items to which they have not yet been exposed. In addition to examining participants' abilities to generalize to untrained items and talkers, as discussed in the section on HVPT training above, studies have also examined whether listeners are capable of generalizing to new phonetic environments. In one such study, Thomson (2011) trained listeners on 10 Canadian English vowels in environments following bilabials and found that learners were able to successfully generalize improvements to the discrimination of vowels that followed alveolar fricatives but not those that followed velar stops. Relatedly, Thomson (2012) found that learners correctly generalized the same 10 Canadian English vowels from labials to velars, but not to alveolar fricatives. In other instances, however, generalizability to new phonetic environments has not been empirically demonstrated. HVPT did not fully facilitate the generalization of /ɹ/-/l/ to different syllable positions for L1 Japanese learners of English (Iverson et al., 2005).

In addition to generalizing to analogous phones, there is also evidence that training can facilitate generalization to analogous POAs. For example, Nielsen (2008) used a word naming imitation paradigm to demonstrate that participants not only imitated lengthened VOTs, but also generalized these VOTs to additional POAs, pointing to participants extracting a more general feature rule for voicing from the input. McClaskey et al. (1983) found that after training participants on a three-way voicing contrast in bilabials, results also transferred to untrained alveolars. Given this, McClaskey et al. concluded that new categories can be created in the lab. Importantly, however, participants were informed that they would be listening for new sounds that they hadn't heard in training. Given this, it is difficult to discern whether the researchers would

have had the same findings without explicitly prompting participants. Tremblay, Krause, Carrell, & McGee (1997) extended McClaskey et al.'s findings by also including mismatched negativity (MMN) responses to measure to what degree results could also be seen neuropsychologically. Their stimuli were modeled after McClaskey et al., and they found that training resulted in improved identification and discrimination of the trained bilabial and untrained velar contrasts. These results were also mirrored in the MMNs, indicating that training resulted in not only behavioral, but also neuropsychological shifts.

In another extension of McClaskey et al. (1983), Maye and Gerken (2001) also explored markedness (Eckman, 2008) in relation to participants' ability to generalize to an analogous POA following training on a single POA in an L2. Participants were trained on either velars or dentals and then tested on the second, untrained POA. With some POAs being more common than others cross-linguistically, the presence of the less common contrast (i.e., velars) typically indicates the presence of the more common contrast (i.e., dentals). Maye and Gerken predicted that because of the directionality of markedness, participants trained on velars (the marked contrast) should be able to generalize to the dentals (unmarked) contrast, but those trained on dentals should have a more difficulty generalizing to the marked velar contrast. Maye and Gerken found that training did not generalize to the untrained contrast in either case. The authors attribute this difference between their results and those of McClaskey et al. to participants not being explicitly told they would be generalizing their discrimination to new segments. Without explicit attention being paid, participants did not become aware of the new contrast. They also noted that it is also possible that training on a single POA is not sufficient for the formation of phonological feature representations, such as voicing, that are necessary for abstraction and generalization.

Relatedly, researchers have examined generalization through the lens of speakers with different accents and pronunciations (Idemaru & Holt, 2020; Kraljic & Samuel, 2007), ambiguous phones and lexical contexts (Kraljic & Samuel, 2005, 2006, 2007; Norris et al., 2003), and in speech therapy for aphasiacs (Coppens and Patterson, 2017), though these studies have not tended to focus on L2 learning and for that reason are not explored in greater depth in this review of the literature.

Pedagogically, the implications of generalization following training are clear – if learners are able to analogize to new environments or POAs, it might be possible to conclude that learners have acquired a more general, phonological feature. If this turned out to be the case, less time could be spent teaching each individual contrast in classroom or laboratory settings. On the other hand, if learners do not successfully analogize, time and attention must be devoted to equally distributing instruction across phones. From a more theoretical point of view, whether learners analogize likely informs the organization of their mental representations. Psychologically valid representations of natural classes in the L2 should facilitate generalizability to new phonetic environments (Nielsen, 2011). This theoretical account of mental representations is explored below with regard to two models of speech perception.

1.1.2 Models of L2 Speech Sound Learning

Given the presence of important phonetic and phonological cross-linguistic differences such as with VOT described above, several models of speech sound learning have been created to attempt to explain the ways in which the learning of speech sounds in a language other than one's L1 can be conceptualized. In this literature review, I discuss two such models: the Speech Learning Model (SLM) and the Perceptual Assimilation Model (PAM).

Flege's SLM was created to account for learning speech sounds in a second language (Flege 1988, 1992b, 1995b, 2002, 2003). As opposed to other competing theories invoking the notion of critical periods (Lenneberg, 1967), the SLM posits that L2 learners maintain the ability to reorganize their perceptual systems across the lifespan (Flege, 1995b; Flege, Munro, MacKay, 1995). Additionally, the SLM postulates that perceptual difficulty is based on phonetic similarity and a perceived equivalence between sounds in the L1 and L2 (Flege, 1995a). L2 phones that are similar but not prototypical exemplars of L1 categories should cause more perceptual difficulty than sounds that are notably distinct. This is because similar phones assimilate to the existing L1 category, whereas a new category is created for phones that are perceived as distinct. For example, Spanish and English differ in their production of /t/. Whereas Spanish uses a dental POA, the English realization is typically alveolar. Whether L2 learners of Spanish learn to accurately produce this difference in the L2 depends on whether they can first perceive it as distinct. If not, the perceived similarity of the two phones may actually lead to increased long-term inaccuracy in production and perception, since learners (wrongly) categorize Spanish and English /t/ as being equivalent.

This perceptual assimilation for similar sounds occurs due to equivalence classification, which assumes that listeners identify a wide range of phones as belonging to a particular category (Flege, 1987, 1995). This allows listeners to cope with variance in the stimulus, but also creates the challenge of forming separate representations for L2 sounds that are assimilated to L1 categories. The more similar (without being the same) that an L2 sound is to the L1 equivalent, the more difficult it will be for a new L2 category to form. A possible result of this is that an L2 sound may map onto an existing L1 category, effectively merging the L1 and L2 values into an intermediate value. Evidence for these shared, modified representations can be found in the

compromised VOTs of /t/ of late L1 Spanish - L2 English bilinguals (Flege, 1991). Whereas early bilinguals had two distinct representations of /t/ - one for each of their languages as shown by distinct VOTs - late bilinguals showed intermediate VOT values that mirrored those of neither English nor Spanish monolinguals. Alternatively, the L1 and L2 phones may move away from one another in order to become maximally contrastive and allow the listener to maintain separate categories for each language, as was the case with the early bilinguals just mentioned (Flege, 1991; Flege, Schirru, & MacKay, 2003).

Whereas the SLM focuses on experienced L2 learners, Best's (1995) PAM was originally meant to account for the perception of non-native speech sounds by naïve listeners, although it has since been expanded to also account for L2 learning in a revised model, the PAM-L2 (Best & Tyler, 2007). The revised model accounts for learners' development away from naïve listening and toward more native-like perception of L2 contrasts. The model examines pairs of phones (as opposed to the single segments dealt with by SLM) and makes predictions about how they perceptually assimilate to the existing phonetic and phonological architecture of the L1. Like the SLM, the PAM also assumes a shared phonetic space between the L1 and L2. Several mappings between the L1 speech sounds and speech sounds of additional languages are presented to attempt to conceptualize the ways in which contrasts in an unfamiliar language are assimilated to the L1. First, speech sounds – and in particular, contrasting pairs – are described as being either categorized or uncategorized. Categorized sounds are classified as such because they are able to find a home within the existing L1 categories. Categorized pairs can be further broken down into the following types: two-category assimilation, single-category assimilation, and category-goodness difference.

Contrasts of the two-category assimilation type are those that map neatly onto two separate L1 categories, thus allowing the listener to maintain a highly accurate ability to discriminate the members of the pairs as they would in the L1. For example, L1 English speakers assimilate Zulu phonemes /ɮ/ and /ɬ/ to two separate English categories: one of several possible fricatives (such as /s/ or /z/) and a lateral (Best, McRoberts & Goodell, 2001). Single-category assimilation occurs when two contrasting phones in a new language are mapped onto a single L1 category and both are considered equally good representations of that L1 category, resulting in a lower overall ability to discriminate the two. For example, in L1 Japanese learners of English /ɹ/ and /l/ both assimilate to the L1 category, /r/, resulting in poor discrimination of the contrast (Best & Strange, 1992). The final type of categorized sound outlined by the PAM is category-goodness. This is similar to single-category assimilation, except that the two phones are not equally good representations of the L1 category to which they have assimilated. Thus, listeners are able to somewhat discriminate the contrast on the basis of the goodness of fit. For example, in L1 English learners of Spanish, /r/ and /ɾ/ are assimilated as English /ɹ/, but /r/ is a poorer exemplar of /ɹ/ than /ɾ/ (Rose, 2010).

Uncategorized sounds are those which are perceived as speech sounds but do not assimilate to any existing L1 phone, similar to perceiving infant babbling, which sounds speech-like but remains foreign (Best, 1995). If only one member of the contrast is uncategorized, the PAM predicts good discrimination. If both are uncategorized, the PAM predicts poor discrimination. The final type of assimilation outlined by the PAM is non-assimilated, where neither phone is perceived as a speech sound at all, such as is the case with African click sounds for English speakers (Best, 1995). Non-assimilated contrasts tend to result in good to excellent discrimination (Best, 1995).

Using these classifications, the PAM provides testable predictions for speech sound learning. In the research questions and hypotheses of the current study, I will return to the specific predictions made by both the SLM and the PAM. Like the SLM, the PAM posits that the mechanisms responsible for phonetic and phonological learning remain intact over the lifespan, and therefore that learning of an additional language and its phonological categories can take place successfully at any age.

1.2 The Relationship between L2 Perception and Production

Following the SLM's postulation that accurate perception is a necessary (though insufficient) element of accurate production (Flege 1992a, 1992b), researchers have sought to better understand the relationship between the two modalities. A growing literature has also examined improvements in L2 production as a product of perception-only training (for an overview, see Sakai & Moorman, 2018). Successful gains in production following no training in this modality provide strong evidence for the dependence of production on perception. Bradlow et al. (1997) trained L1 Japanese learners of English on the /ɹ/- /l/ perceptual distinction and tested them using both identification and production tasks. Participants accurately transferred knowledge learned in the perception training to their productions, leading to more accurate productions of the contrast.

Even without training, learners have been shown to develop perception and production in tandem. In a longitudinal study following L2 Spanish learners from 2nd semester until 4th semester, Nagle (2017) tested the perception and production of VOT over five sessions to examine the relationship between the two modalities. In particular, Nagle examined whether this relationship

is synchronous (changes in one modality co-occur with changes in the other), time-lagged (changes in perception occur systematically prior to equivalent changes in production), or asymptotic (high proficiency is required before a relationship between the two modalities emerges). Perception and production of Spanish /p/ became more native-like, but /b/ underwent less production change despite undergoing perceptual changes, perhaps due to the articulatory difficulty posed by prevoicing. These findings suggest that the relationship between perception and production is time-lagged, with production emerging after perception.

Additional studies have also shown support for production lagging behind perception. Flege (1993) found that L1 Chinese learners of English differed from an L1 English comparison group in both perception and production of /t/ - /d/ contrasts but were overall more accurate in perception than production. Similarly, Casillas (2016) examined L2 Spanish learners participating in a domestic immersion program and found that the VOT of /p/ and /b/ decreased in a target-like manner in both modalities, although more immediate shifts occurred in perception.

Nevertheless, others have failed to identify a relationship between perception and production. Zampini (1998) examined the VOT of advanced L2 learners of Spanish in both modalities in a non-immersion setting. Although some participants approximated NS VOT norms for either perception or production, there was considerable variation between participants. Furthermore, no relationship was found between gains in perception and production, as many participants improved in one area but not in the other.

Still other studies have shown support for production leading perception. Such a phenomenon manifests as learners producing a contrast that they cannot yet accurately perceive (Flege, Bohn, & Jang, 1997; Zampini & Green, 2001). Importantly, Escudero (2007) notes that

such findings may be attributed to methodological concerns and therefore should be considered with some skepticism.

Improvements in production via perception-only training are beyond the scope of the present work. Nevertheless, the likelihood of a link between the two bears mentioning since a significant portion of the literature on training studies – especially those that take place in a classroom setting – examines production (Lord, 2005; Nagle, 2017; Saito, 2018 for classroom learning; Casillas, 2016; Jacobs, Fricke & Kroll, 2016; Lord, 2010 for immersion). Additionally, the SLM's assertion that inaccurate production is a reflection of inaccurate perception (Flege, 1977), points to the importance of examining production studies, which can provide critical insight into the organization of those participants' perceptual systems. More specifically, inaccuracies or difficulties in L2 production studies may still inform our understanding of where parallel *perceptual* difficulties lie.

1.3 Distributional and Minimal Pair Learning

When considering the acquisition of the perception of L2 phones, it is important to consider how and why changes in perceiving or producing L1 or L2 phones actually occur. Changes to a learner's mental organization must take place at least in part due to the non-invariance problem. This is the idea that the acoustic realization of a single phone can differ greatly depending on factors such as speaker, environment, vocal tract size, dialect, etc. (Klatt, 1979; Stokes, Venezia, & Hickok, 2019). Learners nevertheless do eventually learn to successfully perceive and categorize speech sounds. In part, this ability to learn to cope with a lack of invariance stems from

adaptation. Listeners must learn to adapt and to filter out irrelevant information in the stimuli (for an overview, see Weatherholtz & Jaeger, 2016).

Given the immense variability seen even within a single phonemic category, both L1 and L2 learners must somehow learn to accurately perceive and categorize the input in the language they are learning. Although newborn infants can discriminate the sounds of any of the world's languages, by about 12 months of age this ability is lost and infants show adult-like sensitivity to phonemic contrasts of the L1 (Kuhl et al., 2006). For example, at six to eight months of age, Japanese and American infants show a similar ability to discriminate between /r/ and /l/. By 10-12 months of age, however, American infants correctly discriminate /r/ and /l/ at significantly higher rates than Japanese infants (74% versus 60%) (Kuhl et al., 2006). Thus, the phonemic status of /r/ and /l/ in English and the allophonic status of /r/ and /l/ in Japanese is acquired early on. In particular, whether and to what degree word-learning plays a role in phonetic learning is best explored by examining two accounts of phonetic learning – distributional and minimal pair-based learning.

1.3.1 Distributional Learning

Language learning may be a product of distributional learning, a type of bottom-up processing in which learners unconsciously analyze the input for patterns. In the case of phonetic learning, clusters of exemplars separated by sparsely populated areas of phonetic space indicate category boundaries (Olejarczuk, Kapatsinski, & Haayen, 2018). This type of probabilistic learning has been captured by exemplar models (Johnson, 1997; Pierrehumbert, 2003) and prototype models (Feldman, Griffiths, & Morgan, 2009; Flannagan, Fried, & Holyoak, 1986), which both posit that each token contributes to the mental representation equally; the greater

number of exemplars of a specific type encountered in the input, the stronger the representation. Olejarczuk et al. (2018) provide an alternative account of distributional learning in which they emphasize the role of surprise over the role of frequency. Although frequency is said to play a role in the acquisition of phonetic categories, frequency as a tool has diminishing returns for higher frequency exemplars. Instead, Olejarczuk et al. argue that learners are more highly influenced by an unexpected event and update their representations disproportionately more for events that are surprising as compared to what was anticipated on the basis of the existing representation.

Distributional learning better accounts for infant learning as compared to minimal pair learning, since infants have limited vocabularies and thus cannot rely on word-learning as a means of phonetic learning. It has been proposed that, instead, infants use distributional learning to determine the statistical peaks in the phonetic input they are exposed to and use this information to inform the creation of phonemic categories in their L1s (Kuhl, 2004; Maye, Werker, & Gerken, 2002). The native language magnet theory (NLM) set forth by Kuhl (1991) describes how infants acquiring their L1s learn to organize their perceptual systems around prototypes. Although tokens produced on the periphery may be perceived as being less than ideal members of the category, they are still able to be classified as belonging to the category thanks to a magnet effect that attracts nearby members of the category. Importantly, this effect takes place in infants prior to learning word meanings or phonemic contrasts (Kuhl, 1993).

There is also evidence that L2 learners incorporate these statistical tendencies in a similar way as in L1 acquisition. In studies that have shown this, learners are not provided with meanings associated with stimuli, but are rather instructed to simply listen to a stream of input. Maye and Gerken (2000), for example, exposed adult participants to either bimodal or unimodal stimuli from a /ta/-/da/ (aspirated-unaspirated) continuum. The group that received bimodal input heard four

times as many tokens from endpoints of the continuum, whereas the group that received unimodal input heard tokens primarily from the center of the continuum. Following exposure to the input, participants that had received bimodal input were more accurate in a discrimination task than those that had received unimodal input, despite neither group being trained on meaningful minimal pairs. Although both groups had heard the tokens from all points of the continuum, hearing the majority of the tokens at either end of the continuum for the bimodal group provided stronger evidence for the existence of two separate categories (short-lag and aspirated). The unimodal group, on the other hand, did hear tokens from the endpoints of the continuum, but much fewer of them, leading them to conclude, albeit subconsciously, that these tokens were simply peripheral exemplars of a singular phonetic category. This study demonstrated that even without word meanings, learners are nonetheless analyzing the input to which they are exposed and using these statistical tendencies to inform their mental representations of phonetic and phonological categories. In a similar experiment with infants, Maye et al. (2002) also found support for statistical learning via the bimodal distribution. Together, Maye and Gerken (2000) and Maye et al. (2002) suggest that L2 learners still have access to the same phonetic learning mechanisms as infants.

Importantly, young children acquiring their L1s do not yet have robust vocabularies that include minimal pairs before about 12 months of age and yet are still able to perceive essential phonemic contrasts (Caselli et al., 1995). This provides additional evidence for a distributional account of phonetic learning in children. Adult learners, however, are not necessarily confined by a limited lexicon in the way that small children are. Given this, it bears revisiting the role of minimal pairs in the word learning and phonetic learning of adult L2 learners.

1.3.2 Minimal Pair Learning

Because exemplars rarely fit the category prototype, learning to group instances of a phone with the appropriate phonological category can be a daunting task for language learners. Thus, several researchers have posited that phonetic discrimination may be at least partially acquired as a result of word learning (Best, 1995; Bisson, Kukona, & Lengeris, 2020; Jusczyk, 1985; Lalonde & Werker, 1995; MacKain, 1982; Werker & Pegg, 1992). In this vein, the minimal pair hypothesis posits that language learners will begin to attend to phonemic differences between sounds when they realize that the two members of a given minimal pair can be used to make a distinction between two separate concepts (Maye & Gerken, 2000). A great number of training studies, including many of those mentioned above, used a minimal pair-based training paradigm (Aliaga-García & Mora, 2009; Bradlow et al., 1997, 1999; Curtin et al., 1998; Lively, Logan, & Pisoni, 1993; Lively et al., 1994; Logan et al., 1991; Pater, 2003; Strange & Dittmann, 1984). This may be because, despite some arguments against minimal pair-driven phonetic learning in L1 learners (namely the inability to rely on an impoverished lexicon to drive phonetic learning), the same dilemmas do not apply to L2 learning, which is frequently centered around vocabulary learning. Further evidence for this claim comes from Cutler, Weber, and Otake (2006), who used eye-tracking to examine L1 Japanese learners of English in their perception of the English /l-r/ contrast. When instructed to click on a rocket, participants looked towards a locket, but the reverse was not true. Weber and Cutler (2004) found similar results for L1 Dutch learners of English, who demonstrated an asymmetric perception of vowels. This mismatch of phonetic information with the lexicon points to a difference in goodness of fit for either sound, as predicted by PAM (Best, 1995). For the learners in Weber and Cutler's study, both English phonemes assimilated to the existing Japanese representation, but one was a better representation than the other, which allowed

participants to maintain some level of discrimination, albeit asymmetrically. Weber and Cutler's findings also demonstrate that learners do in fact have separate, non-homophonous representations for each lexical item. More research is still needed regarding the relationship between phonetic and lexical learning. In particular, it remains unclear whether phonetic learning takes place as a byproduct of word learning (Best, 1995; Bisson et al., 2020; Jusczyk, 1985; Lalonde & Werker, 1995; MacKain, 1982; Werker & Pegg, 1992), or whether these two systems are more effectively trained separately in the L2 learner (Hayes-Harb & Masuda, 2008; Jarvi, 2008; Thomson & Derwing, 2016).

1.3.3 The Role of the Lexicon in Phonetic Learning

Whether lexical items are included in a training paradigm may influence the degree to which participants' phonetic representations are affected by the training, providing a better understanding of the role of the lexicon in phonetic learning. A number of training studies have used real words, but it is important to differentiate between *employing* lexical items, such as by presenting the written form of the word on the screen and *evoking lexical access* by asking the participant to make a connection between a piece of information given in the experiment and the corresponding item in the lexicon. In this section, I discuss the role of the lexicon with regard to phonetic learning, first with infants before turning to adult learners.

Regarding infants, Werker and colleagues (Fennell & Werker, 2003; Pater, Stager, & Werker, 2004; Stager & Werker, 1997) have used duration of looks toward a visual target as evidence of 14-month-old infants having a decreased ability to notice switches from a word to its minimal pair as compared to 8-month-old infants who had longer looks during switch trials. Stager and Werker (2007) tested both age groups on /b/ - /d/ minimal pair nonce words by habituating the

infant to a single word-object combination and then switching the label to its minimal pair. Unlike the older infants, the younger infants perceived the difference between minimal pairs, as evidenced by longer looks during switch trials. The researchers attribute this difference to the younger infants doing a purely discrimination task, whereas the older infants reserved their discrimination of phonetic information for word learning tasks. Without a new object to assign a new label to, the older infants saw no reason to relearn a label for the original object. These findings were found consistently, even when using a variety of phonetic contexts (Pater et al., 2004). When using already known minimal pairs (*ball* and *doll*), however, the 14-month-olds did notice the switch, with longer looks during switch trials (Fennell & Werker, 2003).

Thiessen (2007) found that training infants on distinct lexical items (*dabo* and *tagu*) led to improved discrimination between minimal pairs as compared to infants trained directly on minimal pairs (*dagu* and *tagu*). In Thiessen's study, these labels were always paired with visual objects, which may account for some of the learning. In a follow up experiment though, Thiessen (2011) showed that even without visual referents, distinct lexical items are more facilitative for learning than minimal pairs, since distinct labels allow learners to experience acquired distinctiveness. Rather than being able to focus on the phonetics, when infants are presented with a minimal pair, they are forced to devote attentional resources to forming an association between an object and its label. This may be due to the additional cognitive and attentional resources required to discriminating minimal pairs as compared to pairs of words that are more noticibly distinct. The findings of both Thiessen (2007, 2011) and Werker and colleagues (Fennell & Werker, 2003; Pater et al., 2004; Stager & Werker; 1997) call into question the assumptions underlying of the minimal pair hypothesis, which predicts just the opposite, at least with regard to infant L1 acquisition. Specifically, the minimal pair hypothesis predicts that presenting language learners with minimal

pairs that illustrate a phonemic contrast should direct their attention to the phonetic cues that distinguish the two phones.

Studies on L2 learning have presented mixed findings regarding the role of the lexicon in phonetic learning. Abramson and Lisker (1970) tested L1 English speakers' ability to identify synthetic stimuli with VOT values that corresponded to Thai, specifically the three-way phonemic voiced, voiceless, aspirated bilabial contrasts (/b, p, p^h/). They found that participants identified voiced and voiceless stimuli as /b/ and aspirated stimuli as /p/ respectively, suggesting that participants were better able to perceive aspiration than voicing contrasts, in line with their L1 categories. Curtin et al. (1998) trained participants on the same Thai three-way contrast but used a word learning paradigm to do so. They found that after training, L1 English listeners could better perceive voicing contrasts than aspiration contrasts. The researchers attributed this difference in perceptual ability to the L1, in which voicing is a phonemic contrast, but aspiration is allophonic. They also attributed the discrepancy between their findings and those of Abramson and Lisker to their use of lexical items to access the learned Thai phonology. According to this account, Abramson and Lisker's phoneme identification task tapped into surface level phonetics since it did not involve the lexicon, where such surface level phonetics are not stored or represented. Curtin et al. accounted for these findings by explaining that since English aspiration is only phonetic (and therefore not stored in the lexicon), the participants in Abramson and Lisker's study were able to accurately perceived aspiration. On the other hand, Curtin et al. had learners tapping into lexical representations through a word learning paradigm. Since these English lexical items are unspecified for surface level phonetic variation such as aspiration, listeners were not as sensitive to aspiration.

In a follow up study to better examine the distinction between lexical and surface representations, Pater (2003) used the exact same stimuli and training paradigm as in Curtin et al. (1998) but used XAB discrimination tasks in which X could be either a sound or picture (Curtin et al. presented sound and picture simultaneously). Pater found the opposite of Curtin et al., in that discrimination by L1 English listeners was better for aspiration than for voicing, indicating that the Curtin et al. finding is generalizable across differing experimental conditions and that more research in the area is still needed.

Hayes-Harb and Masuda (2008) conducted an experiment in which they trained native English speakers on Japanese consonant length contrasts. The training began with a word learning phase in which auditory labels for nonce words were assigned to visual objects. Importantly, the singleton and geminates formed minimal pairs (/pete/-/pette/). Participants were then tested using a matching task, in which they were asked to match the auditory label with the corresponding picture, requiring participants to detect differences in consonant length that formed minimal pairs. Participants were able to discriminate with high levels of accuracy between members of the singleton-geminate minimal pairs, indicating that phonetic learning can occur as a byproduct of lexical learning. In a related study, Hayes-Harb and Barrios (2019) examined not only consonant singleton / geminate pairs, but also included vowel pairs. Although consonants were correctly perceived according to duration, vowels did not show evidence for contrastive encoding. It remains unclear the degree to which contrastive encoding of any kind could be generalized to untrained contexts, such as novel POAs or speakers, since these generalizations were not included in the testing phase of these studies.

In an additional study that examined differential training at the word versus sound level, Thomson and Derwing (2016) used HVPT to train L2 English learners on 10 vowel sounds using

either real words or nonsense words consisting of only open syllables. They found that training at the syllable level was more effective than the word level, and concluded it was because participants were able to attend to the phonetic detail without needing to split attentional and cognitive resources between phonetic information and the lexicon. Additional support for the greater efficacy of phonetic learning separate from lexical learning that was reported by Thiessen (2007, 2011) comes from Jarvi (2008), who discussed at length the mismatch between the frequently decontextualized laboratory training and real-world language learning. In an effort to explore the role of context – both lexical and semantic – Jarvi tested L1 English learners of Ukrainian on their ability to acquire a nonnative phonemic (palatalized versus non-palatalized) contrast using nonword minimal pairs. He compared a lexical access task (participants saw an object and heard an auditory label of either the object shown or its minimal pair and had to determine if the auditory label corresponded to the displayed object) and a no-lexical-access condition (a production-based word reading task). The results of the study showed that requiring lexical access led to a decreased ability to discriminate the trained contrast, as compared to the no-lexical access task.

The lexical-distributional hypothesis helps to explain these findings by predicting that minimal pairs might, in fact, impede phonetic learning (Feldman et al., 2013a, 2013b). Instead, distributional learning without minimal pairs should provide greater evidence for the distinction (both phonetic and semantic) between tokens, and therefore non-minimal pairs should lead to the creation of separate phonetic category representations. Feldman and colleagues (Feldman et al., 2013a; Feldman et al., 2009), attempted to model some of the differences between distributional learning only and distributional learning with a lexical component, and found evidence to support minimal pairs impairing learning since the model occasionally mistakenly identified minimal pairs as two tokens of the same word.

In an additional test of whether word level information helps to constrain phonetic learning, Feldman et al. (2013b) presented vowel sounds [ɑ] and [ɔ] to two groups of native English speakers, either in minimal pairs or non-minimal pairs. As predicted by the lexical-distributional hypothesis, the non-minimal pair group was better able to differentiate between the vowel sounds as a result of receiving input containing these sounds in contrasting words. Similar to Thiessen (2007, 2011), researchers found that both children and adults learned phonetic categories better when they were presented within different (non-minimal pair) words, thus concluding that phonetic learning may not occur in isolation, but rather is facilitated by the lexicon in that lexical items can help learners to associate the distinctiveness of the words and their differing meanings with meaningful phonetic contrasts contained by these words. However, this conclusion does not hold for minimal pairs, which, without such distinctiveness elsewhere in the word, do not facilitate an acquired distinctiveness for fine phonetic detail that distinguishes the members of the pair. Overall, invoking the lexicon leads to less effective phonetic learning, presumably as a product of greater cognitive demand and split attentional resources (Hayes-Harb & Masuda, 2008; Jarvi, 2008; Thomson & Derwing, 2016). What's more, minimal pairs in particular may be especially disadvantageous when it comes to phonetic learning (Feldman et al., 2009, 2013a, 2013b; Thiessen, 2007, 2011). Taken together, these studies present mounting evidence for the efficacy of distributional learning over minimal pair learning.

1.3.3.1 Sleep Consolidation

Learning in general – and in particular lexical learning – has been shown to benefit from sleep consolidation. For newly acquired lexical items to experience lexical competition as do existing lexical items, a period of sleep must take place. Dumay and Gaskell (2007) taught participants new words at either 8 am or 8 pm and tested them on these new lexical items

immediately and again 12 hours later. Participants of both groups showed knowledge of phonological information of the lexical items immediately, but only those who slept between sessions showed evidence of lexical competition. Participants who were tested after 12 waking hours, on the other hand, did not show signs of lexical competition for the novel items.

Others have called into question the importance of sleep consolation, however, positing that other paradigms are equally effective. For one, the Hebb (1961) repetition effect describes the improved recall experienced by participants exposed to lists with covert repetition as opposed to randomly ordered lists. This effect is considered to be implicit, since participants are not aware of the repetition in the sequencing, are not asked to learn it, and are not aware that they do learn it. In fact, this paradigm has been shown to be so effective that Szmalec, Page, and Duyck (2012) called into question the necessity of the previously established role of sleep consolidation. Szmalec et al. found that learners experienced lexical competition for words using this implicit means of presentation with the simple passing of time, regardless of whether participants had slept. Sobczak and Gaskell (2019) compared recognition, recall, and lexical integration by participants trained either using the Hebb paradigm or with an explicit phoneme monitoring task containing isolated tokens. Sobczac and Gaskell found no benefit for the more implicit Hebb paradigm and therefore reiterated the importance of sleep consolidation for lexical integration.

1.4 Implicit and Explicit Learning

The constructs of explicit and implicit learning and their relationships to phonetic and phonological learning have been discussed briefly in the above sections, but in this section, I return to these concepts to further explore this relationship and its importance in L2 learning in greater

detail. Researchers differ in what they consider to be the necessary components of implicit and explicit learning. Implicit learning was first defined by Reber (1967) as learners in an experimental setting making use of statistical tendencies in the input without a willingness to or an awareness of doing so. Dekeyser (2008) expands this definition to learning without an awareness of what is being learned, regardless of being in an experimental setting. Implicit learning also critically involves 'chunk learning' over 'rule learning' (Ellis, 2015). In summary, it is learning without any intention or conscious effort to do so.

According to Hulstijn (2005), explicit learning is just the opposite – it is "an intentional effort to uncover the rules of the system underlying the implicit data". In other words, learners must first be aware that there is a rule to acquire and then must come to recognize what is happening in the input, before finally constructing a rule that describes the pattern, resulting in knowledge of what has been learned. One thing shared by each of these definitions is the notion of awareness, a concept to which I will return in a later section devoted specifically to the role of this critical construct.

Related to implicit and explicit learning is the Declarative/Procedural Model, which makes a distinction between knowledge that is verbalizable (declarative) and knowledge that is more rapid and automatic (procedural) (for an overview, see Ullman, 2015). Not only do these memory types differ in verbal recall and in recall speed, but they also involve important physiological differences, with each memory type activating different regions of the brain, providing some physical evidence for the observable behavioral differences exhibited by learners engaged in either implicit or explicit learning. Declarative memory is associated with the medial temporal lobe, hippocampus, and Brodmann's Areas. Although declarative memory does not solely include explicit learning, all explicit learning necessarily takes place within the declarative memory

system. Greater attention or explicit instruction tends to result in declarative memory, whereas a lack of either, as well as increased complexity, typically results in procedural memory. Procedural memory involves only implicit learning and involves the basal ganglia and frontal regions of the brain. Declarative and procedural memory complement one another, can inform one another, and can interact with one another. Additionally, declarative can become procedural knowledge, a process known as proceduralization, through a gradual automatization of the skill (Dekeyser, VanPatten, & Williams, 2007).

Researchers frequently try to invoke explicit learning through explicit instruction, such as pointing out patterns or instructing learners to be on the lookout for such patterns in the input. For example, phonetic contrasts are often taught in ways that are thought to invoke explicit learning in that learners are presented with stimuli that are so inherently decontextualized, devoid of sentential or even word context, with the input instead consisting of a single CV syllable or at most a full word. This is thought to free up learners' attention to focus on uncovering the underlying rules (Best, McRoberts, & Goodell, 2001; Lively et al., 1993; Strange & Dittmann, 1984). In fact, HVPT is thought to be so effective in part due to the decontextualization of the stimuli (Thomson & Derwing, 2016). In that way, learners may fully focus their attention on the phonetic variable in question without also simultaneously attending to either the lexicon or semantics. Notably, however, this makes HVPT inherently different from language acquisition as it would take place in a classroom or immersion setting and does not address the degree to which such training might be applicable to more meaningful contexts.

Many have called for an increase in explicit pronunciation training, arguing that increased awareness of phonological form gives learners greater awareness of how their own productions differ from those of native speakers (Derwing & Munro, 2005; Thomson & Derwing, 2014;

Venkatagiri & Levis, 2007). Explicit instruction has been shown to be especially beneficial in specific contexts – in particular, pronunciation training. Although the primary focus of this dissertation is the examination of the role of implicit and explicit training on *perception*, it is important to recall that the two systems are inherently linked, with production gains reflecting a target-like shift in perceptual category representation (Flege, 1987). In a meta-analysis of 77 studies, Saito and Plonsky (2019) found that pronunciation instruction was most effective when it was focused on specific segmental or suprasegmental features, as opposed to global pronunciation training. Saito (2011) found that explicit instruction had a significant impact on comprehensibility but was less beneficial for accentedness. Gordon, Darcy, and Ewert (2013) trained ESL learners on explicit segmental features, explicit suprasegmental features, or a combination of the two without explicit instruction. They found that explicit instruction led to greater noticing by the learners and thus called for additional explicit phonetic training. Additionally, they found significantly better results for explicit instruction on suprasegmentals as compared to segments.

Such studies on pronunciation training have been carried out both in the classroom and in the lab. Studies conducted in a laboratory setting have the benefit of increased control over variables, stimuli, and exposure that classroom settings can sometimes lack. A great number of laboratory studies have already been reviewed in the section on HVPT, and a number of studies have examined the benefit of explicit pronunciation training without necessarily using HVPT to do so (Munro & Derwing, 2005). Neri, Mich, Gerosa, and Giuliani (2008) used either in person instruction or computer assisted pronunciation training to train 11-year-olds learning English as a foreign language. The two groups made comparable improvements, lending support to both classroom and laboratory pronunciation training. Although results from studies conducted in the lab have shown a successful ability to train participants' productions of particular phones, these

studies do not necessarily share the ecological validity of studies conducted in the classroom. For this reason, a great number of studies examining the role of explicit instruction have also taken place in the classroom, as outlined by Lord and Fionda (2014). Couper (2006) supplemented an English as a foreign language class with explicit pronunciation instruction on segments. After two weeks of training, the training group experienced a significant decrease in error rate, which persisted in a delayed posttest. A control group which received only the regular classroom instruction did not experience these gains in pronunciation.

Similarly, Zhang and Yuan (2020) provided explicit pronunciation training of either segments or suprasegmentals to L1 Chinese learners of English. Improvements in comprehensibility were compared to a control group that received no explicit instruction, but rather only the standard classroom instruction. After eighteen weeks of instruction, both the segmental and suprasegmental groups experienced improved comprehensibility in a sentence reading task, but only the suprasegmental group improved in a spontaneous speaking task. Lord (2005) found that L2 Spanish learners enrolled in a phonetics course improved their pronunciation over the course of the semester, providing additional support for the benefits of explicit instruction.

An important element of both implicit and explicit instruction concerns feedback, which has been shown to be effective for target language development (Lyster & Saito, 2010; Mackey & Goo, 2007; Russell & Spada, 2006). Saito and Lyster (2012) examined English /ɹ/ production in L2 English learners following form-focused instruction with corrective feedback, without corrective feedback, and a control group that received neither feedback nor form-focused instruction. They found that corrective feedback – and not form-focused instruction – was instrumental in leading to improvements in accurate English production of the phone. In this case, corrective feedback took the form of recasts, or corrected restatements of a learner's incorrect

utterance (Nicholas, Lightbown, & Spada, 2001), aimed at implicitly drawing participants' attention to the relevant acoustic dimensions of the phone. Although not particularly effective with morphosyntactic errors, recasts have been shown to be effective at targeting other areas of language acquisition, such as pronunciation errors. Mackey, Gass, and McDonough (2000) provided learners with conversational feedback on morphosyntactic, lexical, semantic, and phonological errors and later asked participants to reflect on the feedback received. Participants reported being aware of receiving feedback targeted at lexical, semantic, and even phonological errors, but did not recall receiving morphosyntactic feedback. If awareness of feedback is equated as attention, these findings point to the utility of feedback on phonology.

More explicit forms of feedback have been shown to be even more effective than recasts, however (for an overview, see Loewen, 2012 and Loewen & Sato, 2018). Carroll and Swain (1993) administered different types of feedback to different participant groups as they were learning English dative alternations. They found that explicit, metalinguistic feedback trumped the control group (no feedback), as well as more implicit forms of feedback such as recasts or simply being told they had made a mistake. Similar results were also found by Ellis, Loewen, and Erlam (2006) when training participants on the English past tense *-ed* morpheme. Although there are, of course, decisive differences between phonetic and morphosyntactic learning, these studies provide additional support for the role of explicit learning in language learning mechanisms.

Nevertheless, explicit instruction has not always been shown to be more beneficial than a control (implicit) condition. Untutored immigrants also manage to acquire an L2, often without instruction, providing further evidence for successful implicit learning (Andersen, 1991; Klein, 1995; Klein & Dimroth, 2009). Even within classroom settings, implicit instruction can also be sufficient for inducing improvements. For example, Kissling (2013) compared a phonetics

instruction group and a control group of Spanish L2 learners. Both groups received similar input, practice, and feedback, but the phonetic instruction group additionally received explicit instruction on the phonetics, phonology, and articulatory gestures involved in the production of Spanish [p, t, k, b, d, g, β, ð, ɣ, r, ɾ]. Kissling found a non-significant difference in production between the groups at the posttest, suggesting that the targeted nature of input, practice, and feedback is more important than explicit phonetic instruction, at least for production. Returning briefly to perception, Zampini (1998) and Nagle (2017) both examined the role of implicit learning by tracking L2 Spanish learners' perception of /p/ and /b/ over the course of one or more semesters of Spanish language classes without any explicit instruction on the sounds. Recall that some participants approximated native VOT norms by the end of the study, albeit with a high amount of individual variability.

Overall, however, language teachers are not confident in their ability to implement pronunciation teaching (Nagle, Sachs, Zárate-Sández, 2020). It is perhaps for this reason that teachers have not, for the most part, prioritized explicit pronunciation teaching in the classroom, instead providing feedback only when related to comprehensibility, and relying on learners to mostly implicitly acquire information about L2 speech sounds (Huensch, 2019).

Since L2 immersion is often assumed to be the most likely setting for naturalistic improvement (Flege, 1998), it follows that most production studies involving implicit learning have taken place in immersion settings. In fact, learners have been shown to improve in various phonological realms including intonation (Henriksen, Geeslin, & Willis, 2010) and awareness of phonotactics (Díaz-Campos, 2004; Face, 2018, 2021; Lord, 2010). Additionally, L1 production has also been shown to be sensitive to L2 input in an immersion setting, with learners' L1 VOT moving towards that of the L2 target (Chang, 2012; Chang, 2013). Time spent studying abroad in

Spain has been shown to increase learners' awareness of the dialect-specific variant /θ/ (Ringer, Hilfinger, 2012), as well as an increase in the pronunciation of the variant, even without any explicit pronunciation instruction (Willis, Geeslin, & Henriksen, 2009; c.f. George, 2013). Similarly, Schmidt (2018) found that Spanish learners with two semesters of classroom study or less perceived aspirated-/s/ according to the closest English category, /h/. However, participants who had studied for a greater number of semesters or spent time abroad showed a perceptual movement towards native speaker perception of aspirated-[s] as an allophonic form of /s/. Importantly, however, this finding was only true for those who had spent time abroad in a place where the target dialects used aspirated-[s]. Those who had studied abroad in [s]-preserving dialects did not share this advantage. Of particular relevance to the current study, immersion has also been found to significantly improve learners' *perception* of voicing in the L2. In one case, L1 English learners' perceptual boundary for the /p-b/ contrast in Spanish shifted towards Spanish following only two weeks of domestic immersion (Casillas, 2016).

It is important to consider the ways in which explicit and implicit learning interact and contribute to the overall acquisition of L2 phonetics and phonology. Rather than viewing one as superior to the other, the literature suggests that both implicit and explicit learning are essential. In the present investigation, explicit and implicit learning are conceptualized and operationalized as minimal pair and non-minimal pair learning respectively. Minimal pairs in this study are presented during the training phases to participants in the explicit condition since minimal pairs on either side of the VOT category boundary provide explicit evidence that the phones contrast. In the implicit condition, distributional learning takes place without additional information from minimal pairs since no minimal pairs exist in the training that could illustrate a direct meaningful relationship between voiced/voiceless pairs of phones. Rather, learners must implicitly infer the

lower VOT category boundary from the non-minimal pair items. In a similar vein, Hayes-Harb (2007) trained participants in either an explicit lexical condition, which included minimal pairs (pot-cot), or an implicit statistical condition, in which participants heard only non-contrastive pairs (pot-calm). An additional group received both lexical and statistical input. The group that received both lexical and statistical training experienced the most improved perception on the discrimination task, pointing to the importance of both implicit and explicit learning.

Further evidence that some minimum level of both implicit and explicit learning is necessary for learners' improvements in production comes from Lord (2010), who examined learners who had received explicit phonetic instruction prior to immersion followed by implicit learning via a study abroad program. She found that the combination of the implicit and explicit learning was most impactful, with neither one being as beneficial in isolation.

Unfortunately for researchers, an attempt to provide implicit or explicit instruction does not necessarily equate to implicit or explicit learning on the part of the student or participant. This creates a challenge for assessing what is actually happening in the mind of the learner as a product of either implicit or explicit instruction, since there is no guarantee that such instruction translates to learning of the same type. One solution is to simply ask learners about their strategies for learning in a retrospective verbal report (see Rebuschat, 2013 for an overview). However, learners are not always fully aware of the ways in which they have learned, since task performance and verbalization have been shown to belong to different cognitive systems (Berry & Dienes, 1993), and therefore can result in inaccurate reporting (Williams, 2005).

Explicit learning is typically operationalized by learners' explanations of patterns, whereas implicit learning is generally examined through procedural incorporation of said patterns in their productions. A resulting difficulty of examining explicit versus implicit instruction is the lack of

parallelism between the two; learners exposed to implicit instruction have the possibility for forming an explicit rule, but explicit instruction cannot yield implicit learning, which by definition requires a lack of formation of such a rule. Given this, it is important that researchers be mindful of this divide between implicit and explicit learning, as well as between instruction and learning, and carefully interpret results of studies which aim to measure implicit or explicit learning as a product of implicit or explicit instruction.

1.5 The Role of Attention

A final critical factor in L2 perceptual acquisition for the current study is the role of attention. Tomlin and Villa (1994) explored attention using three separate systems – alertness, orientation, and detection – with each one involving a greater role of consciousness than the last. Evidence for these systems was then provided by Leow (1998). First, alertness involves a general readiness to modulate attentional resources toward a stimulus. Importantly Tomlin and Villa note that alertness is not necessarily related to – and may in fact be inversely related to – accuracy and is therefore of minimal importance to SLA.

The second system of attention, orientation, involves the commitment of attentional resources to a stimulus and importantly, away from other stimuli. Accordingly, the ways in which stimuli are presented can impact the degree to which learners orient themselves toward a stimulus. Evidence for such orientation comes from differences in stimuli presentation using either exogenous or endogenous orienting. Exogenous orienting involves the attraction of attention through an external stimulus, such as a flash of light or explicit feedback. Endogenous orienting, on the other hand, allows participants to orient their attention towards a stimulus through goals or

expectations, as with presenting varying types of instructions for the same set of stimuli. Guion and Pederson (2007) trained participants on Hindi stop consonant contrasts and told different groups to attend either to meaning or sound but were otherwise presented with the same stimuli. The meaning-attending group experienced greater improvements on a semantics test and the sound attenuating group improved more as measured by a phonetic discrimination task of the Hindi dental and retroflex contrast, demonstrating that endogenous orienting through varying the instructions given was enough to significantly sway participants' perception, although notably only for the most difficult contrast; easier to learn contrasts did not reflect group differences. In a similar study, Pederson and Guion-Anderson (2010) trained groups by orienting them towards either consonants or vowels using the same set of Hindi stimuli. Only the consonant-oriented group improved their perception of Hindi consonants from pretest to posttest. Similarly, Alves and Luchini (2017) used HVPT to train L1 Spanish learners of English on word initial stop consonants. They compared a standard training group to a group that was told explicitly to focus on the acoustics, and found that both groups improved in identification of the stop consonants, but the explicit group made greater gains in production transfer.

Finally, the third system of attention, detection, is the noticing, awareness, and processing of a particular variable within the speech signal. Detection is the system of attention responsible for the intake of input. The strong form of the noticing hypothesis stated that input cannot become intake unless it is attended to by the learner (Schmidt, 1990), but this claim has since been attenuated such that the weak form of the hypothesis posits that noticing is, at the very least, facilitative of learning (Schmidt, 1994, 1995, 2001, 2012). Relatedly, attention need not necessarily be thought of as a dichotomy, but rather can be considered to exist along a continuum (Leow, 2000). Recent studies have attempted to operationalize attention, especially using eye

tracking methodologies. For example, Godfroid and Schmidtke (2013) operationalized attention as a participant's ability to recall a novel word from a previous task at the post-test. Additionally, Brusnighan and Folk (2012) found that increased reading time on sentences led to higher accuracy at defining novel words in the target sentences at the posttest. Noticing is different still from understanding; noticing refers to consciousness of specific instances whereas understanding refers to a more global generalization of awareness across instances.

Other conceptualizations of noticing have been put forward to attempt to explain the phenomenon, including the Modular Online Growth and Use of Language (MOGUL) framework, which posits that learners become aware of mental representations and those representations with higher activation levels are more likely to be noticed (Truscott & Smith, 2004, 2011). Additionally, working memory has been cited as a contributor to noticing (Mackey, Philp, Egi, Fujii, & Tasumi, 2002; Trofimovich, Ammar, & Gatbonton, 2007).

Linguistic variables that are more perceptually salient are naturally more likely to be detected or noticed by learners (Henrichsen, 1984; Schmidt & Frota, 1986). In fact, attention has been shown to improve phonetic learning in studies where listeners were trained to notice and rely more heavily on particular acoustic cues and direct their attention to dimensions not previously relied upon (Christensen & Humes, 1997; Francis & Nusbaum, 2002). Novel stimuli are typically more perceptually salient, thus producing novelty effects in studies with infants such as an increase in sucking speed (Mehler et al., 1988) or looks toward a visual target (Dietrich, Swingley, & Werker, 2007). Adults have also been shown to be biased towards novel stimuli, and in particular with novel stimuli embedded within stimuli with which the listener has become accustomed (Folstein, Van Petten, & Rose, 2008; Yang, Chen, & Zelinsky, 2009), a phenomenon described by Johnston and colleagues as novel popout (Hawley, Johnston, & Famham, 1994; Johnston, Hawley,

& Farnham, 1993; Johnston, Hawley, Plewe, Elliott, & DeWitt, 1990; cf. Christie & Klein, 1996).

1.6 Research Questions

1.6.1 Gap in the Literature

Despite the large body of existing research explored throughout this literature review, no study to my knowledge has examined the connections among implicit and explicit learning, the lexicon, and the outcomes of phonetic training. More specifically, the degree to which phonetic learning occurs as an implicit byproduct of lexical learning is not yet fully understood. Without this understanding, it is difficult to make recommendations with confidence for classroom instruction, despite the importance of perception (and consequently production) on the teaching of L2 speech sounds. Additionally, it is crucial for educators to have a better sense of the degree to which individual contrasts need to be taught, or whether analogy can be relied on after teaching a handful of exemplar contrasts. Given this, the current study seeks to examine whether generalization to an analogous POA is facilitated more by implicit or explicit training within an HVPT paradigm. Additionally, this dissertation seeks to provide a deepened understanding of the relationship between phonetic learning and lexical learning using a training paradigm that reflects naturalistic or classroom-based second language acquisition to a greater degree than many of the decontextualized HVPT studies that have come before.

1.6.2 Research Questions

Therefore, in light of this gap in the literature, the current study is guided by the following research questions and predictions, which are presented separately below:

1. What is the role of minimal pair words in phonetic training?

   The answer to this question can also shed light on the issue of explicit versus implicit instruction, since explicit evidence for a lowered VOT category boundary can be provided through minimal pairs. In the implicit, non-minimal pair condition, learners must rely entirely on distributional learning, since no minimal pairs exist in the training that could illustrate a direct meaningful relationship between voiced/voiceless pairs of phones. Rather, learners must implicitly infer the lower VOT category boundary from the non-minimal pair items.

2. Do articulatory differences between the L1 and L2 (i.e., alveolar → dental /t/) facilitate or impede trainability? To what degree do the SLM and the PAM account for these differences?

3. Do learners generalize to untrained phones (/k, g/) differently following either explicit or implicit training of analogous phones?

1.6.3 The Current Study

To attempt to address each of these research questions, the current study uses a word learning HVPT paradigm to train participants on an artificial language that has a lower VOT category boundary than does participants' L1, English. By using two different lexical conditions, the present investigation is able to explore the question of whether or not using minimal pairs

provides any benefit to these learners with regard to learning to discriminate tokens across a lower VOT category boundary.

Participants are trained on 24 lexical items designed to present evidence for the lower VOT category boundary using bilabial and dental POAs. Since the /t/ - /d/ contrast in the training language, Numana, is dental unlike its alveolar English counterpart, this difference in POA may affect the learnability of the contrast, depending on the theoretical stance taken. This idea will be explored in greater depth in the following section, along with the rationale for using an artificial language. Finally, after being trained on bilabial and dental POAs, participants will be tested on a novel, velar POA in order to examine their ability to generalize to analogous contexts as a function of the type of training received.

1.6.4 Predictions

Based on the lexical-distributional hypothesis, which predicts that minimal pairs should actually impede the learning and perception of separate phonetic categories, it was predicted that the implicit group would more quickly and accurately acquire the novel VOT category boundary on which they were trained (Thiessen, 2007, 2011; Feldman et al., 2013a, 2013b). The explicit group, with attention directed towards both word meaning and towards the novel, lower VOT category boundary, was expected to encounter the type of cognitive overload that has led researchers to report the hindrance of using minimal pairs in their previous experiments (Feldman et al., 2013a, 2013b; Hayes-Harb & Masuda, 2008; Jarvi, 2008). On the other hand, directing attention towards the category boundary, as in the explicit group, is more likely to result in participants noticing the novel Numana VOT. Since noticing is a prerequisite for acquisition (Schmidt, 1990, 1994, 1995, 2001, 2012), this phonological awareness could then also lead to

improved perception. Without such awareness being required of the implicit group, lesser gains were predicted by the noticing hypothesis. Given this, the paradigm of the current study will shed light on both the lexical-distributional hypothesis and the noticing hypothesis.

Regarding the second research question, the SLM and the PAM offer contrasting predictions. The SLM predicts that the participants of the current study (who are necessarily "late learners" of Numana) will have difficulty creating separate Numana categories with lower VOT boundaries. In the case of discriminating between prevoiced – short-lag stops in Numana, the SLM predicts that both will be perceived as English voiced stops, making discrimination of any of the three contrasts challenging for Numana learners.

Although clear predictions can be made regarding the shift in VOT category boundary, it is less clear what to expect regarding participants' perception of the change in POA from alveolar to dental. Specifically, it is unclear whether this change in POA should result in equivalence classification, especially since previous studies have not addressed this shift in POA through the lens of the SLM. The SLM predicts that for sounds that undergo equivalence classification, accurate perception – and in this case, discrimination – of sounds that are not exactly equivalent in the L1 and L2 will be more challenging than if L2 phones are not perceived as equivalent to sounds with an assumed preexisting an L1 category match. If equivalence classification does take place, and dental stops are assimilated to the English alveolar category, dentals should experience identical improvements in perception and discrimination as bilabials, since the differing dental/alveolar POAs will likely lead to an assimilation of Numana dentals to the existing English alveolar category (Flege, 1995). If the current study finds that dentals are, in fact, slower to experience improvement than bilabial stops, this will provide evidence against equivalence

classification taking place, since this would suggest that learners were more aware of a distinction between the representation that informs their expectations and the actual stimuli heard.

The PAM, on the other hand, predicts single category assimilation for the Numana /p/ - /b/, /t/ - /d/, and /k/ - /g/ contrasts, since all would likely be perceived as belonging to the English voiced category. However, given that the Numana /t/ - /d/ contrast is dental, it will be assimilated to English alveolar /t/ - /d/ - albeit a less good version of either. All in all, the PAM does not predict any difference in ease of learnability of any of these three contrasts, since all should assimilate as single category. In summary, an equivalent improvement on both POAs would provide support for the PAM, while a difference between the bilabial and dental stops would be consistent with differential equivalence classification according to the SLM.

The third research question asked about the ability of each of the groups to take the information learned in the training phrase and apply it to a new contrast (velar stop consonants) in the subsequent testing phase, as was demonstrated in McClaskey et al. (1983) and Nielsen's (2008) studies. It was predicted that the minimal pair group, which was predicted to gain more phonological awareness from the training phases, would therefore be better able to generalize new information regarding the location of the VOT category boundary and apply it to an analogous phone (Schmidt, 1990, 1994, 1995, 2001, 2012). Further support for phonological awareness facilitating learning comes from the discrepancy in findings between McClaskey et al.'s study and Maye and Gerken's (2001) study, as described earlier. In short, the participants of McClaskey et al.'s study were able to generalize to an untrained POA, whereas Maye and Gerken's participants could not, which Maye and Gerken attributed to the verbal prompting that took place only in McClaskey et al.'s study. Although neither group of the current study will receive explicit verbal prompting, the study's design, for which the explicit group alone will see side-by-side

presentations of minimal pair objects, is in many regards a non-verbal equivalent of an explicit prompt, meant to raise participants of the explicit group's phonological awareness. The implicit group, without such phonological awareness, was not predicted to be able to generalize to analogous phones with as much accuracy.

2.0 Methodology

The current study employs a word learning methodology to examine participants' perception of phonetic information and their reorganization of the existing phonological system, as speech perception is necessarily related to the recognition of not only sounds, but words and the meanings they convey (Broersma & Cutler, 2011). Participants were trained on an artificial language, Numana. The motivation for training on an artificial language rather than an existing natural language is so that participant's exposure to the language and its variation could be controlled (Curtin et al., 1998; Gomez & Gerken, 1999). Furthermore, artificial languages have been shown to provide an ecologically valid window through which researchers can examine the learning of natural languages (Moreton & Pater, 2012).

Training learners on lexical items provides more contextualization and ecologically valid instruction than is found, for example, in training paradigms that use isolated CV syllables. Also, because orthography has been shown to interact with L2 phonology (Bassetti, 2008), this study does not use graphemic representations of any of the phones examined. Although the instructions were presented in the participants' L1, beyond that, the focus of the study was on L2 speech perception alone. Training learners on lexical items allowed this study to test their perception of the phones contained within these items. Importantly, although learners' focus was on vocabulary acquisition, much like in a naturalistic L2 context, the input that they received in word learning also provides evidence for a lowered VOT category boundary (relative to English).

2.1 Numana

Although Numana is based loosely on Spanish, only phones that are both phonemic and roughly articulatorily equivalent in the L1, English, were included in the phonetic inventory of the language. For that reason, /m, n, s, l, tʃ / were included, but Spanish /ɲ, r, ɾ/ were excluded. Given the research questions, which address the perception of VOT, this requirement was not held for the target phones, /p, t, k, b, d, g/, which complete the consonant inventory of Numana. Four vowels /e, i, o, u/ were also included. The full consonant and vowel inventories can be seen in Tables 2 and 3 respectively.

Table 2 Numana Consonant Inventory

CONSONANTS	Bilabial	Dental	Alveolar	Postalveolar	Velar
Plosive	p b	t̪ d̪			k g
Nasal	m		n		
Fricative			s		
Lateral			l		
Affricate				tʃ	

Table 3 Numana Vowel Inventory

VOWELS	Front	Back
Closed	i	u
Closed-mid	e	o

Whereas English voiceless stops are produced with 30 ms or more of aspiration, Numana voiceless stops have between 0 and 20 ms of aspiration. Numana voiced stops are produced with 20-40 ms of prevoicing, or negative VOT, whereas English voiced stops are typically produced with a VOT of approximately 0 (Lisker & Abramson, 1964). Thus, the phonetic overlap between Numana voiceless and English voiced stops results in an area of ambiguity in this 0-20 ms range. Participants were trained on lexical items with voiced /b, d, g/ with a VOT of -20, -30, or -40 ms and voiceless /p, t, k/ with a VOT of 0, 10, or 20 ms. They were then tested on items with five steps along the VOT continuum: -40 ms, -20 ms, 0 ms, 20 ms, and 40 ms. Table 4 below provides a summary of English and Numana prevoiced, short-lag, and aspirated stops and their phonological categorizations in each language.

Table 4 VOT Continuum of English and Numana

<table>
<tr><td>Full Prevoicing</td><td>Some Prevoicing</td><td>Short-lag</td><td>Some Aspiration</td><td>Full Aspiration</td></tr>
<tr><td>-40 ms VOT</td><td>-20 ms VOT</td><td>0 ms VOT</td><td>20 ms VOT</td><td>40 ms VOT</td></tr>
<tr><td colspan="2">Numana Voiced /b d g/</td><td colspan="2">Numana Voiceless /p t k/</td><td></td></tr>
<tr><td colspan="2"></td><td colspan="2">English Voiced /b d g/</td><td>English Voiceless /p t k/</td></tr>
</table>

Although this VOT category boundary does bear many similarities to Spanish, by examining VOT devoid of the sociophonetic variation present in the alternation in many Spanish dialects (Corral & Wiedemann, 2009; Solon, Linford, & Geeslin, 2018), this study provides a more controlled view of the boundary and its learners' approximation to it, independent of other linguistic or social factors. For example, VOT is known to differ in duration depending on POA

in many languages, with velar stops having longer VOTs than alveolar stops, for example (Lisker & Abramson, 1964). One variable that will be maintained from Spanish, however, is the dental POA for the stops /t/ and /d/. This subtle difference in POA between Numana dentals and English alveolars could present either a benefit or an additional challenge to learners. For this reason, I examine whether learners can adjust their VOT category boundary and whether they can do so using a POA that only approximately mirrors that of their L1.

2.1.1 Lexical Items of Numana

For each of the two training conditions, 24 lexical items comprised the training stimuli, each corresponding to a picture of a nonce object taken from the Novel Object and Unusual Name (NOUN) Database (Horst & Hour, 2016). The NOUN objects are provided in Appendix A and were selected for their novelty in that participants would not have an existing L1 word onto which they could map the object, thus avoiding unintended word association. The 24 phonological forms were carefully designed to train learners to be sensitive to Numana's VOT category boundary of approximately 0 ms, while at the same time avoiding teaching any additional unintended rules or patterns. The phonological grammar is shown using a finite state automaton in Figure 1 below. In it, the phototactically legal combinations of segments are visualized. Each of the Numana lexical items begins with one of the consonants at the start position, accepts a vowel at the second state, and another consonant at the third state, At that point, the fourth state permits an optional consonant cluster, or alternatively allows for bypassing this consonant and instead allows the final vowel at the fifth state.

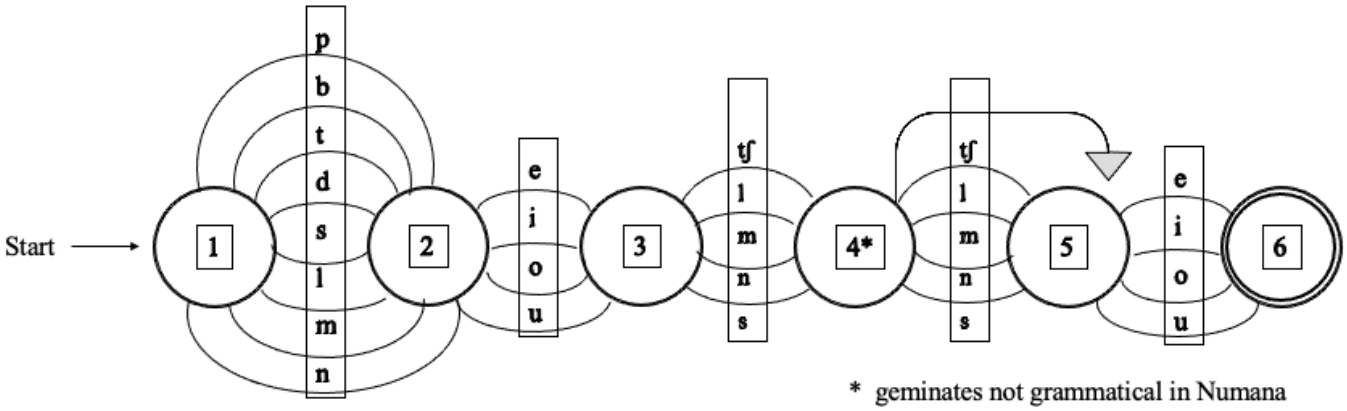

Figure 1 Finite State Automoton of Numana Phonology

The two conditions differed slightly in their lexical items, as described below, but in either case, were made up of three types: bilabial-initial, dental-initial, and distractors. Velar items were not included in the training phases but were tested at pretest and posttest to measure the degree to which training on bilabial and dental items facilitated generalizability. Thus, the bilabial-initial and dental-initial items (referred to collectively from here on as 'critical items') provided evidence for Numana's lower VOT category boundary. Distractors existed to balance the segment and syllable frequencies and to prevent participants from becoming aware of the purpose of the experiment.

Critical items differed between the two conditions such that explicit items were made up of voiced and voiceless minimal pairs (i.e. /*penso*/ and /*benso*/). This was not the case for the implicit condition (i.e. /*pelso*/ and /*benle*/), although the same number of segments and their distribution across the items overall was held constant across conditions. Both testing and training items are shown below in Tables 5 and 6.

Table 5 Explicit Condition Items

	Training Items	Generalized Testing Items	Novel Testing Items
Bilabial-initial	penso, benso	tenso, denso, kenso, genso	pilse, bilse, tilse, dilse, kilse, gilse
	pilsu, bilsu	tilsu, dilsu, kilsu, gilsu	
	posle, bosle	tosle, dosle, kosle, gosle	
	pultʃi, bultʃi	tultʃi, dultʃi, kultʃi, gultʃi	
Dental-initial	tomse, domse	pomse, domse, komse, gomse	
	tenme, denme	penme, denme, kenme, genme	
	tunso, dunso	punso, bunso, kunso, gunso	
	timlu, dimlu	pimlu, bimlu, kimlu, gimlu	
Distractors	netʃu		
	noli		
	memi		
	lime		
	none		
	tʃulo		
	tʃuso		
	misu		

Table 6 Implicit Condition Items

	Training Items	Generalized Testing Items	Novel Testing Items
Bilabial-initial	pelso, benle	telso, delso, kelso, gelso	pilse, bilse, tilse, dilse,
		tenle, denle, kenle, genle	kilse, gilse
	pilsu, binsu	tilsu, dilsu, kilsu, gilsu	
		tinsu, dinsu, kinsu, ginsu	
	posle, boltʃi	tosle, dosle, kosle, gosle	
		toltʃi, doltʃi, koltʃi, goltʃi	
	pulsu, bultʃi	tulsu, dulsu, kulsu, gulsu	
		tultʃi, dultʃi, kultʃi, gultʃi	
Dental initial	tensu, denmi	pensu, bensu, kensu, genus	
		penmi, benmi, kenmi, genmi	
	timtʃo, dimne	pimtʃo, bimtʃo, kimtʃo, gimtʃo	
		pimne, bimne, kimne, gimne	
	tonse, donlu	ponse, bomse, komse, gomse	
		ponlu, bomlu, komlu, gomlu	
	tunmi, dumso	punmi, bunmi, kunmi, gunmi	
		pumso, bumso, kumso, gumso	
Distractors	netʃu		
	suli		
	memi		
	lime		
	none		
	tʃulo		
	liso		
	misu		

All critical items were CVCCV and all distractors were CVCV. This was done to include syllabic variety in the training items. Importantly, the stimuli were balanced such that all

consonants appeared in all consonant-permitting positions. Vowels were also balanced in both syllable positions and across items types. In other words, both groups had critical items that contained the same number of occurrences of each vowel in the first syllable and in the second syllable. The same was true of the distractors for the two groups. The only exception was for bilabials and dentals, which appeared in word initial position only in order to provide greater control over the phonetic environment. Otherwise, every effort was made to balance each segment for overall frequency and the frequency of the neighboring vowel.

2.1.2 Stimuli Construction

Stimuli were recorded by five trained linguists, two men and three women, in a sound-attenuated booth using a high-quality microphone and Praat phonetics software at a 44,100 Hz sampling rate. Pre-voiced, short-lag, and aspirated versions of each item were read three times from a list that was randomized for each speaker. One woman speaker (Woman 1) was used in pretest, training, and post-test portions of the experiment. The training was composed of Woman 1 from the pretest, one additional woman, and two men speakers. For the post-test, Woman 1 and a third, novel woman's voice were included to test for the effect of a familiar versus a novel speaker. Table 7 provides a brief summary of the tasks in which all five speakers appeared.

Table 7 Summary of Speakers

Pretest	Training	Post-test
Woman 1	Woman 1	Woman 1
	Woman 2	Woman 3
	Man 1	
	Man 2	

2.1.3 Acoustic Properties of Stimuli

Recorded stimuli were acoustically manipulated using Praat to ensure that their VOTs fell within a specified range: -40, -30, -20, 0, 10, 20, and 40 ms. Specifically, the VOT of the short-lag tokens was modified to be exactly 0 ms. To create an aspirated token, aspiration was taken from an aspirated token and spliced onto the short-lag token at the first zero-crossing after the stop burst. The duration of the aspiration of this token was then modified to be either 10, 20, or 40 ms. The same procedure was followed with prevoicing to create the prevoiced tokens; prevoicing was taken from a prevoiced token, spliced onto the last zero-crossing before the stop burst of a short-lag token with exactly 0 ms of voicing, and then the prevoicing was modified to be 40, 30, and 20 ms in duration (of negative VOT).

2.2 Participants

Participants (n=44) were randomly assigned to one of two training conditions, implicit (n=23) and explicit (n=21). All were L1 English speakers and had no more than two years of

experience learning foreign languages other than American Sign Language. Information about the participants is summarized in Table 8 below. Participants were recruited from undergraduate courses and from flyers posted around the university. Participants were compensated $40 at the completion of all five sessions.

Table 8 Summary of Participant Language History

Implicit (n=23)	**Explicit** (n=21)
16 women, 7 men	13 women, 8 men
Avg. Age (years) = 21.00 (SD=3.65)	Avg. Age (years) = 20.68 (SD=2.69)
Avg. Years Experience Studying Spanish = 1.19 (SD=1.09)	Avg. Years Experience Studying Spanish = 1.41 (SD=1.36)

2.3 Experiment Procedure

Learners participated in five training sessions spread across five days. In all cases, all five sessions were completed within two weeks. Prior to coming into the lab, participants completed a language history questionnaire. Participants who met the desired criteria were invited to participate. On the first day in the lab, prior to training, participants completed the pretest, composed of two tasks. The first, an imitation task, presented participants with an auditory stimulus and asked that they repeat it as accurately as possible. Imitation items in the pretest included items that were similar to the training items, but no training items were included. These were created by combining attested sounds taken from the training set in a novel, phonotactically grammatical order (for example, training items ‘posle’, ‘pultʃi’ and ‘benso’ were recombined into

imitation item ‘polso’). Twenty-four items were included, and the task took approximately two minutes to complete. An imitation task was selected in order to avoid bias from orthography. Although the data from this imitation task will not be analyzed within this dissertation, later analyses will explore the role of purely perceptual training on production (for an overview, see Sakai & Moorman, 2018).

The second pretest task was an AX discrimination task. The pre-training location of participants’ VOT category boundary was examined by asking participants to decide whether each AX auditory pair contained the same word twice or two different words. For each trial, words A and X were spoken by Woman 1. Participants completed a three-item supervised practice block to confirm their understanding of the task before continuing. An item parallel in structure to (but not present in) the critical training words (refer to Tables 5 and 6 above) was presented at five locations along the VOT category boundary: -40 ms, -20 ms, 0 ms, 20 ms, 40 ms. Each of the five VOTs was paired in the following way: -40/-40, -40/-20, -40/0, -40/20, -40/40, such that each AX pair was presented forwards (AX) and backwards (XA), resulting in 25 trials for each item.

These same 25 items were tested with all three POAs by substituting the critical segment of the testing item to test bilabials, dentals, and velars on the same rhyme word. This resulted in 75 trials in total, and the task required approximately 5 minutes to complete.

For the training, occurring over the course of five days, participants completed three training phases on each day. Phase A was an exposure phrase, during which nonce visual objects from the NOUN database (Horst & Hout, 2016) were presented in turn in conjunction with a corresponding auditory stimulus. Throughout the training phases, the auditory stimuli were randomly distributed among VOT productions at 0, 10, or 20 ms for the voiceless and -20, -30, or -40 ms for the voiced stimuli. Participants were instructed to study the labels of the objects and

were informed that they would be tested on them later. Participants' attention was not explicitly drawn to VOT at any point during the first training phase. All objects and their labels were presented once in a random order in block one and then again in a different random order in block two. With 24 training items and distractors each presented twice, there were a total of 48 exposures, which required approximately 10 minutes to complete.

After all lexical items were presented with their auditory labels twice, participants began training Phase B, the first of two two-alternative forced-choice tasks. In this phase, participants were presented with two visual objects on the screen and heard one auditory stimulus from the set of 24 to which they had been exposed in Phase A. Their task was to select the object that corresponded to the stimulus that they heard. For example, Figure 2 provides a visualization of the training task. Two objects are pictured; the left object was labeled "penso" and the object on the right was "timlu". Participants heard either "penso" or "timlu" and were asked to select the appropriate object.

Figure 2 Example of Training Task

Within this phase, all object pairs were non-minimal pairs. The intent was for learners in both conditions to continue to acquire labels for each of the visual objects, without an explicit focus on VOT. At this stage, it was expected that participants in the explicit condition would assume that some of the minimal pair labels of the objects were homophonous. For example, participants whose VOT category boundary is around 20 ms as might be expected for native English speakers, may assume that [penso] (0 ms) and [benso] (-40 ms) were homophones rather than minimal pair words. The nonce object associated with each of the 24 critical lexical items was presented four total times within this training phase. Twice it appeared as the left picture and twice as the right picture. This object was the target of the auditory stimulus once per position on the screen. In other words, participants heard each lexical item spoken aloud two times within this training block. Additionally, participants received feedback, which is described in detail below, providing additional exposures to the lexical items. Given that each of the 24 items was presented twice, this phase required approximately eight minutes. On testing days, when sessions were considerably longer due to other tasks, this training phase ran as described above. On training-only days (days 2-4), however, this phase was set to double the number of trials, and therefore required approximately 16 minutes.

It was expected that participants in the explicit condition would incorrectly deduce that VOT voiced/voiceless minimal pairs were homophones based on their training in Phase B. Thus, in the final training task, Phase C, these VOT minimal pairs were presented to learners in the explicit condition only in a more explicitly contrasting way. Rather than pairing non-minimal pairs as was done in Phase B, in Phase C each visual object was presented with the visual object of its corresponding minimal pair. In other words, participants heard [penso] and were asked to select between the objects that correspond to [penso] and [benso]. The purpose of this task was to force

participants in the explicit condition to attend to the fine-grained phonetic differences between the minimal pairs, and in doing so, it was expected that participants would lower their VOT category boundary in the direction of the Numana boundary. Importantly, however, this minimal pair, phonetically explicit pairing of Phase C only occurred for the critical stimuli and not for the distractors, which did not form minimal pairs.

In the implicit condition, items were not (and could not be) presented in minimal pair form. Identical to Phase B in number of trials, this phase took approximately 8 or 16 minutes, depending on the training day. For Phases A, B and C, the voices of the four speakers used in the training tasks (two men and two women) were selected randomly for each trial on each training day. In this way, participants were exposed to all four voices in the greatest number of contexts. In addition, the VOT of the voiced and voiceless stimuli varied randomly within each category, with voiced segments having either -40, -30, or -20 ms of voicing, and voiceless segments having either 0, 10, or 20 ms of aspiration.

As noted above, both Phases B and C provided participants with feedback regarding their performance on each trial. Following Iverson and Evans (2007), for correct responses, participants saw a green box appear (as shown in Figure 3), framing the object they correctly selected, and then heard a repetition of the auditory stimulus.

Figure 3 Feedback for 'Correct' Responses

Incorrect responses resulted in an auditory repetition of the target lexical item, with the correct object framed in green, followed by a red box framing their incorrect selection and an auditory rendition of it, and then finally a return to the auditory rendition and green frame of the object they should have selected. A portion of this feedback sequence for incorrect responses is shown in Figure 4.

Figure 4 Feedback for 'Incorrect' Responses

Training phases A-C were repeated each day, for a total of five times per phase. Following participants' completion of five days of training, a post-test was administered on the fifth day. The post-test was composed of the repetition of the imitation and discrimination tasks from the pretest. In addition to testing the trained items and the generalized POAs, an untrained item for both the bilabial-initial and dental-initial condition was tested in the same manner. Although the untrained items were novel, they matched the trained items in terms of syllable structure and segment co-occurrence. In other words, no new syllables or segments were presented, but the order in which the familiar syllables were combined resulted in novel words. Generalization to a novel speaker was also assessed and as such, a new speaker, Woman 3, as well as the speaker from the pretest and training, Woman 1, were included in this task. These total 750 trials, broken into three blocks, required approximately 45 minutes to complete.

Participants also repeated the imitation task from the pretest but the words used were not repeated. Instead, participants were tested on a sampling of training items (one minimal pair from each POA outlined above) as well as the same items analogized to either bilabial or dental and velar POAs. Phones other than stop consonants were included in the imitation stimuli list and will be analyzed in future studies but will not be reported here. Like in the pretest, the imitation task contained 24 total items and required approximately two minutes to complete.

Finally, participants completed a brief questionnaire targeting their experience with the study and the strategies they could identify that they used to learn the Numana words. This questionnaire required no more than five minutes to complete. All tasks were administered using E-Prime version 3.0.

In total, day one including both the pretest and training phases lasted approximately 35 minutes. Days two through four consisted of only training phases A-C, and phases B and C were

twice the duration of day one. Thus, days two, three, and four required approximately 30 minutes, but participants did tend to finish more quickly as they became more and more accurate with the tasks and less feedback on trials was required. Day 5, including training, post-tests, and the exit questionnaire required approximately 90 minutes. The tasks for each phase and day are summarized in Table 9.

Table 9 Procedure Summary

Sessions	Phase:	Tasks	Phones	Approximate Time
1	Pretest:	Imitation	[p, t, k, b, d, g]	2 minutes
		Discrimination		10 minutes
1-5	Training A:	Exposure to items - presentation only	[p, b, t, d]	10 minutes
1-5	Training B:	Two alternative forced choice (2AFC) task with no minimal pairs + Feedback	[p, b, t, d]	8 minutes on sessions 1 & 5 16 minutes on sessions 2-4
1-5	Training C:	2AFC task + Feedback; Explicit group with minimal pairs, Implicit group with no minimal pairs	[p, b, t, d]	8 minutes on sessions 1 & 5 16 minutes on sessions 2-4
5	Post-Test	Discrimination	[p, t, k, b, d, g]	45 minutes
		Imitation		2 minutes
		Exit Questionnaire		3 minutes

3.0 Results

Results of the current study are separated first into results from the training phases B and C, presented as accuracy scores during each of the five training days. Following accuracy scores, I present the results of the discrimination tasks given at the pre- and posttests and discuss the changes that occurred across test times as a function of group, training condition, and additional linguistic variables.

3.1 Accuracy

Accuracy of lexical decisions (correctly identifying which of two visual objects corresponded to the auditory stimulus they heard for any given trial) at each training session was calculated by including only the critical items (i.e., excluding distractors) and is summarized with mean and standard deviation values in Tables 10 and 11 below. Recall that data were not collected for Training A, since this task was presentation only and participants did not log any responses.

Table 10 Mean (SD) Accuracy for Training B 2AFC Task

	Session 1	**Session 2**	**Session 3**	**Session 4**	**Session 5**
Implicit	.887 (.109)	.952 (.119)	.971 (.090)	.980 (.092)	.970 (.109)
Explicit	.853 (.097)	.950 (.089)	.972 (.102)	.980 (.084)	.966 (.108)

Table 11 Mean (SD) Accuracy for Training C 2AFC Task

	Session 1	Session 2	Session 3	Session 4	Session 5
Implicit	.880 (.121)	.927 (.138)	.949 (.158)	.931 (.161)	.950 (.163)
Explicit	.692 (.043)	.814 (.087)	.889 (.090)	.921 (.091)	.906 (.160)

In analyzing the factors that best predicted participants' accuracy, the current study followed Barr, Levy, Scheepers, and Tily (2013) in using nested model comparisons to determine which grouping factors significantly improved model fit. First, Training B 2AFC responses were examined using a baseline generalized linear mixed effects model that examined accuracy as a function of session with a random intercept for subject. A random intercept for item did not converge, but recall that the items in both groups were matched for segment and syllable frequency across conditions. The addition of group as a fixed effect did not significantly improve the model fit (χ^2 (1) = 1.143, n.s.). Thus, the final model for the Training B 2AFC task examined accuracy as a function of session and included a random intercept for subject.

Significant differences in accuracy were found between Session 1 and Session 2 (*ß*=0.970, *z*=6.574, p<.001) and between Session 2 and Session 3 (*ß*=0.702, *z*=3.866, p<.001), with accuracy improving in later sessions. The difference between Session 3 and Session 4 approached significance (*ß*=0.441, *z*=1.933, p=.053), again with participants improving their accuracy scores in the later session. That being said, the difference between Session 4 and Session 5 was significant (*ß*=-0.495, *z*=-2.100, p<.05), in that participants were less accurate at Session 5 than at Session 4.

Nested models were also used to compare accuracy across groups and training sessions for Training C. A baseline model examined accuracy as a function of session and included a random intercept for subject. Again, a random intercept for item did not converge. The addition of group as a fixed effect significantly improved the model fit (χ^2 (7) = 17.704, p<.001). A significant

difference between groups was found such that the implicit group was more accurate than the explicit group overall (ß=1.072, *z*=4.501, p<.001). Significant differences were also found between Session 1 and Session 2 (ß=0.773, *z*=8.494, p<.001), between Session 2 and Session 3 (ß=0.652, *z*=6.392, p<.001), and between Session 3 and Session 4 (ß=0.238, *z*=2.021, p<.05). No significant difference was found between Session 4 and Session 5 (ß=-0.067, *z*=-0.469, n.s.).

Difference scores were also calculated for each group to attempt to compare the efficacy of the training that each group received. These scores were calculated by subtracting the score at Session 1 from the score at Session 5. Greater gains in accuracy should be reflected in higher difference scores. Mean difference scores and standard deviations for each group for Training in Phases B and C are presented below in Table 12.

Table 12 Difference Scores for Training B and C TACF Tasks

	Training B 2AFC Difference Score	**Training C 2AFC Difference Score**
Implicit	0.076 (.068)	0.070 (.082)
Explicit	0.143 (.141)	0.232 (.096)

To compare Difference Scores across groups for either training task, a linear fixed effects model with Difference Score as a function of group was used. A random intercept for subject was not included in the model, since each subject was represented in the data by exactly one Difference Score. There were no significant differences between groups for the Training B 2AFC task (ß=-0.067, *t*=-1.303, n.s.), which was the same for both groups. For the Training C 2AFC task recall that only the explicit group saw the side-by-side presentation of objects whose labels were minimal pairs. For this task, the explicit group had significantly higher difference scores than the implicit group (ß=-0.163, *t*=-4.482, p<.001.). In other words, the explicit group experienced more

improvement over the course of the training. Models comparing the effectiveness of Training B and Training C 2AFC tasks for any one group were not run, since participants in either group completed both, meaning that it is impossible to separate the effects that one training had on the accuracy results of the other.

Because the explicit task was inherently more difficult than the implicit task for Training C TACF task where minimal pairs were used for the explicit group only, it follows that scores at all sessions (and especially at Session 1) were lower than the implicit scores for the same sessions. Given this, comparing Differences Scores does not accurately reflect the efficacy of the training received by either group. For that reason, Relative Improvement scores were calculated by subtracting the score at Session 1 from 1.0, to indicate change relative to the amount of room to improve each group had at Session 1. Then, Difference Scores were divided by Room to Improve scores.

Difference Score = Session 5 Score – Session 1 Score

Room to Improve Score = 1.0 - Session 1 Score

Relative Improvement Score = Difference Score / Room to Improve Score

This measure then indicates the amount of improvement relative to how much improvement was possible. Mean Relative Improvement scores and standard deviations are summarized in Table 13 below.

Table 13 Relative Improvement Scores (SD) for Training B and C 2AFC Tasks

Group	Training B Relative Improvement Score	Training C Relative Improvement Score
Implicit	.477 (.344)	.568 (.653)
Explicit	.645 (.885)	.760 (.240)

To compare Relative Improvement scores across groups for either Training task, a linear fixed effects model with Relative Improvement as a function of group was used. A random intercept for subject was not included in the model, since each subject was represented in the data by exactly one Relative Improvement score. There were no significant differences between groups for Training B ($ß$=0.167, t=0.508, n.s.) or for Training C ($ß$=-0.192, t=-1.022, n.s.) 2AFC tasks. Models comparing the effectiveness of either task for any one group were not run, since participants in either group completed both, meaning that it is impossible to separate the effects that one training had on the accuracy results of the other.

Although these Relative Improvement scores help to visualize the improvement within the window of possible improvement, it is important to bear in mind that as participants approached ceiling levels of accuracy, there was less room for improvement overall. This is especially true for the implicit group who were at 88% and 89% accuracy at Session 1 for Training B and C 2AFC tasks respectively. It follows that it is more challenging for learners to continue to make gains as they approach ceiling levels of accuracy.

It is also noteworthy that both groups performed most accurately for the Training B 2AFC task on Day 4, with a slight decrease in accuracy occurring on Day 5. A similar pattern is found for the Training C 2AFC task, with the implicit group demonstrating their highest levels of accuracy on Day 3 and the explicit group on Day 4.

3.2 Discrimination

Next, I turn to discrimination data that were collected at the pretest and posttest for AX pairs. Due to small token counts in the pretest, the predictors of interest could not all be included in a single model, so several different models were fit, each with one to two variables. In this way, cells of at least 10 for each participant for each comparison were maintained, and pertinent interactions between variables were still able to be examined.

Participants' accuracy of 'same'/'different' responses were converted into d' scores in order to take into account both the participants' sensitivity to the actual differences between the AX pairs, as well as their false alarm rate, or their bias to indicate a difference where there is none. d' scores of zero indicate responses at chance levels of accuracy. Scores of 1 represent accuracy levels of approximately 70%. Overall, higher d' scores indicate higher accuracy and less false alarms.

The following variables will be discussed in detail: the absolute value of the change in VOT between stimuli A and X, the direction of change between the VOT of the two stimuli, the relationship between the stimuli A and X with regards to category boundary and voicing, and the relationship between the stimuli A and X with regards to category boundary and ambiguity. Additionally, variables related to analogy will be discussed, namely speaker, item, and POA. Finally, interactions between speaker and item, item and POA, and speaker and POA will be examined.

3.2.1 Absolute Change in AX

To analyze sensitivity to the differences in various amounts of change in VOT between AX pairs, a linear mixed effects model was used to examine d' scores as a function of the absolute difference in VOT between AX pairs, test time, an interaction between the absolute difference and test time, and included a random intercept for subject. Here and throughout the results section, random intercepts cannot be included for item, since d' scores average over all items. Recall, however, that items were controlled for by design. No significant difference between groups was found at the pretest (ß=-0.144, t=-0.912, n.s.), so it was not included in the final model. There were significant differences between trials in which AX differed by 20 ms as compared to 40 ms (ß=0.755, t=10.662, p<.001), as compared to 60 ms (ß=1.231, t=17.381, p<.001), and as compared to 80 ms (ß=1.392, t=19.658, p<.001). There was also a significant difference between trials differing by 40 ms as compared to 60 ms (ß=0.465, t=6.465, p<.001) and as compared to 80 ms (ß=0.637, t=8.928, p<.001). Finally, there was a significant difference between trials with a 60 ms difference as compared to 80 ms (ß=0.172, t=2.396, p<.05). In all cases, an increased duration difference between A and X stimuli led to increased d's, or a greater ability to differentiate.

There was a significant difference between test times, with posttest d' scores being higher overall (ß=0.462, t=4.770, p<.001). No significant difference was found between the groups' responses at posttest with respect to the absolute difference between AX stimuli (ß=0.027, t=0.165, n.s.). Significant differences were found between trials differing by 20 ms as compared to 40 ms (ß =0.792, t=11.248, p<.001), as compared to 60 ms (ß =1.404, t=19.936, p<.001), and as compared to 80 ms (ß =1.740, t=24.712, p<.001). There were also significant differences between trials differing by 40 ms as compared to 60 ms (ß=0.611, t=8.689, p<.001) and 80 ms

(ß=0.948, t=13.464, p<.001). Again, in all cases, increased duration of change in VOT between stimuli A and X resulted in increased sensitivity as indexed by d' scores.

There was also a significant interaction between the absolute change between A and X and test time, such that the change between test times was significantly greater for trials differing by 80 ms as compared to 20 ms (ß=0.348, t=2.541, p<.05) and as compared to 40 ms (ß=0.311, t=2.272, p<.05). The d' scores as a function of the absolute difference (in ms of VOT) between stimuli A and X is represented in Figure 5 below. As in each of the plots below, the independent variable being examined – in this case, absolute difference – is represented on the x-axis, and d' scores are represented on the on the y-axis. Higher d' scores indicate increased sensitivity to the differences between the stimuli for each of the two groups.

All together, these results indicate that AX pairs that are further apart on the VOT continuum are more identifiable as distinct. This finding simply points to the stimuli being perceived as they were intended by design. Overall, these results also show significant improvements across test times, suggesting that the training paradigm employed was successful in increasing accuracy at identifying distinct AX pairs as being distinct.

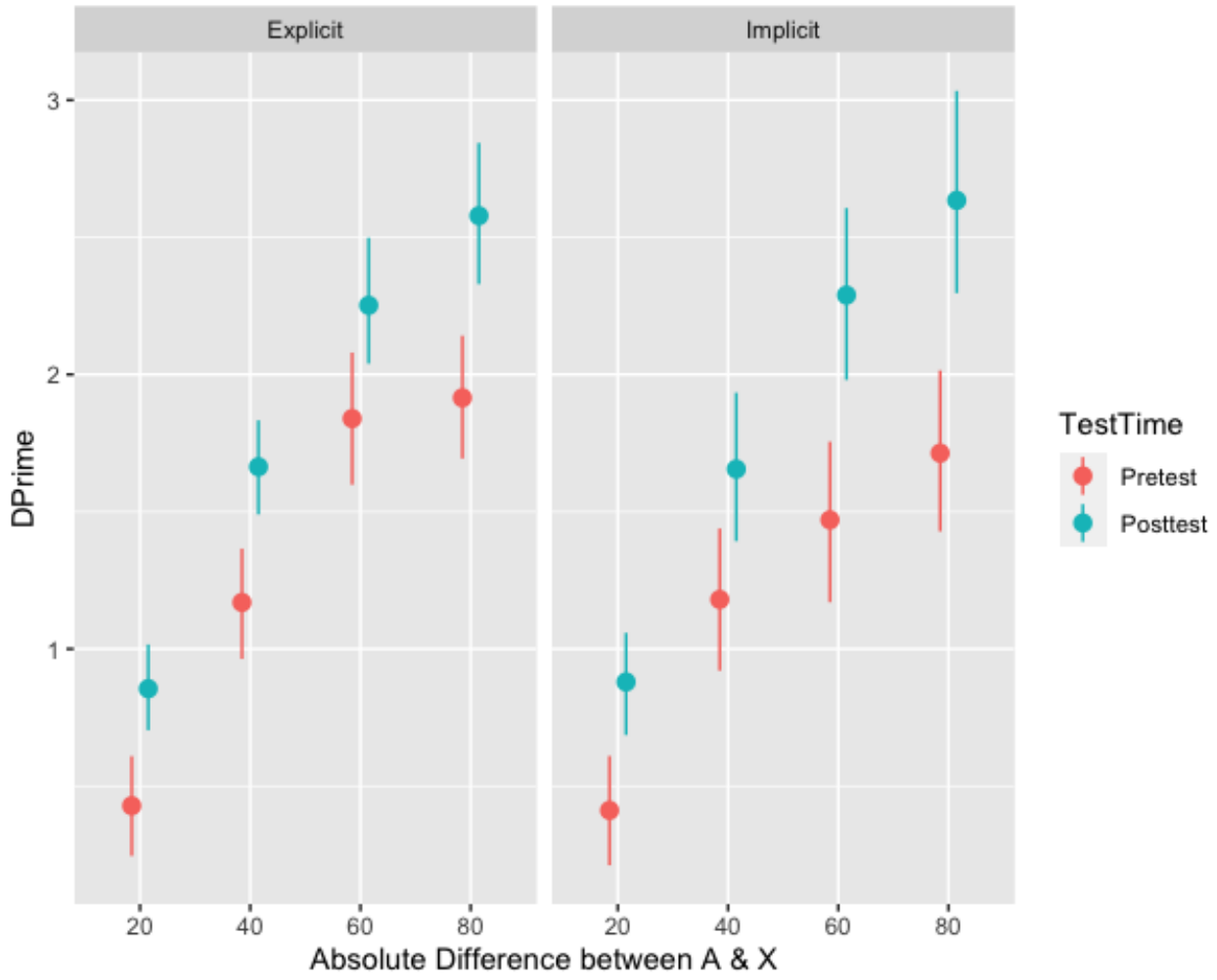

Figure 5 d' Scores for the Absolute Difference between A and X

3.2.2 Direction of Change of AX

In a second model, the direction of change between the VOT of A and X stimuli was examined. Again, a linear mixed effects model was used with d' scores as a function of the direction of change in VOT between AX pairs, test time, and an interaction between the two. The model also included group and a random intercept for subject. No significant difference was found between groups at the pretest (ß= -0.086, *t*= -0.588, n.s.) or the posttest (ß= -0.105, *t*= -0.749, n.s.), however there was a significant difference between increase and decrease trials at the posttest for

implicit condition (ß= 0.208, *t*=5.628, p<.001) that was not found for the explicit group (ß= -0.023, *t*=-0.658, n.s.). As shown in Figure 6, both groups responded overall more accurately when the direction of change between A and X stimuli was an increase (ß=0.319, *t*=6.591, p<.001).

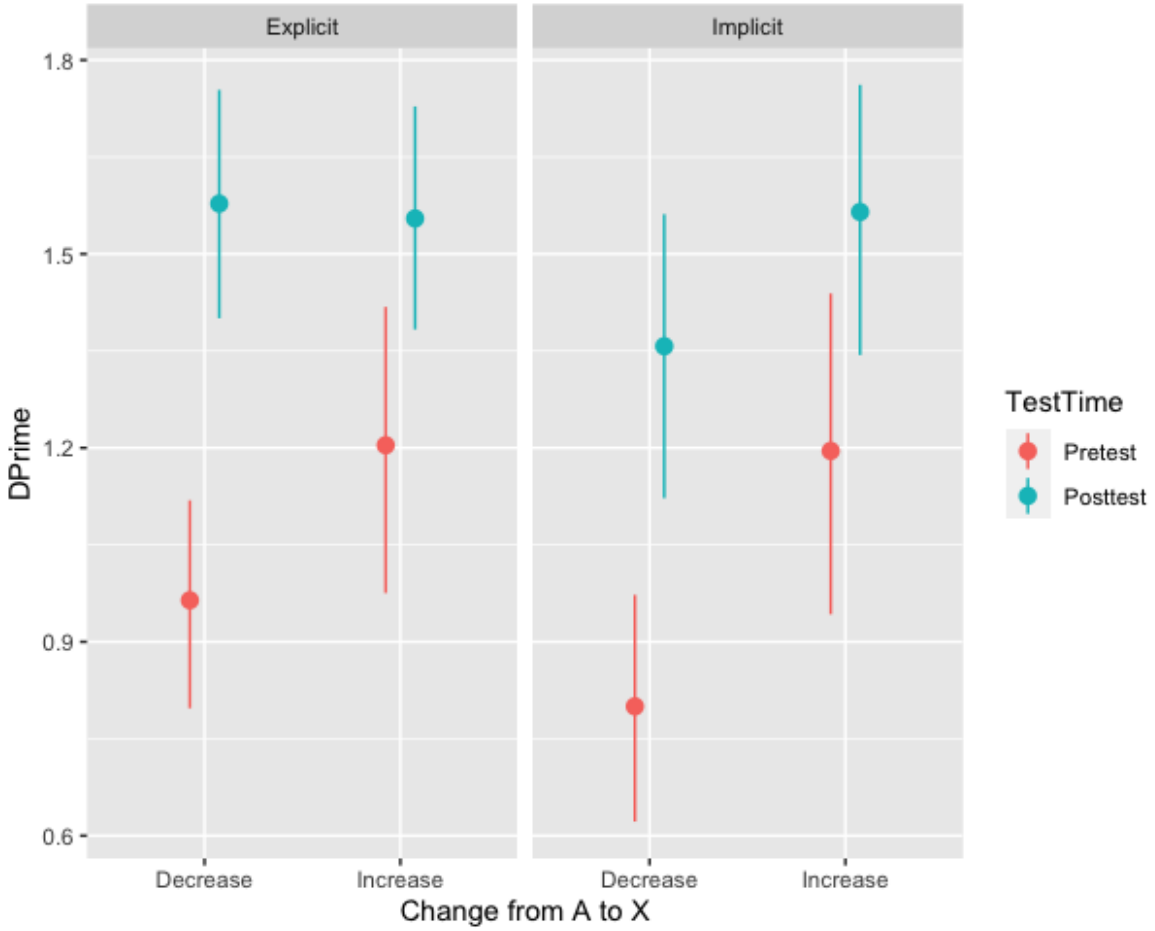

Figure 6 d' Scores for the Direction of Change between VOT of A and X Stimuli

There was a significant difference between test times, with higher d' scores at the posttest (ß=-0.598, *t*=-7.976, p<.001). There was also a significant interaction between the direction of change between the VOT of A and X stimuli and test time such that less change across test times occurred for trials in which the change from A to X was an increase (ß=0.224, *t*=2.122, p<.05).

The rationale for participants' differentiated responses to AX pairs on the basis of the direction of change is beyond the scope of the current study's research questions.

3.2.3 Category Boundary and Voicing

Next, the relationship between the stimuli A and X with regards to category boundary and voicing was examined. Trials were coded with regard to the category and voicing as being either 'Across', 'Within Voiced', 'Within Voiceless', or 'Within Aspiration'. Specifically, 'Within Voiced' trials were those for which both A and X had either 20 or 40 ms of prevoicing (-40/-40, -20/-20, -40/-20, -20/-40). 'Within Voiceless' trials were those for which either A or X had 0 ms of aspiration and the other stimuli had either zero, 20, or 40 ms of aspiration (0/0, 0/20, 20/0, 40/0, 0/40). 'Across' trials were those that straddled the (0 ms) VOT category boundary and those that compared 0 ms tokens to prevoiced tokens (-40/40, 40/-40, 20/-40, -40/20, -20/20, -20/40, 20/-40, 20/-20, 40/-20, -40/0, 0/-40, 0/-20, -20/0). Finally, 'Within Aspiration' trials were those with either 20 or 40 ms of aspiration for both A and X stimuli (40/40, 40/20, 20/40, 20/20). Sum coding was used to code this variable, since none of the levels were a baseline against which to compare the others.

A third linear mixed effects model was employed, with d' scores as a function of the categorization type, test time, an interaction between the two, and included a random intercept for subject. There was no significant difference between groups at pretest with regard to this categorization variable ($ß$=0.114, t=0.661, n.s.). There was a significant difference between Within Voiced trials and the overall average d' score ($ß$=0.077, t=2.201, $p<.05$), with these trials being overall significantly higher than the average of all trials. No significant differences were found

between the overall average d' score and the Across trials (ß=-0.045, *t*=-1.294, n.s.) or the Within Voiceless trials (ß=0.033, *t*=0.944, n.s.).

No significant difference between groups was found at the posttest (ß=-0.055, *t*=-0.393, n.s.). There were, however, significant differences between the overall average d' score and the d' scores for Within Voiced (ß=-0.149, -4.699, p<.001) and Within Voiceless trials (ß=0.188, *t*=5.927, p<.001) such that d' scores were lower than average for Within Voiced trials, and higher than average for Within Voiceless trials. There was also a significant difference between test times, with higher d' scores at posttest (ß=0.534, *t*=5.764, p<.001).

There was also a significant interaction between this categorization variable and test time such that Within Voiced trials improved significantly less across test times as compared to Across trials (ß=-0.274, *t*=-2.099, p<.05). Additionally, there was more change across test times in d' scores for Within Voiceless trials than for Within Voiced trials (ß=0.382, *t*=2.919, p<.01). Finally, an interaction between categorization and test time bordered on significance such that Within Aspiration trials improved marginally more than Within Voiced trials across test times (ß=0.250, *t*=1.911, p=0.057). These results are shown in Figure 7.

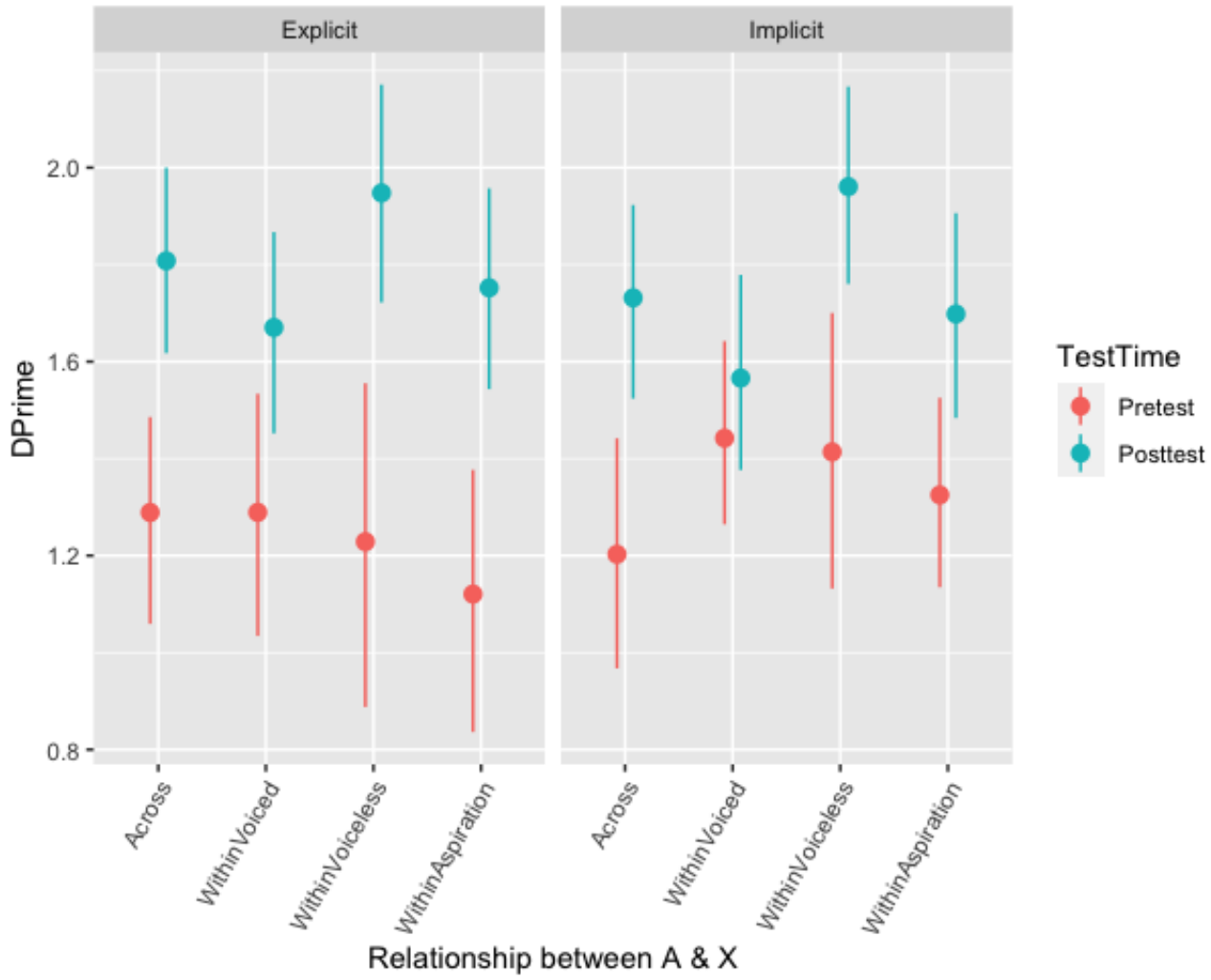

Figure 7 d' Scores for Boundary / Voicing Categorization

The Within Voiced category is among the most interesting findings of the current study thus far. The implicit group did not improve significantly across test times for trials of this type (*ß*=0.139, *t*=1.050, n.s.), but the explicit group did (*ß*=0.563, *t*=8.658, p<.001). This finding that the explicit training alone led to an improvement in discrimination within the prevoiced portion of the continuum provides direct evidence for the effectiveness of the training using minimal pairs.

3.2.4 Category Boundary and Ambiguity

Trials were also coded for the relationship between A and X stimuli with regard to their relationship to the category boundary and ambiguity. Levels of this categorization variable included 'Across Ambiguous', 'Within Ambiguous', 'Within Unambiguous', and 'Across Unambiguous'. These pairs were categorized in a similar way to the previous categorization variable with regard to 'across' and 'within'. Additionally, however, ambiguity was taken into account such that trials for with either A or X (but not both) had 0 ms of VOT were classified as ambiguous. Accordingly, 'Across Ambiguous' trials contained the following AX pairs: -40/0, 0/-40, 0/-20, -20/0. 'Across Unambiguous' trials contained -40/40, 40/-40, 20/-40, -40/20, -20/20, -20/40, 20/-20, 40/-20 pairs. 'Within Ambiguous' trials were those with AX pairings of 40/0, 0/20, 0/40, 20/0. Finally, 'Within Unambiguous' trials were the following AX pairs: 40/40, 40/20, 20/40, 20/20, -40/-40, -20/-20, 0/0, -40/-20, -20/-40.

Unlike with the previous categorization variable, this variable was treatment coded, with Within Unambiguous as the base level, given its approximation to a 'Same' classification.

An additional linear mixed effects model examined d' scores as a function of the categorization type, test time, and included a random intercept for subject. At the pretest, there was no significant difference between groups ($ß$=-0.024, t=-0.159, n.s.), so this variable was not included in the model. Significant differences were found between Within Unambiguous and Across Ambiguous trials with Across Ambiguous trials being responded to significantly less accurately ($ß$=-0.822, t=-11.360, p<.001). Participants were also less accurate in responding to Within Ambiguous trials than they were with Within Unambiguous trials ($ß$=-0.257, t=-3.545, p<.001). When compared to Within Unambiguous trials, participants were significantly more accurate, however, in responding to Across Unambiguous trials ($ß$=0.024, t=2.813, p<.01). Across

Ambiguous trials were responded to significantly less accurately than Within Ambiguous trials (ß=0.566, *t*=7.814, p<.001) and than Across Unambiguous trials (ß=1.027, *t*=14.173, p<.001). Finally, Within Ambiguous trials were responded to significantly less accurately than Across Unambiguous trials (ß=-0.461, *t*=-6.358, p<.001). Notably, Across Ambiguous (when set as the intercept) is significantly different from zero (ß=0.563, *t*=6.488, p<.001).

Again, at the posttest, there was no significant difference between groups (ß=-0.007, *t*=-0.048, n.s.). Significant differences were found, though, between Within Unambiguous and Across Ambiguous trials (ß=-0.828, *t*=11.608, p<.001), with higher d' scores for Within Unambiguous. Across Unambiguous d' scores were significantly higher than Within Unambiguous trials (ß=0.496, *t*=6.956, p<.001). There was also a significant difference between Across Unambiguous and Across Ambiguous (ß=1.324, *t*=18.564, p<.001). Significant differences were found between Across Unambiguous and Within Unambiguous trials (ß=1.028, *t*=14.173, p<.001), and between Across Unambiguous and Within Ambiguous trials (ß=0.566, *t*=7.814, p<.001). In each case, Across Unambiguous trials were responded to more accurately than other trial types. No significant difference was found difference between Within Ambiguous and Within Unambiguous trials (ß=-0.018, *t*=-0.246, n.s.).

As compared to the pretest, d's were significantly higher overall at the posttest (ß=-703, *t*=-7.538, p<.001). A significant interaction was found between test time and categorization such that Across Ambiguous (ß=0.297, t-2.257, p<.05) and Within Unambiguous trials (ß=0.292, *t*=2.221, p<.05) improved significantly less across test times than Across Unambiguous trials. A marginally significant interaction was found between Across Ambiguous and Within Ambiguous trials and test time, such that Within Ambiguous trials improved marginally more from pretest to posttest (ß=-0.244, *t*=-1.855, p=0.065). An additional marginally significant interaction was

found, such that marginally more gains were made for Within Ambiguous trials as compared to Within Unambiguous trials (ß=0.239, *t*=1.819, p=0.070). These results are shown in Figure 8.

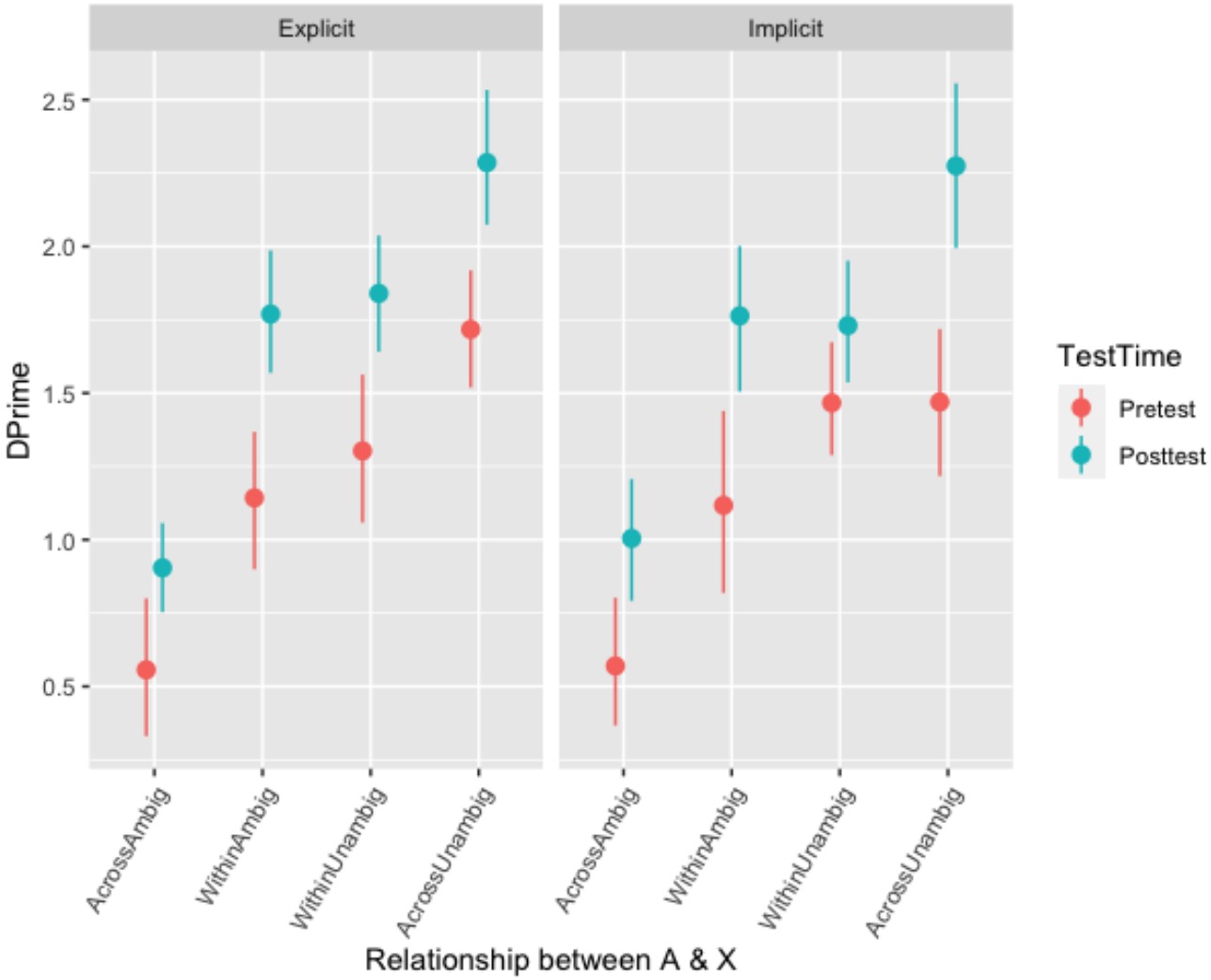

Figure 8 d' Scores for Boundary / Ambiguity Categorization

3.2.5 Speaker

Next, d' scores as a function of whether the speaker of the trial was a new speaker (not heard at the pretest or during training) or a familiar speaker (heard at the pretest and during

training) was examined using a linear mixed effects model that examined d' scores as a function of speaker and test time, and included a random intercept for subject.

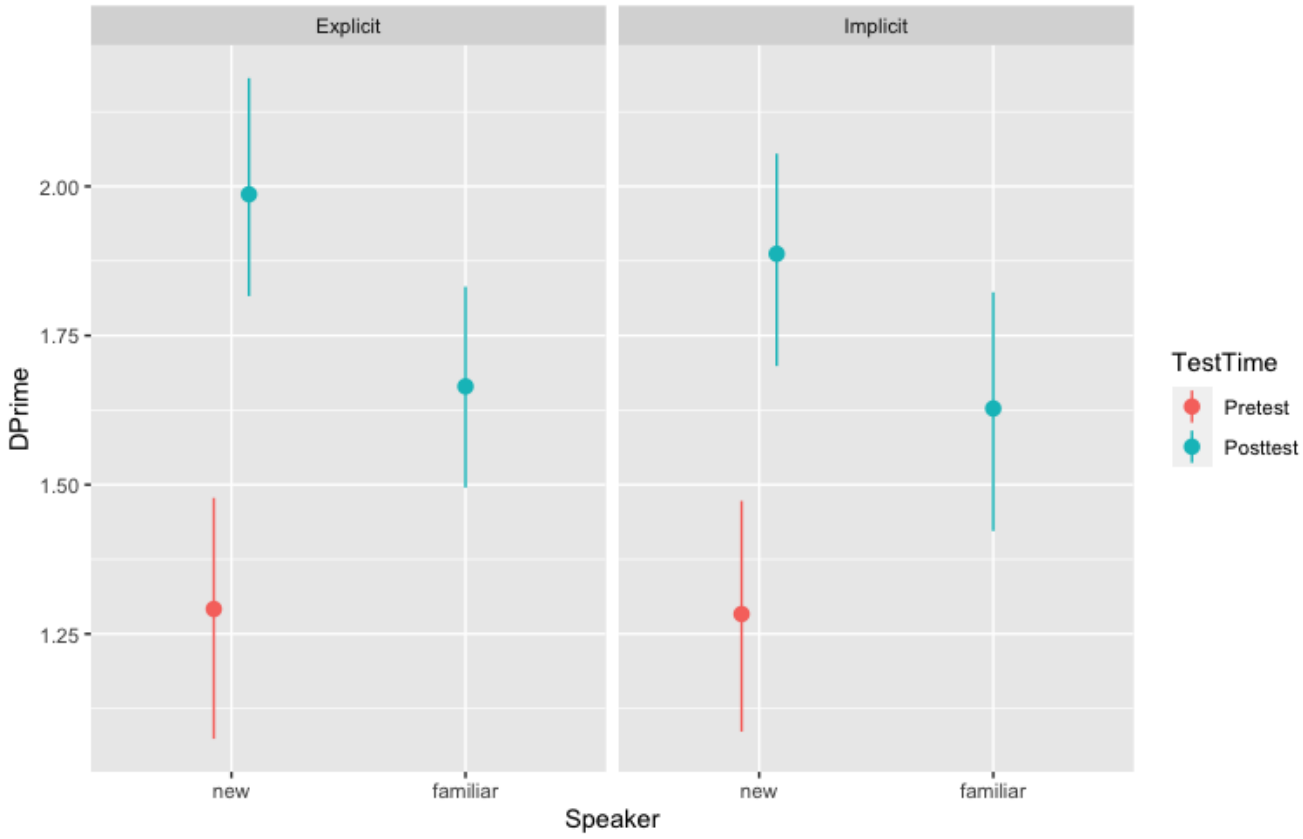

Figure 9 d' Scores for New versus Familiar Speaker

As shown in Figure 9, with regards to speaker, groups were not significantly different at the pretest (*ß*=-0.008 *t*=-0.057, n.s.) or the posttest (*ß*=-0.068, *t*=-0.507, n.s.) and thus were not included in the final model. At the posttest, both groups were significantly more accurate at discriminating tokens produced by the new speaker as compared to the familiar speaker (*ß*=0.290, *t*=10.129, p<.001). For both groups, d' scores were significantly higher following training, both for tokens produced by the new speaker (*ß*=0.659, *t*=8.985, p<.001) as well as by the familiar speaker (*ß*=0.986, *t*=9.768, p<.001). That the new speaker was perceived with higher accuracy by

both groups than the speakers on which they were trained is an unexpected finding that will be explored in depth in the discussion section.

3.2.6 Item

Whether the item of a given trial was familiar to the participant from the training phases was also examined using a linear mixed effects model that examined d' scores as a function of item. The model included test time and item as fixed effects, and included a random intercept for subject.

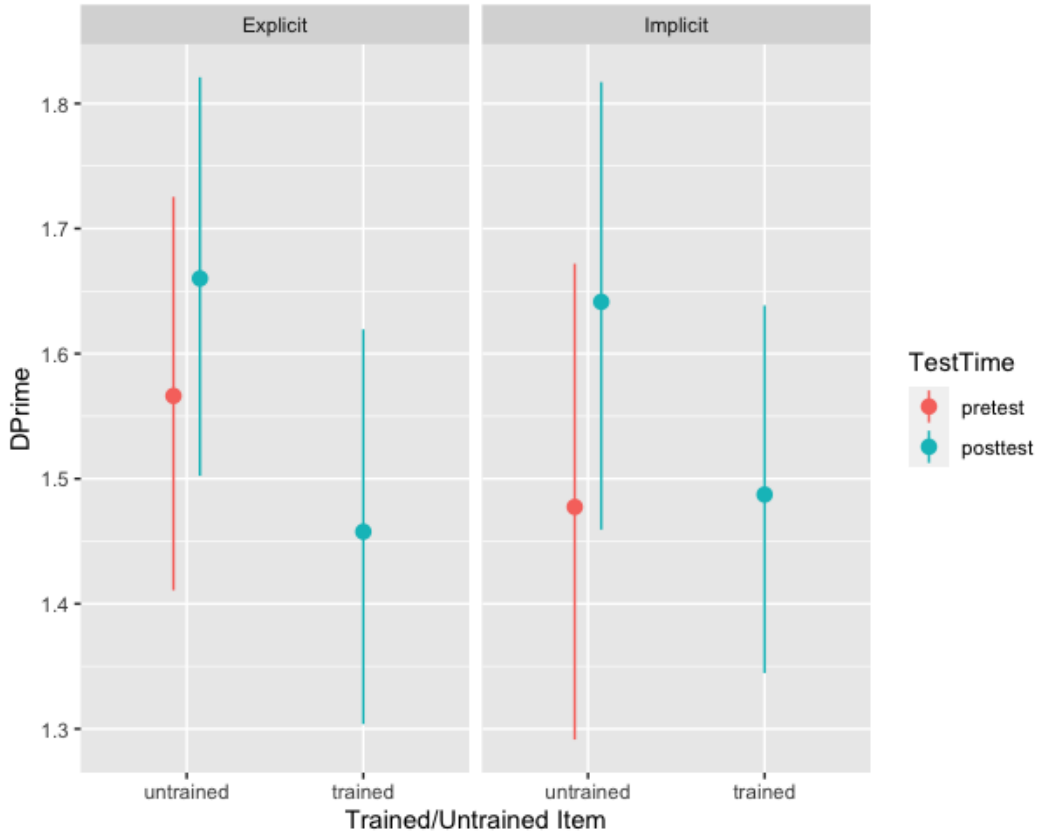

Figure 10 d' Scores for Trained versus Untrained Items

Groups were not significantly different at the pretest (ß=-0.041, *t*=-0.362, n.s.) or posttest (ß=-0.081, *t*=-0.592, n.s.) so they were not included in the final model. As shown in Figure 10, there was, however, a significant difference between test times, with improved d' scores at the posttest (ß=0.136, *t*=3.823, p<.001). There was also a significant difference between trained and new posttest items, with higher d' scores for the untrained items for both groups (ß=0.178, *t*=5.046, p<.001). Although this finding that new items were responded to more accurately than trained items at the posttest is surprising, it is consistent with the finding for speaker.

3.2.7 POA

Next, the role of each of the three POAs on participants' accuracy in discrimination was examined. The model used was an additional linear mixed effects model that examined d' scores as a function of POA and test time, and included a random intercept for subject. These variables can be visualized below in Figure 11.

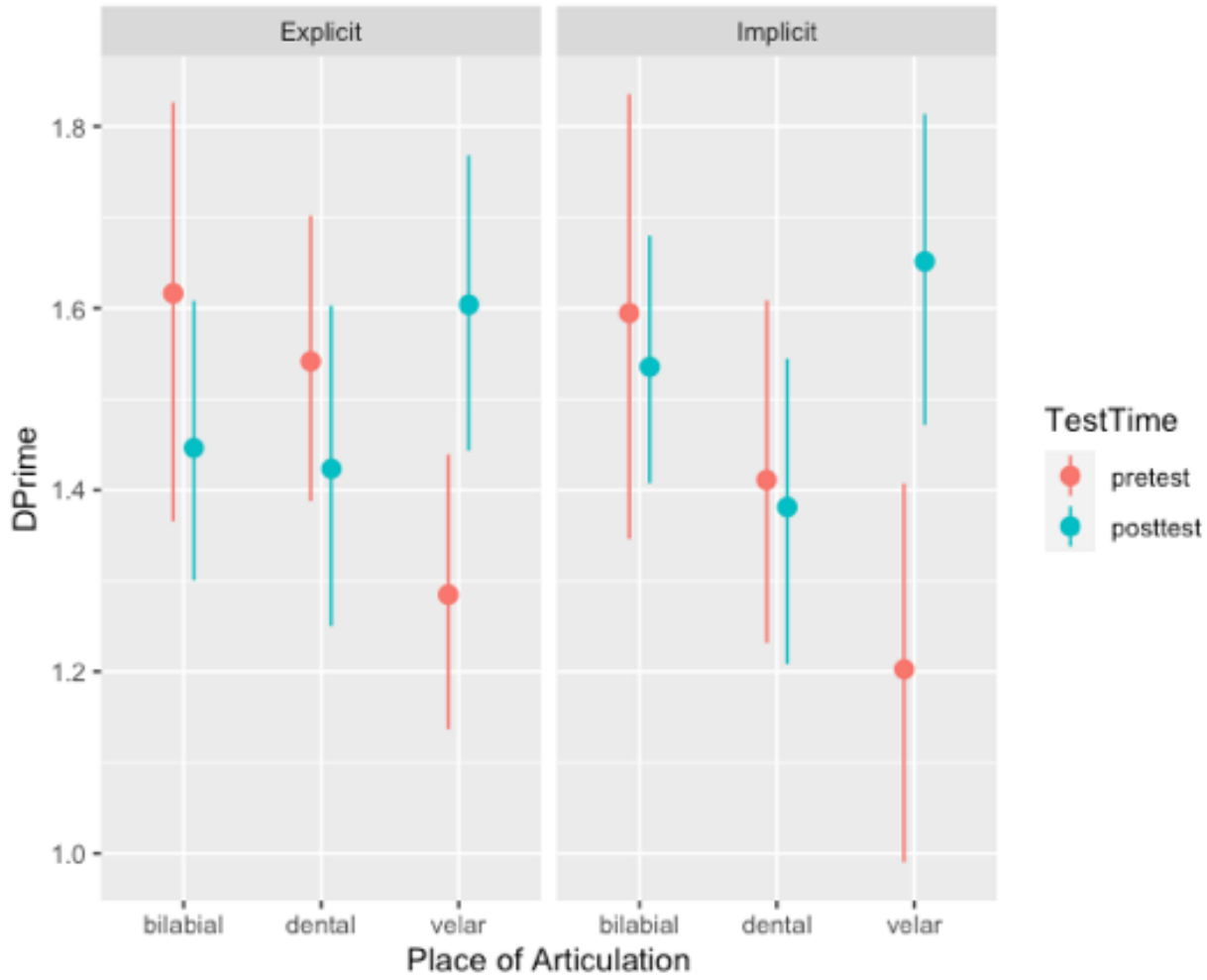

Figure 11 d' Scores for POA by Group

There were no significant differences between groups at pretest (*ß*=-0.078, *t*=-0.576, n.s.) or posttest (*ß*=0.032, *t*=0.272, n.s.), so groups have been combined in Figure 12 below and differences between the different POAs will be discussed in greater detail without including group as a variable. For this same reason, group was not included in the final model.

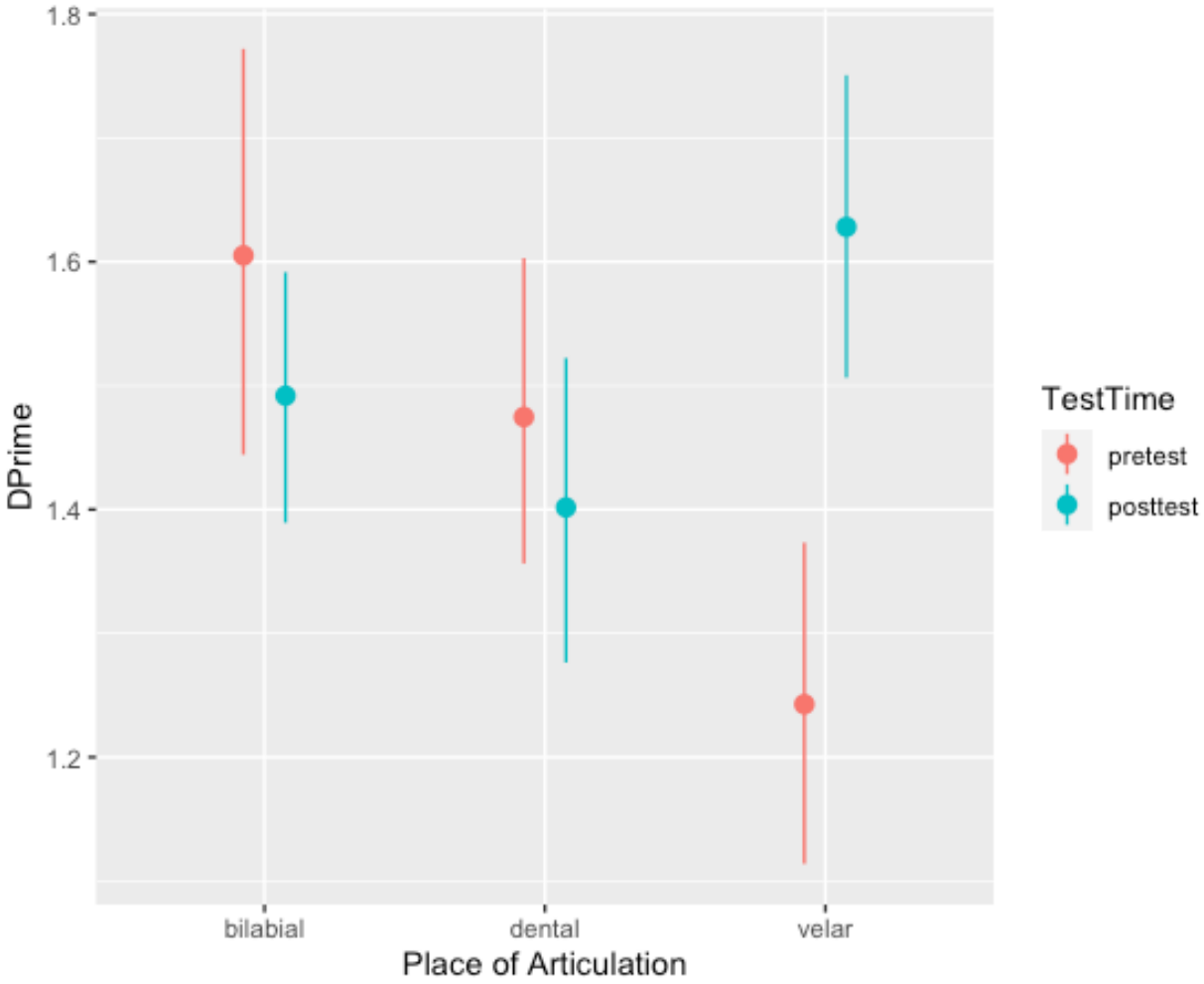

Figure 12 d' Scores for POA with Groups Combined

At the pretest, bilabial trials were responded to significantly more accurately than dental trials (*ß*=0.130, *t*=2.330, p<.05), and dental trials more accurately than velar trials (*ß*=0.232, *t*=4.144, p<.001). Bilabial trials were also responded to more accurately than velar trials (*ß*=0.362, *t*=6.474, p<.001). Overall, the two groups showed significant improvement across test times (*ß*=-0.068, *t*=-2.051, p<.05).

At the posttest, bilabial trials were still responded to more accurately than dental trials (*ß*=-0.090, *t*=-2.554, p<.05), but velar trials improved substantially (*ß*=-0.382, *t*=-8.995, p<.001),

yielding higher d' scores at the posttest than dental (ß=0.226, *t*=6.411, p<.001) and bilabial (ß=0.236, *t*=3.857, p<.001) trials. The other POAs did not improve from pretest to posttest as substantially as did the velar trials; bilabial trials improved marginally (ß=0.114, *t*=1.947, p=0.058) and dental trials did not improve significantly (ß=0.066, *t*=1.271, n.s.). With velars improving more than bilabials and dentals, this finding again demonstrates that untrained features improved more than trained features, a pattern that will be explored in detail in the discussion.

3.2.8 Speaker and Item

In addition to the models mentioned above which included a single predictor in addition to group and/or test time, the interaction between specific variables was also examined. To begin, item and speaker as well as the interaction between the two will be discussed. The lmer model employed examined d' scores as a function of speaker, item, an interaction between speaker and item, and test time, and included a random intercept for subject.

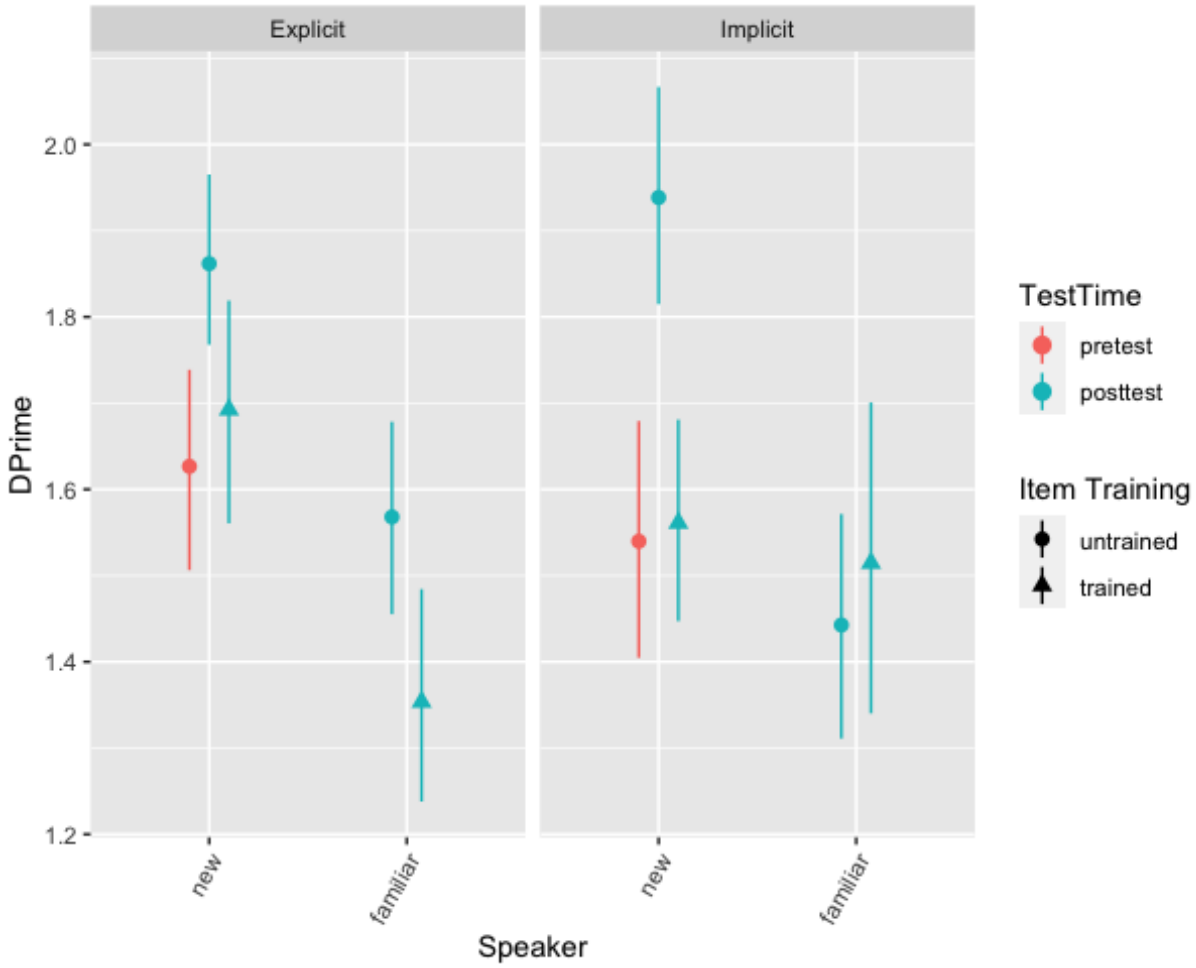

Figure 13 d' Scores for Item and Speaker Faceted by Group

As can be seen above in Figure 13, no significant differences existed between groups at the pretest (*ß*=-0.041, *t*=-0.342, n.s.) or posttest (*ß*=-0.009, *t*=-0.072, n.s.). Given the lack of differences between groups at either test time, group is omitted from the model, and the plot in Figure 14 below illustrates this omission to instead emphasize the changes that occurred from pretest to posttest.

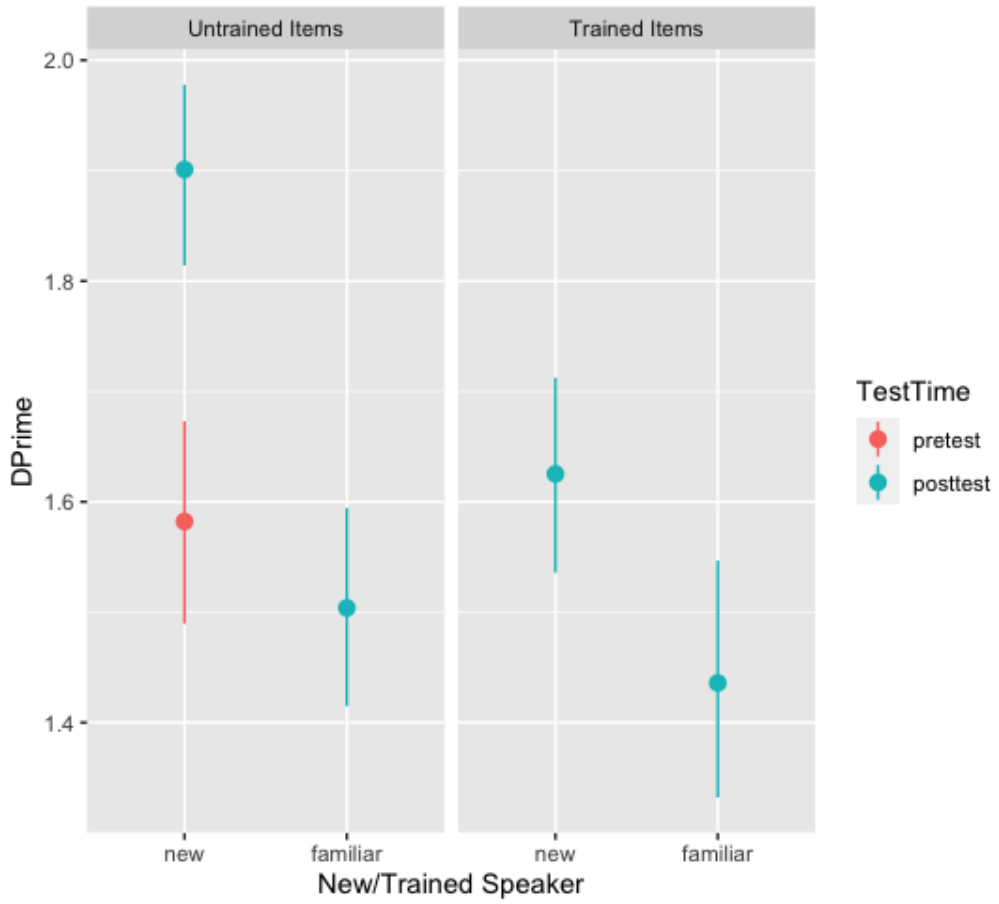

Figure 14 d' Scores for Item and Speaker with Groups Collapsed

D' scores for untrained items produced by a new speaker improved significantly from pretest to posttest (ß=0.284, *t*=7.198, p<.001). Significant differences between new and familiar speakers for trained items (ß=0.189, *t*=3.557, p<.001), as well as for untrained items (ß=0.239, *t*=6.576, p<.001) were also found. There was a significant interaction at the posttest between speaker and training, with higher d' scores for items that were both untrained and produced by the new speaker (ß=-0.178, *t*=-2.686, p<.01).

3.2.9 POA and Item

Next, POA and item and the interaction between the two were examined. The model that was used examined d' scores as a function of POA, item, an interaction between POA and item, group, and test time, and included a random intercept for subject. Overall, there were no significant differences between groups at the pretest (ß=-0.090, *t*=-0.695, n.s.) or posttest (ß=-0.022, *t*=-0.19, n.s.) with regards to POA and item, as shown in Figure 15. For the dental items, however, there was a significant interaction between groups and training. The implicit group responded significantly more accurately to untrained than trained dental items at the posttest (ß=-0.166, *t*=-3.535, p<001.). The explicit group, on the other hand, responded with similar accuracy to untrained items at both test times (ß=-0.167, *t*=-1.408, n.s.).

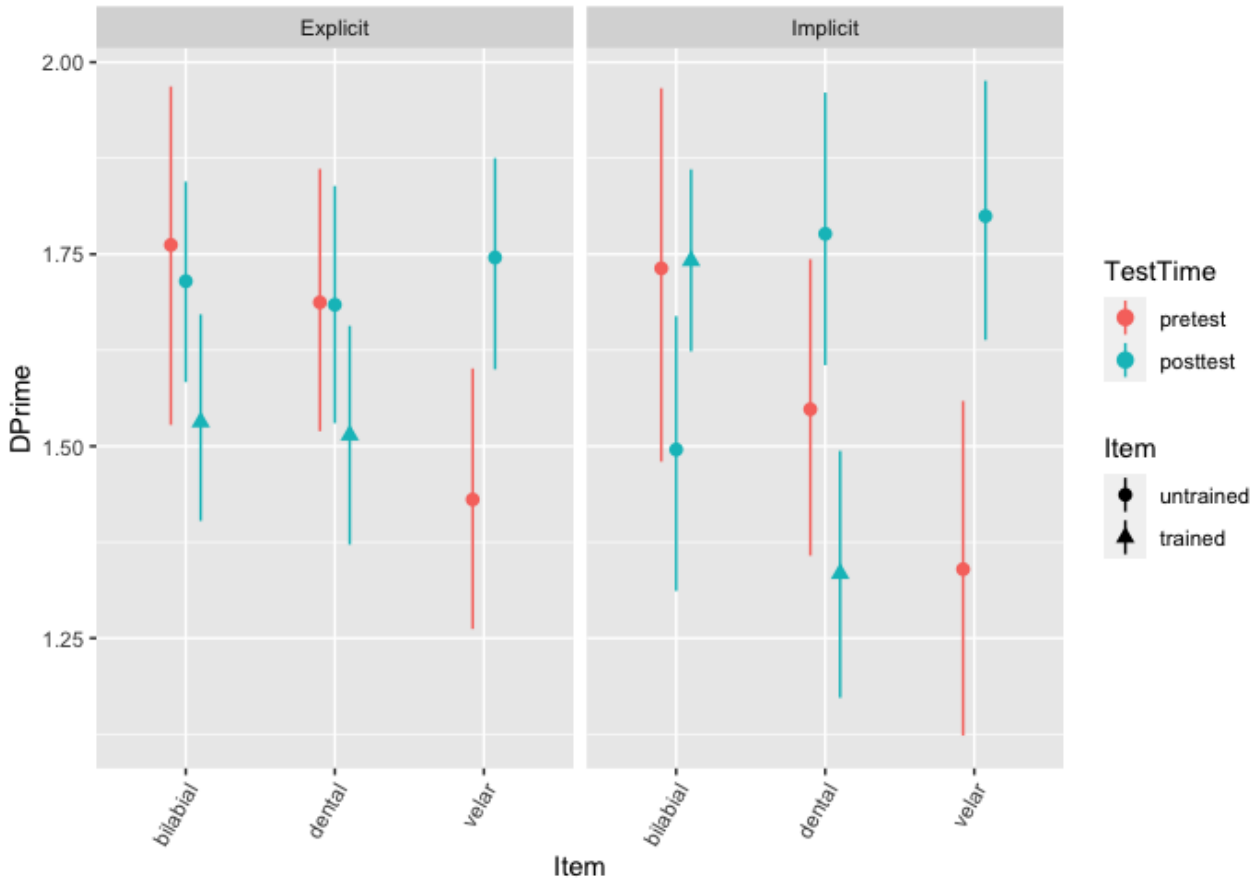

Figure 15 d' Scores for Item and POA Faceted by Group

Overall, d' scores calculated for POA and item worsened following training (ß=0.123, *t*=2.511, p<.05). Both the explicit (ß=0.092, *t*=1.584, n.s.) and implicit (ß=-0.022., *t*=-0.226, n.s.) groups experienced a decreased sensitivity to dental items they were trained on. The same was true for bilabial items, though only for the explicit group (ß=0.132, *t*=2.094, p<.05). There was no significant change for bilabial items across test times for the implicit group (ß=0.120, *t*=1.45, n.s.).

Interestingly, both groups made the most gains with velar items, the one POA on which they were not trained. These gains were significant for both the explicit (ß=-0.319., *t*=-6.273., p<.001) and the implicit (ß=-0.444., *t*=-6.696, p<.001) groups. This led to a significant interaction between POA and training, with higher d' scores at the posttest for velar items as compared to

other POAs (ß=-0.512., *t*=-6.868, p<.001). There was also a significant interaction between POA and training at posttest such that trained dental items were responded to with higher accuracy than trained bilabial items, with the reverse being true for untrained items (ß=0.367, *t*=5.469, p<.001).

A three-way interaction between groups at the posttest, POA, and training was found. The implicit group responded less accurately to untrained items as compared to the explicit group (ß=-0.326., *t*=-5.505, p<.001). At the same time, the implicit group responded significantly more accurately to trained bilabial items as compared to the explicit group (ß=0.436, *t*=7.370, p<.001).

3.2.10 Speaker & POA

Finally, the interaction between speaker and POA were examined using an lmer model that examined d' as a function of POA, speaker, an interaction between the two, group, test time, and included a random intercept for subject. The only significant difference between groups with regards to speaker and POA stemmed from an interaction between group and speaker when examining velar tokens only. The implicit group responded significantly more accurately to the new speaker for these velar items at the posttest than did the explicit group for the same type of items (ß=0.329, *t*=2.462, p<.05). Otherwise, there were overall no significant differences between groups at pretest (ß=-0.091, *t*=-0.685, n.s.) or posttest (ß=0.020, *t*=-0.179, n.s.) with regards to speaker and POA, as shown in Figure 16.

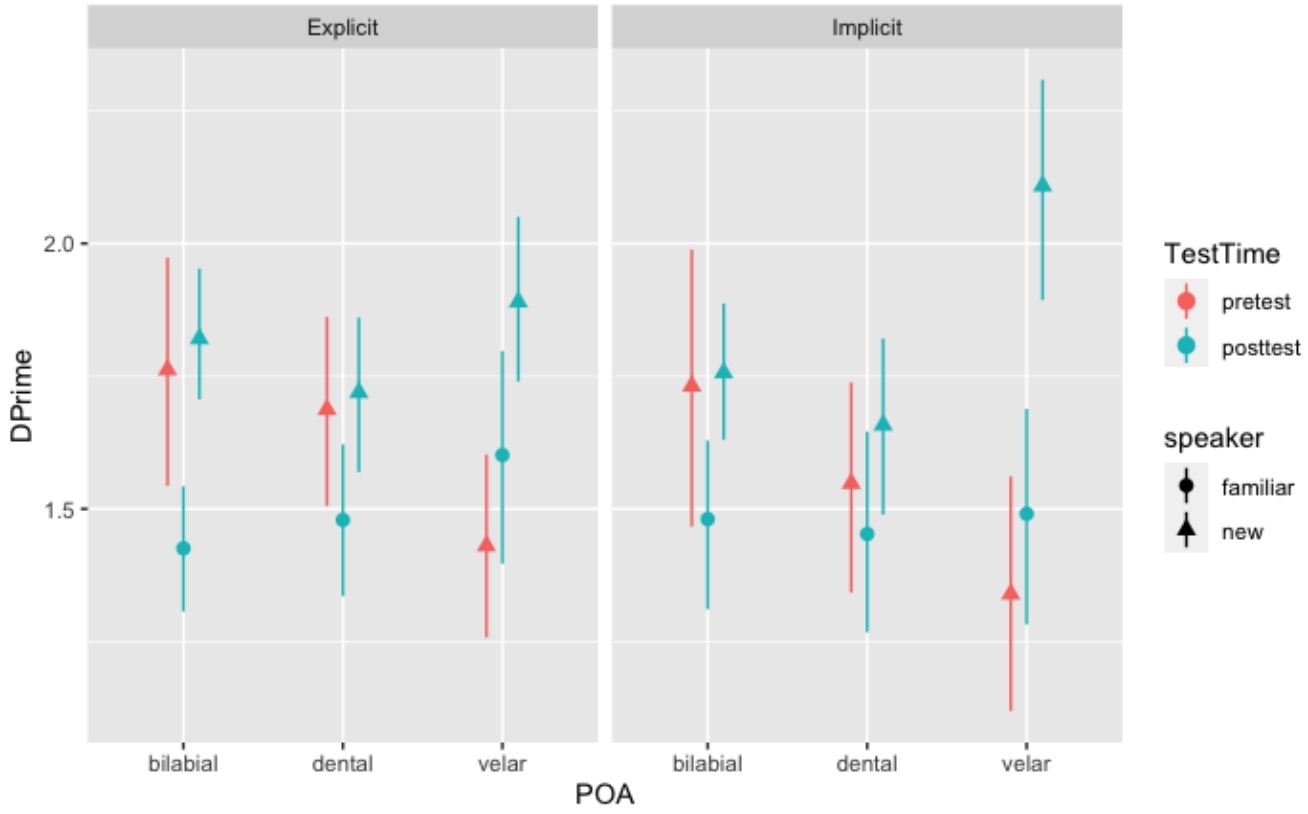

Figure 16 d' Scores for Speaker and POA Faceted by Group

Thus, groups have been collapsed in subsequent analyses of these variables, as in Figure 17. There was a significant difference at the pretest between bilabial and dental (*ß*=-0.130, *t*=-2.330, p<.05), between dental and velar (*ß*=0.232, *t*=4.414, p<.001), and between bilabial and velar (*ß*=-0.362, *t*=-6.474, p<.001) trials.

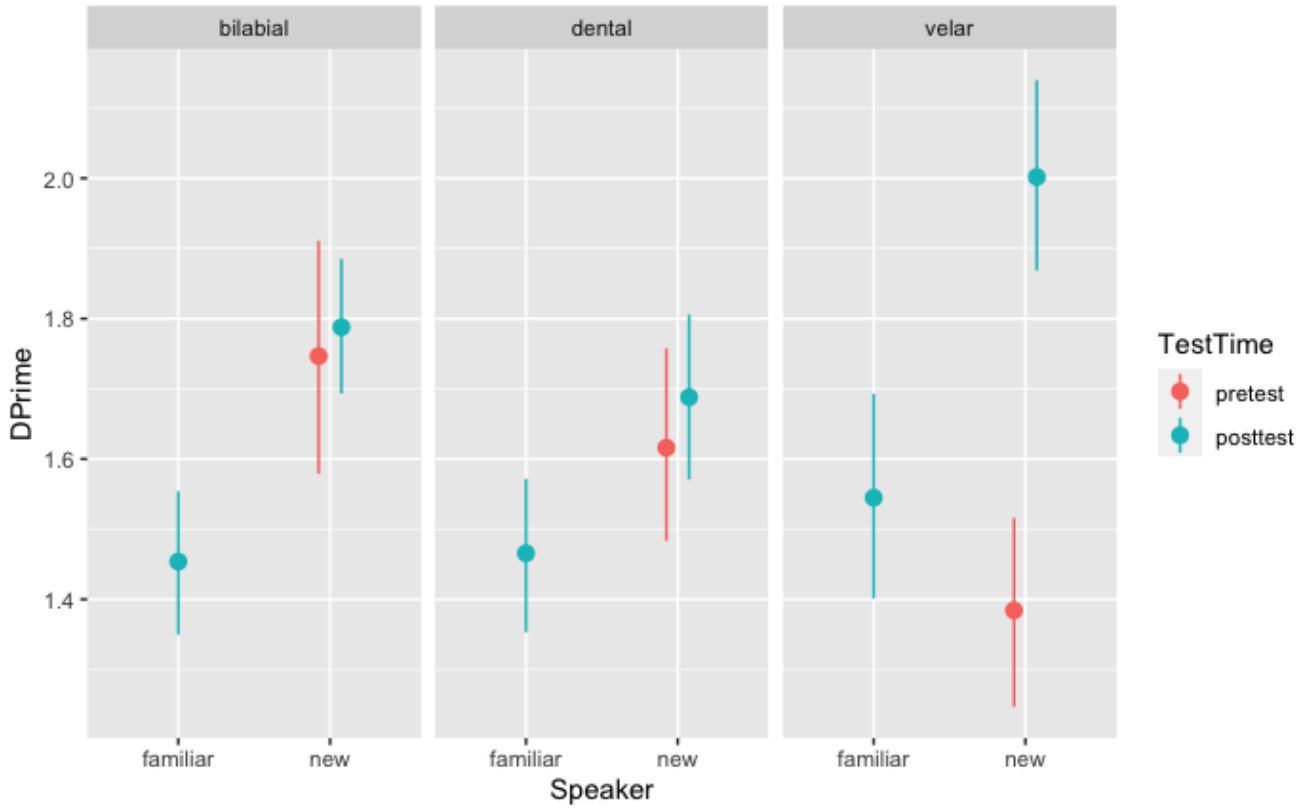

Figure 17 d' Scores for Speaker and POA with Groups Collapsed

At the posttest for trials spoken by the familiar speaker, there was a significant difference in accuracy between dental and velar trials, with higher d' scores for velar trials (*ß*=-0.133, *t*=-2.467, p<.05). For these same trials, there was no significant difference between bilabial and dental trials (*ß*=-0.081, *t*=-1.509, n.s.), nor between bilabial and velar trials (*ß*=0.052, *t*=0.958, n.s.).

There were also significant differences between different POAs at the posttest for new speaker trials. In particular, bilabial trials spoken by the new speaker were responded to more accurately than were dental trials (*ß*=-0.096, *t*=-2.095, p<.05). Similarly, velar trials led to higher d' scores than dental trials (*ß*=-0.328, *t*=-7.160, p<.001) and than bilabial trials (*ß*=0.232, *t*=5.064, p<.001).

In summary, the results of the discrimination task as described above seem to indicate that participants improved overall as a result of either type of training, providing evidence for the statistical distributional learning hypothesis, but only improved in their perception of prevoicing in the explicit condition, providing greater support for the minimal pair hypothesis. What's more, for the three types of analogy examined – item, speaker, and POA – participants at the posttest tended to perform better for analogous trials as compared to trials and conditions on which they were trained. In the discussion section, I consider some explanations for this finding, as well as possible implications.

4.0 Discussion and Conclusions

The experiments presented in this dissertation examined perceptual adjustments to a novel VOT category boundary following training using a word learning paradigm through which participants were exposed to either implicit (distributional) training or explicit (minimal pair) training. Accuracy data from the training phases also allowed the current study to examine lexical acquisition through the lens of two different lexicons – one with minimal pairs (explicit) and another without (implicit). From this data, it was found that when participants were asked to listen to a word and select the visual object that corresponded, when choosing between non-minimal pairs, participants approached ceiling levels of accuracy very quickly, typically within the first three training sessions. When participants in the explicit group were asked to choose between objects represented by minimal pairs, thus invoking phonetic knowledge in addition to lexical knowledge, their accuracy scores suffered, lagging behind the implicit group by about two training sessions.

The phonetic discrimination tasks in the pretest and posttest allowed the present investigation to delve into questions regarding phonetic learning, and, in particular, examine sensitivity to phonetic cues as a product of HVPT and different learning conditions. Although there were few differences in performance between the different training conditions, one important difference did emerge – the explicit group alone improved in their ability to use prevoicing to inform their perception. Additionally, both groups showed robust ability to generalize to novel items, a novel speaker, and a novel POA. In fact, these generalizations were so robust that both groups performed higher when presented with novel components as compared to the components on which they were trained. Explanations for this unexpected finding will be explored in depth

below. To do so, lexical learning will be examined first, before then turning to phonetic learning and the relationship between the two.

4.1 Lexical Learning

First, to examine lexical learning in the current study, accuracy scores were calculated for Training Phase B and Training Phase C on all five days of training. During Training B, both groups heard the name of an object, were presented with two visual objects on the screen, and were asked to decide which of the two objects corresponded to the auditory stimulus. During Training B, neither group saw visual representations of minimal pair items presented side by side, in essence allowing participants to focus on more global properties of the lexical items rather than on fine-grained phonetic cues. The results presented in the results section reflect the critical items only, with distractor items removed. It is, however, still important to understand any differences in accuracy between critical and distractor items. No significant difference was found between the distractors and the critical items during Training B 2AFC task for the implicit group, suggesting that the critical items themselves were not more difficult overall than the distractors. Nevertheless, at Session 1, the explicit group did demonstrate greater difficulty with critical items as compared to distractor items. Since the implicit group was not impacted in the same way, this difference presumably stems from the existence of minimal pairs in the explicit group's lexicon, even when the corresponding images are not presented side by side. Previous studies on the effects of neighborhood density on word learning have instead found higher neighborhood density to have a facilitatory effect on word learning in the long term (Storkel, Armbruster, & Hogan, 2006; Storkel, Bontempo, & Pak, 2014). This difference in the findings of Storkel and colleagues and those of

the present investigation likely stems from the current study's manipulations of VOT such that, without training, minimal pairs were likely to be perceived as homophonous. Overall, however, the lack of difference between critical items and distractors became especially apparent for both groups in the later days of training as participants approached ceiling rates of accuracy for both item sets, as can be seen in Figure 18 below.

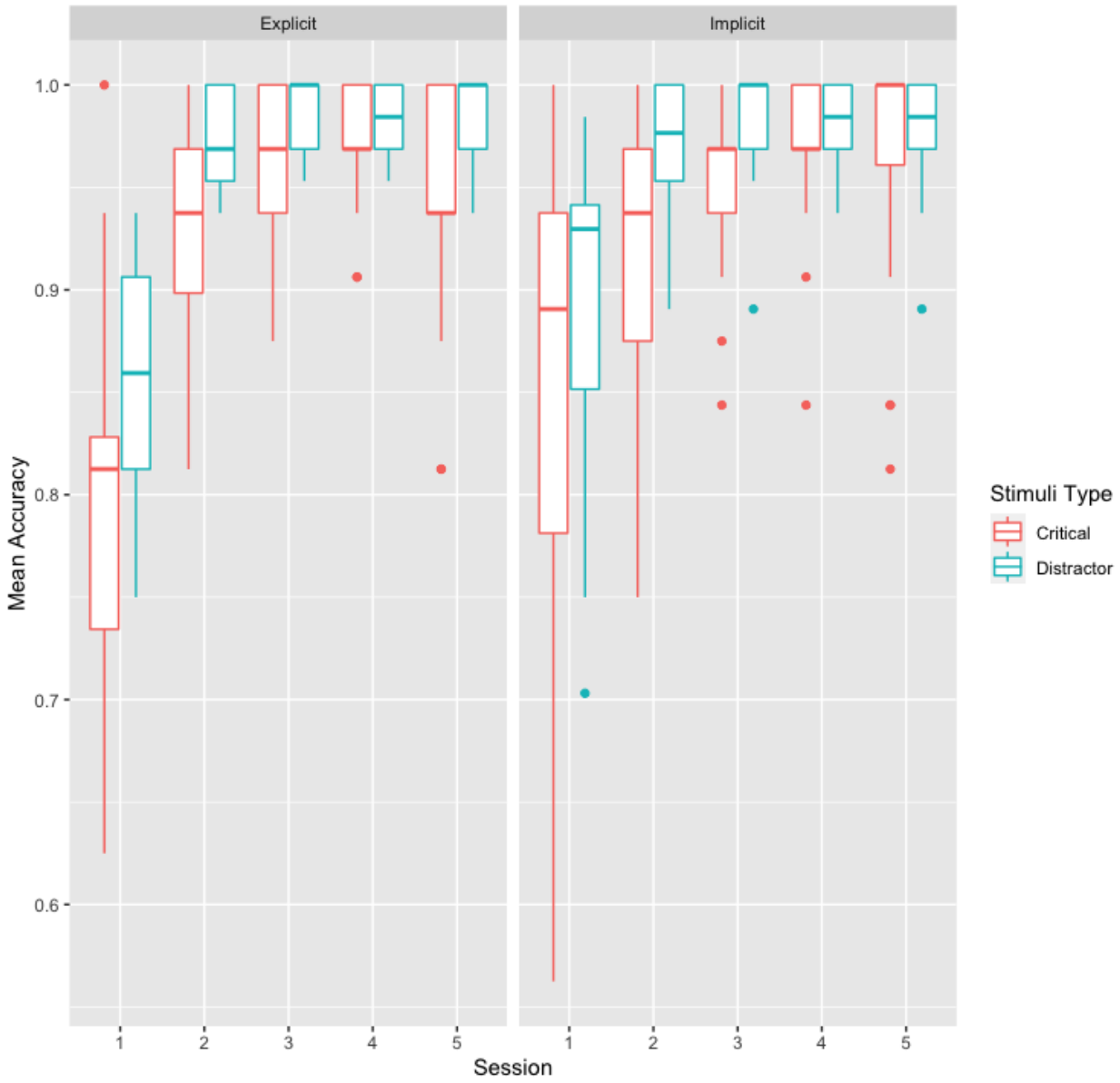

Figure 18 Training B 2AFC Task Critical vs Distractor Items

The results from the Training B 2AFC task illustrated that at Session 1, the explicit group had a lower overall accuracy rate with the critical items presumably as a result of the existence of minimal pairs in the lexicon. Given this, it came as little surprise that the explicit group also had lower accuracy than the implicit group at Training C, which homed in on phonetic learning for the explicit group through the side-by-side presentation of pictures whose labels were minimal pairs. This difference between groups was to be expected, since identifying the corresponding minimal pair was a much more difficult task than that of the implicit group, whose task was effectively a repetition of Training B. This difficulty is also reflected in the explicit group's lower scores for critical items as compared to distractor items, a difference that was far less pronounced in the implicit group. Given the nature of the task, this difficulty for the explicit group in the Training C 2AFC task can be attributed to difficulty with perceiving a novel, lower VOT category boundary. There is little reason to assume that the implicit group was better equipped to perceive this boundary; rather, it follows that this information was not evoked by the task of selecting between non-minimal pairs objects.

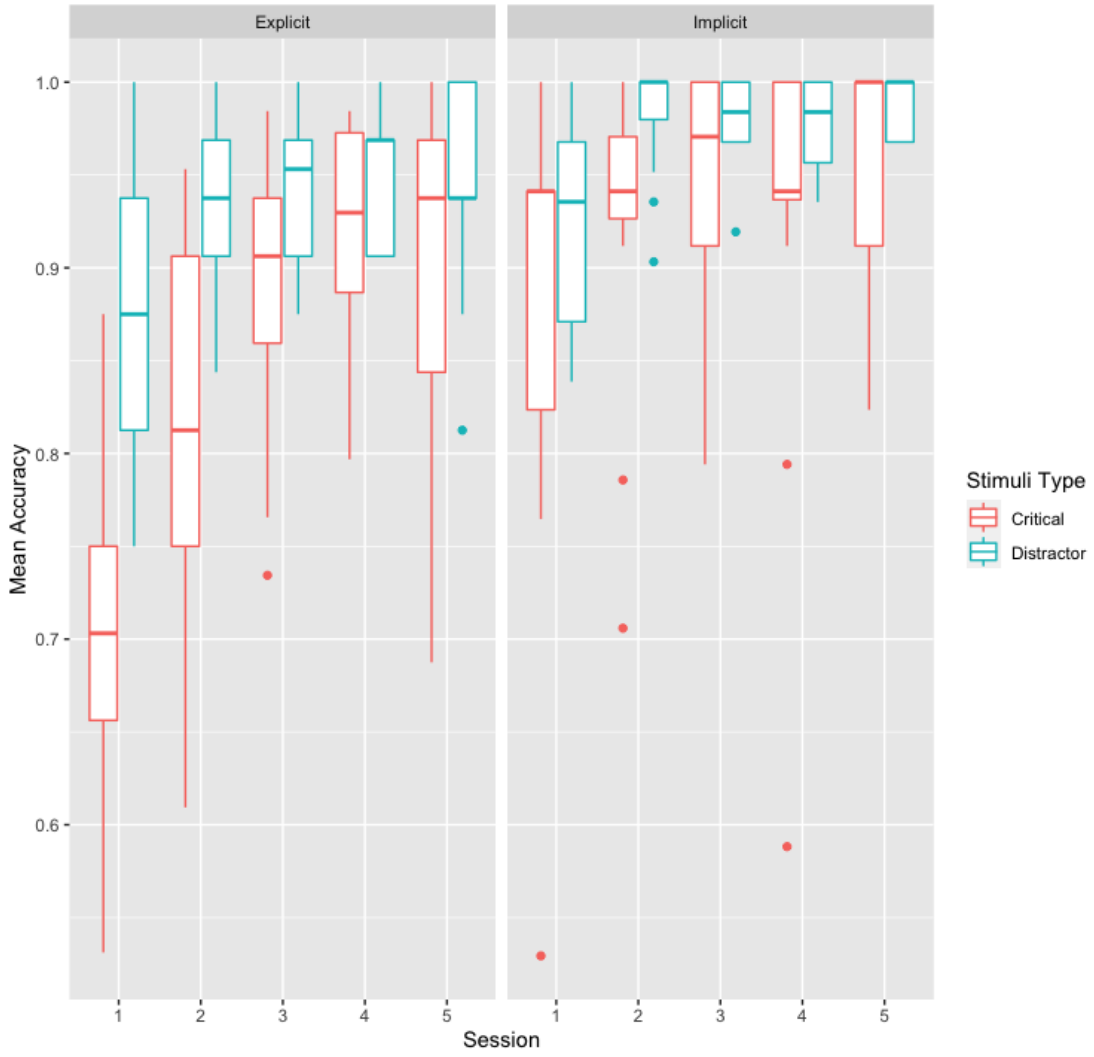

Figure 19 Training C 2AFC Task Critical vs Distractor Items

As in Training B, however, the differences between groups at the Training C 2AFC task were most pronounced at the earlier training sessions, particularly 1-3, and were mitigated in later sessions, as shown in Figure 19 above. Interestingly, ceiling effects are reached around Session 4, but then accuracy scores dipped at Session 5, suggesting that participants may have been overtrained. This result was surprising, given the extensive literature on HVPT, which frequently employs a far greater number of training sessions and for which overtraining has not been reported (Logan et al., 1991; Lively et al., 1993, 1994; Bradlow et al., 1997, 1999; Thomson & Derwing,

2016, etc.). This difference between the current study and those that have come before may to some degree be attributed to the length of the posttest. Previous studies have reported posttest durations of thirty minutes or less. Therefore, it remains possible that the attention of the participants of the current study had waned before the completion of the task. Even so, however, it remains the case that participants also performed significantly worse on the fifth day of training as compared to the day before, so inattention due to a long posttest is only one factor among many to explain overtraining.

Previous research has examined overtraining through the lens of the color-word contingency learning paradigm (Schmidt, Crump, Cheesman, & Besner, 2007), in which participants are presented with words written in different colors. Words that are most frequently presented in a consistent color show an effect of decreased reaction times and increased contingency effects, both suggesting that overtraining resulted in a more established relationship between a word and its associated color (Schmidt, De Houwer, & Moors, 2020). These results are generally positive and facilitative, but the opposite was found when the color contingency was removed during a counterconditioning phase in which words were reassigned to new colors. To my knowledge, however, research on overtraining has not demonstrated any such negative effects as were found in the current study. Since no such counterconditioning took place in the present study, another explanation is needed. In a later section of this discussion, I present evidence for the roles of attention and saliency as leading factors for this drop off in accuracy rates in the later training sessions.

Due to the differing natures of the lexicons in the two groups in term of the presence or absence of minimal pairs, the task required by the explicit group was inherently more difficult, as reflected by overall lower accuracy scores. Given this, it is challenging to assess the relative

effectiveness of either training paradigm using accuracy scores alone. The explicit group had a more difficult task to complete and therefore had lower initial scores, allowing for more room to improve overall, whereas the implicit group reached ceiling levels of accuracy much more quickly. Given these differences, relative improvement scores could be a more accurate indicator of which training paradigm was most effective for lexical learning. Recall that relative improvement scores represent the amount of improvement made as a proportion of the amount of improvement possible, and were calculated using the formula below:

Difference Score = Session 5 Score – Session 1 Score

Room to Improve Score = 1.0 - Session 1 Score

Relative Improvement Score = Difference Score / Room to Improve Score

Higher relative improvement scores were found for the explicit group for both phases Training B and Training C 2AFC tasks, providing some evidence for the effectiveness of the paradigm. It is important to recall, however, that as the implicit group approached ceiling levels of accuracy, it became more and more difficult for this group to improve their score by the same amount as the explicit group. Future studies might address this limitation by including a third training condition that includes minimal pairs in the lexicon but does not present them side by side as is done with the explicit group in Training C. Such an inclusion would allow researchers to disentangle the difference in lexicons from the difference in object presentations in the training phases.

Although the explicit group generally performed poorer as measured by the accuracy scores for Training C, this was expected, given the more difficult nature of the task assigned. In

the next section, I address the question of whether this increased difficulty associated with the Training C 2AFC explicit task was transferred into increased gains in phonetic learning.

4.1.1 The Role of Feedback

Participants in the current study received feedback for each trial of the 2AFC tasks. Recall that for correct responses, participants saw a green box appear framing the object they correctly selected, and then heard a repetition of the auditory stimulus. For incorrect responses, participants heard an auditory repetition of the target lexical item, saw the correct object framed in green, followed by a red box framing their incorrect selection and an auditory rendition of it, and then finally a heard the auditory rendition of the object they should have selected, with an accompanying green frame around the object.

A primary consideration with regard to feedback is its effect on explicitness. Although the implicit group was considered to be receiving implicit input due to the lack of minimal pairs that could provide any explicit evidence of the location of the VOT category boundary, this group did still receive targeted feedback. This feedback was inherently somewhat explicit in that participants' attention was being drawn to the stimuli, making the implicit condition somewhat less than purely implicit. However, it was necessary to provide feedback to both groups; otherwise, it would have been impossible to discern whether any differences between groups were a product of the lexical items themselves or the presence or absence of feedback.

An additional consideration with regard to feedback is that of frequency. Because the explicit group's 2AFC tasks were more difficult due to the presence of minimal pairs, the explicit group selected the incorrect object more often than the implicit group, and therefore experienced the incorrect feedback sequence more frequently. Although the feedback sequence was identical

for the two training conditions, this resulted in the explicit group having overall more exposure to the stimuli, since the incorrect sequence provided two tokens of the correct word and one token of the incorrect word whereas the correct sequence provided only one token of the correct word.

Finally, in addition to raising the question of whether the explicit group's improvement was due to a difference in the frequency of exposure between the two groups, it is also important to consider the role of error-driven learning. Recall that error-driven learning posits that learning is more likely to occur as a function of surprise or prediction error than simply due to the distribution of the input (Olejarczuk et al., 2018). The incorrect feedback sequence presented in the current study was likely an unexpected event that could serve to update the category representations being formed. Given this, additional exposures to the incorrect feedback sequence may have ultimately led the explicit group to greater gains from pretest to posttest with regard to lexical learning. It is also possible that error-driven learning may have accounted for the explicit group's improvement in discriminating prevoicing. This and other phonetic variables will be explored in the upcoming section.

4.2 Phonetic Learning

To examine phonetic learning, the current study examined the results from the discrimination task given as the pretest and posttest on sessions 1 and 5 respectively. Participants heard two aural stimuli and were asked to decide whether they were the same word or two different words. Their responses informed the degree to which their VOT boundaries for Numana shifted as a result of the five days of training completed between pretest and posttest.

4.2.1 The Role of Attention in Distributional vs Minimal Pair Learning

The first research question asked about the role of minimal pairs in learning to perceive the lower Numana VOT category boundary. Both the explicit minimal pair group and the implicit non-minimal pair group were exposed to the same number of items and segments and were simply given different tasks regarding lexical identification. Overall, there were no significant differences between implicit and explicit groups at the pretest. Both groups improved substantially as a result of the training tasks, but contrary to predictions, groups were not significantly different from one another at the posttest according to most measures.

One important difference between groups, however, was found in the d' scores for the 'Within Voiced' category, as is shown in Figure 20. The implicit group did not improve significantly at identifying a difference when one existed within the prevoiced portion of the continuum across test times ($ß$=0.139, t=1.050, n.s.), whereas the explicit group did ($ß$=0.563, t=8.658, p<.001). In essence, participants in the explicit group improved their ability to perceive prevoicing as a useful cue for distinguishing prevoiced tokens. Although the implicit condition was meant to attend to the prevoicing throughout the course of their training, this wasn't strictly necessary in order to distinguish between items given that there were no minimal pairs. Therefore, the group remained unaware of prevoicing's utility as a cue for voicing, even after five training sessions.

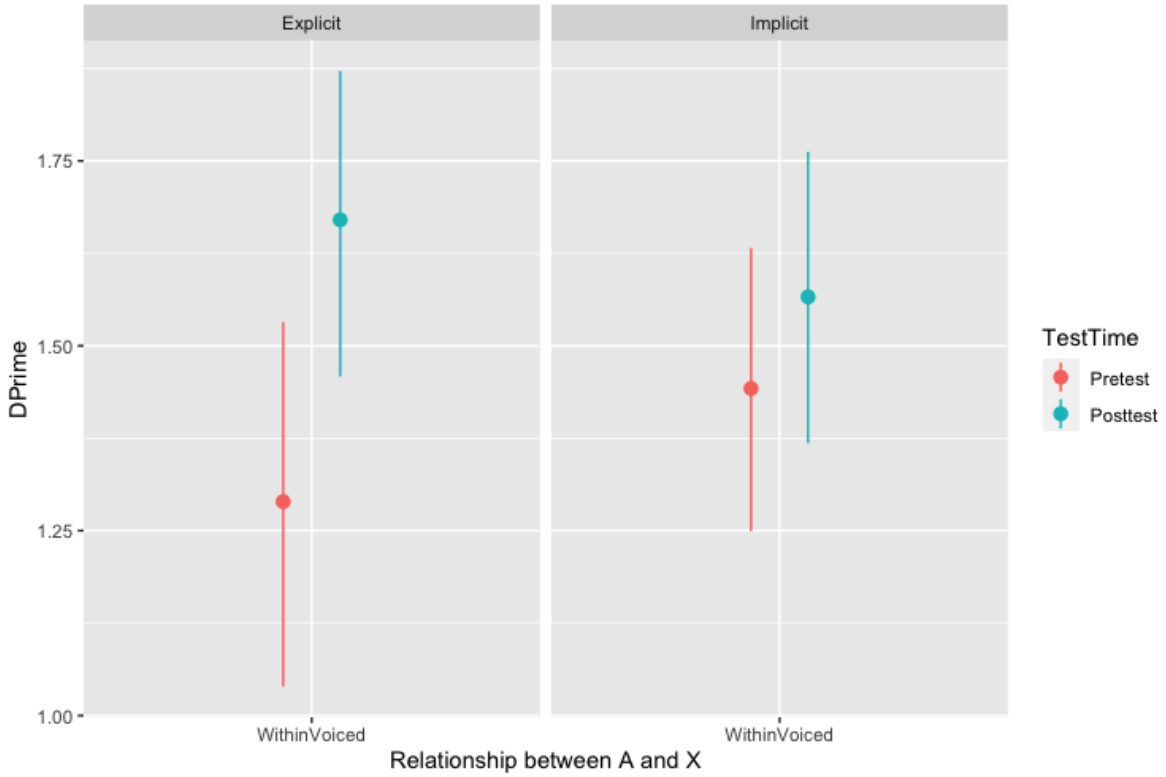

Figure 20 d' Scores for 'Within Voiced' Items

Because this difference in training resulted in the improved perception of prevoicing for the minimal pair group only, this study provides support for the noticing hypothesis (Schmidt, 1990, 1994, 1995, 2001, 2012) and the minimal pair hypothesis (Maye & Gerken, 2000). Schmidt's (1990, 1994, 1995, 2001, 2012) noticing hypothesis states that noticing facilitates input being converted into intake. Learners in this study who were not made explicitly aware of prevoicing as a meaningful voicing cue did not notice it and therefore did not intake this information, despite being flooded with input containing it. Attention is thus a crucial component of both phonetic, and consequently, lexical learning. Although the greater difficulty of the explicit task did appear to draw participants' attention towards particular acoustic cues (namely

prevoicing), it did not appear to hold participants' attention insofar as maintaining interest in the task. This can be seen in the dips in accuracy scores by both groups at Session 5.

This study's findings contradict the predictions of the lexical-distributional hypothesis, which states that minimal pairs should impede the learning of new phonetic categories (Feldman et al., 2013a, 2013b). Feldman et al. (2013b) exposed participants to one of two training conditions. In the minimal pair condition, participants listened to input containing pseudoword minimal pairs in the training phase. The non-minimal pair group listened to input consisting of only non-minimal pair pseudowords. Participants were only told that they would be listening to words and would later be asked questions about the sounds in the language. Then, in the testing phase, participants completed a discrimination task that evaluated their ability to perceive fine grained phonetic differences between minimal pair stimuli. They found that learners were more able to perceive phonetic differences when contrasting vowels were consistently presenting in differential lexical contexts. Contrary to Feldman et al.'s findings and the lexical-distributional hypothesis, the participants of the current study were more successful at homing their attention into the relevant acoustic cue that differentiated the members of each minimal pair. Without such attention being focused, the implicit group of the present investigation did not learn to attend to prevoicing over the course of the five training sessions. The findings of the current study differ from those of Feldman et al., and reasons for the differences in our findings will be discussed in greater depth below.

That being said, the implicit group still improved on the perception of other variables such as discriminating "Across" and "Within Voiceless" trails, so there is still some argument to be made for the effectiveness of distributional learning. Importantly, individual phonetic variables and the ways in which they interact with the L1 appear to be more or less susceptible to particular

types of training. This is in line with the contradictory findings of Curtin et al. (1998), who on the one hand found that participants experienced greater improvements in perceiving prevoicing, whereas on the other Pater (2003) and Abramson and Lisker (1970) found that participants instead improved on aspiration when the type of training was tweaked. As a result, the most effective type of training may be contingent on the particular phonetic variable in question.

Although the implicit group did not match the explicit group in their improvement in the perception of prevoicing, the improvement of perceiving other independent variables, such as discrimination among within-category short-lag and within-category aspirated stops as well as between-category tokens did provide some evidence for phonetic learning, nonetheless. In particular, the implicit groups' improvement of such measures seems to have occurred as a byproduct of word learning, since participants were not required to focus on the phonetics in order to learn the labels of the objects in their purely distributional condition. In these cases, explicit attention to the onset and to the VOT, specifically, were not necessary components of either word learning or of perceiving fine grained phonetic differences in VOT between non-prevoiced tokens. This provides some weak support for the distributional learning of phonetic variables such as VOT and for the implicit learning of phonetics in L2 learners more generally. Despite the implicit learning of phonetics having taken place to some degree in this study, only particular acoustic cues seem to have been trainable in this way; prevoicing, at least within five training sessions, proved resistant to implicit training.

Interestingly (and anecdotally), participants of the implicit group frequently expressed that their training tasks were a fun challenge and enjoyed reporting their accuracy rates to me after each training session. The explicit group, on the other hand, regularly expressed their frustrations with the task. I suspected that this difficulty was an important component of what would eventually lead

this group to greater gains. This type of desirable difficulty is thought to cause a variable to be more difficult to learn initially, but leads to greater long-term retention (Bjork, 1994). In this case, desirable difficulty does seem to have amounted to greater improvements in the explicit group as compared to the implicit group, at least regarding the perception of prevoicing.

4.2.2 Models of L2 Speech Perception

The second research question asked to what degree the PAM and the SLM could account for any difference in trainability between bilabial and dental phones, and whether this difference would be impacted by the type of training received by the two groups. Recall that both the bilabial and dental stop consonants taught in Numana differed from English by virtue of having a lower VOT category boundary. Additionally, the Numana /t/ - /d/ contrast differed from that of English in that Numana was dental and English is alveolar. The PAM predicted single category assimilation for each contrast, since both phones were predicted to be perceived as the English voiced component of either pair. The SLM made a similar prediction with regard to the perception of voicing, but additionally predicted that the differences in POAs could lead to differential difficulty for dental and bilabial items, depending on whether equivalence classification would lead to dental stops assimilating to the English alveolar representations (Flege, 1987, 1995). Although it was not clear whether equivalence classification would take place, equal improvements with dentals as for bilabials would provide evidence for it in fact doing so. On the other hand, if participants experienced more improvement for bilabial items than dental items, this would suggest that equivalence classification had not taken place for the dentals and that participants were aware of the difference in POA and used this awareness to inform a separate Numana dental POA.

No significant differences between groups were found at either test time with regard to POA, so it can be concluded that the type of training provided did not have an impact on participants' bias towards classifying phones as being similar to or the same as their existing L1 representations. Both groups responded more accurately to bilabial trials than dental trials at both test times. However, the accuracy of responses to trials of both POAs improved by about the same amount from pretest to posttest, which suggests that similar methods of forming and maintaining mental representations were used for both phones. Since there is no reason to believe that a new representation was created for bilabials, it can consequently be concluded that no new representation was created for dentals. Instead, this finding is in agreement with equivalence classification and suggests that Numana /t/ was likely classified according to the existing English alveolar representation. It also suggests that the established representation was a strong enough fit that the training provided was not enough to impose a new, separate Numana dental category. If this were not the case and the phone were perceptibly distinct from English, one might expect greater differentiation between bilabials and dentals at the posttest, with more accurate responses for bilabial items overall.

In any case, the discrimination of the contrasts of all three POAs improved for the participants of both training conditions, as was predicted by both models, and affirms that the phonological system remains malleable across the lifespan. Although native-like perception (of course suspending the disbelief that such a thing could exist for Numana) would not necessarily be expected, perceptual shifts towards the lower VOT category boundary target are seen.

4.3 Generalization and Analogy

In addition to the prediction that both groups would improve their perception of the trained bilabials and dentals, it was also predicted that groups would experience improvements – although to a lesser degree than the trained items – on an untrained place of articulation. Participants' ability to do so would indicate not only their ability to learn to perceive a new VOT category boundary but would crucially indicate their ability to extract and generalize phonological feature representations such as voicing.

Perhaps the most surprising finding of this study was that participants generalized to velars to a degree that far surpassed the POAs on which they were trained. This finding is largely in contrast with Maye and Gerken (2001), who, recall, trained participants by having them listen to either a unimodal or bimodal distribution of one of two possible POAs - dental or velar. After the training phase, participants were then tested their ability to generalize to the untrained contrast using a discrimination task. The group trained on dentals was expected to have more difficulty generalizing to the more marked velars, but ultimately Maye and Gerken were unable to induce generalization in either group, regardless of markedness or whether participants were trained on a unimodal or bimodal distribution. That participants of the current study were able to generalize to such a significant degree speaks to the efficacy of the training paradigm used, irrespective of condition. The findings of the current study are much more in line with those of McClaskey et al. (1983), who found that training participants on bilabial items resulted in generalization to an untrained place of articulation. Participants in McClaskey et al.'s study were prompted that they should be listening for something new in the posttest. In the case of the current study, such generalization to untrained POAs can take place even without explicitly prompting participants and directing their attention towards the analogous stimuli.

One explanation as for why velar items were responded to with such high accuracy at the posttest could be that velars have the least ambiguous cues to POA, with formant transitions, a compact concentration of energy, and strong stop bursts. While these acoustic cues may have provided some assistance for participants, their scores at the pretest do not support this explanation. At the pretest, velar items were by far the most challenging POA and were responded to by both groups significantly less accurately than both dentals and bilabials.

Given this, a more plausible explanation for the stark contrast between trained and untrained POAs at the posttest is a novelty effect that comes from overtraining. This is consistent with the finding that accuracy scores fell at Session 5 of the identification task comprising Phase B and Phase C of training, as well as with other variables of generalization measured by the discrimination task. In addition to the untrained POA, the untrained speaker and untrained items were both responded to more accurately at the posttest than their trained counterparts. What's more, the explicit group actually was slightly less able to discriminate between trained items at the posttest than untrained items at the pretest. By all other accounts though, participants improved in their perception and discrimination of the items, which again supports the explanation that participants' attention had waned by the end of the fifth training session. Hearing a new item, a new speaker, or a new POA was enough to raise their attention, even momentarily, and allow them to implement the phonological feature rule they had learned in training and apply it to the novel items, consistent with the novel popout phenomenon described by Johnston and (Hawley et al., 1994; Johnston et al., 1990, 1993).

That participants generally – and counterintuitively – more accurately perceived voicing differences for untrained variables suggests that the task at hand was not challenging enough to warrant all five training sessions, or perhaps that participants needed an additional cognitive

challenge such as solving math problems at the same time as their training tasks (e.g., Engonopulos, Sayeed, & Demberg, 2013). The rapidness of these results should not be entirely surprising given the reported success of HVPT at inducing perceptual changes remarkably quickly. Recall that Alves and Luchini (2017) trained participants in only three thirty-minute sessions and Schertz et al. (2015) induced perceptual changes in under 100 exposures to the stimuli. Despite this, the current study is the first HVPT study to my knowledge to report an effect of overtraining resulting in a decreased accuracy for the trained stimuli. An important difference between this study and others in the literature is the current study's use of a discrimination task. Previous studies have typically used identification or categorization tasks, and these have typically required less time than the discrimination task used in the current study. There were two main advantages to using a discrimination task: first, to avoid orthography (otherwise it was expected that participants would have been biased towards selecting the voiced member of each contrast due to influence from English), and second, to provide a more sensitive test of phonetic cue learning. One result of using discrimination rather than categorization, however, was that the task simply took a long time to complete (around 45 minutes). Although the task was divided into three blocks in an effort to ease the burden of testing, the length of the task did lead to an overall decrease in d' scores from Block 1 to Block 2 ($ß$=-0.136, t=-2.948, $p<.01$), but no additional significant decrease from Block 2 to Block 3 ($ß$=0.040, t=0.875, n.s.). This effect was equal for both training conditions, with no significant difference across groups ($ß$=-0.090, t=0.637, n.s.).

That discrimination decreased over the course of the testing session suggests that participants may have become bored or disengaged, an effect which may also have been exacerbated by overtraining on the contrast. Discrimination (d') scores for dental and bilabial items may have been higher had they been recorded during Session 4, at the peak of participants' lexical

accuracy (as measured by Training B and Training C 2AFC tasks) and before attention appears to have fallen. Evidence for this comes from Coppens and Patterson (2017), who state that "generalization cannot be expected if the treatment itself is unsuccessful. Furthermore, if application of a treatment protocol yields significant improvement, the success of generalization is always of a lesser extent (quantitatively)". Additionally, Dickey and Yoo (2010) state that the rate of generalization is necessarily slower than the rate of acquisition of trained items. If these statements are taken to be true, accuracy to bilabial and dental trials must have been artificially lowered as a result of an attentional mechanism, or otherwise velar trials must have improved as a function of some process other than generalization.

Nevertheless, that the participants were overtrained still speaks to the ability of the tasks used to effectively train the lexical items and phonetic variable of interest. Given this, it is apparent that the participants of this study were able to generalize beyond the variability in speakers, items, and POAs to apply a phonological feature of voicing in Numana.

It is worth noting that there were no significant differences between training conditions with regards to participants' ability to generalize to a new POA. This provides some additional support for the validity of distributional learning; the minimal pair group did not have an apparent advantage at generalizing to an untrained POA; participants of the implicit group were equally able to synthesize a phonological rule for voicing given the implicit presentation of the stimuli. It is important to keep in mind, however, that as L1 English speakers, participants were well poised to generalize to a velar POA. This POA is not new to participants, only new to them as listeners of Numana.

4.4 The Role of the Lexicon in L2 Phonetic Learning

An important contribution of the current study is the examination of the role of the lexicon in adult L2 phonetic learning. By examining the accuracy scores and the discrimination data together, the present investigation is able to comment on whether and how lexical learning informs phonetic learning. Understanding the degree to which phonetic learning and lexical learning are linked is integral to creating effective training paradigms for adult learners and for making recommendations for pedagogy.

Overall, the current study did not find a pervasive effect of training condition on participants' ability to discriminate along the VOT continuum. Recall, however, that one important exception emerged, which was that the minimal-pair group alone learned to accurately perceive and discriminate prevoiced tokens. Since both groups were exposed to the same number of items and the same distribution of segments and varied only in the presence and presentation of minimal pairs, this suggests that the minimal pairs were, in fact, the catalyst for this change.

The lexical-distributional hypothesis predicted that minimal pairs would impair phonetic learning since learners may mistakenly identify minimal pairs as homophonous or as two tokens of the same word. Instead, distributional learning without minimal pairs should have provided greater evidence for a phonetic and semantic distinction between tokens, and therefore non-minimal pairs should better facilitate the creation of separate phonetic category representations (Feldman et al., 2009, 2013a, 2013b). Given this hypothesis and the existing research supporting it, it was therefore predicted that the implicit, distributional learning group would more quickly and accurately acquire the novel VOT on which they were trained. The explicit group, with attention directed towards both word meaning and towards the novel, lower VOT category boundary was expected to encounter the type of cognitive overload that has led researchers to

report the hindrance of using minimal pairs in their previous experiments (Feldman et al., 2009, 2013a, 2013b). Taken one step further, previous literature has found that invoking the lexicon in any sense is detrimental to the creation of new phonetic categories, since attentional resources are then split between phonetic and lexical mechanisms (Hayes-Harb & Masuda, 2008; Jarvi, 2008; Thomson & Derwing, 2016).

Meanwhile, the minimal pair hypothesis posits that language learners should begin to attend to phonemic differences between sounds once they realize that the two contrasting phones of a given minimal pair can be used to make a distinction between two separate concepts (Maye & Gerken, 2000). It follows that the simplest way to ensure that learners come to this realization is to present the members side by side and highlight the semantic distinction by invoking the lexicon. This study's support for the minimal pair hypothesis comes in contrast to the predictions made at the start of the investigation, as well as in contrast to the literature that states that minimal pairs are disadvantageous to phonetic category formation (Feldman et al., 2009, 2013a, 2013b).

Studies that have previously sought to compare minimal pair and distributional learning have found little to no evidence for the benefit of minimal pairs. For example, Maye and colleagues (Maye & Gerken, 2000; Maye et al., 2002) compared the two hypotheses by exposing participants to either bimodal or unimodal stimuli along a continuum. Participants were simply exposed to a stream of aural stimuli but were not presented with any visual correlates or other methods for invoking the lexicon. In this distributional paradigm, participants were successfully able to classify the stimuli as belonging to one category in the unimodal group and two distinct categories in the bimodal group. Additionally, Feldman et al. (2013b) trained adult participants on a uniform distribution of either a non-minimal pair corpus or a minimal pair corpus. They too found a positive effect of differentiation of context on participants' ability to learn to perceive new vowel

categories, providing further support for the lexical-distributional hypothesis. Through this presentation of solely auditory stimuli, participants in the studies presented above (Feldman et al., 2013b; Maye & Gerken, 2000; Maye et al., 2002) were implicitly exposed to minimal pairs in the bimodal condition but did not have word meanings to help orient participants to this distribution. An important difference between the findings of these previous studies and those of the current study is the current study's use of pictures to provide evidence of the existence of distinct lexical items. Providing a forced choice between visual objects provides additional evidence that the words likely aren't homophonous, and that the auditory stimulus should only apply to one of the pictures displayed. An additional difference between previous studies and the current one is the current study's use of feedback throughout the learning process for both groups. The improvements across sessions on the lexical identification training tasks (Phase B and Phase C) may to at least some degree reflect testing effects such as feedback and repetition, since participants completed the Phase B and Phase C task on all five training sessions.

The formation of this link between the aural stimuli and the lexicon appears to be a strong driver for the formation of phonetic categories and helps to explain why the findings of the current study differ so greatly from the findings presented above from previous studies. A number of studies have found evidence for syllable-level training over word-level training, or non-lexical word-level training (such as reading from a word list or using nonce words without a corresponding referent) over invoking the lexicon and have therefore suggested phonetic training without the additional cognitive demand of juggling word meanings (Hayes-Harb & Masuda, 2008; Jarvi, 2008; Thomson & Derwing, 2016). The current study provides some nuance for this recommendation. While the data presented cannot address the particular dichotomy of lexical versus non-lexical learning, in cases where the lexicon *is* invoked – as was always the case in this

study – minimal pair training has been shown to be more effective than a more implicit, distributional condition that does not include minimal pairs.

Taken together, the lexicon does appear to be important in phonetic learning, but knowledge of word meanings – at least for adults – is critical for the formation of separate phonetic categories. Without semantic information, minimal pairs may hinder phonetic learning by suggesting to learners that minimal pairs are simply homophonous or various tokens of the same word as was predicted by the lexical-distributional hypothesis. With semantic information, minimal pairs home participants' attention into the relevant acoustic variables that are necessary for creating a phonemic contrast, as predicted by the minimal pair hypothesis and the noticing hypothesis.

4.5 Pedagogical Implications

The data of the current study contribute to a clearer understanding of the role of the lexicon in L2 phonetic learning and in particular provide evidence for the utility of minimal pairs for drawing learners' attention towards a phonemic contrast that might otherwise be overlooked. Although this study was conducted in a laboratory, there are still valuable lessons to be learned and applied to the language classroom.

Whether minimal pairs are a necessary component of the curriculum appears to depend on the nature of the variable being acquired. For lexical learning more broadly, the current study found few differences between the implicit and explicit groups, suggesting that minimal pairs were not necessarily integral, although they did initially impede word learning. The same is true for learning to generalize to new words, speakers, and POAs. Importantly, learners of both groups

were able to generalize to velar items despite receiving no training on them, which provides promising evidence for learners' ability to generalize to analogous contexts without needing to train every POA in the target language, at least in cases where the L1 has a similar phonological inventory from which learners can draw their L2 representations. What's more, learners were able to generalize to a more marked POA than the ones on which they were trained, suggesting that the generalization is in fact quite robust, even after relatively few training sessions. Teachers, therefore, may not need to devote classroom time to teaching every individual contrast. More implicit instruction may in fact be sufficient for many linguistic variables, but explicit minimal pair-based instruction does appear to be important for teaching phonetics when the contrast is within-category in the L1 but phonemic in the L2. Without minimal pairs, learners may not devote the necessary cognitive resources and attention to the phonetic variable in order for the input to become intake and contribute to learners' perception. It seems that simply hearing minimal pairs in and of itself is not necessarily sufficient, however, to induce this perceptual shift. Rather, the minimal pairs must have meanings assigned in order to invoke the lexicon. Additionally, feedback and side-by-side presentation of the minimal pairs may have also attributed to learning in the current study, although without comparison conditions for these particular variables, the degree to which each of these elements facilitated learning remains an empirical question.

Although minimal pairs may not be necessary for all phonetic variables, the data of the current study did not find evidence that they were in any way disadvantageous. By no measure did the implicit group significantly outperform the explicit group in the discrimination task. Additionally, despite some initial slowness in lexical acquisition, both groups reached accuracy rates of above 90% by the final training session. Given this, in combination with a number of calls for explicit instruction (Derwing & Munro, 2005; Gordon et al., 2013; Thompson & Derwing,

2014; Venkatagiri & Levis, 2007), it follows that minimal pairs may be an important component of explicitly directing learners' attention towards linguistic variables being taught, at least for matters of speech perception.

The good news is that the recommendation to incorporate meaningful minimal pairs is a natural addition to classrooms, and in many cases is already present. The training paradigm used could also be very easily adapted for language learning apps or Computer Assisted Language Learning so that learners could practice with vocabulary and undergo perception training simultaneously. This electronic means of delivery also provides the benefit of the ease of incorporating the high variability of speakers that can be more challenging in a language classroom with only one teacher. These implementations of the paradigm are relatively simple by design – this training procedure was developed with real life language learning in mind. Although this study took place in a laboratory, it was designed to be as ecologically valid as possible through its use of meaningful word learning and consistent correct/incorrect feedback. HVPT has been said to be as effective as it is due to being devoid of meaningful context and lexical access, but this study shows that learning can still take place without stripping these important components from the input. Future HVPT studies will do well to also incorporate meaning and context and to not shy away from implementing lexical access and especially minimal pairs, since with this study there is now greater support for the imperative role of the lexicon in phonetic learning.

It is worth noting that the aforementioned recommendations are based on changes to a singular and particular phonetic variable, VOT. Moreover, the recommendations are specific to training participants to perceive the prevoiced portion of the VOT continuum, which is used in English, especially when speaking emphatically, but is not used phonemically to distinguish between words. Given this, participants were not learning to listen for an entirely new phonetic

cue. Whether the findings and resulting recommendations of the present investigation would hold with additional phonetic variables is an empirical question. A number of variables may influence this, including the type of phonetic variable and its status in the L1 of the learner.

Cues varying along another acoustic dimension, such as F0 or other spectral contrasts, may or may not follow these learning patterns. Nevertheless, I expect that the general learning mechanisms that aided in the improved perception of prevoicing over the course of the experiment are the same ones that would aid in the learning of whichever acoustic cue were being trained. Given this, feedback, real word stimuli with referential meaning, and minimal pair presentations are likely still all beneficial for phonetic learning in general.

4.6 Limitations and Future Directions

Despite its contributions, this study was limited in its lack of inclusion of a third training condition with a lexicon that contains minimal pairs but does not present them side by side as was done in the current study's Training C 2AFC task. This control group would help to further understand whether learning still takes place due to the simple existence of minimal pairs in the lexicon, or whether it was their presentation side by side that directed learners' attention to them. I hypothesize that such a group would struggle to attend to the lower VOT category boundary without a side-by-side presentation of the minimal pair object, since, as predicted by the lexical distributional hypothesis, participants may simply assume the labels of these objects were simply homophonous. The setup of the current study left little doubt for the participants that there was a correct and incorrect answer; this would not have been possible given homophones.

It also remains to be seen whether the two training conditions result in differential production results. Although production data were collected, they were not analyzed for the purposes of this dissertation but will be explored moving forward. It stands to reason that the explicit group, having demonstrated an improved ability to *perceive* prevoicing, may also be better situated to more accurately *produce* prevoicing as well. This hypothesis is supported by Sakai and Moorman (2018), but it still warrants investigation with this particular set of sounds and this particular training paradigm.

As alluded to above, this study is somewhat limited in its ability to generalize to phonetic acquisition more broadly, given the inclusion of a single type of phonetic learning – VOT. Recommendations for classroom instruction are based off of the data of the current study, which solely examined learners' perception of the VOT category boundary in an L2. It is important to keep in mind that while these findings may speak to more widespread linguistic processes and language acquisition mechanisms, there is no concrete evidence that phonetic variables beyond VOT would respond to training in the same way or within the same number of exposures. Future studies will do well to continue to explore the trainability of additional phonetic variables. Additionally, the same study could be replicated to examine the efficacy of training on learners who have differential levels of experience learning languages with differing VOT boundaries, such as Spanish. The current study examined only learners with minimal or no experience with languages with a similar VOT category boundary to Numana. Perhaps, though, learners with intermediate or advanced levels of foreign language instruction would be better positioned to convert the input from the study into intake that would have a bearing on their perception, or perhaps would be able to make more effective use of the training of the implicit condition.

Finally, though importantly, it would not necessarily be expected to see global or fixed shifts to the participants' phonological systems after only five training sessions, especially when examining naïve learners. Instead, the results outlined in this study likely indicate a superficial perceptual shift towards a particular acoustic cue – in this case prevoicing or lack thereof. Whether participants could incorporate this information meaningfully in contexts larger than a single word or even in a production task remains to be seen. Nevertheless, this study has illustrated that participants can be trained to notice perception and incorporate it into their prevoicing in a relatively short amount of training under particular lexical conditions.

4.7 Conclusion

In conclusion, the present investigation examined the efficacy of two distinct vocabulary training conditions for instantiating adult L2 phonetic learning: minimal pair-based training, and purely distributional training. The experiment examined the effect of these conditions on the perception of a fine-grained phonetic variable, VOT, that differed between the training language and the participants' L1. This research adds to our understanding of the role of minimal pairs in phonetic learning, as well as the role of attention in generalizing to analogous contexts. The data suggest that, when invoking the lexicon, minimal pairs can help, rather than hinder, phonetic learning in an L2. The results of the current study also revealed that attention and the lexicon are both crucial components of both phonetic and lexical learning.

Appendix A NOUN Objects








Bibliography

Alaiga-García, C. & Mora, J.C. (2009). Assessing the effects of phonetic training on L2 sound perception and production. In M.A. Watkins, & B.O. Baptista (Eds.), *Recent research in second language phonetics/phonology: Perception and production.* 2-31.

Alves, U. K., & Luchini, P. L. (2017). Effects of Perceptual Training on the Identification and Production of Word-Initial Voiceless Stops by Argentinean Learners of English. *Ilha do Desterro*, *70*(3), 15-32.

Andersen, R. W. (1991). Developmental sequences: The emergence of aspect marking in second language acquisition. *Crosscurrents in second language acquisition and linguistic theories*, 305-324.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language*, *68*(3), 255-278.

Bassetti, B. (2008). Orthographic input and second language phonology. *Input matters in SLA*, 191-206.

Best, C. T., & Strange, W. (1992). Effects of phonological and phonetic factors on cross-language perception on approximants, *Journal of Phonetics, 20,* 305–330.

Berry, D.C. and Dienes, Z. (1993). Towards a working characterisation of implicit learning. In D.C. Berry and Z.Dienes (Eds.), *Implicit Learning: Theoretical and Empirical Issues* (pp.1-18). Hove: Lawrence Erlbaum Associates.

Bisson, M. J., Kukona, A., & Lengeris, A. (2020). An Ear And Eye For Language: Mechanisms Underlying Second Language Word Learning. *Bilingualism: Language and Cognition*, 1-20.

Bjork, R. A. (1994). Memory and metamemory considerations in the. *Metacognition: Knowing about knowing*, *185*.

Brusnighan, S. M., & Folk, J. R. (2012). Combining contextual and morphemic cues is beneficial during incidental vocabulary acquisition: Semantic transparency in novel compound word processing. *Reading Research Quarterly*, *47*(2), 172-190.

Carroll, J.S. and Johnson, E.J. (1990). *Decision Research: A Field Guide.* Newbury Park, CA: Sage.

Best, C. T. (1995). Learning to perceive the sound pattern of English. *Advances in infancy research*, *9*, 217-217.

Best, C. T., McRoberts, G. W., & Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *The Journal of the Acoustical Society of America*, *109*(2), 775-794.

Best, C. T., & Tyler, M. (2007). Nonnative and second-language speech perception. *Language experience in second language speech learning: In honour of James Emil Flege*, 13-34.

Bradlow, A. R. (2008). Training non-native language sound patterns: Lessons from training Japanese adults on the English /r/-/l/ contrast. In J. G. Hansen Edwards, & M. L. Zampini (Eds.), *Phonology and Second Language Acquisition* (pp. 287-308). John Benjamins Publishing Company.

Bradlow, A. R., Akahane-Yamada, R., Pisoni, D. B., & Tohkura, Y. I. (1999). Training Japanese listeners to identify English/r/and/l: Long-term retention of learning in perception and production. *Perception & psychophysics*, *61*(5), 977-985.

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., & Tohkura, Y. I. (1997). Training Japanese listeners to identify English/r/and/l: IV. Some effects of perceptual learning on speech production. *The Journal of the Acoustical Society of America*, *101*(4), 2299-2310.

Broersma, M., & Cutler, A. (2011). Competition dynamics of second-language listening. *Quarterly Journal of Experimental Psychology*, *64*(1), 74-95.

Carroll, S., & Swain, M. (1993). Explicit and implicit negative feedback: An empirical study of the learning of linguistic generalizations. *Studies in second language acquisition*, *15*(3), 357-386.

Caselli, M. C., Bates, E., Casadio, P., Fenson, J., Fenson, L., Sanderl, L., & Weir, J. (1995). A cross-linguistic study of early lexical development. *Cognitive Development*, *10*(2), 159-199.

Casillas, J. (2016). Longitudinal development of fine-phonetic detail in late learners of Spanish. Unpublished doctoral dissertation, University of Arizona.

Chang, C. (2012). Rapid and multifaceted effects of second-language learning on first-language speech production. *Journal of Phonetics*, *40*(2), 249-268.

Chang, C. (2013). A novelty effect in phonetic drift of the native language. *Journal of Phonetics*, *41*(6), 520-533.

Christensen, L. A., & Humes, L. E. (1997). Identification of multidimensional stimuli containing speech cues and the effects of training. *The Journal of the Acoustical Society of America*, *102*(4), 2297-2310.

Christie, J., & Klein, R. M. (1996). Assessing the evidence for novel popout. Journal of Experimental Psychology: General, 125(2), 201–207.

Coppens, P., & Patterson, J. (2017). Generalization in aphasiology: What are the best strategies? *Aphasia Rehabilitation: Clinical Challenges: Clinical Challenges*, 206.

Couper, G. (2006). The short and long-term effects of pronunciation instruction. *Prospect, 21* (1), 46-66.

Corral, J. A. M., & Wiedemann, E. J. G. (2009). La elisión de/d/intervocálica en el español culto de Granada: factores lingüísticos. *Pragmalingüística*, (17), 92-123.

Curtin, S., Goad, H., & Pater, J. V. (1998). Phonological transfer and levels of representation: the perceptual acquisition of Thai voice and aspiration by English and French speakers. *Second Language Research*, *14*(4), 389-405.

Cutler, A. (2015). Representation of second language phonology. *Applied Psycholinguistics*, *36*(1), 115-128.

Cutler, A., Weber, A., & Otake, T. (2006). Asymmetric mapping from phonetic to lexical representations in second-language listening. *Journal of Phonetics*, *34*(2), 269-284.

DeKeyser, R. (2008). 11 Implicit and Explicit Learning. *The handbook of second language acquisition*, *27*, 313.

DeKeyser, R. (2015). Skill acquisition theory. In VanPatten B. & Williams J. (Eds.), Theories in second language acquisition: An introduction (2nd ed., pp. 94-112). Routledge.

Derwing, T. M., & Munro, M. J. (2005). Second language accent and pronunciation teaching: A research-based approach. *TESOL quarterly*, *39*(3), 379-397.

Derwing, T. M. (2010). Utopian goals for pronunciation teaching. In *Proceedings of the 1st pronunciation in second language learning and teaching conference*, 24-37.

Díaz-Campos, M. (2004). Context of learning in the acquisition of Spanish second language phonology. *Studies in second language acquisition*, *26*(2), 249-273.

Dickey, M., & Yoo, H. (2013). Acquisition versus generalization in sentence production treatment in aphasia: dose-response relationships. *Procedia-Social and Behavioral Sciences*, *94*, 281-282.

Dietrich, C., Swingley, D., & Werker, J. F. (2007). Native language governs interpretation of salient speech sound differences at 18 months. *Proceedings of the National Academy of Sciences*, *104*(41), 16027-16031.

Dmitrieva, O., Llanos, F., Shultz, A. A., & Francis, A. L. (2015). Phonological status, not voice onset time, determines the acoustic realization of onset f0 as a secondary voicing cue in Spanish and English. *Journal of Phonetics*, *49*, 77-95.

Dumay, N., & Gaskell, M. G. (2007). Sleep-associated changes in the mental representation of spoken words. *Psychological Science*, *18*(1), 35-39.

E-Prime (Version 3.0) [Computer program] (2016).

Eckman, F. R. (2008). Typological markedness and second language phonology. *Phonology and second language acquisition, 36*, 95-115.

Ellis, R. (2015). Form-focused instruction and the measurement of implicit and explicit L2 knowledge. *Implicit and explicit learning of languages*, 417-441.

Ellis, R., Loewen, S., & Erlam, R. (2006). Implicit and explicit corrective feedback and the acquisition of L2 grammar. *Studies in second language acquisition*, *28*(2), 339-368.

Engonopulos, N., Sayeed, A., & Demberg, V. (2013). Language and cognitive load in a dual task environment. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 35, No. 35).

Escudero, P. (2007). Second-language phonology: The role of perception. In *Phonology in context* (pp. 109-134). Palgrave Macmillan, London.

Escudero, P., Benders, T., & Wanrooij, K. (2011). Enhanced bimodal distributions facilitate the learning of second language vowels. *The Journal of the Acoustical Society of America*, *130*(4), EL206-EL212.

Face, T. L. (2018). Ultimate attainment in Spanish spirantization. *Spanish in Context*, *15*(1), 27-53.

Face, T. L. (2021). What does advanced L2 pronunciation look like? *Advancedness in Second Language Spanish: Definitions, challenges, and possibilities*, *31*, 143.

Feldman, N. H., Griffiths, T. L., Goldwater, S., & Morgan, J. L. (2013a). A role for the developing lexicon in phonetic category acquisition. *Psychological review*, *120*(4), 751.

Feldman, N., Griffiths, T., & Morgan, J. (2009). Learning phonetic categories by learning a lexicon. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 31, No. 31).

Feldman, N. H., Myers, E. B., White, K. S., Griffiths, T. L., & Morgan, J. L. (2013b). Word-level information influences phonetic learning in adults and infants. *Cognition*, *127*(3), 427-438.

Fennell, C. T., & Werker, J. F. (2003). Early word learners' ability to access phonetic detail in well-known words. *Language and speech*, *46*(2-3), 245-264.

Flannagan, M. J., Fried, L. S., & Holyoak, K. J. (1986). Distributional expectations and the induction of category structure. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*(2), 241.

Flege, J. (1987). The production of “new” and “similar” phones in a foreign language: evidence for the effect of equivalence classification. *Journal of Phonetics, 15*, 47-65.

Flege, J. E. (1988). The production and perception of foreign language speech sounds. *Human communication and it's disorders-a review*, 224-401.

Flege, J. E. (1991). Age of learning affects the authenticity of voice-onset time (VOT) in stop consonants produced in a second language. *The Journal of the Acoustical Society of America, 89*(1), 395-411.

Flege, J. E. (1992a). The intelligibility of English vowels spoken by British and Dutch talkers. *Intelligibility in speech disorders: Theory, measurement, and management, 1*, 157-232.

Flege, J. E. (1992b). Speech learning in a second language. *Phonological development: Models, research, implications, 565*, 604.

Flege, J. (1993). Production and perception of a novel, second-language phonetic contrast. *Journal of the Acoustical Society of America, 93*(3), 1589-1608.

Flege, J. E. (1995a). Second language speech learning: Theory, findings, and problems. *Speech perception and linguistic experience: Issues in cross-language research*, *92*, 233-277.

Flege, J. (1995b). Two procedures for training a novel second-language phonetic contrast. *Applied Psycholinguistics*, 16, 425-442.

Flege, J. (1998). Second-language learning: The role of subject and phonetic variables. In *STiLL-Speech Technology in Language Learning*.

Flege, J. E. (2002). *Interactions between the native and second-language phonetic systems* (pp. 217-244). Trier, Germany: Wissenschaftligher Verlag.

Flege, J. E. (2003). Assessing constraints on second-language segmental production and perception. *Phonetics and phonology in language comprehension and production: Differences and similarities*, *6*, 319-355.

Flege, J. E., Bohn, O. S., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of phonetics*, *25*(4), 437-470.

Flege, J. E., Munro, M. J., & MacKay, I. R. (1995). Factors affecting strength of perceived foreign accent in a second language. *The Journal of the Acoustical Society of America*, *97*(5), 3125-3134.

Flege, J., Schirru, C., & MacKay, I. R. (2003). Interaction between the native and second language phonetic subsystems. *Speech Communication*, *40*(4), 467-491.

Folstein, J. R., Van Petten, C., & Rose, S. A. (2008). Novelty and conflict in the categorization of complex stimuli. *Psychophysiology*, *45*(3), 467-479.

Francis, A. L., & Nusbaum, H. C. (2002). Selective attention and the acquisition of new phonetic categories. *Journal of Experimental Psychology: Human perception and performance, 28*(2), 349.

George, A. (2013). The development of /θ/, a variable geographic phonetic feature, during a semester abroad: The role of explicit instruction. In J. Levis & K. LeVelle (Eds.). Proceedings of the 4th Pronunciation in Second Language Learning and Teaching Conference. Aug. 2012. (pp. 120- 128). Ames, IA: Iowa State University.

Godfroid, A., & Schmidtke, J. (2013). What do eye movements tell us about awareness? A triangulation of eye-movement data, verbal reports and vocabulary learning scores. *Noticing and second language acquisition: Studies in honor of Richard Schmidt*, 183-205.

Gomez, R. L., & Gerken, L. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition, 70*(2), 109-135.

Gordon, J., Darcy, I., & Ewert, D. (2013). Pronunciation teaching and learning: Effects of explicit phonetic instruction in the L2 classroom. In *Proceedings of the 4th pronunciation in second language learning and teaching conference* (pp. 194-206).

Guion, S. G., & Pederson, E. (2007). Investigating the role of attention in phonetic learning. *Language experience in second language speech learning*, 57-77.

Hawley, K. J., Johnston, W. A., & Farnham, J. M. (1994). Novel popout with nonsense strings: Effects of predictability of string length and spatial location. *Perception & Psychophysics*, *55*(3), 261-268.

Hayes-Harb, R. (2007). Lexical and statistical evidence in the acquisition of second language phonemes. *Second Language Research*, *23*(1), 65-94.

Hayes-Harb, R., & Barrios, S. (2019). Investigating the phonological content of learners'"fuzzy" lexical representation for new L2 words. In *Proceedings of the 10th Pronunciation in Second Language Learning and Teaching conference*. *Ames, IA: Iowa State University* (pp. 55-69).

Hayes-Harb, R., & Masuda, K. (2008). Development of the ability to lexically encode novel second language phonemic contrasts. *Second Language Research*, *24*(1), 5-33.

Hebb, D. O. (1961). Distinctive features of learning in the higher animal. *Brain mechanisms and learning*, *37*, 46.

Henrichsen, L. E. (1984). Sandhi-variation: A filter of input for learners of ESL. *Language Learning*, *34*(3), 103-123.

Henriksen, N. C., Geeslin, K. L., & Willis, E. W. (2010). The development of L2 Spanish intonation during a study abroad immersion program in León, Spain: Global contours and final boundary movements. *Studies in Hispanic and Lusophone linguistics*, *3*(1), 113-162.

Horst, J. S., & Hout, M. C. (2016). The Novel Object and Unusual Name (NOUN) Database: A collection of novel images for use in experimental research. *Behavior research methods*, *48*(4), 1393-1409.

Huensch, A. (2019). The pronunciation teaching practices of university-level graduate teaching assistants of French and Spanish introductory language courses. *Foreign Language Annals*, *52*(1), 13-31.

Hulstijn, J. H. (2005). Theoretical and empirical issues in the study of implicit and explicit second-language learning: Introduction. *Studies in second language acquisition*, *27*(2), 129-140.

Idemaru, K., & Holt, L. L. (2020). Generalization of dimension-based statistical learning. *Attention, Perception, & Psychophysics*, 1-19.

Iverson, P., & Evans, B. G. (2007). Learning English vowels with different first-language vowel systems: Perception of formant targets, formant movement, and duration. *The Journal of the Acoustical Society of America*, *122*(5), 2842-2854.

Iverson, P., Hazan, V., & Bannister, K. (2005). Phonetic training with acoustic cue manipulations: A comparison of methods for teaching English/r/-/l/to Japanese adults. *The Journal of the Acoustical Society of America*, *118*(5), 3267-3278.

Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Kettermann, A., & Siebert, C. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition*, *87*(1), B47-B57.

Jacobs, A., Fricke, M., & Kroll, J. F. (2016). Cross-Language Activation Begins During Speech Planning and Extends into Second Language Speech. *Language Learning*, *66*(2), 324-353.

Jarvi, A. O. (2008). *Effect of Lexical Access and Meaningful Linguistic Context on Second Language Speech Perception* (Doctoral dissertation, Department of Linguistics, University of Utah).

Johnson, K. (1997). Speech perception without speaker normalization: An exemplar model. In *Talker variability in speech processing* (pp. 145-165). Burlington, MA: Morgan Kauffman.

Johnston, W. A., Hawley, K. J., & Farnham, J. M. (1993). Novel popout: Empirical boundaries and tentative theory. *Journal of Experimental Psychology: Human Perception and Performance*, *19*(1), 140.

Johnston, W. A., Hawley, K. J., Plewe, S. H., Elliott, J. M., & DeWitt, M. J. (1990). Attention capture by novel stimuli. *Journal of Experimental Psychology: General*, *119*(4), 397.

Jusczyk, P. W. (1985). On characterizing the development of speech perception. In Erlbaum. J. Mehler & R. Fox (Eds.), *Neonate cognition: beyond the blooming, buzzing, confusion* (pp. 199–229). Erlbaum

Kissling, E. M. (2013). Teaching pronunciation: Is explicit phonetics instruction beneficial for FL learners? *The modern language journal*, *97*(3), 720-744.

Klatt, D. H. (1979). Speech perception: A model of acoustic–phonetic analysis and lexical access. *Journal of phonetics*, *7*(3), 279-312.

Klein, E. C. (1995). Second versus third language acquisition: Is there a difference? *Language learning*, *45*(3), 419-466.

Klein, W., & Dimroth, C. (2009). Untutored second language acquisition. In *The new handbook of second language acquisition* (pp. 503-522). Emerald.

Kondaurova, M. V., & Francis, A. L. (2010). The role of selective attention in the acquisition of English tense and lax vowels by native Spanish listeners: Comparison of three training methods. *Journal of phonetics*, *38*(4), 569-587.

Kraljic, T., & Samuel, A. G. (2005). Perceptual learning for speech: Is there a return to normal? *Cognitive psychology*, *51*(2), 141-178.

Kraljic, T., & Samuel, A. G. (2006). Generalization in perceptual learning for speech. *Psychonomic bulletin & review*, *13*(2), 262-268.

Kraljic, T., & Samuel, A. G. (2007). Perceptual adjustments to multiple speakers. *Journal of Memory and Language*, *56*(1), 1-15.

Kuhl, P. K. (1991). Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perception & psychophysics*, *50*(2), 93-107.

Kuhl, P. K. (1993). Innate predispositions and the effects of experience in speech perception: The native language magnet theory. In *Developmental neurocognition: Speech and face processing in the first year of life* (pp. 259-274). Springer, Dordrecht.

Kuhl, P. K. (2004). Early language acquisition: cracking the speech code. *Nature reviews neuroscience*, *5*(11), 831-843.

Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental science*, *9*(2), F13-F21.

Lalonde, C. E., & Werker, J. F. (1995). Cognitive influences on cross-language speech perception in infancy. *Infant Behavior and Development*, *18*(4), 459-475.

Lambacher, S. G., Martens, W. L., Kakehi, K., Marasinghe, C. A., & Molholt, G. (2005). The effects of identification training on the identification and production of American English vowels by native speakers of Japanese. *Applied Psycholinguistics*, *26*(2), 227.

Lenneberg, E. H., (1967). *Biological foundations of language.* New York: Wiley.

Leow, R. P. (1998). Toward operationalizing the process of attention in SLA: Evidence for Tomlin and Villa's (1994) finegrained analysis of attention. *Applied Psycholinguistics*, *19*(1), 133-159.

Leow, R. P. (2000). A study of the role of awareness in foreign language behavior: Aware versus unaware learners. *Studies in second language acquisition*, 557-584.

Loewen, S. (2012). The role of feedback. In S. M. Gass & A. Mackey (eds.), *The Routledge handbook of second language acquisition*. New York: Routledge, 24–40.

Loewen, S., & Sato, M. (2018). Interaction and instructed second language acquisition. *Language teaching*, *51*(3), 285-329.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological review*, *74*(6), 431.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, *21*(1), 1-36.

Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, *20*(3), 384-422.

Lively, S. E., Logan, J. S., & Pisoni, D. B. (1993). Training Japanese listeners to identify English/r/and/l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. *The Journal of the acoustical society of America*, *94*(3), 1242-1255.

Lively, S. E., Pisoni, D. B., Yamada, R. A., Tohkura, Y. I., & Yamada, T. (1994). Training Japanese listeners to identify English/r/and/l/. III. Long-term retention of new phonetic categories. *The Journal of the acoustical society of America*, *96*(4), 2076-2087.

Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English/r/and/l: A first report. *The Journal of the Acoustical Society of America*, *89*(2), 874-886

Lord, G. (2005). (How) can we teach foreign language pronunciation? On the effects of a Spanish phonetics course. *Hispania*, 557-567.

Lord, G. (2010). The combined effects of immersion and instruction on second language pronunciation. *Foreign language annals*, *43*(3), 488-503.

Lord, G., & Fionda, M. I. (2013). Teaching Pronunciation in Second Language Spanish. *The Handbook of Spanish Second Language Acquisition*, 514-529.

Lyster, R., & Saito, K. (2010). Oral feedback in classroom SLA: A meta-analysis. *Studies in second language acquisition*, 265-302.

MacKain, K. S. (1982). Assessing the role of experience on infants' speech discrimination. *Journal of Child Language*, *9*(3), 527-542.

Mackey, A., Gass, S., & McDonough, K. (2000). How do learners perceive interactional feedback? *Studies in second language acquisition*, *22*(4), 471-497.

Mackey, A., & Goo, J. (2007). Interaction research in SLA: A meta-analysis and research synthesis. In A. Mackey (Ed.), *Conversational interaction and second language acquisition: A collection of empirical studies* (pp. 407-452). Oxford: Oxford University Press.

Mackey, A., Philp, J., Egi, T., Fujii, A., & Tatsumi, T. (2002). Individual differences in working memory, noticing of interactional feedback and L2 development. In P. Robinson & P. Skehan (Eds.), *Individual differences in L2 learning* (pp. 181-208). John Benjamins.

Martínez-Celdrán, E. M. (1993). La percepción categorial de /b-p/ en español basada en las diferencias de duración. *Estudios de fonética experimental*, 223-239.

Maye, J., & Gerken, L. (2000). Learning phonemes without minimal pairs. *Proceedings of the 24th annual Boston university conference on language development*, *2*, (522-533).

Maye, J., & Gerken, L. (2001). Learning phonemes: How far can the input take us. In *Proceedings of the 25th annual Boston University conference on language development* (Vol. 1, p. 480). Somerville, MA: Cascadilla Press.

Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, *82*(3), B101-B111.

McClaskey, C. L., Pisoni, D. B., & Carrell, T. D. (1983). Transfer of training of a new linguistic contrast in voicing. *Perception & psychophysics*, *34*(4), 323-330.

Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoncini, J., & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition*, *29*(2), 143-178.

Mora, J. C. (2008). Learning context effects on the acquisition of a second language phonology. *A portrait of the young in the new multilingual Spain*, 241-263.

Moreton, E., & Pater, J. (2012). Structure and substance in artificial-phonology learning, part I: Structure. *Language and linguistics compass*, *6*(11), 686-701.

Morgan-Short, K., Steinhauer, K., Sanz, C., & Ullman, M. T. (2012). Explicit and implicit second language training differentially affect the achievement of native-like brain activation patterns. *Journal of cognitive neuroscience*, *24*(4), 933-947.

Nagle, C. L. (2017). Individual developmental trajectories in the L2 acquisition of Spanish spirantization. *Journal of Second Language Pronunciation*, *3*(2), 218-241.

Nagle, C., Sachs, R., & Zárate-Sández, G. (2020). Spanish teachers' beliefs on the usefulness of pronunciation knowledge, skills, and activities and their confidence in implementing them. *Language Teaching Research*, 1-27.

Neri, A., Mich, O., Gerosa, M., & Giuliani, D. (2008). The effectiveness of computer assisted pronunciation training for foreign language learning by children. *Computer Assisted Language Learning*, *21*(5), 393-408.

Nicholas, H., Lightbown, P. M., & Spada, N. (2001). Recasts as feedback to language learners. *Language learning*, *51*(4), 719-758.

Nielsen, K. Y. (2008). *Word-level and Feature-level Effets in Phonetic Imitation*. University of California, Los Angeles.

Nielsen, K. (2011). Specificity and abstractness of VOT imitation. *Journal of Phonetics*, *39*(2), 132-142.

Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive psychology*, *47*(2), 204-238.

Olejarczuk, P., Kapatsinski, V., & Baayen, R. H. (2018). Distributional learning is error-driven: The role of surprise in the acquisition of phonetic categories. *Linguistics Vanguard*, 1-9.

Pater, J. (2003). The perceptual acquisition of Thai phonology by English speakers: task and stimulus effects. *Second Language Research*, *19*(3), 209-223.

Pater, J., Stager, C., & Werker, J. (2004). The perceptual acquisition of phonological contrasts. *Language*, 384-402.

Pederson, E., & Guion-Anderson, S. (2010). Orienting attention during phonetic training facilitates learning. *The Journal of the Acoustical Society of America*, *127*(2), EL54-EL59.

Pierrehumbert, J. B. (2003). Phonetic diversity, statistical learning, and acquisition of phonology. *Language and speech*, *46*(2-3), 115-154.

Pisoni, D. B., Aslin, R. N., Perey, A. J., & Hennessy, B. L. (1982). Some effects of laboratory training on identification and discrimination of voicing contrasts in stop consonants. *Journal of Experimental Psychology: Human perception and performance*, *8*(2), 297.

Pollock, M. (2020). Did you say peso or beso? *Variation and Evolution: Aspects of language contact and contrast across the Spanish-speaking world*, *29*, 127.

Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of verbal learning and verbal behavior*, *6*(6), 855-863.

Rebuschat, P. (2013). Measuring implicit and explicit knowledge in second language research. *Language Learning*, *63*(3), 595-626.

Ringer-Hilfinger, K. (2012). Learner acquisition of dialect variation in a study abroad context: The case of the Spanish [θ]. *Foreign Language Annals*, *45*(3), 430-446.

Rose, M. (2010). Differences in discriminating L2 consonants: A comparison of Spanish taps and trills. In Y. Watanabe, M. Prior, & S. –K. Lee (eds.), *Selected Proceedings of the 2008 Second Language Forum*, 181-196. Sommerville, MA: Cascadilla Proceedings Project.

Saito, K. (2011). Examining the role of explicit phonetic instruction in native-like and comprehensible pronunciation development: An instructed SLA approach to L2 phonology. *Language awareness*, *20*(1), 45-59.

Saito, K. (2018). The role of aptitude in second language segmental learning: the case of Japanese learners' English/r/pronunciation attainment in classroom settings. *Applied Psycholinguistics*.

Saito, K., & Lyster, R. (2012). Effects of form-focused instruction and corrective feedback on L2 pronunciation development of/ɹ/by Japanese learners of English. *Language learning*, *62*(2), 595-633.

Saito, K., & Plonsky, L. (2019). Effects of second language pronunciation teaching revisited: A proposed measurement framework and meta-analysis. *Language Learning*, *69*(3), 652-708.

Sakai, M., & Moorman, C. (2018). Can perception training improve the production of second language phonemes? A meta-analytic review of 25 years of perception training research. *Applied Psycholinguistics*, *39*(1), 187-224.

Schertz, J., Cho, T., Lotto, A., & Warner, N. (2015). Individual differences in phonetic cue use in production and perception of a non-native sound contrast. *Journal of Phonetics*,*52*(Sep), 183-204.

Schmidt, L. B. (2018). L2 development of perceptual categorization of dialectal sounds. *Studies in Second Language Acquisition*, *40*(4), 857-882.

Schmidt, R. W. (1990). The role of consciousness in second language learning1. *Applied linguistics*, *11*(2), 129-158.

Schmidt, R. (1994). Implicit learning and the cognitive unconscious: Of artificial grammars and SLA. *Implicit and explicit learning of languages*, *22*, 165-209.

Schmidt, R. (1995). Consciousness and foreign language learning: A tutorial on the role of attention and awareness in learning. *Attention and awareness in foreign language learning*, *9*, 1-63.

Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3-32). Cambridge: Cambridge University Press.

Schmidt, R. (2012). Attention, awareness, and individual differences in language learning. In W. M. Chan, K. N. Chin, S. Bhatt, & I. Walker (Eds.), *Perspectives on individual*

characteristics and foreign language education (pp. 27–50). Boston, MA: Mouton de Gruyter.

Schmidt, J. R., Crump, M. J., Cheesman, J., & Besner, D. (2007). Contingency learning without awareness: Evidence for implicit control. *Consciousness and cognition*, *16*(2), 421-435.

Schmidt, J. R., De Houwer, J., Moors, A., & Vazire, S. (2020). Learning habits: Does overtraining lead to resistance to new learning? *Collabra: Psychology*, *6*(1).

Schmidt, R., & Frota, S. (1986). Developing basic conversational ability in a second language: A case study of an adult learner of Portuguese. *Talking to learn: Conversation in second language acquisition*, *237*, 326.

Schoonmaker-Gates, E. (2017). Regional variation in the language classroom and beyond: Mapping learners' developing dialectal competence. *Foreign Language Annals*, *50*(1), 177-194.

Simonet, M. (2012). 34 The L2 Acquisition of Spanish Phonetics and Phonology. *The handbook of hispanic linguistics*, 729.

Stager, C. L., & Werker, J. F. (1997). Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature*, *388*(6640), 381-382.

Stokes, R. C., Venezia, J. H., & Hickok, G. (2019). The motor system's [modest] contribution to speech perception. *Psychonomic bulletin & review*, *26*(4), 1354-1366.

Storkel, H. L., Armbrüster, J., & Hogan, T. P. (2006). Differentiating phonotactic probability and neighborhood density in adult word learning.

Storkel, H. L., Bontempo, D. E., & Pak, N. S. (2014). Online learning from input versus offline memory evolution in adult word learning: Effects of neighborhood density and phonologically related practice. *Journal of speech, language, and hearing research*, *57*(5), 1708-1721.

Sobczak, J. M., & Gaskell, M. G. (2019). Implicit versus explicit mechanisms of vocabulary learning and consolidation. *Journal of Memory and Language*, *106*, 1-17.

Solon, M., Linford, B., & Geeslin, K. L. (2018). Acquisition of sociophonetic variation: Intervocalic/d/reduction in native and nonnative Spanish. *Revista Española de Lingüística Aplicada/Spanish Journal of Applied Linguistics*, *31*(1), 309-344.

Strange, W., & Dittmann, S. (1984). Effects of discrimination training on the perception of /rl/ by Japanese adults learning English. *Perception & psychophysics*, *36*(2), 131-145.

Szmalec, A., Page, M. P., & Duyck, W. (2012). The development of long-term lexical representations through Hebb repetition learning. *Journal of Memory and Language*, *67*(3), 342-354.

Thiessen, E. D. (2007). The effect of distributional information on children's use of phonemic contrasts. *Journal of Memory and Language*, *56*(1), 16-34.

Thiessen, E. D. (2011). When variability matters more than meaning: the effect of lexical forms on use of phonemic contrasts. *Developmental psychology*, *47*(5), 1448.

Thomson, R. I. (2011). Computer assisted pronunciation training: Targeting second language vowel perception improves pronunciation. *Calico Journal*, *28*(3), 744.

Thomson, R. I. (2012). Improving L2 listeners' perception of English vowels: A computer-mediated approach. *Language Learning*, *62*(4), 1231-1258.

Thomson, R. I., & Derwing, T. M. (2014). The effectiveness of L2 pronunciation instruction: A narrative review. *Applied Linguistics*, *36*(3), 326-344.

Thomson, R. I., & Derwing, T. M. (2016). Is phonemic training using nonsense or real words more effective. In *Proceedings of the 7th Pronunciation in Second Language Learning and Teaching Conference* (pp. 88-97). Ames, IA: Iowa State University.

Tomlin, R. S., & Villa, V. (1994). Attention in cognitive science and second language acquisition. *Studies in second language acquisition*, *16*(2), 183-203.

Tremblay, K., Kraus, N., Carrell, T. D., & McGee, T. (1997). Central auditory system plasticity: generalization to novel stimuli following listening training. *The Journal of the Acoustical Society of America*, *102*(6), 3762-3773.

Trofimovich, P., Ammar, A., & Gatbonton, E. (2007). How effective are recasts? The role of attention, memory, and analytical ability. *Conversational interaction in second language acquisition: A collection of empirical studies*, 171-195.

Truscott, J., & Smith, M. S. (2004). Acquisition by processing: A modular perspective on language development. *Bilingualism*, *7*(1), 1.

Truscott, J., & Smith, M. S. (2011). Input, intake, and consciousness: The quest for a theoretical foundation. *Studies in Second Language Acquisition*, 497-528.

Ullman, M. T. (2015). The declarative/procedural model: A neurobiologically motivated theory of first and second language. In VanPatten B. & Williams J. (Eds.), *Theories in second language acquisition: An introduction*, (2nd ed., pp. 135-158). Routledge.

Venkatagiri, H. S., & Levis, J. M. (2007). Phonological awareness and speech comprehensibility: An exploratory study. *Language awareness*, *16*(4), 263-277.

Weatherholtz, K., & Jaeger, T. F. (2016). Speech perception and generalization across talkers and accents. In *Oxford Research Encyclopedia of Linguistics*.

Werker, J. F., & Pegg, J. E. (1992). Infant speech perception and phonological acquisition. *Phonological development: Models, research, implications*, 285-311.

Williams, J. N. (2005). Learning without awareness. *Studies in Second Language Acquisition*, *27*, 269–304.

Willis, E., Geeslin, K., & Henriksen, N. (2009, October). The acquisition of /θ/ by study abroad learners in León, Spain. In *13th Hispanic Linguistics Symposium, San Juan, Puerto Rico*.

Yang, H., Chen, X., & Zelinsky, G. J. (2009). A new look at novelty effects: Guiding search away from old distractors. *Attention, Perception, & Psychophysics*, *71*(3), 554-564.

Zampini, M. (1998). The Relationship between the Production and Perception of L2 Spanish Stops. *Texas Papers in Foreign Language Education*, *3*(3), 85-100.

Zampini, M. L., & Green, K. P. (2001). The voicing contrast in English and Spanish: The relationship between perception and production. *One mind, two languages: Bilingual language processing*, 23-48.

Zhang, R., & Yuan, Z. M. (2020). Examining the effects of explicit pronunciation instruction on the development of L2 pronunciation. *Studies in Second Language Acquisition*, *42*(4), 905-918.