SPANISH NATIVE-SPEAKER PERCEPTION
OF ACCENTEDNESS IN
LEARNER SPEECH

A Dissertation
Submitted to
the Temple University Graduate Board

In Partial Fulfillment
of the Requirements for the Degree
DOCTOR OF PHILOSOPHY

by
Kara Moranski
May 2012

Examing Committee Members:
Dr. Paul Toth, Advisory Chair, Department of Spanish and Portuguese
Dr. Augusto Lorenzino, Department of Spanish and Portuguese
Dr. Jonathan Holmquist, Department of Spanish and Portuguese
Dr. Brian McHugh, External Member, Temple University Department of Communication Sciences and Disorders
ABSTRACT

Building upon current research in native-speaker (NS) perception of L2 learner phonology (Zielinski, 2008; Derwing & Munro, 2009), the present investigation analyzed multiple dimensions of NS speech perception in order to achieve a more complete understanding of the specific linguistic elements and attitudinal variables that contribute to perceptions of accent in learner speech. In this mixed-methods study, Spanish monolinguals (n = 18) provided information regarding their views of L1 American English (AE) speakers learning Spanish and also evaluated the extemporaneous production of L2 learners from this same population. The evaluators’ preconceived attitudinal notions of L1 AE speakers learning Spanish negatively correlated with numerical accentedness ratings for the speech samples, indicating that evaluators with more positive perceptions of the learners rated their speech as less accented. Following initial numerical ratings, evaluators provided detailed commentary on the individual phonological elements from each utterance that they perceived as “nonnative.” Results show that differences in the relative salience of the nonnative segmental productions correspond with certain phonetic and phonemic processes occurring within the sounds, such as aspiration, spirantization and lateralization.

Keywords: native-speaker perception, nonnative accent, SLA, attitudinal factors, L2 phonology
ACKNOWLEDGEMENT

During my time at Temple University I have had the opportunity to learn from a very talented group of scholars. It is through their instruction and guidance that I have acquired the theoretical and technical knowledge that made this research possible. Dr. Paul Toth exemplifies excellence as an advisor, mentor, and researcher. His is the standard to which I will hold myself as I continue in my career. Dr. Brian McHugh and Dr. Augusto Lorenzino literally taught me everything I know about phonetics and phonology. It was because of Dr. Jonathan Holmquist that I learned to truly appreciate the complex social nature of language. Dr. Jennifer Cromley taught me the art of using rigorous statistical methodology in social-science research and gave generously of her own time in order to consult with me on this project. I am also very appreciative of the help that I received from my friends and colleagues while I was piloting this study and preparing speech samples, in particular that of José Luis Montiel, Cristina Anaya García, Matthew Fischetti, Pedro Rubio, Milca Dubón, Gabriela Ramírez-Dávalos, Moira Álvarez, and Marta Hidalgo Romero. I cannot thank them enough for their support, time, and patience.

My good fortune extends beyond the university; I have been blessed with a wonderful and supportive family. My parents, Martha and Ron Pristas and Nick and Beverly Moranski, have always encouraged me. I have looked to their professional success as a model for my own life. My grandfather, Frank “Moose” Moranski, taught me the value of hard work and consistency, and for that I am forever grateful. Finally, my fiancé Andrew has been a steady source of support and inspiration. His intellectual curiosity and dedication to his own work motive me every day.
# TABLE OF CONTENTS

ABSTRACT.................................................................................................................. iii

ACKNOWLEDGEMENT................................................................................................ iv

LIST OF TABLES......................................................................................................... viii

LIST OF FIGURES...................................................................................................... ix

CHAPTER

1. SECOND-LANGUAGE PHONOLOGICAL ACQUISITION AND
   THE “LEARNAIBILITY” DEBATE................................................................. 1
   Historical background...................................................................................... 2
   Early Views on L2 Pronunciation and the Effects of Pronunciation
   Instruction........................................................................................................ 7
   CPH, SLA and L2 Phonology........................................................................... 10
   L2 Phonological Instruction: Making Gains.................................................. 13
   Models for L2 Phonological Instruction......................................................... 16
   L2PA: Varying Degrees of Difficulty.............................................................. 23
   Learner Data Elicitation................................................................................... 26
   Conclusion and Research Questions............................................................... 29

2. NATIVE-SPEAKER PERCEPTION OF NONNATIVE L2
   PHONOLOGY................................................................................................. 30
   NNS Perception: the Phonology Problem....................................................... 32
   Studies on NS Perception of Individual Components of NNS
   Phonological Production.................................................................................. 35
   Semantic Value of Segmental Features......................................................... 47
   Weighting Perception of Segmentals in Learner Speech.............................. 49
      The Functional Load Principle.................................................................. 49
   Markedness considerations............................................................................ 52
      Salience Of Unmarked Segments............................................................... 53
      Salience Of Marked Segments................................................................. 53
   Affective Variables and the NS Listener......................................................... 53
Summary of NS Perception and Research Questions

1. “Ambiguous Accent” as an Anachronism
2. Segmental Versus Suprasegmental Features
3. Semantic Segmentals
4. Weighted Segmentals
5. Attitudinal Adjustments

3. METHODOLOGY AND DATA COLLECTION

Phase I: Collection and Analysis of Nonnative Speech Samples

Learner Participants
Elicitation Of Learner Speech Samples
Speech Sample Splicing
Error Analysis Of Learner Speech
Commentary On Error Distribution In Learner Samples

Phase II: Data Elicitation from NS Evaluators

NS-Listener Participants
Online Survey Tool: Overview
Instructions And Practice Speech Sample Ratings
Accentedness Data Collection
Attitudinal Data Collection

4. RESULTS

NS Likert Scale Ratings and Attitudinal Data
Measures Of Central Tendency
Likert-Scale Attitudinal Data And Correlation With Speech-Sample Ratings
Additional Attitudinal Data
Results From Free-Response Data
Inter-rater Reliability
Categories
Quantitative analysis of segmental data
Chi-square Analysis
Summary of Results
### DISCUSSION AND CONCLUSION

NS Listener Profile and Affective Factors

Results of NS Observations of Non-Target Segmental Production

1. Sounds with Phonemic and Manner-of-Articulation Differences
2. Over-aspirated Plosives: /p ʰ k/
3. /l/ Velarization
4. Unspirantized Allophones of Voiced Plosives /b d g/

Theoretical Considerations

Markedness

Functional Load

Pedagogical Implications

Limitations and Implications for Further Research

Accent Versus Intelligibility

Participant self-selection

Follow-up Interviews

Limitations of Speech Samples

Conclusion

REFERENCES

APPENDICES

A: VOCABULARY TERMS PRESENTED TO LEARNERS

B: FULL SET OF SPEECH SAMPLE STIMULI
## LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>L2 Consonant Production Errors in Learner Speech by Sentence</td>
<td>74</td>
</tr>
<tr>
<td>3.2</td>
<td>Summary of Consonant Errors</td>
<td>74</td>
</tr>
<tr>
<td>4.1</td>
<td>NS Listener Accentedness Ratings for Learner Speech</td>
<td>89</td>
</tr>
<tr>
<td>4.2</td>
<td>One-Way Analysis of Variance of NS Likert-scale Accentedness Rating by NS Control</td>
<td>89</td>
</tr>
<tr>
<td>4.3</td>
<td>t-Test for the Difference between NS-Listener Accentedness Ratings for NS Controls</td>
<td>90</td>
</tr>
<tr>
<td>4.4</td>
<td>Likert Scale Attitudinal Data from Spanish NSs</td>
<td>91</td>
</tr>
<tr>
<td>4.5</td>
<td>Correlation Between Speech Sample Ratings and Attitudinal Data...</td>
<td>93</td>
</tr>
<tr>
<td>4.6</td>
<td>NS Listener Number of Friends who are NSs of AE</td>
<td>93</td>
</tr>
<tr>
<td>4.7</td>
<td>NS Willingness to Converse with Someone with a Strong AE Accent</td>
<td>94</td>
</tr>
<tr>
<td>4.8</td>
<td>Category Distributions for Free-Response Data</td>
<td>96</td>
</tr>
<tr>
<td>4.9</td>
<td>Examples of Token Distribution among Categories</td>
<td>98</td>
</tr>
<tr>
<td>4.10</td>
<td>Chi-Square Results for Free-Response Observations of Accented Consonants</td>
<td>103</td>
</tr>
<tr>
<td>5.1</td>
<td>Chi-Square Results for Free-Response Observations of Accented Consonants, Separated into Four Categories</td>
<td>121</td>
</tr>
</tbody>
</table>
### LIST OF FIGURES

<table>
<thead>
<tr>
<th>Number</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Spectrogram for SS04 (Male Learner Low) “ella limpia” (with Formants)</td>
<td>6</td>
</tr>
<tr>
<td>3.2</td>
<td>Spectrogram for NS05 (NS Male) “limpia su casa” (with Formants)</td>
<td>72</td>
</tr>
<tr>
<td>3.3</td>
<td>Spectrogram for NS10 (NS Female) “está limpia” (with Formants)</td>
<td>72</td>
</tr>
<tr>
<td>3.4</td>
<td>IP Address Map for NS Evaluators</td>
<td>80</td>
</tr>
<tr>
<td>3.5</td>
<td>Screenshot from Online Survey</td>
<td>83</td>
</tr>
<tr>
<td>3.6</td>
<td>Screenshot of Attitudinal Questionnaire</td>
<td>86</td>
</tr>
<tr>
<td>5.1</td>
<td>Hierarchy of Saliency of Non-Target Segmental Consonantal Productions</td>
<td>116</td>
</tr>
</tbody>
</table>
CHAPTER 1

Second-Language Phonological Acquisition and the “Learnability” Debate

This introductory chapter provides a comprehensive review of several nuanced aspects of the current study’s independent variable of L2 learner phonological production. In this investigation extemporaneous learner speech served as the basis from which NS listeners developed their perceptions of L2 learner phonological production. Since learner production played a pivotal role in this research, it was imperative to examine some of its major characteristics. The identification of individual components of and trends within the learners’ speech stimuli facilitate comparison with patterns emergent in the NS evaluations of L2 speech production. Put simply, an analysis of the NS-listener perceptual data in isolation could answer the question of how the stimuli were perceived but would not provide complete insight as to why they were perceived in that way.

Though the perceived saliency of certain characteristics within learner speech is due in part to factors relating to the NS listener (affective factors, L1), the composition of the speech sample itself undoubtedly plays a role in shaping listener perception. Factors such as L1 transfer and age-related L2 phonological developmental limitations have important implications for understanding (a) the distribution and frequency of specific non-target articulatory patterns in learner speech and (b) the potential for pedagogical intervention. The following pages discuss past and present scholarly understanding of the ability that adults have to learn, retain and reproduce a range of phonological aspects of a second language.
Historical Background

Over the course of the last several decades, various theories have emerged with the common goal of deconstructing the processes through which adult learners acquire (or fail to acquire) L2 phonology. Though almost half of a century has passed since the publication of the first seminal articles on this topic (Lenneberg, 1967; Asher & García, 1969), many of the questions raised in this body of research remain points of controversy within the field. Some of the most central and persistent questions are as follows:

1. Why is it that the great majority of adults who take up learning an L2 have end-state L2 phonologies that fall short of being native or near-native?

2. What role, if any, does pedagogical intervention have in L2 phonological acquisition (L2PA)?

3. Is native-like L2 pronunciation a realistic and/or achievable goal for learners?

4. Is there a greater theoretical framework, such as typological markedness or phonological natural class, that can account for research that indicates that some aspects of second-language phonology (L2P) are more difficult to acquire than others?

The first issue, which explores the apparent inability of adult learners to achieve native-like end-state L2 phonologies, has long been intertwined with the notion of a “critical period” for language acquisition. The Critical Period Hypothesis (CPH)\(^1\) proposes that certain physiological changes that occur within the human brain at the onset of puberty, such as lateralization and the resulting loss of plasticity, severely limit the post-pubescent learner from achieving native or near-native acquisition of many of the
linguistic elements of an L2 (Lenneberg, 1967; Johnson and Newport, 1989). Lenneberg first alluded to the notion of a CPH in 1967, when he stated that:

Adults ‘inevitably’ speak foreign languages with an accent if L2 learning begins after childhood, because the ability to learn new forms of pronunciation is inhibited as the result of the ‘firm structuring’ of neural processes through cerebral lateralization. (p. 377)

L2PA in particular has been singled out by several research studies and theorists as being especially vulnerable to a critical period (Asher & García, 1969; Oyama, 1982; Tahta, Wood and Loenthal, 1981). This research suggests that, after the onset of puberty, inherent phonological abilities in an L2 deteriorate before other linguistic abilities, such as syntactic or lexical competence. According to the CHP, native-like L2 phonology is an unattainable goal for the adult language learner.

While the notion of a critical period (CP) gained prominence during the 1970s and 1980s within the young discipline of second-language acquisition (SLA), its parent field of general linguistics was undergoing a paradigm shift that would affect SLA and L2PA theory for decades to come. Previously, language acquisition theory and pedagogy had been heavily influenced by behavioral psychology theories prevalent in the 1950s and 1960s. Behaviorism posited that language was a learned behavior, an idea that manifested itself within SLA via the popular language memorization and “drilling” pedagogies of the era. These methodologies actually afforded L2PA a place of prominence in that pronunciation exercises composed many of the classroom L2 “drills.” However, the behaviorist tenants for language learning were soon to be directly challenged by an emergent theory from within theoretical linguistics. Noam Chomsky’s
groundbreaking theory of Universal Grammar (UG) (Chomsky, 1965, 1966) would influence the direction of scholarship in linguistics - and later in SLA - for years to come.

One of Chomsky’s most notable early affronts to behaviorist accounts of language acquisition is found in his critique of behaviorist B.F. Skinner’s *Verbal behavior* (1957), in which Chomsky argues that learners do not simply memorize and repeat the utterances they hear, as posited by the behaviorist perspective (Chomsky, 1959). According to the UG paradigm, humans have a biological predisposition for language acquisition, whose “parameters” eventually set to the distinct rules of the native language’s (L1) grammar. This theory accounted for the fact that children (a) produce utterances that they have not previously heard and (b) overgeneralize rules to produce non-target-like utterances such as “Mommy goed to the store.” Chomsky argued that children are very unlikely to have memorized such utterances from the speech of others, thus making language acquisition a process that cannot be explained as a strictly learned behavior. UG theory postulates that language acquisition occurs as input and verbal exchanges in the L1 shape the parameters of one’s UG until they are perfectly formed to suit the distinct linguistic rules of the native language. (Chomsky, 1966)

The intense influence of Chomsky’s theory of UG in the field of SLA, along with the developing notion of a critical period for language learning, had strong negative implications for the behaviorist theories of language learning and their accompanying pedagogical methodology of drills and repetition exercises. In behaviorism’s wake, new “communicative” methodologies would come to dominate in the foreign language classroom. These approaches, such as Hymes’s (1971) “communicative language teaching” or Krashen and Terrell’s (1983) “natural approach” had communication at their
core, and maintained that learners would naturally acquire other aspects of the L2, such as pronunciation, as they progressed. Learners were encouraged to use their L2 in meaningful and creative ways, in situations that simulated real communication instead of rote memorization and drill exercises. The new communicative approach has been widely accepted within SLA pedagogy - a fact easily attested to by any brief review of currently successful Spanish language textbooks such as *Dos Mundos* (Terrell, Andrade, Egasse & Muñoz, 2010) or *Gente* (Martín Peris, Sans Baulenas & Cabello, 2006). Communicative language teaching, however, proved problematic for L2PA, in that it deemphasized explicit phonetic and phonological instruction in the L2. As put by Lafford and Salaberry (2003):

> With the advent of more communicative approaches to language learning and instruction in the 1980s, explicit instruction in target language pronunciation became a thing of the past and was generally limited to the preliminary chapters of language textbooks or found in the back pages of the textbook appendices. While it may appear that researchers had examined almost every facet of language acquisition in relation to the Communicative Approach, the acquisition of pronunciation fell to the wayside and consequently suffered from serious neglect in the communicative classroom. (p. 32)

The shrinking presence of L2 phonological instruction in the foreign language classroom under communicative instructional approaches allowed for an important juxtaposition between Chomsky’s UG theory and the CPH. Chomsky’s UG theory posits that human beings have an innate biological predisposition toward all aspects of language acquisition, citing said predisposition as the reason as to why virtually all children
demonstrate perfect acquisition of their L1. Evidence supporting the CPH, however, suggests that the brain develops in such a way that disallows this type of “perfect” learning after a certain point in the individual’s lifetime, a phenomena most noticeable in (“imperfect”) adult L2 phonological acquisition. This incompatibility between the two theories lends itself to a series of questions regarding the full application of UG theory to the processes of SLA. Some of the most pressing questions are as follows: Is access to UG deactivated at a certain stage of development? What is to be made of L2 learners that do in fact show gains in their phonological acquisition, or for that matter the small number of post-pubescent L2 learners that do achieve native-like L2 phonological competence?

These questions motivated a resurgence of interest in the possibility of adult L2 phonological acquisition that began in the 1990s and has been steadily gaining momentum. This movement represented a pendulum swing of sorts, following two decades of focus on communicative language teaching methods that devoted little to no classroom time on phonological training. The learners were just not “picking up” L2 pronunciation as a result of being in the communicative classroom, and the SLA research community began to take note. A variety of studies were conducted on the effects of pedagogical intervention on L2PA.

The following sections of this chapter examine in greater detail the theory of UG and CPH as they relate to L2PA and provide an overview of more recent work that has explored the effects of explicit pedagogical intervention via phonological training. The review presented here does not claim to be exhaustive but rather illustrates several
seminal studies and ensuing trends that have influenced the way L2PA is seen both within the realm of the foreign language classroom and the field of SLA.

**Early Views on L2 Pronunciation and the Effects of Pronunciation Instruction**

Once the behaviorist paradigm for L2 acquisition was overshadowed by more communicative methods, the presence of any type of phonological instruction in many foreign language classrooms was relegated to marginal importance, if not eliminated entirely. As stated by Derwing and Munro (2009):

The decline of audiolingualism led to a concomitant marginalization of pronunciation research and teaching. It was believed that pronunciation instruction could not be effective, in part because of the unrealistic goal of native-like speech in L2 learners, and also because of research findings that suggested that instruction had a negligible impact on oral production. (p. 476)

Several major studies provided empirical evidence to support the arguments that (a) L2 phonology cannot be acquired without great difficulty and many resulting imperfections, and (b) even if phonological instruction were implemented in the classroom, it would not lead to acquisition.

Perhaps the most influential early study to argue the former of these two arguments is Asher and García’s (1969) investigation on “The optimal age to learn a foreign language.” The impact of this study on SLA has been undeniable, with modern scholars such as Piske, MacKay, and Flege (2000) marking this work as the beginning of the investigation into foreign accent. Asher and García tape-recorded 71 Cuban immigrants between the ages of seven and nineteen as they read sentences aloud in English. The recordings were intermixed with recordings of children born in the United
States reading similar sentences. English native-speaker (NS) evaluators were then asked to listen to the recordings and indicate the degree of accent that they perceived for each speaker. The ratings were correlated with several variables, with age of arrival (AOA) to the United States being found to have the strongest correlation with pronunciation ability. Asher and García stated that “the implication is that if a child was under six when he came to the United States he has the highest probability of acquiring a near-native pronunciation of English and if the child was older than 13 he had the lowest chance of near-native speech.” (p. 336)

In general, Asher and García found that none of the Cuban children had achieved native competency in their L2 phonology, though many of those who had arrived in the United States at younger ages were reported as having native-like L2 pronunciation. These findings complemented the work of other scholars who were also investigating the AOA question. Two years before the Asher and García study, Lenneberg (1967) had theorized that there existed a correlation between pronunciation and hemispheric lateralization occurring at the onset of puberty. Scovel (1969) also blamed foreign accent on the loss of brain plasticity.

Several years later another study would support the notion of the “unlearnability” of L2 phonology, this time from a pedagogical standpoint. Suter’s (1976) study measured the English pronunciation of learners from various L1 backgrounds who had reported on a language history questionnaire that they had received pronunciation instruction in English. A panel of NSs of English evaluated the pronunciation of these learners along with the pronunciation of learners who reported that they had not received any phonological training. No correlation was found between pronunciation instruction and
L2 phonological competence. This study “has been widely cited as showing the futility of pronunciation instruction.” (Pennington, 1998, p. 325)

The results of the Suter (1976) study have, however, been questioned. Piske, McKay and Flege (2001) point out that the participants in the study were not matched for AOA, length of residency (LOR) or amount of L2 use. The lack of data on these variables, especially AOA, is problematic since it has significantly correlated with L2P ability in other influential L2P studies (Asher & García, 1969; Snow & Hoefnagel-Hoehle, 1978; Oyama, 1982; Johnson & Newport, 1989). In a more lengthy criticism of this work, Pennington (1998) argues that the study is fundamentally flawed because there was no investigation into the type of phonological instruction the students had received and no verification of their self-reports. In addition, the evaluators did not have an agreed-upon list of criteria for determining pronunciation accuracy, leading Pennington to speculate that they very well could have mistaken overall fluency for accuracy (in certain L2 aspects of phonology).

A curious aspect of these grievances is that they are from researchers who conducted their work decades after the study’s original publication. A possible explanation as to why the notion of the “futility” of second-language pronunciation instruction remained relatively uncontested for so long within SLA may lie within Lenneberg’s theory of a critical period for language learning, as cited in Asher and García (1969), which argued that the window for phonological acquisition closed at six years of age or earlier. With this theoretical backing, phonological instruction had dismal prospects for positively affecting learner’s L2 phonological competence. The learners in
Suter (1976) could not have possibly experienced gains from L2 phonological instruction: their brains were already developmentally closed off to such a possibility.

The notion of a CP for SLA did not, however, close the door to L2PA research. Over the past four decades, SLA research has delved into the question of a CP, at times with varying results. The next section will provide a comprehensive summary of what have been considered some of the most pivotal developments in the theory of a CP within SLA, with special emphasis on its relation to the notion of L2PA.

**CPH, SLA and L2 Phonology**

In *Fossilization in Adult Second Language Acquisition*, Han (2004) asserts that “the relevance of CPH to SLA was first discovered in the realm of phonology” (p. 49), citing Scovel (1969) as the first researcher to draw connections between developmental hemispheric brain dominance and the ensuing decrease in one’s ability to acquire certain linguistic elements in a native-like way. According to Scovel, “practice cannot make perfect what nature has already made permanent.” (1988, p. 159)

Following Scovel’s theoretical contributions to the notion of a critical period for language acquisition, various empirical studies investigated the feasibility of an application of this theory to SLA. Over the course of an eleven-month-long longitudinal study, Snow and Hoefnagel-Höhle (1978) found that adult NSs of English learning Dutch in an immersion context initially showed an advantage over younger learners in the pronunciation of Dutch lexemes but were later “passed out” by the younger learners: After a period of just eleven months, the younger learners began to receive higher scores from the Dutch NS evaluators. The adult learners were never able to “catch up” to the younger learners; once they had been surpassed in performance they were unable to
outperform them in post-tests. The results of this study have been employed to argue that, though adults may initially show greater competency than children due to their increased abilities to reason and think critically, children demonstrate an innate ability to acquire aspects of an L2 that quickly surpass the adults’ greater capacity for critical thinking. This reiterated a pattern noticed by Krashen, Long and Scarcella (1979), who stated in their review of the literature of ultimate L2 attainment that:

Adults and older children in general initially acquire the second language faster than young children (older-is-better for rate of acquisition), but child second language acquirers will usually be superior in terms of ultimate attainment (younger-is-better in the long run). (p. 574)

Another study by Oyama (1976) supported the notion of a critical period for SLA. In the study 60 NSs of Italian were divided into two independent variables: age of arrival (AOA) and length of residency (LOR). The subjects were taped reading in their L2 (English), and their pronunciations were scored by NSs of the L2. AOA was found to have a strong correlation with phonological competence but LOR was not, leading Oyama to arrive at the conclusion that “a sensitive period exists for the acquisition of a non-native phonological system.” (p. 261)

The case made by the CPH for incomplete L2 phonology acquisition was not, however, without counterexamples. A study by Ioup, Boustagui, Tigi and Moselle (1994) examined the language learning process of an individual learner named Julie. The subject was an English woman who, while in her twenties, moved to her spouse’s native country of Egypt, where she learned Egyptian Arabic (EA). Despite all predictions made by the CPH, Julie was successful in acquiring EA and was virtually undistinguishable from NSs,
even in the realm of pronunciation. This last fact is especially relevant, given the typological differences between the phonetic inventories of EA and English.

Research into cases like Julie’s is increasingly more common, and several studies have examined individuals whose L2 phonologies are recognized as “native-like” by NSs (Bongaerts, Planken and Schils, 1997; Moyer, 1999; Bongaerts, Mennen and van der Slik, 2000). Though the great majority of professionals in the field of SLA support the notion that language acquisition and especially L2 phonological acquisition is quite challenging for adults, it seems that nearly everyone happens to know a person who “has a knack for languages,” and many language practitioners and researchers themselves have achieved very high levels of proficiency in L2P. According to Selinker (1972), 5% of adult language learners will achieve what he calls “absolute success” in an L2.

What then is to be made of adult L2PA? In regards to the first question presented in this chapter, pertaining to the paucity of adult L2 learners that achieve native or native-like phonological competence in their L2, a critical period appears to be a possible if not likely culprit. However, the empirical evidence of counterexamples to the CPH demonstrates that the hypothesis is not absolute. Instead of believing that adult L2PA is impossible due to developmental limitations, the consensus within the field of SLA has shifted during the last several decades to allow for the possibility of adult L2PA, though it may indeed be severely limited by a critical period. According to Pennington (1998), “Although adults clearly cannot master a system of pronunciation in the (apparently) easy way children do, it is equally clear that they can make progress in their acquisition of the phonology of a second language over time.” (p. 323)
This shift to the recognition of the possibility of some degree of L2PA has in turn generated new ideas on the teaching of L2P in the L2 classroom. The following section examines some of the more recent studies that have been conducted in adult L2 phonological acquisition, focusing on the results of explicit L2 phonological instruction.

**L2 Phonological Instruction: Making Gains**

The possibilities for pronunciation instruction are numerous. An instructor may choose to focus on segmental phonological elements, such as a specific phoneme or minimal pair in the L2, or instead phonological instruction may be directed at suprasegmental factors (traditionally categorized as secondary features in phonology theory), such as intonation or pitch.\(^5\) In addition, instruction may be either implicit or explicit. The studies chosen here represent some of the most recognized and influential works in Spanish L2PA from the last two decades, with the majority of these studies focusing mainly on explicit pronunciation instruction conducted with Spanish as the L2.

One of the first major studies of this kind conducted with L2 learners of Spanish was carried out by Elliott (1995). Elliott tested the effects of explicit phonological instruction on 66 third-semester college Spanish students. Prior to treatment, participants completed a four-part pronunciation test in which they were evaluation in: (1) mimicking pronunciation of certain allophones at word level and (2) sentence level, (3) pronouncing allophones in written words, and (4) “overall accuracy in spontaneous communication” (p. 533). The students were then given formal instruction on Spanish pronunciation for the first 10 to 15 minutes of every class period for a span of 21 classes. After the treatment was complete, the four-part pronunciation evaluation was administered a second time:
The result was that the students instructed in phonology showed significant improvement in overall pronunciation.\(^6\)

González-Bueno’s (1998) study of the acquisition of Spanish stops also showed a correlation between formal instruction and improved pronunciation. In the study, 20 native English-speaking university students were given five to ten minutes of formal instruction on the pronunciation of the Spanish stops /p g b t d k/ over a period of 15 weeks. Recorded data from the participants was collected both before and after the treatment. The results showed that the subjects in the experimental groups showed a significant decrease in the voice onset time (VOT) for /p/ and /g/, consistent with native Spanish pronunciation. The improvement in the VOT for the other phonemes, however, was not significant.

Ganschow and Sparks’ (1995) yearlong study of the effects of Spanish pronunciation instruction on “at-risk” students also showed the positive effects of explicit phonological training. In this study “at-risk” learners were defined as students who performed at high levels in other academic areas but scored poorly on language aptitude assessments. The participants consisted of 14 “at-risk” women and 19 “not-at-risk” women. After a year of explicit instruction in phonology and orthography using a multisensory approach, the “at-risk” participants showed significant gains. (They did, however, continue to lag behind the “not-at-risk” control group.)

The conclusion that adult learners have the ability to progress in their L2P development is also supported by studies conducted on languages other than Spanish. Hardy (1993) found that Spanish NSs learning English showed significant L2P improvement after six months of training in minimal pairs in the L2.\(^7\) Volker (1996)
found that after fifteen weeks of training L1 AE speakers learning German had significantly improved L2 pronunciation. Derwing, Munro and Wiebe (1998) showed that ESL students made gains in both segmental and suprasegmental areas of English production following just eleven weeks of treatment.

In addition to the empirical studies that demonstrate adult learners’ ability to improve their L2 pronunciation, Levis’s (2005) work differentiating between the “nativeness principle” and the “intelligibility principle” is changing the conceptualization of successful adult L2PA. The “nativeness principle” states that successful L2PA results in pronunciation equivalent to that of a NS. The practice of using NS pronunciation as the objective for the language learner was one of the trademark characteristics of behaviorist methods of language instruction. Audiolingual language instruction methodologies gave L2PA a place of primacy, but L2P instruction was composed of repetition and mimicry drills. After behaviorist and audiolingual methodologies fell out of favor, the communicative language teaching (CLT) methods that replaced these paradigms did not directly challenge the nativeness principle in that CLT left L2PA largely ignored. Further complicating the matter was the fact that research on the CPH (referred to in the previous section of this chapter) supported the notion that native-like L2P was unrealistic if not impossible for adult language learners, thus seemingly supporting the notion that L2P should be left out of language instruction. As a result, explicit pronunciation instruction fell out of favor under communicative pedagogy. The result of these factors led to an impasse of sorts for L2P instruction: Explicit pronunciation was not considered a component of “communicative” language teaching, and CPH research describing the rarity of native-like L2PA made any effort in L2P training seem futile.
Levis (2005) correctly asserts that the nativeness principle is an unfortunate remnant of behaviorism that is no longer appropriate given current understanding of L2PA. Levis posits that the nativeness principle be replaced by the “intelligibility principle,” which holds effective, intelligible communication as the goal of L2PA. Levis’s work is based in part on research (Derwing & Munro, 1995; Derwing, Munro and Rossiter, 2003) that shows that learner speech rated by NS listeners as highly accented may also be rated by the same NS listeners as highly intelligible and/or comprehensible. The intelligibility principle recognizes that, although adult language learners can indeed make gains in L2P, native-like L2P is an unrealistic and unnecessary goal.

The introduction of the intelligibility model has been pivotal in shaping the current understanding of L2PA. This model acknowledges that the great majority of adult learners may very well not develop native-like L2 pronunciation. However, these learners can develop their L2P such that they achieve the more realistic objective of intelligibility. Since intelligible speech is a pillar of effective communication, L2P and L2P instruction are reemerging as areas of interest under the current communicative language-teaching methodologies.

**Models for L2 Phonological Instruction**

The emerging body of research on the positive effects of L2 phonological instruction for adults has allowed the concept of L2 phonological instruction to break free from its former role as one of the most neglected linguistic elements of SLA under the influence of the CPH and Communicative Language teaching. The belief that adult learners are unable to progress in the L2P development has been amended in light of
research that demonstrates that learners demonstrate some ability for this type of acquisition, with a small percentage of adult learners achieving high levels of competence in their L2 phonologies.

Once the effectiveness of L2 phonological instruction was established, a new debate arose within SLA regarding how to best implement L2 phonological instruction within the foreign language classroom. Some methodologies emphasize suprasegmental features (Morley, 1991) while others focus heavily on instruction at the segmental level (Fernández & de Santiago Guervós, 1997). Over the past fifteen years, suprasegmental instruction has been favored over instruction at the segmental level, though the dominance of suprasegmental-level instruction has not been accepted by all SLA researchers (Levis, 2005). In addition, there is evidence that learner segmental production may carry more perceptual weight than suprasegmental production in certain linguistic contexts, but that the inverse is true in different linguistic contexts (Riney, Takagi, & Inutsuka, 2005).

Further complicating this matter is the fact that, in the typical university classroom, instructors generally are required to abide by the guidelines of a prescriptive syllabus that dictates the topics covered within each class meeting and in most cases does not allow much time for L2 phonology (Elliott, 1995b). Even should L2 phonological training be integrated into the syllabus - a process that has been painfully slow at the level of the university foreign language classroom - there is a limited amount of time to be devoted to it. Thus, the role for a pedagogical model for teaching L2 phonology is invaluable.
One major caveat, however, is that an agreed-upon methodology for L2 phonological instruction does not exist within the field of SLA. There are differing theories regarding which phonological elements should be taught, arguing that certain aspects of L2 phonology are more pertinent than others for overall language acquisition. One of the current issues being explored is whether it is more beneficial to teach L2 phonology at the suprasegmental level or at the segmental level.

Derwing, Munro and Wiebe (1998) provide evidence to support what they deem a “broad framework” for pronunciation instruction, which favors instruction at the suprasegmental level. In their study ($n = 48$), two experimental groups of participants in English second language (ESL) classes were provided with two different types of pronunciation instruction, one focusing on a “broader” suprasegmental level, while the other group’s instruction was focused at the segmental level. A third control group did not receive any L2 phonological instruction. The authors state that “many materials and programs are primarily segment-based, that is, they focus on the perception and productions of individual phones.” (p. 395) Another flaw that the authors found with prior studies contrasting methods of L2 phonological training was the short duration of the instruction; Derwing, Munro and Wiebe rectify this in their study by providing both experimental groups with ten full weeks of explicit pronunciation training. The segmental group received a more traditional instruction based upon articulatory properties of the individual phones. In contrast, the suprasegmental group received instruction that introduced several novel elements, such as the use of Jazz Chants, which involved participants tapping the beat while repeating nonce words in the L2.
At the end of the instruction, all groups were evaluated by a panel of NSs on the measures of “accentedness, comprehensibility, and fluency.” Both experimental groups showed improvement over the control group, though of particular interest is the comparison of the participants’ results in each individual area evaluated. The participants were analyzed after performing two tasks, the first of which consisted of reading sentences aloud in the L2. Participants in the segmental group outscored the suprasegmental group for this portion of the evaluation. The second task, in which the participants produced extemporaneous narratives, showed improvement in comprehensibility and fluency only in the suprasegmental group. The authors propose that these results support a model for pronunciation instruction within a broad framework, i.e. focused on suprasegmentals, stating that learners were able to “transfer learning” of suprasegmental elements to spontaneous production, but the segmental group was not.

One should reexamine certain elements of the study, however, before accepting the conclusion that suprasegmental instruction is more beneficial than segmental instruction. First, the authors’ definition of comprehension is “a subjective assessment of ease or difficulty of comprehension as opposed to a measure of actual intelligibility.” (p. 396) It is quite possible that a read-aloud task and an extemporaneous speech task are inherently different, and that the latter has a higher “ease of comprehension.” It should also be noted that though the participants could very well avoid suprasegmental patterns that they have not yet mastered (thus creating the illusion of greater fluency), this type of avoidance is simply not possible in segmental phonetic production.

Arguments that favor segmental over suprasegmental instruction concern notions of intelligibility at the phonemic level. An error in a minimal pair is very likely to impair
intelligibility on the part of the listener, since the production of an incorrect phoneme alters the learner production at the semantic level. It is for this reason that the segmental properties that distinguish one phoneme from another are labeled primary phonetic features (such as voicing, sonority, etc.), while the suprasegmental features (intonation, pitch) are considered secondary, since they do not alter the semantic properties of the utterance.

Elliott (1995a), a proponent of segmental instruction, proposes a segmental model for L2 phonological instruction with the following components:

1. Teaching concrete rules about point, place, and manner of articulation
2. Providing students with ample drill and practice exercises
3. Giving immediate feedback in order to prevent phonological fossilization
4. Designing class presentations on pronunciation that appeal to all learning styles and preferences through aural input
5. Employing both deductive and inductive modes of teaching pronunciation of the L2

Elliott (1995a) employed this five-part model over the length of an academic semester in his study examining the success of formal instruction of Spanish pronunciation, as presented earlier in this chapter. This segmental approach was also utilized by González-Bueno (1997) in the previously cited study. These segmental models concentrate on the teaching of the production and perception at the phonemic level.
The present study does not advocate the teaching of segmentals over suprasegmentals, but rather takes the position that both areas should be represented in explicit L2P instruction. Segmental production and perception is the focus of the analyses presented here, but this is due to practical and not ideological motives. One of the overarching goals of this study is to investigate the possibility of a “hierarchy of difficulty” of Spanish sounds, from the perspectives of both (a) non-native speakers (NNS) learning these sounds and (b) NS listeners attempting to process learner speech. Here, segmental analyses are preferred due to the fact that notions of typological markedness and functional load (discussed in Chapter 2) are much more developed at the segmental level than at the suprasegmental level (Eckman, 1977; Castino, 1991; Derwing & Munro, 2008).

A logical question that arises is, “Why not just adopt Elliott’s model for all L2 segmental phonological instruction and teach as many L2 sounds as possible?” The answer is that while Elliott’s model is very effective at prescribing how to teach L2 phonology, it is not very specific in terms of describing exactly what to teach. Even if language instructors have the means to implement L2 phonology training in their classrooms, how are they to know which phonemes to concentrate on? Are certain phonemes more important to communication between NNSs and NSs than others? Moreover, is there a hierarchy of difficulty from the perspective of the learner, with certain sounds more challenging to acquire than others? A complete model of L2 phonological instruction needs to take these questions into account, prescribing not just how something should be taught, but also, what should be taught. Elliott’s model does assert that mode and place of articulation information must be included for the individual
phones and phonemes; however, it stops short of telling the instructor whether there should be emphasis placed on individual phonemes and whether certain phonemes will prove more difficult for learners.

The investigation at hand seeks to expand upon Elliott’s model for Spanish explicit pronunciation instruction by investigating which phones’ non-target productions are (a) most frequent within the data set presented and (b) most salient to NS listeners. The implication is that these phonemes/minimal pairs should be given particular attention within the structure of an explicit pronunciation instruction model, given their potential importance in transmitting semantic information and achieving communication. To arrive at an L2 “hierarchy of difficulty” for Spanish phones, two main issues must be taken into account:

(1) From the perspective of the learner, which sounds are most challenging from a production standpoint? 

(2) From the perspective of the NS, which sounds’ non-target productions are most salient to NS listeners?

Both of these questions are addressed in this current study. Though the learner speech samples serve as the independent variable, they are carefully analyzed in order to determine which L2 sounds have the most frequent occurrences on nonnative productions. It is also crucial to arrive at an understanding of the sounds whose nonnative productions are most salient to NS listeners. For example, a sound that is very difficult for a learner to acquire may not be very salient to the NS listener in comparison to other sounds, which may have implications for its pedagogical prioritization. A more detailed account of NS listener perception is presented in Chapter 2.
L2PA: Varying Degrees of Difficulty

One of the first attempts to create a hierarchy of phonological difficulty for learners of an L2 came from Stockwell, Bowen and Martin (1965), during the era when the contrastive analysis hypothesis (CAH) and behaviorism had a stronghold on L2PA. The CAH was based upon the principle that the degree of difficulty learners faced in L2PA could be accounted for solely by contrasting the presence and use of sounds in the L1 versus the L2. Stockwell et al.’s hierarchy contrasting the sounds of English and Spanish had eight levels of difficulty, separated by whether a phoneme was obligatory, optional, or absent in the L1 as compared with the L2. Phonemes that were considered to be the most difficult to acquire were those that were absent in the L1 but obligatory in the L2, such as the Spanish trill /r/ when presented to NSs of English. At the other end of the hierarchy, phonemes that were obligatory in both the L1 and the L2, such as nasals /n/ and /m/, were considered to have the least degree of difficulty.

However, Stockwell et al.’s hierarchy was strongly criticized. Kleinmann (1977) introduced what he called “novelty effect,” which directly contradicted Stockwell et al.’s theory. The novelty effect predicts that elements that are not present in the L1 but obligatory in the L2 will actually be easier for the learner to acquire, based on their saliency and the fact that learning is “the most difficult where the most subtle distinctions are required either between the target and native language, or within the target language.” (Oller & Ziahosseiny 1970, p. 127). The notion of a “novelty effect” has been supported by studies in both morphosyntax (Kleinman, 1977) and L2PA (Elliott, 1997). Kleinman’s (1977) study showed that L1 Arabic speakers were able to learn the progressive tense in English, even though it is not a component of Arabic morphosyntax. Elliott (1997) found
that it was more difficult for adult L1 English speakers learning Spanish to learn spirantized fricatives [β ð γ] than it was to learn trill [r], which would have a higher novelty effect for the L1 AE learners.

Another theory for classifying phonemes by potential difficulty for learners was introduced by Eckman (1977; 1995) as the Marked Difference Hypothesis (MDH). Eckman draws upon the theory of typological markedness to explain the order of acquisition of phonemes in the L2. Phonological markedness was first proposed by Trubetzkoy (1939) and Jakobson (1941), and is based upon the notion of binary or distinctive features, such as voiced versus voiceless obstruents or oral versus nasal vowels. Typological markedness posits that the two features of a binary pair are not equal in their distribution across languages. One of the features is considered “unmarked” and more frequently distributed across languages, while the “marked” feature is less frequently distributed across languages. Furthermore, the presence of the marked feature in a particular language implies the presence of its corresponding unmarked feature, but the presence of the unmarked feature does not necessarily imply that of the marked. Typological markedness, as summarized by Eckman (1977) states that “A phenomenon A in some language is more marked than B if the presence of A in a language implies the presence of B but the presence of B does not imply the presence of A.” (p. 320) For example, nasalized vowels and oral vowels reflect a binary feature contrast in which the nasal vowels are the more marked member of the pair. This means that languages that have nasal vowels will also have oral vowels. The phonological inventory of Portuguese, for example, contains both unmarked oral /a/ and marked nasal /ã/. In contrast,
Hawaiian’s phonological inventory contains the typologically unmarked oral /a/ but not the marked nasal /ā/.

Eckman (1977) applied the principles of typological markedness to SLA with the Marked Difference Hypothesis (MDH), which compares the relative degree of typological markedness in both the L1 and the target language in order to predict the relative difficulty that learners face when learning certain sounds in the L2. According to the MDH:

1. Those areas of the target language (TL) which differ from ones in the NL and are more marked than those in the native language will be difficult to acquire.
2. The relative degree of difficulty of acquisition of given areas of the TL that are more marked than parallel ones in the NL will correspond to the relative degree of markedness of each.
3. Those areas of the TL which are different from corresponding ones in the NL, but are not more marked than the latter will not be difficult to acquire. (p. 321)

Castino (1991) applied Eckman’s MDH to Spanish phonology to investigate the degree of difficulty learners faced while learning the marked phones [r x β δ γ] versus the unmarked phones [p t̪ k]. He deemed the Spanish phonemes /ɾ/, /x/ as being overtly marked and the allophones [β δ γ] as also being marked for NSs of American English (AE), since their existence implies the existence of the non-spirantized and therefore less marked [b d g]. Castino surmised that the phones [p t̪ k] [-aspiration] were less marked than the AE phones [pʰ tʰ kʰʰ] [+aspiration] and therefore should present a lower degree of difficulty for the NSs of AE. Castino then recorded and transcribed data for 40 third- and fourth-year university students, all NSs of AE, hypothesizing that their speech would
show more incorrect pronunciations of the marked forms than of the unmarked forms. The students were recorded both during a read-aloud task for a list of words and a narration task in which they were asked to invent stories for pictorial depictions of events. His hypothesis proved correct; students showed the highest levels of incorrect pronunciation for /r/, /x/ and [β ð γ] and much less error in their pronunciation of /p t k/.

The present study hypothesizes that, once learner errors are identified and calculated, there will be a strong correlation between degree of markedness of a particular segment and the instances of learners’ non-target-like productions of that particular segment. Drawing once more upon the example of the stops /b d g/ and their marked allophones [β ð γ], it is predicted that English L1 learners of Spanish will produce less target-like tokens of the more marked [β ð γ] than they will of the relatively less-marked [b d g]. These distinctions, if present, will support the development of a hierarchy of difficulty for learner production of Spanish sounds.

**Learner Data Elicitation**

A final consideration for research in L2PA is the question of which types of learner stimuli are most appropriate for evaluation. Munro and Derwing (1994) posed this question in terms of understanding the “‘best’ or ‘worst’ production that a talker was capable of.” (p. 254) Ideally, the NS would evaluate learner speech in real time or “online”, during an actual conversation. This method, however, poses several critical obstacles to an investigation. During a conversation between a learner and an NS, the learner may give visual cues that adequately convey meaning even though the spoken component of their message is unclear. It is also not feasible to “pause” a conversation in progress in order to ask the NS to transcribe a portion of it or give verbal feedback to an
objective third party – important investigational tools used by researchers. Since speech samples derived from on-line conversations often produce so many intervening variables, researchers frequently opt to have NSs evaluate prerecorded learner speech that has either been read from a text or produced extemporaneously (Johnson and Jenks 1994; Isaachs, 2008; Zielinski, 2008; Derwing and Munro, 1997, 2010).

The option of preparing a text passage for a speaker to read aloud and later be evaluated can appear advantageous since it is generally accepted that NSs produce fewer errors when reading in their L1 as opposed to extemporaneous production (Remez et al., 1991, Munro & Derwing, 1994). At first glance it seems reasonable to hypothesize this generalization for NSs may be transferable to learners reading in their L2, but the application of sociolinguistic principles to language acquisition research is problematic. Written texts often contain complex morphosyntactic and phonetic structures that are problematic for learners and thus lead to pauses or errors in production. As observed by Munro and Derwing (1994), SLA researchers will at times “load” a text with phonological or morphosyntactic structures that are known to be avoided by learners, such as trill /ɾ/ or past subjunctive in Spanish, as a last resort for collecting tokens of these structures. Munro and Derwing (1994) argue that this practice of “loading” a text with structures that are known to be difficult for the learner leads to artificially low NS evaluations when compared to evaluations of learners’ extemporaneous production.

Munro and Derwing (1994) examined NS evaluations of read versus extemporaneous learner speech, L1 speakers of Mandarin learning Canadian English were recorded as they invented a storyline to accompany a wordless comic strip (extemporaneous speech). These narrations were then transcribed; at a later date, each learner was again recorded as
they read the transcript of their own narrative aloud. English L1 speakers \( (n = 44) \) rated the accentedness of both the extemporaneous and read speech (in randomized order) on a nine-point Likert scale. The authors found that there was no significant difference in the NS ratings of accentedness for the extemporaneous and the read speech.\(^\text{10}\) Munro and Derwing thus concluded that stimuli from read samples are valid for “assessing pronunciation only if it can be established that the speakers are familiar with all vocabulary and grammatical constructions present in the material to be read.” (p. 257)

The speakers in their study had equal degrees of familiarity with both read and extemporaneously produced speech since they were reading transcripts of their own previously recorded speech.

Short of replicating Munro and Derwing’s read-aloud task for prerecorded extemporaneous speech, it would be exceedingly difficult to assure that learners have a similar degree of familiarity with all morphosyntactic and phonetic structures presented in a read-aloud task. For this reason much research into NS perception of learner speech is conducted with prerecorded stimuli of extemporaneous production. NS evaluation of extemporaneous learner speech is often at the paragraph level (Rubin, 1992; Johnson and Jenks, 1994; Kang, O., Rubin, D. & Pickering L., 2010; Cargile, Maeda, Rodriguez & Rich, 2010), or alternatively at the sentence or utterance level (Atechi, 2008; Zielinski, 2008), meaning that NS listeners hear an uninterrupted utterance, sentence or paragraph-length stimulus of prerecorded extemporaneous listener production. Sentence and utterance-level stimuli are often spliced from longer narrations, as in the case with Zielinski (2008).
The goal for the sample stimuli used in this study was to retain as many of the characteristics of on-line speech as possible and yet have the function of being analyzable in an objective context in which the NS evaluator could pause to annotate their observations. Thus, an extemporaneous speech elicitation model like the one used in Munro and Derwing’s (1994) research was chosen, and learners created a story from storyboard illustrations. Later, utterance-level recordings were spliced from the data and presented to the NS evaluators. The process of stimuli elicitation and collection and its preparation for presentation to the NS evaluators will be discussed further in Chapter 3.

**Conclusion and Research Questions**

Historically, the notion of L2 phonological acquisition has been viewed from a variety of theoretical perspectives with varying implications for L2 phonological instruction’s inclusion in the foreign- or second-language classroom. As the initial belief that L2PA was physiologically impossible was gradually replaced by the understanding that adult learners can in fact make progress in phonological acquisition, L2 phonological instruction has enjoyed a resurgence of interest in SLA research. However, much of this research has failed to make its way into current practice, i.e. the majority of Spanish textbooks or syllabi. It is the goal of this study to continue the work of scholars like Elliott and contribute the process of developing a comprehensible methodology for L2 phonological instruction in the foreign language classroom by answering the question of what needs to be taught. This question will be answered by taking into account (a) the segmental phonological elements of Spanish that are most often produced in non-target ways by NSs of English, and (b) the segmental non-target learner productions that are more salient to NSs of Spanish (Chapter 2).
CHAPTER 2

Native-Speaker Perception of Nonnative L2 Phonology

The last chapter explored what is known as the “learnability” problem of L2 phonology, addressing the fact that L2 language learners often fall short of native or near-native phonological acquisition, and the effects of pedagogical intervention on learners’ L2 phonological development. Though a thorough understanding of learnability issues is fundamental to the study of L2 phonology acquisition, the study of these issues alone leads to an incomplete conception of the L2P acquisition process in that the sole focus is on the language learner. Communication does not consist of the learner producing L2 speech in a vacuum, but rather includes the NS interlocutor as an integral component of the communication process. This chapter shifts the focus from learner production and investigates NS perception of learner phonology. The two main questions explored in this section are as follows:

(1) Which factors determine whether or not an NS finds learner speech intelligible and/or accented?

(2) What role does phonology play in NS perception of learner speech?

(3) Is there a theoretical model, such as typological markedness or the functional load principle, that can predict the segmental components of learner speech whose non-native productions will be salient to NS listeners?

The idea that the role of the “hearer” is at times neglected in speech studies in not novel. In 1921, Walter Peterson stated:

That language is a social product, which presupposes both the speaker and the hearer, is a linguistic commonplace, the acceptance of which in theory is self-
evident. Nevertheless it is equally true that in discussions of individual problems and in explanations of individual linguistic phenomena the hearer is almost completely ignored even at the present time. (p. 16)

In the near century that has transpired since Peterson’s remarks, interest in the role of the hearer has been generated in the fields of socio- and psycholinguistics. However, in-depth research into the NS listener has not been widely conducted within the relatively young linguistic sub-discipline of SLA, and the vast majority of prominent research agendas within SLA pertain in some way to the learner. Some examples include promising types of pedagogical methods such as task-based instruction (Lee, 2000; Long and Norris 2009), sociocultural theory and learner development (Lantolf & Thorne, 2006; Lantolf & Beckett, 2009), and the effect of individual differences on acquisition (Robinson, 2001; Dörnyei & Skehan, 2003, Winke, 2005). When NS listeners are included in SLA research, they often are cast in the role of interlocutor in order to produce an interaction with the learner that will be recorded and later used as data. The analyses of these interactions are generally learner-focused and include notions of scaffolding\textsuperscript{11}, communication breakdown, repair, and communication strategies, (Varonis & Gass, 1985; Mackey & Silver, 2005; Gass & Mackey, 2006). Alternatively, NS listeners are frequently recruited to rate learner performance in a certain grammatical aspect before and after a particular treatment has been applied. This second method of incorporation of the NS listener also focuses on the learner, relegating the NS listener to the role of measurement instrument.

The objective here, though, is not to denounce the learner-centric research agenda within SLA. On the contrary it is the learner who is “acquiring” language and thus merits
a place of primacy within the research. However, many studies that consult NS listeners for Likert-scale judgments and nothing more forfeit two research opportunities. First, there are affective variables such as prejudices or in-group affiliations that may color NS perception of accented learner speech and influence the way in which results are interpreted (Rubin, 1992; Cargile & Giles, 1997). Second, the NS listener is a potential source of valuable information on the specific components of learner speech behind the Likert-scale ratings. A small but growing number of studies have already begun to examine this potential by deconstructing NS evaluations of learner speech by means of more in-depth interviews with and questions for NS listeners (Zielinski, 2008; Derwing & Munro, 2009). If the overall goal of communication is to remain in focus, it is imperative to arrive at a more nuanced understanding of NS reactions to and perceptions of learner output. This chapter will examine research that has been conducted in NS perception of learner output, with a concentration on the perception of the phonological aspects of said output.

**NNS Perception: the Phonology Problem**

Dating back to the early 1990s, a small number of studies have investigated NS perception for its own sake. The majority of these recent studies concentrate on NS perception of learner phonology. In his study on NS perception of NNS varieties of English, Atechi (2008) states, “the majority of researchers agree that phonology stands out as the most troublesome area with regard to matters of intelligibility.” (p. 262). This exploration of NS perception of learner phonological output first manifested itself in research of NS reactions to nonnative “accent,” and many researchers have since asserted that foreign accent is “an especially salient aspect of learner speech.” (Derwing & Munro,
2009a, p. 476; Scovel, 1988). According to Johnson and Jenks (1994), the bulk of the early “accent studies” that focus on the hearer have tended to concentrate on the area of perception of nonnative accent (Carranza & Ryan, 1975; Ryan, Carranza & Moffie, 1977; Brennan & Brennan, 1981), many from a sociolinguistic or attitudinal basis.

Though research in perception of L2 accent is quite relevant to socio- and psycholinguistics, the results generated by this branch of research can prove challenging to interpret within SLA, given that the notion of “accent” is an umbrella term for several phonological phenomena that occur simultaneously. The multidimensionality of the notion of “foreign accent” can be illustrated by a hypothetical example of a learner production at the phrasal level that includes several common non-target productions. For example, when a beginner-level English L1 learner of Spanish attempts to convey the semantic message that “the wine is expensive,” it is unlikely that their message will have the same phonetic content as that of an NS of Spanish, who would produce [el.bi.noes.ká.ro] (“el vino es caro”) or some regional variation thereof. A common production of a beginner-level learner could be [lɔ.ɪ.vi.now.es.kʰɔ.ɹow]. The learner’s utterance diverges from that of the NS at several linguistic levels, as discussed below.

Morphosyntactically, the feminine singular definite article “la” is used instead of the masculine singular definite article “el.” Several segmental differences are present: the labiodental fricative [v] replaces the bilabial stop [b], most of the vowels have been diphthongized, and the single-tap [ɾ] is replaced by the alveolar approximant [ɹ]. The use of [ɹ] is especially problematic for a Spanish NS listener in that /ɹ/, since it is not a phoneme in Spanish, has the potential to be understood as either single-tap /ɾ/ as in caro (expensive) or trill /ɾ/ as in carro (car). In addition, the increased voice-onset time in the
learner production of /k/ is phonetically represented as [kʰ] instead of the native production [k].

The hypothetical learner production differs from a target production suprasegmentally as well. The contrast of stress-timing in the learner’s L1 English to the syllable-timing of Spanish will likely cause undue syllable elongation and misplaced stress. Also, elongation of the final /o/ sounds of vino will further compound the perceived foreign accent in this utterance. Unlike English, Spanish does not have long vowels, so elongation of tense Spanish vowels can cause misplaced stress.

A final “sub-segmental” or phonational difference could also be heard between learner production and an NS target-like production should the learner use “creaky voice,” which refers to irregular vibration of the vocal folds several octaves lower than in normal speech. Creaky voice is prevalent in many dialects of American English but not used in Spanish.

As evidenced by the example of the very brief utterance presented above, perceived foreign accent is certainly not a one-dimensional issue but rather a general term for a myriad of linguistic processes that occur within speech. When the various elements of nonnative phonology are studied as individual components, it is readily apparent that “accent” is not always a sufficient label to give to the aggregate whole of all of these individual processes. Moreover, it should not be readily assumed that all of these linguistic elements are equally weighted in their contribution to what a NS listener hears as “accentedness.” One of the major goals of the present study is to arrive at a more nuanced understanding of the individual components of learner speech that most heavily contribute to perceived accentedness. The next section of this chapter reviews several
empirical studies that have investigated the respective weighting of individual linguistic elements in the perception of foreign accent.

Studies on NS Perception of Individual Components of NNS Phonological Production

Johnson and Jenks (1994) confronted the issue of “ambiguous accent,” referring to the multiple components of perceived foreign accent in learner speech and the inappropriateness of the umbrella term “accent.” They proposed that the hearer’s perception be divided into four main categories, as follows:

1. **segmental** - phonemes and minimal pairs, such as the English approximant /ɹ/ when uttered in place of the Spanish multiple tap /ɾ/
2. **suprasegmental** - prosody and intonation, such as the implementation of English stress-timing when Spanish syllable-timing is required in the target language
3. **subsegmental** - voice quality, such as using “creaky-voice” in a language that typically does not have this linguistic element
4. **grammar** - morphosyntactic errors, such as an article of incorrect gender used in a language that requires noun-article gender agreement

Johnson and Jenks’ model marked an important step in the division of the ambiguous notion of “accent” into measurable components and promoted inquiry into which individual components of nonnative phonological production have the greatest impact on NS perceptions of foreign accent and intelligibility in learner speech. Johnson and Jenks (1994) used their model of NS perception as the basis for an investigation into the perception of two of its components: segmentals and grammar. The goal of the study was
to research whether learner errors in segmental phonological production or learner grammatical errors had the greater impact on NS perception of foreign accent.

In their study three nonnative speakers were recorded speaking in English. These three participants were NSs of Spanish, German, and Arabic, respectively, but all could also speak English “without a detectable accent.” (p. 10) As a preliminary step, the researchers studied representative Speaking Proficiency English Assessment Kit (SPEAK) samples from test takers with the same L1s as the recorded speakers. The SPEAK test is an oral evaluation of proficiency in English developed by the Educational Testing Service (ETS) and is often used in undergraduate and graduate entrance exams and in testing of international teaching assistants. The researchers identified common phonological and grammatical errors of lower-level learners in order to instruct the recorded speakers to intentionally commit these types of errors in their recorded speech samples. The speakers then each read a paragraph out loud, as is done in the SPEAK test. A first recording from each speaker was carefully controlled to contain only segmental errors, while a second recording of the same paragraph contained only grammatical (and no segmental) errors. Recordings did not contain suprasegmental or sub-segmental errors, the other areas listed by Johnson and Jenks in their original model. A fourth NS of English served as a control, reading a paragraph out loud in English with no phonological or grammatical errors.

NSs of English were recruited from three local universities to evaluate the recordings, with the evaluator population controlled to represent the various ethnic groups living in the area. The participants were asked to listen to the recordings and fill out a questionnaire inquiring how successful they thought the individual speaker in the
recording would be at the university. The participants were not told that each bilingual speaker recorded two samples; rather, they believed that they were listening to seven recordings from seven different speakers (each of the three bilingual speakers had one segmentally- and one grammatically-flawed recording, while the monolingual English speaker produced the control recording).

As expected, the results showed that the recording of the English NS received the highest scores. The scores for the Spanish and German speakers were comparable to each other, and both scored higher than the Arabic speaker. The results also showed an interesting outcome: the scores for the samples with grammatical errors were significantly higher than those for the samples with segmental errors. This fact is especially relevant, since the samples with segmental errors were controlled to meet the so-called “threshold of comprehensibility,” meaning that they did not excessively inhibit comprehension (Anderson-Hsieh, J., Johnson, R., & Koehler, K., 1992), whereas the samples with grammatical errors were not: “the phonetic errors did not hinder comprehension and some of the grammatical errors did.” (Johnson & Jenks, 1994, p. 24)

From these results the authors conclude that NSs as hearers dislike segmental phonetic inaccuracy more than they do grammatical errors. The implications for this conclusion underscore the importance of phonology in NS perception of NNS speech: The listeners in this study were less forgiving of segmental phonological errors, even when these errors did not necessarily lead to communication breakdowns. This investigation has value in that it identifies segmental phonetics as one of the key problem areas for NS-learner communication. However, one should be cautious not to overgeneralize that all segmental errors are more salient than all grammatical errors,
especially since the relative salience of the individual segmental and grammatical errors was not one of the research foci of their study. Indeed, it is possible that the segmental errors in all of the sample stimuli were especially salient, thus causing them to be perceived more frequently and with greater intensity than grammar errors that were only moderately salient.

Another, more recent study conducted by Isaachs (2008) also attempts to arrive at a better understanding of the aggregate parts of learner phonology, this time from the perspective of intelligibility. In the investigation 19 college-student raters, all NSs of English, completed a complex evaluation process in which they listened to a recorded Test of Spoken English (TSE) for eight international teaching assistants (ITAs) of varying L1s. First, the raters were instructed to evaluate each speaker in terms of perceived overall intelligibility and indicate whether or not they believed that each speaker should teach a college-level course. The raters were then instructed to check off boxes to indicate which of six possible pronunciation features “hindered” their ability to understand each speaker. Last, the raters were instructed to draw a box (in the transcript) around three features of the learner speech that “most hindered” their understanding of said speech. Not surprisingly, results showed that the speakers with the highest intelligibility ratings across the three tasks were most often labeled as competent to teach college courses and vice-versa.

Several patterns emerged in the results regarding the pronunciation features that most hindered the raters’ understanding of learner speech. For the task of checking boxes to indicate which pronunciation features proved most problematic to the listeners, a frequency count of the results showed the following ranking: (1) individual sounds
(segmentals), (2) speech clarity (mumbling or “over-pronouncing” words), (3) word stress, (4) sentence rhythm, (5) rate of speech and (6) pitch. For the task in which raters were instructed to draw a box around and rank three individual pronunciation features that were problematic for their understanding of the speaker, the following rank-order of these six categories was produced: (1) individual sounds, (2) word stress, (3) speech clarity, (4) sentence rhythm, (5) rate of speech and (6) pitch. Distribution of NS ratings of these categories for each individual speaker, however, showed that the NS ratings varied by speaker. For three of the eight speakers, the NS listeners most often identified pronunciation features categorized as “speech clarity” (defined as “mumbling” or “over pronouncing” words) as most hindering to their understanding of the speech samples. Two of the eight speakers’ most problematic pronunciation features pertained to “individual sounds,” two to “word stress” and one to “rate of speech.” This distribution of categories among the speakers is most likely accounted for by the fact that they were from a variety of L1 backgrounds, thus having distinct difficulties with English pronunciation dependent upon their L1. The suprasegmental categories of “word stress” and “speech clarity” were consistently listed by the NS listeners as highly problematic, whereas “rate of speech” and “pitch” were the least frequently cited categories.

Though the differing individual results from speakers of different L1s indicate that results regarding individual categories of problematic pronunciation features need be interpreted with caution, several patterns emerge from the results. For both tasks in which NS raters were asked to choose the features that hindered their understanding of the speakers, the frequency count results showed the category of “individual sounds” to be most problematic. Isaachs includes in this category the subcategories of “substituting
sounds” and “adding/deleting sounds.” Thus, in the aggregate data for this study, segmental issues were most prevalent in NS listeners’ indications of individual aspects of learner production that proved problematic. Also highly problematic in the frequency-count results for both tasks were the suprasegmental categories of word stress and speech clarity, ranking second and third to the segmental features. Thus, segmental and certain suprasegmental components of the learner speech signal were identified by the NS listeners as hindering their understanding of learner speech. The categories of rate of speech and pitch were least often identified by the NS listeners in both tasks.

Isaachs (2008) illustrates important progress in the task of discerning which non-native phonological productions in the learner speech signal are most salient to the NS listener. This study is especially valuable in providing insight into which suprasegmental areas most contribute to NS perceptions of unintelligibility, since several of the categories used - word stress, speech clarity, and pitch - represent various suprasegmental components of speech. However, the fact that the speech sample stimuli were collected from learners of various L1s might be a confounder; it is possible that learners with different L1 backgrounds produce suprasegmental errors in the L2 with varying frequencies dependent on their L1. For example, it is possible that the group of learners recorded here was composed of several individuals with a shared L1 that had a high propensity for committing word-stress errors in the L2, and it is that propensity that is reflected in the high frequency of NS observations of word-stress errors.

Similar to Johnson and Jenks (1994), this study does not discern between individual segmentals beyond the subcategories of “substituting sounds” and “adding/deleting sounds.” This is most likely due to the fact that the speech sample
stimuli were gathered from learners of various L1 backgrounds, and such an analysis would be inappropriate in that these learners’ segmental errors would almost undoubtedly differ in type and frequency depending upon their L1. In order to effectively investigate the relative saliency of certain segmentals, the speech stimuli would need to be collected from learners with a common L1 in order to ensure that a certain feature did not appear especially salient due to differences in learners’ L1s.

Another attempt at deciphering the various components that contribute to perception of learner pronunciation was carried out by Atechi (2008). As in the Isaachs (2008) study, Atechi used the notion of intelligibility as a point of departure for his research, measuring the perceived intelligibility between three groups of English-speakers. Two of these groups, British English (BrE) and American English (AE) were labeled by the author as NSs of English. A third group of speakers of Cameroonian English (CamE) were labeled as NNSs. Various recordings were collected from the participants of each group: monologues of extemporaneous speech (participants talking about their daily lives and routines), readings of various texts, and the readings of minimal pairs. After an initial phase in which the participants were recorded, each group was then asked to listen to the recordings from the two other groups and identify areas of the recordings in which they failed to understand the speaker.

The areas measured were divided by the researcher into the following three groups: (1) segmental, (2) suprasegmental and (3) phonotactic. Atechi lists examples of segmental differences such as the monophthongization of native English diphthongs, vowel shortening, final consonant devoicing and yod deletion. Examples of suprasegmental differences include differences in the nucleus placement of words and
differences in rhythm and/or intonation. Finally, differences in elision, linking and assimilation are categorized as phonotactic differences.

When the NSs of AE and BrE listened to the recordings of the nonnative CamE speakers, he found the following breakdown among the lead causes of unintelligibility: 50.5% suprasegmental, 39.1% phonotactic and 10.3% segmental. NS perception of NNS speech also showed a trend towards suprasegmentals as causing the greatest number of communication breakdowns. Atechi uses these results to arrive at the conclusion that since suprasegmental phonological elements have the greatest correlation with occurrences of unintelligibility, these particular phonological components merit the most pedagogical attention.

However, there are several reasons to question such a conclusion. One of the main concerns with this study can be seen in the design itself, particularly in the categorization of phonological elements into the three categories of suprasegmentals, segmentals and phonotactics. While the use of segmentals and suprasegmentals as separate categories is hardly controversial, the addition of the phonotactics category is somewhat questionable. The original model from Johnson and Jenks did not include a separate grouping for phonotactics, perhaps with good reason. Phonotactics is described as “the sequential arrangements (or tactic behaviour) of phonological units which occur in a language - what counts as a phonologically well-formed word” (Crystal, 2008, p. 366). Phonotactics refers to phenomena such as the types of consonant clusters permitted in a language. A good example of a phonotactically motivated phenomenon is seen in the English word-initial combination of the alveolar fricative /s/ and a voiceless stop /t/ or /p/ (“study,” “Spanish”) that is permissible in English but, due to phonotactic constraints, not in
Spanish, where this fricative-stop combination must be preceded by a vowel, such as in *estudiar* or *español*.

The processes that Atechi labels as phonotactic - linking, elision and assimilation - occur frequently at the segmental level. This is not to say that “phonotactic” is itself an invalid category; the issue lies in its use as a mutually exclusive category alongside separate categories of “segmentals” and “suprasegmentals.” Though it can apply to one or various sounds, assimilation is also segmental in nature. For example, in the common English expression “Don’t be rude,” when spoken at a conversational pace, the /t/ of “don’t” is deleted and /n/ assimilates in place of articulation to the following /b/ , changing the expression to /dom.be/. Thus, the “phonotactic” differences that Atechi notes are also segmental phonological changes motivated by phonotactic constraints and could be more aptly categorized under “segmental” differences. If the categories are altered in this way, the results show that, from the perspective of the NS, 50.5% of the errors that led to unintelligibility were suprasegmental in nature, while 49.4% were segmental. Also complicating the comparison between suprasegmental and segmental features is the fact, stated by Atechi himself that, “with regard to the pronunciation of consonant phonemes, it should be emphasized here that the consonant charts of native and non-native varieties of English are basically the same,” thus mitigating the frequency of possible segmental differences in this study. (p. 277)

Though the argument for the predominance of suprasegmentals versus segmentals for promoting intelligibility is indeed problematic, the value in this study can be found in the fact that like Isaachs (2008), Atechi (2008) demonstrates (a) progress in the study of NS perception of individual components of learner L2P and (b) the fact that both
segmental and suprasegmental phonological features play a central role in arriving at the overall goal of intelligibility in nonnative speech.

One of the most intricate and well-known investigations into NS perception of the individual components of L2P was conducted by Zielinski (2008). According to Munro and Derwing (2011), “Zielinski was the first researcher to analyse individual listeners’ understanding of accented speech by determining which aspects of the speech stream they favoured.” (p. 327) Zielinski echoed the sentiment of Peterson (1921) that the listener was often overlooked in linguistic studies, and titled her investigation, “The listener: No longer the silent partner in reduced intelligibility.” In this exploratory study, speech sample stimuli were collected from three NNSs of English, whose L1s were Korean, Mandarin, and Vietnamese, respectively. The researcher then edited the recorded samples so that they contained only utterances that she believed would be problematic for English-NS listeners. Then, three NSs of Australian English evaluated the recordings. Each NS listener participated in three different evaluation sessions, for a total of nine sessions for all NS evaluators. The researcher was present during all sessions and asked follow-up questions at each session. During these sessions the researcher identified what were labeled as “sites of reduced intelligibility,” which were defined as “any pause group containing a section of speech where a listener was unable to, or had difficulty in identifying the speaker’s intended words.” (p. 73) Sites of reduced intelligibility were identified via a listener’s (1) actual transcription of the speech sample, (2) behavior while transcribing, including facial expressions and/or laughing, and (3) comments. A total of 402 responses were analyzed for number and strength of syllables and segments produced at the sites of reduced intelligibility. The results showed that NS listeners were most often
misled by both misplaced syllable stress and non-standard production of segmentals. Careful study of the phonetic transcription of the learner productions and the NS listeners’ transcriptions showed that the NS listeners relied heavily on segmental production and syllable stress patterns in strong syllables when transcribing. Syllable-initially consonants were labeled as particularly important, due to the fact that they were preserved in 75% of the NS-listener transcriptions.

Zielinski’s (2008) work signaled several key advancements in the study of NS perception of learner speech. Zielinski employed the work-intensive technique of meeting with each NS evaluator over multiple sessions and collected data not only from the NS transcriptions but also from information gathered based on NSs’ behavioral cues and the information provided by their individual commentaries. This holistic approach to analyzing NS-listener data marks an important development in the conceptualization of the NS listener as not solely a source of numerical-scale ratings but as a complex individual whose evaluations should be gathered via multiple collection techniques, both quantitative and qualitative. Additionally, Zielinski does not attempt to portray segmentals and suprasegmentals as competitors for attention in pedagogy or research. Rather, this research depicts both segmentals and suprasegmentals as complementary components of NS perception of learner L2P and asserts that it is imperative that both continue to remain foci of research in this area. Like Isaachs (2008) and Atechi (2008), Zielinski’s study also analyzed data from learners of different L1s. However, this does not appear to have been problematic in Zielinski’s investigation, since the finding of NS-listener reliance on stress and segmental production in strong syllables was present in transcriptions of NS-listener transcriptions of all three of the NNS learners.
Though the number of empirical studies conducted in this line of research is relatively small, these research efforts indicate a growing interest in Zielinski’s “silent partner.” The study presented in this paper expands upon this body of research in several key ways. In contrast to the prior studies reviewed here, the present study uses speech sample stimuli from learners with a common L1. This will eliminate the possibility that learner errors in the speech sample stimuli are distributed unevenly due to L1 differences. In addition, the current study investigates the possibility that segmental components of the learner speech stimuli have varying perceptual “weighting,” meaning that certain non-target segmental productions will be more salient to NS listeners than others. Finally, this study distinguishes itself from much of the previous research on this topic in that it proposes to draw upon theoretical perspectives in order to arrive at a “hierarchy of segmental perceptual difficulty” to NS listeners of Spanish as they process speech from L1 English speakers learning Spanish. In order to achieve this last objective, both the Marked Difference Hypothesis (Eckman, 1977) and Derwing and Munro’s application of the Functional Load Principle (Derwing and Munro, 2008; Catford, 1987) are investigated as possible frameworks from which to base a hierarchy of perceptual difficulty.

The next section of this chapter focuses on NS-listener perception of the segmental aspects of learner phonology. As stated in Chapter 1, this focus on segmental elements of L2P is in no way meant to show a preference for segmentals over suprasegmentals. Rather, segmental elements of the speech signal were chosen because the theoretical approaches of typological markedness and functional load in SLA are
much more robust at the segmental level than at the suprasegmental level (Eckman, 1977; Castino, 1991; Derwing & Munro, 2008).

**Semantic Value of Segmental Features**

A useful source of information on NS perception of learner speech at the segmental level can be found in interactionist studies that investigate how communication breakdowns occur and the repair tactics used by speakers. Though these types of studies tend to focus heavily on the learner, the transcripts of the conversations examined can be helpful in investigating NS perception. Though not the main focus of their study, the dialogues presented in Varonis and Gass (1985) offer valuable insight into NS perception of learner speech, especially at the segmental phonological level.

One of the NS-learner conversations presented in the Varonis and Gass study is especially useful for illustrating just how segmental phonological errors by the learner can drastically alter the semantic message and thus the entire conversation. The following is an exchange between an L2 speaker of English and his NS instructor.

Luis: I want to ask you about somethin'

NS [teacher]: Yes, Luis, what is it?

Luis: I don't understand about urinals.

NS: What is it you don't understand about them?

Luis: I don't understand what you're supposed to do with them. (p. 330)

According to Gass and Varonis:

Luis, a NNS of English had come to see his English teacher with a problem: He didn't know what he was supposed to do with urinals. It was only after some time into the conversation that the teacher (a woman)
realized that he was talking about the journal assignment that she had made. (p. 330)

Luis’s unfortunate segmental phonological error is clear: he substituted the voiced on-glide /j/ for the voiceless postalveolar affricate /dʒ/. The result was that the learner’s intended message underwent a drastic semantic change.

A second example from the Varonis and Gass conversation data underscores the possible effects of segmentally-induced communication breakdowns in NS-learner interactions. This example is in the context of Sally and Akihiro, a couple that is in a romantic relationship. Sally is a NS of English and Akihiro is not. In this case, the unintelligibility led to an alteration in some very significant semantic material:

Akihiro: I want to make a life with you.

Sally: [understood - 'I want to make love with you.']. (p. 330-31)

Varonis and Gass explained “in this example, a correction was made a full year later. The male, a nonnative speaker of English, uttered what he intended as a proposal of marriage: ‘I want to make a life with you.’ Sally understood: ‘I want to make love with you.’” (p. 331) Though the misunderstanding was resolved and the couple was eventually married, it is easy to see how this communication breakdown could have been disastrous. Akihiro most likely produced something along the lines of [meikaraifu] or [meikulaifutu] for make a life with, and it appears that Sally perceived [ai] as [ʌ] and either /f/ or /ɸ/ as /v/. In addition, it is probable that Sally perceived Akihiro’s production of the indefinite article a as the epenthetic vowel [ɯ], which is routinely used by L1 speakers of Japanese. Thus, this seems to be a case of both erroneous production and perception.
Though the examples found in the Varonis and Gass data are quite colloquial and certainly comical, they underscore the practical importance of segmental features within conversation: non-target segmental production can drastically alter the semantic content of a message. Several of the components of this present study that focus on segmental production are motivated by this semantic weight carried by segmental features. It should be understood that this stance is in no way intended as a position that posits segmental features as superior to suprasegmentals. Indeed, both are imperative for effective communication in an L2.

**Weighting Perception of Segmentals in Learner Speech**

The **functional load principle**. The learners’ non-target L2P productions in the examples presented above were undoubtedly quite salient to the NS listeners. However, not all non-target L2P productions result in serious communication breakdowns, and it is reasonable to hypothesize that some non-target L2 segmental productions are more salient than others to the NS listener. Munro and Derwing (2006) investigated the possibility of NS segmental perceptual “weighting” using the theoretical framework of the functional load principle (FL) (Meyerstein, 1970; Catford, 1987; Brown, 1991), focusing in particular on Catford (1987) and Smith’s (1991) applications of the FL principle to ESL. According to Catford, “the functional load of a phoneme or phonemic contrast is represented by the number of words in which it occurs in the lexicon, or in the case of a phonemic contrast, the number of pairs of words in the lexicon that it serves to keep distinct.” (p. 88) For example, Catford presents the phonemic contrast between /i/ and /ɪ/ ("sheep"/"ship") as having a higher FL than the phonemic contrast between /u/ and /ʊ/ ("fool"/"full") for two reasons. First, according to Catford, the /i/ - /ɪ/ minimal
pair serves to distinguish more words in English than the /u/ - /ʊ/ pair. Second, the /u/ - /ʊ/ distinction is not maintained in some native dialects of English, while the /i/ - /ɪ/ distinction is maintained in almost all native dialects of English. Catford (1987) used the principle of FL to devise a hierarchy of FL for minimal pairs in English, in which final consonants, initial consonants, and vowel pairs were given a score between 1-100 to represent their FL in English. Shortly thereafter, Brown (1991) proposed a similar hierarchical rating based on functional load, and used a percentage-based instead of a point-based system to rank sound pairings.

Though the FL principle was a promising theoretical basis for study in L2P and NS perception of L2P, Munro and Derwing (2006) pointed out that it had not been tested empirically. They sought to apply the principle of FL, designing an exploratory study to investigate the ways in which FL errors impacted NS perceptions of nonnative speech. In their study twenty-three samples of Cantonese-accented English speech of differing functional loads were evaluated by NSs for both accentedness and comprehensibility. In order to create these speech sample stimuli, a database of 2,300 utterances was examined and the most frequently made consonantal substitutions (consonantal minimal pairs) were determined. The authors used the ranking systems from Catford (1987) and Brown (1991) to determine the relative FL of the most frequent substitutions and assigned each substitution to one of two FL categories. The substitutions deemed as high-FL were /l/-/n/, /ʃ/-/s/, /n/-/l/, /s/-/ʃ/ and /d/-/z/, while the substitutions deemed as low FL were /ð/-/d/ and /θ/-/f/. Twenty-three individual speech samples from nineteen different speakers were chosen to represent all of the possible minimal pair substitutions identified above. The sentences chosen contained eight possible combinations with varying degrees of high and
low FL errors: sentences had between zero and three high FL load errors or between zero and two low FL errors (high and low FL errors only appeared together in one combination). The sentences all used high-frequency words and did not include any overt subphonemic or intonation irregularities. English NS evaluators then completed a listening task in which they heard each individual sentence and rated it on two different nine-point Likert scales - one for accent and one for comprehensibility.

Results showed a statistically significant relationship between the number of high FL errors in a sentence and accentedness ratings. There was also a “cumulative effect” for accentedness ratings, meaning that as the number of high FL errors in a sentence increased, the degree of perceived accent also increased significantly. The comprehensibility ratings, however, did not show the same pattern as the accentedness ratings; the reported comprehensibility of sentences with more low FL errors was not significantly different from that of sentences with less low FL errors. The authors concluded that the results supported the notion that “listeners were clearly sensitive to high functional load errors…however, these same errors had a smaller, non-cumulative impact on their comprehensibility ratings.” (p. 529-53)

This exploratory research by Munro and Derwing (2006) was extremely valuable in showing that (a) not all segmental production errors contribute equally to NS listeners’ perception of accentedness and (b) NS perception of accentedness may be more sensitive to segmental errors than NS perception of comprehensibility. This current study considers the FL principle as a viable explanation as to why certain nonnative productions of Spanish phones could be more salient than others to NS listeners. For example, a finding that the NS evaluators in this study indicated that substitutions of single-tap /ɾ/ for trilled
/ɾ/ were more salient to them then substitution of [d] for [ð] would in part support the FL principle. This is due to the fact that in Spanish the /ɾ/ - /ɾ/ substitution carries a much higher FL in that it constitutes a phonemic difference with possible semantic implications, while the [d] - [ð] substitution carries a lower functional load since it is not phonemic but rather represents the failure to use the correct allophone of a phoneme.

This current study differs from the one conducted by Munro and Derwing (2006) in that in this study the NS listeners are instructed to provide Likert-scale accentedness ratings, as well as indicate the individual sounds from the speech samples that most contributed to their ratings. This difference reflects one of the main goals of this study, which is to arrive at a hierarchical categorization of the sounds or categories of sounds that most contribute to NS listeners’ perceptions of accent in learner speech. Such a hierarchy may need to be more in-depth than a separation of sounds into categories of high and low FL, and thus the FL principle is considered with the notion of typological markedness (see following sections) in the analysis presented. Another difference between the present study and Munro and Derwing (2006) is that this study focuses on accentedness ratings and does not include ratings of comprehensibility. This difference is due to the fact that in pilot studies of the speech-sample stimuli used, NS listeners reported very high to perfect levels of comprehension.

**Markedness considerations.** Another possible theoretical framework that could be used to interpret NS-listener perceptions of segmental elements of learner speech is that of typological markedness. Though markedness has been proposed as a possible theoretical framework for predicting and explaining errors in learner production (Eckman, 1977, 1995; Castino, 1990), it has not received much consideration as a viable
theoretical framework for interpreting NS-listener perceptions of non-target productions in learner speech. In the study presented here, the data provided by the NS listeners are analyzed, and the segmental sounds that are most frequently identified as highly accented are examined in order to determine possible patterns of typological markedness. Like Munro and Derwing’s (2006) study that investigated the viability of FL as a theoretical framework, this study is exploratory in nature. Two of the possibilities for NS segmental data and typological markedness categorization are as follows:

**Salience of unmarked segments.** Since unmarked segments are considered more prevalent crosslinguistically, it is a strong possibility that the NS listeners will be especially sensitive to non-target productions of unmarked segmentals. NSs may anticipate non-native speakers to produce non-target pronunciations of linguistically marked sounds such as trill /r/, but not unmarked sounds such as certain nasals or stops. Examples of common non-target productions of unmarked Spanish segments by L1 AE speakers included in Castino (1991) include the productions [pʰ tʰ kʰ] as over-aspirated, non-target productions of [p t̪ k].

**Salience of marked segments.** A second possibility for the NS segmental data is that non-target productions of marked segments will be more salient to the NS listeners, since marked sounds are by nature more salient. According to Castino (1991), L1 AE speakers have more difficulty producing Spanish segmentals that are considered more marked, such as trill /r/ and [β δ й] (See Chapter 1).

**Affective Variables and the NS Listener**

A final consideration for the NS listener has to do with affective or attitudinal variables involved in the act of processing learner speech. Just as the speech signal is not
homogenous, the NS listener must also be understood as complex and dynamic. NS listeners’ perceptions, while certainly modified by the individual properties of learner speech, can also be colored by these affective variables. Indeed, language-attitudes research is a productive area of investigation with a strong tradition that spans several decades and disciplines (Cargile, Maeda, Rodriguez & Rich, 2010). One of the most famous studies on affective variables in NS perception conducted by Rubin (1992) implied that simply identifying a speaker as “foreign” could have a detrimental effect on NS listener comprehension.

In Rubin’s (1992) seminal study, 62 North American college students, all NSs of AE, evaluated nonnative speech from an attitudinal perspective. Each participant listened to one of two prepared recordings, both approximately four minutes in length and both taken from articles in The New York Times. One was representative of the sciences and described helium scarcity, while the other, representative of the humanities, discussed the Mahabarata in Indian society. While participants listened to their assigned recording, they were (randomly) shown one of two photos. One photo was of a Caucasian woman and the other of an Asian woman. The women were matched for dress, attractiveness, and were posed in similar settings. What the participants did not know was that the woman they saw in the photograph was not the speaker they heard in their recording. All recordings were of single speaker: she was “a doctoral student in speech communication, a native speaker of English raised in central Ohio, and was well regarded by her own undergraduate students for especially effective and clear classroom delivery.” (Rubin, 1992, p. 515)
After hearing the lecture, participants completed a cloze test for comprehension (filling in every seventh word of the transcript) with only exact recall considered correct. Participants also completed a “homophily instrument.” Homophily, from Greek and meaning literally “love of the same,” is defined within the context of this research as “a construct that pertains to listeners’ perceptions of similarity to message sources.” (Rubin, 1992, p. 513) The instrument itself was comprised of 16 items that were part of four categories: (1) attitude homophily, (2) background homophily, (3) value homophily, and (4) appearance homophily. For example, on the “values homophily” instrument, participants indicated on a scale to what degree they felt that the speaker shared their values. In addition to the homophily items, participants were presented with “additional semantic differential items,” which included questions regarding the speaker’s ethnicity, level of foreign accent, and effectiveness as a teacher.

In the results an analysis of variance (ANOVA) test with the dependent variables as the cloze test scores, the four homophily sub-scales, perceptions of accent, ratings of teaching competence, and a manipulation check scale showed that instructor ethnicity emerged as the only main effect that was statistically significant. Then, a stepwise discrimination analysis showed that the comprehension test score was the first variable to enter: “using comprehension test scores alone, 61.2% of participants could be correctly classified into groups (in this case the grouping factor being whether they viewed an Asian instructor or a Caucasian instructor).” (p. 518) Degree of perceived accent entered at the second step of the analysis, when “an additional 9.8% of participants were correctly classified into treatment groups when perceived accent was added to the model.” (p. 518) In other words, just knowing which picture was shown to a participant (the Caucasian or
the Asian woman) would allow for one to predict with a relatively high degree of accuracy that participant’s (1) comprehension score on the cloze test and (2) accentedness ratings for their speaker. It should be reiterated that all participants heard the same speaker (the NS of AE).

The results of this study garnered national attention and were even featured in the national newspaper *USA Today* (Carroll, 1993). The implications are striking: perceived foreignness is correlated with an actual decrease in comprehension. According to Rubin, “Evidence from the discriminant analysis suggests that participants stereotypically attributed accent differences - differences that did not exist in truth - to the instructors' speech. Yet more serious, listening comprehension appeared to be undermined simply by identifying (visually) the instructor as Asian.” (p. 519)

Another study on language attitudes by Cargile, Maeda, Rodriguez and Rich (2010) provided insight into NS perceptions of speech that had a wide variety of accents. For this investigation 65 college students at a large urban North American university listened to audio recordings from 14 different male speakers who read the same text in English. These recordings were taken from “The Speech Accent Archive,” a large database of phonetically transcribed recordings of speakers with different accents reading the same paragraph in English (Weinberger, 2011). The recordings chosen for this study were from speakers of seven different L1s: Spanish, German, Italian, Mandarin, Hindi, Vietnamese, and (Midwestern) American English (AE). Two speakers from each L1 were used, one was of a speaker with a higher proficiency in English and the other was of a speaker with a lower English proficiency. The higher-proficiency recordings were matched for rate of speech and LOR across accents, and the same was done for the lower-
proficiency recordings. Two recordings from NSs of AE served as a control. Raters completed a 12-scale inventory for items regarding their impressions of the speakers that included ratings for status, attractiveness and foreignness. The raters were also asked about the strength of each speaker’s accent and instructed to assign an L1 to each speaker. Unlike the Rubin (1992) study, these NS raters were not provided with photographs of the speakers, and all perceptions were based solely on the audio cues.

A single, repeated-measures MANOVA was conducted to analyze the effect of speaker accent, and a significant main effect was found. Univariate tests showed significant differences within ratings for foreignness, status and attractiveness. Results showed a distinct ranking for perceived degree of “foreignness” for the speech samples: Vietnamese and Mandarin speakers were considered the most “foreign,” while Spanish, Hindi and Italian speakers were considered moderately “foreign.” German speakers were ranked as the least “foreign,” second only to the NSs of English. The ranking for perceived status closely resembled that of perceived foreignness. The NSs of English were rated as the most status-possessing, followed by the Germans. Hindi and Mandarin speakers were ranked third for status, while the Vietnamese and Spanish speakers were rated as the least status-possessing. Results for the Italian speakers were omitted from the results due to the fact that the raters consistently mislabeled them as L1 speakers of Spanish, perhaps a result of the raters’ lack of familiarity with Italian-accented speech. Though the published tables of the univariate results for the attractiveness variable had several statistically significant results, the authors stated that, “what emerged here was largely a pattern of few (and little) differences,” since the mean differences were small across all speakers. (p. 68) The authors conclude that several factors external to the
speech signal itself may indeed have contributed to NS listeners’ perceptions of accent, including perceived foreignness and status of the speaker.

Similar to findings from the Rubin (1992) study, these results indicate that there is a high correlation between an NS listener’s identification of a speaker as “foreign” and the NS’s perceptions of the speaker’s degree of accent. This correlation is seen whether said accent is indeed present, as in Cargile, Maeda, Rodriguez and Rich (2010), or completely imagined by the NS listener, as in Rubin (2010). Cargile, Maeda, Rodriguez and Rich (2010) speculate that higher ratings of foreignness “could lead listeners to hear a heavier accent and/or to provide lower status evaluations.” (p.73-74)

This current investigation draws from these developments in research on affective variables in NS perception of foreign accent in order to achieve a more comprehensive understanding of that perception. The measurement tools used to collect data on NS perception of the speech signal are coupled with questionnaires that elicit data regarding NS attitudinal variables. In the case of this study, NSs of Spanish are asked about their perceptions of NSs of AE learning Spanish, as well as about their perceptions of the United States in general. This attitudinal data is compared with numerical speech sample ratings of accentedness in order to see if there is any correlation between accentedness ratings and attitudinal variables, as was the case in Cargile, Maeda, Rodriguez and Rich (2010).

A final concern regarding affective factors and NS perception of nonnative speech deals with the NS’s affiliation to a particular group. According to Cargile and Giles (1997), an individual’s affiliation to a group may become more or less salient dependent upon the situational context. For example, an NS of Castilian Spanish may not normally
strongly identify herself as such, but that identity may change drastically if she is in a study-abroad context in a South American country where the majority of the population is composed of NSs of Latin American Spanish. Cargile and Giles (1997) state that research has shown that individuals “with robust in-group identities are more likely to dislike and derogate outgroup members…” (p. 199) Thus, when presented with accented speech, it is quite possible that an NS’s awareness of his or her in-group identity is heightened and the “foreignness” of the nonnative speech is made more salient. As evidenced by Rubin (1992) and Cargile, Maeda, Rodriguez and Rich (2010), heightened perceptions of “foreignness” often correlate with stronger perceptions of accent and even decreased comprehension.

For these reasons the attitudinal questionnaires presented to the NS evaluators in this study also inquired into the number of friends the NSs have who are NSs of AE. It is posited that Spanish-NS participants who have a high number of friends who are NSs of AE will be less likely to see NSs of AE as members of an “outgroup” and, according to Cargile and Giles (1997), will be less likely to “dislike” them. In addition, a high number of AE-speaking friends could indicate that a Spanish NS-listener participant associates a lower degree of “foreignness” to these nonnative speakers. Since lower levels of perceived foreignness have corresponded with decreased perception of strength of foreign accent (Cargile, Maeda, Rodriguez & Rich, 2010), it is reasonable to hypothesize that Spanish NSs with high numbers of L1 AE-speaking friends could show a tendency to rate NSs of AE as having overall less-accented L2 speech.

**Summary of NS Perception and Research Questions**
The review of the literature presented in this chapter illustrates the dynamic and multifaceted nature of the study of NS perception of nonnative speech. Though this field of inquiry has often been overshadowed by learner-centric focus within SLA, a growing number of scholars and researchers are taking interest in the role played by the NS listener as an integral component of the communication process. The research presented in this chapter can be summarized into the following conclusions:

1. “Ambiguous accent” as an anachronism. NS perception of NNS phonology has many individual components, and as Johnson and Jenks (1994) have shown, it is no longer appropriate in SLA research to label these all under the “ambiguous” term “accent.” Just as “groceries” can describe a broad list of comestibles from steak to popcorn, “accent” refers to any combination of linguistic features of nonnative speech that differ from target productions. Components of perceived accent include a variety of linguistic features, such as morphosyntactic, segmental and suprasegmental elements.

2. Segmental versus suprasegmental features. There is disagreement as to which individual component of NNS phonology has the greatest potential to contribute to the occurrence of communication breakdowns. However, it has been established that both segmental and suprasegmental phonological elements are an integral part of achieving successful NS-learner communication. The studies featured in this chapter do not show one of these areas within phonology to greatly overshadow the other in terms of effect on NS listener perception - a fact that attests to the importance of both.
3. **Semantic segmentals.** Segmental phonological elements, when uttered in a non-target-like manner, have the capacity to change the semantic value of the learner’s message as perceived by the NS. The effects of such semantic alterations could be quite detrimental to the relationship between these two parties.

4. **Weighted segmentals.** Recent research in L2P (Munro & Derwing, 2006) on the Functional Load Principle (Catford, 1987) has indicated that certain segmental productions are more prone than others to pose difficulty to the NS listener. The “weighting” of these segments in terms of their salience to the NS listener and potential for negatively affecting communication depends on specific properties of each segmental, such as frequency of a minimal pair in the target language. Munro and Derwing (2006) point out there is a paucity of research in this area, a sentiment that is seconded by others in the field (Zielinski, 2008).

5. **Attitudinal adjustments.** A holistic understanding of the NS listener as an individual whose L1 background and in-group identity affect their perception of learner speech is paramount to the study of NS perception. The studies presented in this chapter demonstrate the power of affective variables. All research in NS perception of nonnative speech should include data on affective variables for the NS listener(s) studied. Such information can be invaluable when interpreting results.

The present study focuses on points (4) and (5), while simultaneously considering the other points presented above. The current investigation begins to answer the call made by Munro and Derwing (2006) and Zielinski (2008) for more “fine-grained” research into NS perception of segmental elements of learner phonology. In addition to
an in-depth analysis of Spanish NS perception of individual segmental components of learner speech, this study also includes data on NS affective variables. In this way, this research presents a detailed analysis of Spanish NS-perception of learner speech from two complementary perspectives: (a) linguistic (segmental) variables internal to the speech sample and (b) affective variables that, though internal to the NS listener, are external to the speech sample itself. To the author’s knowledge, this is the first study of its type that has been conducted on NS-listener perception of Spanish L2P.
CHAPTER 3

Methodology and Data Collection

The points presented in the summary in Chapter 2 were taken into careful consideration when developing the research questions for this current study’s dependent variable of Spanish NS perception of learner speech. This study confronts the notion of “ambiguous accent” by asking the question:

1. “Which aspects of Spanish L2 learner extemporaneous speech do NSs most often identify as accented?”

This study expands upon the current body of research in several key ways. The first major differentiating factor is that in this research the speech-sample stimuli are all gathered from L2-learners of Spanish who share the common L1 of AE. In previous studies (Johnson & Jenks, 1994; Isaachs, 2008; Zielinski, 2008) the use of learner speech-sample stimuli from learners of varying L1s limited or confounded the degree to which segmental data could be analyzed, as reviewed in Chapter 2. This study looks to the model presented in the exploratory study of segmental weighting by Munro and Derwing (2006) and controls for L1 in the learner speech that is to be evaluated. Thus, it is proposed that differences observed by the NS evaluators are due to differences between (a) the segmental elements themselves and/or (b) affective factors internal to the NS listeners, since there is no difference in the learners’ L1.

Additionally, this study is unique in that it presents L1 speakers of English as “the nonnative speaker.” The great majority of studies conducted on NS perception of learner speech consist of NSs of English evaluating the speech of ESL learners (Johnson & Jenks,
A second way in which this current investigation builds upon the existing body of research is via its treatment of segmentals. Previous studies have been limited to the labeling of various phonological components of foreign accent at the segmental and suprasegmental levels (Johnson & Jenks, 1994; Zielinski, 2008; Isaachs, 2008). Conducting the study with nonnative speakers of the same L1 allows for an analysis of the individual phones whose non-target productions are most often identified as “accented” by NSs. Thus, the phones identified can be studied in order to look for patterns based upon such theoretical notions as markedness or functional load (see Chapters 1 & 2). This leads to the second research question for the dependent variable of NS perception:

2. Is there a hierarchical order in Spanish that ranks sounds in terms of their potential for contributing to NS perceptions of accent? Would this possible order reflect an established theoretical framework, such as the notion of markedness or functional load?

Again, this focus on segmental phonology is not meant to show a preference for segmentals over suprasegmentals. Such a debate is outside the scope of this investigation. Instead, this research takes advantage of the opportunity to conduct an in-depth investigation into the role played by segmentals in NS perception of nonnative speech. In addition to their semantic value, segmentals lend themselves much more easily to an analysis based upon FL or markedness, whereas such an analysis with suprasegmental features would be far less reliable, if not impossible.
Lastly, the current study carefully considers the role of affective variables in NS perception, asking the question:

3. “How do NS attitudinal variables correlate with the way in which they evaluate accentedness of learner speech, if at all?”

As evidenced by this review of the literature, linguistic and affective factors in past research into NS perception of nonnative speech have been mutually exclusive; this study will consider both. NS perceptions of learner speech will not be the only dependent variables in this study. The NS participants will also provide information regarding their views of NSs of AE and of NSs of AE learning Spanish in particular, which will prove invaluable for appropriately interpreting results.

The current investigation addresses the research questions presented above by means of a mixed-methods study in which Spanish NS listeners \((n = 18)\) evaluated the pre-recorded utterance-level speech of L1 AE speakers learning Spanish. The NSs provided numerical Likert-scale ratings of accentedness and also completed a survey that asked Likert-scale questions regarding their perceptions of L1 speakers of AE. The Likert-scale accentedness ratings were correlated with the Likert-scale data on the affective variables in order to see if there was a correlation between attitude toward the learners and the NSs’ accentedness ratings. In addition, the NS evaluators answered free-response questions in which they described the individual non-target productions from each of the specific speech samples that they perceived as most salient. The NS-evaluator responses focusing on individual segments were then analyzed within the theoretical principles of both FL and typological markedness, in keeping with the goal of arriving at
a hierarchical order in Spanish that ranks sounds in terms of their potential for contributing to NS perceptions of accent in L1 AE learner speech.

**Phase I: Collection and Analysis of Nonnative Speech Samples**

Prior to collecting data on the NSs’ perceptions of nonnative phonology, sample stimuli of learner speech was carefully collected and analyzed. This analysis was conducted in order to explore the phonological nature of non-target productions in the learner speech samples. Given that one of the overarching goals of this research is to investigate the types of learner productions that NSs most frequently identify as accented, it was necessary to first arrive at an understanding of the sounds produced with non-target productions within the sample stimuli.

**Learner participants.** Speech samples were collected from third- and fourth-semester students of Spanish at a large public east coast North American university. The participants consisted of two females and two males, with a mean age of 20 years. They were NSs of American English and did not speak other languages in their homes. None of the participants had advanced knowledge of linguistics or of any language other than English. All of the participants had lived in the Northeastern United States for the majority of their lives, and all were native speakers of a regional variation of northeastern American English. Two of the learners, one male and one female, were enrolled in Elementary Spanish 102, and the other two, one male and one female, were enrolled in a higher-level Spanish Conversational Review course. All participants were taking their Spanish courses for college credit. In addition to the learner recordings, speech samples were gathered from two NSs of Spanish: a female NS of Peninsular Spanish and a male NS of Latin American Spanish.
**Elicitation of learner speech samples.** After completing an initial survey to determine eligibility based upon the criteria described above, participants met individually with the researcher for their recorded sessions. Participants were informed that they would be creating narrations for two stories, each of which would be shown to them in sequentially-ordered storyboard drawings that depicted a day in the life of a fictional character. They were instructed to narrate creatively but to speak only in Spanish, with all verbs in the present tense. Since all second-, third-, and fourth-semester Spanish classes undergo an extensive review of verb tenses at the beginning of the semester, participants would have had a great deal of exposure to this verb tense. Thus, the present tense was selected in order to allow learners to devote more cognitive resources to meaning than to (morphosyntactic) form.

Participants were permitted to omit a block of the storyboard if they did not understand the action or event depicted in that illustration. Before each narration they were shown a short list of basic vocabulary items that they had learned previously in their Spanish classes and were given the opportunity to ask the researcher questions about word meaning or verb conjugations. Participants were informed that it was not obligatory to use the words on the vocabulary list; rather, they were suggestions that could be useful in their narration or would serve to remind them of vocabulary associated with one’s daily routine. After each participant indicated that the researcher had answered all questions concerning the list of suggested vocabulary, the list was taken out of the participant’s sight and they were not able to refer to it during the narration tasks. This was done to prevent the participants from reading directly from the vocabulary list and
thus ensure that speech recorded was truly extemporaneous production and not read text (refer to Chapter 1).

For the first of the two narration tasks, the researcher sat across from the participant and informed him or her that it was a “practice” narration; the participants were told that they were not being recorded at that point. During the practice narration, the participant was encouraged to ask the researcher any questions they had regarding task instructions. Upon completing the first narration task, the participant was asked to put on a Logitech “ClearChat Premium PC” headset with noise-cancelling microphone that rested approximately two-and-a-half inches from the mouth. The headset was connected to a Sony digital audio recorder. As with the practice narration, the participant briefly viewed a list of relevant vocabulary items before the researcher took the list away. The audio recorder was turned on, and the learner was presented with a second storyboard illustration. Upon presenting the participant with the materials, the researcher exited and left the participant alone in the room for this second recording. The storyboard included sixteen separate illustrations, and each learner produced the same number of utterances - one to correspond to each illustration. The speech elicited contained high-frequency lexical items, including many of the same lexemes from the sample vocabulary list presented to the learner.

**Speech sample splicing.** Once the paragraph-level narrations were collected from the four participants, utterance-level samples were spliced from these recordings in preparation for presentation to the NS evaluators. The original recordings were converted from .dvf format to .wav files with a sample rate of 44.1 mHz. Each participant’s narration was transcribed, and the utterances that were to be evaluated by the NSs were
chosen. Certain utterances were excluded because they contained elements with the potential to distract the NS evaluator from the phonological components of learner speech. Possible distracting elements included morphosyntactic errors, extensive false starts and non-linguistic sounds, such as coughing or sneezing.

In all, twenty utterances from the learner recordings were chosen to be evaluated by the NS listeners (Appendix B). The final twenty utterances were distributed fairly evenly amongst the four learners: five utterances from the lower-proficiency male learner, six utterances from the lower-proficiency female learner, five utterances from the higher-proficiency male learner and four utterances from the higher-proficiency female learner. The Praat (Boersma & Weenink, 2011) spectrogram program was used to splice these utterances from the longer recordings and store them as new digital files. These twenty utterances were given Likert-scale ratings by NS listeners for both accentedness and affective variables. However, the Chi-squared analysis of NS observations of the individual segmentals required that some of these original twenty utterances be excluded from that test, as is described in greater detail in the following section of this chapter.

With the exception of some accentuation errors that were morphosyntactic in Spanish, speech samples were not controlled for segmental errors. In addition, it was decided that the speech samples would also contain suprasegmental errors. Two main factors motivated the decision to not control for suprasegmental errors. First, non-target suprasegmental productions (pitch, intonation, etc.) were observed in almost every utterance elicited from the learners. Thus, any attempt to control for these errors would have led to a very reduced and artificial set of speech-sample stimuli. Second, since one of the main goals of this study is to deconstruct the various components of NS perception
of learner accent, it was imperative that NS evaluators also have the option of indicating any suprasegmental aspects of the speech samples that were problematic for them in their free response data. Though the main focus of this analysis is at the segmental level, it was necessary to investigate to what degree the suprasegmental errors observed by the NS evaluators contributed to the total perception of accentedness. Many researchers have already attested to the importance of both segmental and suprasegmental aspects of learner speech (Levis, 2005; Munro & Derwing, 2006; Zielinski, 2008). A finding that the segmental learner errors were substantially more or less salient to the NS listeners than the suprasegmental errors would be relevant in that it would both (a) contextualize the analysis of the segmental data presented here and (b) contribute to the larger discussion within L2PA research on the role of both segmental and suprasegmental non-target productions in NS perception of accent in learner speech.

**Error analysis of learner speech.** Prior to NS evaluations, the selected learner utterances were analyzed for phonetic and phonological errors at the segmental level. First, the author phonetically transcribed the twenty utterances that had been selected for NS evaluation. Then, any problematic areas of the transcriptions were revised under the guidance of an expert in phonetic transcription, who is a full university professor of linguistics at the time of this writing. Though many of the non-target productions were discernable to the naked ear, the Praat (Boersma & Weenink, 2011) spectrogram program was used to analyze any sound whose transcription was unclear to either of the evaluators. For example, productions of plosives [p t k] were labeled as “target-like” if their voice-onset time (VOT) was .03 seconds or less, a benchmark derived from the research of Lisker and Abramson (1964) and Moranski (2007). The lower VOT in the native-speaker
productions of these plosives indicates a low degree of aspiration in target productions, whereas the higher VOT often seen in spectrograms of learner speech is indicative of over-aspirated learner productions. Attempts at approximants [ð β ɣ] were also examined in the spectrogram to confirm that these productions lacked the full closure of their counterpart plosives [d b g].

In the case of [l], spectrogram readings were employed in order to discern the position of the second formant (F2), which indicates “backness” of tongue body position. A very low F2 reading is indicative of [l], a velarized or “dark” [l], which is a common Anglophone mispronunciation of the Spanish “clear l,” whose F2 is typically positioned higher in spectrogram readings (Hagiwara, 2009). Figures 3.1-3.3 below illustrate the differences in F2 height for these pronunciations in both learner and NS productions from the speech samples. In all three cases presented below, the liquid is in word-initial position and is followed by /i/ or /ɪ/. The Praat analyses show that the F2 for the learner’s pronunciation (Figure 3.1), at 1339 Hz, has more backness than the two NS productions in the same context (Figures 3.2 and 3.3), whose readings are at 2101 Hz and 1913 Hz, respectively. Thus, this particular learner production was labeled as non-target.

Figure 3.1 Spectrogram for SS04 (Male Learner Low) “ella limpia” (with Formants)
Spectrogram readings were also used to label fricatives as target or non-target. Voiceless fricatives /ʃ/ and /s/ were distinguished from voiced /z/ by the presence of what Hagiwara calls the “voicing bar,” (2009, sec. “Fricatives”) across the bottom of the spectrogram. Though there were several instances of learners using the alveopalatal fricative [ʃ] as an allophone of /s/, this was not considered a non-target pronunciation due to the fact that this allophonic palatalization exists in several dialects of Spanish, especially in southern Peninsular Spanish.¹⁶

Once complete, the phonetic transcriptions of the learner utterances were compared to standard target-language pronunciations. For example, the transcriptions below represent (1) the learner’s actual recorded speech and (2) a standardized target-language production of the phrase se levanta a las nueve (she wakes up at nine o’clock).
The above transcriptions illustrate the many possible differences in target versus learner pronunciations, in both consonant and vowel production. For the purposes of this study, however, the error analysis focused on learner production of consonants. The reasoning behind this decision is rooted in the data itself: once the NS evaluations were collected, it became clear that an overwhelming amount of the NS commentary on segmental features focused on learner production of L2 consonants. The high number of tokens from this category best supported the Chi-squared analysis used, which requires approximately twenty tokens.

Of the twenty learner sentences used in the NS-listener Likert-scale analyses of numerical accentedness and affective variable ratings, ten produced the number of tokens needed to be included in the Chi-squared analysis that examined the individual segments identified by the NS evaluators in the free-response data. These ten utterances were distributed among the four speakers as follows: three utterances from the lower-proficiency male learner, four utterances from the lower-proficiency female learner, two utterances from the higher-proficiency male learner and one utterance from the higher-proficiency female learner. It is not surprising that a greater number of tokens were generated in the speech samples from the two lower-level learners, since lower-level learners would be expected to produce more non-target L2P productions than higher-level learners. Below, Table 3.1 describes in detail the ten utterances used in the Chi-square analysis of the segmental tokens and shows the target and not-target consonantal productions from each utterance (from the complete consonant inventory from the speech...
samples). Table 3.2 summarizes the total number of target and non-target productions of each consonant, from the complete consonant inventory of the ten speech samples.

**Table 3.1**

*L2 Consonant Production Errors in Learner Speech by Sentence*

<table>
<thead>
<tr>
<th>Sample</th>
<th>Speaker</th>
<th>Transcription</th>
<th>Non-target</th>
<th>Target</th>
</tr>
</thead>
<tbody>
<tr>
<td>SS01</td>
<td>low male</td>
<td>se levanta a las nueve</td>
<td>t̪, k</td>
<td>s(2), l(2), n(2)</td>
</tr>
<tr>
<td>SS02</td>
<td>low male</td>
<td>toma un café</td>
<td></td>
<td>m, n, f</td>
</tr>
<tr>
<td>SS04</td>
<td>low male</td>
<td>ella limpia su casa</td>
<td>j, k, l</td>
<td>s, p</td>
</tr>
<tr>
<td>SS08</td>
<td>low female</td>
<td>limpia su carro</td>
<td></td>
<td>l, s</td>
</tr>
<tr>
<td>SS09</td>
<td>low female</td>
<td>ella habla con su amigo</td>
<td>j, k, r, β, l, n(2), s, m</td>
<td></td>
</tr>
<tr>
<td>SS10</td>
<td>low female</td>
<td>nada en la piscina</td>
<td>δ, l, p</td>
<td>n(3), s</td>
</tr>
<tr>
<td>SS13</td>
<td>low female</td>
<td>mira la televisión</td>
<td>r, θ, β, l</td>
<td>m, n, s, l</td>
</tr>
<tr>
<td>SS17</td>
<td>high male</td>
<td>ellos van a un restaurante</td>
<td>j, β, r</td>
<td>s(2), n(3), t̪(2)</td>
</tr>
<tr>
<td>SS20</td>
<td>high male</td>
<td>mira un poco de televisión</td>
<td>r, p, k, δ, t̪, β</td>
<td>m, n(2), l, s</td>
</tr>
<tr>
<td>SS25</td>
<td>high female</td>
<td>ella baila con un amigo</td>
<td>j, β, k, γ</td>
<td>n(2), m</td>
</tr>
</tbody>
</table>

**Table 3.2**

*Summary of Consonant Errors*

<table>
<thead>
<tr>
<th>Phone</th>
<th>Target-like tokens</th>
<th>Non-target tokens</th>
<th>Total tokens</th>
<th>Percentage target-like</th>
</tr>
</thead>
<tbody>
<tr>
<td>p</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td>0%</td>
</tr>
<tr>
<td>t̪</td>
<td>2</td>
<td>4</td>
<td>6</td>
<td>33.3%</td>
</tr>
<tr>
<td>k</td>
<td>0</td>
<td>6</td>
<td>6</td>
<td>0%</td>
</tr>
<tr>
<td>β</td>
<td>0</td>
<td>7</td>
<td>7</td>
<td>0%</td>
</tr>
<tr>
<td>δ</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>0%</td>
</tr>
<tr>
<td>γ</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>0%</td>
</tr>
<tr>
<td>r</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>0%</td>
</tr>
<tr>
<td>r̂</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td>0%</td>
</tr>
<tr>
<td>j</td>
<td>0</td>
<td>4</td>
<td>4</td>
<td>0%</td>
</tr>
<tr>
<td>l</td>
<td>5</td>
<td>4</td>
<td>9</td>
<td>55.6%</td>
</tr>
<tr>
<td>m</td>
<td>5</td>
<td>0</td>
<td>5</td>
<td>100%</td>
</tr>
<tr>
<td>n</td>
<td>16</td>
<td>0</td>
<td>16</td>
<td>100%</td>
</tr>
<tr>
<td>f</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>100%</td>
</tr>
</tbody>
</table>
Commentary on error distribution in learner samples. The unequal distribution of the number of tokens for each individual sound is attributed to the extemporaneous nature of learner production and was to be expected in these speech samples. Though the small numbers in this distribution do not allow for any inferential statistical tests such as Chi-square, to be performed on this subset of the speech-sample stimuli, several patterns can be discerned from the descriptive data presented in Tables 3.1 and 3.2. For example, the learners had no difficulty producing target-like tokens of the nasals [m] and [n]. Since these nasals occur in both Spanish and English phonemic inventories and in similar linguistic environments, target-like productions of these phones could be expected.17

Unlike nasals, the fricatives [β ð γ] were never produced in a target-like fashion in the speech samples. These results support Eckman’s MDH, which states that these sounds would be among the most challenging for learners to produce because they represent: “areas of the target language (TL) which differ from ones in the NL and are more marked than those in the native language.” (Eckman, 1977, p. 321) Spanish fricatives [β ð γ] are both (a) different from sounds in the L1 in that [ð] appears in completely different contexts in Spanish than in English, and /β γ/ are not part of the English phonological inventory, and (b) more marked in Spanish because their plosive counterparts /b g d/ imply the presence of their respective allophones [β ð γ] but not vice-versa. The complete lack of target productions for [β ð γ] is consistent with Castino’s (1991) findings, which found that L1 AE speaking learners to have high levels of non-target productions of [β ð γ].
The absence of native-like tokens of trill /r/ also supports Castino’s findings and Eckman’s MDH. The MDH states that trill /r/ would be at the highest level of difficulty for learners because it is (1) different from any sound in the AE phonological inventory and (2) universally considered a marked phone (Castino, 1991). Like the speech samples used in this study, Castino’s work (1991) also showed trill /r/ to be among the most frequently mispronounced sounds by L1 AE speakers learning Spanish.

The plosives /p t k/, though unmarked in comparison to [β δ ü], had very few native-like productions in these speech samples. According to the MDH, these sounds should be easier for learners to produce in a target-like fashion than [β δ ü], since they are less marked. However, these sounds will prove more difficult than nasals because the phonemes /p t k/ have different phonetic manifestations in English and Spanish. In many AE linguistic environments, these sounds are often highly aspirated and are represented phonetically as [pʰ tʰ kʰ]. The Spanish counterparts lack aspiration, and Spanish /t/ has a dental place of articulation ([t̪] in comparison to the English /t/’s alveolar place of articulation).

Even though Eckman’s MDH posits that [pʰ tʰ kʰ] will be less difficult for learners than [β δ ü] - a supposition supported by Castino’s research - the learners in this study had great difficulty producing target-like tokens of /p t k/. With the exception of [t̪], which had a very small number of native-like pronunciations, the speech samples studied here lack native-like production of these sounds. One possible explanation for the lack of native-like tokens of /p t k/ considers the role of interference of the L1, which the MDH does not take into account. Though it takes more articulatory force to aspirate a plosive,
the influence of the speakers’ L1 English backgrounds could have maintained this tendency to aspirate, which ultimately resulted in the nonnative productions.

Finally, it is imperative to note that the observations presented here are drawn from the small number of speech samples and are in no way considered exhaustive. The main purpose of this investigation of the learner speech samples was to arrive at a better understanding of the stimuli used in segmental analysis. The goal of this short analysis of the speech stimuli was to verify whether they shared characteristics of other speech samples used in perception research and to examine whether Eckman’s MDH was applicable to these speech samples, as it had been in Castino’s (1991) study. In short, the segmental non-target productions of these learner speech samples did in part support the first tenant of Eckman’s MDH, which holds that sounds in the L2 that are more marked and different from their counterparts in the L1 (should they exist) present the greatest degree of difficulty to learners. There were in fact no target-like pronunciations of the sounds corresponding with this first degree of difficulty ([r β ð γ]) in this sample. However, there were also very few target-like tokens of the several sounds that have been classified as less difficult according to the MDH ([p t k]). It is possible that L1 interference accounted for the lack of target-like productions of the less marked sounds. Here, it should also be reiterated that these speech samples served as the independent variable used in the segmental analysis, and the NSs’ analyses of them will be the main dependent variable of the present investigation. The Likert-scale evaluations of accentedness and attitudinal data included the full corpus of the twenty utterances.
Phase II: Data Elicitation from NS Evaluators

The second phase of the data collection was dedicated to gathering data on the study’s dependent variable - NS evaluations. In order to answer this investigation’s three research questions, three main sets of data were gathered from the NS listeners. NS evaluators listened to the full set of speech-sample stimuli (twenty utterances) and provided Likert-scale ratings of accentedness as well as free-response data on the aspects of each individual utterance that they perceived as most salient. This was carried out in order to address the research questions of (1) “Which aspects of Spanish L2 learner extemporaneous speech do NSs most often identify as accented?” and (2) “Is there a hierarchical order in Spanish that ranks sounds in terms of their potential for contributing to NS perceptions of accent?” It should be noted that the analysis of the second research question required that a reduced set of the NS listener data be used in order to complete a Chi-square analysis, as detailed in the previous section of this chapter.

In addition to the Likert-scale and free-response data collected on NS perception of accentedness, NS listeners also completed a Likert-scale questionnaire regarding their perception of NSs of AE learning Spanish and of the United States in general. The attitudinal data were correlated with the Likert-scale accentedness data, in order to address the third research question of “How do NS attitudinal variables correlate with the way in which they evaluate accentedness of learner speech, if at all?”

All three sets of NS-evaluator data were collected via an online survey in which the NS evaluators heard the learner utterances and provided Likert-scale accentedness ratings and free-response evaluations. Each online survey concluded with a Likert-scale
attitudinal questionnaire. The following section describes in detail (a) the demographic information from the NS-evaluators and (b) the design of the online-survey task.

**NS-listener participants.** Of the 40 original respondents, 23 completed the survey in its entirety, but five were excluded due to prior training in linguistics. The remaining participants \((n = 18)\) were all NSs of Spanish, were relatively balanced in gender (11 females and 7 males), and had a mean age of 28.2 years (four participants failed to report age). The participants were from varied educational backgrounds. When asked to report their highest level of education, two reported high-school, ten reported that they had either completed a bachelor’s degree or had some college education, four had master’s degrees, and one participant reported five years of professional training. In contrast to the five excluded respondents, the NS listeners whose data were analyzed reported having no specialized knowledge or training in language instruction. Upon completion of the online-survey task, all participants were given an honorarium in the form their choice of either a digital gift card of 10 dollars from Amazon.com or of a digital gift card of 10 euros from fnac.es. When asked their nationality, the NS listeners reported as follows: thirteen Spanish, two Columbian, one Mexican, one Chilean, and one Guatemalan. It should be noted that the predominance of Spanish NS listener participants is not viewed negatively within the context of this particular investigation. The university in which the research took place has a strong study-abroad tradition in Spain and a long-standing affiliation with a Spanish university; consequently, the majority of students that study abroad do so in Spain. Therefore, it was advantageous to obtain a majority of NS evaluations of learner speech from the population with which many of these learners would eventually interact.
Since the participants were able to take the survey online, much of the data collected was from outside the United States. Figure 3.4 below illustrates the internet protocol (IP) address map for the NS participants.

![Figure 3.4 IP Address Map for NS Evaluators](image)

Though the majority of the NS evaluators were Spanish, the IP addresses from Figure 3.4 show that the NS participants completed the online survey in two principal geographic areas: fourteen were from within Spain and four from within the United States. The IP addresses from within the United States suggest that these participants would have some degree of familiarity with English, though it was unclear if they were residing permanently within the United States or whether they were there for a finite amount of time (for work, studies, family issues, etc.) However, the participants with IP addresses from within the United States were not excluded from the analysis because they had indicated that their “dominant” language was Spanish and that they were not native or native-fluent speakers of English. When asked to self-report their proficiency in English, the eighteen NS evaluators reported the following indicated the following: two beginner level, eleven intermediate level, and four advanced level. One participant
indicated that she had no prior experience with English. Of the NS participants that self-reported their English proficiency level as advanced, three conducted the survey from IP addresses within Spain and one from an IP address within the United States.

**Online survey tool: Overview.** In order to reach the maximum number of Spanish-dominant participants, an online survey tool was employed. These types of online surveys have been used by prominent scholars in the field of L2 phonology acquisition (Anderson & Souza, 2011) and serve as a valuable data collection tool, capable of efficiently collecting large amounts of listener data. The survey used in the current investigation was sent to potential participants via a link generated by the online survey company SurveyGizmo.com.\(^{19}\) The online survey took approximately 25 minutes to complete and had the following structure:

I. Overview of survey and digital consent form

II. Participant background questionnaire

III. Instructions and practice speech sample evaluation task

IV. Speech sample evaluation task

V. Attitudinal questionnaire

**Instructions and practice speech sample ratings.** Once participants completed the overview, consent, and background questionnaire sections, they were given detailed instructions as to how to complete the speech evaluation task. These instructions consisted of incremental directions for the rating task, which included screenshots of the windows they would see when completing the actual evaluations. They were also shown a variety of sample responses to the open-ended portions of the task. These sample responses included both segmental and suprasegmental observations. The purpose of
exposing participants to the sample responses was not to prime them, but rather to show
the level of detail needed in the open-ended portions and thus avoid very broad comments
of “good” or “bad” accent that would provide little insight into the NSs’ perceptions.

Special concern was given to the topic of regional accent. Given the abundance
and diversity of regional accents within the Spanish language, it was imperative that the
NS listeners understood that, in this case, “accent” did not refer to regional accents or
dialects of Spanish but rather to nonnative accent. The study’s introductory pages
informed participants that they would be evaluating the speech of nonnative speakers,
and they were again reminded of this in the stepwise instructions. At the end of the
instruction pages, a full page was dedicated to this point. In red letters participants were
reminded that they would be commenting on “nonnative aspects” of speech samples and
given the example of a NS of German learning Spanish. On this page they were again
reminded that they were not to consider aspects of regional dialects of Spanish as
“nonnative.”

After finishing the instructions section, the NS listeners completed three practice
evaluations. Of the speech samples used in these practice evaluations, two were from
nonnative speakers and one was from a NS of Spanish. The purpose for the practice
evaluations was twofold: the NS raters had the opportunity to practice the task, and the
researcher could also refer to the ratings for these practice evaluations to ensure that the
NS raters understood the evaluation task.

For the evaluation task, each speech sample appeared on its own screen (window),
as is illustrated in Figure 3.5 below:
As shown in Figure 3.5, participants were given the transcription of the utterance with the corresponding sound clip below the transcription. This study was piloted in three different formats: with utterance transcription, without utterance transcription, and with illustrations instead of transcription. It was decided to utilize the utterance transcription, after participants in the pilot study reported confusion with the other two versions.

After listening to each sample, participants indicated on a six-point Likert scale the degree of foreign accent they perceived in the speech sample. Though some researchers have indicated a preference for longer Likert-scales (Fletcher & Flege, 1992), the six-point Likert scale was ultimately chosen based on feedback from the pilot study.
participants who commented that the longer scales were overwhelming and confusing. A rating of zero, at the left extreme of the scale, indicated that the listener found the utterance to be “without accent/of a native speaker [of Spanish].” A rating of five, at the right extreme of the scale, indicated that the NS rater perceived the utterance as having “much nonnative accent.”

Following the Likert-scale rating, NS raters were asked to identify the aspects of the speech sample that they perceived as “más alejados del habla de un nativo” (“most different from the speech of a native-speaker”). NS raters were given two text boxes in which they could write up to two observations. If they had additional observations, they had the option of including them in the section labeled “observaciones generales” at the end of each page. There was no limit to the length of the comments, but the program was configured such that the participants had to complete the Likert-scale question and write at least one open-ended response before they were allowed to advance to the next speech sample. The data elicited in the free-response sections were analyzed in order to determine which segmentals were most frequently noticed by the NS listeners, in accordance with the second research question.

**Accentedness Data Collection.** The actual accentedness data-collection task had the exact same format as the practice task. Once the NS listeners successfully completed the background questions of the online survey and the three practice questions, NS listeners were told that they were exiting the practice stage and that they would begin their actual evaluations. In total, the NS raters evaluated thirty utterances: twenty from the nonnative-speaker learners and ten from the two NS control participants. As in the practice evaluations, each utterance was represented in a separate screen as shown in
Figure 3.5. Each NS rater’s survey presented the thirty speech samples in a different, randomized order, as configured by the online survey program. The NS evaluators were able to progress through the task at their own pace.

**Attitudinal data collection.** The final component of the online survey was a required attitudinal questionnaire that asked questions regarding the NS listeners’ perceptions of NSs of AE learning Spanish and of the United States in general. The NS raters were also presented with an item that asked them to indicate their overall willingness to converse with someone that spoke Spanish with a pronounced AE accent. An additional item asked participants to indicate the number of American friends that they had, as this would be indicative of their “in-group” affiliation (see Chapter 2). With the exceptions of the questions regarding willingness to communicate with someone with highly accented speech and their number of American friends, the other attitudinal questions were in the same 6-point Likert-scale format that was used for the speech sample ratings. A screenshot from the attitudinal questionnaire portion of the survey is shown below in Figure 3.6.
Figure 3.6 *Screenshot of Attitudinal Questionnaire*

The attitudinal data was of particular relevance to this study for two reasons. First, using the same six-point Likert scale to collect both speech sample ratings and attitudinal data would allow statistical testing for a correlation of these two variables. It was possible that higher, more positive perceptions of NSs of AE learning Spanish and their home country would correlate with higher ratings of the speech samples themselves. Should this be the case, it would indicate that there were others factors contributing to the NS listener ratings outside of the properties of the actual speech samples. Additionally, the attitudinal data were gathered in anticipation of any possible outliers in the speech sample ratings.
CHAPTER 4

Results

The results for the procedures described in the last chapter are divided into two main sections: (1) numerical listener ratings for speech samples, attitudinal data, and correlations between the two, and (2) listener commentary on specific aspects of learner speech. The analysis for the former is quantitative in nature and addresses the research question of “How do NS attitudinal variables correlate with the way in which they evaluate accentedness of learner speech, if at all?” The latter lends itself to mixed-methods analyses and addresses the research question of “Which aspects of Spanish L2 learner extemporaneous speech do NSs most often identify as accented?” and “Is there a hierarchical order in Spanish that ranks sounds in terms of their potential for contributing to NS perceptions of accent?”

NS Likert-Scale Ratings and Attitudinal Data

Measures of central tendency. As discussed in Chapter 3, each of the 18 NS listeners rated the 30 speech samples on a Likert scale ranging from 0 to 5. A rating of 0 indicated that they perceived the speech sample was from a “native speaker/without accent,” while a rating of 5 indicated that that speaker had a “strong nonnative accent.” Thus, a sample with a lower numerical score meant that the sample was rated as more native-like, and a higher numerical score indicated less native-like speech. As mentioned in detail in the methodology sections, the twenty learner speech samples were collected from four nonnative speakers of Spanish whose L1 was AE, two males and two females. Ten additional control speech samples were collected from NSs. Before further descriptive or inferential statistics were calculated, the Likert-scale ratings of the learner
speech samples were analyzed for internal consistency reliability. Cronbach’s alpha was found to be $\alpha = .95$, indicating a very high level of internal consistency within the NS listener ratings. This high level of internal consistency in the accentedness ratings meant that each utterance was receiving relatively equally low or high ratings from each NS listener. In other words, there was general agreement among the NS evaluators as to which utterances they perceived as highly accented and which they perceived as native-like. It was especially important to verify internal consistency in these ratings before performing the segmental analysis of the free-response commentary; if the NS listeners differed greatly in their accentedness ratings that would have negative implications for the feasibility of finding representative patterns in the free-response commentary data.

Measures of central tendency were calculated for the NS-evaluator Likert-scale ratings of both learner and NS control speech samples. Since the speech samples provided by the speakers were elicited during an extemporaneous production task, each speaker had unique linguistic qualities. In the majority of analyses conducted in this study, the uniqueness of the samples across participants did not allow for average ratings to be collapsed into more general groups. For example, the average accentedness scores are not collapsed into groups of higher-level and lower-level speakers because, due to the extemporaneous nature of the task, the linguistic content of each speaker’s speech samples differed greatly. Instead, average NS listener ratings for the full set of the twenty learner utterances were calculated for all tokens analyzed from each individual speaker. Results are presented in Tables 4.1 and 4.2.

In Table 4.1, the average NS ratings for the listener speech samples are grouped according to the four speakers: male lower-level, female lower-level, male higher-level
and female higher-level. In this case, \( n \) refers to the total number of speech samples analyzed by NS listeners for each learner. For example, five utterances from one learner were used in the stimuli and then evaluated by the eighteen NS listeners, then \( 5 \times 18 = 90 \) utterances were evaluated from that learner.

Table 4.1

*NS Listener Accentedness Ratings for Learner Speech*

<table>
<thead>
<tr>
<th>Gender</th>
<th>Lower-level learners</th>
<th>Higher-level learners</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( n^* )</td>
<td>( M )</td>
</tr>
<tr>
<td>Male</td>
<td>90</td>
<td>4.13</td>
</tr>
<tr>
<td>Female</td>
<td>108</td>
<td>3.93</td>
</tr>
</tbody>
</table>

Table 4.2

*One-Way Analysis of Variance of NS Likert-scale Accentedness Rating by NS Control*

<table>
<thead>
<tr>
<th>Source</th>
<th>( df )</th>
<th>( SS )</th>
<th>( MS )</th>
<th>( F )</th>
<th>( p )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between groups</td>
<td>3</td>
<td>21.191</td>
<td>7.064</td>
<td>8.484</td>
<td>.000</td>
</tr>
<tr>
<td>Within groups</td>
<td>68</td>
<td>56.613</td>
<td>.833</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>71</td>
<td>77.804</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The average scores for the individual learners showed that the learners that were enrolled in the higher-level Spanish course were rated as having more native-like pronunciations than those enrolled in the lower-level Spanish course. The female learner in the lower-proficiency group was rated as sounding more native-like than the male from this group, and the male learner in the higher-proficiency group was rated as sounding more native-like than the female from that group. A one-way ANOVA showed a significant effect of learner category on NS speech-sample ratings \([F(3, 68) = 8.484,\)
MSE = .833, $p < .000$, $\eta = .27$. A post-hoc Tukey HSD test (at $p < .05$) indicated that the mean score for the higher-level male learner ($M = 2.73$, $SD = .99$) was significantly different from the scores of both the male lower-level ($M = 4.13$, $SD = .85$) and female lower-level ($M = 3.4$, $SD = .9$) learners and that the lower-level male learner was significantly different from the higher-level female learner.

The mean ratings for the NS control samples, presented in Table 4.3 below, were much lower than the mean ratings for any of the learners. Here, it should be noted that the NS control stimuli were balanced for both gender and regional accent: The female NS control spoke a northern peninsular dialect of Spanish, while the male NS control spoke a Caribbean dialect of Latin American Spanish. Though the great majority of the NS evaluators were Spaniards, the ratings for the NS controls were very similar. An independent-samples t-test was conducted to compare NS-evaluator accentedness ratings of the male NS control speech sample and the female NS control speech sample. There was no significant difference in NS accentedness ratings for the male NS control ($M = .042$, $SD = .65$) and the female NS control speech ($M = .047$, $SD = .85$); $t(34) = -.129$, $p = .898$, $d = -.001$. This indicates that the NS evaluators did not base their accentedness ratings on regional accent (as they were instructed to ignore regional accent in the practice ratings).

Table 4.3

<table>
<thead>
<tr>
<th>Gender</th>
<th>n</th>
<th>M</th>
<th>SD</th>
<th>t</th>
<th>df</th>
<th>p</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>90</td>
<td>.042</td>
<td>.65</td>
<td>-.129</td>
<td>34</td>
<td>.898</td>
<td>-.001</td>
</tr>
<tr>
<td>Female</td>
<td>90</td>
<td>.047</td>
<td>.85</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Likert-scale attitudinal data and correlation with speech sample ratings.

Attitudinal data were collected from the NS listeners upon completion of the speech sample-rating task, at the end of the survey. Participants were asked several questions regarding their perception of (a) NSs of AE and (b) the United States in general. Participants answered these questions using a six-point Likert scale, with a score of 0 indicating a strongly negative perception and a score of 5 indicating a strongly positive perception. Results are presented in Table 4.4.

Table 4.4

<table>
<thead>
<tr>
<th>Likert Scale Attitudinal Data from Spanish NSs</th>
<th>n</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perception of NSs of AE</td>
<td>18</td>
<td>4</td>
<td>1.08</td>
</tr>
<tr>
<td>Perception of United States</td>
<td>18</td>
<td>3.78</td>
<td>.81</td>
</tr>
</tbody>
</table>

Though the participants indicated that they had slightly more positive perceptions of NSs of AE than of the United States in general, ratings for both categories were well above neutral. The underlying distributions for these variables showed positive skew, reflecting the positive ratings for these two variables.

The attitudinal data were then correlated with the average accentedness rating that each NS listener gave for all twenty of the learner speech samples. Though the distinct learner categories were not collapsed for gender or proficiency level in the above section on central tendency, global ratings of accentedness by each NS listener were used to correlate each NS evaluator’s Likert-scale ratings for accentedness with their Likert-scale data for the attitudinal variables. This allowed investigation into whether NS evaluators’ overall accentedness ratings correlated with the positive or negative perceptions of the learners.
It was hypothesized that high global ratings for accentedness may correspond with strongly negative perceptions of NSs of AE and of the United States, and conversely low ratings of accentedness may correspond with more positive perceptions. It was also hypothesized that possible outliers in the NS accentedness ratings could correspond to very strong positive or negative reports for the attitudinal variables.

Due to the positive skew present in the underlying distributions of the attitudinal variables, a Spearman’s Rho rank-order correlation was carried out on the following variables: perception of NSs of AE, general perception of the United States, and Likert-scale speech sample ratings. A one-tailed test was used to correlate the attitudinal data with the speech sample ratings, since it was reasonable to anticipate positive perceptions of the NSs of AE and the United States to negatively correlate with speech sample ratings (lower speech ratings were more “native-like”), while it would not be anticipated for these variables to positively correlate (raters with negative perceptions of NSs of AE and the United States favorably rating the speech samples). In contrast, a two-tailed test was utilized for the correlation between the NS evaluators’ general perception of the United States and their perception of L1 AE speakers learning Spanish, due to a lack of anticipated directionality in the results. For example, a rater could have positive/negative perceptions of both or conversely could maintain a positive perception of Americans while disliking the United States (due to political leadership, economic policies, international policies, etc.).

The test revealed a statistically significant negative correlation between general perception of the United States and speech samples ratings $r_s (16) = -.467, p = .025, d = -.24$, indicating that participants who reported more positive perception of the United
States also rated the learner speech as more native-like. There was no significant
correlation between perception of NSs of AE (the learners) and speech sample ratings $r_s$
$(16) = -.258, p = .151, d = .435$. As could be expected, there was a significant positive
correlation between perception of NSs of AE and general perception of the United States,
$r_s (16) = .651, p = .003, d = .232$. These results are shown in Table 4.5.

Table 4.5

**Correlation Between Speech Sample Ratings and Attitudinal Data**

<table>
<thead>
<tr>
<th>Speech Sample Ratings</th>
<th>Perceived NSs of AE</th>
<th>Perception of United States</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speech Sample Ratings</td>
<td>$n = 18$</td>
<td></td>
</tr>
<tr>
<td>Perception of NSs of AE</td>
<td>-.258</td>
<td></td>
</tr>
<tr>
<td>Perception of United States</td>
<td>-.467*</td>
<td>.651**</td>
</tr>
</tbody>
</table>

* *$p < .05$*
**$p < .01$**

Additional attitudinal data. Participants were also asked to indicate the number
of friends that they had who were NSs of AE. Participants responded by choosing one of
four categories, as shown in Table 4.6. Since this data was collected as a multiple-choice
question, data are presented in percentages. As evidenced by the data, the majority of NS
raters reported having a moderate to high number of American friends.

Table 4.6

**NS Listener Number of Friends who are NSs of AE**

<table>
<thead>
<tr>
<th>Number of American Friends</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>11%</td>
</tr>
<tr>
<td>1-2</td>
<td>6%</td>
</tr>
<tr>
<td>3-7</td>
<td>22%</td>
</tr>
<tr>
<td>7+</td>
<td>61%</td>
</tr>
</tbody>
</table>
The final attitudinal variable assessed the willingness of the NS raters to interact with NSs of AE whose Spanish was heavily accented. This question, presented in multiple-choice format, asked participants which response they would have should a stranger with a strong AE accent approach them and attempt to speak with them in Spanish. Results are shown in Table 4.7 below.

Table 4.7

<table>
<thead>
<tr>
<th>NS Willingness to Converse with Someone with a Strong AE Accent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Response</td>
</tr>
<tr>
<td>No perdería tiempo en intentar comunicarme con él.</td>
</tr>
<tr>
<td>(I would not waste my time trying to communicate with him.)</td>
</tr>
<tr>
<td>El acento me molestaría bastante pero todavía haría un esfuerzo para hablar con él.</td>
</tr>
<tr>
<td>(The accent would bother me a lot but I would still make the effort to speak with him.)</td>
</tr>
<tr>
<td>El acento me molestaría un poco pero haría un esfuerzo para hablar con él.</td>
</tr>
<tr>
<td>(The accent would bother me a little but I would make an effort to speak with him.)</td>
</tr>
<tr>
<td>El acento no me molestaría y hablaría con él.</td>
</tr>
<tr>
<td>(The accent would not bother me and I would speak with him.)</td>
</tr>
</tbody>
</table>

All of the participants indicated that they would make an effort to converse with the stranger, with the majority stating that the strong AE accent would not bother them at all. It should be noted that this question was posed to the NS listeners after they had completed the 25-minute task of listening to heavily accented learner speech.

Results from Free-Response Data

After listening to each speech sample, the NS evaluators also wrote commentary as to which aspects of the sample they perceived as having the strongest foreign accent. They had two text boxes in which they could write their commentary, as well as a
“general comments” text box at the end of each page of the survey. Respondents were required to fill in at least one of the commentary text boxes before they could proceed to the next page of the survey.

**Inter-rater reliability.** The 606 NS commentaries were coded into one of seven general categories, as shown in Table 4.8. In order to develop the coding scheme, 20% of the free-response data was analyzed, and seven major categories were determined to be representative of the data. After all of the tokens were coded by the first (primary) rater, a second rater who had been trained in the coding scheme coded 25% (111) of the tokens. After the second rater’s first coding, agreement between the two raters was 79%, which was considered unacceptable. The second rater and the first rater met several times, and it was found that the majority of the disagreement in the coding was due to a misunderstanding of the “segmental” and “general” categories. After reconciling this difference, the second rater coded a different portion of the data, which represented 28.5% (173) of the tokens. Agreement for this second portion of coded data was 87%, with 151 agreements and 22 disagreements between the raters. Cronbach’s kappa was very high at $\kappa = 0.801$, $p = .000$. The raters met again and reconciled the differences for the 22 disagreements to reach 100% agreement. The first rater then used this coding scheme to recode the data set.
### Table 4.8

**Category Distributions for Free-Response Data**

<table>
<thead>
<tr>
<th>Category</th>
<th>Sub-category</th>
<th>Definition</th>
<th>Tokens</th>
<th>Category</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Segmental</strong></td>
<td>Seg</td>
<td>Negative commentary on segmental consonants</td>
<td>203</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SegV</td>
<td>Negative commentary on vowels</td>
<td>82</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SegREG</td>
<td>Segmental production identified as incorrect due to regional/dialectical differences</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SegPOS</td>
<td>Positive commentary on segmental production</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Category total:</strong></td>
<td></td>
<td><strong>302</strong></td>
<td><strong>(50%)</strong></td>
</tr>
<tr>
<td><strong>Suprasegmental</strong></td>
<td>Sup</td>
<td>Negative commentary on suprasegmental production (such as intonation, linking, etc.)</td>
<td>152</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SupPOS</td>
<td>Positive commentary on suprasegmental production</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Category total:</strong></td>
<td></td>
<td><strong>154</strong></td>
<td><strong>(25%)</strong></td>
</tr>
<tr>
<td><strong>General</strong></td>
<td>Gen</td>
<td>Negative general comments (“bad pronunciation,” “sounds nonnative,” etc.)</td>
<td>34</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GenPOS</td>
<td>Positive general comments (“good pronunciation,” “perfect,” etc.)</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Category total:</strong></td>
<td></td>
<td><strong>47</strong></td>
<td><strong>(8%)</strong></td>
</tr>
<tr>
<td><strong>Word</strong></td>
<td>Word</td>
<td>Negative commentary on a specific word</td>
<td>66</td>
<td></td>
</tr>
<tr>
<td></td>
<td>WordPOS</td>
<td>Positive commentary on a specific word</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Category total:</strong></td>
<td></td>
<td><strong>66</strong></td>
<td><strong>(11%)</strong></td>
</tr>
<tr>
<td><strong>Syllable</strong></td>
<td>Syll</td>
<td>Negative commentary on a specific syllable</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SyllPOS</td>
<td>Positive commentary on a specific syllable</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Category total:</strong></td>
<td></td>
<td><strong>14</strong></td>
<td><strong>(2%)</strong></td>
</tr>
<tr>
<td><strong>Ambiguous</strong></td>
<td>N/A</td>
<td>Comment ambiguous/could not be coded</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td><strong>Syntax</strong></td>
<td>N/A</td>
<td>Commentary on syntactic elements of sample</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Overall total:</strong></td>
<td></td>
<td><strong>606</strong></td>
<td></td>
</tr>
</tbody>
</table>
Categories. Each commentary token analyzed was assigned to one of the seven major categories identified in Table 4.8, as well as to a subcategory when applicable. Examples of classifications for tokens from one of the speech samples are presented in Table 4.9. As is shown by the results, segmental changes were the most frequently identified category among the commentaries (tokens) provided by the NS listeners. In order for a token to be considered segmental, a direct reference to a non-target production of a particular segment was required. Some of these observations were of non-target productions that were phonetic in nature, such as the segment /t/ referenced in Table 4.9 below. This phoneme’s non-target production is [θ] instead of the target [k], due to the fact that the learner uses an alveolar place of articulation instead of the target dental place of articulation, thus causing a higher, non-native VOT. Other observations categorized as segmental referred to non-target pronunciations that, in Spanish, were phonemic in nature. An example of this is found in one of the NS listeners’ comments for SS08, in which the learner produces the single-tap /ɾ/ in a linguistic environment that necessitates multiple-tap /ɾ/. The result is that the lexeme /ká.ro/ “car” is produced as /ká.ɾo/ “expensive.” While the context of the sentence limpia su carro (she washes her car) allows for the NS listeners to comprehend its intended meaning, the phonemic-level non-target production of /ɾ/ corresponds with comments such as “‘rr’ de ‘carro’ como una ‘r’” (“the ‘rr’ of ‘carro’ is like an ‘r’”) and “‘rr’ no suena como debe ser en ‘carro’” (“the ‘rr’ does not sound like it should in ‘carro’”).

The segmental data for this study showed that the NS evaluators identified non-target consonantal productions more than twice as much as they did non-target vowel productions. This difference could be attributed to the contrasting vowel inventories of
American English and Spanish. While dialects of American English can have between fifteen and twenty vowels, the great majority of Spanish dialects have five vowels. Thus, it is possible that non-target productions of Spanish vowels assimilated into the NS evaluators’ perceptual vowel space, while non-target consonantal productions were more salient, since the consonantal inventories of English and Spanish are similar in size.

Table 4.9

*Examples of Token Distribution among Categories*

<table>
<thead>
<tr>
<th>Category</th>
<th>Sample NS Listener Commentary on Learner Production of SS01 [se.le.ban.t3o-∫a.las.nu.wej.ve], <em>Se levanta a las nueve</em> (target production: [se.le.βan.ta.a.las.nue.βe])</th>
</tr>
</thead>
</table>
| Segmental  | *t de “levanta” similar a una “cha”*  
(t of “levanta” is similar to a “cha”)                                                                                                                                                           |
|            | *la t de levanta no tiene buena pronunciación*  
(the t of *levanta* is not pronounced well)                                                                                                                                                     |
| Suprasegmental | *poco fluido*  
(not very fluent)                                                                                                                                                                           |
|            | *intonación cantarina*  
(sing-song intonation)                                                                                                                                                                         |
| General    | *la pronunciación de “levanta”*  
(the pronunciation of “*levanta*”)                                                                                                                                                             |
| Word       | *nueve suena inglés*  
(“*nueve*” sounds English)                                                                                                                                                                        |

It should be noted that in all cases tokens categorized as segmental were closely examined in the learner speech samples to ensure that no overt suprasegmental non-target productions in areas such as pitch, intonation, rate of speech, etc., were also present. NS-listener tokens that identified segmental consonants as having too much or too little “stress” were categorized as either segmental or suprasegmental, depending upon the circumstances. For example, a commentary for SS01 that “*La ‘t’ en la palabra levanta tiene mucho enfasis*” (there is a lot of emphasis on the ‘t’ in *levanta*) is categorized as segmental, since it is clear that this comment focuses on the over-aspirated production of
In contrast, a NS-listener comment for SS10 of “NADA. Falta enfasis [sic] en la primera ‘A’” (NADA. Lacks emphasis on the first “A”) is categorized as suprasegmental, as it is in reference to the fact that the learner stresses the second syllable instead of the first, thus incorrectly making this paroxytone (stress on penultimate syllable) word an oxytone word (stress on last syllable).

Other comments labeled as suprasegmental included those that referenced pausing, intonation, fluency, rate of speech and linking.

Tokens categorized as “general” included global commentaries or evaluations of accent, pronunciation or stress, or comments pertaining to the learners’ level of confidence, persona, etc. A very small number of comments were categorized as “general” under the subcategory of “position,” meaning that they addressed place of articulation in a general manner. For example, one of the tokens for SS09 labeled in this way commented that “todo con la misma posición de la lengua, muy interno sin sacar las palabras fuera” (“everything is with the same position of the tongue, very internal, without bringing out the words”).

Tokens were classified in the word category only when the commentary could not be broken down past the level of an individual lexeme. In the example provided in Table 4.8, “la pronunciación de ‘levanta’” (the pronunciation of “levanta”), it would be impossible to infer which component of the learner’s production motivated the comment. Though it could seem reasonable to extrapolate the NS evaluator’s intended meaning for certain word-level commentaries, such as those containing non-target phonemic-level productions like that of /ká.ro/ for /ká.ro/, it was ultimately decided that such a high level of inference was not appropriate for this type of study.
Similar to NS-commentary tokens that were categorized at the word level, the analysis of the data showed a small percentage of comments that could not be analyzed beyond the level of the individual syllable. For example, a comment for SS08 identified the “ca de carr[o]” (the ‘ca’ of ‘carr[o]’) as an aspect of the learner’s speech that stood out as being non-native. Given the spectrogram information regarding the increased VOT for the /k/, it is quite reasonable to assume that this segment motivated the comment and categorize it accordingly. However, as was the case with the word-level comments, this very high level of inferential coding was deemed inappropriate.

A very small number of tokens could not be classified into any of the major categories and were labeled as ambiguous. In addition to differing from the main categories, these tokens were also quite dissimilar from one another, thus preventing more precise categorization. Examples of these ambiguous tokens can be found in the comments for SS25 “sonido especial en ‘ella’ y ‘amigo’” (special sound in ella and amigo) and SS03 (not featured in Table 3.4) “la terminación de la palabra ‘amigo’” (the ending of the word ‘amigo’).

Even though the listeners were instructed to report the aspects of the learners’ pronunciation that seemed least native-like to them, there was one case of syntactic commentary. In this token the listener commented on the learner’s word choice: “Se diría se ve no se mira” (you would say that you watch, not that you look at).

Finally, it should be noted that the speech samples used in this study were taken from extemporaneous speech and were not balanced to contain an equal number of non-target productions from each category. For this reason, the proportions of comments in each of
the distinctive categories do not evidence the salience of one over another in terms of NS perception of learner speech

**Quantitative analysis of segmental data.** In order to address the research question of determining the relative weight of specific segmental elements in NS perception of accent, the analysis of the free-response data focuses on commentary of consonantal segmentals. Segmental consonants were chosen due to their prevalence within the segmental category (more than twice as many consonants as vowels were identified), which best lent itself to statistical analysis. Since each utterance was unique, the 10 utterances were considered as individual cases for the analysis. As described in Chapter 3, ten of the original twenty speech-sample stimuli were used in the chi-square analysis due to the fact that only these ten utterances produced the number of NS segmental commentary tokens needed to conduct the chi-square test.

**Chi-square analysis.** A chi-square or “goodness of fit” test was conducted on each of the utterances in order to determine whether the distribution of NS comments on a particular segmental was what was to be expected by chance, based upon the learner’s non-target productions from that sample. To conduct this analysis, the transcripts of learner utterances were referenced to determine an expected distribution for each utterance. For example, in SS01, the learner produces [sɛˑ.lɛˑ.bɒn.tʰəˑ.lɛˑ.nЮ.wɛˑ.vɛˑ] instead of the native pronunciation [sɛˑ.lɛˑ.bɒn.tʰəˑ.lɛˑ.nЮ.βɛˑ]. The distribution of nonnative consonant segmental production is as follows: bilabial fricative [β] is pronounced in a non-target manner twice and dental stop [t̪] once. If all segmental errors were equally salient to NS listeners, it would be expected that 67% of consonantal segmental observations would reference [β], while 33% would reference [t̪]. This
expected underlying distribution was compared to the observed distribution from the consonantal segmental commentary. In this case, the observed results differed from the expected: 79% of the NS commentary focused on pronunciation of \([\ddot{k}]\), whereas 21% focused on \([\ddot{p}]\). The chi-squared analysis showed this difference to be statistically significant, meaning that the NS observations differed significantly from the observations that would be expected should all nonnative production of consonantal segmentals be equally salient to the NS listeners.

Results for the chi-squared analysis for each of the utterances are shown below in Table 4.10. All results are corrected for continuity. Utterances that had less than fourteen consonantal segmental commentaries were only considered significant at the \(p < .01\) level. Utterances with less than ten consonantal segmental commentaries were not included in this analysis would not provide reliable chi-square results due to the low number of items (commentaries). This necessitated exclusion of ten of the twenty original utterances.
Table 4.10

*indicates that p is significant at the .05 level
**indicates that p is significant at the .01 level
***due to a reduced number of cases, these results were only considered significant for p < .001

<table>
<thead>
<tr>
<th>Speech sample</th>
<th>Over-aspirated plosives /p t k/</th>
<th>Unspirantized allophones of /b d g/</th>
<th>Both phonemic and manner differences</th>
<th>Velarization of /l/</th>
<th>chi square</th>
<th>$p$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td>SS01</td>
<td>exp. 33.3%</td>
<td>exp. 66.3%</td>
<td>exp. 21%</td>
<td>exp. 10.94</td>
<td>.0009**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>obs. 79%</td>
<td>obs. 21%</td>
<td>obs. 21%</td>
<td>obs. 10.94</td>
<td>.0009**</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SS02***</td>
<td>exp. 50%</td>
<td>exp. 50%</td>
<td>exp. 33.3%</td>
<td>exp. 9.13</td>
<td>.0104*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>obs. 70%</td>
<td>obs. 30%</td>
<td>obs. 78%</td>
<td>obs. 9.13</td>
<td>.0104*</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SS04</td>
<td>exp. 33.3%</td>
<td>exp. 33.3%</td>
<td>exp. 33.3%</td>
<td>exp. 19.73</td>
<td>.0001**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>obs. 19%</td>
<td>obs. 4%</td>
<td>obs. 78%</td>
<td>obs. 19.73</td>
<td>.0001**</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SS08</td>
<td>exp. 33.3%</td>
<td>exp. 33.3%</td>
<td>exp. 33.3%</td>
<td>exp. 23.37</td>
<td>.0001**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>obs. 18%</td>
<td>obs. 4%</td>
<td>obs. 78%</td>
<td>obs. 23.37</td>
<td>.0001**</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SS09</td>
<td>exp. 20%</td>
<td>exp. 20%</td>
<td>exp. 20%</td>
<td>exp. 33.3%</td>
<td>.0007**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>obs. 6%</td>
<td>obs. 0%</td>
<td>obs. 31%</td>
<td>obs. 33.3%</td>
<td>.0007**</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SS10***</td>
<td>exp. 33.3%</td>
<td>exp. 33.3%</td>
<td>exp. 33.3%</td>
<td>exp. 14.6</td>
<td>.0007**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>obs. 0%</td>
<td>obs. 0%</td>
<td>obs. 10%</td>
<td>obs. 14.6</td>
<td>.0007**</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SS13***</td>
<td>exp. 25%</td>
<td>exp. 25%</td>
<td>exp. 25%</td>
<td>exp. 11.92</td>
<td>.0077**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>obs. 30%</td>
<td>obs. 0%</td>
<td>obs. 0%</td>
<td>obs. 11.92</td>
<td>.0077**</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SS17</td>
<td>exp. 33.3%</td>
<td>exp. 33.3%</td>
<td>exp. 33.3%</td>
<td>exp. 7.43</td>
<td>.0244*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>obs. 0%</td>
<td>obs. 0%</td>
<td>obs. 0%</td>
<td>obs. 7.43</td>
<td>.0244*</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SS20***</td>
<td>exp. 16.6%</td>
<td>exp. 16.6%</td>
<td>exp. 16.6%</td>
<td>exp. 6.8</td>
<td>.2359</td>
<td></td>
</tr>
<tr>
<td></td>
<td>obs. 30%</td>
<td>obs. 0%</td>
<td>obs. 10%</td>
<td>obs. 6.8</td>
<td>.2359</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SS25***</td>
<td>exp. 25%</td>
<td>exp. 25%</td>
<td>exp. 25%</td>
<td>exp. 14.09</td>
<td>.0028**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>obs. 18.2%</td>
<td>obs. 0%</td>
<td>obs. 0%</td>
<td>obs. 14.09</td>
<td>.0028**</td>
<td></td>
</tr>
</tbody>
</table>
The top row of Table 4.10 shows the non-target segmental consonant productions from the ten speech samples used in the chi-square analysis. These ten consonants have been organized into four main groups reflective of the non-target way in which each sound was pronounced: (1) over-aspirated productions of plosives /p t k/, (2) unspirantized allophones of /b d g/, (3) productions with both phonemic and manner differences and (4) /l/ velarization.

A blank column under a specific consonant signifies that there were no non-target productions of that consonant for that particular sample. For example, in SS01 (se levanta a las nueve), there was one non-target production of [t̪] and two of [β] but none of any of the other possible consonants. Each speech sample has a row dedicated to the expected (“exp.”) distribution of commentary for each non-target segmental production present in the sentence. The exp. (“expected”) row for this speech sample indicates the expected proportions of NS commentary for each sound: a purely chance distribution for SS01 would have 66% of the commentary on [t̪] and 33% on [β]. Below the expected distribution the observed (“obs.”) distribution demonstrates the actual proportion of NS commentary dedicated to each non-target production. For example, though non-target productions of [p] comprised one-third of the non-target consonant productions in SS10, none of the NS listeners commented on it. A p-value of less than .05 indicates that there is less than a 5% probability that the differences between the expected and observed NS commentary distribution were due to chance alone.

Of the ten speech samples analyzed, eight presented statistically significant results for the chi-square test of the distribution of the consonantal segmental non-target pronunciations identified by the NS listeners. This indicates that these listener
observations could not be attributed to chance alone, since they did not show “goodness of fit” with the expected pattern of identified non-target pronunciations derived from the speech samples.

A pattern emerged from the observed distribution in the NS-listener commentary: The NS evaluators followed a certain categorical order when describing the non-target productions that they perceived as most salient. With the exception of SS10 (discussed in Chapter 5), the distribution of the commentaries on the saliency of certain accented sounds always displayed the following descending order: (1) phonemic and manner differences, (2) over-aspirated plosives or /l/ velarization, and (3) unspirantized allophones of /b d g/. This meant that in any given speech sample, the greatest number of NS-listener segmental comments was dedicated to phonemic and manner differences, followed by over-aspirated plosives or /l/ velarization, and then followed by unspirantized allophones of /b d g/. For example, the NS listeners produced a total of twenty-two segmental commentaries for SS08 (limpia su carro). Of these twenty-two commentaries, seventeen cited the non-target production of /r/ (from carro) and five cited the over-aspirated plosives [kʰ] (from carro) and [pʰ]. The chi-square analysis of this speech sample showed that this observed distribution in the commentary could not be attributed to chance. Though all four groups of sounds were not all present in each speech sample, all of the commentary distribution (except for SS10) was distributed in this order. For example, there are no non-target pronunciations from the “phonemic and manner” category in SS01, but there is a non-target production of an over-aspirated plosive [tʰ]. The non-target production of the over-aspirated plosive thus becomes most salient to the
NS-listeners, as is evidenced by the fact that it was the subject of 79% of the comments. This pattern and its implications are discussed in further detail in Chapter 5.

**Summary of Results**

The results for the investigation into the research questions presented in this study can be divided into (a) NS-listener Likert-scale evaluations of accentedness and attitudinal variables and (b) the analysis of the NS-listener commentary on the learners’ non-target productions. In regards to the former, the NS listeners showed very positive perceptions of both L1 AE speakers that are learning Spanish and of the United States in general. However, this positive skew was not found in the NS-listener Likert-scale ratings for accentedness, which were distributed along the normal curve. Correlation data for the Likert-scale accentedness and attitudinal variables showed a statistically significant negative ($p < .05$) correlation between the NS listeners’ perceptions of the United States and their Likert-scale accentedness ratings, indicating that NS listeners with more positive perceptions of the United States rated the learner speech as less accented. In addition to the Likert-scale data, the NS-listener participants from this study showed other indications of positive perceptions of L1 AE speakers learning Spanish, such as the number of friends they had from this population and the participants’ willingness to facilitate communication with L1 AE speakers who are learning Spanish.

The analysis of the 606 NS-listener commentaries of non-target productions from ten speech samples showed that commentary on the segmental aspects of learner speech was the most prevalent (302 comments). NS-listener commentaries on suprasegmentals (154 comments) and global aspects of learner pronunciations (54 comments) represented the other two major categories of NS-listener commentaries.
Chi-square analyses conducted on the individual speech samples showed that the distribution of the NS-listener comments on the non-target productions of the segmental consonants was not attributable to chance alone. Further investigation of the distribution of the NS-listener comments showed a distinctive hierarchical pattern underlying the data. Non-target productions that altered a sound phonemically and in manner were always identified first, when present. Over-aspirated plosives and /l/ velarization ranked second in this order. Unspirantized allophones /b d g/ were the last category in this hierarchy, meaning that these non-target productions were identified in the lowest proportions in the NS-listener commentary. These results and their theoretical implications are discussed in detail in Chapter 5.
CHAPTER 5
Discussion and Conclusion

This chapter is divided into two main parts in order to correspond with the two main sections presented in Chapter 4. First, the results from the Likert-scale accentedness ratings are considered in conjunction with those from the attitudinal Likert-scales, thus returning to the research question of “Which aspects of Spanish L2 learner extemporaneous speech do NSs most often identify as accented?” Second, the results for the chi-square analysis of the segmental aspects of the NS listener commentary are discussed in terms of the research questions of “Which aspects of Spanish L2 learner extemporaneous speech do NSs most often identify as accented?” and “Is there a hierarchical order in Spanish that ranks sounds in terms of their potential for contributing to NS perceptions of accent?”

NS-Listener Profile and Affective Factors

One of the overarching goals of this study was to present the NS listener as an integral part of L2 communication. Upon doing so, it becomes clear that NS perceptions of learner speech are influenced not only by characteristics found in the speech signal itself but also by the various affective factors unique to the NS listener. A strong body of research in the area of listener affective variables (see Chapter 2) shows that the NS interlocutor is not merely a means of gathering data on learner production, but rather a complex individual whose aggregate life experiences have formed a distinct set of biases and beliefs that have the power to impact the way in which “foreign” speech is perceived. The affective variables for the NS listeners examined in this study provide valuable
insight into the NSs’ perceptions of learner speech and need be taken into careful
cornerstone before analyzing the NS perception data of the speech itself.

The profile of the eighteen NS-listener participants is that of a group displaying
primarily positive attitudes towards the L1 AE-speaking learners. As shown in Tables 4.6
and 4.7, the majority of the NS listeners indicated that they had a high number of
American friends and displayed a great degree of willingness to facilitate communication
with a learner whose Spanish was heavily accented. The latter of these two details is
especially striking, given the fact that the NS listeners completed the inventory on
affective factors after the somewhat tedious task of providing their detailed evaluations of
heavily accented learner speech. When asked directly about their perceptions of NSs of
AE, the evaluators’ scores averaged well above neutral: $M = 4.00$ ($SD = 1.08$) on a scale
that ranged from zero to five. The NS-evaluators’ ratings for their perceptions of the
United States in general were only slightly lower, with $M = 3.78$ ($SD = 0.81$) on a zero-
to-five scale.

Though the NS-listener profile could lead one to the conclusion that the NS
evaluators would be sympathetic to the learners and perhaps even forgiving in their
speech evaluations, this was not the case. The Likert-scale ratings for the level of
perceived accent presented in Table 4.1 show that the mean ratings for all four speakers
fell above neutral, denoting high levels of perceived foreign accent. Means ranged from
2.73 to 4.13, or just above the midpoint almost to the upper end of the 5-point scale. The
NS listeners’ consistently low ratings for accentedness in the NS-speech samples used as
controls (Table 4.3) show that this group of raters was not trending toward higher ratings
for all samples. This lack of leniency in accentedness ratings is further illustrated by the
distributions of the accentedness and attitudinal ratings: the NS-listeners’ average Likert-scale accentedness ratings (the overall average for each NS evaluator) were distributed along the normal curve, while the data for the attitudinal variables showed positive skew. For this reason it was necessary to use a Spearman’s Rho rank-order correlation between these variables instead of the more common Pearson correlation. In short, this group of NS listeners had positive perceptions of the learners but did not display a great deal of leniency in their ratings of the learners’ speech.

The correlation carried out between the attitudinal perception data and the accentedness ratings (Table 4.5) further explained the relationship between these variables. A statistically significant positive correlation was found between the two attitudinal variables of perceptions of the United States and perceptions of L1 speakers of AE. This correlation was not surprising; it is logical that individuals with a positive perception of a country would also perceive its people in a favorable manner. In addition, the statistically significant negative correlation between perceptions of the United States and accentedness ratings indicated that NS evaluators with more positive (higher-rated) perceptions of the United States also rated learner speech as less accented. Though the $r_s$ value for the correlation between speech-sample ratings and perceptions of L1 speakers of AE was not statistically significant at -.245, its high effect size ($d = -.435$) indicated that this correlation would likely become statistically significant with a larger sample size.

Based upon the above observations, a two-fold answer can be posited for the research question of “How do NS attitudinal variables correlate with the way in which they evaluate accentedness of learner speech, if at all?” The fact that the positive skew was present in the attitudinal ratings curve but not in the accentedness ratings curve
indicates that, for this group of listeners, favorable perceptions of and familiarity with learners from a shared L1 background does not automatically correspond to overwhelmingly lenient scores in the area of accentedness. However, there was a statistically significant correlation between positive perceptions of the United States and lower accentedness ratings. In addition, the high effect size for the correlation between perceptions of L1 speakers of AE and accentedness ratings suggests that a larger sample size would most likely produce a statistically significant correlation. In short, the NS listeners’ positively skewed attitudinal perceptions of the learners did not correspond with similar skew in their Likert-scale accentedness ratings, but positive affective perceptions did significantly correlate with lower (more native-like) accentedness ratings.

These results can also be interpreted in regards to the concept of “foreignness,” as presented in the Rubin (1992) and Cargile, Maeda, Rodriguez and Rich (2010) studies. In these previous studies (Chapter 2), a statistically significant relationship was found between the NS listeners’ perceptions of a given speaker’s “foreignness” and NSs’ accentedness ratings for that speaker. Findings from the present study show that NSs’ positive attitudes of the nonnative speakers and their country correlate with a decrease in the NSs’ ratings of degree of “foreign accent” for the nonnatives’ speech. It may be that that the NSs’ strongly positive attitudes towards the learners and their culture facilitated a greater amount of NS interactions with L1 speakers of AE, thus decreasing the NS listeners’ perceptions of the nonnative speakers’ level of “foreignness”. Support for this idea of a high degree of NS-NNS interaction can be found in the fact that the majority of the NS listeners reported having a relatively large number of L1 AE-speaking friends, even though most of these NS listener participants lived outside of the United States.
Last, the NS-listener affective data have positive practical implications for L1 speakers of AE learning Spanish, especially for those planning to study or work in Spain. The positive NS attitudes towards AE speakers and the NSs’ high number of AE-speaking friends suggest that learners of this profile could have ample opportunity to form friendships with NSs of Spanish. The potential for sustained communication in Spanish is further supported by the NSs’ reported willingness to communicate even when the learners’ speech is highly accented. Granted, this study has a limited NS participant pool, and the results presented here are not generalizable to all L1 AE speakers learning Spanish or to all NSs of Spanish interacting with L1 AE learners. However, such positive implications for NS-learner communication are certainly encouraging, even when presented on a smaller scale.

**Results of NS Observations of Non-Target Segmental Production**

The second major component of this study’s results addressed NS comments on segmental consonant production. As stated in Chapter 4, the results for the chi-square analysis of the listener observations of non-target segmental production presented a distinct pattern of the different categories of phones and phonemes. The ten sounds identified by the NS listeners can be separated into four groups, according to the ways in which learner production differed from target production. As shown in the results, NS-listener comments on the non-target segmental productions were not distributed randomly. Instead, they followed a hierarchical order. Non-target productions of the phonemes /ɾ r ʝ/ (when present) received the greatest proportion of comments in each speech sample. The second place on the hierarchy was shared by two categories: over-aspirated plosives /p t k/ and /l/ velarization. Non-target productions of unspirantized allophones of /b g d/
represented the lowest place in the hierarchy, meaning that these non-target productions were commented on in the smallest proportions. The following section discusses in detail each of these categories and their different manifestations within the speech-sample stimuli.

1. Sounds with phonemic and manner-of-articulation differences: /ɾ r j/. In Spanish, alveolar trill /ɾ/ and alveolar tap /ɾ/ contrast phonemically, as is evidenced by minimal pairs such as /ká.ro/ (“car”) and /ká.ɾo/ (“expensive”) or /pé.ɾo/ (“dog”) and /pé.ɾo/ (“but”). As is common of L1 speakers of AE learning Spanish, the speakers recorded for this study substituted Spanish /ɾ/ and /ɾ/ with the alveolar approximant [ɾ], which does not form part of the Spanish phonological inventory. In the ten sentences used in this analysis, the learners in substituted [ɾ] for both /ɾ/ and /ɾ/ five in five out of five possible environments, or 100% of the time. These non-target pronunciations have the potential to cause confusion at the phonemic level and proved to be especially salient to the listeners in this study.

Another sound whose non-target pronunciations presented differences at the phonemic level was the palatal /ʝ/, found in words such as /é.ʝa/ “she/her”. Following a pause, lateral or nasal, /ʝ/ is produced as its affricate allophone [ʝ], and is produced as [ʝ] in all other linguistic environments. The learner speech samples analyzed for the chi-square analysis presented four separate attempts at /ʝ/ by four different learners, but in none of the cases did the learners produce either allophone of /ʝ/. Instead, the sounds produced by the learners differed from /ʝ/ at the phonemic level. None of the non-target productions, however, resulted in [l], as is often common with beginner-level read-aloud speech data. Instead, the non-target productions of /ʝ/ presented here were all cases of
either complete ellipsis or vowel substitution. For example, in SS04 the learner produced [e.i.a] instead of the target pronunciation [é.ja], replacing /j/ with a syllabic /i/. As was the case with /r/ and /ɾ/, these phonemic-level non-target pronunciations were especially salient to the NS listeners.

2. Over-aspirated plosives: /p t k/. Though both English and Spanish have stops /p t k/, these sounds have a much longer VOT in many linguistic environments in English (see Chapter 1) than do their Spanish counterparts and are often represented in the English phonetic inventory as [pʰ tʰ kʰ], allophones of /p t k/. In addition, Spanish /t/ has a dental place of articulation ([ʃ]), whereas the place of articulation for English /t/ is alveolar. Though these differences can be salient to the naked ear, they are not phonemically contrastive in Spanish. The learners’ recordings contained pronunciations of plosives [p ŋ k] whose VOT was higher than 0.04 seconds, which is considered non-target (Lisker & Abramson, 1964; Moranski, 2007). Even though these non-target pronunciations lacked the phonemic contrastiveness of the non-target pronunciations of /r ɾ j/, they were frequently commented on by the NS listeners when they were present in the speech samples. For example, eleven of the fourteen segmental comments (79%) from SS01 were about the over-aspirated production of /t/.

3. /l/ velarization. The learner speech samples produced several examples of non-target pronunciations of the alveolar lateral /l/. In Spanish, /l/ is commonly referred to as “clear,” in that it has an alveolar point of articulation, towards the front of the mouth. While “clear /l/” is a component of the AE phonetic inventory, it also has a velarized allophone in various positions, depending on dialect, referred to as “dark /l/” or [ɬ]. L1 speakers of AE learning Spanish often produce the velar dark [ɬ] in place of clear [l], as
was the case in several of the learner speech samples examined in this analysis. As is the case for [pʰ tʰ kʰ] and [ɾ], dark [l] does not form part of the Spanish phonetic inventory. However, /l/ velarization and aspiration of plosives were of comparable saliency to the NS evaluators, while the dark [l] substitutions were noticed much less often by the NS listeners than the non-target pronunciations of /r̩ j/. This difference in saliency between the non-target productions of /l/ and those of /r̩ j/ is most likely attributable to a change in manner of articulation that takes place in the /r̩ j/ productions but not in those of /l/.

The non-target pronunciations of /r̩ j/ were all manifested in one of three possible ways: different phonemes which also form part of the Spanish inventory, ellipsis of the target phoneme, or substitutions of [ɾ] for both /r/ and /ɾ/. Of these three cases, only the third is comparable to /l/ velarization, in that neither the non-target phoneme produced as [ɾ] not dark [l] form part of the Spanish phonemic or phonological inventory. However, the substitution of [ɾ] for both /r/ and /ɾ/ marks a change in manner of articulation to approximant from trill and tap, respectively. In contrast, the [l] velarization does not modify manner of articulation, which remains a lateral liquid for both [l] and [ɾ]. For this reason, /l/ velarization presented a less severe alteration than the differences in the rhotics, and instances of /l/ velarization are grouped separately from /r̩ j/ in this analysis. Since neither /l/ velarization nor aspiration of /p t k/ denote a change in manner, it stands to reason that the over-aspirated productions of /p t k/ and /l/ velarization were of comparable saliency but were both much less salient than the non-target productions of /r̩ j/.

4. Unspirantized allophones of voiced plosives /b d g/. Spanish phonemes /b d g/ have the allophonic realizations [b ɾ̩ g] and [β ð̩ ɣ], dependent upon linguistic
environment. For example, plosive /d/ is realized as its allophonic fricative [ð] in any position other than those following a pause, nasal or lateral. As is shown in the analysis of learner speech samples (Tables 3.1 and 3.2), the utterances analyzed for the chi-square analysis did not contain any instances in which any of the four learners produced target-like lenition of the allophones of /b d g/ as required by linguistic environment (or otherwise). Instead, the learners produced the plosives in environments that required the spirantized allophones. These differences were not phonemic but were still identified by the NS listeners in three of the ten speech samples.

Hierarchy of non-target pronunciations salient to NS listeners. Figure 5.1 illustrates the application of the categories presented above to the results of the chi-square analysis (Table 4.10).

![Hierarchy of Saliency of Non-target Segmental Consonantal Productions](image)

Figure 5.1 *Hierarchy of Saliency of Non-target Segmental Consonantal Productions*

When present in the speech samples, phonemic-level non-target pronunciations of /r r j/ were identified by the NS listeners in the highest proportions. Though the individual
phonemes from this category were commented on in varying proportions, their non-target productions were always the most frequently identified for any particular speech sample. In the chi-square analysis, observed identifications of the phonemes from this category were higher than expected in all cases, and in some cases observed values nearly tripled expected values (SS09, SS25).

When this pattern first emerged, the very high number of observations of non-target pronunciation of [ð] in SS10 appeared as a counterexample to this hierarchy: this pronunciation was identified in the same high proportions that were present in the phonemic differences of /ɾɾʝ/. The pattern present from the other speech samples showed that the unspirantized allophones of /b d g/ were the least often identified by the NS listeners, and in some cases were not identified at all. Upon closer investigation of this speech sample’s transcription, it was discovered that this non-target pronunciation of /d/ should not be classified in the same manner as the other instances of unspirantized allophones of /b d g/ because it was not in fact perceived by the NS listeners as an unspirantized allophone, though it had not been transcribed in this way. In SS10, the learner did not produce the expected [d] but rather the “flap” or single-tap allophone [ɾ] of AE /d/. This was due to the non-target sound’s very short duration, as well as its misplaced alveolar place of articulation (/ɾ/ and /ɾ/ are the only voiced Spanish phonemes with an alveolar place of articulation). The learner had not produced the unspirantized allophone of /d/, but rather flap [ɾ]. This was most likely due to transfer from the learner’s L1, American English, where intervocalic alveolar flapping commonly changes intervocalic /t/ or /d/ to [ɾ], such as in the pronunciation of [bʌɾɹ] for “butter.” In the case of SS10, the learner had produced the intervocalic /d/ of nadar as [ɾ]. The NS
listeners’ perceptions of the sounds as a flap are made clear by their commentary: “la ‘d’ de ‘nada’ similar a una ‘r’” (the “d” of “nada” is similar to an “r”), “la letra ‘d’ de ‘nada,’ la ha pronunciado como una ‘r’” (the letter “d” of “nada,” it was pronounced as an “r”), and “‘nara’???” Since this substitution crosses the boundary between two phonemic categories of Spanish, it is included with the rhotic non-target pronunciations whose non-target productions denote a change in manner of articulation. The other non-target unspirantized productions of /d/ were reviewed, and no other instances of flapping were found.

The over-aspirated pronunciations of the plosives /p t k/ and velarization of /l/ producing dark [l] ranked second on the hierarchy. These non-target pronunciations were always identified less frequently than the phonemic group /ɾ ɾʝ/ but more than the unspirantized allophones (when present in a speech sample). Results indicated that the over-aspirated and velarized categorizes appeared to be in fluctuation with one another. In SS04, the non-target pronunciation of plosive /k/ was identified at a slightly higher percentage than velarized /l/, but the reverse was true of SS13. In SS10, NS listeners identified over-aspirated /p/ and velarized /l/ in equal proportions. Given the overlap present in these two categories, they are presented on the same “rung” in the hierarchy table in Figure 5.1. Though it is indeed possible that a larger sample size might better differentiate these two categories, the size and scope of the data set collected from the extemporaneous learner speech did not allow for such an analysis.

Also of note is that within the speech samples whose chi-square values were statistically significant, the observed non-target productions of /t/ always exceeded the expected values. This was not the case for the other over-aspirated plosives /p/ and /k/,
whose observed values were always lower than expected values in the samples with significant chi-square $p$ values. The prominence of commentary of these non-target productions of /t/ as compared to that of the other over-aspirated plosives is most likely due to the fact that the over-aspirated production of [tʰ] did not denote just a change in VOT, but in place of articulation as well. While Spanish [t] has a dental place of articulation, the [tʰ] produced by the learners has an alveolar place of articulation. This change in place of articulation in addition to the increase of VOT appeared to make the non-target productions of /t/ the most salient of the over-aspirated plosives. Though the small size of this data set did not allow for a definite hierarchy of salience to be proposed within the category of over-aspirated plosives, it is likely that a larger data set would provide more clear evidence for non-target productions of /t/ as being the most salient among the over-aspirated plosives.

The chi-square results for the over-aspirated plosives and velarized /l/ contrasted with those of the first phonemic group in terms of both observed and expected values. The phonemic difference in the /ɾ r ʝ/ group’s observed identifications by the NS listeners surpassed expected values in all instances; this was not the case with the over-aspirated plosives and velarized /l/. The observed values for the over-aspirated plosives only exceeded the expected values in three of thirteen possible cases (23%). In none of the cases did the observed identification values for velarized /l/ exceed expected values.

The effectiveness of the groupings and the respective ranking of each group is further illustrated in Table 5.2, which presents the same chi-square analysis conducted in Table 4.10 but with the expected and observed values for each of the four groups of sounds instead of for the individual sounds. This table also presents the non-target
attempt at /d/ in SS10 as belonging to the /ɾ ɾ ʝ/ group with both phonemic and manner of articulation differences. The same samples whose non-target pronunciation distributions produced statistically significant $p$ values in Table 4.10, in which distributions were analyzed at the level of the individual sounds, also have statistically significant $p$ values when the sounds are categorized into their respective groups.
Table 5.1

*Chi-Square Results for Free-Response Observations of Accented Consonants, Separated into Four Categories*

<table>
<thead>
<tr>
<th></th>
<th>Over-aspirated plosives /p t k/</th>
<th>Unspirantized allophones of /b d g/</th>
<th>Phonemic and manner differences in /r r j/</th>
<th>Velarized /l/</th>
<th>chi square</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>SS01</td>
<td>exp. 33.3% obs. 79%</td>
<td>66.7% 21%</td>
<td></td>
<td></td>
<td>10.94</td>
<td>.0009**</td>
</tr>
<tr>
<td>SS02***</td>
<td>exp. 100% obs. 100%</td>
<td></td>
<td></td>
<td></td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>SS04</td>
<td>exp. 33.3% obs. 19%</td>
<td>33.3% 69%</td>
<td>33.3% 12%</td>
<td></td>
<td>9.13</td>
<td>.0104*</td>
</tr>
<tr>
<td>SS08</td>
<td>exp. 66.7% obs. 22%</td>
<td></td>
<td>33.3% 78%</td>
<td></td>
<td>17.9</td>
<td>.0000**</td>
</tr>
<tr>
<td>SS09</td>
<td>exp. 20% obs. 6%</td>
<td>40% 94%</td>
<td>40% 94%</td>
<td></td>
<td>19.47</td>
<td>.0000**</td>
</tr>
<tr>
<td>SS10***</td>
<td>exp. 33.3% obs. 0%</td>
<td></td>
<td>33.3% 90%</td>
<td>33.3% 10%</td>
<td>14.6</td>
<td>.0007**</td>
</tr>
<tr>
<td>SS13***</td>
<td>exp. 25% obs. 30.5%</td>
<td>25% 0%</td>
<td>25% 61.5%</td>
<td>25% 8%</td>
<td>11.92</td>
<td>.0077**</td>
</tr>
<tr>
<td>SS17</td>
<td>exp. 33.3% obs. 0%</td>
<td></td>
<td>33.3% 66.7%</td>
<td>100%</td>
<td>5.5</td>
<td>.0182*</td>
</tr>
<tr>
<td>SS20***</td>
<td>exp. 50% obs. 50%</td>
<td></td>
<td>33.3% 16.7%</td>
<td>40%</td>
<td>4.9</td>
<td>.0863</td>
</tr>
<tr>
<td>SS25***</td>
<td>exp. 25% obs. 18.2%</td>
<td></td>
<td>50% 9.1%</td>
<td>25%</td>
<td>13.91</td>
<td>.001**</td>
</tr>
</tbody>
</table>

*indicates that p is significant at the .05 level
**indicates that p is significant at the .01 level
***due to a reduced number of cases, these results were only considered significant for p < .0
Theoretical Considerations

**Markedess.** As Figure 5.1 and Table 5.1 make clear, the various non-target learner productions differ in their "perceptual weighting." However, the underlying cause of these distinctions, such as the one between the over-aspirated plosives /p t k/ and the unspirantized stops /b d g/, was not as clearly delineated. As applied to Eckman’s MDH, the unspirantized stops are “areas of the target language that…differ from ones in the NL and are more marked than those in the native language” and thus present the highest degree of difficulty to L1 AE speakers learning Spanish. (Eckman, 1977, p. 321) The notion that the unspirantized stops are problematic for learners is supported by the aforementioned research (Elliott 1995; Castino, 1991), as well as by the investigation of the learner speech samples in this current investigation. Yet the high level of difficulty that the learners apparently faced in producing these sounds was in no way reflected in the proportions with which the NSs commented on the non-target productions of /b d g/.

Conversely, the over-aspirated productions of /p t k/ were identified in much higher proportions than those of the unspirantized stops. Under Eckman’s MDH, the over-aspirated stops would possess the smallest degree of relative difficulty to the NS-AE learner since “those areas of the TL which are different from corresponding ones in the NL, but are not more marked than the latter will not be difficult to acquire.” (p. 321)²³ This notion is supported by previous research (Moranski, 2007) and to a small degree by the fact that in this study [ŋ] was the only non-nasal non-lateral that had any instances of target-like production in the learner speech samples. As was the case with the unspirantized stops, degree of production difficulty did not correspond with salience from the perspective of the NS listeners. Though the over-aspirated stops presented less
production difficulty to the learners and a few native-like tokens were produced in the speech-sample stimuli, the NS listeners identified the non-target pronunciations of the over-aspirated stops in much higher proportions than those of the unspirantized stops. As shown in Table 4.10, when an over-aspirated stop and an unspirantized stop appear within the same utterance, the NS listeners always identified the over-aspirated stop more frequently (SS01, SS09, SS13, SS25). In this case, markedness was an effective predictor of learner difficulty but not of perceived accent.

The motivation for the greater saliency of the over-aspirated plosives is likely due to allophonic distribution in Spanish. The non-target pronunciations of /p t k/ produce the phones [pʰ tʰ kʰ], which are not part of the Spanish phonetic inventory. Thus, it is quite probable that the NS listener will hear these non-target productions as saliently different from the original phonemes but still have the ability to categorize them as variants of /p t k/. In contrast, the unspirantized pronunciations of /b d g/ produce [b d g], which do exist within the language’s phonetic inventory as allophones of /b d g/. As a result, an NS listener may still perceive the non-target pronunciation, but the resulting sounds would be familiar to them.

It is quite possible that the over-aspirated productions of /p t k/ and the velarization of /l/ were of comparable salience to the NS listeners because all of these non-target pronunciations produced phones that do not normally belong to the Spanish phonetic inventory.

Functional load. The weighting of the segmental comments can also be interpreted in terms of the functional load (FL) principle (Catford, 1987; Munro & Derwing, 2006). The non-target pronunciations that were the most salient were those that
differed phonemically from the target pronunciations and also changed in their manner of articulation. The sounds in this category had the greatest FL due to their potential for conveying semantic meaning. The fact that the phonemic and manner category far surpassed the other three in perceptual weight supports work done by Munro and Derwing (2006), which found a statistically significant positive relationship between NS accentedness ratings and the amount of high-level FL load errors within an utterance.

The results for the category with the lowest FL, the unspirantized allophones of /b d g/, also support previous research conducted on the FL principle. As mentioned in the previous section, [β δ γ] and [b d g] are both allophones of /b d g/ in Spanish. The allophonic nature of these phones indicates a relatively low FL, since they do not distinguish between words (a tenant of the FL principle mentioned in Chapter 2). The low perceptual weighting of these low-FL sounds reflects the results of Munro and Derwing’s (2006) study, which found no significant correlation between the number of low-FL errors in an utterance and NS-accentedness ratings.

**Pedagogical Implications**

It is of note that among the non-phonemic categories, the sounds that are most difficult for the learner to acquire ([β δ γ]) from a markedness perspective are also the sounds that rank last on the hierarchy presented in Figure 5.1. In contrast, the sounds that were less marked and posed less relative difficulty to the learners ([p ɹ k]) were always identified before the non-target pronunciations of [β δ γ]. This is potentially encouraging for L2 Spanish learners, in that the sounds whose non-target productions are more salient to the NS listeners are also the sounds that will present less acquisition difficulty.
Perhaps more striking than the markedness differences are the overall saliency of the phonemic non-target productions, especially those that included changes in manner and carried a high FL. Not only did these non-target productions rank highest on the FL hierarchy, but they also were consistently identified by the NS listeners in much higher proportions than expected in the chi-square analysis. Thus, if a language instructor has limited class time to devote to pronunciation instruction, it appears that these common high-FL, phonemic-level distinctions ought to take priority.

Limitations and Implications for Further Research

Though this investigation provides key insight into Spanish NS perception of learner speech, it is by no means exhaustive. Several of this study’s limitations and their possible solutions are discussed below.

Accent versus intelligibility. Perhaps the most apparent limitation of this study is that it examines NS-listener perception of learner speech in terms of nonnative accent, but not in regards to intelligibility. In the context of this study, nonnative accent is understood as the degree to which a learners’ production differs from that of a native variety. Intelligibility, on the other hand, “may be broadly defined as the extent to which a speaker’s message is actually understood by a listener.” (Derwing & Munro, 1995, p. 77) According to Derwing and Munro (1997), “accent and intelligibility are related but partially independent dimensions.” (p. 2) The implication is that high-levels of foreign accent do not always correlate with low levels of intelligibility (Derwing & Munro, 1995). This discrepancy between accent and intelligibility was seen in the pilot study of this present investigation: all 20 learner speech-sample stimuli consistently received perfect intelligibility ratings from NS listeners while the accentedness ratings varied. It was for
this reason that intelligibility was not introduced as a separate variable to be evaluated by the NS listeners.

Thus, this study should be seen as a first step in evaluating the components of L1 AE learner speech that are most problematic for communication with NSs of Spanish. Future research should also include data on intelligibility of learner speech. A possible direction for subsequent studies would be to examine whether or not the saliency hierarchy of non-target consonant productions (Figure 5.1) found for accentedness is also applicable to intelligibility. Since the learners who provided the speech samples used in this study were all in their second-semester of university-level language instruction or higher, future studies including intelligibility as a variable would most likely require speech stimuli from lower-level learners.

**Participant self-selection.** Given the nature of this study’s design, there is a possibility that some degree of self-selection occurred in the NS-listener participant population. Of the forty participants that began the online survey, only twenty-three finished, despite the fact that they received various reminder emails and were aware that they would not receive their honorarium until they completed the survey. Though the reasons as to why these participants did not finish the online survey remain unclear, there is a chance that they abandoned the survey due to an aversion to the task itself (that of listening to accented speech). However, since the attitudinal survey appeared at the end of the speech-rating tasks, it is impossible to know these would-be participants’ motivations for abandoning the study, and further inferences will not be made. In order to eliminate the possibility of self-selection in future research, the investigator could administer the study in-person to ensure the participants completed the online survey.
**Follow-up interviews.** Though the online survey tool proved tremendously effective in recruiting Spanish monolingual participants (a somewhat difficult task within the United States), it was limited in that its design did not allow for the researcher to ask any necessary clarification questions after the initial survey was conducted. In terms of category assignment for the commentary tokens, this meant that 13% of the commentaries could not be broken down into categories more specific than those of “word” or “syllable” (see Table 4.7), and an additional 4% were left as “ambiguous.” One means of rectifying this situation would incorporate brief, one-on-one interviews with the individual participants shortly after they completed the speech-evaluation task. While the participant completed the attitudinal survey questions, the researcher would have the opportunity to review the participant’s free-response data and prepare questions regarding ambiguous commentary. This interview could certainly be conducted in video format over the internet and would not greatly alter the design of the study.

**Limitations of speech samples.** Though the extemporaneous nature of the learner speech sample stimuli allowed for the samples to be more reflective of spontaneous production, it did present difficulties within the recordings themselves. Unlike other data elicitation procedures, such as the reading of minimal pairs, this task did not allow for the linguistic environment of the segments studied to be kept identical. Thus, it was difficult to control for the co-occurrence of other speech. Several of the segmental commentaries could not be included in the analysis for these reasons.

**Conclusion**

In this study, NS-listener perception is recognized as multifaceted with influences drawn from both (a) affective factors that have developed within the listener and (b) the
speech signal itself. Detailed analyses of the results from these two areas further illustrate the complex and nuanced nature of NS-learner perception.

The attitudinal data demonstrate that the NS listeners are not objective evaluators but rather individuals whose perceptions directly correlate with their judgments of learner speech. In addition, the NS-listener participants in this study were found to have strongly positive perceptions of the learners and of their home country. However, the positive skew found in the attitudinal data was not present in the numerical ratings of the speech samples. This indicates that although the NS listeners reported very positive perceptions of the learners, they were able to maintain some degree of objectivity when asked to analyze the learner speech. Though these findings have positive implications for the common SLA research practice of relying upon NS speakers for objective judgments of learner speech, one should be cautious about interpreting these results. Further research on the evaluations from NS listeners with negative and/or neutral perceptions of L1 AE-speaking learners is needed in order to investigate potential skew in numerical speech rating from these groups. A group of NS evaluators with more neutral or negative perceptions of the learners may rate their speech samples as being more accented. In addition, NS evaluators with less bias towards the learners could also be more sensitive to the non-target productions of the unspirantized allophones, which were rarely mentioned in the free-response commentary and in some cases were not mentioned at all by the NS evaluators (SS09, SS13, SS20).

On its own, the attitudinal data from this study have encouraging implications for language learners. The possibility of NS-participant self-selection has a positive interpretation: The NSs that chose to participate in this study are very likely to choose to
interact with actual learners when given the chance (as indicated by the NSs themselves in this study). Even though the learners’ speech was rated as possessing a strong AE accent, the NS listeners reported that they would attempt to communicate with the learners if presented with the opportunity.

The value of the in-depth analysis of NS-listener perceptions is undeniable. The data from this study are not collected from generic, nebulous NSs: they are collected from a group of NS participants with very positive perceptions of the learners and of the United States, whose speech-sample ratings did not share the same positive skew of the attitudinal ratings. In addition, it was found that speech-sample ratings significantly correlated with NS perceptions of the United States. It is within the context of these findings that the speech-sample ratings and commentary should be interpreted.

The free-response commentaries about the learners’ speech samples displayed the same degree of intricacy as the attitudinal data. In the categorical breakdown, the high percentage of both segmentals and suprasegmentals spoke to the salience of both elements in NS-listener perception of L2 phonology. The chi-square analysis of the consonantal segmental commentary showed that the distribution of NS listeners’ perceptions was not attributable to chance. A detailed analysis of the consonantal segmental data revealed that (in this data set) there is indeed a hierarchical order in Spanish that ranks sounds by their potential to contribute to NS perceptions of accent. The most salient non-target productions were those that involved a high-FL phonemic contrast along with a change in manner (/r r j/). These differences were not only reported in higher proportions as compared to the other sounds, but the chi-square observed differences were consistently much higher than expected for these sounds. Ranked
second were the sounds whose non-target productions were less divergent from target production than those of the first group but still manifested themselves as sounds that were not components of the Spanish phonetic inventory ([pʰ tʰ kʰ] for /p t k/ and “dark” [l] for /l/). Ranked last were the unaspirantized allophones of stops ([b d g] for /b d g/).

Notions of markedness-based difficulty were inverted for learner production and NS perception of the non-phonemic non-target pronunciations. Though, according to the MDH, the spirantized allophones of stops /b d g/ are more marked and therefore more difficult for NSs of AE, these sounds’ non-target productions were the least salient to the NS listeners. Even though the stops /p t k/ are considered less marked for NSs of AE and therefore present a lower degree of difficulty for NSs of AE according to the MDH, the over-aspirated productions of these stops were much more salient to the NS listeners. Put simply, in this study, markedness was a viable theory for predicting which sounds would be difficult for the learners to produce but it did not predict which non-target learners productions would be most salient to the NS listeners. This is most likely accounted for by the fact that the non-target productions of /b d g/ manifest themselves as phones that form part of the Spanish phonetic inventory ([b d g]), whereas the non-target productions of /p t k/ as [pʰ tʰ kʰ] are not part of the Spanish phonetic inventory. This finding implies that L1 speakers of AE learning Spanish will encounter a high degree of difficulty producing the spirantized allophones of /b d g/, but that these non-target productions have much lower “weighting” to NS listeners when compared to phonemic-level non-target productions and non-target productions of /p t k/. The low perceptual weighing of the unaspirantized allophones can most likely be attributed to their low FL.
The predominant result from the consonantal segmental analysis was the relative saliency of the non-target productions with phonemic and manner differences. Not only were these high-FL sounds identified in NS-listener commentary in much higher proportions than the other sounds, but also their chi-square observed values consistently well exceeded expected values. These results suggest that pedagogical priority should be given to the teaching of these sounds. However, certain sounds whose non-target productions manifest themselves as different phonemes with manner changes will be difficult for the learner to acquire if they are highly marked, such as is the case with trill /r/, which is universally marked. Regardless, the overwhelming saliency of these non-target productions for the NS listener should be taken into consideration when designing explicit pedagogical models for Spanish L2 phonology instruction.

Though this research in no way claims to be exhaustive, it provides a holistic account of NS perception of learner speech by devoting equal emphasis to the three key components of (1) affective variables of the NS listeners, (2) the composition of the learner speech sample, and (3) the actual NS perceptions. The results of this study showed that affective variables regarding NS-evaluator perception significantly correlated with the accentedness ratings that they gave to learner speech samples. Careful analysis of the speech samples evidenced that the distribution of the non-target phonetic productions within each sample significantly differed from the distribution of NS-evaluator commentary on these non-target productions. Further investigation into the NS free-response commentary showed that specific linguistic processes present in each particular non-target production, such as aspiration, lack of spirantization, or lateralization, corresponded to the proportion in which the non-target production was
commented on by the NS evaluators. Future research in this area should consider all three of these factors in order to promote a complete and holistic understanding of NS perception of learner speech.
Many scholars refer to this developmental period as the “critical period” (Johnson and

Throughout this study the term Second Language Acquisition (SLA) is used to refer to
the general area of study that encompasses both second language acquisition in an L2
setting and foreign language acquisition within a classroom setting. The majority of the
applications for the empirical research discussed in this literature review are, however,
intended for classroom foreign language acquisition.

Asher and García also found length of residency (LOR) to be positively correlated with
the subject’s L2 phonological competence. However, since the subjects were children, the
ones with the lower AOAs would also have higher LORs.

Earlier work had been done on the AOA question. In 1958, for example, Penfield and
Roberts asserted that the brain “stiffens” after age nine.

A case can be made that the plethora of choices for phonological instruction in the L2
classroom can be quite burdensome for an instructor that “does not know where to
begin.” For example, with precious little time available for phonological instruction, is it
of more value to the learners in a Spanish FL classroom to focus on the trill /r/ or rather
the stops /b d g/?

It should be noted that Elliott’s study also focused on the effect on certain individual
difference among learners, such as attitude or hemispheric dominance, as they correlated
to subjects’ acquisition of Spanish phonology.

Without continued phonological instruction (after the initial six months) the initial
improvements in the learners’ L2P was no longer apparent. These results support the
notion that L2P training need be an integral and continuous component of L2 instruction.
The learner’s perception is, of course, a central component of the SLA process. However, being that the current study limits itself to NS perception of NNS production, NS perception will not be included as an integral part of this study or its background information.

The authors employ the term “extemporaneous” and not the often-used phrase “spontaneous speech,” stating that speech samples gathered under experimental conditions are never truly “spontaneous” but rather may aptly be described as “extemporaneous,” in the sense that this speech is not rehearsed (Munro & Derwing, 1994).

The only statistically significant finding of the study was that NSs consistently rated later samples worse than the first samples they heard, even though the content was exactly the same. Munro and Derwing hypothesized that this could be in part due to listener familiarity with the speaker or with the content itself (once they heard the content the first time, they had more opportunity to detect pronunciation errors in the second stimulus).

“Scaffolding,” a term commonly associated with sociocultural theory, refers to the assistance provided to language learners that enables them to achieve a task that would otherwise be too difficult for them to complete independently.

“Pause group boundaries were identified as pauses in the speech signal, located perceptually by me and confirmed on the graphic representation of the speech signal provided by Sonic Foundry ACID Pro 3.0 software as zero amplitude.” (Zielinski, 2008, p. 73)
This is the approximate age of undergraduate students that participate in this university’s summer and semester-long study-abroad programs.

The two illustrations were taken from the Dos Mundos Edition 7 Spanish language textbook and depicted a routine day in the text’s characters’ lives.

dvf is a proprietary sound file used by Sony on digital audio recording devices.


Both /n/ and /m/ alternate in Spanish, dependent upon the sounds that follow them. For example, “un bebé” is pronounced [um beβe], with /n/ assimilating to the bilabial place of articulation of [b]. These Spanish allophones are generally not problematic for NSs of AE learning Spanish since English nasals also quite often assume the place of articulation of a following sound, as is seen in the example of “don’t be,” pronounced in fast speech as [dom.bi].

14 of the participants indicated that they best identified themselves as Hispanic or Latino, and three identified themselves as Mestizo (of mixed Indigenous and European descent). One participant’s response was excluded due to what appeared to be a misunderstanding of the question (he identified himself as all of the races listed).

Email address lists of potential participants were gathered by the researcher through various contacts both in the United States and abroad. Participant identifications were verified by these contacts to guard against identity fraud and against participants taking the survey a second time under an assumed name.
Although Flege’s body of research argued for a 9-point Likert scale for pronunciation ratings, Devellis (2003) posits that, for research in the social sciences, alpha values of much higher than 0.9 could be indicative of a scale that is in fact too long.

The AE alveolar approximant /ɹ/ also has retroflex and postalveolar regional variations. /ʝ/ has several dialectal variants, such as /ʒ/ and /ʃ/ in Argentine Spanish. The grapheme “ll” is also pronounced as /ʎ/ in (older) varieties of Peninsular Spanish.

Comparatively, aspirated AE stops are more marked than (unaspirated) Spanish stops because the presence of aspirated stops in a language generally implies the presence of unaspirated stops, but not vice-versa.
REFERENCES


**Vocabulario II (Repaso)**

These are possible word choices that could help you with your narration. Ask if do not know what any of these words mean or not know/remember how to conjugate them in the present tense.

<table>
<thead>
<tr>
<th>Verbos:</th>
<th>Sustantivos (nouns):</th>
</tr>
</thead>
<tbody>
<tr>
<td>despertarse</td>
<td>el periódico</td>
</tr>
<tr>
<td>leer</td>
<td>el perro</td>
</tr>
<tr>
<td>dormir</td>
<td>el carro</td>
</tr>
<tr>
<td>ir de compras</td>
<td>la lámpara</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>nadar</td>
<td></td>
</tr>
<tr>
<td>limpiar</td>
<td></td>
</tr>
<tr>
<td>bailar</td>
<td></td>
</tr>
<tr>
<td>jugar</td>
<td></td>
</tr>
</tbody>
</table>
**APPENDIX B: FULL SET OF SPEECH SAMPLE STIMULI**

<table>
<thead>
<tr>
<th>Sample</th>
<th>Speaker</th>
<th>Transcription</th>
</tr>
</thead>
<tbody>
<tr>
<td>SS01</td>
<td>low male</td>
<td>se levanta a las nueve</td>
</tr>
<tr>
<td>SS02</td>
<td>low male</td>
<td>toma un café</td>
</tr>
<tr>
<td>SS03</td>
<td>low male</td>
<td>almorzar con su amigo</td>
</tr>
<tr>
<td>SS04</td>
<td>low male</td>
<td>ella limpia su casa</td>
</tr>
<tr>
<td>SS05</td>
<td>low male</td>
<td>baila con su novio</td>
</tr>
<tr>
<td>SS06</td>
<td>low female</td>
<td>habla con su amiga</td>
</tr>
<tr>
<td>SS08</td>
<td>low female</td>
<td>limpia su carro</td>
</tr>
<tr>
<td>SS09</td>
<td>low female</td>
<td>ella habla con su amigo Ramón</td>
</tr>
<tr>
<td>SS10</td>
<td>low female</td>
<td>nada en la piscina</td>
</tr>
<tr>
<td>SS11</td>
<td>low female</td>
<td>limpia la lámpara en su casa</td>
</tr>
<tr>
<td>SS13</td>
<td>low female</td>
<td>mira la televisión</td>
</tr>
<tr>
<td>SS14</td>
<td>high male</td>
<td>la mujer se despierta</td>
</tr>
<tr>
<td>SS16</td>
<td>high male</td>
<td>lava su coche.</td>
</tr>
<tr>
<td>SS17</td>
<td>high male</td>
<td>ellos van a un restaurante</td>
</tr>
<tr>
<td>SS19</td>
<td>high male</td>
<td>va de compras</td>
</tr>
<tr>
<td>SS20</td>
<td>high male</td>
<td>mira un poco de televisión</td>
</tr>
<tr>
<td>SS23</td>
<td>high female</td>
<td>bebe el café</td>
</tr>
<tr>
<td>SS24</td>
<td>higher female</td>
<td>van a un resaurante</td>
</tr>
<tr>
<td>SS25</td>
<td>high female</td>
<td>ella baila con un amigo</td>
</tr>
<tr>
<td>SS26</td>
<td>high female</td>
<td>lee un periódico</td>
</tr>
</tbody>
</table>

*Shaded speech samples were used in the chi-square analysis of segmental data*