The relative degree of difficulty of L2 Spanish /d, t/, trill, and tap by L1 English speakers: Auditory and acoustic methods of defining pronunciation accuracy

Jeremy C. Waltmunson

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2005

Program Authorized to Offer Degree:
Department of Linguistics

UMI Number: 3163413

University of Washington
Graduate School

This is to certify that I have examined this copy of a doctoral dissertation by

Jeremy C. Waltmunson

and have found that it is complete and satisfactory in all respects,
and that any and all revisions required by the final
examining committee have been made.

Chair of Supervisory Committee:

Alicia Beckford-Wassink

Reading Committee:

Alicia Beckford-Wassink

Sharon Hargus

Ellen Kaisse

Date: 14 March 2005

University of Washington

Abstract

The relative degree of difficulty of L2 Spanish /d, t/, trill and tap by L1 English speakers: Auditory and acoustic methods of defining pronunciation accuracy

Jeremy C. Waltmunson

Chair of the Supervisory Committee:

Assistant Professor Alicia Beckford-Wassink
Department of Linguistics

This study has investigated the L2 acquisition of Spanish word-medial /d, t, r, ɾ/, word-initial /r/, and onset cluster /ɾ/. Two similar experiments were designed to address the relative degree of difficulty of the word-medial contrasts, as well as the effect of word-position on /r/ and /ɾ/ accuracy scores. In addition, the effect of vowel height on the production of [r] and the L2 emergence of the svarabhakti vowel in onset cluster /ɾ/ were investigated. Participants included 34 L1 English speakers from a range of L2 Spanish levels who were recorded in multiple sessions across a 6-month or 2-month period. The criteria for assessing segment accuracy was based on auditory and acoustic features found in productions by native Spanish speakers. In order to be scored as accurate, the L2 productions had to evidence both the auditory and acoustic features found in native speaker productions. L2 participant scores for each target were normalized in order to account for the variation of features found across native speaker productions. The results showed that word-medial accuracy scores followed two significant rankings (from lowest to highest): /r <= d <= ɾ <= t/ and /r <= ɾ <= d <= t/; however, when scores for /t/ included a voice onset time criterion, only the ranking /r <= ɾ <= d <= t/ was significant. These results suggest that /r/ is most difficult for learners while /t/ is least difficult, although individual variation was found. Regarding /r/, there was a strong effect of word position and vowel height on accuracy scores. For productions of /ɾ/, there was a strong effect of syllable position on accuracy scores. Acoustic analyses of taps in onset cluster revealed that only the experienced L2 Spanish participants demonstrated svarabhakti vowel emergence with native-like performance, suggesting that its emergence occurs relatively late in L2 acquisition.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude and appreciation to my advisor, Dr. Alicia Beckford-Wassink, who has shown me so much support and guidance over the past five and a half years. She has an uncanny ability to combine intellectual antagonism with social grace and tact, making her both an extraordinary researcher and mentor. She has been a role model for me in how one should strive for excellence in intellectual depth, integrity, clarity of reasoning, and personal character. The Department of Linguistics and its students are truly blessed to have her among faculty.

I would also like to thank Sharon Hargus and Ellen Kaisse for their insightful questions and suggestions to this research. Their expertise in phonology and their appreciation for acoustic study have made them fantastic professors to work with. I'd also like to thank Richard Wright for his role in teaching me about phonetics and for all his support. I feel very fortunate to have had so many kind mentors who have simultaneously provided me with challenging insights.

None of this research, especially the soliciting of volunteers, would have been possible without two brilliant mentors in the Department of Spanish. My career as a linguist started with Paloma Borreguero who first provided me with the opportunity of teaching Spanish. In addition, Maria Gilman has been a gem, both in her positive personality and her encouragement for this research.

I'd like to thank a few of my professors over the past few years: Karen Zagona, Heles Contreras, Fritz Newmeyer, Richard Wright, Toshi Ogihara, Soowon Kim, Jurgen Klausenburger, and Julia Herschensohn. I am honored to have had the opportunity to learn about so many aspects of linguistics from so many distinguished scholars.

I am very grateful to the students and instructors that volunteered for these studies. I hope that their participation and the results of this study have provided them with some insightful benefit.

It's been a pleasure to have had the opportunity to work the past few years at the Language Learning Center. Paul Aoki, Bob Majors, Larry Lesage, and Sherri Huber have been great colleagues and friends. Graduate work is more than just a dissertation.

I feel very fortunate to have made so many wonderful friends and colleagues in the department. The sharing of ideas, socializing, and getting people to laugh at my jokes has been the glue that has held my graduate work together. To Susie Levi, Chia-Hue Huang, Lesley Carmichael, Darik Olson, Benjamin Barret, Aixa Heller, Jeff Stevenson, Zack Bohnencamp, Laura Partanen, Sylwia Tur, Duane Blanchard, Anya Dormer, Annemarie Walsh, and Julia Miller (to name but a few) – thanks for being great people.

Last and not least, I'd like to thank my loving wife, Kymber Waltmunson, for her patience over these last few years. I promise that I won't work so hard on the weekends anymore. Unless it's in the garden, of course.

DEDICATION

To my father, Edwin S. Munson.
He always told me not to become
a doctor. But I'm pretty sure
he meant the other kind.

And to my mother, Priscilla A. Hardin.
Thanks for always encouraging
me to be an adventurer.

1. Introduction

In Spanish, word-medial /d, t, r, ɾ/ are contrastive as in *codo* 'elbow', *coto* 'enclosure', *corro* 'ring of persons', and *coro* 'chorus' (Harris 1969). Although a four-way contrast between the targets is less common, word-medial contrasts between pairs of the members of this set are widespread and often found in frequently used words (e.g., *carro* 'car'– *caro* 'expensive', *bota* 'boot' – *boda* 'wedding', *miro* 'I look at' – *mito* 'myth'). In American English, [ð, t, ɾ] are present in the language; however, the trill [r] is altogether absent. The [ð] and [t] are allophones of American English /ð/ and /t/, respectively. Typologically, taps, trills, and flaps are related to stops because they often alternate with stops, all include some form of closure, and "might be considered to be very short stops." (Ladefoged and Maddieson 1996: 47) Regarding orthography, Spanish <r> is realized as a tap (*pronto* 'soon', *pero* 'but') or trill (*rato* 'short while') depending on its syllabic position. In syllable-final position (including word-final), trilling is optional and <r> may be realized as either a tap or a trill (Harris 1969, Quilis 1981). In American English, orthographic <r> is typically realized as an approximant [ɹ].

The goal of this study is to examine the second language (L2) phonological acquisition patterns of Spanish /d, t, r, ɾ/ by adult first language (L1) American English speakers. It is hoped that the results of this study may further insights into the mechanisms of L2 acquisition and, specifically, the relative degree of difficulty that these related targets pose for the L1 English learner of L2 Spanish. This investigation establishes auditory and acoustic criteria, based on native-speaker productions, for assessing L2 production accuracy in order to score and compare the productions of the following L2 targets: 1) word-medial intervocalic /d, t, r, ɾ/, 2) word-initial trill /#r/ vs. word-medial intervocalic trill /VrV/, and 3) word-medial tap /VɾV/ vs. tap in onset cluster /#Cɾ/.

The stated goal of this study was achieved through a 2-part experimental phonetic investigation. Experiment A was designed to address the following three questions: First, *is there an acquisition order (i.e., relative degree of difficulty) of word-medial intervocalic* /d, t, r, ɾ/? It is well known that English speakers will flap orthographic <t, d> word-medially in English and that orthographic <r> is realized as an approximant; however, word-medial <d, t, rr, r> represent phonological contrasts in Spanish. The Spanish trill has often been cited as acquired late in L1 acquisition (Bosch 1983, Carballo et al. 1997, Gonzalez 1989, Jimenez 1987, Stoel 1974) as well as for L2 learners (Arteaga 2000, Bergen 1974, Elliot 1995a, Simoes 1996, Stockwell and Bowen 1965). The difficulty of L2 trill acquisition may result from the new burden placed on the L2 learner to acquire both a new articulatory gesture as well as a new abstract representation (Flege 1980). In contrast, Lado (1957) argued that learning an allophonic split, whose L1 allophones are overlapping and redistributed in L2, is the most difficult process in L2 phonological acquisition (i.e., more difficult than learning a new phoneme). For the current study, Lado's claim predicts lower scores for /d, t, ɾ/ than for /r/. The second research question addressed in Experiment A asked, *does trill accuracy emerge with higher scores word-medially over word-initial trill scores?* A trill can be identified word-medially by its orthography as in *churro* 'pastry'; however, the single grapheme <r> represents a trill word-initially as in *rico* 'rich.' Hence, this study may be able to reveal a correlation between spelling and accuracy scores. Differences in acquisition of trills by word position may also be predicted if L2 trill acquisition is affected by prosodic structure; Zampini (1998) found that L2 productions of intervocalic /b/ were more likely to be spirantized in word-medial position over /b/ in word-initial position. The third research question asked, *are tap accuracy scores higher in word-medial positions (e.g.,* pero *'but') than in onset clusters (e.g.,* pronto *'soon')?* Unlike the trill, orthographic <r> represents the tap word-medially and the tap in onset cluster. Typological study of markedness suggests that a two member onset is more marked than a single consonant onset (Greenberg 1978) which is supported by certain facts of L1 English acquisition

(Vanderweide 1994, cited by Archibald 1998). Hence, one would expect learners to receive higher scores for word-medial taps as opposed to scores for taps in onset cluster.

Experiment B was conducted to address three additional questions and to provide additional data for the questions in Experiment A. In Spanish, /t/ is unaspirated; however, /t/ in English is aspirated (Lisker and Abramson 1964). Thus, the fourth research question asked, *does the use of durational criteria in the assessment of L2 accuracy for* /t/ *affect accuracy scores?* Almost all L2 participants in Experiment A received greater accuracy scores for /t/ than for scores in other categories; however, the amount of aspiration in /t/ productions did not affect /t/ accuracy assessment, suggesting that L1 transfer of L1 English aspirated [t] could be transferred to the L2 Spanish target. Thus, a post-hoc analysis of durational measures for /t/ productions may lower accuracy scores for /t/ and yield a different relative ranking of /t/ scores. In effect, greater difficulty in the acquisition of /t/ may be found over other targets when durational criteria for /t/ are considered. Hence, the effect of word position on trill accuracy was retested in Experiment B. The fifth question asked, *does preceding vowel height affect scores for trill targets?* Vowel height was not varied preceding word-initial trills in Experiment A whereas vowel height preceding word-initial trills was varied in Experiment B. Thus, accuracy scores for trills in the environment of a low vowel are expected to be higher due to a more open oral cavity as well as the fact that [ra] would not place competing demands on one single articulator (i.e., the realization of [ri] involves the tongue tip for both [r] and [i]). The sixth question of this study asked, *is the emergence of the svarabhakti vowel a phenomenon that is observed relatively late in L2 Spanish acquisition?* The svarabhakti vowel is a vowel-like element realized between a consonant and tap as in the realization of the word *prisa* 'haste' as [p^{ə}ɾisa]. In Experiment A, very few L2 productions of taps in onset clusters revealed the emergence of the svarabhakti vowel despite its frequent presence in native speaker productions. Thus, a group of highly experienced L2 learners were included in Experiment B in order to assess the emergence of the svarabhakti vowel in their productions.

Regarding participants, Experiment A included L1 English speakers designated as "beginning," "intermediate," and "advanced" students of Spanish at the University of Washington. Participants were enrolled in Spanish throughout the study and were recorded in three sessions during a 6-month period. All beginning level participants were first recorded in Spanish 101 (first-year) while the intermediate and advanced level participants were first recorded at the 201 (second-year) and 301 (third-year) levels, respectively. Experiment B also included L1 English speakers of L2 Spanish designated as "beginning," "advanced," and "experienced" learners. The beginning and advanced learners were recorded in two occasions over a 2-month period while the experienced speakers were recorded in a single session. Participants in the experienced group had all studied Spanish longer than 10 years and all were Spanish instructors at the time of recording. In contrast to learners in other groups, the experienced speakers were only recorded on one occasion as there was no reason to suspect changes in their pronunciation over several months. The purpose of multiple recording sessions was to provide longitudinal information about students' development, strategies, and a more informative measure of acquisition development than a single recording session might reveal. In addition, in the comparison of scores across targets, the data will be analyzed inferentially by yielding an average score for each participant across session scores.

This study's research is important as it relates to the mechanism behind L2 phonological acquisition, the methods by which it is evaluated, and how its difficulties can be addressed. Several important issues arising in previous L2 research are discussed, including: When researchers discuss the difficulty of acquiring a trill (e.g., Carballo and Mendoza 2000, Elliot 1997), is it accurate to categorize trills in all word positions as being equally difficult to produce? If trills are acquired differently by word position, as the results of this study suggest, the evaluation and teaching of L2 phonological acquisition should take those differences into account. Similarly, if taps are more difficult to produce in onset clusters as opposed to in syllable-initial position, then universal markedness may be a factor in L2 as well as L1 acquisition. This study's results will enable researchers to better understand L2 phonological acquisition while language teachers may better recognize which sounds to emphasize in pronunciation instruction and become more adept at evaluating pronunciation development.

The methods of the current study differ from previous L2 pronunciation studies' methods on two levels. First, native speaker productions in the current study were auditorily and acoustically analyzed to provide a clearer basis for the definition and measurement of *accuracy* in L2 productions. Both auditory and acoustic criteria were used to score L2 accuracy. The use of both acoustic measures and auditory transcription is important, as the validity of auditory transcription alone, even for phoneticians, can be affected by psycho-acoustic factors and requires an explicit transcription strategy (Kerswill and Wright 1990). Second, the current study compares production accuracy in trills and taps by word position. Previous pronunciation studies have determined L2 accuracy based solely on auditory impressions (Elliot 1995a, Hammond and Flege 1989, Reeder 1997, Suter 1976) and have collapsed productions across word positions (Elliot 1995a). Several cross-language studies of Spanish vs. English stops have used acoustic measures (Flege 1986, Flege 1991, Gonzalez-Bueno 1997, Lisker and Abramson 1964); however, their analyses have been limited to word-initial productions.

The following paragraphs describe the ordering and content of the chapters of this thesis. In each chapter, discussion of the categories always follows the order /d/, /t/, /r/, then /ɾ/, for consistency in presentation.

Chapter 2 presents the background literature relevant to this investigation. Section 2.1 presents a phonological perspective of /d, t, r, ɾ/ in English (i.e., the learner's starting point) and in Spanish (i.e., the L2 target). Section 2.2 describes the concept of *interlanguage* and evidence for it in previous studies of the L2 acquisition of Spanish /d, t, r, ɾ/. In Section 2.3, factors known to affect successful L2 pronunciation and acquisition are discussed in conjunction with their predictions for the current study. Section 2.4 addresses previous approaches to defining L2 accuracy and argues for the use of acoustic measures in accuracy assessment.

Chapter 3 presents additional background that sets out the acoustic properties of Spanish /d, t, r, ɾ/ and previous approaches to their acoustic measurements. Section 3.2

concludes the chapter with a detail of this study's questions and the hypotheses that have arisen out of the literature review found in Chapters 2 and 3.

Chapter 4 presents the methods for two related experiments. Section 4.1 and Section 4.2 address the participant backgrounds, word lists, and recording procedures incorporated into Experiment A and Experiment B, respectively. In Section 4.3, the native speaker backgrounds are outlined. Section 4.4 details the acoustic measures that were taken and applied to each production. Statistical tests are detailed in Section 4.5.

Chapter 5 sets out the native speaker results and methods for L2 accuracy scoring. The dual presentation is necessary since the definition of accuracy is based on the productions of native speakers. Section 5.1 describes the auditory and acoustic features found in native speaker productions. For each target category, example waveforms and spectrograms are provided. In addition, overall measurement averages by experiment and gender are detailed. Based on the most common auditory and acoustic features, each subsection presents the required features for scoring an L2 production as accurate. Section 5.2 demonstrates the application of the accuracy criteria to scoring L2 productions. The section is divided among the target categories as described above for /d, t, r, ɾ/, with the exception of word-final <r> which was not scored for accuracy due to the extreme variability in the auditory and acoustic features of the native speaker productions. For each target category, the scoring of L2 productions is explained with reference to waveform and spectrograms found at the end of the chapter.

Chapter 6 presents the results for the L2 productions of word-medial Spanish contrasts /d, t, r, ɾ/. Section 6.1 describes the range of auditory and acoustic features found in the productions of /d/ and /t/, independently of accuracy scores. Variation types are identified based on the combinations of auditory and acoustic features. Section 6.2.1 compares the accurately scored productions of /d, t, r, ɾ/ by overall accuracy score rankings in order to address the question of overall degree of difficulty. The analysis shows significantly lower scores for /r/ and highest scores for /t/. The role of voice onset time for /t/ as an accuracy criterion is presented in Section 6.2.2. The analysis shows that some participant scores for /t/ are affected by the inclusion of a duration criterion;

however, a significant ranking of scores still demonstrates lowest scores for /r/ and highest scores for /t/.

Chapter 7 presents the results for the L2 realizations of /r/ targets by word position. In Section 7.1.1, productions are classified into variant types based on auditory and acoustic features. In Section 7.2.1, trills scored as accurate are compared by word-initial and word-medial position, revealing a significant pattern such that scores for word-initial /r/ are lower than word-medial /r/ scores. Section 7.2.2 addresses the question of the effect of vowel height on /r/ accuracy. It is shown that scores for /r/ are significantly lower when /r/ is in the context of a high vowel environment vs. scores for /r/ in a mid or low vowel environment.

Chapter 8 describes the results of the L2 /ɾ/ realizations by word position. Section 8.1 categorizes all productions of /ɾ/ into variant types based on different combinations of auditory and acoustic features. In Section 8.2, scores for /ɾ/ productions are compared by word position. Section 8.2.1 shows that accuracy scores are significantly ordered such that scores for /ɾ/ in onset cluster are lower than scores for /ɾ/ word-medially, suggesting an effect of word position on L2 tap production. Section 8.2.2 addresses the emergence of the svarabhakti vowel by participant level. The results show that only the experienced level L2 participants demonstrate svarabhakti vowel emergence with native-like performance.

Chapter 9 discusses the overall results of this investigation. Section 9.1 addresses this study's six questions and the results in light of their hypotheses. Section 9.2 addresses interlanguage development in the context of the L2 variants and their patterns. Section 9.3 discusses observed patterns of scores for Spanish level and gender. Section 9.4 discusses the implications of the results for L1 and L2 Spanish phonology. In Section 9.5, the benefits and implication of a two-tiered assessment of determining pronunciation accuracy are presented. Section 9.6 addresses the implications of the results for teachers and pronunciation instruction. Ideas for future related studies are presented in Section 9.7. A summary of this study is presented in Section 9.8.

2. Background

The goal of this chapter is to provide background that will help address the questions posed in this study.[1] The chapter begins with a phonological view of the starting and ending points that an L1 English learner must follow in acquiring L2 Spanish sounds. Between the starting and ending points of acquisition, the chapter presents a discussion of the development stage between the two points. Having established three stages in learner acquisition, the chapter then presents the factors that have been shown to affect successful acquisition of pronunciation which in turn allow for a variety of predictions for /d, t, r, ɾ/ degree of difficulty. The second half of this chapter is dedicated to this study's second goal, specifically, the justification of a two-tiered method of scoring L2 target accuracy. Thus, previous methods that have been used to assess a learner's location within the overall acquisition process are presented. Previous knowledge of the acoustic properties of the target sounds in question are also summarized as they relate to the methods used to assess accuracy in this study.

This chapter is organized into the following four sections: Section 2.1 addresses the phonological status of the sounds [ð, t, r, ɾ] in American English and Spanish and the phonological development that an L1 English learner must undergo in order to acquire the L2 Spanish targets /d, t, r, ɾ/. Section 2.2 presents the concept of *interlanguage* and describes previous studies of the acquisition of L2 Spanish /d, t, r, ɾ/. Section 2.3 discusses the factors that have been shown to successfully predict L2 pronunciation. In Section 2.4, previous auditory and acoustic approaches to assessing L2 pronunciation accuracy are described.

---

[1] See Section 3.2 for a summary of the questions and hypotheses.

2.1 The phonology of [ð, t, r, ɾ] in American English and Spanish

In order to understand how the L1 English speaker acquires L2 Spanish sounds, it is necessary to understand both a learner's starting and ending points in a successful acquisition process. This section begins with the analyses of underlying [ð, t, r, ɾ] from an L1 American English point of view in an attempt to answer the question, *what is the abstract knowledge of a learner at the starting point of the acquisition process?* As will be described in Section 2.3, L1 transfer into L2 has been identified as a starting acquisition process. The second question addressed in this section is, *what is the targeted abstract knowledge that should eventually be acquired by the L2 learner?* Thus, the L1 Spanish representations reflect the end point of successful L2 phonological acquisition.

Section 2.1.1 briefly presents the phonological status of [ð, t, r, ɾ] from an L1 American English perspective. The L1 Spanish representations of these targets are divided into two sections. In Section 2.1.2, Spanish /d, t/ have received little discussion in the literature as their underlying status is undisputed; however, the underlying nature of trills and taps in Spanish has been the focus of much debate. Therefore, Section 2.1.3 is separately dedicated to the phonemic issues and arguments surrounding Spanish trills and taps. Section 2.1.4 relates the preceding sections by suggesting the phonological process that a learner must undergo in the successful L2 acquisition of the Spanish targets.

2.1.1 Phonological status of /d, t, r, ɾ/ in American English

The process of transfer is well documented in L2 acquisition (see Section 2.3.1). The current study assumes that a beginning learner of L2 Spanish who speaks L1 English must to a certain extent fall back on L1 American English when first starting the acquisition process. Thus, a learner must create a new L2 phonological system based to some extent on underlying representations in L1 English.

In L1 English, the sounds [d, t] are underlyingly /d, t/ while the tap [ɾ] is an allophone of both /d, t/ (Chomsky and Halle 1968); however, the trill [r] is not an allophone of any dialect of American English.[2] The grapheme <t> is typically realized as a voiceless aspirated stop in word-initial position (e.g., *teen* [t^hin]) or as a voiceless unaspirated stop in onset cluster (e.g., *stick*). The grapheme <d> is normally realized as short-lag unaspirated stop word-initially (e.g., *dip*). Preceding an unstressed syllable, <t, d, tt, dd> can all be realized as a tap [ɾ], although the flapping rule in English in optional (Bloch 1941, Haugen 1938, Sharf 1960). The flapping rule has also been shown to be optional in formal elicitation tasks (de Jong 1998, Zue and Laferriere 1979). As noted by Haugen (1938), <t, d> can also surface as a tap before a stress syllable if the stressed syllable is part of the following word (e.g., *put on*, *what else*).

2.1.2 Phonological status of /d, t/ in Spanish

In most dialects of Spanish, /d/ is realized as a voiced dental fricative [ð] following a vowel (e.g., *madre* 'mother', *dado,* 'given'); however, [ð] may be voiceless when followed by a voiceless obstruent (e.g., *adquirir* [að̥kiɾir] 'to acquire') or when in final position (e.g., *amistad* [amistað̥] 'friendship') (Harris 1969)[3]. In more careful speech, /d/ is realized as [ð] in utterance-initial and word-initial position unless preceded by [l] or [n] (Harris 1969).[4] Quilis (1981) describes /d/ being realized as its stop variant [d] after a pause, lateral, and even a nasal.[5]

---

[2] For the purpose of brevity, all future references to English will reflect American English unless specified otherwise.

[3] Harris transcribes the voiceless variant of [ð] as [$ð^θ$].

[4] Harris labels more careful or formal speech as *Andante* while "moderately fast, casual, or colloquial" speech is labeled *Allegretto*; however, for the present analyses the distinction is not relevant since all /d/ targets appear word-medially are predicted to be realized as [ð] in any speaking rate.

[5] Although I can find no words with tautosyllabic /md/ sequences in Spanish, there are a dozen or so words that end in /m/. Thus, [md] combinations in normal speech could be found between words as in *Miriam dió un paseo* 'Miriam took a walk,' *álbum de música* 'music album,' or *currículum de history* 'history curriculum.'

Intervocalic /d/ has even been shown to have an allophonic tap [ɾ] variant in parts of the Dominican Republic and a small region in Colombia (Nuñez Cedeño 1987). This realization of /d/ as a tap only occurs in intervocalic position, regardless of stress. The resulting form is homophonous with normally occurring taps and its distinction from an underlying tap is difficult even for native speakers to perceive outside of a given context (e.g., *toro* [toɾo] 'bull' vs. *todo* [toɾo] 'all'). The allophonic tap realization of /d/ also emphasizes the particularly close relationship of stops, taps, and trills in Spanish.

The segment /t/ is typically realized as a voiceless unaspirated dental stop in all positions (e.g., *bo[t]a* 'boot', *[t]ono* 'tone', *a[t]le[t]a* 'athlete') except that voicing may be assimilated when the segment is followed by a voiced obstruent as in *a*[t^{d}]*mósfera* 'atmosphere', *é*[t^{d}]*nico* 'ethnic', or *fú*[t^{d}]*bol* 'soccer' (Harris 1969). Quilis and Fernández (1973) extend Harris's observation slightly further stating that /t/ can surface as [ð] in *any* post-nuclear position (e.g., *atlas* 'world' [að.las]). Quilis and Fernández suggest that this realization may be due to the Spanish's "tendencia a formar sílabas abiertas"(82), implying that [að] is more open than [ad]. Thus, it is suggested that [t] assimilates voicing due to a loss of articulatory tension rather than the presence of a following voiced consonant.

In sum, intervocalic <d> is typically realized across Spanish dialects as [ð] while intervocalic <t> is systematically realized as [t].

2.1.3 Phonological status of the tap and trill in Spanish

The Spanish tap and trill are concurrently discussed in this section due to the predictability of their distributions and because some phonological approaches have categorized them as having the same underlying phoneme.

The distribution of the Spanish tap (*la vibrante simple* 'simple vibrant') and trill (*la vibrante múltiple* 'multiple vibrant') is as follows: The tap can appear word-medially (*cara* 'face') and in an onset cluster (*pronto* 'soon', *cuatro* 'four'), but not word-initially.

The trill is found word-initially (*rico* 'rich') and word-medially (*gorra* 'cap'). In syllable-final position, <r> can be realized by a tap or trill. In sum, the tap and trill are in complementary distribution in word-initial onsets, in contrastive distribution word-medially, and in free-variation syllable-finally.[6] Furthermore, although /ɾ/ typically corresponds to a rapid occlusion of the tongue tip against the alveolar ridge (Quilis 1981), it can also be realized as an assibilated variant [ř] (Cárdenas 1958, Quilis and Carril 1971, Quilis 1981, Torreblanca 1984, Widdison 1998).

Quilis and Fernandez (1973) posit that the tap and trill are separate phonemes based on their distribution and their contrast found word-medially; however, in order to account for the neutralization in syllable-final position, the authors posit an archiphoneme /R/ syllable-finally. Thus, /R/ in syllable-final position can be realized as a tap, trill, or even a fricative-approximant variant. Many other investigations concerning the tap and trill have assumed their underlying distinction as well (Moreno de Alba 1972, Quilis and Carril 1971, Quilis 1981, Quilis 1993).

In contrast to Quilis and Fernandez, Stockwell and Bowen (1965) treat the medial contrast as underlyingly a doubling of the same phoneme /ɾɾ/. The authors support their argument by noting that some dialects of Spanish have other underlyingly doubled continuants as in /kanne/ for *carne* 'meat' and /kallos/ for the name *Carlos*; however, Stockwell and Bowen do not provide specific dialects for these forms.

Harris (1969, 1983) has also proposed that that the trill and tap are underlyingly geminate vs. singleton forms of the same phoneme, providing the most detailed arguments to their underlying status. Word-initially, <r> is underlying /ɾ/ and a strengthening rule is posited for its realization as a trill. In rhyme-final position, the strengthening rule applies optionally to underlying /ɾ/. Word-medially, a trill production is underlyingly two taps /ɾɾ/. Three rules are posited: First, both underlying taps are re-

[6] Quilis and Fernández (1973) point out 15 minimal pairs that are contrasted with word-medial tap and trill.

syllabified into two syllables. Second, the 2nd underlying tap is strengthened to a trill. Third, the 1st tap is deleted.[7]

Harris poses three arguments that the trill, in intervocalic position only, is underlyingly a series of two taps /ɾɾ/. The first argument is based on the future verb morpheme (e.g., [-ɾé] for the first person singular). For example, given the infinitive *hablar* 'to speak', the first person singular future form is *hablaré* (*habla* +[ɾ]*é*) 'I will speak'; however, given a class of infinitives, as in *poder* 'to be able to', the same morpheme is realized without the theme vowel as in *podré* (*pod* + [ɾé]). Harris argues that *querer* patterns in the same way with *poder* which effectively accounts for the realization of *querré* (*que*[ɾ]+[ɾ]*é*) 'I will want' as two underlying taps. The second argument relates to the plural formation of nouns and adjectives and, specifically, the occurrence of *–e* after the trill in singular noun form of *torre* 'tower'. Harris proposes that *–e* is present underlyingly after a majority of forms and not deleted when preceded by two consonants. Thus, *pan* [pan] 'bread' would be /pane/ underlyingly while *carne* 'meat' and *grande* 'large' retain the final [e] in their surface forms due to the preceding two consonants /rn/ and /nd/, respectively. Because *torre* is also realized as [tore], support is provided in favor of [r] as /ɾɾ/ underlyingly. Harris's third argument relates to the absence of antepenultimate stress on Spanish nouns when the penult is a heavy syllable. For example, proparoxytones are found in Spanish nominals as in *fábrica* 'factory', *fórmula* 'formula', and *lágrima* 'tear'; however, proparoxytones with a heavy penult are not found (e.g., **te.lé.fos.no*) and are judged ungrammatical by native speakers.[8] The positing of the trill as /ɾɾ/ underlyingly (alongside its re-syllabification) helps to account for the observed phonotactics of proparoxytones in Spanish since there are no proparoxytones with a trill in the onset to the final syllable (e.g., **cá.ma.*[r]*a*),.

Nuñez Cedeño (1989) also argues for the trill as a series of two underlying taps with evidence from hypercorrection in Caribbean Spanish. While speakers will

---

[7] The deletion rule also prevents the additive realizations of *[ɾ#r] which would otherwise occur in *hablar rápido* 'to speak quickly'.

[8] However, the few exceptions include the Spanish place name *Frómista* and the English loan words *Wáshington* and *Mánchester*.

sometimes hypercorrect and insert /s/ into varying rhymes, Nuñez Cedeño observes that /s/ is never inserted into the rhyme of a penult when the antepenult is stressed (in support of Harris's observation that the grammar does not allow for proparoxytones with a heavy penult). Furthermore, Nuñez Cedeño points out that if the tap and trill were separate phonemes that one might expect clusters of a trill plus a tap which do not exist in Spanish, although, its absence can be accounted for by a deletion rule proposed by Harris.

In contrast to Harris and Nuñez Cedeño, Lipski (1990) argues against positing the tap and trill underlyingly as a singleton vs. geminate representations, suggesting that both the tap and trill are underlyingly the same phoneme and that separate realizations are due to a constrained syllabic template. Arguing against Harris, Lipski suggests that the lack of proparoxytones is a remnant of their absence in Vulgar Latin and attributes native speaker grammaticality judgments against these forms (e.g., *te.l.é.fos.no) to a lack of familiarity on the part of the native speaker rather than a violation of internal grammar. According to Lipski's proposal, the Spanish syllabic template only allows for two consonantal elements in syllable onset. Underlying /ɾ/ is posited to have a skeletal slot adjoined to its left in all cases where possible, except in word-medial tap forms which are marked to prevent the adjunction. The effect of Lipski's proposal is to suggest that syllable-final <r> must typically be realized as a trill. In sum, Lipski is able to differently account for the word-medial trill vs. tap contrast by suggesting that the medial tap is underlyingly marked to only allow one consonant element in its skeletal structure.

Bonet and Mascaró (1997) also suggest that the medial tap is underlyingly marked, although they take a different approach than Lipski. Bonet and Mascaró use sonority structure to account for the underlying nature of trills and taps. The authors propose that the tap is as sonorous as a glide, while the trill patterns with the sonority of an obstruent. It is argued that Spanish syllable formation follows Clements's (1990) theory of Core Syllabification (i.e., that syllables obey a rising falling sonority curve) as well as the Theory of Dispersion in which the first demisyllable (onset + nucleus) is maximal while the second demisyllable (nucleus +coda) is minimized. Hence, the trill always occurs when it is a single onset element because it better satisfies the Maximal

Onset Principle. Similarly, the /Cɾ/ cluster is preferred in order to appease syllable structure through the maximization of sonority rise in the onset. In effect, intervocalic flaps are marked as [+f]. Thus, except for intervocalic taps, all tap and trill realizations are underlyingly /R/ and realized based on their attempt to conform to Core Syllabification.[9]

In sum, the purpose of this section is not to resolve the debate over the underlying status of taps and trills in Spanish but rather to summarize the descriptions of these segments in the literature. Importantly, a representational difference must underlie trills and taps in medial position regardless of one's theoretical view. It is this representational difference that must somehow be acquired by the L2 learner of Spanish.

2.1.4 Phonological processes a learner must undergo in acquiring /d, t, r, ɾ/

To summarize the previous sections, American English [d] and [ð] are contrastive in distribution representing allophones of separate phonemes (e.g., *there* [ðɛɹ] vs. *dare* [dɛɹ]) while in Spanish they are allophones of the same phoneme (e.g., *dado* [daðo]). In many dialects of American English, the tap is an allophone of both /t/ and /d/ (e.g., *writer* [ɹaɪɾəɹ] and *rider* [ɹaɪɾəɹ]); orthographic <d, t> are phonetically equivalent for English speakers preceding an unstressed syllable (e.g., *edit, water, sanity*). The trill is altogether absent from American English while <r, rr> in any context is realized as an approximant (e.g., *starry, tree, rich, far*). Table 2.1 presents the Spanish targets investigated in this study and their underlying representations in L1 English and Spanish.

Thus, the Spanish contrasts pose a challenge to the L1 English learner of L2 Spanish, albeit in different ways. The English phoneme /ð/ needs to be reassigned as an

---

[9] One weakness of the proposal is that Bonet and Mascaró must posit a post-lexical rule of tensing to account for word-final trill realizations (that occur "in some varieties"); however, the implication is that word-final trilling is dialect dependent whereas my own impression is that the phenomenon occurs regardless of dialect. Furthermore, the authors admit that they cannot account for the absence of a trill in the onset of a final syllable of a proparoxytone (e.g., *cán.to.[r]a).

allophone of /d/ while English /t/ must not be realized in its flapped variant. In the case of [r], the learner must either create a new phonological category /r/ or associate a special feature onto word-medial /R/ in order to produce the trilled variant. For /ɾ/, the learner must disassociate the English allophone [ɾ] from /t, d/. This study will suggest the relative degree of difficulty of these processes and the influence of word position on L2 phonological development. Table 2.2 outlines the L2 targets and the acquisition processes that the L1 English learner of L2 Spanish must accomplish.

Table 2.1 Phonology of [ð, t, r, ɾ] in English and Spanish.

Category	English Phonology	English word positions	Spanish Phonology	Spanish word positions
[ð]	[ð]is an allophone of /ð/	word-initial: [ð]ese word-medial: fa[ð]er word-final: ba[ð]e	[ð] is an allophone of /d/	intervocalic: yo [ð]igo 'I say' comi[ð]o 'eaten' word-final: ciu[ð]a[ð] 'city'
[t]	[t] is realized as a tap between vowels.	preceding unstressed syllable: wa[ɾ]er hi[ɾ] it	[t] is an allophone of /t/ in all positions	word-initial: [t]ono 'tone' word-medial: pi[t]o 'string'
[r]	NA	NA	[r] is an allophone of /r/	word-initial: [r]opa 'clothing' word-medial: pe[r]o 'dog' syllable-final: ha[r]to 'full' habla[r] 'to speak' after /l,n,s/: al[r]ededor 'around' en[r]ollar 'to wrap' Is[r]ael 'Israel'
[ɾ]	[ɾ] is an allophone of both /t/ and /d/	preceding unstressed syllable: e[ɾ]it bo[ɾ]y hi[ɾ] it	[ɾ] is an allophone of /ɾ/	onset cluster: c[ɾ]espo 'curly' word-medial: toro 'bull' syllable-final: ha[ɾ]to 'full' habla[ɾ] 'to speak'

Table 2.2 Phonological acquisition of L2 Spanish /d, t, r, ɾ/ by L1 English speakers.

Category	Learner development from L1 English to L2 Spanish
[ð] - /d/	Learner must merge allophones (i.e., negative transfer or convergence) of different English phonemes /d/ and /ð/ into allophones of a single phoneme /d/ and avoid the flapped allophone of /d/.
[t] - /t/	Learner must disregard the English flapping rule (i.e., allow [t] to be produced intervocalically). Alternatively, one could also argue for the process of divergence in which [ɾ] and [t] must be disassociated from /t/.
[r] - /r/	Learner must create a new phoneme /r/ or acquire a special length feature to incorporate into /ɾ/.
[ɾ] - /ɾ/	Learner must make an allophonic split (i.e., negative transfer or divergence) of [ɾ] from /t/ and /d/ and associate [ɾ] with a new phoneme /ɾ/.

2.2 Interlanguage and variation

The acquisition of L2 sounds is a process which manifests through a range of systematic variation in attempted L2 productions. This process represents a learner's developing grammar and has been described by Selinker (1969) as "interlanguage." Interlanguage has also been described as an *idiosyncratic dialect* (Corder 1967), an *approximative system* (Nemser 1971), *language-learner language* (Corder 1978), and *evolving linguistic competence* (Flege 1980). For the purposes of the current investigation, this developing system will henceforth be referred to as interlanguage. Interlanguage is a series of evolutionary grammatical stages that are constructed while a learner attempts to acquire the target language. Thus, an interlanguage production is an attempt at an L2 target using the rules and processes of the interlanguage grammar. The purpose of this section is to address the previous studies on the interlanguage development of L2 Spanish [ð, t, r, ɾ]. For a summary of the L1 Spanish studies of /d, t, r, ɾ/, see Section 2.3.2.

2.2.1 Previous studies on the interlanguage of Spanish targets [ð, t, r, ɾ]

Major (1986) conducted research specifically addressing the interlanguage production of taps and trills. In his study, L1 English speakers (n=4) of L2 Spanish were recorded seven times over a seven week period. Major analyzed the L2 productions of five L2 targets: word-initial [#r], word-medial [VrV], and word-final <r>; post-consonantal [Cɾ] and word-medial [VɾV]. Although acoustic measures were not incorporated, productions were transcribed and labeled into one of three categories: transfer error, developmental error, or correct target. Productions labeled as "transfer" included any tap or trill production realized as an approximant or a trill realized as a tap. Developmental errors included any realization judged not to be the English tap, English approximant or target language form. The goal of Major's study was to asses his Ontogeny Model of language development which hypothesizes that transfer errors linearly decrease over time while developmental errors (i.e., interlanguage variations) increase, then decrease, over time.[10] The study results did indeed find in the case of taps and trills that transfer errors decreased over time while developmental errors increased, and then decreased, over the seven week period, lending support for the Ontogeny Model. Major's study provides insight into interlanguage existence, its forms, and variation in the L2 productions of taps and trills; however, we are not able to quantify the range of variation without more stringent transcription strategies and phonetic measures.

Although Major did not specifically point out the relation between the acquisition of taps and trills by word position, his data suggested an effect of word position on accuracy. Two subjects started off without any correct tap realizations and subsequently demonstrated correct tap realizations word-medially [VɾV] before correct tap realizations in onset cluster [Cɾ]. In the case of trills, the only subject evaluated as showing some correct trill productions systematically showed correct trills medially, before evidence of

[10] It seems as if this prediction could more simply be stated that L2 realizations progressively go from L1 forms, to IL forms, to native-like forms. Furthermore, without the use of acoustic measures, it is not clear to me if the forms labeled as developmental errors are actually a result of developmental processes or merely some sort of compromised form between the NL and TL targets.

correct target forms for trills word-initially. Major suggested that the tap and trill targets in medial position may have been easier due to "the advantage of already having an air flow present from the preceding vowel."(492) Major's explanation may suffice for the target words produced in isolation; however, some target words did appear in sentences read by the participants.

Zampini (1998) also found differences in interlanguage development of L2 Spanish stops by word position. In the analysis of the Spanish target /b/ by L1 English speakers, her results demonstrated that learners were more likely to spirantize /b/ word-medially than between word boundaries. Zampini's study suggests that one might expect different realizations of other segments as well, such as the trill and tap targets analyzed in the current investigation.

Munson (2001) analyzed the short term effects of formal pronunciation instruction on the pronunciation of L2 Spanish stops, taps, trills, and laterals. Production accuracy was assessed based on auditory and acoustic measures in order to assign a target accuracy score. Munson's analyses included intervocalic [ð, r, ɾ]; productions of [t] were also investigated but only in word-initial contexts. At the end of a nine-week intensive study, scores for [p, t, ð] remained significantly different from scores prior to formal pronunciation instruction. In particular, students showed the most difficulty in producing the trill, despite their own observations of knowing when they should be producing it. In the context of the current study, the results demonstrated a range of variation for L2 productions. Furthermore, Munson identified the difficulty that beginning learners had for pronouncing the tap and especially the trill.

The difficulty of acquiring the trill has long been noted by researchers of L2 Spanish (Arteaga 2000, Elliot 1995b, Stockwell and Bowen 1965).[11] Simoes (1996) provided his impressions of interlanguage trill and tap realizations by L1 English learners of L2 Spanish. In both word-medial position and in an onset cluster, the grapheme <r> was produced as a retroflex approximant which Simoes attributed to the process of language transfer (see Section 2.3.1). Part of the aim of this study is to analyze the

[11] See Section 2.3.2 for a discussion of the difficulty of acquiring the trill in L1 Spanish.

relative difficulty of the trill in relation to the contrasts of [d,t,ɾ] and to also analyze acquisition of trills and taps by word position.

In sum, previous investigations have analyzed the interlanguage production of the L2 Spanish targets [d, t, r, ɾ]; however, the contrasts have yet to be studied simultaneously. The results of previous studies also suggest that the trill may be most difficult in L2 acquisition. Furthermore, the results of Major and Zampini suggest that L2 native-like realizations may show an effect of word position. Thus, for the current study, trill scores are predicted to have overall lower scores than other targets; within trill targets, trill scores are predicted to have lower scores word-initially.

2.2.2 The role of variation in L1 and interlanguage

The goal of this section is to summarize previous approaches to variation in both L1 and L2 and their implications for variation in interlanguage. In the current study, the assumption is that native speaker performance is the learner's target. That is, if we want to determine if a learner has acquired native-like performance then we must first establish what it is that native Spanish speakers are doing under the same set of conditions (e.g., recording procedures). Thus, there is an attempt to infer the underlying form (i.e., competence) from the phonetic representation (i.e., performance). In a sense, one must work backwards from a surface form in order to understand the state of the underlying form.

From a traditional Generative view, as characterized by Chomsky and Halle's *Sound Pattern of English* (1968), it has been argued that underlying structure is best revealed through native speaker competence and not performance. That is, underlying phonological representations can best be identified through native speaker knowledge (e.g., intuition), rather than surface forms which may be altered by other linguistic components or universal language rules outside of the grammar (e.g., universal phonetic rules). While the identification of unknown underlying patterns argues for the study of competence (e.g., perception testing and grammatical judgments) and not performance, in the case of interlanguage, the underlying target structure is already known to the researcher. In the context of the current study, there is little debate (with the exception of the exact nature of taps and trills) to the underlying forms in Spanish and the rules that govern their surface forms. Thus, the comparison of L2 speech with native speaker speech, allows one to arguably gain insight into interlanguage structure.

One particular challenge in studying speech patterns is the variation found across forms that are considered to be the same underlyingly. In the *Sound Pattern of English,* the phonological component of the grammar contains phonetic-detail rules that operate on binary valued features to generate the phonetic transcription. Surface variation can result from optional rules within the grammar. As stated by Chomsky and Halle, "it is not necessarily the case that each deep structure determines a single phonetic representation; if the grammar contains optional rules or analyses, a given deep structure can underlie two or more phonetic transcriptions." (294) Alternatively, an aberrant surface form may be the result of a performance error. The difference between the two is that in one case a variant is the result of the component contained in the grammar while in the second case the variant is a mistake resulting from cognitive or gestural interference.

In the field of sociolinguistics, surface form variation across groups of speakers has been the center of investigation. In particular, sociolinguistic studies have correlated the systematic realization of particular phonetic variants of linguistic variables with gender, task formality, social situations, and the social groups with which speakers may identify themselves. As discussed by Labov (1972), the realization of a form can be treated as a function which includes sociolinguistic variables as its input. For example, the realization of /-ing/ as [ɪŋ] vs. [ɪn] has been shown to depend on socioeconomic class, contextual style, age, sex, and ethnic group (Labov 1966). Importantly, studies have shown that native speakers' knowledge of their L1 entails a knowledge of how to make sense of systematic variation. That is, listeners know how to relate surface variants to the same underlying linguistic form.

More recently, even approaches to Optimality Theory have attempted to account for variation in surface forms. Boersma and Hayes (2001) proposed the Gradual Learning Algorithm in which there is a stochastic (i.e., probabilistic) component to constraint rankings. In their proposal, rankings lie along a continuous scale as opposed to being categorical in nature. That is, "at every evaluation of the candidate set, a small noise component is temporarily added to the ranking value of each constraint, so that the grammar can produce variable outputs."(1) Rankings are considered to have Gaussian distributions over a portion of the constraint scale. Overlapped rankings not only help to explain the variation found in native speaker productions, but also explain intermediate wellformedness decisions by native speakers.

Whereas Contrastive Analysis has been able to ignore variation by merely focusing on the beginning and end stages of L2 acquisition, any study of interlanguage development has been forced to address the range of variant forms.[12] For example, Dickerson (1975) used transcription to identify the variants (i.e., the sounds produced for a particular target segment) that were realized by L1 Japanese speakers of L2 English in attempting to produce /z/. Her results suggested that all subjects nearly used the same set of variants in their productions. In addition, subjects similarly used the same variants in the same environments, and similarly changed their patterns depending on elicitation task. Dickerson's study demonstrated that behavior across subjects was consistent and that the interlanguage system must be modeled in terms of variable rules. In effect, Dickerson's results argued for the extension of the sociolinguistic variability model to the learning of second language phonology. Beebe (1980) also found an effect of speech style on the frequencies of different variants realized by L1 Thai speakers of L2 English. Furthermore, studies of interlanguage variation have not been limited to phonology; Tarone (1983) summarizes interlanguage data from syntax and morphology to show that interlanguage behavior varies in a systematic way with elicitation task. More recently, Young (1988) used a multivariate model to account for the variation found in the L2 English acquisition of the *–s* pluralization morpheme by L1 Chinese speakers.

---

[12] See Section 2.3.1 for a discussion of Contrastive Analysis.

In the current study, it is assumed that variation exists not only across L2 productions, but also in the productions of the native Spanish speaker participants. Although the formal elicitation task (i.e., reading from word lists) may reduce the amount of variability (as suggested by Tarone, 1983), as well as the potential production errors that might be found in more spontaneous tasks, even the native Spanish speakers in this study show intra-speaker variability (see Section 5.1). In addition, native speaker variability is different for different segments, and even different for the same segment in different word-positions. Thus, in the attempt to assess L2 acquisition, that is, to ask whether a participant has acquired the grammatical knowledge of a particular segment, it is necessary to compare the L2 patterns to the patterns of the L1 speakers under the same conditions.

In sum, variation has been found in both L1 and interlanguage. In a study of second language speech production, L2 segments and their variation must be compared to the productions of native speakers under the same set of circumstances. Relevant to this study, these conditions include task, segment word position, and syllable structure. In order to account for the variability found in the native speaker productions, L2 accuracy scores will be adjusted by a variability constant (see Section 5.2).

2.3 Factors of successful L2 acquisition: predicting target difficulty

This section presents the general body of cross-linguistic research that has identified factors affecting successful L2 phonological acquisition and pronunciation. Based on these factors, the end of this section aims to make predictions of the relative degree of difficulty in the acquisition of L2 Spanish /d, t, r, ɾ/. This section identifies many predictors that have been associated with successful L2 pronunciation even though not all factors presented in this section were able to be analyzed in this study. In reality, controlling for all factors would have been methodologically difficult and arguably impossible; however, identifying factors from previous studies may help to understand interlanguage development, rate of acquisition, and ultimate attainment.

The term "pronunciation" may appropriately be defined as a system of features which include segmental, voice-setting characteristics, and prosodic levels (Pennington and Richards 1986); however, in the current investigation, the segmental level is the focus of L2 acquisition and pronunciation. Nevertheless, this section includes evidence of factors that have been shown to successfully affect any level of L2 pronunciation.

2.3.1 Language transfer and universal markedness

Contrastive Analysis (CA) predicts that the ease or difficulty of second language learning can be predicted through the comparison of native (i.e., L1) and target language (i.e., L2) forms (Lado 1957). A *transfer* may be described as producing an L2 segment with one or more features from a similar segment in the L1. In addition, CA can predict both positive and negative transfers. An example of a positive transfer would be the transfer of L1 English /f/ to L2 Spanish /f/.[13] A negative transfer can occur in one of two ways: *convergence* is a negative transfer in which a learner must merge two L1 phones into a single L2 phone while *divergence*, or allophonic split, describes the process by which two allophones of the same L1 phone must be associated with 2 phones in the L2. For example, convergence describes the process in which a learner's separate L1 English /ð/ and /d/ must both be associated with L2 Spanish /d/. An example of divergence includes the case of the L1 English [d] and [ɾ] (allophones of English /d/) which must be associated with separate L2 Spanish phonemes /d/ and /ɾ/. It might also be claimed that the association of L1 English [t, ɾ] (allophones of English /t/) also falls under the process of divergence as [t, ɾ] must be diverged from their association with English /t/ and be associated with L2 Spanish /t/ and /ɾ/, respectively.

---

[13] It would seem that a positive transfer implies the accidental production of an L2 target by using an L1 sound; however, it appears impossible in such cases to distinguish between an L1 positive transfer and a successfully acquired L2 target.

Lado (1957) has claimed that divergence is not only more difficult than convergence, but also the most difficult learning process. As stated by Lado:

> [T]he kind of problem in which part of a phoneme in the native language can pass as a separate phoneme in the foreign language, and other parts of the same native-language phoneme pass as a different phoneme in the foreign language- that kind of problem is by far the most difficult one to overcome.(15)

Hence, investigations into target language learning have used the notions of CA, convergence, and divergence to predict target language errors (Tarone 1978). Corder (1978) suggested that transfer processes are in essence communication strategies used by the learner when the L2 has yet to develop (i.e., the learner has to fall back on the L1 when there is not yet competence in the L2). The idea of the L1 being the starting point for interlanguage has more recently been argued for by the Full Access/Full Transfer hypothesis (Schwartz and Sprouse 1996). In the context of the current study, if phonemic splitting is indeed the most difficult acquisition process, scores for /d, t, ɾ/ are expected to be lower than scores for /r/.

Eckman (1977) incorporated universal principles of markedness into CA in order to predict the relative difficulty of target language forms. He proposed the Markedness Differential Hypothesis (MDH) in which typological markedness and implicational universals can predict degree of difficulty for certain target environments. Specifically, the MDH is able to account for the seemingly different results of two particular studies. In the first study, the results are predicted by CA alone: Dinneson and Eckman (1975) showed that L1 German speakers learning L2 English had more difficulty learning the word-final voicing contrast when compared to another group of L1 English counterparts learning word-final devoicing in German; however, in a second study of L2 French, the results are not predicted by CA: Gradman (1971, cited by Eckman 1977) found that L1 English learners of L2 French had no difficulty learning word-initial [ʒ], even though /ʒ/ in English only occurs word-medially and word-finally. Thus, while CA could account for the results of the L2 German subjects, it could not predict the ease with which L1

English speakers learned the L2 French word-initial contrast. Eckman reasoned that the maintenance of the obstruent voicing contrast is more marked in English than it is in German because typologically, the word-final voicing contrast implies word-initial voicing contrast, but not vice-versa. Hence, the MDH predicts the results of both Dinnsen and Eckman (1975) and Gradman (1971).

Despite the strong influence of L1 transfer in early L2 development, a great deal many other factors have emerged in relation to predicting target language difficulties and contradict CA predictions. As pointed out by Nemser (1971), there are at least three arguments against hypotheses based on CA. First, different analyses yield different predictions. Ambiguous analyses using CA have emerged due to the independent effect of various linguistic levels on production and also due to the sometimes incongruent systems that are compared (e.g., English sound system and Bantu clicks). A second difficulty of the CA hypothesis is that it presupposes what Nemser refers to as the *blinding flash* fallacy which assumes that the target and L1 language come together completely and at the same time; however, exposure to the target language (i.e., L2 input) is a well-known gradual process. Finally, the third argument against CA hypotheses is that they must refer to interlanguage in order to predict interlanguage. Because the premise of CA is that future learning is predicted by current knowledge, each stage of interlanguage development contributes to the learner's total knowledge. In effect, the language learner "is no longer the pristine speaker of L1 assumed by dialinguistic analysis, but also the user of a more recently acquired system."(122) Thus, CA should (but does not) analyze both the current interlanguage stage as well as the L1 background in order to predict difficulties.

In sum, the while CA and the MDH have attempted to predict and account for target language difficulties and realizations, their predictions have only proven to be partially insightful into the range of interlanguage productions. For the current study, Lado's claims for CA would predict lower scores for /d, t, ɾ/ than for /r/ in contrast to the findings from L2 Spanish acquisition studies (presented in Section 2.2) which would predict lowest scores for /r, ɾ/; however, as the following paragraphs explain, a great many factors also correlate and help to predict interlanguage development and L2 acquisition success.

2.3.2 Universal tendencies of acquisition and in L1 Spanish acquisition

Tarone (1980) found evidence not only in support of language transfer and universal preferences, but also for the reactivation of L1 strategies in L2 learning. Tarone presented a pilot study that looked at how interlanguage syllable structure differed from structures of the target language. In her attempt to sort out the influence of L1 transfer vs. the universal open syllable preference for CV structure, Tarone analyzed the productions of L2 English by a variety of L1 Cantonese, Portuguese, and Korean speakers (n=6). The results showed that all speakers made errors that could not be attributed to language transfer. In addition, speakers reactivated L1 strategies such as deletion in the production of consonant clusters. Thus, for the current study, scores for taps in onset are predicted to be lower than scores for taps word-medially.

Macken and Ferguson (1987) suggested that some L1 learning strategies may be a result of L1 acquisition tendencies generalized across speakers of all languages. The authors pointed out that at least two universal acquisition tendencies have emerged from linguistic observations. First, stops tend to be acquired before nasals, nasals before fricatives, and fricatives before liquids. Second, voiceless sounds tend to be acquired before voiced sounds as demonstrated in L1 Hindi acquisition (e.g., voiceless unaspirated stops are produced before pre-voiced and aspirated stops. See Macken (1980) for a cross-linguistic perspective of L1 stop acquisition). If L1 learning strategies are reactivated by the L2 learner, and these strategies tend to be universally observed, then one can make predictions of targets that differ based on their phonological category, as in the current study. Therefore, one might predict for the current study greater accuracy scores for /t/ in comparison to /d, r, ɾ/ scores assuming the same mechanisms are functioning for L1 and L2 acquisition.

Macken and Ferguson's observation that liquids are acquired late in L1 holds true in the case of L1 Spanish. The Spanish trill, usually analyzed as a liquid, is often the last and most difficult sound for children to acquire. Bosch (1983) looked at all Spanish sounds acquired by 293 children aged 3-6 years old in an attempt to determine their mastery of Spanish sounds. Subjects were divided into five groups based on their age. Following a transcription of their productions (elicited by asking the children to describe pictures containing key target word items), the results reflected the percent error of each sound category production. For the present study, Bosch's results yielded several interesting observations: 1) percent error for trills (apart from diphthongs) were higher than for any other category across all subjects, 2) percent error scores for trills were higher than for percent errors for taps up until the age of six years old, and 3) both tap and trill accuracy did not reach 90% until the age of 7. In terms of correct articulations for [ð, t, r, ɾ], subjects first reached 90% accuracy for [t] (age 3), then [d] (age 4), and finally [ɾ] (age 7) and [r] (age 7). Other studies of L1 Spanish acquisition have shown similar difficulty in acquiring the trill and tap in relation to other sounds (Gonzalez 1989, Jimenez 1987).

The difficulty of liquid acquisition in L1 may also be compared to the difficulty of L1 English approximant acquisition. McGowan et al. (2004) also found effects for syllable position in the L1 acquisition of the English liquid [ɹ]. Eight children were recorded every 2 months up until age 26-31 months. Results showed that post-vocalic and syllabic [ɹ] were produced by the end of experiment, but not [ɹ] in prevocalic position. It was suggested that production of the approximant is easier in post-vocalic and syllabic positions as it only requires an oral cavity constriction while in pre-vocalic position, an additional constriction in the pharynx may be required. McGowan's results support previous findings by Stoel-Gammon (1985) in which L1 English [ɹ] was successfully produced word-finally first, and then later in word-initial position. Assuming L1 strategies are in effect for L2 acquisition (as suggested by Tarone 1980), McGowan's findings predict an effect of word position for the current study. Thus, scores for trills are predicted to be higher word-medially than scores for trills word-initially.

The difficulty of learning new motor control may also affect L2 acquisition (as well as suggest why L1 learners of Spanish acquire the trill relatively late). As pointed out by Flege (1980), the learning of a new phoneme requires the acquisition of abstract representation and new motor control. The "establishment of articulatory motor control is itself an important part of second language learning. The language learner, it seems, must acquire complex new sets of highly automatic articulatory gestures or modify existing patterns of phonetic implementation *in addition* to acquiring control of an abstract, reorganized phonology."(118) Thus, because the realization of the trill is arguably the only sound for the L1 English learner in this study that requires the acquisition of a new manner of articulation, we would predict that the trill will be most difficult for the participants in the current investigation.

In sum, L1 strategies of acquisition may predict degree of difficulty of L2 target sounds, assuming that the learners continue to have access to L1 acquisition processes. For the current study, scores of taps are predicted to be higher word-medially than scores in onset clusters. Based on L1 Spanish patterns, scores for trills are predicted to be lower than scores for [ð, t, ɾ] if L2 acquisition follows L1 Spanish development. The trill is also predicted to be most difficult due to its novel manner of articulation for the L2 learner.

2.3.3 Age, maturation, and the Critical Period Hypothesis

The effects of a learner's age in L2 acquisition has been greatly studied since Lenneberg (1967) found that children rather than adults were more able to fully recover language functions from certain types of aphasia. Based on the apparent correlation between neurological development and the difficulty of post-pubescent learners to overcome L2 accents, the Critical Period Hypothesis (CPH) was proposed (Scovel 1969) and has continued to be addressed (Patkowski 1990, Patkowski 2003). The CPH predicts that ultimate L2 proficiency is subject to an optimal period corresponding to biological effects. That is, adults can never fully acquire L2 pronunciation due to a loss of plasticity in the brain. The loss of neural plasticity accompanies a neural lateralization of language functions. As stated by Scovel:

> It is the nature of the human brain, not its nurture, that is essentially involved here – specifically, that the onset of cerebral dominance, which seems to occur around the age of twelve, inhibits the ability of a person to master the sound patterns of a second language without an impinging foreign accent.(245)

While many studies have attempted to test the CPH (e.g., Asher and García 1969, Oyama 1982, Scovel 1969, Snow and Hoefnagel-Höhle 1982b), more recent debate has concluded that the CPH cannot realistically be tested (Birdsong 2004, Flege 1986). Flege (1998) points out that the relationship between neural development and L1 acquisition is extremely confounded. Variables such as the age at which a learner arrives in the target country (AOA), length of residence in the target country (LOR), amount of L1 use, and amount of L2 use have also been shown to be confounded with each other. In effect, it may never be possible to control for all factors in the comparison of child vs. adult learners. For example, one difference between children and adults is that it is difficult to control for the amount of L1 use and the types of social situations in which children and adults use their L2. Cochrane (1977) administered a detailed survey to L1 Japanese speaking children and adults living in the US learning L2 English. The results revealed considerable differences between children and adults in that the children used English more outside the home and in a wider range of social situations than the adults. Thus, due to the confounding nature of age with so many factors and the limits on human subject neural testing, it has been claimed that the CPH is not a true hypothesis since it cannot be tested (Flege et al. 1997). It has also been argued that social and cultural factors may have more to do with post-pubescent inability to acquire native-like L2 pronunciation than does cerebral dominance (Hill 1970).

Despite inconclusive evidence for the CPH, the age at which a person first begins to learn an L2 has been strongly correlated with successful L2 pronunciation. Age of Arrival (AOA), or Age of Learning (AOL), has been found to correlate with the achievement of native-like performance. Krashen et al. (1982) reviewed many studies and concluded that AOA is the best predictor of eventual attainment. In Oyama (1982), the authors looked at 60 L1 Italian speakers of L2 English. Subjects were rated by two judges on a 5 point scale ('no foreign accent to heavy foreign accent'). Judges were "told to restrict their attention to phonological and prosodic aspects of the samples, [although] the possibility remains, of course, that other features of the samples may have influenced the judgments."(24) Each subject read from a paragraph and also spoke in a free elicitation task. The results found no effect of length of residence ("Number of Years in the United States") but effects were found of AOA. Thus, a linear relationship between AOA and degree of foreign accent was suggested.[14] In Flege and Fletcher (1992) L2 learners included 30 native Spanish speakers divided into three groups: Early L2 learners (AOL 5-6 years), Experienced Late L2 learners (averaging 14.3 years length of residence in the U.S. at testing), and Inexperienced Late L2 learners (.7 years average length of residence at testing). The results found that length of residence was potentially confounded with amount of L2 input and L1 use; however, AOL did indeed correlate with degree of perceived foreign accent such that subjects who had started learning English at age 5-6 did not have a foreign accent while those who started at age 7.6 years significantly differed from native English speakers. Chronological age, daily use of L2, and gender were not found to be significant. More recently, Yeni-Komshian et al. (2000) also found AOA effects in the analysis of L1 Korean speakers of L2 English. Subjects that had emigrated by age 6 to the U.S. showed no accent in their L2 English.

In the short term learning of L2 pronunciation, older aged learners may have an advantage. Snow and Hoefnagel-Höhle (1982a) looked at the pronunciation of L2 Dutch

---

[14] Some of Oyama's subjects were rated worse on a supposed more simpler paragraph reading task over the free-elicitation task. In order to account for this unintuitive result, Oyama suggested that these unpredicted results could by explained by the "Yerkes-Dodson law" in which a highly motivated subject may have increased performance on a relatively simple task yet worse performance on a more difficult task.

by L1 English adults and children. In a naturalistic L2 setting, adults showed short term superiority in L2 Dutch pronunciation over children; however, differences were not significant in the long term. It was suggested that adults conform less to peer pressure than children in a naturalistic setting and that inaccurate L2 pronunciation may reflect adults' desire to hold on to cultural-personal identities. The implication is that adult vs. child age differences may be confounded with attitudes and motivation.[15]

In the long term learning of L2 pronunciation, Asher and Garcia (1969) compared productions by Cuban immigrants to the U.S. with productions of native American English children. Subjects read from four sentences thought to be most difficult for native Spanish speakers (e.g., "I had two hot dogs and a glass of orange juice for lunch yesterday."). Subjects were allowed to rehearse the phrases ahead of time. The judges, all L1 English trained linguists, listened to the set of sentences and scored each speaker into one of four categories: A) native English speaker, B) near native speaker, C) slight foreign accent, and D) definite foreign accent.[16] Results showed that immigrants arriving to the U.S. past the age of 13 were most likely to be rated with a "definite accent" vs. immigrants who arrived before the age of 6 were most likely to be rated with "near native" pronunciation. This study has been replicated with Korean immigrants to the U.S. with similar results (Yeni-Komshian et al. 2000).

In sum, AOA and length of residence have been shown to correlate with degree of perceived foreign accent. The current investigation is limited to analyzing any effects of AOL across adult learners; however, it should be made clear that AOL is also confounded with Spanish experience in the current investigation. Thus, any effects for AOL may be inconclusive. In regards to the CPH, it cannot be formally tested in this investigation. As pointed out by Patkowski (2003), the CPH applies to informal and naturalistic learning environments so it is not relevant to discuss the CPH in terms of the learners in the current study.

---

[15] In addition, it is clear that one cannot ignore differences between cognitive development stages.
[16] Without further criteria, I'm not sure what the differences would be between category B) and C).

2.3.4 Influence of a learner's L1 on L2 perception and production

Although L2 participant perception is not tested in the current study, a brief discussion of perception is included in this chapter as a learner's perceptual experience has been argued to affect successful L2 pronunciation.

Kuhl and Iverson (1995) argued that our linguistic experience alters our phonetic perception such that perceived sounds are distorted due to a "magnet effect." The authors proposed the Native Language Magnet (NLM) model to explain altered perceptions of sounds which in turn affect production of sounds. Most importantly, the model attempts to explain why adults with differing native languages perceive sounds differently. According to the NLM, the mind is altered through experience with language. The NLM doesn't specifically make predictions about L2 production; however, the model predicts that difficulty in acquisition is a result of difficulty in the discrimination of sounds. Thus, the implication is that learners may have difficulty producing sounds because they have associated targets with their own L1 prototypes: "The magnet effects exhibited by prototypes of native-language categories render certain foreign-language contrasts less discriminable, making the acquisition of a second language in adulthood more difficult than the acquisition of a primary language."(149)

Best (1995) presented a different model of perception in which speech sounds are perceived in terms of gestural features. The Perceptual Assimilation Model (PAM) is intended to account for differences in the perception of L2 sounds; however, the model also predicts difficulties in L2 production. The implication of PAM for L2 learners is that increased perceptual experience in a foreign language leads to increased "attunement to detecting higher-order invariants of objects, surfaces, and events."(180) Thus, PAM predicts that pronunciation difficulties can result from the perceived gestural similarity of non-native sounds and contrasts with native sounds.

Flege (1995) also has argued for a correlation between perception and production of non-native targets. The Speech Learning Model (SLM) was proposed in order to account for the difficulty late learners face in pronouncing target sounds. The model claims that the correct perception of a target guides the correct pronunciation of the sound. While factors other than inaccurate perception can also affect production errors, in general, the SLM predicts a correspondence between advanced learner perception and production. For the current study, the SLM would predict that if an experienced learner accurately produces a target, then he or she will perceive native Spanish sounds similar to native speakers. Additionally, if an advanced learner inaccurately produces a target, then he or she will not perceive sounds like native Spanish speakers. More precisely, the inability of a learner to produce inaccurate targets is due to his or her inability to group classes of sounds into categories that are phonetically relevant.

In sum, the NLM, PAM, and SLM all suggest that L2 perception of sounds affects L2 production, although the processes are explained through differing models of perception. The NLM and SLM explain perception in terms of phonetic properties while PAM proposes that speech sounds are perceived in terms of gestural features. In the context of the current study, the predictions made by the models cannot be tested as this study focuses solely on L2 productions; however, the inclusion of only L1 English speakers in the study helps to control for L1 perceptual backgrounds and their potential effects on L2 production.

2.3.5 Learning and developmental strategies

Learning and developmental strategies within a learner's interlanguage have also been shown to affect L2 pronunciation. Selinker (1972) identifies several processes affecting L2 performance which function within the interlanguage grammar. *Overgeneralization* can occur after rules are learned with limited constraints, often in the early stages of rule acquisition. For example, students learning the past tense morpheme *–ed* may unknowingly apply the form to irregular past tense verbs. *Transfer of training* has been observed in which learners are so accustomed to the input of certain forms that they fail to produce distinctions even though they may be consciously aware of the differences (e.g., students learning English who fail to distinguish between *he/she* in their performance due to invariable textbook input of *he*). *Second language learning strategies* occur when the target language is reduced to a simpler system, as in treating all verbs as if they were transitive. A final process of s*econd language communication* manifests when forms are reduced or simply not used because they may not be necessary to the intended message being communicated. In addition to the above processes, Selinker identifies spelling pronunciations and hypercorrection as other processes which may affect L2 productions. For example, L1 English speakers have been observed to incorrectly pronounce the Spanish name *Jiménez* as *[dʒímɪnɪz] (Lado 1957). Nemser (1971) suggests the process of *reinterpretation* as another type of learning interference. Nemser provides anecdotal evidence of L1 Serbo-Croatian speakers of L2 English producing [t] in all positions with aspiration. Nemser describes this phenomenon as the phonematization of a non-distinctive feature in English.

Developmental errors such as *avoidance* have also been shown to characterize first and second language productions (Tarone 1978). On the L1 phonological level, Celce-Murcia (1977) found that a child simultaneously learning English and French substituted lexical items in an attempt to avoid pronunciation of the voiceless labio-dental fricative /f/. In L2 pronunciation, Weinberger (1987) found that adult L1 Chinese learners of L2 English avoided the L2 production of complex syllables in a story-telling task. The learners demonstrated higher error rates for complex syllable productions when required to produce the complex syllables in more formal reading task. Regarding syntax, Schacter (1974) concluded that L2 English speakers who had difficulty with relative clauses were avoiding them in their productions.

In sum, developmental and learning strategies may potentially affect individual learners at different stages of their L2 acquisition. Although their emergence may not be predictable, by recording learners on separate recording sessions over a period of input and instruction (e.g., 6 months in the case of Experiment A and 2 months in Experiment B), averaging results across sessions may more reliably represent target difficulty throughout the acquisition process. In addition, if spelling errors affect accurate production, higher accuracy scores for trills medially might be expected over word-initial trill scores as the graphemes of the former (<rr>) are more indicative of multiple articulatory contacts than the grapheme of the latter (<r>).

2.3.6 Instruction and input

The use of visual aids has also been shown to result in more native-like production of L2 sounds. Warsi (2002) examined L2 English productions of /l/ vs. /r/ by L1 Japanese speakers. An experimental group was shown diagrams of articulatory movements and given auditory feedback on the native-likeness of their productions while a control group was given no special training. Both groups were rated by 16 native English speaking judges using a seven point scale. The scores of the experimental group were rated significantly more like native speakers than scores for the control group. Interestingly, no differences were found between groups in the learners' ability to perceive the /l/ vs. /r/ contrast. In another example, Dalby and Kewley-Port (1999) developed systems of automatic speech recognition to provide feedback to L1 English speakers of L2 Spanish or L2 Mandarin. Through the use of a computer monitor and headphones (e.g., hearing a minimal pair and selecting the heard word on the screen with a mouse), the systems were shown to effectively improve L2 pronunciation. Regarding intonation, L1 English speakers of L2 Spanish improved their imitation of L2 Spanish intonation following visual instruction on native speaker F0 patterns (Allison 1993).

Amount of L2 input has also been found to affect native-like pronunciation. For example, Bongaerts (1999) found that L2 learners who were judged to be native speakers by native speakers had experienced "massive" amounts of L2 input (i.e., they were all living in the L2 country and continually exposed on a daily basis to the L2 at the university level). Some L2 English subjects read English sentences and were judged to be native English speakers, suggesting that late L2 learners can be perceived as native speakers by native speakers of the L2; however, as pointed out by Bongaerts, in addition to having lots of L2 input, these subjects appeared to have had high motivation to sound like native speakers and in-depth training in production and perception of the target sounds.

Similar to the amount of L2 input, adult L2 pronunciation ability has also been correlated with regularly overhearing the target language as a child. Knightly (2000) found that adult L1 English learners of L2 Spanish demonstrated significantly more native-like pronunciation of word-initial /p, k/ and word-medial /b, d/ than students who had first been exposed to Spanish in high school. Knightly's results emphasize the importance of collecting language background information for subjects since previous childhood language exposure may affect pronunciation ability. To some extent, it could be argued that overhearing a target language is confounded by and related to 'age of learning' (see Section 2.3.3). Thus, overhearing a target language is a very beginning step towards learning a target language, or at the very least, may prime the learner to more easily learn it later on.

It may be worthwhile to point out that learners have their own beliefs regarding what affects accurate L2 pronunciation. Cenóz and García Lecumberri (1999) surveyed 86 L2 English learners with L1 backgrounds in Castilian-Spanish and Basque. The results showed that learners believed that ear training and contact with native speakers were the two most important factors identified with successful L2 pronunciation. In actuality, it's not clear how ear training and contact with native speakers corresponded to their L2 pronunciation ability since Cenóz and García Lecumberri did not test L2 production; however, further investigations could test the idea that contact with native speakers may correspond to amount or quality of L2 input; however, a learner's own predictions may not be reliable. The results of Lefkowitz and Hedgecock (2002) showed a disparity between subjects' perceptions of their own performance and their observable production.

In the current study, the type and amount of input could not be strictly controlled for across participants; however, by grouping participants into levels based on their experience, the degree of each group's L2 overall language ability was assumed to be constant across speakers at a given level. For this same reason, some participants at the beginning level were excluded from the final analyses due to extensive previous exposure to Spanish instruction; however, one participant did report incidental exposure to spoken Spanish from the age of 5 to 15 years of age (see Section 4.2.1). Thus, bearing in mind the findings reported by Knightly, the scores of this subject might be predicted to higher than scores for the counterparts in her cohort.

2.3.7 Gender

Research into the role of gender in successful L2 pronunciation acquisition has yielded differing results. Some studies have found an effect of gender while others have not. Larsen-Freeman and Long (1991) suggest that females may be better than males in L2 acquisition due to the "generally accepted fact that females enjoy a rate advantage, initially at least [in L1 acquisition]"(204), implying that females demonstrate successful pronunciation sooner than male counterparts. Nevertheless, the authors only cite two studies to support their claim. In the first study (Farhady 1982), which included 800 subjects, females scored better than males on a listening comprehension task during a foreign language placement test. In a second study (Einstein 1982), results demonstrated that females were significantly more able than males to discriminate between L2 prestige dialects.

Gender differences may also be expected when the self-reported importance of language use is shown to differ between males and females. Flege, Takagi, and Mann (1995b) assessed accuracy of L2 English productions by L1 speakers of Japanese. Participants were living in the United States. Although the authors did not originally analyze gender differences in L1 Japanese productions of /ɹ/ and /l/ (granted, their sample size was limited to four males and eight females), Flege (1998) performed a reanalysis of the 1995 subject background responses. The reanalysis showed that the men placed a greater importance on English language use at work rather than English use at home; however, the reverse was true for the women.

Gender has been found to interact with age of learning and the degree of foreign accent (Flege et al. 1995a). Participants included 240 L1 Italian speakers who had begun learning English in Canada between the ages of 2 and 23 years, yielding an overall length of residence of 32 years. The judges included 10 native English speakers. The results found that females received higher scores (i.e., they had less of a foreign accent) when subject age of arrival averaged at 9.6 years; however, males showed higher ratings when comparing groups with an average age of arrival of 21.5 years.

In a more recent study, gender effects were found in the production of the L2 Spanish grapheme <v> (Stevens 2000) by L1 English instructors of Spanish. Although a group of native Spanish speakers (n=10) behaved as a group for gender (i.e., there were no significant pronunciation differences between males and females), the very experienced L2 learners (three females and two males) showed differences between genders. Specifically, females and males favored different realizations of the grapheme <v> such that females were more likely to produce <v> as a labiodental; however, it should be noted that the small sample size may have affected observed male/female differences.

Other studies of L2 acquisition have not shown any affects for gender. In Piske et al. (2001), three native English speakers rated 72 native Italian speakers of L2 English. Age of learning was found to be the most significant factor in predicting degree of perceived foreign accent. Gender, length of residence in target language country, and self-estimated L1 ability were not found to be a significant predictor. Elliot (1995a) assessed the L2 segment accuracy of Spanish targets. Out of 32 males and 34 females, there was no significant effect of gender. Flege and Fletcher (1992) analyzed the L2 English by L1 Spanish speakers (n=30; equal numbers of males and females). Subjects were rated by 10 native English speakers for degree of foreign accent. Length of residence was found to correlate with degree of foreign accent; however, gender, daily use of English, and chronological age were not found to be significant factors.

In sum, gender effects in successful L2 acquisition should continue to be analyzed as differences are not always conclusive or necessarily robust due to small sample sizes. In the current study, the lack of equal samples of men and woman do not allow for appropriate inferential testing of subjects by gender. Nevertheless, qualitative descriptions of subject results by gender are included in the current investigation as gender is included as an independent variable and it has previously been found in some studies as a possible factor in pronunciation success.

2.3.8 Attitudes, empathy, and motivation

A learner's attitude has also been correlated with L2 pronunciation accuracy. Suter (1976) asked native English speakers to subjectively score L2 English learners of L1 Arabic, Persian, Japanese, and Thai. The results were reanalyzed by Purcell and Suter (1980) and showed that s*trength of their concern of pronunciation accuracy* was one of four predictors of accurate pronunciation scores. Hammond and Flege (1989) analyzed subjects' attitudes towards foreign accents (i.e., speaking a language without native-like pronunciation) and subjects' ability to mimic Spanish targets. The results suggested that learners with negative attitudes towards a foreign accent were more likely to imitate their own English L1 with Spanish segment substitutes. In order to explain this inverse relationship, Hammond and Flege suggested that "the more negative an individual feels toward foreign accent, the more closely s/he may attend to it."(674) In a separate study, Elliot (1995) assessed pronunciation attitudes using a Likert-type scale survey. Pronunciation attitude by L1 English speakers was found to be the best predictor of L2 Spanish accuracy when compared to gender, years of L2 instruction, travel to the L2 country, L2 speaking relatives, target language grade point average, overall grade point average, and number of other languages studied. Stokes (2001) looked at 5 factors: 1) length of residence in Spanish environment, 2) length of exposure to Spanish, 3) attitude,

4) music reading ability, and 5) music experience. Results showed that only attitude was a significant factor in predicting global pronunciation accuracy.

The degree of learner empathy may be similar to learner's attitude. Guiora et al. (1972) showed that an empathy score was most accurate in predicting native-like L2 pronunciation. The score was derived from a test in which subjects watched 3 short films, first at a realistic speed and then again in slow motion. Participants were told to press a button each time they noted a "change in facial expression" as a gauge of empathy. Authenticity of L2 pronunciation was judged by a panel of three "trained experts" in the target language who scored varying segments of subject's imitation of L2 forms after hearing native speaker sentences.

Regarding motivation, Smit (2002) found that L1 Austrian German speaker success in an L2 English pronunciation course was positively correlated with their readiness to change and work on their pronunciation. Schumman (1978) looked at affective factors such as language and culture shock of an L2 learner in a naturalistic setting. He also hypothesized a correlation between the socio-economic status of the L2 learner native speaker, and the learner's degree of self-esteem (e.g., fear of being ridiculed); however, Schumann did not formally test his ideas.

2.3.9 Amount of L1 use

More recent studies have shown that the amount of L1 use can affect L2 production accuracy. According to the "single system" hypothesis (Flege et al. 1997), first suggested by Dunkel (1948 cited by Flege et al. 1997), a learner's self-reported use of L1 in an L2 learning environment inversely predicts L2 production accuracy. That is, the lower a learner's L1 use, the more native-like his or her L2 pronunciation is predicted to be. In Flege et al.'s study, the authors analyzed L2 Canadian English productions by two groups of native Italians (n=100) who had emigrated to Canada at the same age (5.6-5.9 yrs); however, the two groups differed in their amount of self-reported use of Italian (36% vs.

3%). Monolingual English listeners completed a foreign accent detection task and identified speakers in both groups has having foreign accented English; however, Italian subjects who had reported high levels of L1 use received significantly higher foreign accent scores than participants who reported low L1 use.

Guion et al. (2000) also found support for the single system hypothesis. In her study, 30 Quichua-Spanish bilinguals were separated into 4 groups based on self-reported use of Quichua. Subjects produced sentences (through an auditory repetition task) in both languages while 10 judges (5 near monolingual Spanish and 5 near monolingual Quichua speakers) auditorily rated the degree of foreign accentedness on a 1-9 point scale. The researchers found no significant difference between group scores when rated on their Quichua productions (i.e., differences in L1 use did not affect degree of perceived accent in the L1); however, a significant difference was found between the high vs. low L1 use groups when producing Spanish. To further test these findings, Guion et al. (2000) reanalyzed Korean-English bilingual production data from Yeni-Komshian, Flege & Lui (2000). The results also yielded a significant correlation between amount of L1 use and degree of perceived L2 foreign accent.

In the context of the current study, amount of L1 use is less of a predictor of L2 acquisition since all learner groups are learning Spanish in a formal classroom environment. Thus, it is assumed that they all use English to a similar extent.

2.3.10 Additional factors

Degree of foreign accent may also be affected by a listener's social background as well. Beardsmore (1979) found that different communities tolerate accented speech differently. 237 listeners representing three different speech communities showed differences in the acceptability of the accented speech of a single L1 Dutch speaker of L2 English. The results suggest that a judge's auditory impressions are affected by their community backgrounds. Although not argued for by Beardsmore, the results support the use of objective acoustic criteria in the assessment of L2 speech. That is, if judges are subjective in their assessment of foreign speech, acoustic criteria may help to increase assessment reliability.

More important to the current study is the nature of the pronunciation task which has yielded different effects on accuracy evaluation. While not necessarily an indication of accuracy, increasingly informal tasks have been associated with increased variability of attempted target productions. For example, Tarone (1983) described how different elicitation tasks affect L2 pronunciation. Citing Dickerson and Dickerson (1977), Japanese learners of L2 English yielded higher accuracy scores for /ɹ/ in a careful speech task in comparison to scores in a casual speech task. Major (1992) also found that L1 Portuguese speakers of L2 English produced less native-like durations of /p, t, k/ when speaking spontaneously rather than in reading from word lists.

In contrast to previous findings, Lin (2001) found that adult L1 Chinese learners of L2 English produced more accurate forms in less formal tasks. Lin analyzed the production of complex consonants and found that subjects increased their use of epenthesis as the style of their production became more formal; however, the incidence of consonant deletion and consonant replacement decreased with greater task formality. Flege et al. (1995b) also found that inexperienced L1 Japanese learners of L2 English demonstrated significantly lower accuracy scores in producing the English /l, ɹ/ distinction; however, accuracy of experienced learner productions did not show an effect of elicitation task suggesting that effects of task may be limited to inexperienced learners.

For the current investigation, if elicitation task does allow learners to more accurately approximate target forms as most studies suggest, then we might expect productions to truly reflect the best of their pronunciation ability. For example, if the use of syllable simplification strategies is indeed weakened in more formal tasks as suggested by Lin (2001), then analyzing the productions of learners in a formal word list reading task may most facilitate the emergence of the svarabhakti vowel in their productions.

To summarize Section 2.3, a variety of processes can influence interlanguage productions and acquisition including native language transfer. Gestural components of phoneme acquisition, universal markedness, tendencies in universal as well as L1 Spanish acquisition, age of learning, length of study, perception, developmental and learning strategies, instruction, input, gender, attitudes, and amount of L1 use have all been correlated with successful L2 pronunciation. The specific predictions that each of these factors makes for the current study and its questions are restated in at the end of this chapter in Section 3.2

2.4 Previous measures and criteria of pronunciation "accuracy"

One goal of this study is to quantify and define *pronunciation accuracy* using auditory and acoustic evaluation strategies based on native speaker productions. This section attempts to justify a two-tiered methodology of accuracy assessment (i.e., auditory and acoustic analysis), present its benefits, and explain how this study's methods differ from previous studies.

Previous pronunciation investigations have judged L2 productions in a variety of ways. The term "accurate pronunciation" has almost as many meanings as there are studies in which it is investigated. With the exception of acoustic waveform measurements of voice onset time (VOT), the majority of pronunciation studies have been limited to subjective transcriptions and rating scales reflecting impressions of accentedness. VOT can be described as the time between the release of a stop consonant (e.g., the coming apart of the lips in the sound [b]) and the beginning of periodic vocal fold vibration (Abramson and Lisker 1973). Kerswill and Wright (1990) demonstrated the importance of a *theory of transcription*. The authors conducted an experiment in which they compared the transcriptions of thirteen phoneticians to electropalatographic and acoustic measures. Transcriptions reflected English minimal pairs (e.g., *bribe* - *bride*). The results showed a lack of validity in some transcriptions. For example, transcribers sometimes noted degree of alveolarity and vowel length distinctions in contrast to the electropalatographic or acoustic data. Additionally, the indication of consonantal features by the transcribers suggested that they were constrained by the IPA's segmental nature. In effect, Kerswill and Wright's study showed that an unambiguous transcription strategy must clearly define the role of the following three areas: articulatory vs. auditory impression, individual segments vs. the continuously varying acoustic signal, and the effects of coarticulation vs. allophonic rule. Kerswill and Wright emphasized the importance of knowing the features and strategy used by the transcriber in order to correctly interpret the transcription. Unfortunately, transcription strategies are taken for granted and rarely discussed in methodologies. This section presents a variety of approaches that have been used to evaluate *accurate pronunciation*.

This section is organized as follows: Section 2.4.1 looks at studies that have assessed L2 pronunciation accuracy using auditory criteria. In Section 2.4.2, investigations which have used acoustic measures and criteria to score pronunciation accuracy are presented.

2.4.1 Auditory assessment of "accuracy"

Several studies have looked subjects' overall pronunciation (i.e., global accent). Suter (1976) assessed L2 English pronunciation accuracy based on an overall impression of word pronunciation (and not on a segmental level). In his study, 14 native English speakers listened to 30 seconds of speech elicited from L1 Arabic, Persian, Japanese, and Thai speakers of L2 English. The speech samples described a holiday or celebration from the subject's native country. Each subject's "overall pronunciation" was rated on a 1-6 point scale from "best to worse." Details about the scale were not provided; however, judges were told to pay attention to rhythm, stress, and intonation and "not to lump scores into the middle range."(242) Neufield (1980) described several experiments in which native French judges attempted to categorize "overall pronunciation." Subjects included both native French and non-native advanced L2 speakers of French who were categorized as 1) francophone from Canada 2) francophone from another country or 3) non-francophone. In a more recent study (Stokes 2001), global accent was scored on 1-10 scale. Subjects included 37 L1 English speakers of L2 Spanish who were scored by 3 native Spanish speakers.

Snow and Hoefnagel-Höhle (1982a) used three judges and a five point scale to rate the pronunciation accuracy of segments rather than overall pronunciation. Subjects included L1 English speakers imitating words in L2 Dutch. Segments thought most likely to cause problems for the subjects were selected for rating. Judges were asked to rate productions based on the following 5 point scale: "1) Uninterpretable as target sound 2) Correct target sound, very strong accent 3) Correct target sound, noticeable accent 4) Correct target sound, slight accent or 5) Indistinguishable from a native speaker's pronunciation."(86)

Flege and Hammond (1989) had two "phonetically-trained listeners" rate the Spanish accent imitations of L1 English subjects in order to test the effects of attitude and experience on a subject's ability to imitate English with a Spanish accent (e.g., with the carrier phrase "The ___ is on the ___." and inserting [bays] for *vice* and [t͡ʃit] for *sheet*). Subjects' productions (i.e., imitation accuracy) were scored into one of three categories: "1) The target sound was produced as it normally occurs in English (e.g., vice [vays]) 2) The target sound was replaced by the expected Spanish-accent variant (e.g., vice [bays]) or 3) The target sound was replaced by some sound other that the expected Spanish-accent variant (e.g., vice [mays])."(672) Each target was attempted twice within the same carrier phrase and the production was judged accurate if only one attempt was rated as shown in (2) above – *replaced by the expected Spanish-accent variant*. Hence, it appears that a subject who would have always produced but one out of two correct Spanish variants would have been rated at 100% pronunciation accuracy. Similarly, a different student who might have produced two out of two correct variants would also have been rated at 100% accuracy which does not appear to be an insightful comparison of scores.

Zampini (1994) evaluated the pronunciation accuracy of spirantized /b, d, g/ by L1 English speakers of L2 Spanish. Subjects read from a paragraph in Spanish and their pronunciation accuracy rating was based on the transcription of the subjects' realizations by two judges. The judges included a native Spanish speaker with a background in linguistics and a native English speaker working on a doctorate in Spanish linguistics. Accuracy scores were calculated only in the instantiations in which both judges agreed on the transcription (which was 95%); however, the criteria for their transcription is not clear from the methodology and appears to be based on auditory impression alone.

Elliot (1995a) used three judges to score the pronunciation accuracy of L1 English speakers of L2 Spanish.[17] Subjects were given three tests: auditory word mimicry, written sentence reading, and a word list reading. Target sounds were selected and judges were asked to rate productions on a scale of "1" (incorrect target sound), "2" (approximation of the target sound), or "3" (correct target sound). A fourth free-elicitation test was assessed following Suter's (1976) methodology for judging accuracy, in which judges rated "overall pronunciation" on a scale of 1-5, from "best to worse."

Southwood and Flege (1999) had 10 L1 English speakers rate the L2 pronunciation by 90 L1 Italian females on a 7 point scale. Six native English speakers were also included in the speech to be rated. Two important results were found in their study: First, ceiling effects by the judges suggested that a 9 or 11 point scale would be more insightful scale to foreign language performance. Second, testing between raters revealed poor inter-judge reliability which again supports the need for clear criteria for auditory judgments.[18]

Several studies make no mention at all of the way in which productions were rated accurate. For example, Eckman and Iverson (1997) rated L2 English productions of [s, θ] by L1 Korean speakers and [ð, d] for L1 Spanish speakers. Kim and Jung (1998) also used transcription to measure the L2 English accuracy of consonant clusters by L1 Korean subjects. Following transcription, the productions were rated as being either "target language form" or "non-target language form" and non-target forms were then

---

[17] Elliot and a second judge were two English L1 speakers with backgrounds in Spanish linguistics. The third judge was a native Chilean speaker working towards a Ph.D. in Spanish.

[18] It might also be argued that more judges are needed; however, despite the number of judges, one would expect clear criteria to reduce aberrant rating differences between judges.

categorized from the transcription as having an articulatory feature change, cluster reduction, or cluster deletion. Carlisle (1997) and another linguistically trained judge transcribed two and three-segment onset clusters by L1 Spanish speakers of L2 English although no mention was made on their transcription strategy. Stevens (2000) transcribed L1 English learners of L2 Spanish in their productions of the Spanish graphemes <b, v> without providing a transcription strategy.

In sum, research demonstrates a range of auditory methods in the assessment of L2 segment accuracy. While some scores are awarded based on linear Likert-type scales, others result from categorical accuracy. While these methods do not necessarily reflect unreliable methods of evaluation, the reproducibility of methods and insight into phonetic variability are more constrained using auditory methods alone and, more so, in the absence of a transcription strategy.

2.4.2 Acoustic-phonetic assessment of "accuracy"

Assessment of "accuracy" may be accomplished under an acoustic phonetic approach by quantifying aspects of the acoustic signal which ideally are independent of the judge as long as consistent measurement criteria are applied. The use of acoustic measures in the study of interlanguage may be more insightful than transcription alone. Flege (1980) provides two arguments for the use of acoustic analysis of interlanguage. First, an experimenter may not be able to reliably transcribe all phonetic variants due to a limited number of transcription categories and the influence of the experimenter's native language perceptual categories. Second, phonetic transcription provides greater detail regarding a particular variant's distribution and the proximity of different variants to native target norms. Simoes (1996) provides additional arguments in favor of acoustic measures: The integration of phonetics in the assessment of L2 acquisition helps to reduce the subjectivity that is so often found in the use of the term "fluency"; and, the use of quantitative measures allows for more reproducible studies.

Several studies have defined L2 accuracy in terms of spectral measures through the assessment of L2 vowels (Mendez 1982, Simoes 1996). Mendez compared the productions of 3 monolingual Midwestern American English speakers with productions of 3 native Puerto Rican Spanish speakers. The first and second vowel formants were measured at the temporal midpoint. Mendez found no significant differences between both groups' productions of /a, i/ but a difference was found in /u/ (significantly lower frequencies were evidenced for Spanish /u/). Simoes also measured the first and second formants of vowels to serve as part of objective measurements of fluency in L2 Spanish. Subjects included five L1 English learners of L2 Spanish in a study abroad context in Costa Rica. In a third study, Kelm (1987) assessed the L2 acquisition of contrastive emphasis. Kelm measured pitch and intensity to compare L1 English Speaker productions of L2 Spanish with productions of L1 Spanish speakers.

More recently, acoustic measures have been used to judge the accuracy of L2 stop consonant productions. The use of acoustic measures has long been used in experimental phonetics to describe the differences in VOT of stop consonants between monolingual Spanish and monolingual English speakers (Abramson and Lisker 1973, Flege and Efting 1986, Flege 1991, Lisker and Abramson 1964, Quilis 1981); however, its use has been limited in L2 phonological acquisition and pronunciation accuracy. Still, Gonzalez-Bueno (1997) and Munson (2001) measured VOT durations for L2 Spanish stop consonants by L1 English subjects in order to assess their interlanguage development. Unlike subjective auditory scoring, VOT measures are clearly defined and allow an objective assessment of stop consonant production, emergence of pre-voicing, and degree of aspiration.

Flege (1991) analyzed the productions of word-initial [t] by L1 Spanish speakers of L2 English. Subjects were divided into a group of early L2 learners (5-6 year olds) and late learners (adults). Flege found that the late learners produced "compromised" VOT values in their realizations of L2 English [t] with values longer than their Spanish [t] yet still significantly shorter in duration than monolingual [t] realizations; however, the VOT durations produced by the early learners were not significantly different from the monolingual English VOT productions. Magloire and Green (1999) also did not find evidence of compromised VOT values by early L2 English learners (i.e., bilingual Spanish and English speakers). In their study, the bilinguals had all started to learn Spanish at 2 years on average and English at 3.1 years. These results suggest that the subjects in the current study may also produce measures that are also compromised although this may be difficult to test for the targets since /r/ has no comparable English sound and [ð] is not known to have a temporal feature difference between Spanish and English; however, the targets [t] and [ɾ] may demonstrate temporal differences (see Section 3.1.2 and 3.1.4). The findings of compromised VOT in L2 have also been found by learners with L1 not related to English and Spanish. Flege (1980) presented data by L1 Arabic speakers producing L2 English stops. Importantly Flege's temporal measurements showed that learners produced a range of phonetic variants.

Knightly (2000) used acoustic measures to assess intervocalic stop consonants /b, d, g/ which are realized in Spanish as [β, ð, ɣ], respectively. Based on the examination of spectrograms of each production, the intervocalic target was assigned one of three degrees of lenition (visible F1 contour, visible F1 and F2 contour, or visible F1, F2 and F3 contours) and one of three degrees of voicing (voiceless, partially voiced, or fully voiced). Native Spanish speakers auditorily rated the L2 segment productions as less accurate than the native realizations even though there were no differences between segmental measures. Knightly's results found that native speaker judges were affected by global pronunciation even when instructed to focus on individual segments, effectively supporting the incorporation of objective acoustic measures when assessing L2 segment accuracy.

Arslan and Hansen (1996) used acoustic measurements to identify temporal features and intonational characteristics of L2 English Speech produced by L1 Turkish, German, and Mandarin speakers. The subjects included 43 speakers from the "general Duke University community" with accented English speech.[19] Through an analysis of 20

---

[19] It is not clear how the authors decided that their subjects evidenced accented speech or if they controlled for experience in L2 English.

isolated words and 4 sentences, the authors measured four groups of temporal durations as potential accent discriminators: word-final stop closures, word-initial voice onset time, average vowel duration, and average word duration. With the exception of VOT, significant differences were found between the native and non-native groups. Significant differences were also found in the intonational slopes produced by the German and Mandarin groups and the native English group. While Arslan and Hansen's findings were not used to score individual accuracy, their acoustic measurements are important as they revealed acoustic feature differences between native and non-native speakers of English.

In sum, criteria for assessing accurate pronunciation vary greatly between studies and are often ill-defined or not defined at all. The term "pronunciation accuracy" has been used to reflect gradient impressions of accentedness of different scales or merely the results of transcription. Part of the aim of this study is to provide quantifiable, measurable definitions of pronunciation accuracy based on native speaker productions through the use of clearly defined auditory and acoustic strategies of evaluation. Section 4.4 details the two-tiered criteria for determining accuracy that are subsequently used for each L2 sound category. Having presented the benefits and previous uses of acoustic measures in L2 evaluation, the following chapter presents the actual properties and measures for [ð, t, r, ɾ] that have previously been addressed in the literature.

3. Acoustic properties and measures for [ð, t, r, ɾ] in Spanish

The goal of this chapter is to complete the background literature review by presenting the acoustic properties of [ð, t, r, ɾ] as found from measurements taken in previous research. In addition, the chapter concludes with a presentation of this study's question and their hypotheses which are based on the background presented in Chapter 3 and 4.

Even though this study performs its own analyses of these targets produced by native Spanish speakers (who were recorded under the same procedures as the productions of L2 learners), the summary of previous findings is important for two reasons. First, it is important for current research to be informed by and build upon the results of previous research; previous descriptions of target segments provide an investigator with clues as to important acoustic features to be identified, even though the same measurements of those features may or may not be possible to incorporate into the current study. Second, summarizing previous methods of acoustic analysis will reveal any differences between previous methodological approaches and the current study.

It is important to note that realizations of acoustic features by native speakers may vary based on a variety of dialects and sociolinguistic factors. For example, Medina-Rivera (1999) showed that productions of taps and trills by native Spanish speakers from Puerto Rico varied based on their social setting, theme, and the speakers relation to the interlocutor. The possibility of variation supports the need to base L2 pronunciation evaluation on native speakers subjected to the same tasks and procedures. It is also important to point out that the nature of the recording task may affect previous analyses of acoustic properties. For example, as found by Umeda (1975), temporal cues may be strongest in a formal setting such as the production of targets from the reading of a word list.

This section is organized as follows: Section 3.1.1 describes the properties identified and the measures that have been taken for [ð]. Section 3.1.2 presents studies that have looked at productions of [t]. In Section 3.1.3, the Spanish trill [r] is described together with measures in previous studies while the Spanish tap [ɾ] is discussed in Section 3.1.4. Section 3.1.5 is separately dedicated to the acoustic properties of the svarabhakti vowel. In each section, experimental studies that consider acoustic or auditory differences between the Spanish and English sound category will be mentioned. The auditory and acoustic properties for the L2 productions of /d,t/, /r/, and /ɾ/ are presented in Sections 6.1, 0, and 8.1, respectively. Section 3.2 concludes the chapter by setting out the questions and their hypotheses for the current study.

3.1 Acoustic properties and measures

3.1.1 Acoustic properties and measures for [ð]

In Spanish, <d> is typically realized as a voiced dental fricative [ð]. Quilis (1981) notes that the amplitude and frequencies of its formants, if present, approximate those of the surrounding vowels in an intervocalic production. Similarly, the degree of intensity inversely corresponds to the degree of constriction in the oral cavity. Native Spanish speaker productions of [ð] may evidence differing degrees of frication.

Manrique and Massone (1981) provided an acoustical analysis of voiced and voiceless fricatives in Argentinean Spanish. Their subjects included four adult male speakers producing disyllabic words of the shape CVCV with two patterns of stress. The authors took four measurements of each production: 1) The range in frequency (Hz) observed in the noise and/or the periodic component, 2) the difference (in dB) between the peak amplitude of the vocalic portion and the peak of frication, 3) the frequency at which F2 and F3 begin to transition into the production as well as the transition duration, and 4) the duration of frication. The physical boundary of a production was determined "by the descent of F1 and the simultaneous decrease of overall intensity."(1146)

The results of the Manrique and Massone study yielded several spectral and temporal characteristics for Spanish [ð]. While F1 was consistently centered around 500Hz for all voiced fricatives, F2 varied across fricatives due to the location of the consonant constriction. The initial frequency of the F2 transition into [ð] was also measured when [ð] proceeded the vowels [i, a, e, o, u]. The results yielded an F2 range from 1100Hz to 2000Hz, depending on the preceding vowel. In regards to segment duration, measures differed based on stress. The duration of [ð] averaged 58ms in a stressed syllable and 104ms in an unstressed syllable. Based on their analyses of voiced fricatives in Spanish, the authors found an effect of voicing and stress on segment duration. In sum, durations for fricatives were longer in non-stressed syllables. Additionally, a voiced fricative tended to be shorter than its voiceless counterpart (e.g., [ð] would be shorter in duration than [θ]).

Knightly (2000) also used acoustic measures to assess L2 productions of [ð]. Through a visual inspection of the spectrogram, Knightly assigned each realization two ordinal values. The first value reflected one of three degrees of lenition based on the visibility of the formants: Visible F1 contour; visible F1 and F2 contour; or visible F1, F2 and F3 contours. The second value represented a production's degree of voicing based on the continuity of the voicing bar beneath the segment: Voiceless, partially voiced, or fully voiced.

In the current study, overall segment duration was measured for all [ð] realizations. In the case of "approximated" productions (i.e., there was not enough constriction to appropriately assess segment start and end points), no measurements were taken; however, the approximated nature of the production was noted. See Section 4.4.1 for details.

3.1.2 Acoustic properties and measures for [t]

Spanish word-medial [t] is acoustically characterized by a stop closure, release burst, and short lag voice onset time. Voice onset time (VOT) is defined as the time from the release of the stop (e.g., the coming apart of the tongue tip from the dental place of articulation in Spanish) to the onset of periodic voicing in the following vowel. Spectrally, the release of the stop corresponds to a "sudden spread of spectral energy."(Williams 1977:172) The stop closure corresponds to a total absence of frequencies throughout the spectrum (unlike its voiced counterpart [d]) and is articulated with a dental place of articulation (Quilis 1981).

While both word-medial and word-initial [t] are described as having positive VOT (e.g., Quilis 1981), measurements of [t] have typically been limited to stops in word-initial position. Lisker and Abramson (1964) analyzed two native Spanish speakers. The average VOT was 7ms for /t/ productions (n=14) produced in sentences; 9ms for productions (n=16) realized in isolated words. While both subjects in Lisker and Abramson's study were speakers of a Puerto Rican dialect, Williams (1977) completed a cross-dialect comparison of VOT produced by Guatemalan, Peruvian, and Venezuelan speakers (n=24) in which words were produced in a reading task. The results yielded [t] with an average VOT of 10ms, 20ms, and 16ms, for the Guatemalan, Venezuelan, and Peruvian dialects, respectively. Flege and Efting (1986) measured word-initial /t/ VOT duration of 10 adult native Spanish speakers from Puerto Rico (5 males; 5 females). Target words were disyllabic and read in isolation. The results showed average VOT of 22.4ms with a standard deviation of 6ms. Thus, while variations across dialects may exist, all speakers tend to produce [t] with a short lag VOT less than 25ms.

The effect of speaking rate on [t] duration has also been investigated. Magloire and Green (1999) analyzed the effects of speaking rate on stop consonant productions in both Spanish and English. Their results found that short-lag sounds (e.g., [t]) were temporally constrained across speaking rates while lead-lag (i.e., prevoicing or negative VOT) and long lag (i.e., relatively long positive VOT) sound durations were robustly affected by speaking rate. Effectively, their measurements only found a small positive effect of decreased speaking rate on short-lag productions for both Spanish and English tokens. Thus, for the current study, normalization of [t] durations across speakers may not be necessary; however, Section 6.2.2 addresses the inclusion of native speaker VOT norms on the accuracy scores of L2 productions.

For the current study, two measures were assessed for each [t] production. First, the stop closure duration was measured as indicated from the disappearance of F2 on the preceding vowel and taken to the stop release burst. Second, the duration of VOT was measured from the stop release to the onset of modal voicing in the following vowel. See Section 3.1.2 for further detail.

3.1.3 Acoustic properties and measures for [r]

Quilis (1981) describes the trill as being typically characterized by 3 occlusions. In such cases, two intervening vocalic elements demonstrate spectral characteristics of one of the surrounding vowels. Spectrally, the location on the spectrogram that corresponds to an occlusion of the tongue tip on the alveolar ridge is characterized by "un espacio casi en blanco"(*a space almost all white*) (Quilis 1993:332). For this study, occlusions will generally be referred to visually as "stripes" in the spectrogram (i.e., the absence of glottal pulses). Measurements of trills have revealed an average duration of 85ms (Quilis 1993).

Solé (2002) studied the aerodynamic requirements for the production of voiced and voiceless apical trills. The subjects included the author (a native Spanish speaker) and another researcher (a native English speaker). Tokens were produced within nonce words /ara/ and /iri/, in which the medial trill was realized with and without voicing. Airflow measurements were collected through a Rothenberg mask and pneumotachograph. The trill productions were measured for oropharyngeal pressure (i.e., the pressure between the soft palate and the epiglottis). The results demonstrated several aerodynamic distinctions between the voiced and voiceless trill variants. Voiceless trills (as opposed to voiced trills) showed both higher oropharyngeal pressure and a higher rate of flow across the laryngeal constriction. In addition, voiceless trills were longer with more contacts, evidenced frication in the context of /i/, and demonstrated faster rates of vibration due to the increased translingual air flow. Solé's results suggested that voiceless trills are preferred aerodynamically to voiced trills due to fewer aerodynamic requirements. Solé also found an effect of vowel height on trill production. Higher pressure was needed in order to make the trill vibrate in the context of /i/ because the higher tongue position offered more resistance to exiting air flow.

It has also been observed that the typical trill considered standard in most Spanish dialects is realized as a voiceless velar fricative [x] by native Spanish speakers in Puerto Rico (Zlotchew 1974); however, Zlotchew did not take acoustic measurements of his observations.

An assibilated variant [ř] may also be realized in place of the apical trill in some parts of Colombia, Chile, and Argentina (Cárdenas 1958) as well in some dialects of Costa Rica, Guatemala, Mexico, Bolivia, and Peru (Quilis 1993). The assibilated variant may be described as not having multiple closures of the tongue tip against the alveolar ridge, but rather, a near-closure that is accompanied by frication. Quilis and Carril (1971) suggest that the realization of the assibilated variant in place of a trill may be a result of too much constriction in the oral cavity. In addition, the assibilated variant may be realized without voicing due to the increased supraglottal pressure. In contrast, a relatively open constriction on the assibilated variant is more likely to result in a voiced production. Quilis and Carril acoustically analyzed the assibilated variant of the trill by four native Spanish speakers: 1 male from Argentina, 1 female from Argentina, 1 male from Costa Rica, and 1 female from Chile. The authors found that in environments where the tap was expected to be realized, the assibilated variant was not found. Several spectral and temporal measurements were taken of assibilated productions and durational differences were found by word, syllable, and stress position. Productions in stressed syllables were longer as were trills word-medially. Word-initial productions averaged

140ms in a tonic syllable and 80ms in a non-stressed syllable. In word-medial position, productions were 181ms in tonic syllables and 122ms in atonic syllables. In the context of the current study, we might expect apical trills to be slightly longer word-initially since trill targets word-initially are always found in tonic syllables while medial targets are always in unstressed syllables.

For the current study, we can make two hypotheses based on the results of Solé's findings. First, L2 learners will produce more voiceless variants of trill productions due to the increased aerodynamic requirements of the voiced over voiceless trill. Second, L2 learners will demonstrate lower trill accuracy in the context of /i/ given the aerodynamic constraints caused by a higher tongue tip position.

In the current study, three measures were collected for word-initial and word-medial trill productions: 1) overall trill duration, 2) degree of voicing (0 = voiceless, 1 = partially voiced, 2 = fully voiced), and 3) number of closures. See Section 4.4.3 for analysis procedures.

3.1.4 Acoustic properties and measures for [ɾ]

Recasens (1987) analyzed the articulation of Spanish taps and trills and found gestural differences apart from the number of occlusions identified between the two realizations. Recasens suggested that the trill and tap require "contrasting degrees of tongue-dorsum constraint, which may be associated with the execution of several vibrations for the trill as opposed to only one vibration for the tap."(306) The trill showed a higher degree of coarticulatory resistance, requiring greater tongue-dorsum constraint than the tap (inferred from lower F2 values during target trill closures). Recasens suggested that the results were consistent with findings in L1 acquisition in which dorsal articulations are learned after non-dorsal articulations. With regards to the current study, Recasens' study predicts that L2 learners may demonstrate higher accuracy scores in taps over trills.

In a study of American English, Zue and Laferriere (1979) measured tap duration to the nearest 5ms by placing boundaries at the point in which the first formant was "visibly excited." In the wordshape VɾV (i.e., with the tap following a stressed syllable), average tap duration was 26ms in the context of <t> (11ms stdev) and 27ms when represented by <d> (10ms stdev). Zue and Laferriere also found that some taps appeared as voiced fricatives in the case of partial closures due to turbulence created at the place of constriction). Regarding the effect of vowels on tap duration, results revealed significantly longer flap durations when the preceding vowel was high and fronted (e.g., *seaṯing*). In such instantiations, the authors suggested that the tongue tip overshoots its temporal target resulting in a longer closure duration.

In Spanish, durations for medial taps have been found to be slightly shorter than durations for English taps. Quilis (1981) measured tap durations based on the length of formant interruption. The reported average constrictions duration for taps in a stressed syllable was 22ms and 18.6ms for taps in an unstressed syllable. The latter measurement and environment is relevant to the current study.

Monnot and Freeman (1972) analyzed x-ray results of so-called English flaps and Spanish taps and did not find a cross-language distinction. Their subjects included three native speakers of Spanish and three native English speakers. All productions demonstrated a "rapid movement against the forward part of the alveolar ridge" but no contact with the posterior part of the alveolar ridge (i.e., the tongue did not retroflex upwards and then strike against the posterior part of the alveolar ridge on its way down); however, in English, the tongue tip demonstrated anticipatory preparation in contrast to native Spanish speaker productions suggesting that English speakers may be using a larger portion of tongue tip for contact. Given this latter observation, one might expect longer closure duration (i.e., absence of glottal pulsing at the location of the tap closure) for the English productions.

Horna (1998) investigated a variety of measures for the productions of English and Spanish taps. In her study, Spanish taps were produced by two speakers in more informal and less controlled tasks (text reading and conversation). Following auditory classification of tap targets as 'probable', 'possible', or 'unlikely', all taps judged 'probable' were analyzed while no taps labeled 'unlikely' were considered. Less complete constrictions that were judged as 'possible' taps were included if acoustic inspection revealed "a small decrease in intensity during one to two pulses [on] the spectrogram."(23) Productions labeled 'possible,' that temporally were not longer than 45ms, were also included. For each tap production, five measurements were taken: 1) the constriction duration and, if present, the release burst duration, 2) the so called 'intensity dip' (in dB) 3) the peak amplitude (in dB)of the vowels directly preceding and following the flap, 4) the V1+flap+V2 sequence duration, and 5) F2 at both the offset of the vowel preceding the flap and at the onset of the vowel following the tap. Constriction duration was measured visually on the spectrogram where there was "a clear reduction in amplitude."(25) The 'intensity dip' was calculated by subtracting the trough intensity value of the tap from either the intensity peak of the preceding or following vowel, whichever was higher.[20]

Horna found some flaps with immeasurable durations (4%) while Price (1981 cited by Horna 1998) found that 10% of tokens were immeasurable. Horna suggested that their differences may be a result of differing analysis methods. Price judged a tap's closure interval based on obvious attenuation in the waveform while Horna visually judged the spectrogram. Horna suggested that the spectrogram may be more sensitive to tap closure duration than the waveform. Thus, in the context of the current study, visual inspection of the spectrogram over the waveform may reveal greater subtleties in tap closer intervals.

In regards to voicing, it has been found that L1 English tap productions are inconsistent across speakers and may sometimes be realized as voiceless (de Jong 1998); however, in L1 Spanish, Gili (1921) was unable to find any devoiced tap productions in any position. Thus, in the context of the current study, native speakers are not expected

[20] It should be pointed out, however, that Horna suggested that a measure of the 'intensity dip' may not be reliable for productions in which the preceding vowel's intensity peak is much higher than the peak of the vowel following the tap production.

to realized voiceless taps; however, if learners are experiencing interference from their L1, a learner may produce some voiceless taps.

The analysis of palatograms has revealed the Spanish tap to have an alveolar place of articulation (Gili 1921). In his study, palatograms revealed that the tap and trill in intervocalic position evidence the same post-dental place of articulation; however, the place of articulation of taps and trills was affected by surrounding vowel height. Gili observed that increased vowel height correlated with a farther forward place of contact (e.g., taps and trills around the vowel [a] were closer to the palate).

The realization of taps by word position has been shown to vary. Concerning taps in onset clusters, Gili (1921) acoustically analyzed 74 tokens with a tap in onset cluster and found that taps were realized with a single occlusion in all tokens. Preceding a consonant, Quilis (1981) found that a tap target typically evidences 2-3 occlusions. Gili suggested that the appearance of a vocalic, schwa-like element following a word-final tap (e.g., *sacar* 'to take'), both very common and normal. Quilis found that the tap in syllable-final position can appear with one or more stripes.

In sum, a cross-study comparison of English and Spanish taps suggests that Spanish taps in word-medial position may be shorter than similarly produced taps in English; however, spectral and articulatory differences between Spanish and English were not found to be significant. For the present study, overall tap duration is measured based on the appearance of a visible white stripe on the spectrogram (i.e., the interruption of glottal pulses as observed in the spectrogram). For details, see Section 3.1.4.

3.1.5 Acoustic properties and measures for the svarabhakti vowel

The occurrence of the svarabhakti vowel has long been discussed in the acoustic literature relating to Spanish taps (Gili 1921, Navarro Tomás 1918). The svarabhakti vowel can be described as an epenthetic vowel found in the context of a tap either preceding or following a consonant. The svarabhakti vowel is of particular interest to the current study as it is evidenced almost systematically by native speakers in the production taps in onset clusters; however, the results of Experiment A revealed a notable absence of the vowel in L2 productions. Effectively, the difficulty of L2 learners of producing the svarabhakti vowel led to the following hypothesis: *The manifestation of the svarabhakti vowel, if it is to occur, will only occur in very experienced L2 learners of Spanish.* Thus, Experiment B included experienced L2 learners in order to test this hypothesis.

Gili (1921) has provided perhaps the most in-depth analysis of the svarabhakti vowel in Spanish. In his analysis of 61 productions of taps in onset clusters, only 3 tokens did not evidence the svarabhakti vowel yielding a 95% occurrence. In word-medial position (e.g., *o.pri.mir* 'to oppress') and taps preceding a consonant (e.g., *bar.ba* 'beard'), the svarabhakti vowel occurred less frequently (67% and 62%, respectively). Gili found the duration of the svarabhakti vowel to be quite varied; however, in 80% of the realizations (n=126) its duration was found to be greater than the duration of the tap occlusion. In 14% of the cases the duration was the same as the tap duration and in 6% of the tokens the duration was shorter than the tap occlusion. In word-initial onset clusters, the svarabhakti vowel averaged 53ms in duration. In word-medial onset clusters and in tap plus consonant position, durations averaged 52ms and 37ms, respectively. Gili's measures are similar to those of Quilis (1970) who found the svarabhakti vowel to fall between 8ms and 56ms, averaging 29ms in duration.

Gili suggests that there might have been some measurement interference of taps preceded by velar consonants (e.g., *grano* 'grain'). The interference is attributed to the relatively long release burst of the velar stop consonant. Thus, in the context of the present study, the analysis of the svarabhakti vowel following [p] (i.e., a non-velar) is supported as its measures may be least influenced by a velar stop release.

In regards to spectral characteristics, Quilis (1970) observed that svarabhakti vowels typically demonstrated similar formant structures to the F1 and F2 of the following nucleic vowel; however, formants above F2 were not normally visible. For the current study, only the duration of the svarabhakti vowel was measured. See Section 4.4.6 for details.

3.2 The current study's questions and hypotheses

Synthesizing the theoretical views and previous investigations presented across Chapter 2 and the current chapter, the present study is able to make several predictions regarding the research questions originally presented in Section 1. In this section, each question is restated alongside a reference to the section containing the data that addresses the question. Following each question, hypotheses based on the background review are presented. Questions 1-3 were originally addressed in Experiment A. Experiment B was designed to readdress questions 1-3 and to answer questions 4-6.

<u>Question 1</u>: *Is there a relative degree of production difficulty of L2 Spanish word-medial /d, t, r, ɾ/ by L1 English speakers? That is, do students show similar score rankings across target accuracy scores?* The data in Section 6.2.1 bear on this question.

> H_0: There is no systematic difference in degree of difficulty. That is, observed accuracy scores will reveal no systematic ranking across participants.
>
> H_1: Observed accuracy scores will reflect rankings such that scores for /d, t, ɾ/ are lower than scores for /r/ based on Lado's (1957) claim that the allophonic splits are the most difficult processes in L2 phonological acquisition.
>
> H_2: Observed accuracy scores will reflect rankings such that scores for /r/ are lowest. This prediction is based on the additional gestural skills required to make the trill (Flege 1980) and the assumption that L2 acquisition patterns similarly to L1 acquisition in which trills are acquired relatively late (Bosch 1983, Gonzalez 1989).

<u>Question 2</u>: *Does trill accuracy emerge with higher scores word-medially over word-initial trill scores?* Section 7.2.1 presents the data that address this question.

H_0: There will be no systematic relationship between the rankings of word-initial /r/ scores and word-medial /r/ scores.

H_1: Accuracy score rankings by word position will be systematic such that word-initial scores are greater than scores for word-medial trill accuracy. This prediction is supported by the fact that word-medial trills have an additional orthographic cue <rr> as opposed to word-initial trills which are represented by a single grapheme <r>. In addition, Zampini (1998) found that learners were more likely to spirantize /b/ word-medially than between word boundaries (i.e., word-initially) suggesting negative prosodic effects on the acquisition of target segments at word boundaries.

<u>Question 3</u>: *Does tap accuracy emerge with higher scores word-medially over scores of taps in onset clusters?* The data that attempt to answer this question is presented in Section 7.2.2.

H_0: There will be no systematic ranking of word-medial tap accuracy scores and onset tap accuracy scores.

H_1: Scores for word-medial taps will be systematically greater than scores for taps in onset clusters due to the increased markedness of taps in onset clusters.

<u>Question 4</u>: *Is the relative ranking of scores for /t/ (i.e., the relative degree of difficulty) affected by length of voice onset time as a criterion for /t/ accuracy?* In Section 6.2.2, the data that relate to this question are found.

> H_0: The relative ranking of accuracy scores for /t/ will not be affected by the inclusion of voice onset time as a criterion for accuracy.

> H_1: The relative ranking of scores for /t/ will be altered with the inclusion of voice onset time as a criterion for accuracy. This prediction is based on the well-known evidence that voice onset time of stops differs between Spanish and English (e.g., Lisker and Abramson 1964) despite the fact that in Experiment A, a subject could produce a [t] with long lag voice onset time (i.e., aspiration) and still receive an a score of "1" (accurate) for that production.

<u>Question 5</u>: *Is there an effect of vowel height on trill productions scored as accurate?* The data in Section 6.2.2 bear on this question.

> H_0: Learners are just as likely to have accurate trill scores in any vowel environment. Flege et al. (1995b) did not find any significant interaction of vowel height or frontness on the accurate production of L2 English /l, ɹ/ by L1 Japanese speakers.

> H_1: Scores for trills will be greater in the context of lower vowels as it has been shown that the aerodynamic requirements for trill production are fewer in the context of low as opposed to high vowels (Solé 2002).

Question 6: *Is the emergence of the svarabhakti vowel limited to highly experienced L2 speakers of Spanish? That is, are experienced L2 learners more likely to evidence the svarabhakti vowel in their productions than less experienced L2 learners?* The analysis of data presented in Section 8.2.2 addresses this question.

H_0: There will be no difference between svarabhakti vowel emergence scores across beginning, advanced, and experienced L2 participant groups; however, the distribution of emergence scores for each group will be significantly different from the distribution of emergence scores for native Spanish speakers.

H_1: The emergence of the svarabhakti vowel is most likely to emerge in experienced L2 learners of Spanish productions. That is, the distribution of svarabhakti vowel emergence scores of highly experienced learners will be significantly different from the distribution of scores for learners in less experienced L2 groups. In addition, assuming the target sound is acquired late in L2 acquisition, differences between the experienced group scores and the native speaker scores will not be significant.

In sum, this study's questions are addressed through several hypotheses. In the following chapter, the methods of two related experiments (Experiment A and Experiment B) are presented. The methods of both experiments were designed to address the above questions.

4. Methods

The goal of Experiment A was to address questions 1-3 of this study. Question 1 asked, *Is there a relative degree of difficulty of word-medial* /d, t, r, ɾ/? Thus, L1 speakers of L2 Spanish from three Spanish levels were recorded reading from word lists on three separate sessions over a six-month period. Target words included word-medial /d, t, r, ɾ/. Percent accuracy for each target was determined based on criteria established on native speaker productions. The comparison of accuracy scores across categories will address the degree of difficulty of the medial targets. In order to address Question 2 of this study (*Does /r/ accuracy emerge with higher scores word-medially over word-initial /r/ scores?*), tokens in Experiment A included word-initial trills. Thus, percent accuracy scores will be compared for trills across two word positions. Question 3 asked, *Does /ɾ/ accuracy emerge with higher scores word-medially over scores of /ɾ/ in onset clusters?* Tokens with taps in onset clusters were additionally included in the word lists in order to compare scores for taps in onset clusters with scores for word-medial taps.

The goal of Experiment B was to provide additional data for questions 1-3 and to address questions 4-6 of this study. Question 4 asked, *Is the relative ranking of scores for /t/ (i.e., the relative degree of difficulty) affected by length of voice onset time as a criterion for /t/ accuracy?* With data from both Experiment A and Experiment B, scores for /t/ were reanalyzed to take into account a voice onset time duration criterion. Question 5 asked, *Is there an effect of vowel height on /r/ productions scored as "accurate?"* Experiment A did not vary vowel height surrounding word-initial /r/ targets. In Experiment B, both word-initial and word-medial /r/ targets varied surrounding vowel height. Thus, one goal of Experiment B was to elicit enough realizations of /r/ scored accurate in order to compare the distribution of scores by surrounding vowel height. Question 6 asked, *Is the emergence of the svarabhakti vowel limited to highly experienced L2 speakers of Spanish?* Based on the inability of the beginning, intermediate, and advance groups to successfully produce the svarabhakti vowel in Experiment A, Experiment B included a group of very experienced L2 speakers of Spanish in order to test the hypothesis that the svarabhakti vowel is acquired very late in L2 Spanish acquisition. Thus, Experiment B aimed at comparing scores for svarabhakti vowel emergence across Spanish levels.

Native Spanish speakers were recorded to provide auditory and acoustic baselines on which L2 accuracy scoring could be based; however, due to the dual result/methods nature of the native speaker productions, their productions, derived scoring criteria, and the scoring of L2 productions are detailed in Chapter 5. Nevertheless, the native speaker backgrounds are presented in Section 4.3 of the current chapter.

This chapter is organized as follows: Section 4.1 describes the subjects, materials and recording procedures of the participants found in Experiment A. In Section 4.2, the subjects and methods in Experiment B are discussed, emphasizing methodological differences from Experiment A. Section 4.3 presents the backgrounds of the 11 native Spanish speakers, five of whom were recorded under Experiment A and six who were recorded in Experiment B. Section 4.4 presents the analysis procedures and measures that were obtained from all native speaker and L2 productions. The statistical methods used in this investigation are explained in Section 4.5.

4.1 Experiment A: Subjects over 6-months of study

4.1.1 Experiment A: Participant recruitment and background

Students were invited to volunteer for this study from first, second, and third-year Spanish courses during the Fall quarter 2002 at the University of Washington. The first-year students were studying Spanish for the first time and were designated in this study as "beginning" level. Students from the second-year courses had previously studied Spanish for the equivalent of three quarters and were designated as "intermediate" level. At the third-year, students had the equivalency of six quarters of study and were labeled in this study as "advanced."

Students were informed of the general research question investigated in this study: *How do adult-learners of Spanish acquire a Spanish accent over time?* However, the specific sounds to be studied were not mentioned. It was explained that participants would benefit from the study by receiving personalized feedback regarding their productions and that they would be able, at the end of the study, to compare their productions to those of native speakers.

The final sample included 13 students from the beginning (n=4), intermediate (n=5), and advanced (n=4) levels. Only those participants who continued taking Spanish throughout the year were included in the final analyses. Other students at each level had dropped out of the study throughout the year of their own accord. Participants from each level were recorded three times, once each quarter; students from the beginning, intermediate, and advanced levels were recorded during the second, third, and fourth weeks of each quarter, respectively.

Table 4.1 lists the participants' backgrounds. Participant cohorts were subdivided by level and cohort; average age is provided. A language background survey was administered before recording in Session 1 (Fall quarter) and was used to gather information on gender, age, previous language study, and foreign language exposure. For example, participant A-Beg-F1 was a female who studied at the beginning level of Spanish throughout the project. She was 19 years old at the first recording (Session 1) and had previously studied French for two years and Spanish for one quarter. Prior to Session 1, she had not spent time in a foreign country. This information is important as a learner's native language may influence how he or she accurately produces sounds in an L2 (e.g., Terrell 1989). Furthermore, information regarding exposure to Spanish, such as travel, living in a Spanish community, and incidental exposure to the target language might account for inter-subject variation and suggest an effect on L2 phonology (Knightly 2000). See Section 2.3 for a summary of factors affecting pronunciation success.

A copy of the background survey can be found in Appendix A. All respondents identified themselves as native speakers of English. Participants A-Beg-F1 and A-Adv-F1 claimed to have grown up in Spanish speaking communities; however, the participants later revealed through oral communication that English was used solely in their homes and at school. With the exception of participant A-Beg-F1 who had studied Spanish for one quarter, participants at all three levels were studying Spanish at

Table 4.1 Experiment A: Summary of participant backgrounds.

Spanish Level	Subject Code	Gender	Age	Age of Learning	Previous language study	Time lived in any foreign country
Beginning	A-Beg-M1	M	19	19	German (4 yrs.)	
	A-Beg-M2	M	23	23	French (3 yrs.)	England: 2 yrs.
	A-Beg-F1	F	19	19	French (2 yrs.) Spanish (1 quarter)	
	A-Beg-F2	F	19	19	French (2 yrs.)	
	average age:		*20*	*20*		
Intermediate	A-Int-M1	M	21	15	Spanish (3 yrs.)	
	A-Int-M2	M	21	15	Spanish (3 yrs.)	
	A-Int-F1	F	20	14	Spanish (5 yrs.)	
	A-Int-F2	F	18	15	Spanish (3 yrs.)	
	A-Int-F3	F	21		Spanish (3 yrs.)	
	average age:		*20*	*15*		
Advanced	A-Adv-M1	M	20	16	Spanish (5 yrs.)	
	A-Adv-M2	M	17	5	Spanish (5 yrs.)	
	A-Adv-F1	F	20	16	Spanish (4 yrs.)	Ecuador: 2.5 ms.
	A-Adv-F2	F	19	14	Portuguese (1 yr.) Spanish (1 yr.)	Brazil: 1 yr.
	average age:		*19*	*13*		

their corresponding level for the first time in Session 1 (during the Fall quarter) (e.g., students at the intermediate level had never before taken a second-year course).

4.1.2 Experiment A: Word lists and targets

At each of the three recording sessions, participants read from four Spanish lists and one list in English. The lists were presented in random order. The first list was always in Spanish and used to adjust recording levels and to correct any immediately evident list effects; however, tokens from the first list were never included in the analyses. Each Spanish list was comprised of 21 words to be analyzed while the English list contained 18 words to be analyzed. In order to minimize list effects, three dummy words were placed at the beginning and end of each list. Three extra dummy words were added to the English list in order to make the number of tokens equal to that of the Spanish lists. The Spanish lists contained the same words with three different randomized orderings so that three repetitions of each target were recorded from each speaker; however, the dummy words used on each list at every session were different. Randomization was performed in Microsoft Excel. All target words were placed in a carrier phrase in order to minimize creaky voice and possible list effects.

Each Spanish word was placed in the carrier phrase *Dice* ______ *también.* 'He/she says ___ also.' Each target word included the target sounds in the following word positions: spirantized /d/ (word-medially), stop /t/ (word-medially), trill /r/ (word-initially and word-medially), and tap /ɾ/ (word-medially and in onset clusters). The vowel [e] in *Dice* before each word was important in order to allow for comparison of the word-medial trill (e.g., *perro* 'dog'), which was preceded by a vowel, with the word-initial trill (e.g., *dice rico* 'he/she says rich'). Each Spanish word was disyllabic and of the phonological shape C_1VC_2V except for the onset-tap targets *prisa* 'haste' and *traje* 'suit' (C_1RVC_2V) and *pronto* 'soon' ($C_1RVC_2C_3V$).[21] The three words for each target category controlled for stress placement by maintaining primary stress on the first syllable. Vowel height was varied for word medial targets since vowel height may affect the aerodynamic conditions required for production of a following consonant, as in the

[21] In Experiment B, the target words *pronto* and *traje* were substituted with the words *preso* 'arrested' and *prado* 'meadow', respectively (see Section 4.2.2).

case of trills (Solé 2002). For example, all three tokens for each category were preceded by high, mid and low vowels (e.g., *píto* 'string', *bóta* 'boot', *dáto* 'fact').[22] In order to avoid possible lexical effects (e.g., unfamiliar words), words were chosen from a first-year Spanish text.[23] Table 4.2 lays out all Spanish words included in the wordlist and their corresponding target sounds. A sample word list for Spanish and English are presented in Appendix B and Appendix C, respectively.

---

[22] In the case of word-initial trills, vowel height was varied *following* the target. Vowel height was not varied preceding the word-initial trill target in order to limit the complexity of the study design; however, in Experiment B, vowel height is varied preceding the initial trill targets.

[23] However, Flege et. al. (1998) were not able to correlate cognate status, text frequency, or lexical familiarity to segmental accuracy for /t/ in L2. The Spanish text was "¿Sabías que…?" (Vanpatten et al, 1996).

Table 4.2 Spanish phones and their positions within target words. All tokens were used in both experiments with the exception of *pronto* and *traje* which were replaced in Experiment B with *preso* and *prado*, respectively. Although absent from the table below, word-final <r> targets were also presented to participants in the target words *subir*, *beber*, and *sacar*.

		Word-initial /r/[2,5]			Word-medial /r/[1,2,5]			Word-medial /ɾ/[1,3]		
	Token:	ri	re	ra	urr	err	arr	ir	or	ar
1	risa	risa								
2	remo		remo							
3	rato			rato						
4	churro				churro					
5	perro					perro				
6	carro						carro			
7	miro							miro		
8	toro								toro	
9	para									para

		Onset cluster /ɾ/[3,6]			Word-medial /d/[1]			Word-medial /t/[1,4]		
		ri	ro	ra	id	od	ad	it	ot	at
10	prisa	prisa								
11	pronto		pronto							
12	traje			traje						
13	pido				pido					
14	todo					todo				
15	dado						dado			
16	pito							pito		
17	bota								bota	
18	dato									dato

[1] Category reflects question #1: Is there an acquisition order of word-medial intervocalic /d, t, r, ɾ/?

[2] Category reflects question #2: Does trill accuracy emerge with higher scores word-medially over word-initial trill scores?

[3] Category reflects question #3: Does tap accuracy emerge with higher scores word-medially over scores of taps in onset clusters?

[4] Category reflects question #4: Are L2 accuracy scores for /t/ affected by length of voice onset time as a criterion?

[5] Category reflects question #5: Is there an effect of vowel height on trill accuracy?

[6] Category reflects question #6: Are experienced L2 learners most likely to evidence the svarabhakti vowel in their productions over participants in other groups?

Table 4.3 English phones and their positions within target words.

		Word-initial /ɹ/			Word-medial /ɹ/			Onset /ɾ/[1,3]		
	Token:	[ɹi]	[ɹe]	[ɹa]	[iɹ]	[eɹ]	[aɹ]	[ɹi]	[ɹe]	[ɹa]
1	relay	relay								
2	raider		raider							
3	robber			robber						
4	teary				teary					
5	payroll					payroll				
6	starry						starry			
10	treaty*							treaty		
11	gray*								gray	
12	draw*									draw

		Word-initial /d//			Word-initial /t/		
		[iɾ]	[ɛɾ]	[aɾ]	[uɾ]	[eɾ]	[aɾ]
13	needy	needy					
14	eddy		eddy				
15	body			body			
16	duty				duty		
17	eighty					eighty	
18	haughty						haughty

*At the second session of Experiment B, the tokens *treaty, gray,* and *draw* were replaced with the words *prequel*, *program*, and *product*, respectively. The replacement was made in order to assure that the L1 English participants were indeed producing the /pr/ cluster with a stop-approximant combination as expected in American English.

Table 4.3 presents the English targets and their environments. Each English word was embedded in the carrier phrase "He said _____ to me." Vowel height was varied and primary stress was controlled in English word shapes. All English words were disyllabic except for the tokens *gray* and *draw*.

4.1.3 Experiment A: Recording procedures and data collection

Each participant was recorded in the sound-attenuated booth at the University of Washington Linguistics Phonetics Laboratory. Participants were recorded in three separate sessions over a 6-month period, referred to in the results as T1, T2, and T3, respectively. Prior to the first recording, participants completed a consent form and the language background survey. At the time of the third recording, during the Spring quarter, students at the beginning, intermediate, and advanced levels were several weeks into their third quarter of first, second, and third-year Spanish courses, respectively.

Each recording proceeded as follows: Upon entering the recording booth, the participant was informed that participation was voluntary and that he or she could stop at any time. The microphone was placed approximately six inches from the participant's mouth at a 45 degree angle. It was explained that the first list would be in Spanish and would be a warm-up list to properly adjust the recording levels. The participant was asked to read each sentence and then was given an example of the desired intonation pattern in English (i.e., I illustrated the intonation and speaking rate of the Spanish carrier phrase with English words substituted). The participant was told to read the Spanish *as he or she thought it ought to be pronounced in Spanish* while the English was to be pronounced *as he or she normally pronounced English*. Following the Spanish warm-up list, three more lists in Spanish and one list in English were randomly presented. If the microphone did not receive a clear signal at any time, the participant was asked to start again from a line indicated by the researcher. Each recording lasted approximately 5 minutes. Spanish was not spoken by the researcher at any time before or during the recording with the participant.

Recordings were made using an Electro/Voice RE20 microphone with digital input to a Macintosh G3 computer running SoundEdit™ signal processing software, enabling automatic digitization. The resulting signal was sampled at 44.1 kHz. At the end of the recording, the entire session was saved as a single *.wav* file and then later segmented using Praat signal processing software (Boersma and Weenink 2003) into smaller files representing each target carrier phrase.

4.2 Experiment B: Subjects over 2-months of study

The methodological differences between Experiment A and Experiment B are summarized here: Regarding subject level, both experiments included L2 participant groups designated as "beginning" and "advanced" levels; however, Experiment A included a group designated as "intermediate" while Experiment B included a group designated as "experienced." In the case of the experienced L2 speakers, all were Spanish instructors with L1 English backgrounds at the time of recording. In terms of materials, all target words in Experiment A were placed in the same carrier phrase while target words in Experiment B were varied across three carrier phrases in order to test the effect of vowel height on word-initial trill accuracy. Furthermore, Experiment B replaced two target words (*pronto* and *traje* with *preso* and *prado*, respectively) with the aim of controlling for word shape and the consonant preceding the tap.

The organization of this section is as follows: Section 4.2.1 presents the three participant groups including a group of experienced L2 speakers of Spanish. Section 4.2.2 describes the revised word lists with altered carrier phrases. Five out of 11 native speakers were recorded under Experiment B conditions; however, their backgrounds are presented in Section 4.3.

4.2.1 Experiment B: Participant recruitment and backgrounds

Similar to Experiment A, participants were recruited according to University of Washington Human Subjects Division's approved procedures. A new set of students was invited to participate drawn from the first and third-year courses at the University of Washington during Summer quarter 2004. The first-year students were studying Spanish for the first time and were designated as "beginning" level like their Experiment A counterparts. The third year students were labeled here as part of the "advanced" group and all had the equivalent of six quarters of college level Spanish instruction. The non-native instructors of Spanish all had at least ten years studying Spanish and were designated as "experienced."[24]

The final analyses from Experiment B included four beginning participants, six advanced participants, and 11 experienced participants. Participants at the beginning and advanced levels were recorded in two separate sessions, the first and last weeks of the Summer quarter, respectively. The experienced level learners (i.e., the non-native instructors of Spanish) were recorded in one session. Several students from the beginning level were not included in the final analyses because English was not their native language or they had extensive prior Spanish instruction. Table 4.4 summarizes demographic information for the Experiment B participants, their backgrounds, and corresponding experience levels.

---

[24] The term "near-native" was also considered; however, although it could easily be argued that the non-native instructors were all very competent in their Spanish experience, knowledge of the language, and its pedagogy, the term "near-native" implies a level of proficiency that may or may not be true of their segmental pronunciation performance. Thus, the term "experienced" was chosen as it more appropriately reflects their backgrounds without presuming performance.

Table 4.4 Experiment B: Summary of participant backgrounds.

Spanish Level	Subject Code	Gender	Age	Age of Learning	Previous language study	Time lived in any foreign country
Beginning	B-Beg-F1	F	32	32	German (1 year)	Germany: 9 ms.
	B-Beg-F2	F	20	20	ASL (1 year)	
	B-Beg-F3	F	32	32	ASL (1 year)	
	B-Beg-F4	F	20	20	French (4 yrs.)	
		average age:	*26*	*26*		
Advanced	B-Adv-F1	F	20	12	Spanish (7 yrs.)	
	B-Adv-F2	F	20	13	Spanish (2 yrs.) French (1 yr.)	
	B-Adv-F3	F	26	19	French (15 yrs.) German (10 yrs.)	
	B-Adv-F4	F	24	18	Spanish (2 yrs.) German (2 yrs.)	
	B-Adv-F5	F	30	12	Spanish (10 yrs.) French (3 months)	
	B-Adv-F6	F	19	15	Spanish (4.5 yrs) Thai (19 yrs)	
		average age:	*23*	*15*		
Experienced	B-ExpL-M1	M	24	14	Spanish (10 yrs) Portuguese (2 yrs.)	Spain (4 ms.)
	B-ExpL-M2	M	26	14	Spanish (15 yrs.)	Spain (2 yrs.) Mexico (6 ms.)
	B-ExpL-M3	M	28	15	Spanish (10 yrs) French (6 yrs.) Portuguese (1 yr.)	Spain (1 yr.) Chile (6 ms.)
	B-ExpL-M4	M	27	11	Spanish (14 yrs.) Portuguese (2 yrs.)	Spain (1yr.)
	B-ExpL-F1	F	23	16	Spanish (6 yrs.)	Spain (1yr.)
	B-ExpL-F2	F	29	5	French (4 yrs.) German (3 yrs.)	Spain (8 yrs.)
	B-ExpL-F3	F	38	9	French (6 ms.) Italian (6 ms.)	Spain (2.5 yrs.)
	B-ExpL-F4	F	56	13	Spanish (42 yrs.)	Spain (1 yr.) Mexico (6 ms.)
	B-ExpL-F5	F	28	13	Arabic (3 yrs.) French (1 yr.)	Spain (2.5 yrs.)
	B-ExpL-F6	F	28	13	Spanish (15 yrs.) French (1 yr.)	Spain (1 yr.)
	B-ExpL-F7	F	26	13	Chinese (4 ms.) French (4 ms.) Portuguese (4 ms.)	None
		average age:	*30*	*12*		

4.2.2 Experiment B: Materials

Experiment B word lists followed the design set out in Section 4.1.2 with the exception of two target words and, additionally, phrasing of the carrier sentences. Modifications were accomplished to better control for factors influencing comparison of word-initial vs. word-medial /r/ accuracy, and to include /ɾ/ targets in onset clusters to enable this study to address the question of whether very experienced learners evidence higher svarabhakti emergence over participants with less L2 experience.

In Experiment A, the Spanish word lists included the target words *pronto* 'soon', *traje* 'suit', and *prisa* 'haste' which allowed for the assessment of tap accuracy in onset syllables; however, in order to control for word shape and place of articulation in the onset consonant, *pronto* and *traje* were replaced with *preso* 'arrested' and *prado* 'meadow', respectively. Thus, in Experiment B the target words *prisa, preso,* and *prado* were used to assess tap accuracy in onset cluster. All three words conformed to the same word shape $C_1C_2V_1C_3V_2$, and together controlled for the consonant (i.e., [p]) preceding each tap target. The consonant [p] was chosen as opposed to [t, k] in order to more effectively address svarabhakti vowel productions since Gili (1921) found velar consonant interference on svarabhakti durational measurements. Furthermore, in Experiment A, native speaker svarabhakti production following [p] was not 100%. Placing [p] before svarabhakti targets had the effect of preventing a ceiling effect on the native speaker svarabhakti emergence. A more normally distributed range of scores would allow for more meaningful inferential comparison of the native speaker results to scores for L2 groups.

The Spanish word lists were also modified to include three variations of carrier phrases instead of one in order to test the effect of preceding vowel height on trill accuracy. In Experiment A, the Spanish target words were always placed within the carrier phrase *Dice* _____ *también.* 'He/she says ___ also.'; however, the [e] preceding all initial-trill targets in *Dice* did not provide variation in vowel height as was found before word-medial and word-final <r> targets (e.g., *Dice carro también*). The results from Experiment A, as well as previous research (Carballo and Mendoza 2000), suggested that trills may be more difficult to produce following high vowels as mentioned in Section 3.1.3. Therefore, in order to fairly compare word-initial vs. word-medial trill accuracy scores, the carrier phrases *Habla* ____ *también* 'He/she speaks ____ also' and *Ya vi* ____ *también* 'I already saw ____ too' were included alongside the carrier phrase *Dice* ____ *también.* Thus, in Experiment B vowel height was varied before both word-initial and word-medial trills. The [i] in *Ya vi* always preceded *rico* 'rich,' the [e] in *Dice* always preceded *remo* 'I row,' and the [a] in *Habla* preceded all three tokens of *rato* 'short while'; however, in the case of all other targets, each of the three tokens appeared on each Spanish list in a different carrier phrase (e.g., the target word *codo* 'elbow' would have been found on one list beginning with *Ya vi,* another with *Dice,* and the third with *Habla*). The carrier phrases were randomly assigned (except in the case of initial trill targets), and evenly distributed across all three Spanish lists using a computer programming script created using the Perl programming language. See Appendix B for a sample Spanish list.

The English word list for Experiment B was also modified from Experiment A in order to visually conform to the multiple carrier phrases found in the Spanish lists. The single carrier phrase in Experiment A, "He said ____ to me", was replaced in Experiment B with three carrier phrases: "He'll see ____ today", "He said ____ today", and "He saw ____ today." In contrast to the Spanish lists, there was no reason to suspect differences in pronunciation of English targets based on the change in carrier phrases from Experiment A to Experiment B. Similar to the English lists in Experiment A, the purpose of the English list was simply to confirm that L1 English participants were producing English targets with standard American English forms.[25] The target words *treaty*, *gray*, and *draw* remained in the English lists at the first recording session S1 (week 1); however, they were replaced with *prequel*, *program*, and *product* for the second

[25] It happened that one original participant was excluded from the final analyses based on her English target productions. Although she had indicated English as her first language, her English productions indicated difficulty with /r/ and /l/. It was later verified that she also grew up speaking mandarin at home. Thus, she was excluded from final analyses due to potential confounds with her L1 background.

recording session S2 (week 9) in order compare [pɾ] targets in Spanish with expected L1 English [pɹ] targets. For a sample English list from Experiment B, see Appendix C.

4.2.3 Experiment B: Recording procedures and data collection

Similar to Experiment A, participants were individually recorded in a sound-attenuated booth at the University of Washington Linguistics Phonetics Laboratory. Participants were recorded in two separate sessions over an 8-week period corresponding to the Summer quarter's 1st and 9th week, respectively. Prior to the first recording, each participant completed a consent form and language background survey. The language background survey was the same form utilized in Experiment A (see Appendix A). The recording procedures and data collection in Experiment B were the same as those in Experiment A (see Section 4.1.3).

4.3 Native Spanish speaker participants

4.3.1 Native Spanish speaker backgrounds

In order to provide native speaker baseline data that would serve as the basis for the L2 accuracy criteria (following reasons discussed in Section 2.4), a set of native speakers was also recorded as part of Experiment B. The native Spanish speakers included 11 speakers (five males and six females) from Spain, Mexico, Venezuela, Colombia, and Uruguay. The average age across all speakers was 34 years. All native speakers were university educated and working in some capacity of Spanish language teaching at the time of recording. While dialectal differences do exist (see Section 3), the purpose of the analysis procedures was to minimize potential dialectal differences by extracting only those auditory and acoustic features present across all native speaker productions. Table

4.5 provides a summary of the native speakers and their backgrounds. For example, speaker NS-F4 was 24 years old at the time of her recording and a native Spanish female from Spain.

Table 4.5 Summary of native Spanish speaker backgrounds.

Spanish Level	Subject Code	Gender	Age	Country of origin
Native Spanish speakers	NS-M1	M	35	Venezuela
	NS-M2	M	41	Mexico
	NS-M3	M	48	Spain
	NS-M4	M	26	Spain
	NS-M5	M	25	Colombia
	NS-F1	F	27	Spain
	NS-F2	F	47	Spain
	NS-F3	F	47	Mexico
	NS-F4	F	25	Spain
	NS-F5	F	29	Mexico
	NS-F6	F	26	Uruguay
		average age:	*34*	

4.3.2 Materials and recording procedures

Native speakers were recorded in a single session. The first five native Spanish speakers were given the same word lists described in Experiment A (see Section 4.1.2). A second group of six native Spanish speakers read from the materials used in Experiment B (see Section 4.2.2). Although all native speakers completed a consent form, unlike the L2 participants, native speakers did not complete the language background survey. Instead, their native speaker status, age, and country of origin were orally verified. In regards to data collection, the first group followed the procedures outlined in Experiment A (see Section 4.1.3) while the second group was recorded using the steps described in Section 4.2.3. In addition, all native speakers read from a corresponding English word list in order to ensure that they were following the same recording procedures as the L2 subjects; however, the English productions by the native Spanish speakers are not addressed in this investigation.

4.4 Acoustic analysis procedures

The purpose of this section is to describe the auditory and acoustic analysis procedures applied to all productions by all groups. Measurement procedures for each phonological category are first presented without reference to accuracy scoring (for details on the auditory and acoustic features used to score L2 productions, see Section 5.2); however, it should be noted that the measurement procedures here were finalized following informal inspection of native speaker realizations. That is, it was first necessary to identify any limitations to the acoustic signal that could potentially be affected by the recording procedures and experimental design (e.g., microphone and gain settings). Following the establishment of the measures for each target category, the native speaker productions were reanalyzed and measured accordingly.

While some measures presented in this section were used to assess segment accuracy, other measures simply allowed for independent temporal and spectral descriptions of productions. These descriptions are aimed at providing an independent assessment of interlanguage development and individual variation without necessarily referencing either the source or target language (e.g., understanding "inaccurate" productions). In addition, productions of word-final <r> were measured across all participants and their characteristics are later discussed; however, the word-final <r> target was not judged for accuracy due to the extreme variability found in the native speaker productions. Thus, this section is applicable to seven target categories although accuracy scores were only determined for six targets. Each of the following sections references example waveforms and spectrograms found at the end of Chapter 5.

For the acoustic analyses, Praat version 4.0.1 (Boersma and Weenink 2003)was used to generate waveforms and spectrograms for each sound. The spectrogram was set to use a 300Hz filter. For females, the formant analysis range extended to 5500Hz; 5000Hz for males. Praat automatically re-sampled the signal to twice the frequency range for spectrogram calculation. Spectrogram intensity levels were set to 55dB. For the native Spanish speakers, a total of 693 Spanish tokens were analyzed (99 tokens for each of seven target categories, including word-final <r>). For the L2 participants, 4410 Spanish tokens were analyzed.

4.4.1 Measurement of word-medial stop /d/ (allophone [ð])

Each segment waveform and spectrogram was visually inspected from a 1 second (s) window. In the case of spirantized productions, overall segment duration was measured on the spectrogram from the abrupt decline in F1 or F2 to the following abrupt increase of overall intensity into the following vowel. In certain cases of little to no vocal tract constriction at the target, little if any change in overall intensity was evident (except for perhaps a brief dip in the intensity line). For these "approximated" instantiations, no durational measure was taken; however, the approximated nature of the production was noted. For sample measurements of /d/ productions, see Chapter 5, Figure 5.1 through Figure 5.6.

4.4.2 Measurement of word-medial stop /t/ (allophone [t])

All waveform and spectrogram productions of [t] were viewed and measured from a 250ms analysis window. Two temporal measures were taken from the spectrogram. The first durational measure reflected the period of stop closure as indicated from the disappearance of F2 in the preceding vowel through the absence of clear formant structure prior to the release burst. A release burst was shown by a spike in the waveform and a corresponding dark line in the spectrogram at the moment of release. The second temporal measure reflected any positive voice onset time as indicated from the release burst to the onset of sinusoidal patterning on the waveform. See Figure 5.7 through Figure 5.9 in Chapter 5 for example waveforms and spectrograms of [t] productions.

4.4.3 Measurement of word-initial and word-medial trill /r/ (allophone [r])

Measurements of word-medial and word-initial /r/ followed the same procedure. All productions were analyzed from a 1s window. For each production, three measures were obtained: 1) overall duration, 2) degree of voicing, and 3) number of stripes. A "stripe" was taken to indicate the absence of glottal pulses across one or more formants. Overall duration was measured from the beginning of the first stripe to the end of the second stripe before the following vowel. Voicing was noted in the voicing bar of the spectrogram; the voicing bar was taken to correspond to a dark band of energy below the first formant, corresponding to the speaker's F0. A value of "0" was used to reflect no voicing, "1" represented partial voicing in which at least one (but not all) tap occlusions were followed by a voicing bar, and "2" characterized the presence of a voicing bar following each tap occlusion. In Chapter 5, Figure 5.10 through Figure 5.13 show measurements and spectral features of word-initial trill productions; Figure 5.14 through Figure 5.17 show word-medial productions.

4.4.4 Measurement of word-final <r>

Similar to word-initial and word-medial trills, all realizations of word-final <r> were analyzed from a 1s window in Praat; however, only an overall duration measure was obtained for these productions due to the extreme acoustic variability found in native speaker productions. The beginning of the segment was indicated by a disruption in one or more formants while the end point for measurement was determined by the simultaneous disappearance of energy in the waveform and spectrogram. Figure 5.18 through 4.22 demonstrate measurement of these features in word-final <r>.

4.4.5 Measurement of word-medial /ɾ/ (allophone [ɾ])

All word-medial /ɾ/ target waveforms and spectrograms were analyzed from a 1s window. For each production, overall segment duration was measured as the period corresponding to a spectral "stripe" as indicated by the absence of glottal pulses across one or more formants. See Figure 5.23 through Figure 5.26 for measurements of word-medial /ɾ/ productions.

4.4.6 Measurement of onset cluster /ɾ/ (allophone [ɾ])

Taps in onset clusters were first inspected from a 1s window to determine the presence or absence of a spectral "stripe" as indicated by the absence of two or more glottal pulses across one or more formants. Following the 1s window inspection, the segment was analyzed from a 250ms window and the following measures were taken: 1) Overall onset cluster duration, as measured from the release of the consonant to the onset of sinusoidal patterning in the following nucleic vowel, 2) if present, the duration of the svarabhakti vowel element as indicated by brief sinusoidal patterning between the initial consonant release burst and spectral "stripe" corresponding to the location of tap constriction, and 3) if present, the distance from the start of the spectral "stripe" to the onset of sinusoidal patterning in the following nucleic vowel. In Chapter 5, Figure 5.27 through Figure 5.30 show example measures for /ɾ/ in onsets.

4.5 Statistical methods

The results of this investigation present the L2 production accuracy scores using descriptive and inferential methods. Descriptively, raw accuracy scores for each category are used to provide a percentage of accurately scored productions out of 9 attempted targets. For example, a raw accuracy score of ".33" in a given session for a given target reflects 33% accuracy, or 3/9 accurately scored productions. At each session, raw accuracy scores were divided by the native speaker *variability constant* (described in Section 5.1) to yield an adjusted accuracy score which is normalized for the native speaker production variation. For example, an overall raw accuracy score of ".88" for /t/ would be divided by ".99" (since ".99" was the native speaker variability constant) to yield an adjusted overall accuracy score of ".89" or 89%. Table 4.6 demonstrates the application of the variability constants to the word-medial overall raw accuracy scores for beginning and intermediate level participants in Experiment A.

Table 4.6 The application of variability constants (VC) to word-medial raw accuracy scores (RAW) and resulting adjusted overall accuracy (RAW/VC = ADJ). Only participants from the beginning and intermediate levels of Experiment A are shown.

	Word-medial /d/		Word-medial /t/		Word-medial /r/		Word-medial /ɾ/	
	RAW	ADJ VC = 99%	RAW	ADJ VC = 99%	RAW	ADJ VC = 89%	RAW	ADJ VC = 93%
A-Beg-M1	26%	26%	96%	97%	0%	0%	81%	87%
A-Beg-M2	4%	4%	81%	82%	0%	0%	74%	80%
A-Beg-F1	56%	57%	93%	94%	67%	75%	7%	8%
A-Beg-F2	67%	68%	100%	100%	0%	0%	30%	32%
A-Int-M1	0%	0%	100%	100%	0%	0%	0%	0%
A-Int-M2	11%	11%	100%	100%	89%	100%	74%	80%
A-Int-F1	100%	100%	100%	100%	0%	0%	45%	48%
A-Int-F2	0%	0%	85%	86%	0%	0%	7%	8%
A-Int-F3	15%	15%	100%	100%	0%	0%	0%	0%

Inferentially, observed rankings to the adjusted overall accuracy scores of each category were tested using a Monte Carlo permutation test. A permutation test evaluates whether the observed ordination or relationship is stronger than chance (Good 2000). The rationale for the permutation test can be explained as follows: The results of the accuracy scores have been presented in an arbitrary order (i.e., the columns for spirant, stop, trill, and tap accuracy could have been placed in any sequence on a spreadsheet). It might be claimed that it was the placement of those columns that revealed a particular ranking or relationship of accuracy scores; however, a permutation test asks, what is the chance that randomly rearranging the columns could produce just as many patterns of *any* ranking and not just the observed pattern? In other words, if the null hypothesis were true (i.e., there is no systematic ranking to the accuracy scores), then a permutation test determines the chance of randomly generating any ranking as extreme as the observed ranking. The test statistic, *m*, is reported for each permutation test and reflects the ratio of observed rankings out of 23 participant rankings (all learner level participants). The Monte Carlo permutation test was completed using a program I created using the Perl programming language (Practical Extraction and Report Language).[26]

A similar Monte Carlo permutation test was used to address the question of the effect of vowel height on accurate /r/ productions. Thus, /r/ accuracy scores are separated for scores for /r/ in high vowel environments (HI), mid vowel environments (MID), and low vowel environments (LOW). Instead of using observed rankings of trill accuracy scores, three test statistics were generated: $q^{(H\text{-}L)}$ reflects the average across all participants of HI scores minus LOW scores, $q^{(H\text{-}M)}$ is the average across all participants of HI scores minus MID scores, and $q^{(M\text{-}L)}$ averages all MID minus LOW scores. For example, 24 non-native subjects (across all learner and experienced L2 subjects) from both experiments evidenced some positive trill accuracy scores in either word-medial or

---

[26] Special thanks to Mathias Drton and Paul Scheet in the Department of Statistics at the University of Washington for all their help and patience in outlining this algorithm. Additionally, Paul Sampson and several graduate students have been very helpful in providing advice on appropriate tests.

word-initial position. Thus, each participant has a value that results from subtracting his or her score of trill accuracy in low vowel environments from his or her score of trill accuracy in high vowel environments (e.g., if "John" has 70% accuracy of trills in a high vowel environment, 82% accuracy for trills in a mid vowel environment, and 90% accuracy of trills in a low vowel environment, his HI-LOW, HI-MID, and MID-LOW values would be -20, -12, and -8, respectively). Averaging the resulting values across all 24 speakers would yield each q test-statistic. A Monte Carlo test trial randomly reassigns all the values in a row (i.e., reassigns the column headers for each participant), recalculates the HI-LOW, HI-MID, and MID-LOW values for each subject, and then averages the resulting values across all subjects. If the average is lower than the originally calculated q test-statistic, then the given trial randomly generated a relationship as strong as the one observed ("same-as-observed"). A given *p value* is then the number of "same-as-observed" trials divided by the total number of trials. The total number of trials used in this study is 10,000.

A chi square test of independence was used to test the null hypothesis that there were no differences between groups (including the native speaker group) for the binary svarabhakti vowel emergence scores (see Section 8.2.2). Svarabhakti vowel emergence scores for each participant ranged from 0% to 100%; however, the distribution of scores between groups was not normal. In addition, scores could not be transformed (e.g., by calculating the arcsine of the square root of each score) to yield a normal distribution. In effect, an ANOVA could not reliably be used as a test. Instead, scores for each participant were converted to a binary value where "1" represented native-like performance and "0" represented non native-like performance. The lowest native speaker svarabhakti vowel emergence score was 44%. Thus, any score higher than or equal to 44% was represented by "1" and any score lower than 44% was represented by "0." A chi square test then tested if the proportions of "1" (native-like SVE) and "0" (not native-like SVE) were independent of participant level. Following a significant result of the chi square test, a simple Bonferroni correction was performed on the various possible tests between each L2 level and the native speaker level (i.e., a chi square test for each 2x2 matrix). Based on the comparison of four pair-wise tests, the Bonferroni correction required a significance level lower than .0125 (.05/4) in order to be considered significant.

5. Native speaker baselines, defining accuracy, and L2 scoring

This chapter presents both native Speaker production results and methods of scoring L2 accuracy. This combined presentation of results and methods is necessary because the criteria for scoring L2 productions (i.e., defining accuracy) were determined based on native Spanish speaker productions. There are three important distinctions between the native speaker productions in Section 5.1 and the ultimate scoring of L2 realizations described in Section 5.2. First, results and averages include productions from all 11 native speakers; however, the criteria used to score L2 productions in Experiment A and B were originally determined based upon the productions of five native speakers who were recorded using Experiment A word lists. The auditory and acoustic features that were established as accuracy criteria in Experiment A were supported by the six additional native speakers in Experiment B and hence, were not altered between experiments. Second, the acoustic analyses of native speaker productions included additional temporal and spectral measures that were not necessarily incorporated into accuracy criteria; however, the additional detail may help to further compare L2 productions to native production features (e.g., the role of voice onset time in the accurate scoring of L2 productions of the medial /t/ target found in Section 6.2.2). Third, Section 5.1.5 includes a description of word-final <r> realizations which are commonly said to be realized as a tap or trill in Spanish; however, L2 word-final <r> accuracy was not calculated due to the wide variability across native speaker productions.

A *variability constant* was established for each target category, based on the variation within features found for the native speaker productions (see Section 5.1). While there was an initial attempt to extract only those features found in all native speaker productions of a particular category, variation across features did exist. A procedure was followed for each category that reflects the percentage of native speaker productions that evidenced both the auditory and acoustic features established for that category. Hence, the variability constant required the recalculation of L2 accuracy scores (i.e., normalization) based on the variation found in the native speaker productions of a particular category. For example, participant A-Beg-F1 received an overall raw accuracy of 67% for word-medial /r/. The variability constant for word-medial /r/ was 89%. Thus, her overall adjusted accuracy score was 75% (.67 / .89 = .75).

This chapter is organized into two major sections that set out results for the 2-part study: Section 5.1 presents the auditory and acoustic features observed in the native speaker productions of all categories across all eleven native speakers. Results are grouped by experiment and gender in order to identify any potential differences between these independent variables. All individual native speaker measurements can be found in Appendix D – I. Following the presentation of auditory and acoustic features, I identify the features most common to all native speaker productions that were used as the requirements for scoring the corresponding L2 productions as "1" (accurate). Finally, I calculate the variability constant that was then used for each category. Section 5.2 demonstrates how the accuracy criteria were applied to scoring the L2 productions.

5.1 Auditory and acoustic features of native speaker productions: Defining accuracy criteria

5.1.1 Native speaker features of word-medial /d/

Auditory features: The two most common auditory features included continual voicing throughout the segment and the absence of an impression of complete obstruction in the oral tract. Auditorily, native speaker productions varied dramatically from the clear impression of an open oral tract to that of a narrow constriction. Auditory analyses of native speaker productions in Experiment B showed yielded similar impressions of oral tract constrictions.

Auditory criterion for word-medial /d/: *An impression of voicing in the production without the impression of stop closure was required for an accuracy score of "1" for this category.*

Acoustic features: Perhaps more than any other category, word-medial /d/ productions showed marked inter- and intra-speaker variation. All waveforms for /d/ were inspected from a 1s window in Praat. Realizations spanned a range from intense frication to no frication throughout the segment. Frication appeared as a reduction of energy throughout the frequency spectrum, and aperiodicity with the occasional absence of formant structure. In almost all productions, however, a voicing bar was present throughout the segment and was accompanied by a positive intensity line across the duration of the phone. In some productions, the formant and energy structure of a [ð] could barely, if at all, be distinguished from that of the surrounding vowels. In these "approximated" instantiations, the waveform remained periodic and spectrograms demonstrated a clear vowel-like formant structure such that both F1 and F2 of the approximated [ð] were interpolated between those values of the flanking vowels, suggesting a fairly open oral tract configuration. A dip in the intensity line across the spectrogram was sometimes the only acoustic indication of the approximated [ð] production.

Although all native speaker productions were able to be categorized into spirantized or approximated instantiations, there was variation with the range of frication or clarity of formant structure present in the productions.

Figure 5.1 shows a waveform and spectrogram for a typical spirantized token for /d/ with an overall segment duration of 45ms. Frication in the higher frequencies (>3000Hz) reveals its spirantized nature and, although not always consistent, this particular production shows a disruption of the formants during the segment.

Figure 5.2 and Figure 5.3 also show waveform and spectrograms of spirantized instantiations that demonstrate exceptionally long and short overall durations, respectively. A typical approximated instantiation is shown in Figure 5.4. Its "approximated" designation is characterized by a continuation of the formant structure

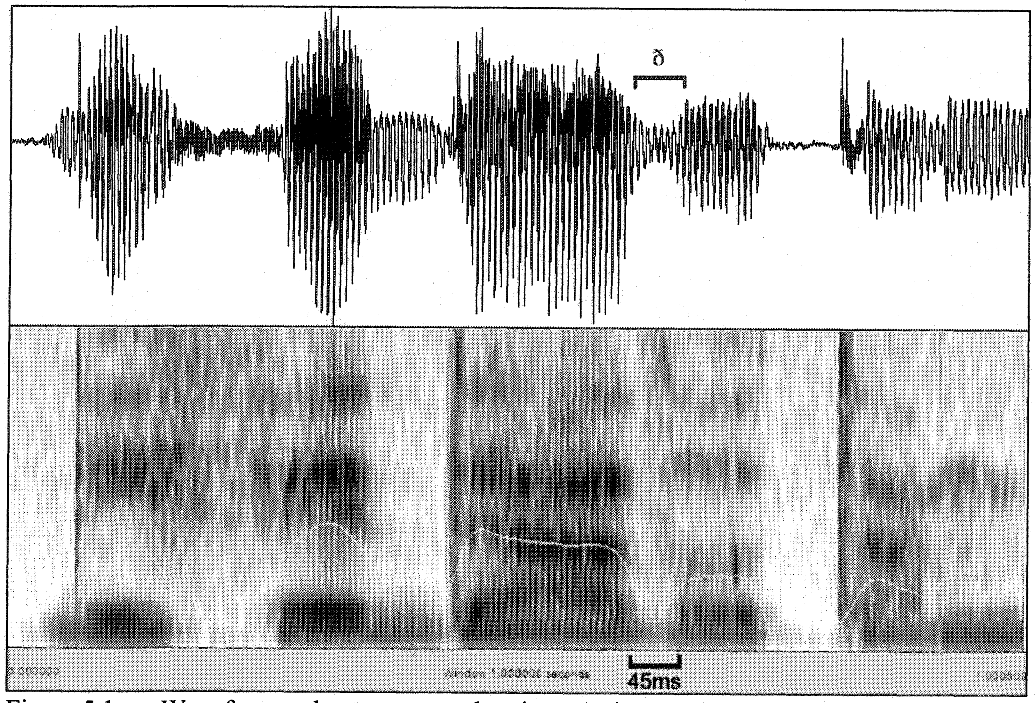

Figure 5.1 Waveform and spectrogram showing a native speaker production of word-medial /d/ designated as "spirantized." Token [dise daðo tam] in *Dice dado también,* participant NS-F1.

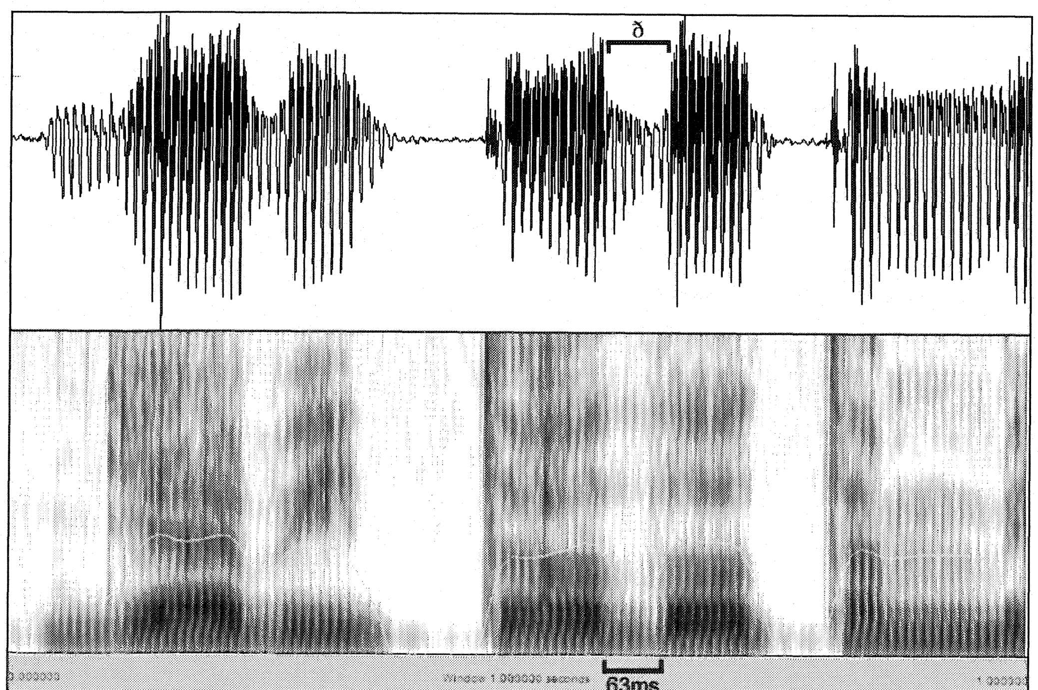

Figure 5.2 Waveform and spectrogram showing a native speaker production of word-medial /d/ with long overall duration. Token [ya βi toðo tam] in *Ya vi todo también,* participant NS-M4.

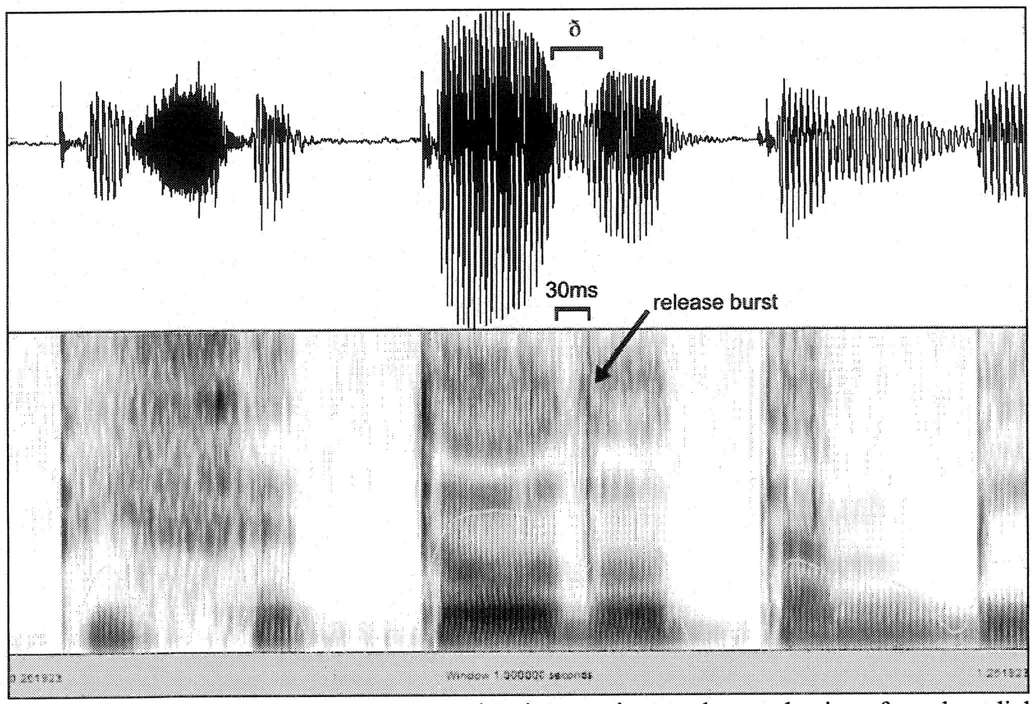

Figure 5.3 Waveform and spectrogram showing a native speaker production of word-medial /d/ with short overall duration and release burst. Token [dise toðo tamby] in *Dice todo también,* participant NS-F3.

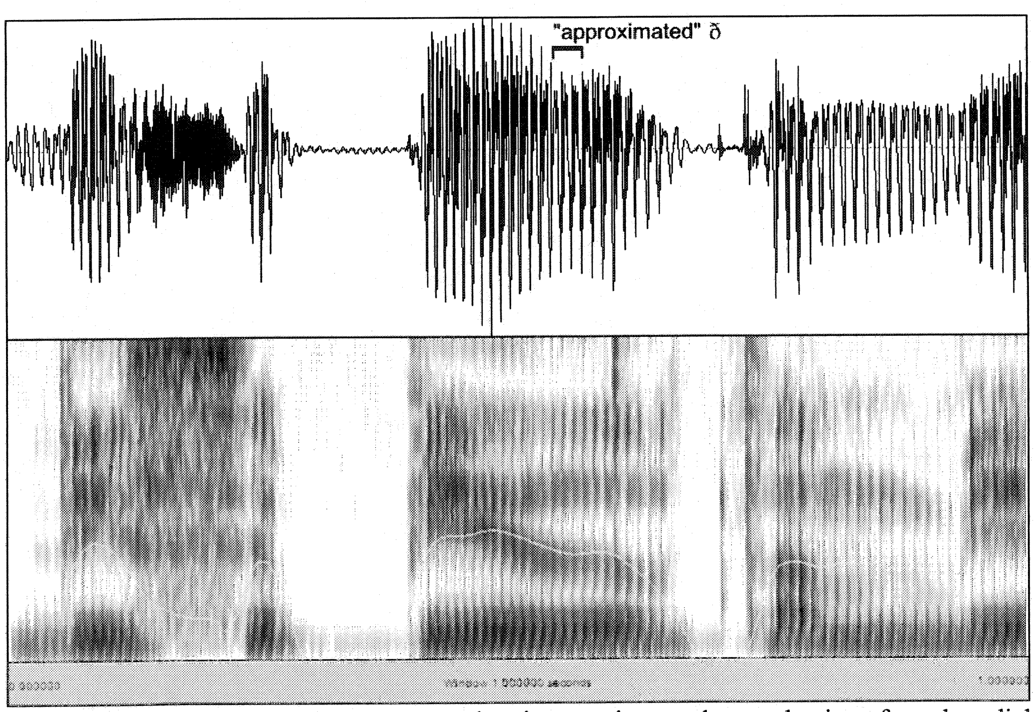

Figure 5.4 Waveform and spectrogram showing a native speaker production of word-medial /d/ designated as "approximated." Token [dise pið̞o tamby] in *Dice pido también,* participant NS-M5.

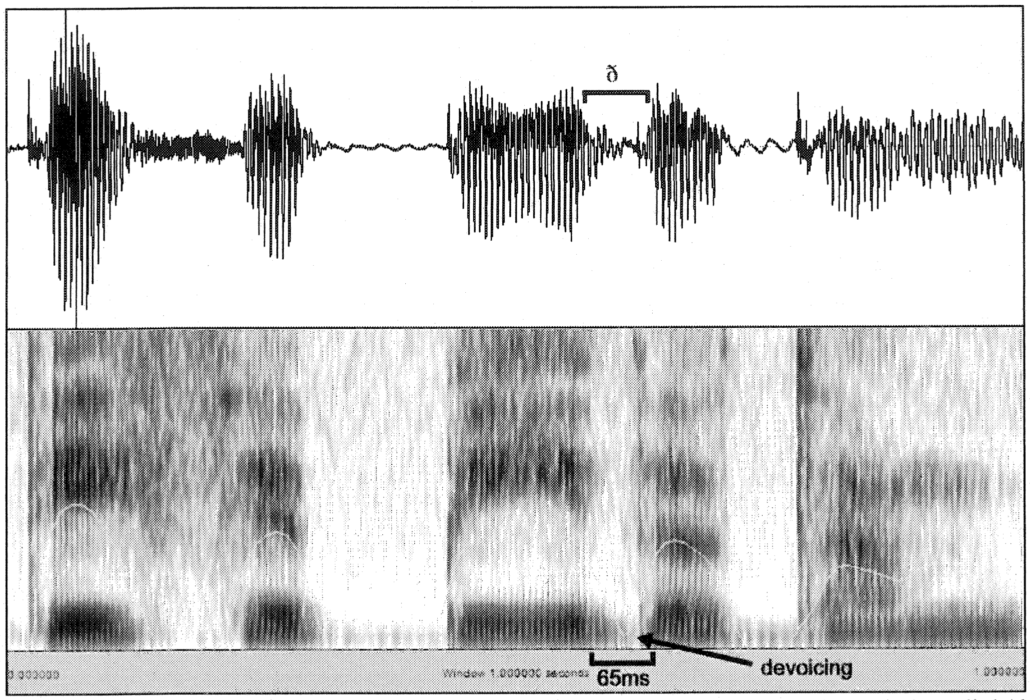

Figure 5.5 Waveform and spectrogram showing a native speaker production word-medial /d/ with partial devoicing. Token [dise piðo tam] in *Dice pido también,* participant NS-F1.

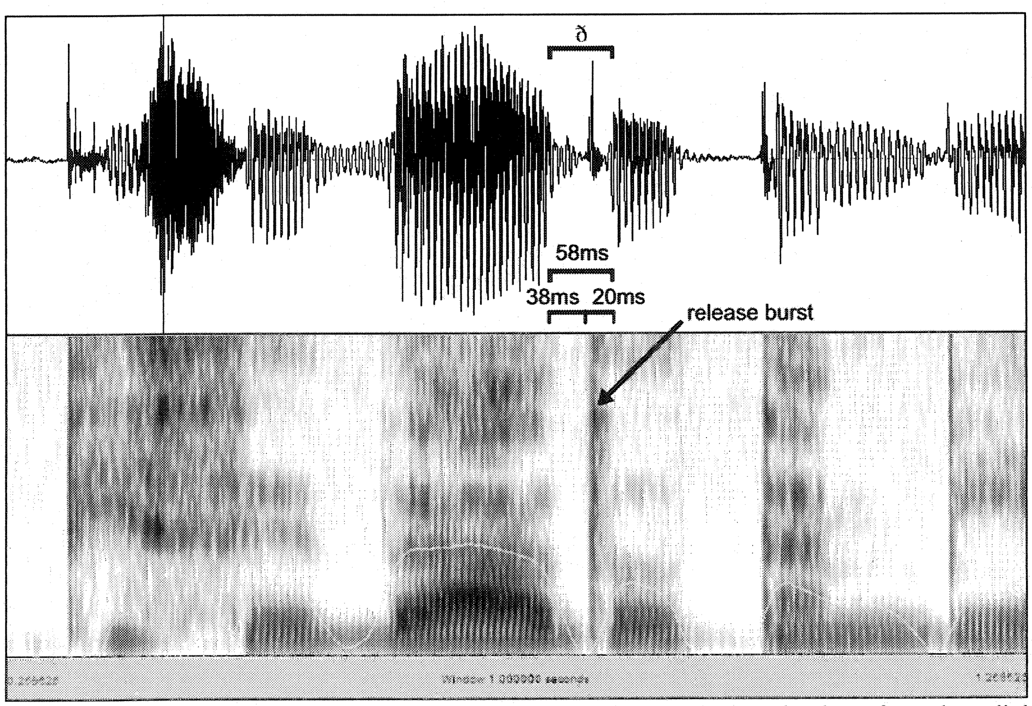

Figure 5.6 Waveform and spectrogram showing a native speaker production of word-medial /d/ with a long overall duration and release burst. Token [dise daðo tamby] in *Dice dado también,* participant NS-F3.

throughout the segment production, an absence of frication in the higher features, and no clear abrupt transition in intensity by which to visually demarcate the segment from the surrounding vowels. For all similar approximated instantiations, overall segment duration was not measured.

Regarding voicing, all but one token across both experiments demonstrated voicing throughout the segment as shown by a voicing bar across the spectrogram.[27] Despite the overwhelming presence of voicing in productions, speaker NS-F1 produced one token with partial devoicing indicated by a disappearance of the voicing bar and aperiodic signal in the waveform (see Figure 5.5). In regards to the presence of release bursts, 6 out of 73 spirantized instantiations were noted to have a release burst. Figure 5.3 and Figure 5.6 demonstrate two spirantized productions with the visual presence of a release burst. In these tokens, overall duration was taken from the start of the segment to the onset of sinusoidal voicing in the following nucleic vowel.

<u>Acoustic Measures</u>: The five native speakers from Experiment A produced 82% of /d/ targets as spirantized productions and 18% as approximated instantiations. In Experiment B, 67% of /d/ productions were spirantized and 33% were approximated. Thus, spirantized instantiations accounted for 74% of all productions and 26% represented approximated instantiations. Regarding productions by gender, 56% of all male (n=5) productions were identified as spirantized and 44% were found to be approximated. Females (n=6) yielded a greater frequency of spirantized instantiations and fewer approximated instantiations compared to male productions; out of 54 productions by female native speakers, 89% were designated as "spirantized" and 11% were designated as "approximated" instantiations. Table 5.1 shows the results of the spirantized vs. approximated instantiations by experiment and Table 5.2 details /d/ production types by gender.

[27] This feature is important as some L2 productions revealed all characteristics of the native speaker realizations for /d/ except voicing. In effect, some L2 productions were realized as voiceless dental approximants and were scored as "0" (inaccurate);

In Experiment A, overall durations of spirantized productions ranged from 30-72ms, averaging 44ms (11ms stdev) duration. Experiment B spirantized productions ranged from 41-73ms with an average duration of 54ms. Across all native speakers, overall duration was 49ms (11ms stdev) with a minimum duration of 30ms and a maximum of 73ms. Analyzing overall duration by gender, male durations were typically slightly longer than female durations. Males yielded average overall durations of 52ms (12ms stdev) in contrast to 48ms (10ms stdev) overall duration for female productions. Table 5.3 presents the overall durations by experiment while durations by gender are shown in Table 5.4.

Acoustic criteria for word-medial /d/: *In the case of spirantized tokens, all but one spectrogram showed an attenuation of the signal longer than 30ms.*[28] *Thus, a durational criterion was used to score acoustic accuracy for the intervocalic target phone /d/.*[29] *All tokens also demonstrated a voicing bar throughout most (>50%), if not all (100%) of the segment. Thus, the presence of voicing (>50% of the segment duration) and durational features (for spirantized productions) were the two acoustic features found in common across all native speaker productions of /d/ required for an accuracy scoring of "1" for this category.*

Variability constant for word-medial /d/: Out of 99 productions, 98 instantiations could be classified as "spirantized" or "approximate" and assigned a score of "1" for both auditory and acoustic accuracy. Thus, the variability constant for L2 productions of this category was 99%.

[28] One token was measured at 30ms.

[29] Although the durational criterion was irrelevant to categorize a /t/ phoneme production, the durational requirement was necessary for /d/ since the visual absence of glottal pulses for less than 30ms was characteristic of native speaker tap productions.

Table 5.1 Frequency of occurrence of "spirantized" vs. "approximated" native speaker productions of word-medial target /d/ by experiment.

		Word-medial /d/	
		Total "spirantized" productions	Total "approximated" productions
Exp. A native speaker productions (n=45)	n	37	8
	frequency	82%	18%
Exp. B native speaker productions (n=54)	n	36	18
	frequency	67%	33%
All native speaker productions (n=99)	n	73	26
	frequency	74%	26%

Table 5.2 Overall duration by experiment of word-medial /d/ spirantized productions.

		Word-medial /d/
		Overall duration (ms)
Exp. A native speaker productions (n=37)	average	44
	stdev	11
	min	30
	max	72
Exp. B native speaker productions (n=36)	average	54
	stdev	8
	min	41
	max	73
All native speaker productions (n=73)	average	49
	stdev	11
	min	30
	max	73

Table 5.3 Frequency of occurrence of “spirantized” vs. “approximated” native speaker productions of word-medial /d/ by gender.

		Word-medial /d/	
		Total "spirantized" productions	Total "approximated" productions
Male (n=5) native speaker productions (n=45)	n	25	20
	frequency	56%	44%
Female (n=6) native speaker productions (n=54)	n	48	6
	frequency	89%	11%

Table 5.4 Overall duration of spirantized productions of word-medial /d/ by native speakers by gender.

		Word-medial /d/
		Overall duration (ms)
Male (n=5) spirantized productions (n=25)	average	52
	stdev	12
	min	36
	max	73
Female (n=6) spirantized productions (n=48)	average	48
	stdev	10
	min	30
	max	72

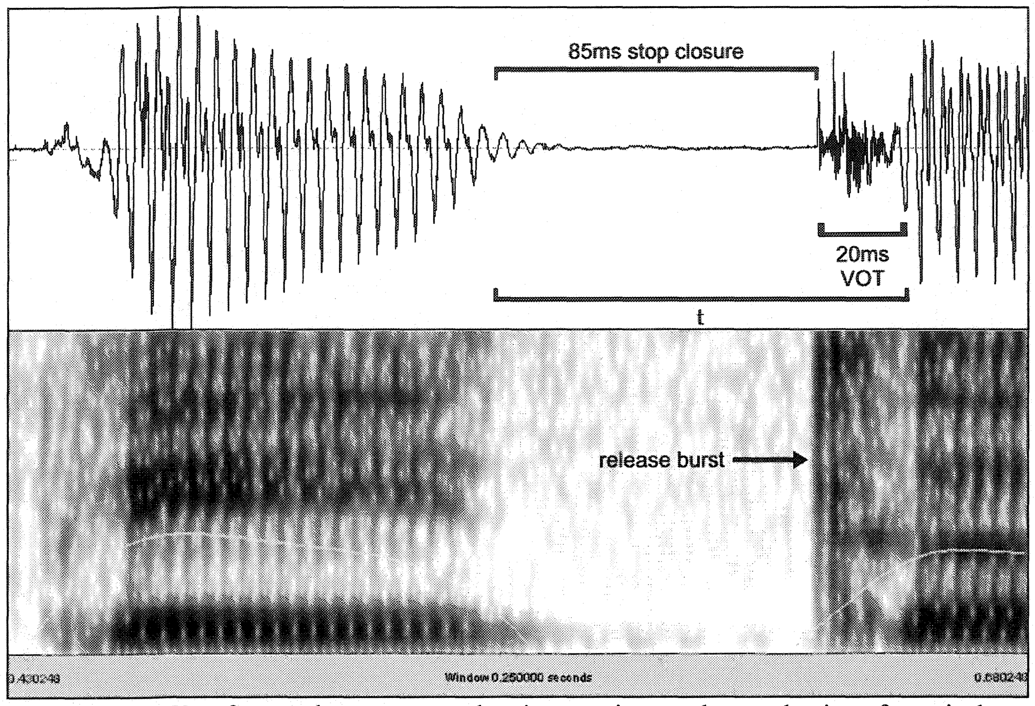

Figure 5.7 Waveform and spectrogram showing a native speaker production of a typical word-medial /t/. Token [pito] in *Dice pito también,* participant NS-F6.

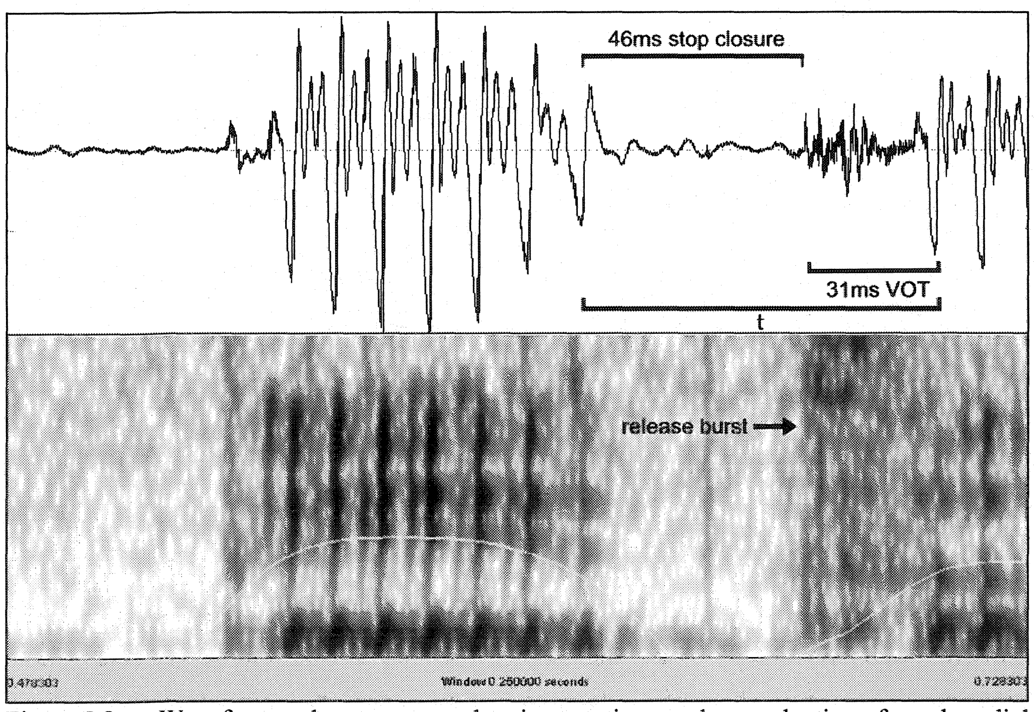

Figure 5.8 Waveform and spectrogram showing a native speaker production of word-medial /t/ with short stop closure duration and long voice onset time (SCD/VOT = 1.5). Token [pito] in *Dice pito también,* participant NS-M1.

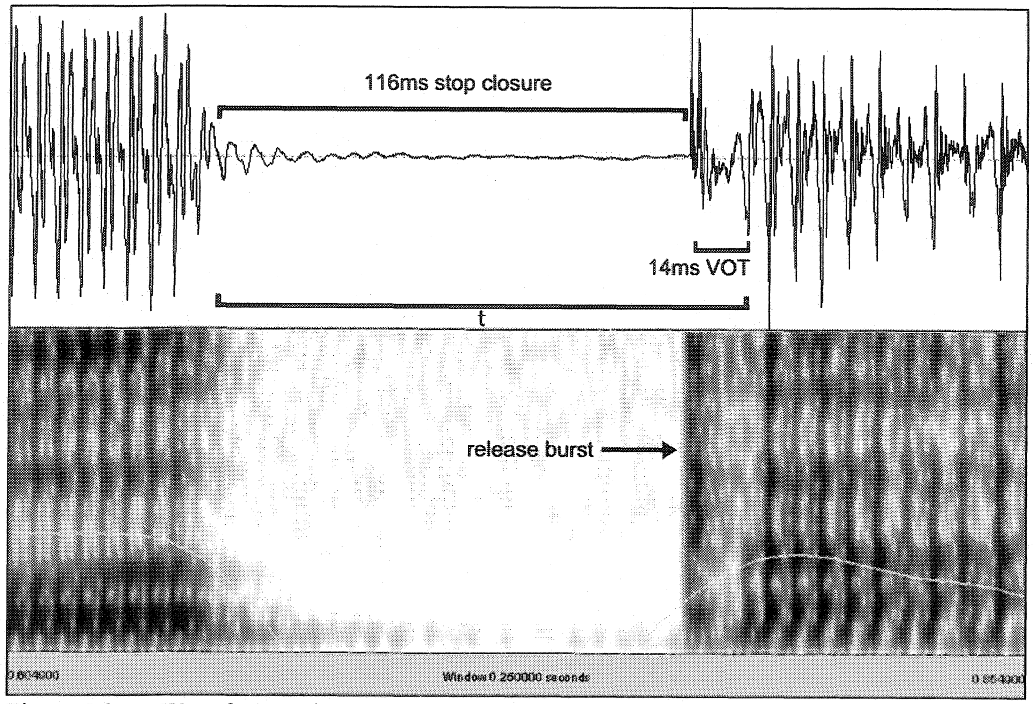

Figure 5.9 Waveform and spectrogram showing a native speaker production of word-medial /t/ with long stop closure duration and short voice onset time (SCD/VOT = 8.3). Token [ato] in *Habla dato también,* participant NS-F2.

5.1.2 Native speaker features of word-medial /t/

Auditory features: In Experiment A productions, 43 out of 45 productions gave the auditory impression of a complete obstruction of the oral tract. In two instantiations (by the same male speaker) a contact was not audible, but the segment seemed narrowly constricted in the vocal tract. In Experiment B, all 45 productions gave the impression of a complete oral tract obstruction.

> ***Auditory criterion for word-medial /t/***: *Therefore, the impression of oral tract obstruction was the common auditory feature among all native speakers and served as the auditory criterion for L2 productions.*

Acoustic features: With the exception of two productions, all native speaker realizations evidenced the acoustic features of a stop closure, release burst, and positive voice onset time.

Following the analyses, a third measurement representing the ratio of stop closure duration to voice onset time (SCD:VOT ratio) was also calculated in an attempt to provide a normalized temporal measure for the relationship between the two distinct portions of the /t/ segment.[30] Figure 5.7 shows a typical /t/ realization with a characteristic stop closure duration, voice onset time duration and SCD:VOT ratio. A production with uncharacteristically long SCD:VOT ratio is shown in Figure 5.8 while a realization with a relatively short SCD:VOT ratio is found in Figure 5.9.

Table 5.5 provides a summary of measurements for /t/ productions by experiment. Native speaker productions in Experiment A yielded 82ms (22ms stdev) of stop closure duration (SCD) compared with 86ms of SCD for Experiment B productions. Across both experiments, average SCD was 84ms (18ms stdev) and ranged from 29-127ms.

---

[30] The initial thought was to use the ratio range as a criterion in the re-evaluation of /t/ accuracy in the L2 productions; however, the native speaker ratio range was so great that it was not meaningful and raw VOT range was used instead. See Section 6.2.2 for the application of the VOT criterion for /t/ accuracy.

Regarding VOT, all productions evidenced positive values ranging from 9-41ms. Productions in Experiment A averaged 22ms (7ms stdev) VOT and in Experiment B, 23ms (6ms stdev). Across all speakers, average VOT was 23ms (6ms stdev). The ratio of SCD:VOT was similar for both groups as well. Values ranged from 0.7:1 to 13.6:1 with an average ratio of 4.2:1 (1.9 stdev).

An analysis of the measurement results by gender revealed slightly longer SCD values for females over values for males in contrast to longer VOT values for males than values for females. The SCD value for males averaged 76ms (16.4ms stdev) compared to 70ms (18ms stdev) for females. In regards to VOT, the average across all male productions was 25ms (5.8ms stdev) in contrast to an average VOT for female productions of 20ms (5ms stdev). SCD:VOT ratios for females varied slightly more than values for males; the average ratio for males was 3.3:1 (stdev 1.3) while the average ratios for females was 5.0:1 (2.0 stdev). See Table 5.6 for measurement results by gender.

> ***Acoustic criteria for word-medial /t/:*** *A stop closure, release burst, and positive voice onset time were the three acoustic features used as baselines for scoring word-medial /t/.*[31] *Due to the difficulty of using acoustic measures to determine dental versus alveolar place of articulation (Jongman and Blumstein 1985), place of articulation was not used as a criterion. Hence, a score of "1" in this category reflects an accurate /t/ production (in contrast to /d, r, ɾ/) and does not reflect range of positive voice onset time or place of articulation.*

---

[31] Although the durations of stop closures varied (29-127ms) as well as voice onset durations (9-41ms), durational measurements were not initially included in the native speaker baselines for the medial /t/ category. Thus, an L2 realization of [t^h] for the target /t/ was considered accurate as long as a contact was auditorily present and the stop closure, release burst, and positive voice onset time were acoustically visible. Even if a /t/ target was aspirated, it was still assumed to be a /t/ underlyingly and not a /d, r, ɾ/; however, question 4 of this study addresses the inclusion of voice onset time as an accuracy criterion and its effect on /t/ accuracy score rankings (see Section 5.2.3 for results).

Table 5.5 Stop closure duration and voice onset time for native speaker productions of word-medial [t] by experiment.

		Word-medial /t/		
		Stop closure duration (SCD) (ms)	Voice onset time (VOT) (ms)	SCD:VOT (VOT = 1)
Exp. A native speaker productions (n=43)	average	82	22	4.3
	stdev	22	7	2.5
	min	29	9	0.7
	max	127	41	13.6
Exp. B native speaker productions (n=54)	average	86	23	4.1
	stdev	15	6	1.4
	min	54	14	1.9
	max	122	38	6.7
All native speaker productions (n=97)	average	84	23	4.2
	stdev	18	6	1.9
	min	29	9	0.7
	max	127	41	13.6

Table 5.6 Stop closure duration (SCD) and voice onset time (VOT) for native speaker productions of word-medial [t] by gender.

		Word-medial /t/		
		Stop closure duration (SCD) (ms)	Voice onset time (VOT) (ms)	SCD:VOT (VOT = 1)
Male native speaker productions (n=43)	average	76	25	3.3
	stdev	16.4	5.8	1.3
	min	29	14	0.7
	max	102	41	6.4
Female native speaker productions (n=54)	average	90	20	5.0
	stdev	18	5	2.0
	min	47	9	1.9
	max	127	34	13.6

Variability constant for word-medial /t/: *The variability constant was determined to be 98% based on the result that 97 out of 99 native speaker productions were in agreement with the auditory and acoustic accuracy criteria.*[32]

5.1.3 Native speaker features of word-medial /r/

Auditory features: In Experiment A, 42 out of 45 native-speaker word-medial /r/ productions were characterized by the impression of a succession of rapid and complete obstructions of the oral tract. In three instantiations, the signal did not demonstrate multiple obstructions, but rather near constriction. In Experiment B, 52 out of 54 instantiations gave the impression of two or more obstructions of the vocal tract.

Auditory criterion for word-medial /r/: *The impression of two or more successive obstructions was identified as the auditory feature in common between native speaker productions because productions varied between the impression of being voiced/voiceless and number of obstruction.*

Acoustic features: All medial trill productions were visually inspected from a 1s window. Although not always present, a "W"-shaped intensity line often accompanied a voiced trill production. In the case of a voiceless trill, a "U"-shaped intensity line was typically evident. The dips in the intensity line corresponded to periods of less energy on the waveform and spectrograms; peaks in the intensity line were consistent with periods of greater energy. Almost all trills demonstrated two or more closures in the oral tract (39/45 in Experiment A; 37/45 in Experiment B). In some tokens, as many as 4 closures could be clearly identified. Some segments showed multiple closures in which the first closure was voiced, while subsequent closures were voiceless as indicated by a rising-to-falling intensity line. In some instantiations of voicelessness, closures did not appear as

[32] Two tokens did not reveal a release burst.

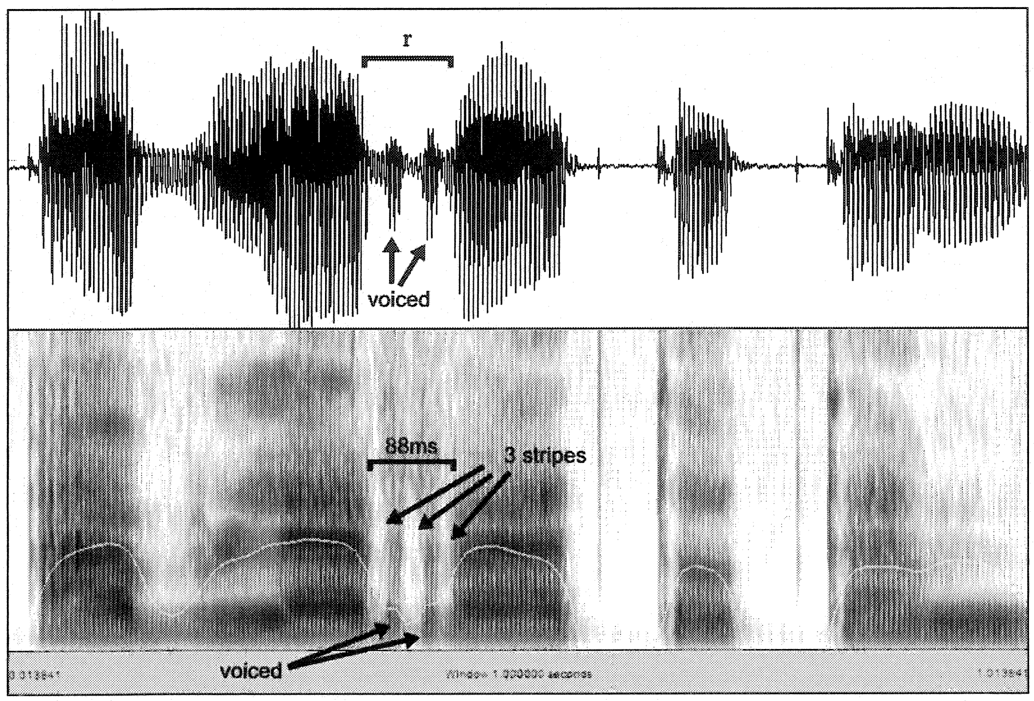

Figure 5.10 Waveform and spectrogram showing a typical native speaker production of word-initial /r/. Token [aβla rato ta] in *Habla rato también,* participant NS-F5, assigned "2" (fully voiced) for voicing value and "3" for number of stripes.

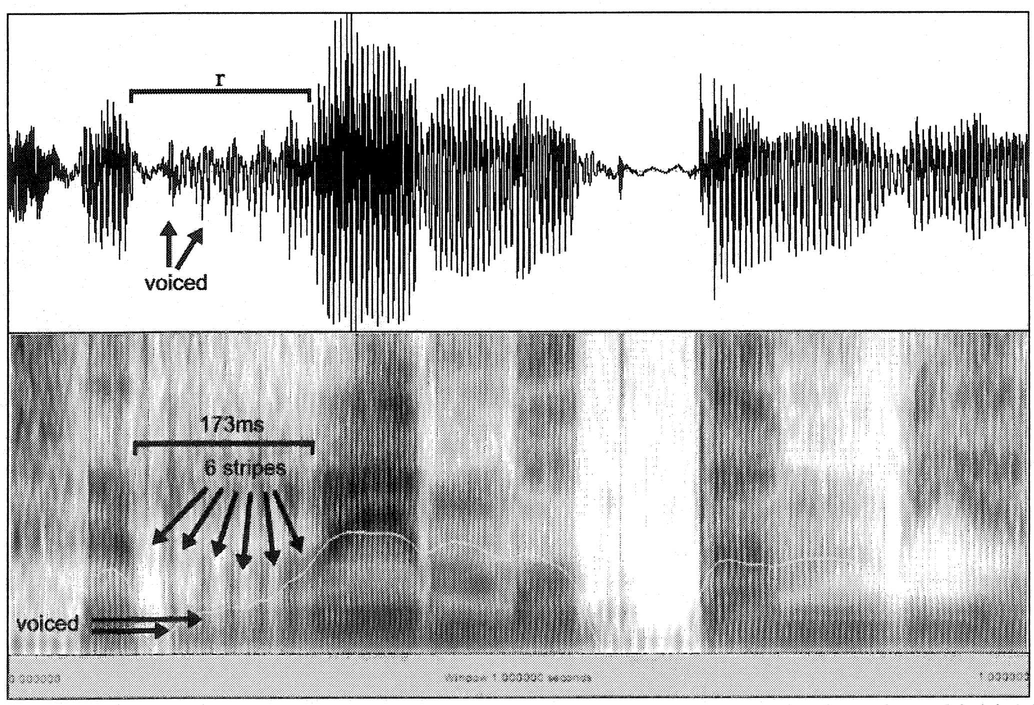

Figure 5.11 Waveform and spectrogram showing a native speaker production of word-initial /r/ realized with 6 stripes (i.e., closures). Token [se remo tambye] in *Dice remo también,* participant NS-F2, assigned "2" for voicing value (fully voiced) and "6" for number of stripes.

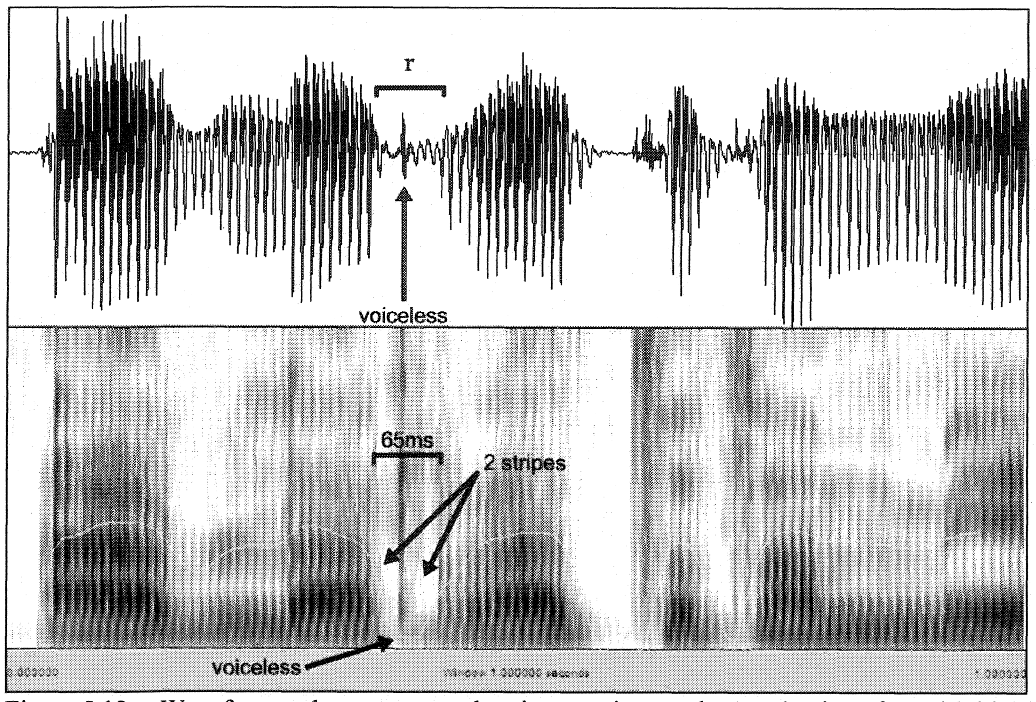

Figure 5.12 Waveform and spectrogram showing a native speaker production of word-initial /r/ realized without voicing. Token [aβla rato tamby] in *Habla rato también,* participant NS-M4.

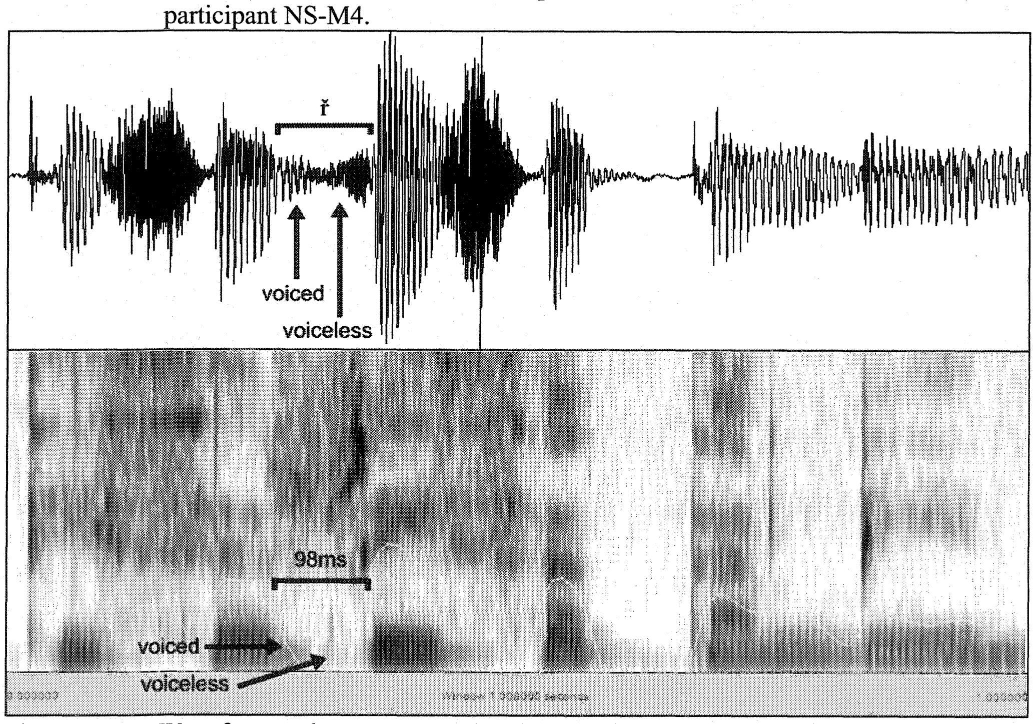

Figure 5.13 Waveform and spectrogram showing a native speaker production of word-initial /r/ realized with partial voicing. Token [dise risa tambye] in *Dice risa también,* participant NS-F3, assigned a voicing value of "1" (partially voiced) and "0" for number of stripes.

spikes on the waveform but rather as invisible "stripes" (i.e., the absence of glottal pulses in the spectrogram). When used in this paper, "stripes" refers to the near absence of visibility of 2 or more glottal pulses on the spectrogram. The typical production was characterized by voicing throughout the segment with 2-3 stripes. See Figure 5.10 for a typical word-medial /r/ production.

In terms of voicing, 81% of all tokens were given a value of "2" (fully voiced), 11% a value of "1" (partially voiced), and 8% were given the value "0" (voiceless). For these voiceless productions 3 out of 8 showed no evidence of visible stripes while 5 out of 8 demonstrated 1-3 stripes. Figure 5.11 shows a voiceless realization with 3 visible stripes.

The presence of frication with the absence of any visible stripes was found in 8% of all tokens, similar to production of some word-initial realizations. The majority of these fricated instantiations (5 out of 8 tokens) evidenced complete voicing throughout the segment as shown in Figure 5.12. In Figure 5.13, a realization characterized as voiceless with no visible stripes is presented. 3% of all word-medial /r/ tokens showed only 1 visible stripe. In sum, 89% of all word-medial productions evidenced 2 or more stripes.

A durational measure was taken of each /r/ production by measuring from the first closure to the onset of voicing in the following vowel. In Experiment A, productions averaged 78ms (18.8 stdev) with a range of overall durations from 53-138ms. Productions in Experiment B similarly averaged 74ms (15.9 stdev) from 52-113ms in range. Across all native speaker productions of word-medial /r/, 76ms was the average duration (17.4ms stdev). Regarding degree of voicing (in which tokens were assigned a scores of 2, 1, or 0 as described above), averaging productions in each experiment yielded a value of 1.7 (0.6 stdev). In terms of number of stripes, realizations in Experiment A resulted in a value of 2.2 (1.1 stdev) while in Experiment B, average number of stripes was 2.1 (0.6 stdev). Across all productions by both groups, 2.2 (0.8 stdev) was the average number of stripes with 2 as the most common number (i.e., the mode). Table 5.7 details word-medial /r/ measurement values by experiment.

Table 5.7 Measurement results of native speaker word-medial /r/ by experiment: Overall duration, degree of voicing and number of stripes.

		Word-medial /r/		
		Overall duration (ms)	Degree of voicing	Number of stripes
Exp. A native speaker productions (n=45)	average	78	1.7	2.2
	stdev	18.8	0.6	1.1
	min	53	0	0.0
	max	138	2	4.0
Exp. B native speaker productions (n=54)	average	74	1.7	2.1
	stdev	15.9	0.6	0.6
	min	52	0	0
	max	113	2	3
All native speaker productions (n=99)	average	76	1.7	2.2
	stdev	17.4	0.6	0.8
	min	52	0	0
	max	138	2	4

Table 5.8 Measurement results of native speaker word-medial /r/ by gender: Overall duration, degree of voicing and number of stripes.

		Word-medial /r/		
		Overall duration (ms)	Degree of voicing	Number of stripes
Male (n=5) native speaker productions (n=45)	average	75	2	2.1
	stdev	14.7	0.6	0.6
	min	52	0	0.0
	max	113	2	3.0
Female (n=6) native speaker productions (n=54)	average	77	1.7	2.2
	stdev	19.4	0.6	1.0
	min	53	0	0
	max	138	2	4

Measurement values by gender yielded no apparent differences. Across all male productions, overall duration demonstrated a range from 52-113ms with an average of 75ms (14.7ms stdev). For female productions, values ranged from 52-138ms, yielding an average of 77ms (19.4ms stdev). Degree of voicing values for male productions averaged 2 (0.6 stdev) while female productions averaged 1.7 (0.6 stdev). All male productions demonstrated 2.2 stripes on average (1.1 stdev) in comparison to female productions that averaged 2.2 stripes (1.0 stdev). In Table 5.8, the measurement results of word-medial /r/ realizations are summarized by gender.

> ***Acoustic criteria for word-medial /r/:*** *Thus, based on these visual and temporal features, a score of "1" (accurate) for L2 productions required two or more stripes in the spectrogram and overall attenuation duration of at least 50ms.*

> ***Variability constant for word-medial /r/:*** *88 out of 99 productions evidenced the auditory and acoustic features established for accuracy in this category.*[33] *Thus, 89% was the variability constant determined for this category.*

5.1.4 Native speaker features of word-initial /r/

Auditory features: In Experiment A, auditory impressions of voicing, frication, and duration varied to some extent; however, in 42/45 instantiations, productions gave the impression of two or more complete obstructions of the oral tract. In Experiment B, 50/54 productions gave the impression of two or more complete oral tract obstructions.

> ***Auditory criterion for word-initial /r/:*** *Hence, the impression of two or more obstructions was selected as the common auditory feature among word-initial trills and required for an accuracy score of "1" for this category.*

---

[33] Three tokens did not evidence any voicing while eight other tokens only demonstrated zero or one stripe.

<u>Acoustic features</u>: Variation in the acoustic features of initial trills was similar to trill productions word-medially. Visually inspected from a 1s window, 41 out of 45 trills showed 2-5 closures as indicated by stripes in the spectrogram. In four tokens, signals appeared with only one stripe followed by frication, or voiced frication throughout the signal.

Native speaker productions of word-initial /r/ evidenced a range of voicing degree and number of visible stripes on the spectrogram; however, the most typical realization was characterized by voicing throughout the segment and 2-3 visible stripes on the spectrogram. See Figure 5.14 for a typical waveform and spectrogram of a native speaker word-initial /r/ production. Figure 5.15 shows an exceptionally long word-initial /r/ realization with 6 visible stripes in the waveform.

Although 90% of realizations showed some degree of voicing within the segment, 10% were identified as voiceless; however, even in a voiceless production the appearance of stripes was still possible. Figure 5.16 demonstrates a word-initial /r/ production that was deemed voiceless yet given a value of "2" for the number of stripes.

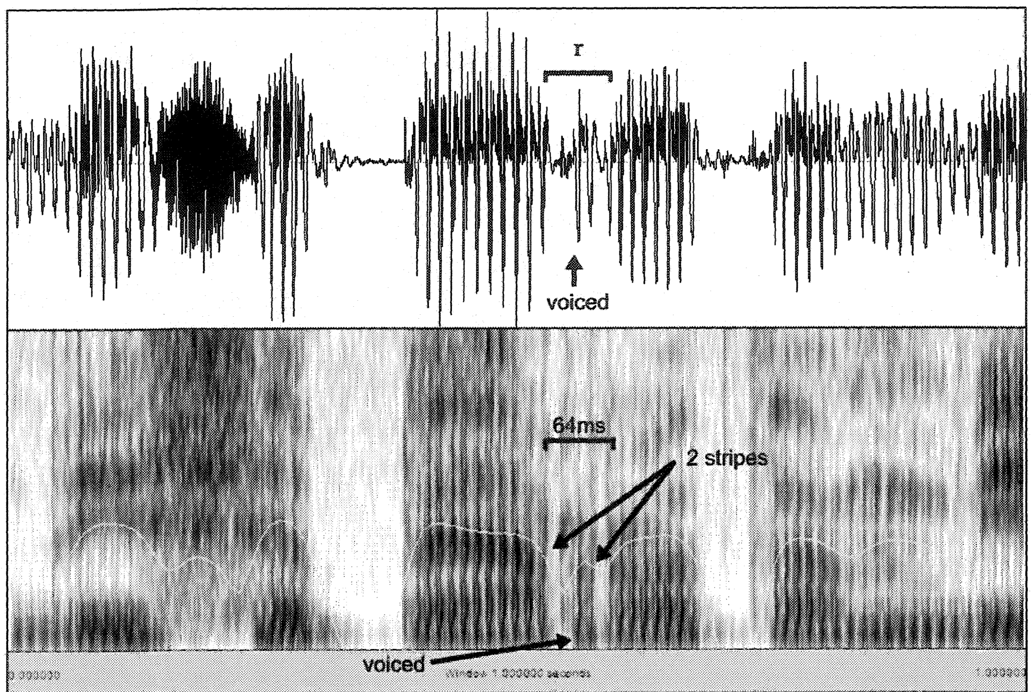

Figure 5.14 Waveform and spectrogram showing a native speaker production of a typical word-medial /r/. Token [dise pero tamby] in *Dice perro también,* participant NS-M1, assigned a voicing value of "2" (fully voiced) and "2" for number of stripes.

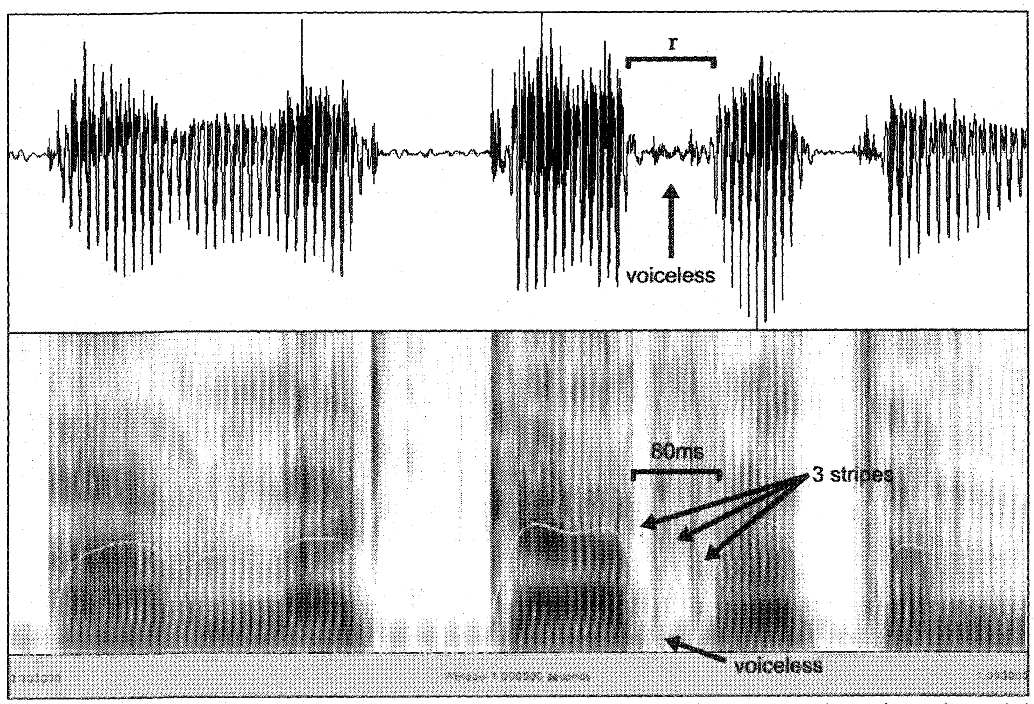

Figure 5.15 Waveform and spectrogram showing a native speaker production of word-medial /r/ realized without voicing. Token [abla karo ta] in *Habla carro también,* participant NS-M4, assigned a voicing value of "0" (voiceless) and "3" for number of stripes.

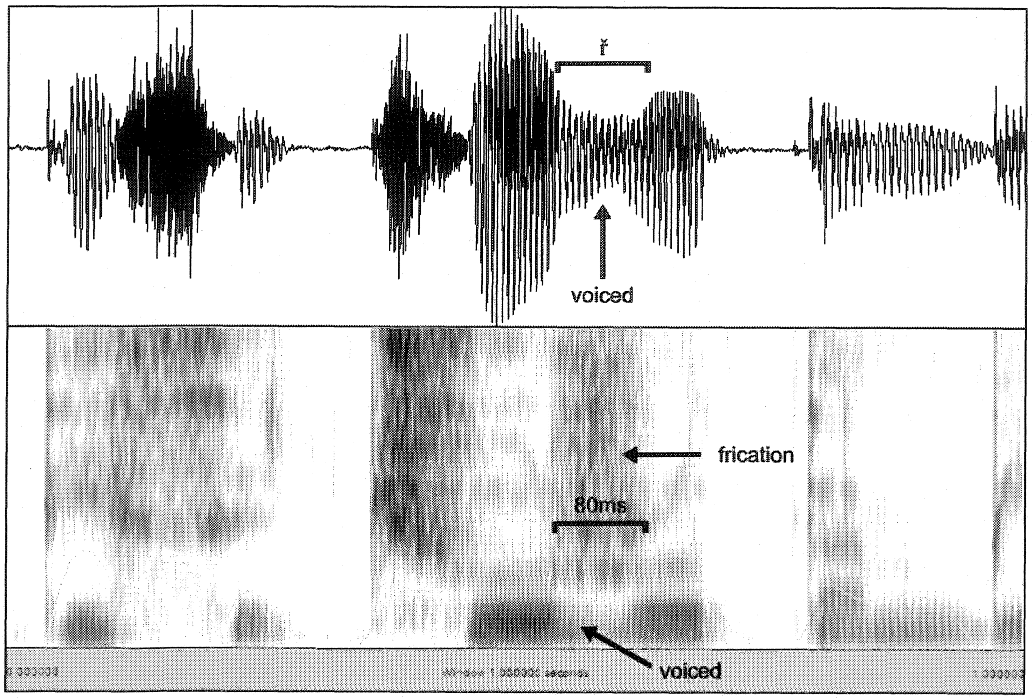

Figure 5.16 Waveform and spectrogram showing a native speaker production of word-medial /r/ realized an assibilated variant. Token [dise ʃtuřo tamb] in *Dice churro también,* participant NS-F3, assigned a voicing value of "2" (fully voiced) and "0" for number of stripes.

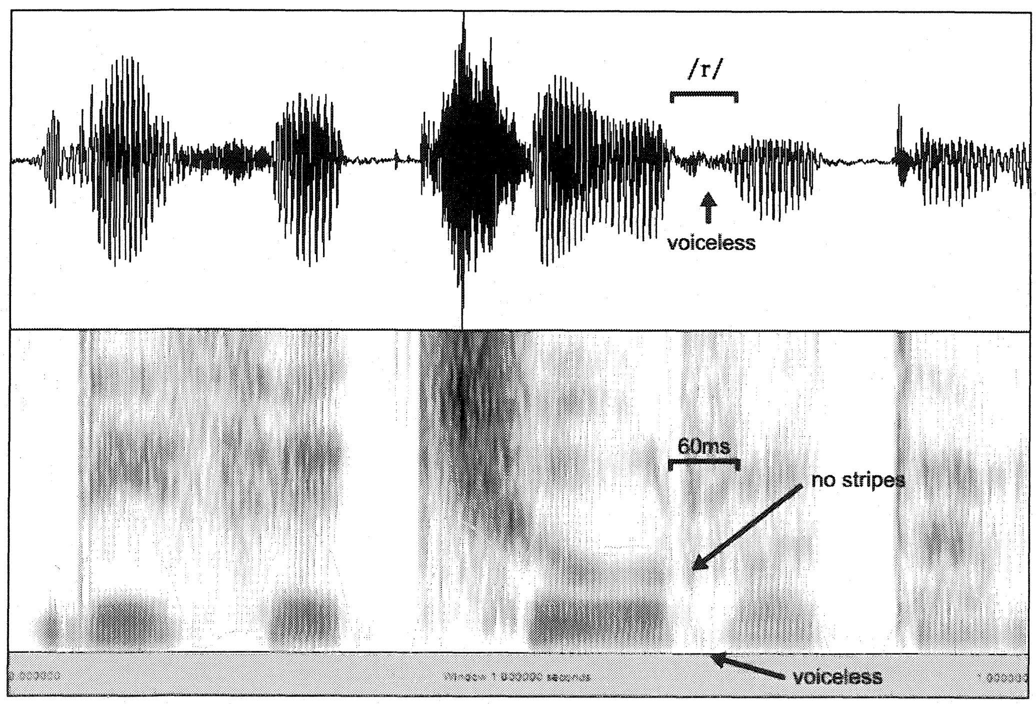

Figure 5.17 Waveform and spectrogram showing a native speaker production of a word-medial /r/. Token [dise ʃtuř̥o ta] in *Dice churro también,* participant NS-F1, assigned a voicing value of "0" (voiceless) and "0" for number of stripes.

In 2 out of 99 productions, the realization was given a value "0" (voiceless) for voicing in addition to "0" for number of stripes.

In terms of visible stripes, values ranged from 0-6 stripes with 3 as the most common value. Across all productions, 10% did not evidence any stripes in the spectrogram while 5% revealed only 1 stripe. For instantiations with fewer than 2 stripes, frication was most commonly observed throughout the segment suggesting assibilated variants of word-initial <r>. See Figure 5.17 for a word-initial /r/ production with frication and partial voicing.

In terms of overall segment duration, segments averaged 88ms in length across both experiments; however, the distribution of values was slightly larger in Experiment A compared to the values in Experiment B (26.3ms stdev vs. 19.9ms stdev, respectively). In Experiment A, values ranged from 49-173ms compared with 61-130ms in Experiment B. Across all native speaker productions, word-initial /r/ averaged 88ms in duration (22.9ms stdev).

Regarding voicing, all tokens were awarded a categorical voicing value of "0" (voiceless), "1" (partially voiced), or "2" (fully voiced). In Experiment A, 1.7 was the average voicing value (0.6 stdev) which was similar to Experiment B's average voicing value of 1.5 (0.7 stdev). Across all word-initial /r/ productions, 10% were identified as voiceless, 22% as partially voiced, and 68% as voiced throughout the segment. In regards to the number of visible stripes, productions in Experiment A yielded 2.7 stripes on average (1.3 stdev) while Experiment B realizations averaged 2.3 stripes (1.1 stdev). Across all native speaker productions, the number of stripes averaged 2.5 with a standard deviation of 1.2. In terms of the most frequently occurring value across all productions, 3 was the mode value. See Table 5.9 for word-initial /r/ measurement values by experiment. These findings are similar to those of studies of native trill productions (as cited by Carballo and Mendoza 2000: Massone 1988, Mota 1990, Quilis 1981, Recasens 1991) who found the average duration of Spanish trills to be between 58.6ms and 103ms with 1-2 apertures and 2-3 occlusions.

Results by gender revealed slight differences between groups. Male word-initial /r/ productions averaged 81ms (19.8 stdev) compared with 94ms (23.9ms stdev) for female productions. Degree of voicing yielded an average score of 1.7 (0.6 stdev) for male productions vs. 1.5 (0.7 stdev) for female productions. For number of stripes, males production values averaged 2.0 (1.0 stdev) in contrast to slightly more stripes for female productions (2.8 average with 1.3 stdev). In

Table 5.10, overall duration, degree of voicing, and number of stripe values are presented by gender.

> ***Acoustic criteria for word-initial /r/:*** *Based on the native speaker productions in this study, the acoustic features used to score L2 realizations were two or more stripes in an overall segment duration of at least 50ms.*

> ***Variability constant for word-initial /r/:*** *A variability constant for each experiment was calculated based on the different vowel environments surrounding word-initial trills in both experiments. In Experiment A, 40 out of 44 tokens were consistent with the established criteria; 48 out of 54 were consistent for Experiment B productions.*[34] *Thus, the variability constant for this category was 91% and 86% for Experiment A and Experiment B, respectively.*

---

[34] In Experiment A, four tokens did not demonstrate any stripes. In Experiment B, six tokens showed zero or one stripe, including two tokens that were voiceless.

Table 5.9 Measurement results of native speaker word-initial /r/ by experiment: Overall duration, degree of voicing and stripes.

		Word-initial /r/		
		Overall duration (ms)	Degree of voicing	Number of stripes
Exp. A native speaker productions (n=45)	average	88	1.7	2.7
	stdev	26.3	0.6	1.3
	min	49	0	0
	max	173	2	6
Exp. B native speakers productions (n=54)	average	88	1.5	2.3
	stdev	19.9	0.7	1.1
	min	61	0	0
	max	130	2	4
All native speaker productions (n=99)	average	88	1.6	2.5
	stdev	22.9	0.7	1.2
	min	49	0	0
	max	173	2	6

Table 5.10 Measurement results of native speaker word-initial /r/ by gender: Overall duration, degree of voicing and stripes.

		Word-initial /r/		
		Overall duration (ms)	Degree of voicing	Number of stripes
Male (n=5) native speaker productions (n=45)	average	81	1.7	2.0
	stdev	19.8	0.6	1.0
	min	55	0	0
	max	130	2	4
Female (n=6) native speaker productions (n=54)	average	94	1.5	2.8
	stdev	23.9	0.7	1.3
	min	49	0	0
	max	173	2	6

5.1.5 Native speaker features of word-final <r>

Auditory features: Auditory impressions of voicing, frication, and duration were greatly varied. Impressions of an oral constriction and complete closure were inconsistent across speakers. In some productions, segments did not give any auditory clues to the presence of these features.

> ***Auditory criteria for word-final <r>:*** *The extreme variation of auditory impressions and acoustic features led to the inability to reliably establish auditory criteria for accuracy in this category. Thus, L2 productions of word-final <r> were not scored for accuracy.*

Acoustic features: Word-final <r> targets demonstrated a wide range of voicing, frication, intensity, and stripe features. Due to the wide range of variation initially observed in these tokens and the subsequent decision not to establish L2 accuracy criteria for this category, quantitative measures for this target were not defined; however, overall segment duration was measured for all native speaker productions. Nevertheless, this section briefly attempts to describe the range of features through example waveforms and spectrograms.

Figure 5.18 shows a waveform and spectrogram of a realization that appears as a tap followed by a voiced element which appears to correspond to the description of word-final <r> provided by Gili (1921). The "tap" portion corresponds to the single stripe at the beginning of the segment while the voiced element is characterized by a periodic waveform and voicing bar in the spectrogram.

Another common realization of word-final <r> suggested an assibilated variant as shown in Figure 5.19 and Figure 5.20. In both instantiations, a dip in the intensity suggests a constriction in the oral cavity while Figure 5.19 evidences a faint stripe at the place of intensity dipping, suggesting a tap or near tap closure. In addition, both figures

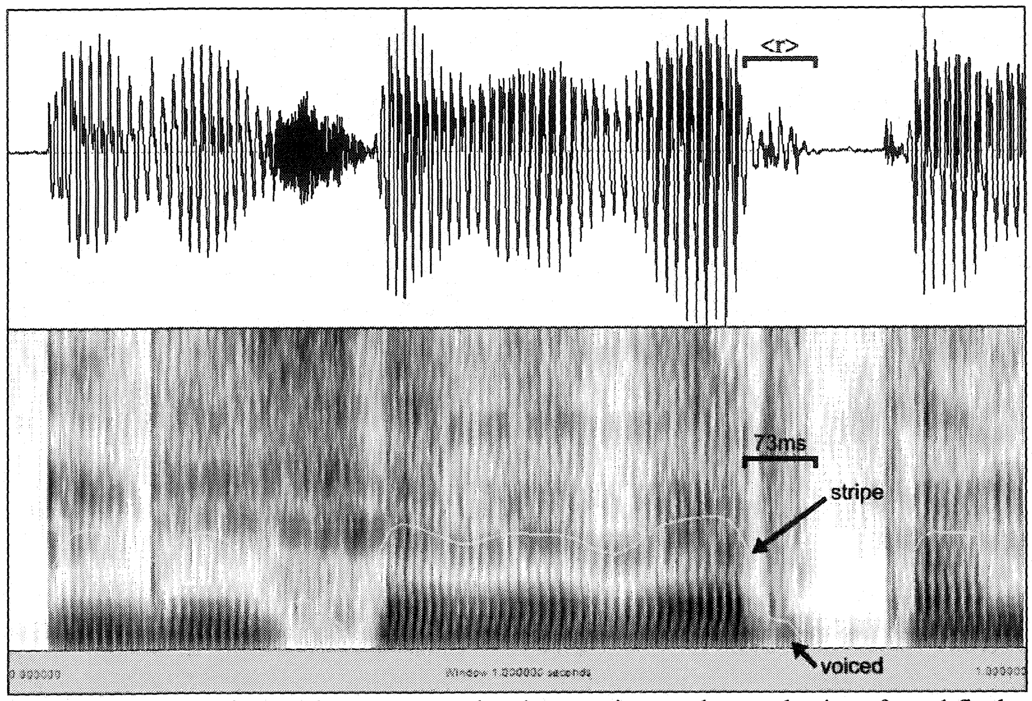

Figure 5.18 Waveform and spectrogram showing a native speaker production of word-final <r> realized as a tap followed by voiced element. Token [dise βeβeɾ ta] in *Dice beber también,* participant NS-M2.

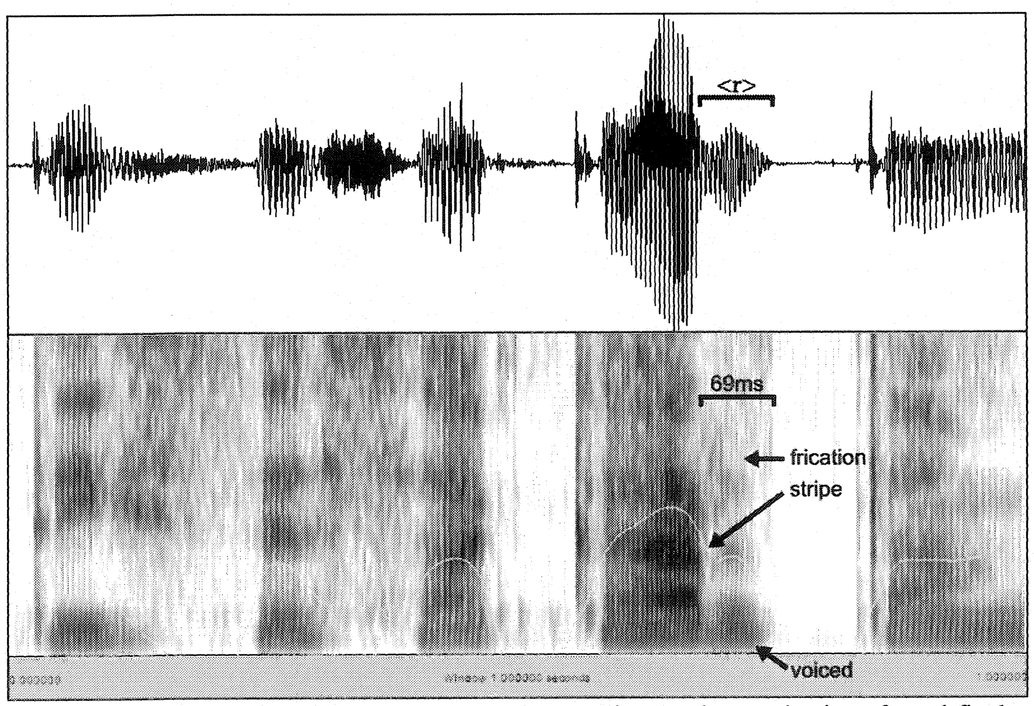

Figure 5.19 Waveform and spectrogram showing a native speaker production of word-final <r> realized as a voiced assibilated variant. Token [dise sakař ta] in *Dice sacar también,* participant NS-F2.

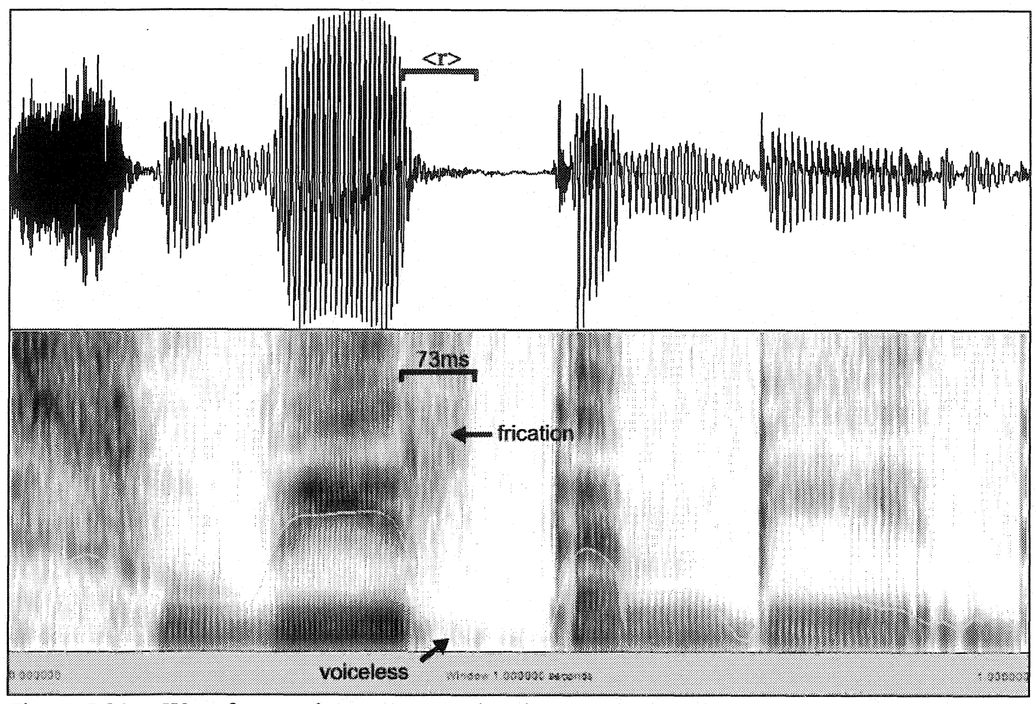

Figure 5.20 Waveform and spectrogram showing a native speaker production of word-final <r> realized as a voiceless fricative. Token [subiř̥ tambyen] in *Dice subir también,* participant NS-F3.

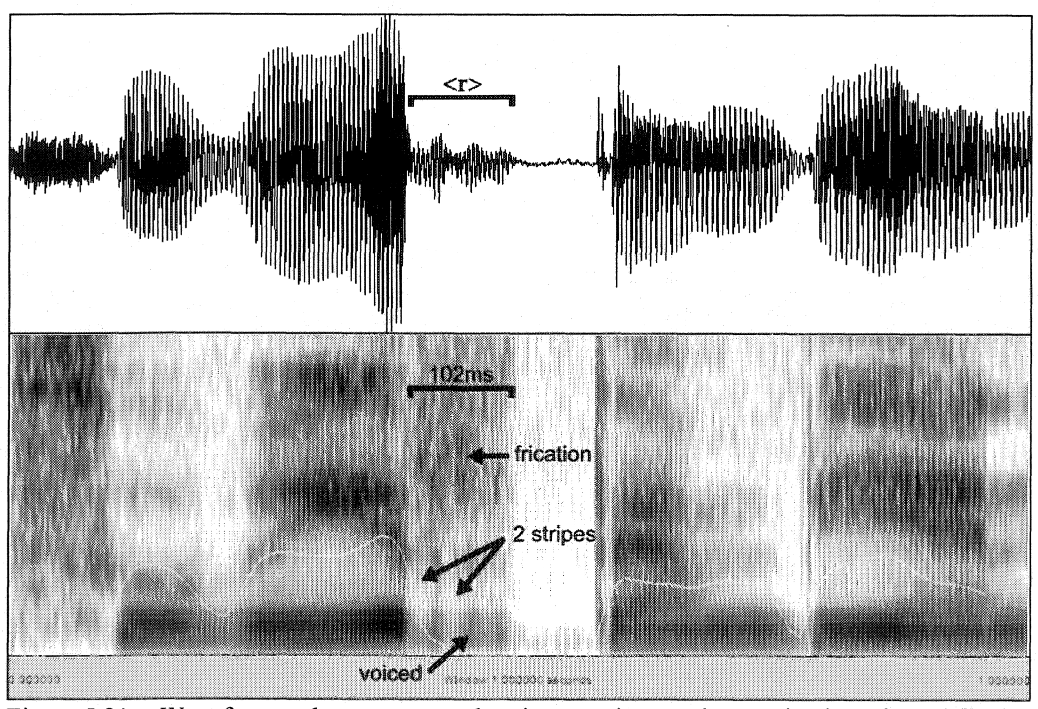

Figure 5.21 Waveform and spectrogram showing a native speaker production of word-final <r> realized as a voiced fricated trill. Token [suβir tambye] in *Ya vi subir también,* participant NS-F5.

are characterized by voicing and frication in higher frequencies. Figure 5.20 shows a voiceless fricated variant of word-final <r>.

Multiple stripes in the realization of word-final <r> were less common; however, Figure 5.21 demonstrates multiple stripes alongside frication suggesting a combination of an assibilated variant and a typical word-initial/word-medial trill realization.

Overall segment duration averaged 67ms (11ms stdev) in Experiment A while overall duration was 81ms (17.9ms stdev) for productions in Experiment B. Across all native speaker productions of word-final <r>, overall average duration was 74ms (16ms stdev) with a range from 40ms-126ms. Table 5.11 shows overall duration by experiment.

In an analysis by gender, overall average duration was similar for males and females. Male productions yielded a 73ms average (11.3ms stdev) while female productions averaged 76ms (19.1ms stdev). Table 5.12 details the overall average duration values for word-final <r> productions by gender.

> ***Acoustic criteria for word-final <r>:*** *The extreme variation in acoustic features and durations did not arguably lead to accuracy criteria for this category. Thus, accuracy scores for L2 productions of word-final <r> were not determined.*

Table 5.11 Overall duration of native speaker productions of word-final <r> by experiment.

		Word-final <r> Overall duration (ms)
Exp. A native speaker productions (n=45)	average	67
	stdev	11.4
	min	40
	max	87
Exp. B native speaker productions (n=54)	average	81
	stdev	17.9
	min	44
	max	126
All native speaker productions (n=99)	average	74
	stdev	16.0
	min	40
	max	126

Table 5.12 Overall duration of native speaker productions of word-final <r> by gender.

		Word-final <r> Overall duration (ms)
Male (n=5) native speaker productions (n=45)	average	73
	stdev	11.3
	min	44
	max	98
Female (n=6) native speaker productions (n=54)	average	76
	stdev	19.1
	min	40
	max	126

5.1.6 Native speaker features of word-medial /ɾ/

Auditory features: Native productions were auditorily more consistent both between and within speakers in comparison to /d/ and /r/ productions. A single obstruction in the oral tract was heard in all productions from both experiments.

> ***Auditory criterion for word-medial /ɾ/:*** *The common auditory feature across all native speaker productions was the impression of a rapid obstruction of the oral cavity and was thus used as the criterion for assigning a score of "1" to L2 productions of this category.*

Acoustic features: Word-medial tap productions were visually inspected from a 1s window. Although Quilis (1981) provides spectrograms of native-speaker tap productions with a complete absence of spectral activity, the native speaker tap productions in this study varied greatly even within speakers. Most native productions showed a spectral stripe across some of, but not the entire frequency range. Furthermore, stripes were accompanied by a drop in intensity; however, frication at the higher frequencies (>3000Hz) was also occasionally present, suggesting an incomplete closure.

Word-medial /ɾ/ productions were typically characterized by a single stripe across one or more of the formants in the spectrogram as shown in Figure 5.22. Figure 5.23 shows a waveform and spectrogram of a production with exceptionally short overall duration while a realization with long overall duration is shown in Figure 5.24. All but one realization out of 99 tokens evidenced a visual stripe in the spectrogram. For the one aberrant token, the realization auditorily gave the impression of an obstruction in the oral tract; however, the visual inspection of the spectrogram gave the appearance of an approximated /d/.

The appearance of a release burst was observed in some productions, manifesting as a dark line across all frequencies in the spectrogram. In such cases, the release period between the release burst and the onset of the following vowel may or may not have been

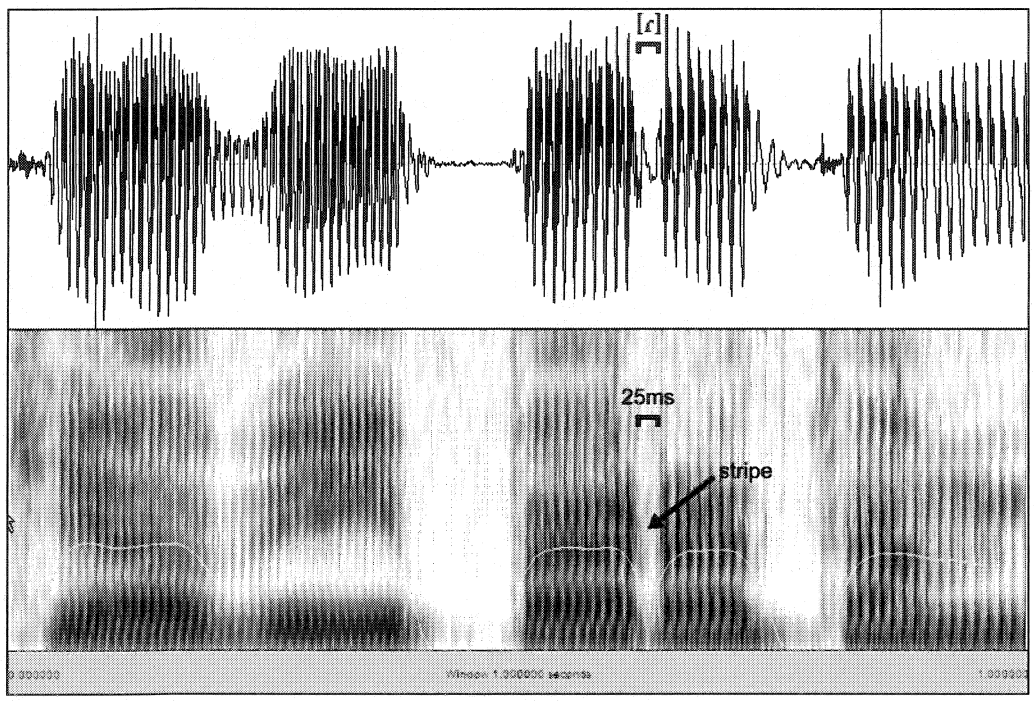

Figure 5.22 Waveform and spectrogram showing a typical native speaker production of word-medial /ɾ/. Token [ya βi paɾa ta] in *Ya vi para también,* participant NS-M3.

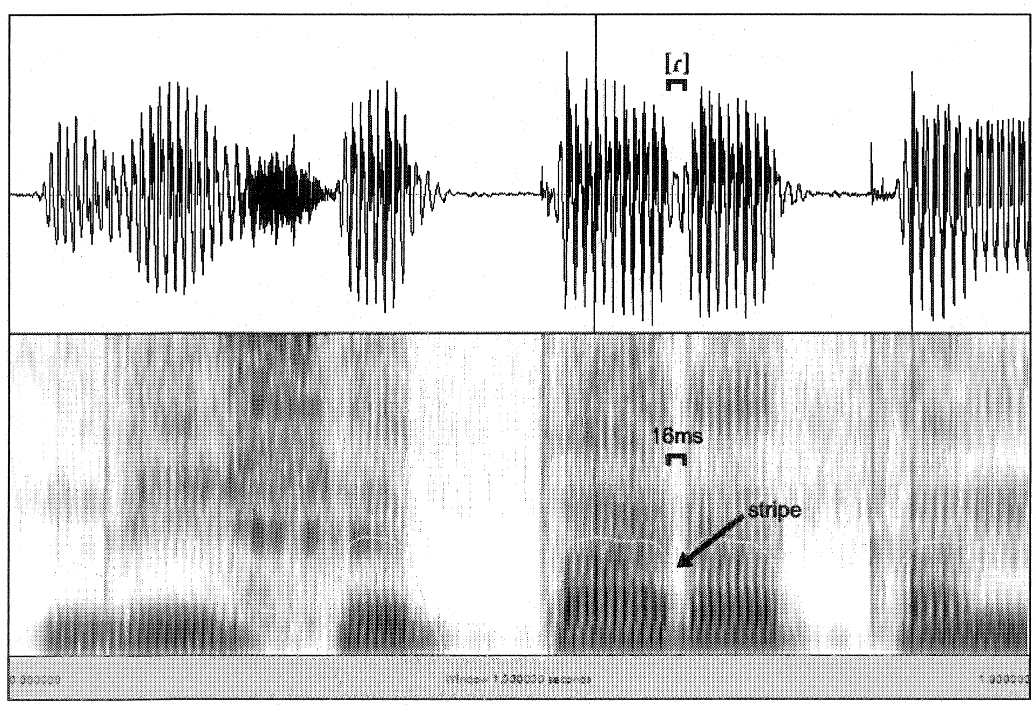

Figure 5.23 Waveform and spectrogram showing a native speaker production of word-medial /ɾ/ with short overall duration. Token [dise paɾa ta] in *Dice para también,* participant NS-M2.

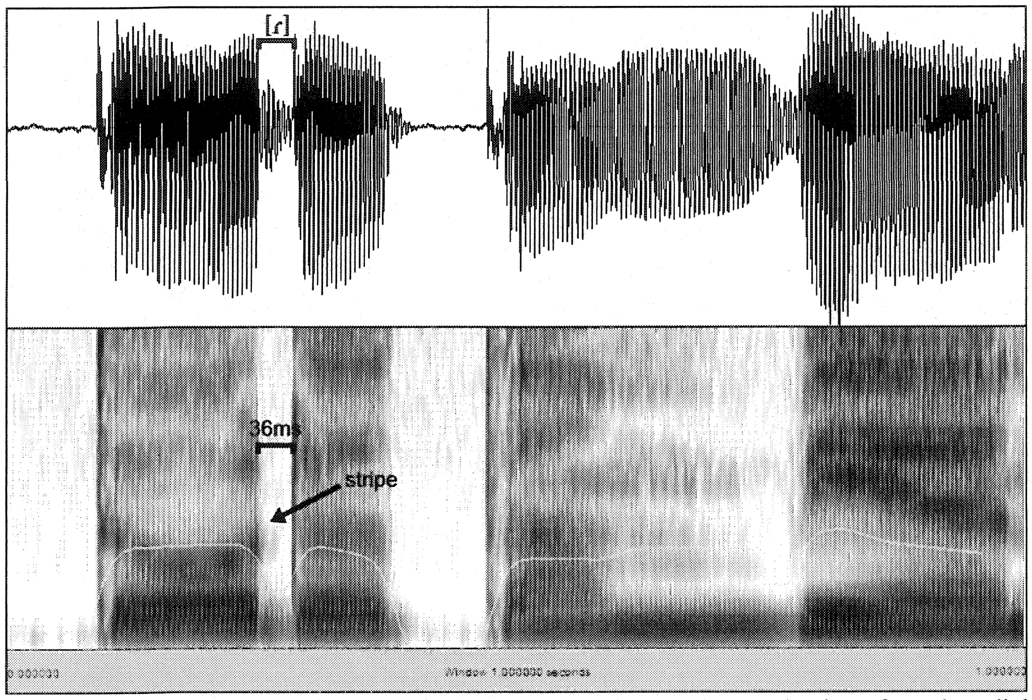

Figure 5.24 Waveform and spectrogram showing a native speaker production of word-medial /ɾ/ with long overall duration. Token [toɾo tambye] in *Ya vi toro también,* participant NS-F6.

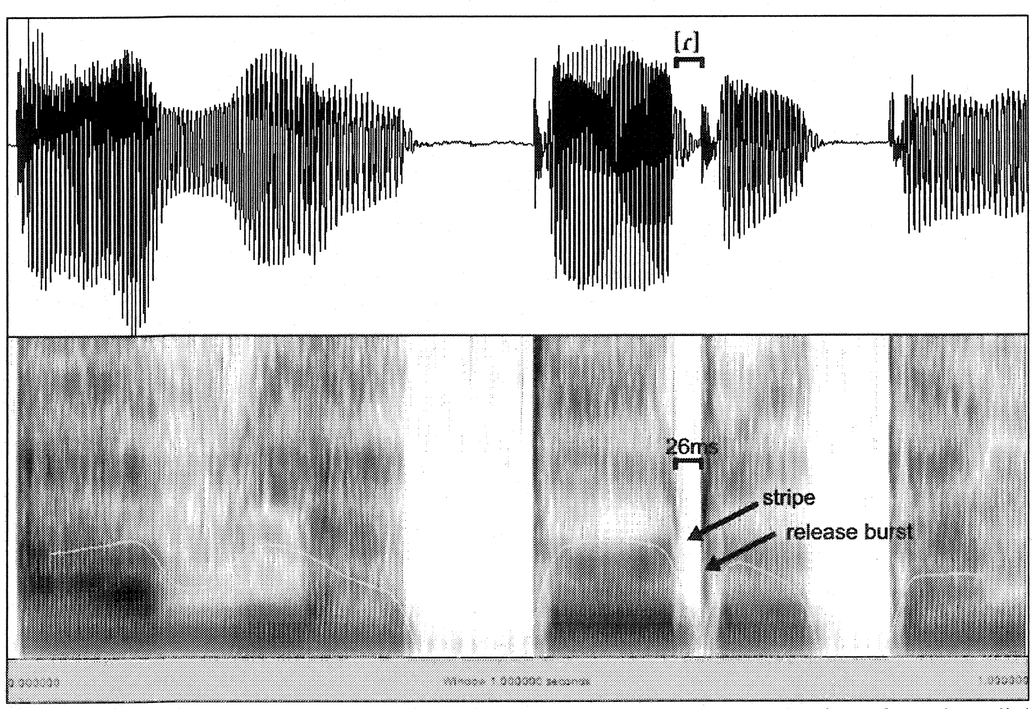

Figure 5.25 Waveform and spectrogram showing a native speaker production of word-medial /ɾ/ with long overall duration, and release burst. Token [ya βi toɾo ta] in *Ya vi toro también,* participant NS-F4.

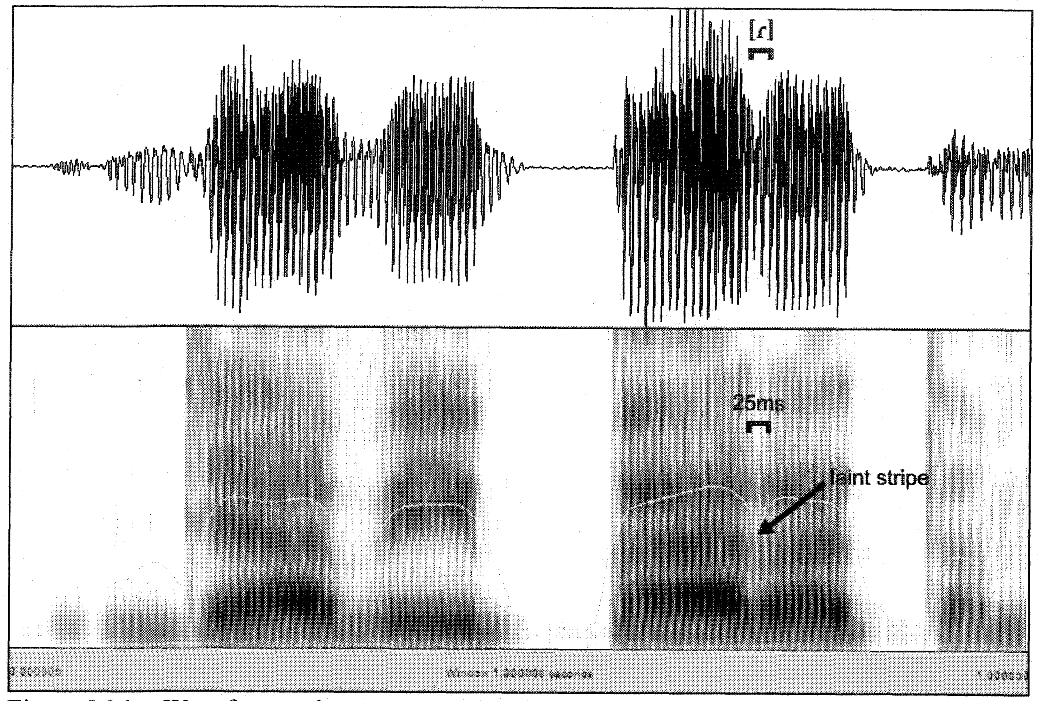

Figure 5.26 Waveform and spectrogram showing a native speaker production of word-medial /ɾ/ realized with slight attenuation. Token [ya βi paɾa ta] in *Ya vi para también,* participant NS-M4.

included in the overall stripe duration measure which was based on the visual absence of glottal pulses across one or more formants. See Figure 5.25 for a relatively long stripe measurement value in which the production included a release burst (although this was not included in the stripe duration measurement).

Word-medial /ɾ/ productions from Experiment A had an average overall stripe duration of 20ms (3.8ms stdev) while overall stripe durations in Experiment B had an average of 26ms (4.7ms stdev). Across all native speaker productions, stripe duration averaged 23ms (5.4ms stdev).

An analysis of stripe durations by gender yielded similar values for female and male productions. Male productions averaged 23ms (5.3ms stdev) with a range from 15-36ms. For female productions, overall stripe durations averaged 23ms (5.5stdev) with a range from 11-36ms.

> ***Acoustic criteria for word-medial /ɾ/:*** *A release burst was not reliably present in the native speaker realizations and hence not identified as a required acoustic feature for accuracy. Thus, acoustic accuracy for word-medial taps was based on visible attenuation in the waveform for 30ms or less and observable stripes across one or more formants.*

> ***Variability constant for word-medial /ɾ/:*** *92 out of 99 native speaker productions evidenced the auditory and acoustic criteria established for this category.*[35] *Thus, 93% represented the variability constant value for this category.*

[35] Six tokens showed durations from 31-36ms. 30ms was selected as the cutoff point to minimize the duration range associated with tokens produced for word-medial /d/.

Table 5.13 Duration of native speaker word-medial /ɾ/ productions by experiment.

		Word-medial /ɾ/ Overall stripe duration (ms)
Exp. A native speaker productions (n=45)	average	20
	stdev	3.8
	min	11
	max	30
	% measurable	100%
Exp. B native speaker productions (n=54)	average	26
	stdev	4.7
	min	18
	max	36
	% measurable	98%
All native speaker productions (n=99)	average	23
	stdev	5.4
	min	11.0
	max	36
	% measurable	99%

Table 5.14 Duration of native speaker word-medial /ɾ/ productions by gender.

		Word-medial /ɾ/ Overall stripe duration (ms)
Male (n=5) native speaker productions (n=45)	average	23
	stdev	5.3
	min	15
	max	36
	% measurable	100%
Female (n=6) native speaker productions (n=54)	average	23
	stdev	5.5
	min	11
	max	36
	% measurable	98%

5.1.7 Native speaker features of onset cluster /ɾ/

Auditory features: In Experiment A, 43 out of 45 productions in onset position gave the impression of a complete obstruction of the oral tract immediately following the word-initial consonant. In Experiment B, 51 out of 54 instantiations gave the impression of an oral tract obstruction.

> ***Auditory criterion for /ɾ/ in onset clusters:*** *Because only a minority of productions did not give the impression of a full obstruction, the common auditory feature in native productions was taken to be the impression of an obstruction between the release of the onset consonant and following nucleic vowel.*

Acoustic features: Taps in a complex onset were visually inspected from a 1s window while durational data were taken from within a 250ms window. The most common and consistent acoustic feature was a stripe across one or more formants with a duration (<30ms) between the release burst of the initial consonant and the onset of the vowel. The majority of native speaker productions contained measurable tap closures immediately preceding the following vowel. The stripe was often accompanied by a visible release burst (indicated by a dark line in the intensity display) at the end of the stripe (i.e., tap closure) and preceding the vowel. A stripe was visible in 83 out of 99 productions. Realizations in which the stripe was not apparent showed the formant structure and voicing of the vowel beginning just after the release of the initial consonant; however, a dip in the intensity trace where the tap closure would normally appear suggested that the tap approximated the place of articulation without closure.[36] A *svarabhakti* vowel, which can be described as a vowel fragment following the onset

[36] It should be pointed out that 4/7 of these aberrations were produced from a single male native speaker (NS-M2) while the remaining three productions corresponded to separate speakers. Interestingly, speaker NS-M2 realized approximants for all three productions of the word *prisa* which may be a result of his particular speech rate or idiolect.

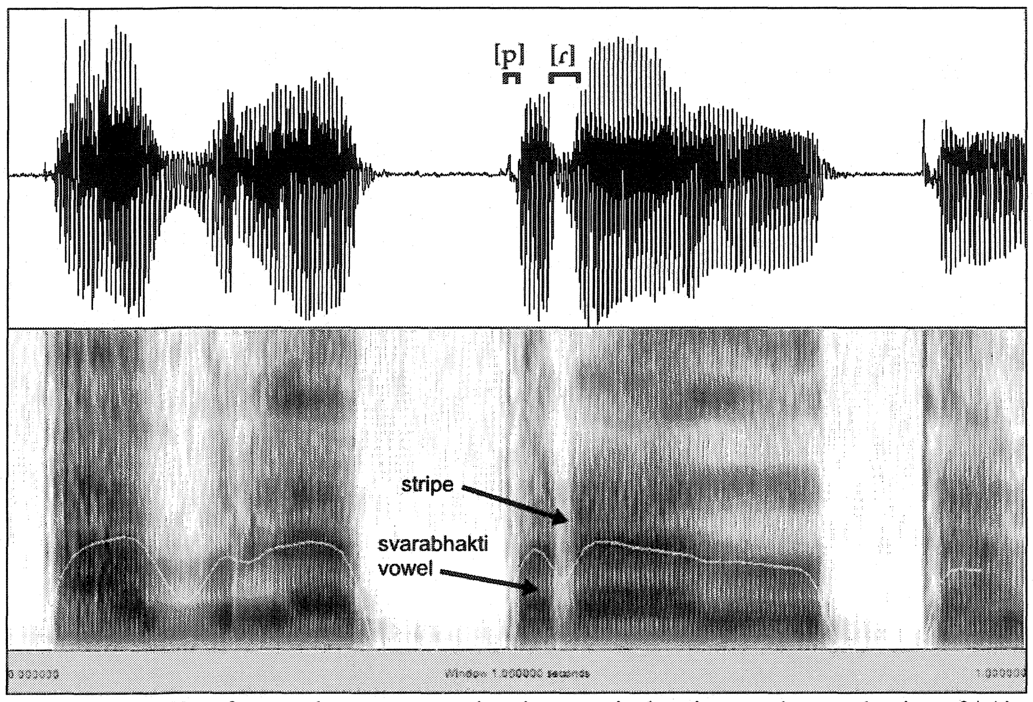

Figure 5.27 Waveform and spectrogram showing a typical native speaker production of /ɾ/ in onset cluster. Token [aβla $p^{ə}$ɾaðo ta] in *Habla prado también,* participant NS-F5.

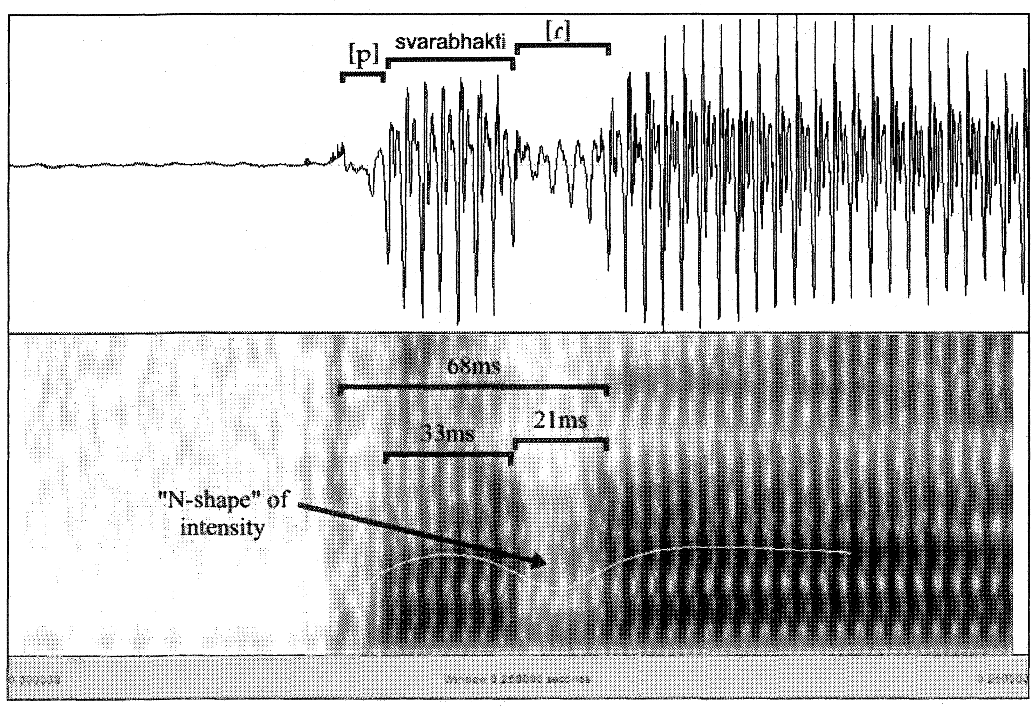

Figure 5.28 Waveform and spectrogram showing a native speaker production of /ɾ/ in onset cluster with emergence of the svarabhakti vowel. Token [$p^{ə}$ɾa] in *Habla prado también,* participant NS-F5.

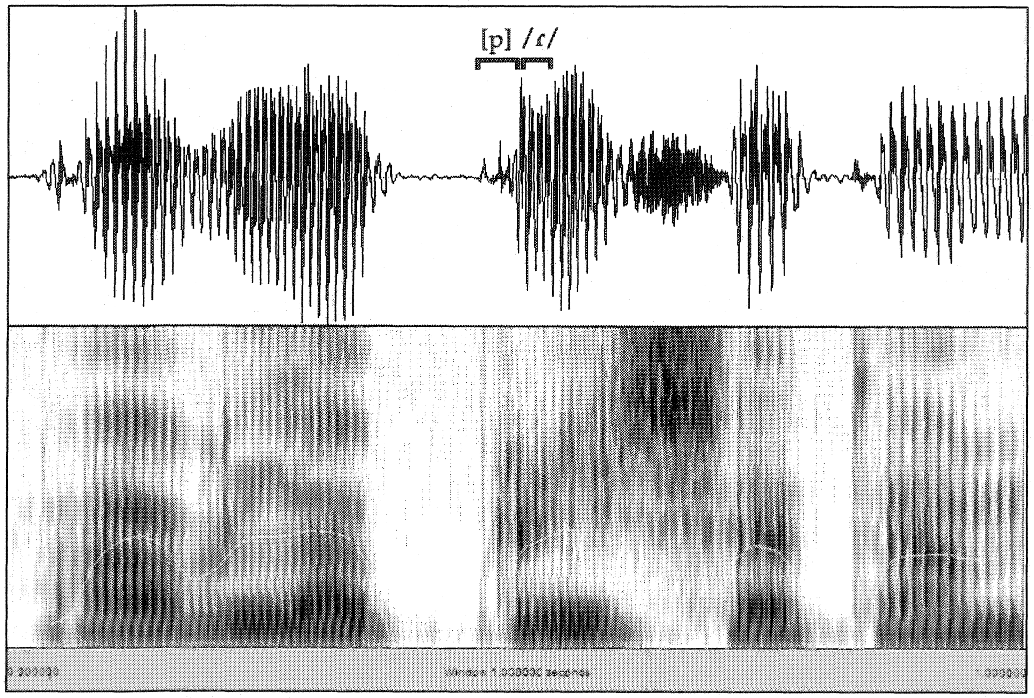

Figure 5.29 Waveform and spectrogram showing a native speaker production of a /ɾ/ in onset cluster realized as an approximant [ɹ]. Token [aβla pɹisa ta] in *Habla prisa también,* participant NS-M3.

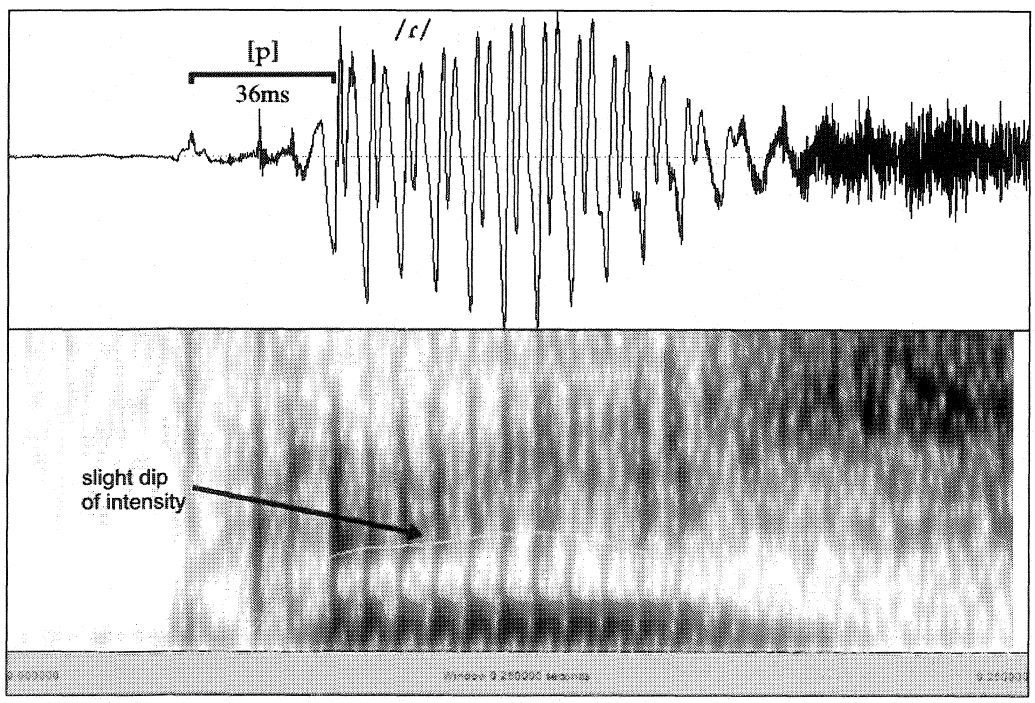

Figure 5.30 Waveform and spectrogram showing a native speaker production of /ɾ/ in onset cluster realized as an approximant without svarabhakti evidence. Token [pɹi] in *Habla prisa también,* participant NS-M3.

consonant and preceding the tap (Bradley and Schmeiser 2002, Gili 1921, Malmberg 1965), was detected in almost all native speaker productions (e.g., *pronto* 'soon' [p^{ə}ɾonto]). Despite its appearance in native productions, the svarabhakti vowel was not selected as a criterion to evaluate tap accuracy as it preceded the tap segment and appeared independently from the portion of the signal corresponding to the tap closure.

In 84% of all productions, visual inspection of the onset cluster revealed the presence of a release burst, followed by a svarabhakti vowel, and then spectral stripe preceding the nucleic vowel. The stripe corresponding to the tap portion of the onset was also visible from a 1s window. Figure 5.27 shows a waveform and spectrogram of a typical /ɾ/ in onset from a 1s window while Figure 5.28 presents the same token from a 250ms window.

In 15% of all productions, the presence of a spectral stripe was not apparent. In such instantiations, there was attenuation in the waveform corresponding to the location of the tap constriction. In addition, a dip in the intensity trace was also an indication of the tap constriction location. As revealed from the measures, these "near-closure" taps only occurred in onset /ɾ/ preceding a high or mid vowel (e.g., *pri̲sa* or *pre̲so*); a stripe absence was never noted before a low vowel (e.g., *pra̲do*). Figure 5.28 demonstrates a waveform and spectrogram of a "near-closure" tap from a 1s window. The same token is presented from a 250ms window in Figure 5.30.

In one production, a stripe was present without the presence of the preceding svarabhakti vowel. This aberrant realization may be of particular interest as similar productions were demonstrated by L2 speakers.

Productions from Experiment A yielded an average overall onset duration of 62ms (12.6ms stdev) and 55ms (9.3ms stdev) in Experiment B productions. Across all native speaker productions, onset durations averaged 58ms (11.5 stdev) with a range from 32-106ms. Values of tap to vowel durations averaged 17ms (5.4ms stdev) in Experiment A and 20ms (6.4ms stdev) in Experiment B. Across both experiments, tap to vowel duration averaged 19ms (6.1ms stdev). In an analysis by gender, similar values were found for both male and female productions. Male production onset durations ranged from 40-87ms and averaged 60ms (11.4ms stdev). For female productions, onset durations averaged 58ms (11.5ms stdev) with a range from 32-106ms.[37] Regarding tap to vowel durations, males averaged 20ms (6.5ms stdev) while female values averaged 18ms (5.7ms stdev). Table 5.15 summarizes onset duration and tap to vowel values by experiment while Table 5.16 presents the same results by gender.

Values for svarabhakti durations yielded an average of 20ms (8ms stdev) for productions in Experiment A and 23ms (7.4ms stdev) for productions in Experiment B. An average of 21ms (7.8ms) was the overall svarabhakti duration mean across all native speaker productions. Productions by males averaged 22ms (6.9ms stdev) in length while productions by females resulted in an average duration of 20ms (8.2ms stdev). Table 5.17 presents the results of svarabhakti durations by experiment and the results by gender are found in Table 5.18.

Acoustic criterion for /ɾ/ in onset cluster: *In sum, the appearance of a stripe (<30ms) preceding the nuclear vowel was required for an accurate score of "1" for taps in onset clusters.*

Variability constant for /ɾ/ in onset cluster*: 81 out of 99 native speaker productions were in complete agreement with the established auditory and acoustic criteria. Thus, 82% was the calculated variability constant.*

Acoustic criteria for svarabhakti vowel emergence: *Because of the inability to auditorily detect the presence of the svarabhakti vowel, only acoustic features were used to determine svarabhakti vowel emergence accuracy. A positive duration of a sinusoidal segment following the release of the onset consonant and*

---

[37] The 106ms onset duration was an apparent aberration in that the second longest female onset duration was 76ms.

preceding a spectral stripe corresponding to the tap closure were required for a svarabhakti vowel emergence score of "1".

Variability constant for svarabhakti vowel emergence: *Based only on the acoustic features of native speaker productions, 84 out of 99 tokens were consistent with the acoustic features for this category. Hence, 85% was used as the variability constant for svarabhakti vowel emergence.*

Table 5.15 Duration of onsets and tap-to-vowel length of native speaker productions of /ɾ/ in onset by experiment.

		Onset cluster /ɾ/	
		Onset duration (ms)	Tap to vowel duration (ms)
Exp. A native speaker attempts (n=45)	average (n=39)	62	17
	stdev	12.6	5.4
	min	32	8
	max	106	41
	% measurable	87%	87%
Exp. B native speaker attempts (n=54)	average (n=44)	55	20
	stdev	9.3	6.4
	min	40	11
	max	76	38
	% measurable	81%	81%
All native speaker attempts (n=99)	average	58	19
	stdev	11.5	6.1
	min	32	8
	max	106	41
	% measurable (n=83)	84%	84%

Table 5.16 Duration of onsets and tap duration of native speaker productions of onset taps by gender.

		Onset cluster /ɾ/	
		Onset duration (ms)	Tap to vowel duration (ms)
Male (n=5) native speaker attempts (n=45)	average (n=32)	60	20
	stdev	11.4	6.5
	min	40	12
	max	87	38
	% measurable	71%	71%
Female (n=6) native speaker attempts (n=54)	average (n=51)	58	18
	stdev	11.5	5.7
	min	32	8
	max	106	41
	% measurable	94%	94%

Table 5.17 Svarabhakti vowel duration by experiment: native speaker productions of onset taps.

		Svarabhakti
		Overall duration (ms)
Exp. A Native speaker attempts (n=45)	average (n=39)	20
	stdev	8.0
	min	10
	max	42
	% measurable	87%
Exp. B Native speaker attempts (n=54)	average (n=44)	23
	stdev	7.4
	min	9
	max	42
	% measurable	81%
All native speaker attempts (n=99)	average (n=83)	21
	stdev	7.8
	min	9
	max	42
	% measurable	84%

Table 5.18 Svarabhakti vowel duration by gender: Native speaker productions of onset taps.

		Svarabhakti
		Overall duration (ms)
Male (n=5) native speaker attempts (n=45)	average (n=32)	22
	stdev	6.9
	min	13
	max	42
	% measurable	71%
Female (n=6) native speaker attempts (n=54)	average (n=51)	20
	stdev	8.2
	min	9
	max	42
	% measurable	94%

In sum, native speakers evidenced a range of variation in their productions of each segment category. Nevertheless, productions in each category showed enough consistency across auditory and acoustic features in order to establish criteria for scoring L2 productions. A variability constant was established for each category to represent the degree of consistency for each category. Table 5.19 presents the variability constant for each category. For example, word-medial /r/ is associated with a 89% variability constant. Word-initial /r/ category has a 91% constant for productions in Experiment A and 86% constant for productions in Experiment B.

Table 5.19 Target category variability constants based on native speaker productions. The constant in parentheses reflects the variability constant for L2 productions in Experiment B only.

Target Category	Variability Constant
Word-medial /d/	99%
Word-medial /t/	99%
Word-medial /r/	89%
Word-initial /r/	91% (86%)
Word-medial /ɾ/	93%
Onset cluster /ɾ/	82%

5.2 Scoring L2 productions

The purpose of this section is to explain the application of the acoustic accuracy criteria established in Section 5.1 to the L2 productions of each category. In order for a participant's production to have received a score of "1" (accurate), the production had to evidence both the auditory and acoustic criteria extracted from the features found in the native productions of the appropriate phonemic category; however, svarabhakti vowel accuracy (i.e., emergence) was determined solely through an acoustic analysis due to the perceptual difficulty of identifying the element. Thus, with the exception of the svarabhakti vowel, the evaluation of accuracy for each production followed a two-step procedure. First, the production was judged for auditory accuracy based on the 44.1kHz sampled signal. Second, I visually inspected the waveform and spectrogram for the required acoustic feature(s).[38]

5.2.1 Scoring L2 productions of word-medial /d/ (target [ð])

Tokens were scored as "1" (accurate) if auditorily the production was voiced throughout the phone (without the impression of complete oral tract obstruction) and if acoustically the segment was voiced throughout with attenuation of energy in all frequencies lasting greater than 30ms, reflecting frication or an approximated closure. Productions were rated as "0" (inaccurate) in the case of auditory disagreement with the acoustic criteria, an absence of voicing, or a stripe less than 30ms in length (i.e., an apparent tap closure). Figure 5.31 and Figure 5.32 show waveforms and spectrograms produced by an L2 Spanish speaker of spirantized and approximated /d/ L2 productions, respectively.

5.2.2 Scoring L2 productions of word-medial /t/ (target [t])

Intervocalic voiceless stops were given a score of "1" if the impression of a contact was auditorily present and if the acoustic signal revealed a stop closure, release burst, and positive voice onset time. Figure 5.33 shows the acoustic features in an L2 production scored as accurate. A score of "1" was awarded despite the length of the release burst as degree of aspiration was not a required criterion. The scoring of productions as "0" (inaccurate) was most often due to the presence of auditory and acoustic cues consistent with production of a tap in place of a voiceless stop.

---

[38] For example, if the production to be analyzed was the word-medial trill target in *perro*, auditorily I would have listened to the entire carrier phrase *Dice perro también*. Then, for the visual inspection, I would have placed the corresponding signal from *perro* in the middle of a one second window.

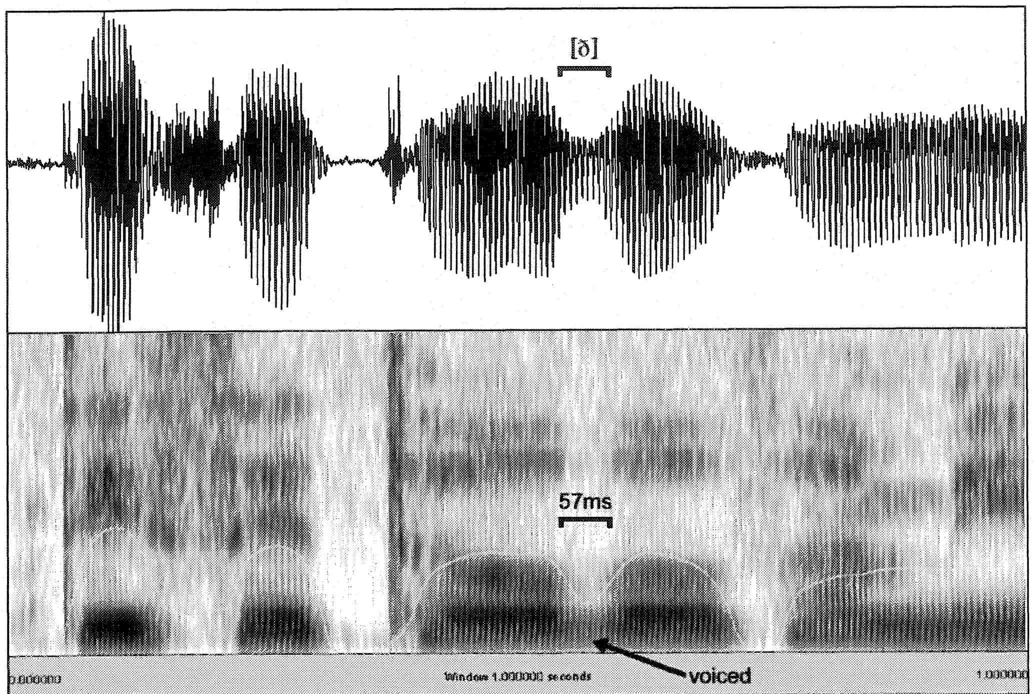

Figure 5.31 Waveform and spectrogram showing an L2 production of word-medial /d/ designated as "spirantized." Token [dise toðo tambye] in *Dice todo también*, Session 1, participant A-Adv-F1, scored as "1" (accurate).

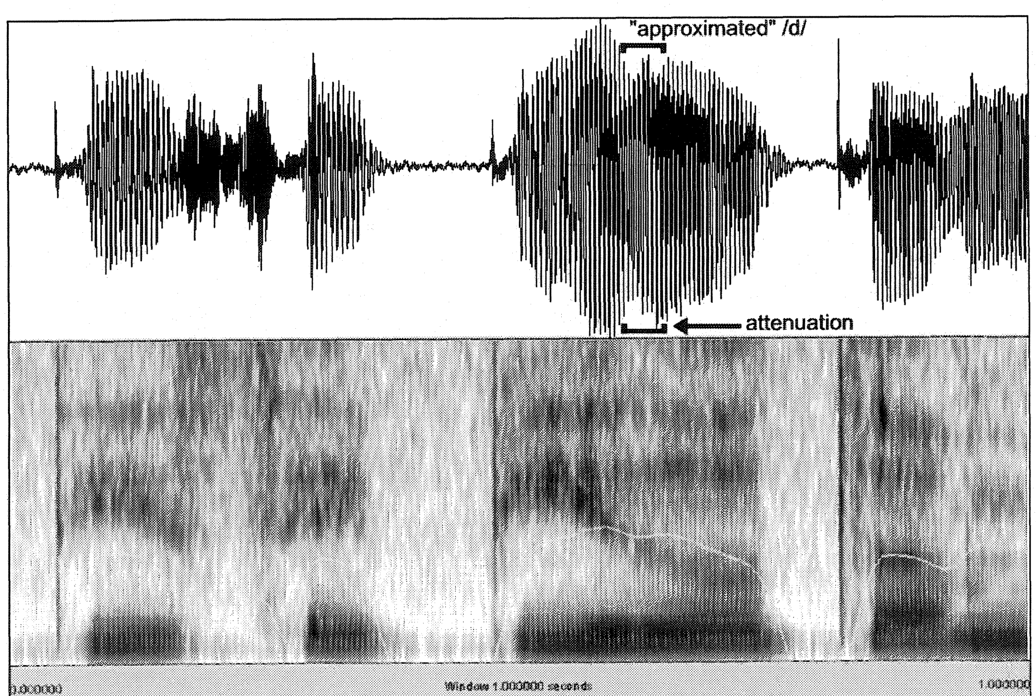

Figure 5.32 Waveform and spectrogram showing an L2 production of word-medial /d/ designated as "approximated." Token [dise piðo tam] in *Dice pido también*, Session 3, participant A-Beg-F1, scored as "1" (accurate).

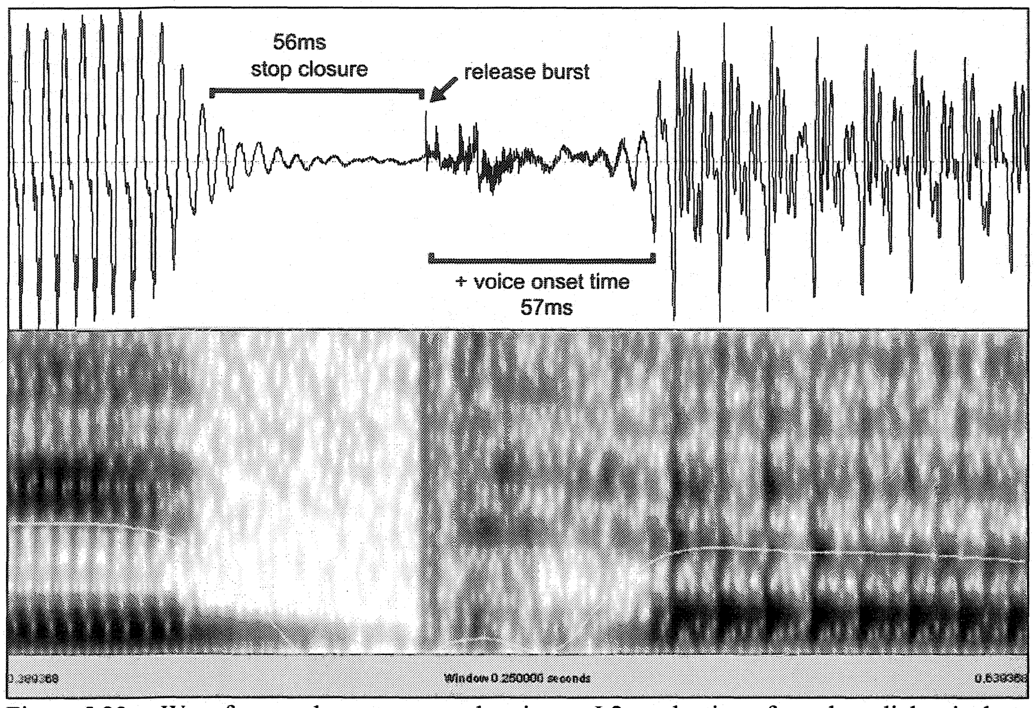

Figure 5.33 Waveform and spectrogram showing an L2 production of word-medial voiceless stop [t]. Token [ito] in *Dice pito también*, Session 1, participant A-Beg-F2, scored as "1" (accurate).

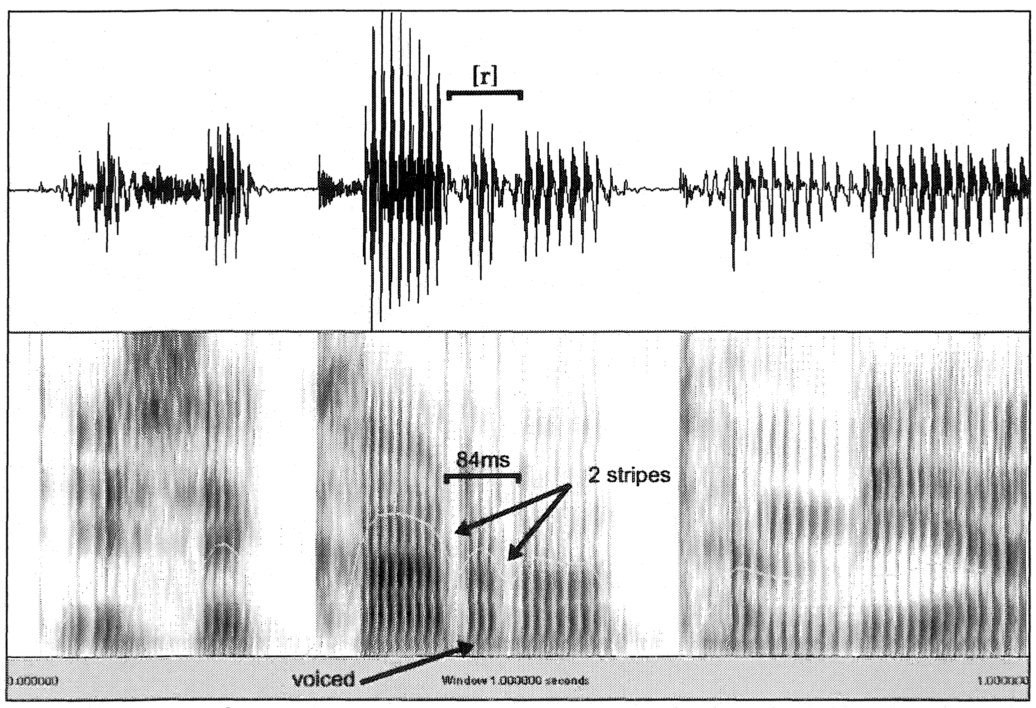

Figure 5.34 Waveform and spectrogram showing an L2 production of word-medial trill [r]. Token [dise karo tambye] *Dice carro también*, Session 2, participant A-Int-M2, scored as "1" (accurate).

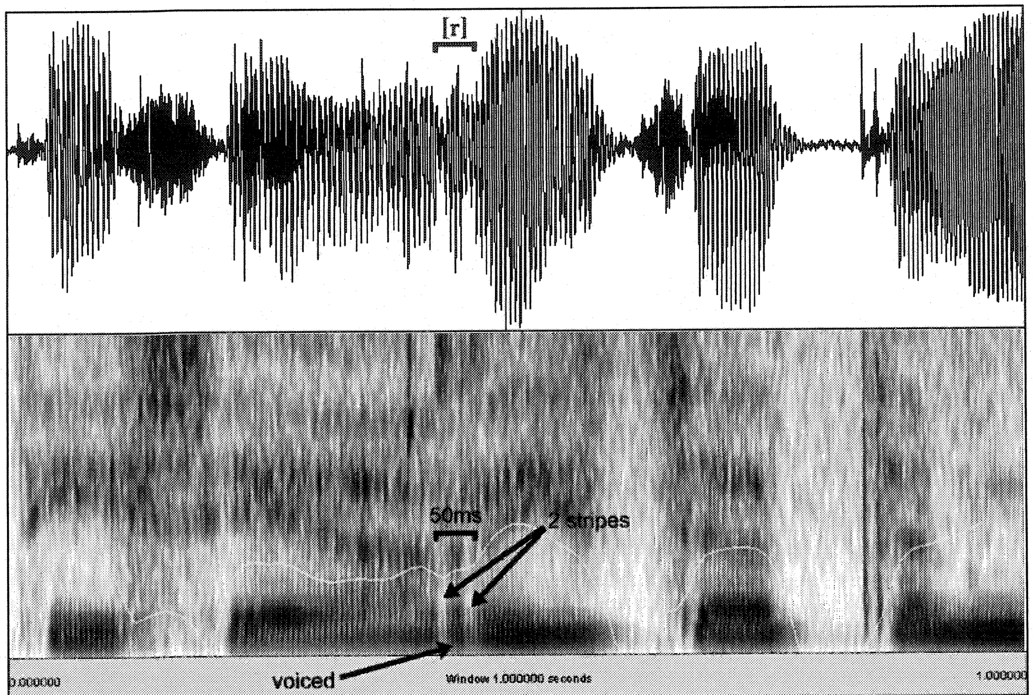

Figure 5.35 Waveform and spectrogram showing an L2 production of word-initial [r]. Token [dise risa ta] *Dice risa también*, Session 3, participant A-Beg-F2, scored as "1" (accurate).

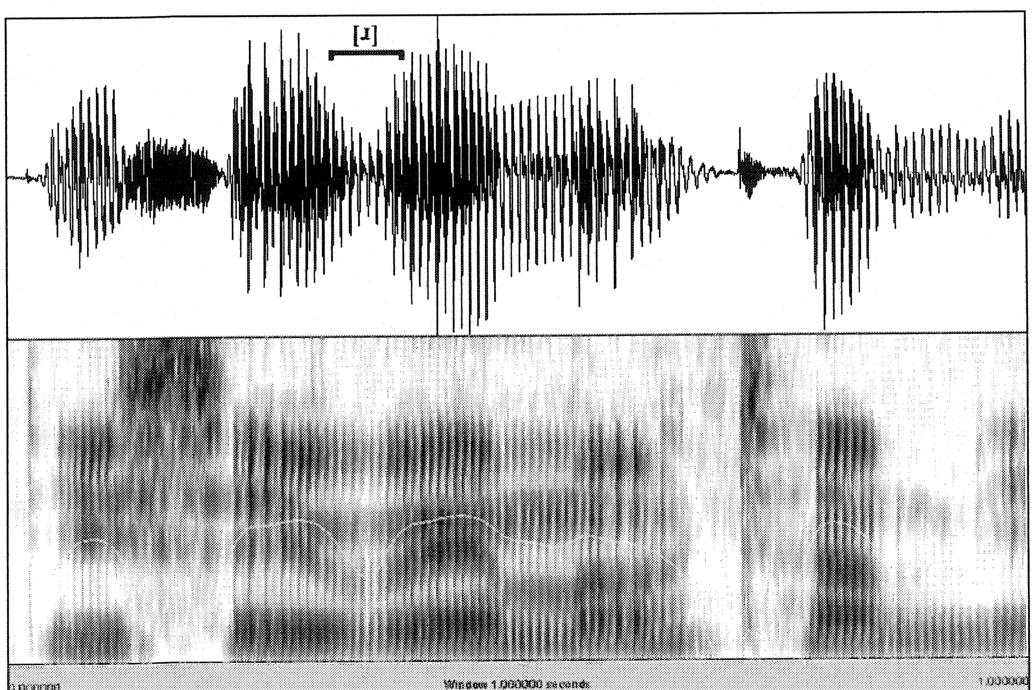

Figure 5.36 Waveform and spectrogram showing an L2 production of word-medial /r/, realized as an approximant [ɹ]. Token [dise ɹemo tam] *Dice remo también*, Session 1, participant A-Adv-M2, scored as "0" (inaccurate).

Section 6.2.2 redefines /t/ accuracy through the incorporation of an additional durational criterion for an accuracy score of "1." A threshold criterion of 41ms was established to require that an L2 production fall within the range of any native speaker VOT duration in order to be scored as accurate.

5.2.3 Scoring L2 productions of word-medial and word-initial /r/ (target [r])

Word-initial and word-medial trills were scored using the same criteria due to the similarities of productions in both environments by native speakers. For an accurate score of "1", the auditory signal had to give the auditory impression of containing a series of two or more obstructions of the oral tract and the spectrogram was required to show two or more visible stripes for an overall segment duration of at least 50ms. Figure 5.34 and Figure 5.35 show accurate L2 productions of word-medial and word-initial trills, respectively. A token was given a score of "0" (inaccurate) if the researcher did not hear a series of two or more contacts (voiced or voiceless), or if the spectrogram failed to show two or more stripes totaling at least 50ms in duration. Many inaccurate productions were judged auditorily to be approximants [ɹ] in place of the trill, as shown in Figure 5.36. Spectrograms of perceived approximants in L2 productions were also auditorily and spectrally similar to approximant productions in participants' L1 productions. As discussed by Boyce & Espy-Wilson (1997), the American English /ɹ/ between sonorants often demonstrates an inverted parabolic trajectory of F3 below 2000Hz. In the case of a "tip-up" tongue shape, the spectrogram may reveal a dip in F2 and not F3 (Hagiwara 1995). Hashi et al. (2003) also analyzed additional literature that has investigated English approximant production and found that a low F3 has been the only consistent correlate identified across studies.

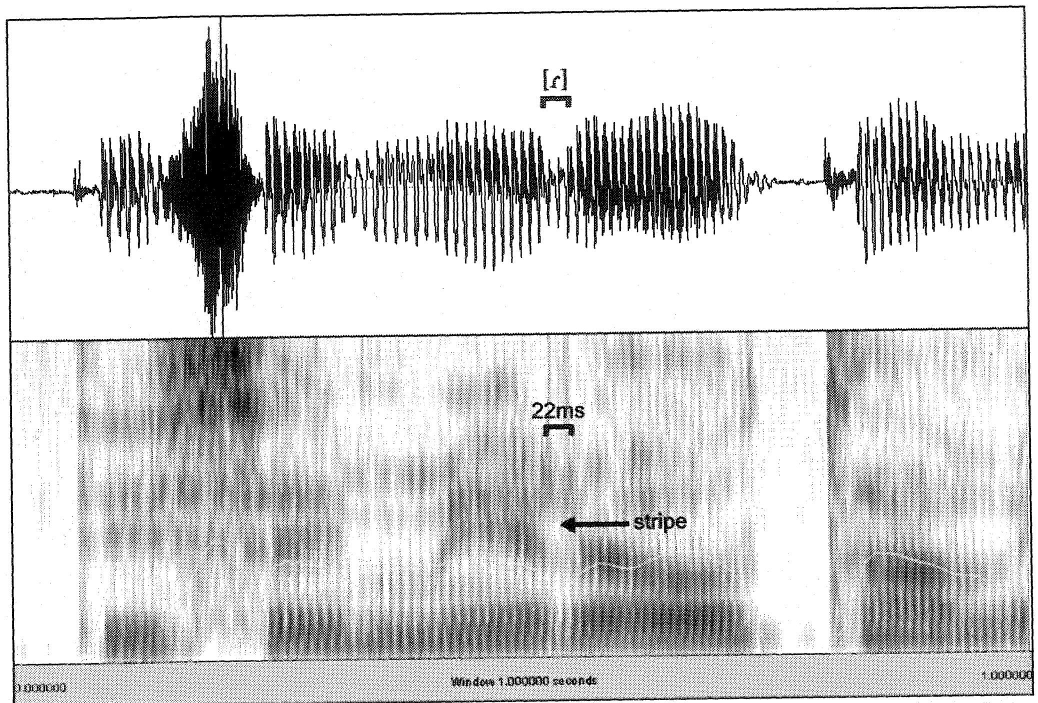

Figure 5.37 Waveform and spectrogram showing an L2 production of word-medial [ɾ]. Token [dise miɾo ta] in *Dice miro también*, Session 2, participant A-Beg-M1, scored as "1" (accurate).

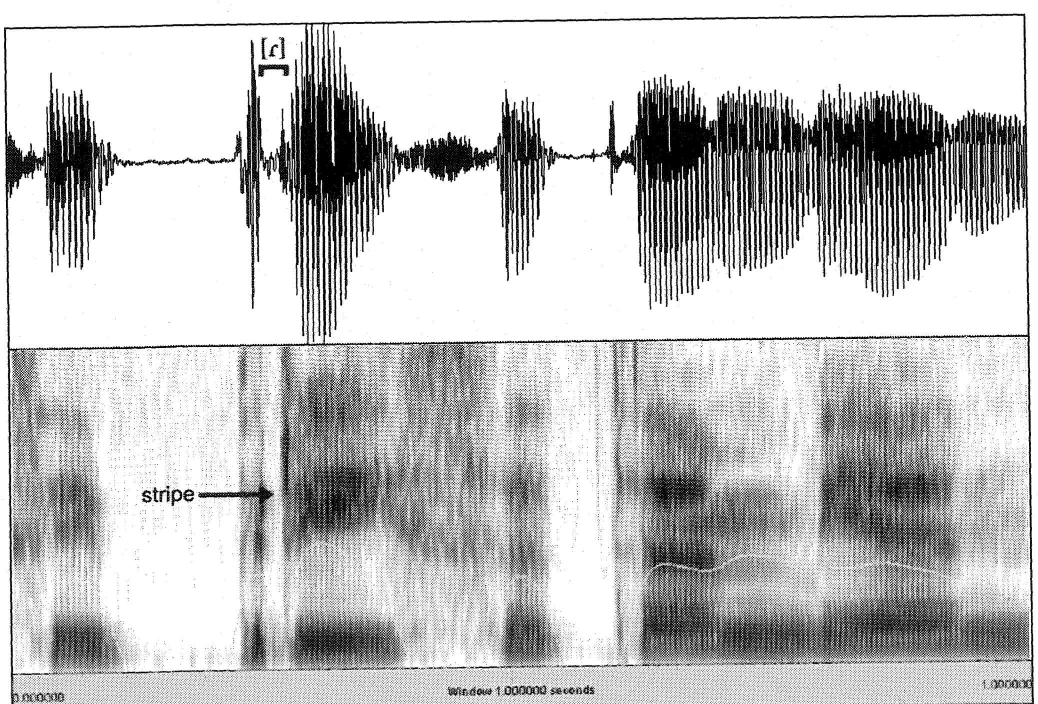

Figure 5.38 Waveform and spectrogram showing an L2 production of /ɾ/ in onset cluster from a 1s window. Token [se pᵊɾisa tambyen] in *Dice prisa también*, Session 2, participant A-Adv-F2, scored as "1" (accurate).

5.2.4 Scoring L2 productions of word-medial /ɾ/ (target [ɾ])

Following the auditory impression of a contact between the tongue and the alveolar ridge, tokens were scored as "1" (accurate) if attenuation in the waveform was visible for less than 30ms with a spectral stripe across one or more formants. The stripe also must have been accompanied by a drop in the intensity line. Figure 5.37 demonstrates a medial tap token scored as "1". In a great number of productions assigned a score of "0" (inaccurate), auditory and acoustic features suggested that students realized approximants [ɹ] in place of taps.

5.2.5 Scoring L2 productions of onset cluster /ɾ/ (target [ɾ])

An accuracy score of "1" required an auditory impression of a single obstruction of the oral tract following the release of the word-initial consonant. In addition, an accurate production was required to evidence a clear stripe for 30ms or less (as measured from a 250ms window) following the consonant release burst and preceding the onset of the vowel. Figure 5.38 shows the waveform and spectrogram of a production scored "1" (accurate) from a 1s window and Figure 5.39 shows the same token from a 250ms window for comparison. In the case of productions scored "0" (inaccurate), participants appeared to be realizing an approximant [ɹ] in place of a tap as shown in Figure 5.40.

5.2.6 Scoring L2 emergence of the svarabhakti vowel

For taps in onset cluster positions, the svarabhakti vowel was scored solely on acoustic criteria, in contrast to the scoring of productions of all other targets. In order to determine the emergence of a svarabhakti vowel in an L2 production of tap in onset cluster, the L2 production was visually inspected from a 250ms window. A token was

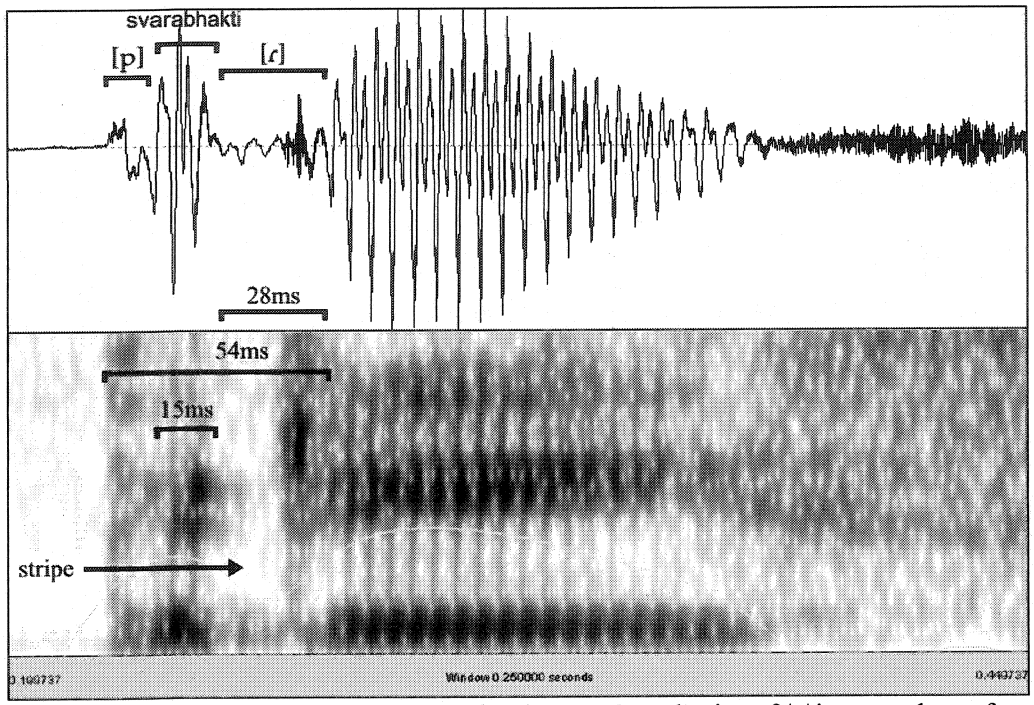

Figure 5.39 Waveform and spectrogram showing an L2 production of /ɾ/ in onset cluster from a .25s window, realized with a svarabhakti vowel. Token [se p^{ə}ɾis tambyen] in *Dice prisa también*, Session 2, participant A-Adv-F2, scored as "1" (accurate) for onset tap accuracy and "1" for svarabhakti emergence.

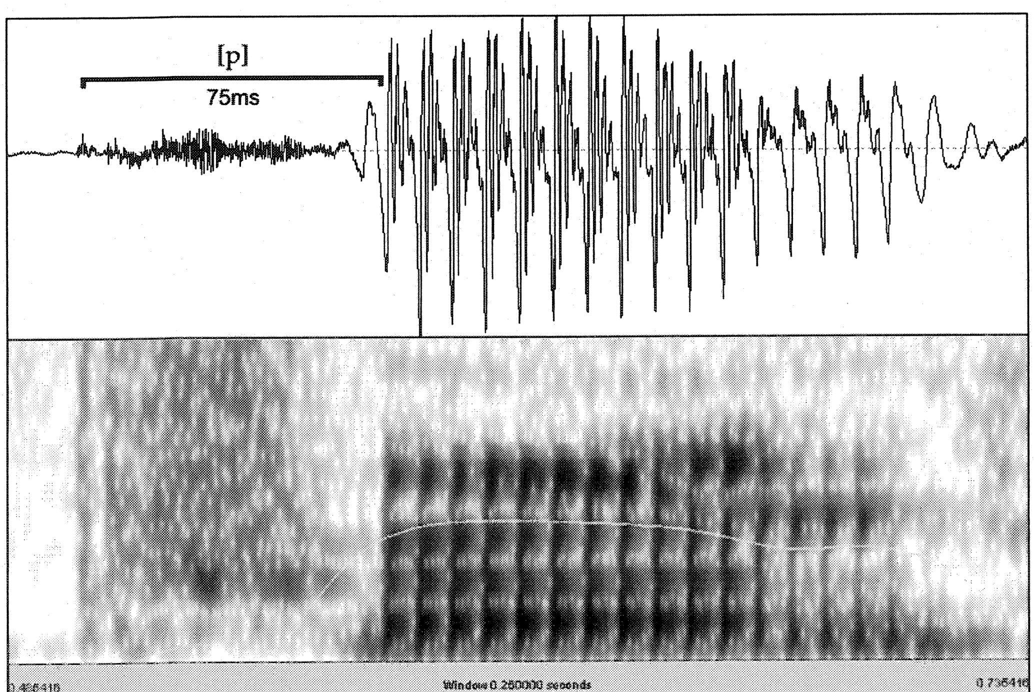

Figure 5.40 Waveform and spectrogram showing an L2 production of /ɾ/ in onset cluster, realized without a svarabhakti vowel. Token [pɹo] in *Dice pronto también*, Session 1, participant A-Adv-M2, scored as "0" for svarabhakti emergence.

awarded a score of "1" for svarabhakti vowel emergence if a sinusoidal waveform with formant structure was present between the release of the onset consonant (e.g., [p] in *prisa*) and a disruption in the sinusoidal waveform and formant structure corresponding to the tap constriction. Figure 5.39 demonstrates an onset tap production scored "1" (accurate) for svarabhakti vowel emergence while Figure 5.40 shows a realization scored "0" (inaccurate) for svarabhakti vowel emergence.

In sum, this chapter has presented the auditory and acoustic features found in the native speaker productions of word-medial /d, t, r, ɾ/, word-initial /r/, word-final /r/, and /ɾ/ in onset cluster positions. For each target category, accuracy criteria were established based on the auditory and acoustic features most common across all native speaker productions; however, due to the slight variation of features found in the native speaker productions, a variability constant was determined for each category in order to normalize the raw accuracy scores of the L2 participants. The second half of this chapter has demonstrated the application of accuracy criteria to the scoring of L2 productions. In the following chapter, the results of the L2 realizations for word-medial Spanish /d, t, r/ ɾ/ are presented.

6. Word-medial results: L2 realizations

The primary purpose of this chapter is to describe the L2 production variants of word-medial /d/ and /t/. Descriptions of word-medial /r/ and /ɾ/ are presented in Chapter 7 and Chapter 8, respectively, in order to allow for their comparison across word positions and because of the similarity of their variants. In the current chapter, all /d/ and /t/ productions, including those productions scored "0" (inaccurate), were assigned to a production variant "type" based on the form's combination of auditory and acoustic features. Following the description of each variant is a presentation of the acoustic measurements for those productions scores as accurate.

The second purpose of this chapter is to address Question 1 and Question 4 of this study. Section 6.2.1 responds to Question 1 which asks: *Is there a relative degree of difficulty of word-medial* /d, t, r, ɾ/*?* In effect, do students show similar score rankings across target accuracy scores? Section 6.2.2 addresses Question 4 of this study: *Is the relative ranking of scores for /t/ (i.e., the relative degree of difficulty) affected by length of voice onset time as a criterion for /t/ accuracy?* Thus, scores for word-medial /t/ are recalculated based on a voice onset time threshold and the relative rankings of scores are reanalyzed.

The organization of this chapter is as follows: Section 6.1 addresses all L2 word-medial /d/ and /t/ productions, independently of scoring. Section 6.1.1 describes the eight variant types identified for word-medial /d/ productions and their frequencies by participant, level, and gender. Section 6.1.2 details the distribution of "spirantized" vs. "approximated" instantiations of /d/ and the acoustic measurements of all "spirantized" productions. In Section 6.1.3, all word-medial /t/ realizations are categorized into one of four types. The acoustic measures for all productions scored "1" (accurate) are then presented in Section 6.1.4. Section 6.2 compares the adjusted overall accuracy scores for word-medial /d, t, r, ɾ/. Section 6.2.1 looks at the rankings of the word-medial contrast scores. In Section 6.2.2, scores for word-medial /t/ are recalculated based on the inclusion of a voice onset time duration criterion and the relative rankings all word-medial contrasts are reanalyzed in light of the new scores for word-medial /t/.

6.1 Description: L2 productions of word-medial /d/ and /t/

6.1.1 Description of word-medial /d/: Variant types

A total of 630 L2 productions of word-medial /d/ were auditorily and acoustically analyzed. In Experiment A, 351 productions were analyzed; 279 productions in Experiment B. Across experiments, the beginning, intermediate, advanced, and experienced levels yielded a total of 180, 135, 216, and 99 realizations, respectively. All L2 productions of word-medial /d/ were categorized into one of eight different variant types which are presented in Table 6.1. The categorization of a type was based on the combination of auditory and acoustic features. Following the establishment of each variant type, each type was associated with an IPA transcription; however, the designated IPA symbol is only presented to aid the reader in differentiating between categories and should not be taken as a narrow transcription. In order to aid the reader, the similar IPA transcription is written following each type reference.

Type 1 (≈[ð̞]) variants were the most common L2 forms realized for word-medial /d/ targets, accounting for 32.4% of all realizations. Productions gave the auditory impression of an absence of complete closure in the oral tract. Visual inspection revealed attenuation in the spectrogram across all frequencies with duration greater than 30ms. Type 1 instantiations demonstrated frication through part of the frequency range as shown in Figure 6.1. Hence, Type 1 variants resembled the "spirantized" tokens realized by the native Spanish speakers and were scored "1" for satisfying the auditory and acoustic criteria.

Table 6.1 Realizations of L2 word-medial /d/ by variant type. Frequency is based on 630 L2 productions.

Word-medial /d/ variant	Similar IPA sound (Aud./ Acou.)	n	Frequency	Auditory impression of oral tract obstruction	Visible presence of voicing bar	Visible presence of continuous spectral energy	Visible attenua-tion of spectral signal > 30ms	Score
Type 1	[ð]	204	32.4%	No	yes	yes	yes	"1"
Type 2	[ð̞]	104	16.5%	No	yes	yes	NA	"1"
Type 3	[d]	152	24.1%	Yes	yes	no	yes	"0"
Type 4	[ɾ]	110	17.5%	yes	yes	no	yes/no	"0"
Type 5	[t]	15	2.4%	yes	no	no	yes	"0"
Type 6	[θ]	7	1.1%	No	no	yes/no	yes	"0"
Type 7	[ɾ] / [ð̞]	32	5.1%	yes	yes	yes	yes	"0"
Type 8	[ð] / [d]	6	1.0%	no	yes/no	no	no	"0"

Type 2 (≈[ð̞]) forms were characterized by a lack of measurable attenuation in the target and represented 16.5% of all productions. These tokens demonstrated a clear continuous formant structure between the formants of the surrounding vowels. Acoustically, a voicing bar was present. In effect, Type 2 variants corresponded to the native speaker "approximated" instantiations. A sample waveform and spectrogram of a Type 2 variant is shown in Figure 6.2. Type 2 forms were also scored "1" (accurate).

Type 3 (≈[d]) realizations accounted for 24.1% of all productions and gave the auditory impression of an obstruction in the oral tract. A visual inspection of the signal revealed three acoustic features: 1) a voiced stop closure as indicated by a sinusoidal waveform without energy across all higher frequencies, 2) a release burst as indicated by a clear band of energy across all frequencies, and 3) short positive voice onset time. Figure 6.3 presents a sample waveform and spectrogram of a Type 3 form which is similar, if not identical, to acoustic features found in some L2 participants' realizations of word-medial English <d>.

Type 4 (≈[ɾ]) variants included realizations characterized by the auditory impression of an obstruction in the oral tract alongside a white stripe in the spectrogram, accounting for 17.5% of all productions. These tokens were auditorily and acoustically similar to native Spanish speaker /ɾ/ productions. In fact, most of these productions would have been scored "1" (accurate) had the target been word-medial /ɾ/; however, in the case of /d/, these instantiations were all scored "0" (inaccurate) due to the auditory impression of an oral tract obstruction and their attenuation in a segment less than 30ms duration. A sample waveform and spectrogram of a Type 4 variant are presented in Figure 6.4.

Type 5 (≈[t]) forms included 2.4% of all productions in which auditory and acoustic features were consistent with native speaker productions of [t]. These tokens gave the auditory impression of an oral tract obstruction and acoustically revealed a stop closure, release burst, and positive voice onset time. A Type 5 variant is shown in Figure 6.6. These instantiations were realized by several participants; however, no speaker consistently realized this form across all /d/ productions. Most of the tokens were represented by the word *dad̲o* 'dart'; however, in several cases the target was represented by *pid̲o* 'I ask for' and *tod̲o* 'all.'

Type 6 (≈[θ]) forms accounted for 1.1% of all productions and demonstrated similar characteristics to Type 1 forms with the exception of voicing. In contrast to Type 1 forms, Type 6 forms did not evidence a voicing bar throughout the segment. These voiceless variants were auditorily impressionistic of the sound represented by <th> in the English word *thin* [θɪn]. A sample waveform and spectrogram of a Type 6 production is shown in Figure 6.6. These instantiations received a score of "0" (inaccurate).

Type 7 (≈[ɾ]; ≈[ð̞]) and Type 8 (≈[ð]; ≈[d]) variants reflected 5.1% and 1.0% of all productions, respectively. Both types are of particular interest in that their forms demonstrated discrepancies between the auditory and acoustic evaluations. For Type 7 realizations, the auditory analysis gave the impression of a complete obstruction of the

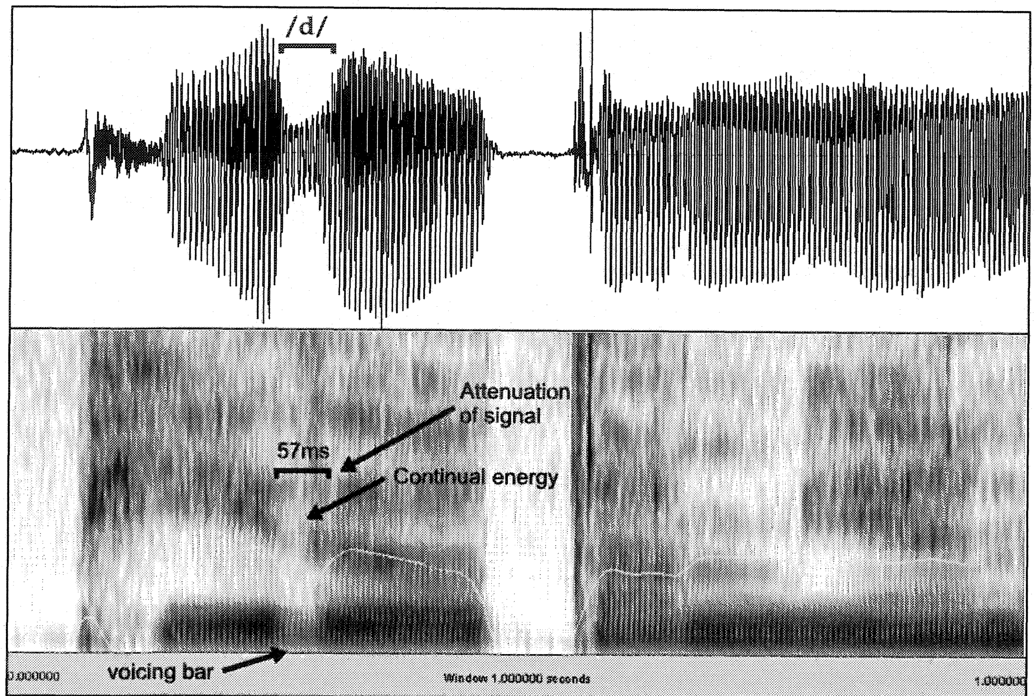

Figure 6.1 Waveform and spectrogram showing a realization of L2 word-medial /d/ designated as a Type 1 variant. Token [piðo tambyen] in *Habla pido también,* Session 1, participant B-Adv-F4, scored as "1" (accurate).

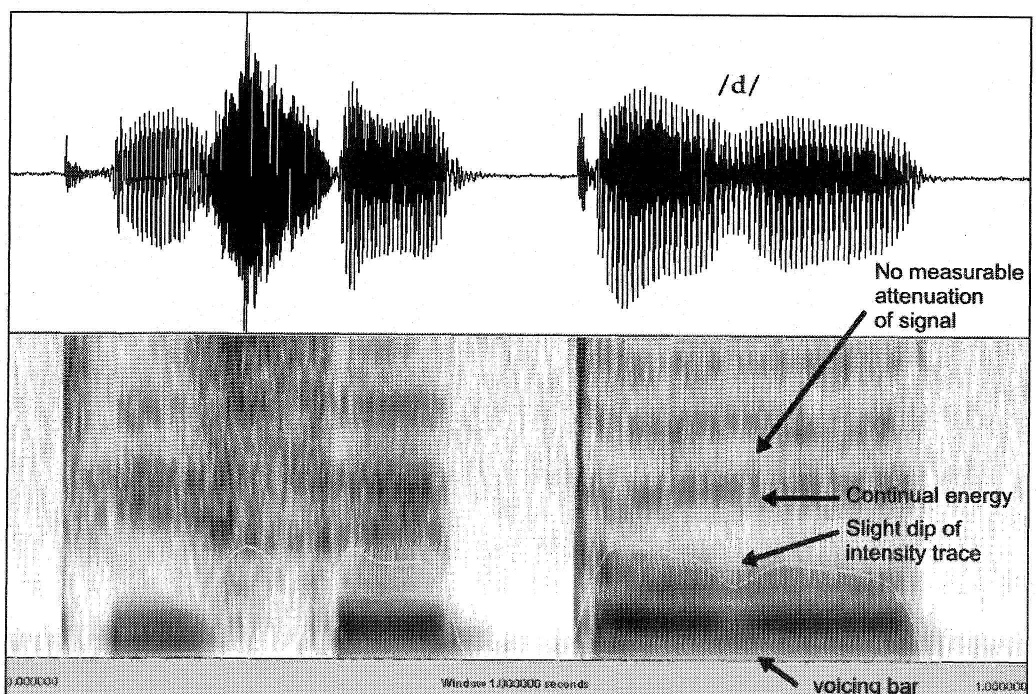

Figure 6.2 Waveform and spectrogram showing a realization of L2 word-medial /d/ designated as a Type 2 variant. Token [dise pio] in *Dice pido también*, Session 3, participant A-Int-F1, scored as "1" (accurate).

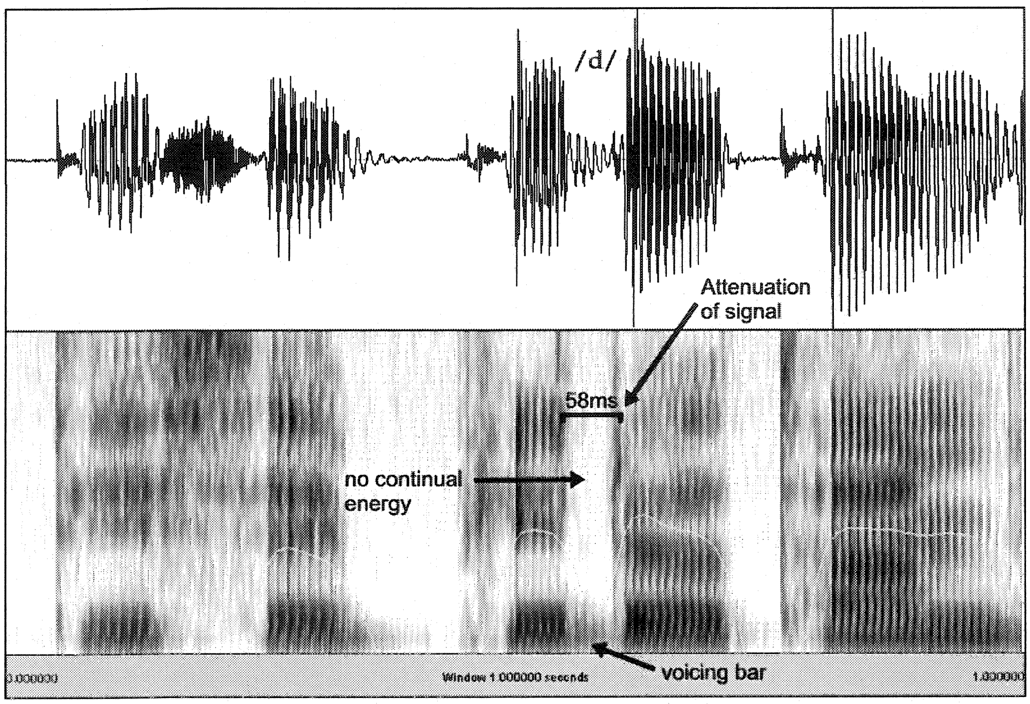

Figure 6.3 Waveform and spectrogram showing a realization of L2 word-medial /d/ designated as a Type 3 variant. Token [dise pido tam] in *Dice pido también*, Session 2, participant A-Beg-M2, scored "0" (inaccurate).

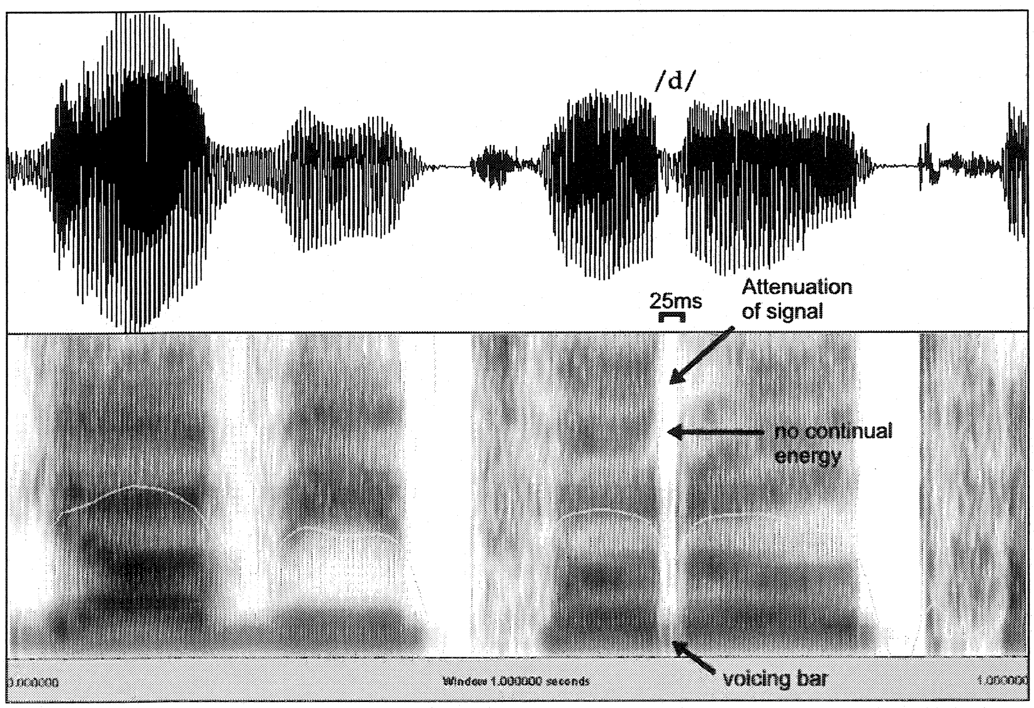

Figure 6.4 Waveform and spectrogram showing a realization of L2 word-medial /d/ designated as a Type 4 variant. Token [ya βi toɾo ta] *Ya vi todo también*, Session 1, participant B-Adv-F6, scored "0" (inaccurate).

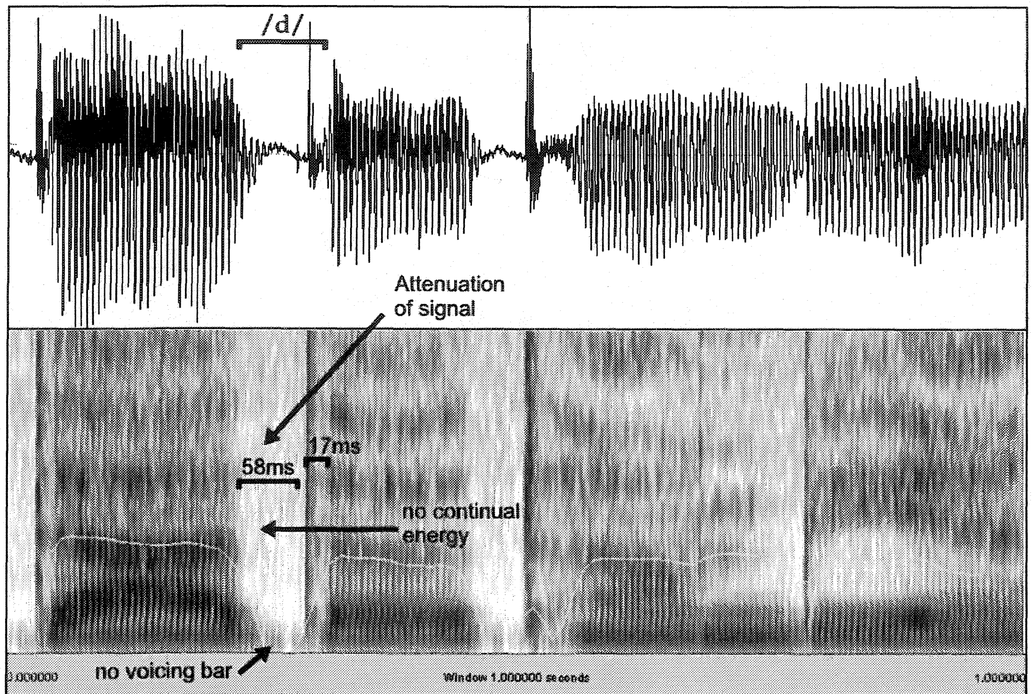

Figure 6.5 Waveform and spectrogram showing a realization of L2 word-medial /d/ designated as a Type 5 variant. Token [dato tambyen] in *Dice dado también*, Session 3, participant A-Beg-F2, scored "0" (inaccurate).

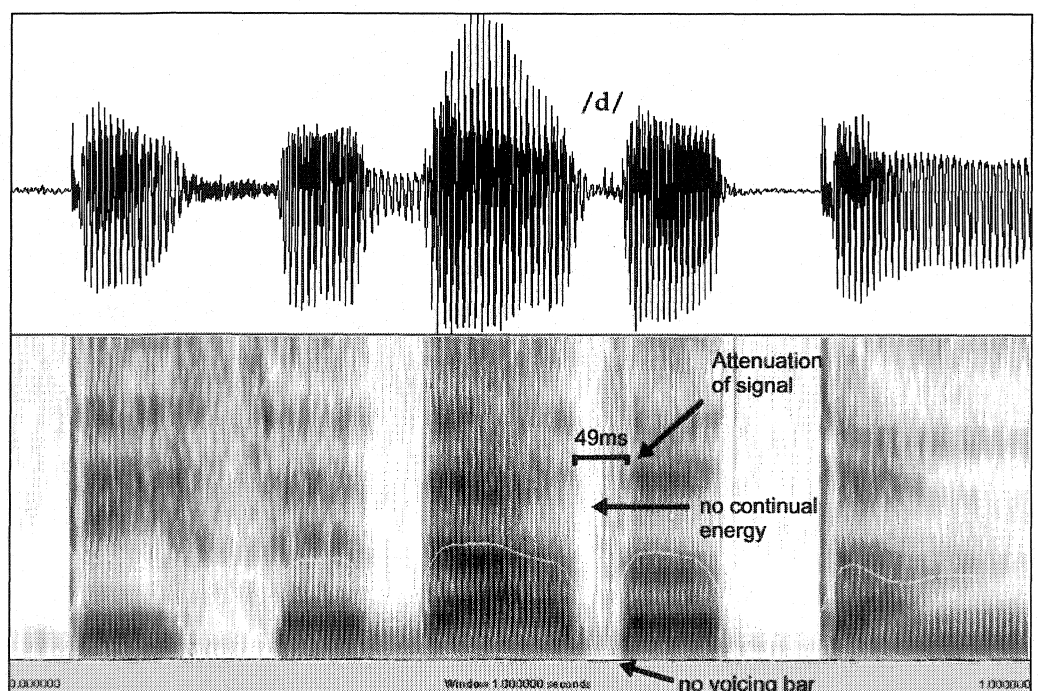

Figure 6.6 Waveform and spectrogram showing a realization of L2 word-medial /d/ designated as a Type 6 variant. Token [dise daθo tam] in *Dice dado también*, participant ExpL-F1, scored as "0" (inaccurate).

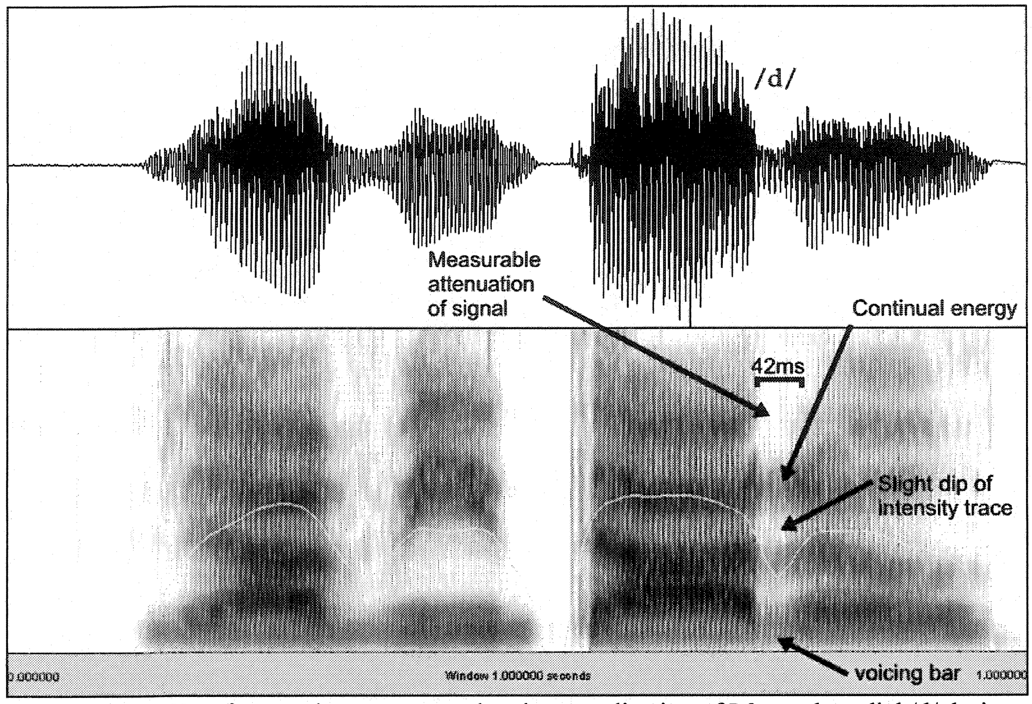

Figure 6.7 Waveform and spectrogram showing a realization of L2 word-medial /d/ designated as a Type 7 variant. Token [ya βi daɾo] in *Ya vi dado también*, Session 1, participant B-Adv-F6, scored "0" (inaccurate).

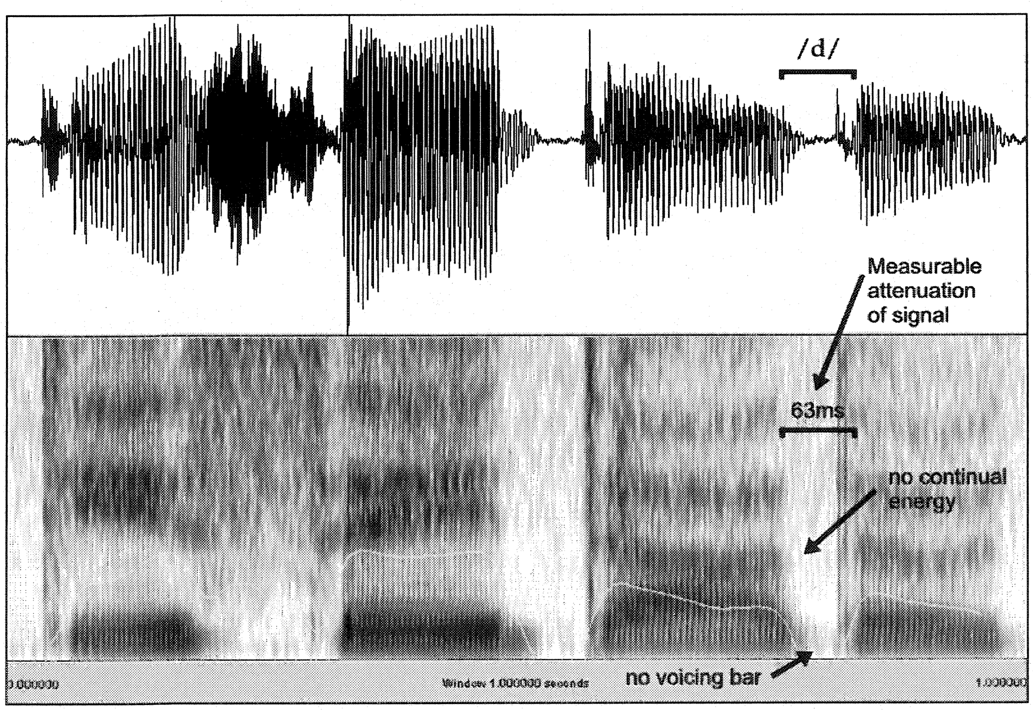

Figure 6.8 Waveform and spectrogram showing a realization of L2 word-medial /d/ designated as a Type 8 variant. Token [dise daθo] in *Dice dado también*, Session 2, participant A-Beg-F2, scored as "0" (inaccurate).

oral tract while the acoustic features were consistent with the criteria to receive a score of "1" (accurate) (i.e., similar to the acoustic features of a Type 2 (≈[ð̞]) variant). As shown in Figure 6.7, the waveform and spectrogram show voicing throughout the segment, clear formant structure throughout the segment, and might have been judged by the acoustic features as an "approximated" production. For Type 8 variants, the discrepancy was reversed: the auditory analysis gave the impression of an absence of oral tract obstruction; however, the acoustic features were not consistent with the criteria to be scored as "1" (accurate) and instead were suggestive of a voiced alveolar stop [d] (i.e., Type 3 (≈[d]) variant). Figure 6.8 shows a waveform and spectrogram of an L2 Type 8 variant.

Table 6.2 shows the frequency of each variant type by Spanish level. The "n" column reflects the number of participants which are included in the overall occurrence. For example, there were 8 participants from the beginning level across both experiments. Averaging across participant frequencies revealed that the Type 3 (≈[d]) variant accounted for 39% of participant productions at the beginning level. The Type 2 (≈[ð̞]) variant was second most frequent at the beginning level with 23% occurrence. Between levels, participant productions were most often categorized as Type 1 (≈[ð]) and Type 2 (≈[ð̞]) variants; 47% and 38% at the Experienced level and 41% and 25% at the advanced level. The frequency of each variant type for each participant can be found in Appendix K.

Table 6.2 Frequency of word-medial /d/ variants by Spanish level. Standard deviation is shown in parentheses. Types 1 and 2 were scored as "accurate."

Level	n	/d/ variant type							
		1 (≈[ð])	2 (≈[ð̞])	3 (≈[d])	4 (≈[ɾ])	5 (≈[t])	6 (≈[θ])	7 (≈[ɾ]; ≈[ð̞])	8 (≈[ð]; ≈[d])
Beginning	8	23% (20%)	5% (10%)	39% (26%)	19% (23%)	7% (11%)	1% (2%)	5% (4%)	3% (4%)
Intermediate	5	21% (34%)	4% (8%)	51% (40%)	21% (29%)	1% (2%)	0% (0%)	1% (2%)	0% (0%)
Advanced	10	41% (37%)	25% (26%)	7% (18%)	19% (23%)	0% (0%)	1% (4%)	7% (8%)	1% (2%)
Experienced	11	47% (25%)	38% (26%)	0% (0%)	4% (13%)	0% (0%)	4% (7%)	5% (17%)	1% (3%)

The frequency of variant types for /d/ by gender group is presented in Table 6.3. The learner groups represent the participants from the beginning, intermediate, and advanced groups; the experienced groups represent participants from the experienced Spanish level. At the learner level, females produced higher frequencies of the Type 1 (≈[ð]) and Type 2 (≈[ð̞]) variants than the frequencies of these types for males. In contrast, the males on average tended to show higher frequencies of variants that were scored "0" (inaccurate). At the experienced level, males demonstrated greater occurrence of Type 2 (≈[ð̞]) forms (44%) than frequency of Type 2 (≈[ð̞])variants for females (35%); however, females were more likely to evidence Type 1 (≈[ð]) forms (57%) than males (31%).

Table 6.3 Frequency of word-medial /d/ variants by gender group. Standard deviation is shown in parentheses. Types 1 and 2 were scored "accurate."

Gender	n	/d/ variant type							
		1 (≈[ð])	2 (≈[ð̞])	3 (≈[d])	4 (≈[ɾ])	5 (≈[t])	6 (≈[θ])	7 (≈[ɾ]; ≈[ð̞])	8 (≈[ð]; ≈[d])
Learner Males	6	6% (9%)	6% (14%)	43% (34%)	38% (25%)	1% (2%)	0% (0%)	7% (7%)	0% (0%)
Learner Females	17	39% (32%)	16% (22%)	23% (30%)	13% (19%)	3% (8%)	1% (3%)	5% (6%)	2% (3%)
Experienced Males	4	31% (21%)	44% (30%)	0% (0%)	11% (22%)	0% (0%)	0% (0%)	14% (28%)	0% (0%)
Experienced Females	7	57% (23%)	35% (25%)	0% (0%)	0% (0%)	0% (0%)	6% (9%)	0% (0%)	2% (4%)

In sum, the L2 productions of word-medial /d/ were categorized into one of eight forms or variant types. These eight variant types were established following auditory and acoustic analyses in order to present the range of variation found across all productions. Type 3 (≈[d]) variants were most frequent for the beginning and intermediate levels while Type 1 (≈[ð]) forms were most common at the advanced and experienced levels. In a description of variant frequencies by gender, learner males were most likely to evidence Type 3 (≈[d]) forms while learner females produced Type 1 (≈[ð])forms most frequently on average. At the experienced level, males most frequently produced Type 2 (≈[ð̞]) forms and females most often evidenced Type 1 (≈[ð]) forms. In the following section, the distribution of Type 1 (≈[ð]) vs. Type 2 (≈[ð̞]) variants is presented and the acoustic measures (i.e., overall segment duration) of Type 1 (≈[ð]) productions (i.e., "spirantized" forms) are summarized.

6.1.2 Measurement results of word-medial /d/ productions scored "accurate"

This section addresses only those word-medial /d/ tokens scored "1" (accurate). First, the frequency of Type 1 ("spirantized": ≈[ð]) vs. Type 2 ("approximated": ≈[ð̞]) variants are detailed. Second, the overall durations of Type 1 (≈[ð]) productions are presented. The description of how spirantized productions were measured can be found in Section 4.4.1.

Table 6.4 summarizes the average frequency of Type 1 (≈[ð]) vs. Type 2 (≈[ð̞]) productions for each level. For example, the average percent of Type 1 (≈[ð]) forms at the advanced level for Experiment A was 73%. Across levels, the intermediate level averaged slightly more Type 1 (≈[ð]) forms; however, the advanced and experienced levels averaged increasingly fewer occurrences of "spirantized" productions. The patterns suggest an increase in the use of approximated forms as L2 Spanish level increases. Appendix L shows the distribution of Type 1 (≈[ð]) and Type 2 (≈[ð̞]) variants of word-medial /d/ for each participant.

Table 6.5 presents the distribution of Type 1 (≈[ð]) and Type 2 (≈[ð̞]) forms based on averaging across participant means in each gender group. The gender groups are divided between the learner and experienced categories. The learner group includes participants from both experiments at the beginning, intermediate, and advanced levels. The experienced group only includes participants from the experienced Spanish level. For the learner group, males demonstrated a greater occurrence of Type 1 (≈[ð]) forms than their female cohort. For the experienced group, the males evidenced fewer Type 1 (≈[ð]) forms than the females. Experienced males evidenced the most Type 2 (≈[ð̞]) forms over any gender group, yielding 59% of all productions scored "1" as "approximated" instantiations.

Table 6.4 Frequency of "spirantized" vs. "approximated" instantiations out of word-medial /d/ productions scored "1" (accurate) by level. Standard deviation is shown in parentheses.

Level	n	Average frequency of Type 1 "spirantized" productions (≈[ð])	Average frequency of Type 2 "approximated" productions (≈[ð̞])
Beginning	7	90% (20%)	10% (20%)
Intermediate	3	94% (11%)	6% (11%)
Advanced	7	67% (29%)	33% (29%)
Experienced	10	56% (23%)	44% (23%)

Table 6.5 Frequency of "spirantized" vs. "approximated" instantiations out of word-medial /d/ productions scored "1" (accurate) by gender group. Standard deviation is shown in parentheses.

Gender	n	Average frequency of Type 1 "spirantized" productions (≈[ð])	Average frequency of Type 2 "approximated" productions (≈[ð̞])
Learner Males	3	100% (0%)	0% (0%)
Learner Females	14	77% (26%)	23% (26%)
Experienced Males	3	41% (6%)	59% (6%)
Experienced Females	7	63% (24%)	37% (24%)

Table 6.6 presents averages across subject duration means, yielding an overall spirant duration by participant level. For example, Experiment B included 5 participants at the advanced level who demonstrated Type 1 (≈[ð]). Across all five participants, the

overall average segment duration was 55ms with 8.6ms standard deviation. Participants at the beginning level demonstrated the shortest spirant durations on average (49ms) while intermediate level subjects showed the longest durations (57ms) on average. In sum, no trends were observed for overall spirant duration by level. Appendix M presents the mean durations of all spirantized productions for each participant with some tokens scored as "accurate."

Table 6.6 Mean overall duration for spirantized productions by level. Standard deviation is shown in parentheses.

Level	n	Mean overall Type 1 (≈[ð]) spirant duration in ms
Beginning	7	48 (9.9)
Intermediate	3	57 (11.3)
Advanced	7	55 (8.6)
Experienced	10	49 (8.3)
Native Speakers	11	49 (11)

In Table 6.7, the overall average spirant durations for gender groups are presented. For both learner and experienced groupings, males yielded an overall segment duration that was less than the overall average for females. Across gender groups, the learner females demonstrated the longest mean overall spirant duration of 54ms (8.6 stdev) while both the learner and experienced male groups yielded an overall mean duration of 46ms.

Table 6.7 Mean overall duration for spirantized productions by gender group. Standard deviation is shown in parentheses.

Gender	n	Mean overall Type 1 (≈[ð]) spirant duration in ms
Learner Males	3	46 (11.2)
Lerner Females	14	54 (8.6)
Experienced Males	3	46 (8.5)
Experienced Females	7	50 (8.5)

In sum, all L2 productions were categorized into one of eight production types which were based on different combinations of auditory and acoustic features. In regards to word-medial /d/ productions scored "1" (accurate), subjects demonstrated varying frequencies of "spirantized" vs. "approximated" tokens; however, results revealed a trend such that lower level learners demonstrated greater percentage of "spirantized" realizations in contrast to the higher level participants. Results for male groups showed a slight trend of having shorter spirant durations on average in comparison with overall average duration for females.

6.1.3 Description of word-medial /t/: Variant types

A total of 630 L2 word-medial /t/ realizations were analyzed. In Experiment A, 351 productions were analyzed while 279 realizations were analyzed in Experiment B. In a breakdown by Spanish level, 180, 135, 216, and 99 productions were analyzed for the beginning, intermediate, advanced, and experienced groups, respectively.

Auditory impressions for word-medial /t/ evidenced little variation across groups and experiments. The impression of a complete closure in the oral tract and the absence of voicing were the two most common auditory impressions which are reflected in the large proportion of tokens scored "1" (accurate). Table 6.8 presents the four variant types which were determined by the different combinations of auditory and acoustic features.

Table 6.8 Variant types for word-medial /t/. Frequency is based on 630 L2 productions.

Word-medial /t/ variant	Similar IPA sound	n	Frequency	Auditory impression of oral tract obstruction	Visible presence of voiceless stop closure	Visible presence of release burst	Visible presence of positive voice onset time	Score
Type 1	[t]	606	96.2%	yes	yes	yes	yes	"1"
Type 2	[ɾ]	9	1.4%	yes	no	yes/no	yes/no	"0"
Type 3	[d]	5	0.8%	yes	no	yes	no	"0"
Type 4	[θ]	10	1.6%	no	no	no	no	"0"

Type 1 (≈[t]) variants were classed as 96.2% of all L2 productions. These forms were all consistent with the auditory and visual criteria based on the native speaker productions and were therefore given a score of "1" (accurate). The productions gave the auditory impression of a complete closure in the vocal tract and the acoustic features demonstrated a voiceless stop closure, release burst, and positive voice onset time. Figure 6.9 shows a sample waveform and spectrogram of a Type 1 variant. Stop closure duration (SCD) and voice onset time (VOT) were the two measures noted for each Type 1 realization; however, each measure demonstrated a wide range of values. For example, Figure 6.10 shows a waveform and spectrogram of a realization with relatively long stop closure duration and short VOT. For comparison, Figure 6.11 reflects a production with relatively short SCD and long VOT.[39]

Type 2 (≈[ɾ]) realizations gave the auditory impression of a complete obstruction in the oral tract obstruction; however, these productions lacked a visible voiceless stop closure within the spectrogram as shown in Figure 6.12. The presence of a release burst

[39] All tokens in Figure 6.9 - Figure 6.11 received a score of "1" (accurate), despite the range of their temporal values; however, in Section 6.2.2, the scoring criteria of word-medial /t/ productions is adjusted and reapplied in order to make the criteria reflect native speaker temporal ranges for /t/.

and accompanying positive VOT may or may not have been present in these productions. Many of these productions would have been scored "1" (accurate) if the target had been word-medial /ɾ/. Type 2 instantiations represented 1.4% of all word-medial /t/ productions.

Type 3 (≈[d]) variants accounted for 0.8% of all productions and gave the auditory impression of an oral tract obstruction; however, an auditory impression of voicing was noted throughout the segment. Visually, a voiced stop closure (represented by a sinusoidal waveform) was present prior to a release burst and positive voice onset time. These realizations were scored "0" (inaccurate) and would not have been scored "1" (accurate) for any L2 Spanish target in this study. A typical Type 3 variant is shown in Figure 6.13.

1.6% of all realizations were classed as Type 4 (≈[θ]) productions and gave the auditory impression of an absence of obstruction in the oral tract. In terms of acoustic features, there was no evidence of a stop closure, release burst, or positive voice onset time. Auditorily and visually, these tokens were similar to spirantized productions of word- medial /d/, without the voicing (i.e., [θ]). Figure 6.14 shows a waveform and spectrogram of a typical Type 4 variant.

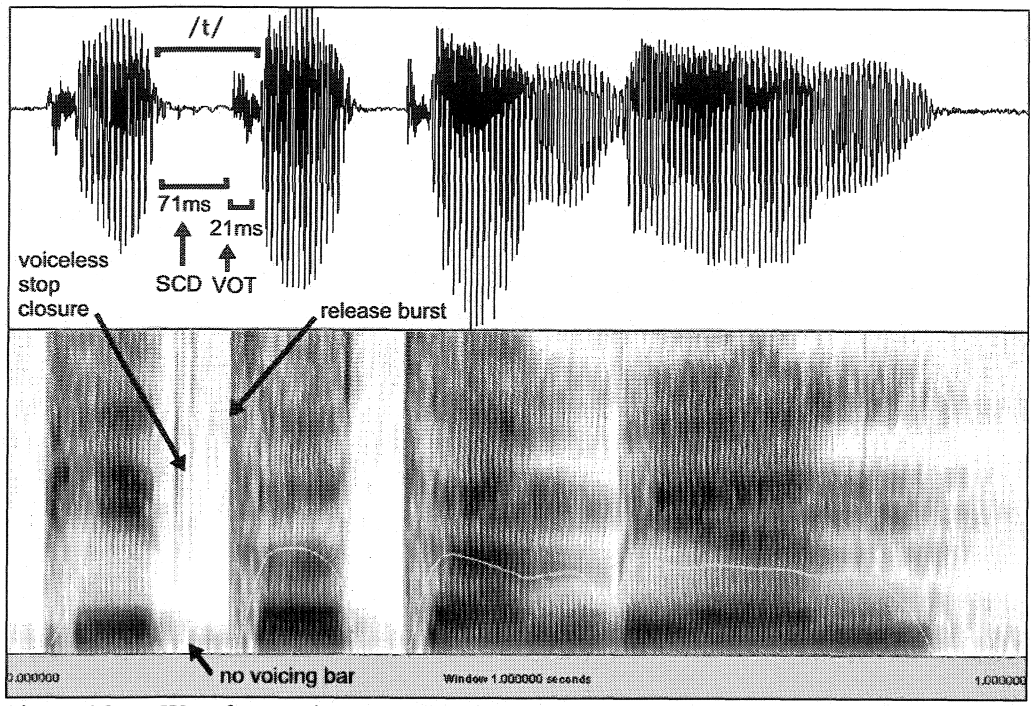

Figure 6.9 Waveform and spectrogram showing a Type 1 realization of L2 word-medial /t/ with typical stop closure duration and voice onset time. Token [pito tambyen] in *Ya vi pito también,* participant ExpL-F7, scored "1" (accurate).

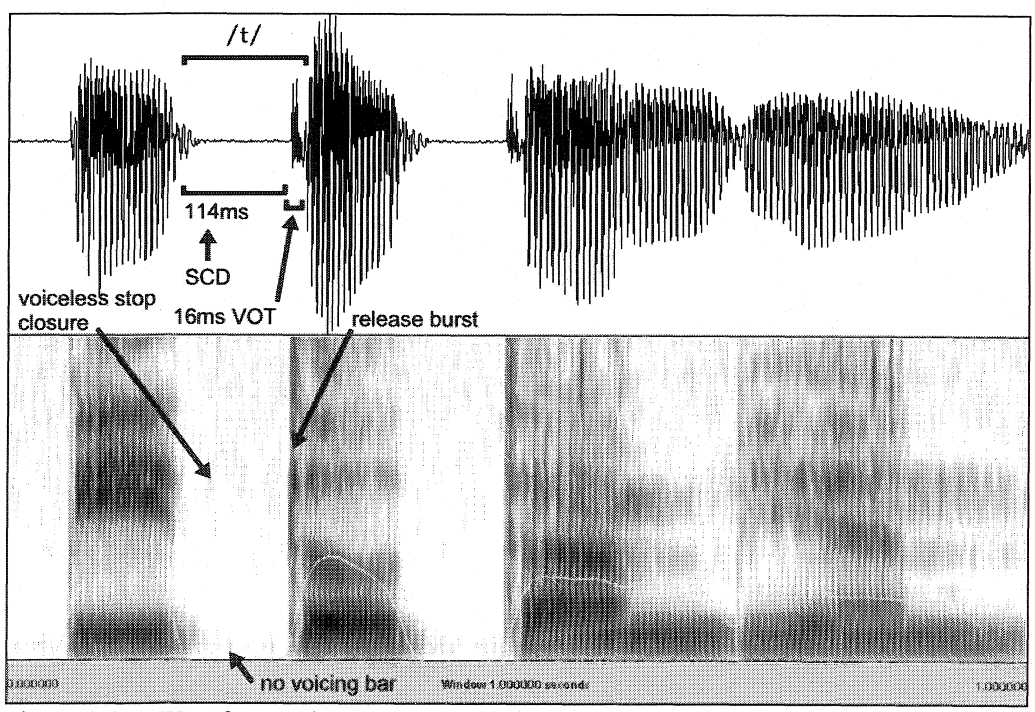

Figure 6.10 Waveform and spectrogram showing a Type 1 realization of L2 word-medial /t/ with relatively long stop closure duration and short voice onset time. Token [pito tambyen] in *Dice pito también*, Session 2, participant B-Adv-F5, scored as "1" (accurate).

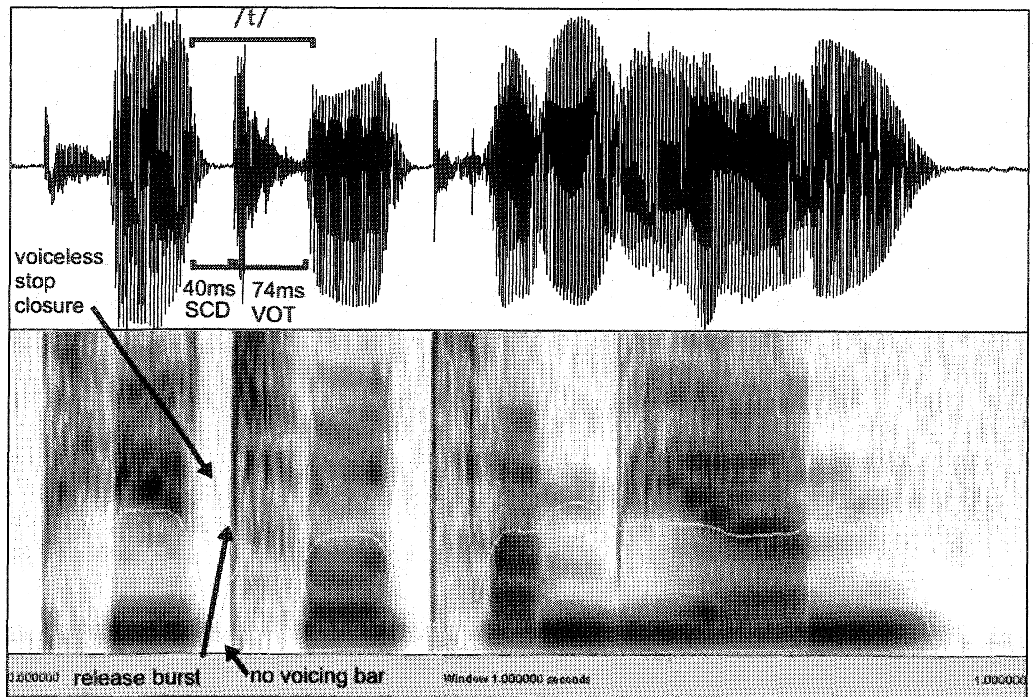

Figure 6.11 Waveform and spectrogram showing a Type 1 realization of L2 word-medial /t/ with relatively short stop closure duration and long voice onset time. Token [pito tambyen] in *Habla pito también*, Session 1, participant B-Beg-F1, scored "1" (accurate).

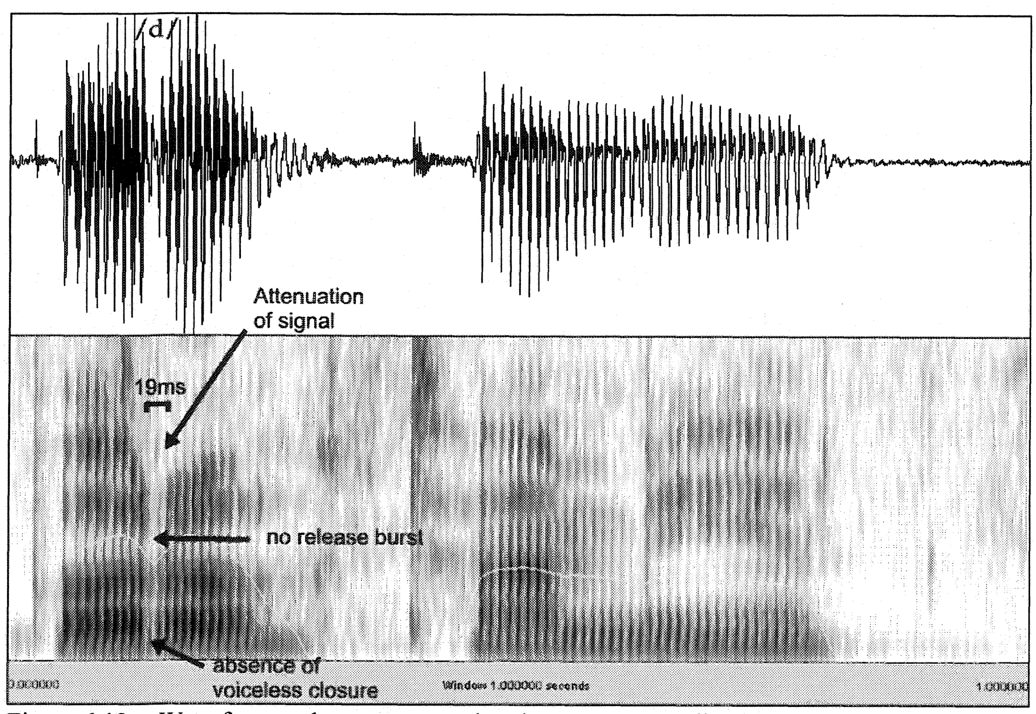

Figure 6.12 Waveform and spectrogram showing a Type 2 realization of L2 word-medial /t/. Token [boɾo tambyen] in *Dice bota también*, Session 1, participant A-Beg-M2, scored "0" (inaccurate).

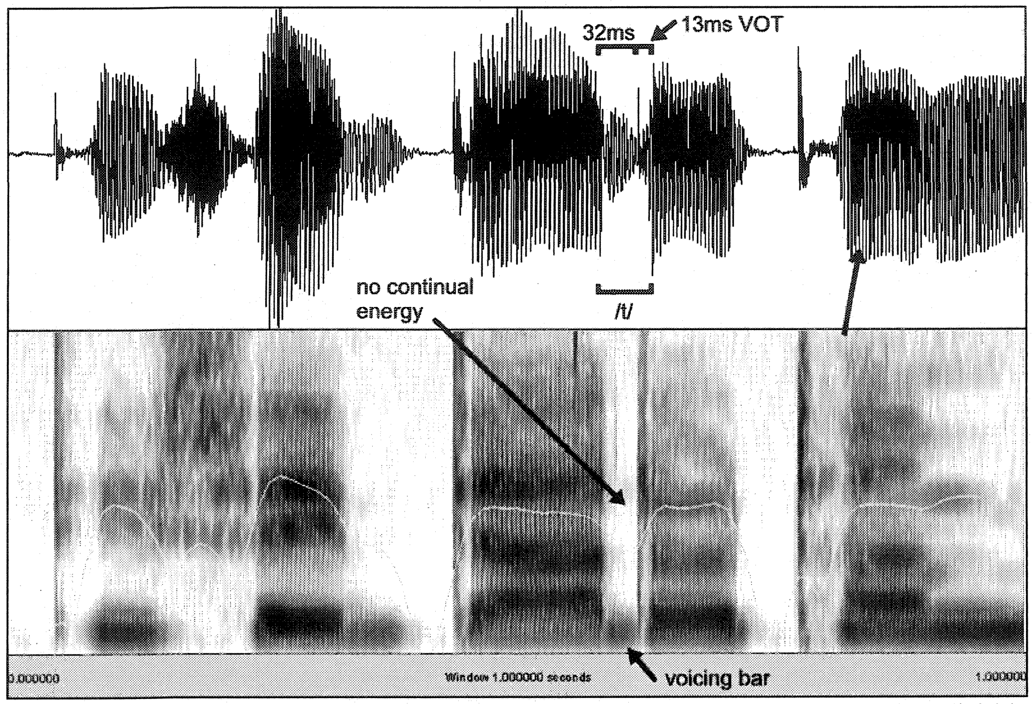

Figure 6.13 Waveform and spectrogram showing a Type 3 realization of L2 word-medial /t/. Token [dise dado tam] in *Dice dato también*, Session 2, participant A-Int-F2, scored "0" (inaccurate).

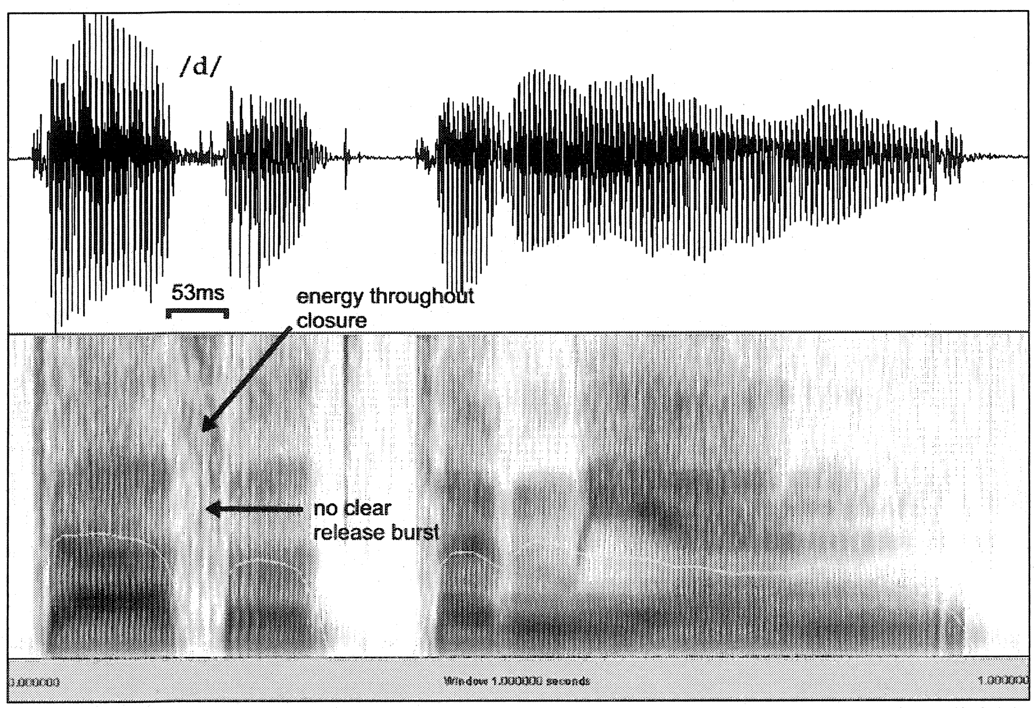

Figure 6.14 Waveform and spectrogram showing a Type 4 realization of L2 word-medial /t/. Token [daθo tambyen] in *Dice dato también*, Session 1, participant A-Adv-F2, scored "0" (inaccurate).

Table 6.9 shows the frequency of /t/ variant types averaged across participant type frequencies at each Spanish levels. For example, all four variant types were evidenced for the beginning and intermediate levels; however, the advanced level only demonstrated Type 1 and Type 4 (≈[θ]) forms. Between levels, the frequency of Type 1 forms (scored as "accurate") increased in accordance with Spanish level. In contrast, Type 4 (≈[θ]) forms (which demonstrated a discrepancy between the auditory and acoustic features) tended to increase in frequency with increased Spanish level, at least up to the advanced level. The frequency of each variant type for each participant can be found in Appendix N.

Table 6.9 Frequency of word-medial /t/ variants by Spanish level. Standard deviation is shown in parentheses.

Level	n	/t/ variant type			
		1 (≈[t])	2 (≈[ɾ])	3 (≈[d])	4 (≈[θ])
Beginning	8	92% (8%)	5% (9%)	2% (3%)	1% (2%)
Intermediate	5	97% (7%)	1% (2%)	1% (3%)	1% (2%)
Advanced	10	98% (4%)	0% (0%)	0% (0%)	2% (4%)
Experienced	11	100% (0%)	0% (0%)	0% (0%)	0% (0%)

Table 6.10 presents the frequency of forms by gender groups. Averaging across the frequency for each participant resulted in an overall learner male frequency of 95.1%, 2.5%, 0%, and 2.5% for Type 1 (≈[t]), 2 (≈[ɾ]), 3 (≈[d]), and 4 (≈[θ]) variants, respectively. For all groups, Type 1 (≈[t]) forms were the most frequent. The Type 3 (≈[d]) variant was only evidenced for the learner female group while Type 2 (≈[ɾ]) and Type 4 (≈[θ]) variants were only found for learner males and learner females. No differences in overall frequency were suggested between experienced males and experienced females.

Table 6.10 Frequency of word-medial /t/ variants by gender group. Standard deviation is shown in parentheses.

Gender	n	/t/ variant type			
		1 (≈[t])	2 (≈[ɾ])	3 (≈[d])	4 (≈[θ])
Learner Males	6	95.1% (7.3%)	2.5% (6.0%)	0.0% (0.0%)	2.5% (3.0%)
Learner Females	17	95.8% (6.5%)	1.5% (5.4%)	1.3% (2.5%)	1.4% (3.0%)
Experienced Males	4	100.0% (0.0%)	0.0% (0.0%)	0.0% (0.0%)	0.0% (0.0%)
Experienced Females	7	100.0% (0%)	0.0% (0%)	0.0% (0%)	0.0% (0%)

To summarize, four variant types were identified for all L2 productions of word-medial /t/. The Type 1 (≈[t]) variant was the most frequent form evidenced for all participants. The beginning and intermediate levels evidenced all four variant types while the advanced level participants only demonstrated Type 1 (≈[t]) and 4 (≈[θ]) variants. At the experienced level, only Type 1 (≈[t]) variants were found. The following section sets out the acoustic measures for the productions designated as Type 1 (≈[t]) forms (i.e., scored as "accurate").

6.1.4 Measurement results of word-medial /t/ productions scored "accurate"

This section reports on the temporal measures collected for all L2 word-medial /t/ realizations that received a score of "1" (accurate) (i.e., all Type 1 (≈[t]) variants). For each production, two measures were noted: stop closure duration and voice onset time. The data here reflect a total of 606 productions and are presented by subject, level, and gender. In addition to mean stop closure duration (SCD) and mean voice onset time (VOT), the mean SCD:VOT ratio is presented as a value to demonstrate the relationship between the SCD and VOT.

Table 6.11 shows the results of word-medial /t/ measures by level. The mean for each measure was calculated by averaging across participant averages in order to prevent an unfair weighting of one participant over another. For example, in both Experiment A and Experiment B, all beginning level participants (n=8) yielded a mean SCD of 73ms (12ms stdev). The average VOT across all beginning level participants was 35ms (11ms stdev) while their mean SCD:VOT ratio was 2.5:1 (0.9 stdev). In regards to SCD, the experienced participants demonstrated the longest mean SCD with 84ms as the average value while the shortest SCD was evidenced on average by students at the intermediate level with 80ms duration. Measures for individual participants are located in Appendix O.

Table 6.11 Mean stop closure duration (SCD) and voice onset time (VOT) for L2 word-medial productions (≈[t]) scored "1" (accurate) by participant level.

Level	n	SCD (ms)		VOT (ms)		SCD:VOT (VOT = 1)	
		mean	stdev	mean	stdev	mean	stdev
Beginning	8	73	12	35	11	2.5	0.9
Intermediate	5	81	23	33	10	3.1	1.6
Advanced	10	80	16	26	8	3.8	1.5
Experienced	11	84	14	21	5	4.4	1.3
Native Speakers	11	84	18	23	6	4.2	1.9

In Table 6.12, the average measures by gender group are presented. The overall average SCD was longest for experienced females and the overall mean VOT was shortest for the experienced females. Thus, the overall average SCD:VOT ratio was greatest for the experienced female group. Within the learner groups, the females evidenced slightly shorter VOT duration on average than males. For the experienced groups, the females also evidenced slightly shorter VOTs on average.

Table 6.12 Table 6.13 Mean stop closure duration (SCD) and voice onset time (VOT) for L2 word-medial productions (≈[t]) scored "1" (accurate) by gender group.

Gender	n	SCD (ms)		VOT (ms)		SCD:VOT (VOT = 1)	
		mean	stdev	mean	stdev	mean	stdev
Learner Males	6	65	8.6	33	9.6	2.4	0.9
Learner Females	17	82	16.1	30	10.4	3.5	1.5
Experienced Males	4	77	13.2	23	4.9	3.6	0.4
Experienced Females	7	88	14.3	20	4.6	4.8	1.4

Mean VOT values across levels demonstrated an inverse relationship between participant and Spanish level. The lower the Spanish level, the longer the mean VOT. For example, all 8 participants at the beginning level yielded a mean VOT of 35ms (11ms stdev); however, at the experienced level, participants demonstrated an overall VOT average of 21ms (5ms stdev). In a similar fashion, the standard deviation of participant averages also revealed a relationship with Spanish level; the range of VOT values decreased with increased Spanish level.

In sum, participants demonstrated a range of stop closure duration (SCD) and voice onset time (VOT) measures. In addition, a SCD:VOT ratio has been provided to reflect the relationship between the two measures. In an analysis by level, SCD tended to increase with increased Spanish level; however, for VOT, values showed an inverse relationship with Spanish level. In an analysis by gender groups, females at the learner and experienced groups demonstrated longer SCD on average and shorter VOTs than their male counterparts. Thus, measures for SCD and VOT suggested a relationship with Spanish level and gender.

6.2 Comparison of accuracy scores for word-medial contrasts

For the participants who recorded in Experiment A, overall accuracy in each category was based on the total number of productions scored "1" out of 27 attempted targets. For the beginning and advanced level participants in Experiment B, adjusted overall accuracy reflects the number of tokens scored "1" (accurate) out of 18 productions. For the experienced level speakers in Experiment B, overall accuracy score were based on 9 realizations total. A total of 2520 word-medial tokens were analyzed across all L2 speakers across word-medial /d, t, r, ɾ/ targets.

6.2.1 Rankings of adjusted overall accuracy scores

Table 6.14 presents adjusted overall accuracy scores for each participant. For example, participant A-Int-M2 was a male participant in Experiment A who was recorded in three separate sessions. Across all three sessions, he yielded an adjusted overall accuracy score of 11% for /d/, 100% for /t/, 100% for /r/, and 80% for /ɾ/. In Experiment B, beginning and advanced level participants were recorded in two separate occasions. For example, participant B-Adv-F3 was a female who received adjusted overall accuracy scores of 67%, 96%, 87%, and 42% for word-medial /d/, /t/, /r/, and /ɾ/, respectively. All scores by session, overall raw accuracy, and adjusted overall accuracy can be found in Appendix D.

Table 6.14 Adjusted overall word-medial target accuracy by beginning, intermediate, and advanced level.

	Participant	Adjusted overall target accuracy			
		/d/	/t/	/r/	/ɾ/
Experiment A	A-Beg-M1	26%	98%	0%	88%
	A-Beg-M2	4%	83%	0%	80%
	A-Beg-F1	56%	94%	75%	8%
	A-Beg-F2	67%	100%	0%	32%
	A-Int-M1	0%	100%	0%	0%
	A-Int-M2	11%	100%	100%	80%
	A-Int-F1	100%	100%	0%	48%
	A-Int-F2	0%	87%	0%	8%
	A-Int-F3	15%	100%	0%	0%
	A-Adv-M1	34%	94%	0%	24%
	A-Adv-M2	0%	100%	0%	0%
	A-Adv-F1	100%	100%	75%	100%
	A-Adv-F2	75%	91%	96%	100%
Experiment B	B-Beg-F1	39%	100%	19%	84%
	B-Beg-F2	11%	79%	0%	72%
	B-Beg-F3	0%	96%	0%	6%
	B-Beg-F4	11%	96%	19%	100%
	B-Adv-F1	67%	100%	0%	66%
	B-Adv-F2	100%	100%	94%	78%
	B-Adv-F3	67%	96%	87%	42%
	B-Adv-F4	100%	100%	100%	90%
	B-Adv-F5	100%	100%	25%	96%
	B-Adv-F6	0%	100%	88%	72%

As shown in Table 6.20, two robust rankings of adjusted overall accuracy scores were observed across all beginning, intermediate, and advanced level participants. Both rankings yielded lowest scores for /r/ and highest scores for /t/. In the first ranking, averages revealed a similar ranking of accuracy scores for 9 out of 23 participants (from lowest to highest): /r <= d <= ɾ <= t/. In the second ranking, averages yielded a similar ranking for an additional 9 out of 23 participants (from lowest to highest) of /r <= ɾ <= d <= t/. Permutation tests yielded a significant result (m=.34; p<.01) on both rankings.

Thus, the chance of randomly generating *any* ranking at least 9 out of 23 times was less than one in 100.

An analysis of rankings between experiments yielded different results than including all participants together. In Experiment A, the most common observed ranking was /r <= ɾ <= d <= t/ which was highly significant (m=.53; p<.001). The second most common ranking was /r <= d <= ɾ <= t/ which was observed in 6 out of 13 participant score rankings, yielding a slightly less robust significant pattern (m=.46; p<.05). In Experiment B, the most common ranking was /r <= d <= ɾ <= t/ and was found only 3 out of 9 times. In effect, there was no significant observed ranking for the participants in Experiment B.

Individual variation was found despite two robust rankings. For example, participant A-Int-M2 demonstrated higher scores for /ɾ/ than for /d/. Participant A-Beg-M2 also showed higher medial /ɾ/ accuracy than scores for /d/. In addition, it should be noted that a participant could have demonstrated more than one ranking if any scores were equal. Table 6.16 shows the raw number of rankings evidenced for each participant. For example, participant B-Adv-F4 demonstrated six rankings due to 100% accuracy for /d, t, ɾ/. Participants with only two equal accuracy scores would have demonstrated two possible rankings (e.g., participant A-Int-F2 who received 0% overall accuracy for both /d/ and /r/). Across all 23 beginning, intermediate, and advanced learners, four participants demonstrated six rankings due to three scores that were equal.

Table 6.15 Rankings of adjusted overall accuracy scores of word-medial contrasts for beginning, intermediate, and advanced level participants.

Ranking (17 total)	Occurrences (50 total)	Participants
/r/ <= /ɾ/ <= /d/ <= /t/	9	A-Beg-F2, A-Int-M1, A-Int-F1, A-Int-F3, A-Adv-M1, A-Adv-M2, A-Adv-F1, B-Adv-F1, B-Adv-F5
/r/<= /d/ <= /ɾ/<= /t/	9	A-Beg-M1, A-Beg-M2, A-Int-M1, A-Int-F2, A-Adv-M2, A-Adv-F1, B-Beg-F1, B-Beg-F2, B-Beg-F3
/ɾ/<= /r/<= /d/ <= /t/	5	A-Int-M1, A-Int-F3, A-Adv-M2, B-Adv-F2, B-Adv-F4
/ɾ/<= /d/ <= /r/<= /t/	5	A-Beg-F1, A-Int-M1, A-Adv-M2, B-Adv-F3, B-Adv-F4
/d/ <= /ɾ/<= /r/<= /t/	4	A-Int-M1, A-Int-M2, A-Adv-M2, B-Adv-F6
/d/ <= /r/<= /ɾ/<= /t/	4	A-Int-M1, A-Int-F2, A-Adv-M2, B-Beg-F3
/r/<= /ɾ/<= /t/ <= /d/	3	A-Int-F1, A-Adv-F1, B-Adv-F5
/ɾ/<= /r/ <= /t/ <= /d/	2	B-Adv-F2, B-Adv-F4
/r/ <= /d/ <= /t/ <= /ɾ/	1	A-Adv-F1
/ɾ/<= /t/ <= /d/ <= /r/	1	B-Adv-F4
/d/ <= /ɾ/ <= /t/ <= /r/	1	A-Int-M2
/ɾ/<= /t/ <= /r/ <= /d/	1	B-Adv-F4
/d/ <= /r/ <= /t/ <= /ɾ/	1	B-Beg-F4
/ɾ/<= /d/ <= /t/ <= /r/	1	B-Adv-F4
/r/ <= /t/ <= /ɾ/<= /d/	1	A-Adv-F1
/d/ <= /t/ <= /r/ <= /ɾ/	1	A-Adv-F2
/r/ <= /t/ <= /d/ <= /ɾ/	1	A-Adv-F1

Table 6.16 Number of ranking patterns for beginning, intermediate, and advanced level participants.

Participant	Number of ranking patterns
A-Beg-M1	1
A-Beg-M2	1
A-Beg-F1	1
A-Beg-F2	1
A-Int-M1	6
A-Int-M2	2
A-Int-F1	2
A-Int-F2	2
A-Int-F3	2
A-Adv-M1	1
A-Adv-M2	6
A-Adv-F1	6
A-Adv-F2	1
B-Beg-F1	1
B-Beg-F2	1
B-Beg-F3	2
B-Beg-F4	1
B-Adv-F1	1
B-Adv-F2	2
B-Adv-F3	1
B-Adv-F4	6
B-Adv-F5	2
B-Adv-F6	1
Total:	59

Participant scores at the experienced level are presented separately since those subjects were recorded on only one occasion and were assumed to have completed their L2 phonological development. Thus, an analysis of their accuracy scores is not assumed to be relevant to the difficulty *within* the acquisition process itself. Table 6.17 shows the observed rankings of experienced level participant scores. Across all 11 participants, all 24 possible rankings were evidenced.[40] The two most common rankings (from lowest to highest) were /ɾ <= d <= r <= t/ and /ɾ <= d <= t <= r/ which were found for five

[40] Given four categories, 4! = 24.

participants; however, a permutation test revealed no significance on either ranking. All experienced level participants satisfied more than one observable ranking due to equal scores across categories. For example, participant B-ExpL-F6 demonstrated 6 observable rankings due to equal scores for /t, r, ɾ/. Participant B-ExpL-M2 demonstrated all 24 possible rankings since all of his scores for /d, t, r, ɾ/ were at 100%. Table 6.18 shows the number of rankings observed for each participant. Individual experienced level participant can be found in Appendix P.

Table 6.19 presents the results of word-medial scores by level. With the exception of the experienced level, each level yielded the same ranking of category scores (from lowest to highest): /r < d < ɾ < t/. The experienced level demonstrated a different ranking (from lowest to highest) of /d < r < ɾ < t/. For the beginning, advanced, and intermediate levels, scores for /r/ were lowest while scores for /t/ were highest; scores for /ɾ/ were higher than /d/ scores. Between levels, the beginning level demonstrated higher scores for /d/ and /ɾ/ than scores for participants at the intermediate level. The advanced level revealed higher scores in all categories in comparison with scores at the intermediate and beginning levels. The experienced level showed higher scores in all categories compared to all lower level scores.

Table 6.17 Word-medial adjusted overall accuracy score rankings and their occurrences for all experienced level participants.

Ranking (24 total)	Occurrences (59 total)	Experienced level participants
/ɾ/<= /d/ <= /r/ <= /t/	5	B-ExpL-M1, B-ExpL-M2, B-ExpL-M4, B-ExpL-F3, B-ExpL-F4
/ɾ/<= /d/ <= /t/ <= /r/	5	B-ExpL-M1, B-ExpL-M2, B-ExpL-M4, B-ExpL-F3, B-ExpL-F4
/ɾ/<= /r/ <= /t/ <= /d/	4	B-ExpL-M1, B-ExpL-M2, B-ExpL-M4, B-ExpL-F3
/ɾ/<= /t/ <= /r/ <= /d/	4	B-ExpL-M1, B-ExpL-M2, B-ExpL-M4, B-ExpL-F3
/ɾ/<= /r/ <= /d/ <= /t/	4	B-ExpL-M1, B-ExpL-M2, B-ExpL-M4, B-ExpL-F3
/ɾ/<= /t/ <= /d/ <= /r/	4	B-ExpL-M1, B-ExpL-M2, B-ExpL-M4, B-ExpL-F3
/r/ <= /d/ <= /ɾ/ <= /t/	4	B-ExpL-M2, B-ExpL-M3, B-ExpL-F1, B-ExpL-F6
/d/ <= /ɾ/<= /r/ <= /t/	3	B-ExpL-M2, B-ExpL-F2, B-ExpL-F5
/d/ <= /r/ <= /ɾ/<= /t/	3	B-ExpL-M2, B-ExpL-M3, B-ExpL-F5
/d/ <= /ɾ/<= /t/ <= /r/	3	B-ExpL-M2, B-ExpL-F2, B-ExpL-F5
/r/ <= /d/ <= /t/ <= /ɾ/	2	B-ExpL-M2, B-ExpL-F6
/r/ <= /ɾ/<= /d/ <= /t/	2	B-ExpL-M2, B-ExpL-F7
/d/ <= /t/ <= /r/ <= /ɾ/	2	B-ExpL-M2, B-ExpL-F5
/r/ <= /ɾ/<= /t/ <= /d/	2	B-ExpL-M2, B-ExpL-F7
/d/ <= /r/ <= /t/ <= /ɾ/	2	B-ExpL-M2, B-ExpL-F5
/d/ <= /t/ <= /ɾ/<= /r/	2	B-ExpL-M2, B-ExpL-F5
/t/ <= /ɾ/<= /d/ <= /r/	1	B-ExpL-M2
/t/ <= /d/ <= /r/ <= /ɾ/	1	B-ExpL-M2
/t/ <= /r/ <= /d/ <= /ɾ/	1	B-ExpL-M2
/t/ <= /d/ <= /ɾ/<= /r/	1	B-ExpL-M2
/r/ <= /t/ <= /d/ <= /ɾ/	1	B-ExpL-M2
/t/ <= /r/ <= /ɾ/<= /d/	1	B-ExpL-M2
/r/ <= /t/ <= /ɾ/ <= /d/	1	B-ExpL-M2
/t/ <= /ɾ/ <= /r/ <= /d/	1	B-ExpL-M2

Table 6.18 Number of ranking patterns for experienced level participants

Participant	Number of ranking patterns
B-ExpL-M1	6
B-ExpL-M2	24
B-ExpL-M3	2
B-ExpL-M4	6
B-ExpL-F1	1
B-ExpL-F2	2
B-ExpL-F3	6
B-ExpL-F4	2
B-ExpL-F5	6
B-ExpL-F6	2
B-ExpL-F7	2
Total	59

Table 6.19 Mean adjusted overall accuracy scores by level. Standard deviation is shown in parentheses.

Level	n	Mean adjusted overall accuracy score			
		/d/	/t/	/r/	/ɾ/
Beginning	8	27% (25%)	94% (8%)	14% (26%)	59% (38%)
Intermediate	5	25% (42%)	97% (6%)	20% (45%)	27% (35%)
Advanced	10	64% (40%)	98% (3%)	56% (44%)	67% (34%)
Experienced	11	84% (30%)	100% (0%)	86% (30%)	90% (12%)

As shown in Table 6.19, the advanced level participants averaged 64% accuracy which included scores for participant A-Adv-F2 who had studied previously for 1 year in Brazil. One might wonder if her interlanguage productions of /d/ were affected by her previous experience due to the nature of standard Brazilian Portuguese to palatalize [d] before a high front vowel (Mateus and d'Andrade 2002). Based on her background, one might expect to find realizations of [dʒ] for /d/; however, Appendix K shows that she produced 74% of all productions as Type 1 (≈[ð]) and Type 2 (≈[ð̞]) which were all scored as "accurate." Additionally, 19% of her productions were categorized as a Type 4 variant (≈[ɾ]). Thus, her variants suggest that the results for the advanced level and females groups were not affected by participant A-Adv-F2's previous experience with Portuguese.

Table 6.20 presents the results of word-medial scores by gender. The male learner group consisted of 6 participants and demonstrated the ranking of /d < r < ɾ < t/. In the female learner group, 17 participants represented subjects from both experiments and yielded a different ranking of /r < d < ɾ < t/. The 4 males from the experienced level demonstrated the same score for /d/ and /r/, yielding the ranking of /d = r < ɾ < t/. The experienced level females demonstrated a fourth ranking across both gender categories of /d < ɾ < r < t/.

Table 6.20 Mean adjusted overall accuracy scores by gender. The "learner" groups include all participants from the beginning, intermediate, and advanced levels. The "experienced" groups include subjects from the experienced level. Standard deviation is shown in parentheses.

Gender	n	Mean adjusted overall accuracy score			
		/d/	/t/	/r/	/ɾ/
Learner Males	6	12% (14%)	96% (7%)	17% (41%)	45% (42%)
Learner Females	17	53% (40%)	96% (6%)	40% (42%)	59% (36%)
Experienced Males	4	75% (50%)	100% (0%)	75% (50%)	91% (8%)
Experienced Females	7	88% (11%)	100% (0%)	96% (6%)	89% (16%)

In sum, an analysis of beginning, intermediate, and advanced level learners resulted in two equally significant rankings (from lowest to highest) of /r <= d <= ɾ <= t/ and /r <= ɾ <= d <= t/; however, an analysis by experiment yielded only significantly observed rankings for participants in Experiment A and not Experiment B. Averaging participant scores by level yielded the same ranking for the beginning, intermediate, and advanced levels: /r < d < ɾ < t/; however, the experienced level demonstrated the ranking of /d < r < ɾ < t/. In an analysis by gender, the male learners revealed the ranking of /d < r < ɾ < t/ while the female learner group yielded the ranking /r < d < ɾ < t/. The males in the experienced level demonstrated the ranking /d = r < ɾ < t/, with equal scores for /d/ and /r/. The experienced level female scores yielded a fourth ranking of /d < ɾ < r < t/. A discussion of these results and the different score rankings can be found in Section 9.1.

6.2.2 The role of voice onset time in scoring word-medial [t]

The purpose of this section is to address Question 4 of this study: *Is the relative ranking of scores for /t/ (i.e., the relative degree of difficulty) affected by length of voice onset time as a criterion for /t/ accuracy?* In the original scoring of L2 word-medial /t/ productions, gradient temporal features were not incorporated into the accuracy criteria. The rationale for omitting a temporal criterion for /t/ was based on the assumption that an L2 production of [t], even with aspiration, was still underlyingly a /t/ and not one of /d, r, ɾ/. Thus, in Section 6.2.1 the scoring of /t/ reflected the presence of categorical auditory and acoustic features rather than the quantity of the VOT feature itself.

It can be argued that complete L2 acquisition of Spanish /t/ requires native-like VOT. As discussed in Section 3.1.2, cross-language differences in length of VOT have been documented for /t/ in Spanish and English (Abramson and Lisker 1973, Lisker and Abramson 1964, Williams 1977); however, comparison to native speaker VOT norms will be limited to the productions of the native speakers in the current study due to the fact that most previous studies looked at productions of VOT word-initially and preceding a stressed syllable.

Because the ultimate scoring of an L2 production is binary (e.g. "1" as accurate vs. "0" as inaccurate), a gradient VOT value must be converted into a categorical criterion/measure of "native-likeness". In addition, the choice of a criterion must be principled. This section takes two approaches to establishing a VOT criterion. In the first approach, the VOT criterion is established based on the complete range of native speaker VOT productions. In effect, the longest native Spanish speaker VOT value is used as a threshold for L2 productions to be scored "1" (accurate). This first approach is motivated based on the assumption that any native speaker VOT duration is equally legitimate. A second more statistically conservative approach treats the distribution of native speaker VOT durations as Gaussian, assuming that outliers and potential speech errors lie at the tails of the distribution. Therefore, the second approach treats any native speaker VOT durations that are more than 2 standard deviations from the mean as non-native-like.

The first part of this section outlines the native speaker VOT range to establish the two VOT duration threshold criteria. Next, both criteria are separately applied to all L2 word-medial /t/ productions, resulting in distinct variability constants and overall accuracy scores for /t/. Lastly, word-medial accuracy scores are recalculated for both criteria and the new relative rankings are presented in light of the original rankings yielded in Section 6.2.1.

As originally presented in Section 5.1.2, native speaker productions of word-medial /t/ evidenced the acoustic features of a voiceless stop closure, release burst, and positive VOT. Native speaker VOT durations ranged from 9-41ms with an average VOT of 22.6ms (n=97; 6.36ms stdev). Figure 6.15 shows a histogram of all native speaker

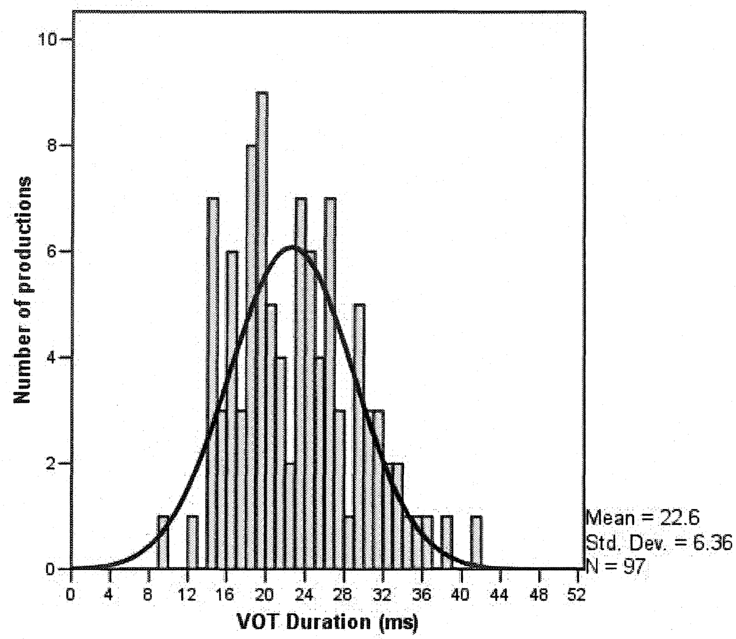

Figure 6.15 Histogram and density curve of voice onset time durations of native speaker word-medial /t/ productions (n=97).

VOT values for word-medial /t/ across all 97 productions.[41] As discussed in Section 3.1.2, these values are slightly longer than those reported by some other studies of word-initial /t/. Lisker and Abramson (1964) found average VOT to be 7ms for productions in sentences and 9ms for isolated word productions. Williams (1977) found VOT to average 10ms, 20ms, and 16ms, for speakers with Guatemalan, Venezuelan, and Peruvian dialects, respectively. The current study's values are most similar to Flege and Efting (1986) who found mean VOT to be 22.4ms (stdev=6ms).

[41] As a reminder to the reader, two of the native speaker productions did not give the impression of a complete obstruction in the oral tract. Acoustically, the tokens were spirantized and did not reveal a clear release burst or positive VOT. In effect, these tokens were excluded as their features contradicted those features needed to maximally account for all native speaker productions of /t/. Thus, based on their exclusion, the variability constant for /t/ scores was .98 (= 97/99).

Figure 6.16 shows that differences between participant VOT duration values (subject levels 1-4) and native speaker values (level 5) are apparent. Consequently, it was necessary to devise a VOT duration criterion that could be used to L2 productions as outside the range of "native-speaker-likeness."

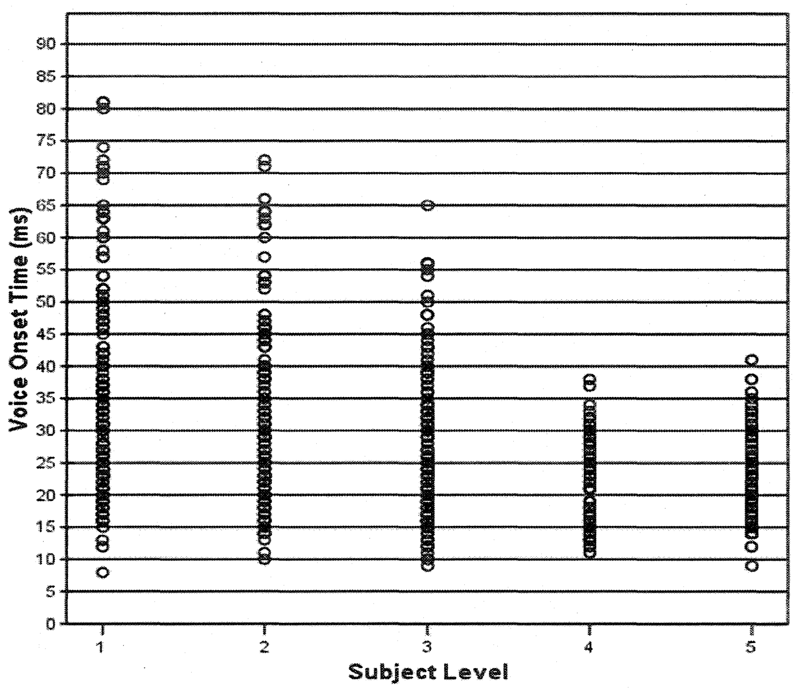

Figure 6.16 Scatter plot showing VOT duration values by subject level. Level 1=Beginning, 2=Intermediate, 3=Advanced, 4=Experienced, and 5=Native Spanish.

The first approach to setting a VOT criterion is motivated by and consistent with the priority given in this thesis to taking all native speaker variability into account. Based on the complete range of native speaker VOT values, a threshold of 41ms was established as the durational criterion (D_{41}) for L2 /t/ productions. Effectively, while previously an L2 production scored as "1" (accurate) could have evidenced *any* positive VOT duration, the application of the D_{41} criterion limited the set of accurate L2 VOT's to those 41ms or less in duration. A variability constant of .98 was used to adjust raw scores calculated with the D_{41} criterion since 97 out of 99 native speaker productions also satisfied all criteria.

The second approach to setting a VOT criterion is principled on the observation that the native speaker VOT durations are normally distributed. A second criterion (D_{35}) is motivated by the reasonable assumption that some of the native speaker variability may be due to speech error or measurement error. Thus, native-likeness may be categorically defined as any VOT duration within 2 standard deviations above the mean, resulting in a threshold duration of 35ms.[42] Thus, any L2 production that was longer than 35ms in duration was scored as "0" (inaccurate) based on the D_{35} criterion. A new variability constant of .95 was established to adjust the D_{35} criterion since 3 additional native speaker productions were excluded from the threshold, yielding a total of 5 out of 99 native speaker productions that did not satisfy the criteria (.95=94/99).

Table 6.21 shows each participant's overall raw accuracy score and adjusted accuracy score without a VOT criterion, with the D_{41} criterion, and with the D_{35} criterion. For example, participant A-Beg-F2 received a raw accuracy score of 100% (e.g., 27 out of 27 productions were scored as "1" - accurate) when VOT duration was not a criterion for accuracy; however, 20 out of her 27 productions demonstrated VOT durations longer than any native speaker production (>41ms). Thus, her overall raw accuracy score was 26% when the D_{41} criterion was applied. With the application of the D_{35} criterion her score dropped to 16%, resulting in a 10% decrease in her accuracy score. A total of 6 participants showed higher adjusted scores with the more stringent D_{35} criterion as a result of having all durations within the 35ms threshold and the application of a lower variability constant for the D_{35} criterion than the D_{41} criterion (.95 vs. .98). In sum, 14 out of 34 participants received lower overall raw accuracy scores with the application of the D_{41} criterion (i.e., they produced at least one production of word-medial /t/ in which the VOT duration was greater than any native speaker VOT duration); 18 participants

[42] Rounding down to the millisecond yields 9ms to be 2 standard deviations from the mean. Since there were no L2 productions less than 9ms (the shortest was 9ms), there were no L2 productions to the left of the mean that were re-scored as "inaccurate." In effect, the criterion for the second approach only affects productions with VOT's greater than 2 standard deviations above the mean.

Table 6.21 Overall accuracy scores of word-medial /t/ without the VOT duration criterion and with the 41ms criterion (D_{41}) and 35ms criterion (D_{35}). Grayed-out cells show higher scores with the application of the more conservative 35ms criterion in comparison to the application of D_{41} criterion.

		Raw score without VOT criterion	Adjusted score (VC=.98)	Raw score w/41ms VOT criterion (D_{41})	Adjusted D_{41} score (VC=.98)	Raw score w/35ms VOT criterion (D_{35})	Adjusted D_{35} score (VC=.98)
Experiment A	A-Beg-M1	96%	98%	96%	98%	96%	100%
	A-Beg-M2	81%	83%	56%	57%	33%	35%
	A-Beg-F1	93%	94%	85%	87%	74%	78%
	A-Beg-F2	100%	100%	26%	26%	15%	16%
	A-Int-M1	100%	100%	48%	49%	33%	35%
	A-Int-M2	100%	100%	96%	98%	93%	97%
	A-Int-F1	100%	100%	100%	100%	93%	97%
	A-Int-F2	85%	87%	33%	34%	11%	12%
	A-Int-F3	100%	100%	85%	87%	78%	82%
	A-Adv-M1	93%	94%	93%	94%	78%	82%
	A-Adv-M2	100%	100%	52%	53%	30%	31%
	A-Adv-F1	100%	100%	100%	100%	100%	100%
	A-Adv-F2	89%	91%	85%	87%	85%	90%
Experiment B	B-Beg-F1	100%	100%	50%	51%	28%	29%
	B-Beg-F2	78%	79%	39%	40%	39%	41%
	B-Beg-F3	94%	96%	83%	85%	67%	70%
	B-Beg-F4	94%	96%	94%	96%	94%	99%
	B-Adv-F1	100%	100%	100%	100%	100%	100%
	B-Adv-F2	100%	100%	94%	96%	94%	99%
	B-Adv-F3	94%	96%	94%	96%	94%	99%
	B-Adv-F4	100%	100%	100%	100%	100%	100%
	B-Adv-F5	100%	100%	100%	100%	100%	100%
	B-Adv-F6	100%	100%	83%	85%	61%	64%
	B-ExpL-M1	100%	100%	100%	100%	100%	100%
	B-ExpL-M2	100%	100%	100%	100%	89%	94%
	B-ExpL-M3	100%	100%	100%	100%	100%	100%
	B-ExpL-M4	100%	100%	100%	100%	100%	100%
	B-ExpL-F1	100%	100%	100%	100%	100%	100%
	B-ExpL-F2	100%	100%	100%	100%	100%	100%
	B-ExpL-F3	100%	100%	100%	100%	100%	100%
	B-ExpL-F4	100%	100%	100%	100%	100%	100%
	B-ExpL-F5	100%	100%	100%	100%	89%	94%
	B-ExpL-F6	100%	100%	100%	100%	100%	100%
	B-ExpL-F7	100%	100%	100%	100%	100%	100%

Table 6.22 Overall raw scores for /t/ by level and change in L2 accuracy scores from no VOT criterion to the use of the D_{41} and D_{35} criteria.

Level	Word-medial /t/ raw accuracy scores				
	Without VOT criterion	With D_{41} criterion	Change in accuracy from no VOT criterion to the D_{41} criterion	With D_{35} criterion	Change in accuracy from no VOT criterion to the D_{35} criterion
Beginning	92%	66%	-26%	56%	-36%
Intermediate	97%	72%	-25%	62%	-35%
Advanced	98%	90%	-8%	84%	-13%
Experienced	100%	100%	0%	98%	-2%

received lower overall raw accuracy scores after the application of the D_{35} criterion (i.e., they produced at least one production in which the duration was greater than 2 standard deviations above the mean).

Table 6.22 presents the overall raw accuracy scores for word-medial /t/ by participant level. For example, averaging across all 10 advanced level participants' overall raw accuracy scores yielded an overall level accuracy of 98% when VOT duration was not included as a criterion. In contrast, the inclusion of the D_{41} criterion reduced this level's raw accuracy score to 90% while the D_{35} criterion reduced the level accuracy score to 84%. The application of either the D_{41} or D_{35} criteria demonstrated a trend in that the lower the Spanish level, the more VOT accuracy criterion reduced level accuracy for word-medial /t/; however, the D_{35} criterion showed greater robustness in lowering accuracy scores for /t/. For the beginning, intermediate, advanced, and experienced levels, the use of the D_{41} criterion reduced overall level raw accuracy by 26, 25, 8, and 0 percentage points, respectively. Similarly, the use of the D_{35} criterion reduced scores by 36, 35, 13, and 2 percentage points for the beginning, intermediate, advanced, and experienced levels, respectively.

Question 4 of this study can now be addressed with the establishment of overall accuracy scores for /t/ that include either the D_{41} or D_{35} VOT criterion. Table 6.23 shows the adjusted accuracy score rankings for all medial contrasts in which the adjusted scores for /t/ include the D_{41} criterion. The ranking /r <= ɾ <= d <= t/ was found 8 out of 23 times and a permutation test revealed this ranking to be significant ($m=.35$; $p<.01$). the second most frequent ranking, /r <= d <= ɾ <= t/, occurred 6 times and was not significant. Thus, the two most frequent rankings were the same as the rankings for scores in which scores for /t/ did not include a VOT criterion (see Section 6.2.1); however, without a VOT criterion both rankings were previously significant while only the first ranking is still significant with the application of the D_{41} criterion.

Table 6.23 Accuracy score rankings for all learner participants in which word-medial /t/ scores reflect the 41ms (D_{41}) criterion.

Ranking (17 total)	Occurrences (48 total)	Experienced level participants
/r/ <= /ɾ/ <= /d/ <= /t/	8	A-Int-M1, A-Int-F1, A-Int-F3, A-Adv-M1, A-Adv-M2, A-Adv-F1, B-Adv-F1, B-Adv-F5
/r/ <= /d/ <= /ɾ/<= /t/	6	A-Beg-M1, A-Int-M1, A-Int-F2, A-Adv-M2, A-Adv-F1, B-Beg-F3
/ɾ/ <= /d/ <= /r/ <= /t/	5	A-Beg-F1, A-Int-M1, A-Adv-M2, B-Adv-F3, B-Adv-F4
/ɾ/ <= /r/ <= /d/ <= /t/	4	A-Int-M1, A-Int-F3, A-Adv-M2, B-Adv-F4
/d/ <= /r/ <= /ɾ/ <= /t/	4	A-Int-M1, A-Int-F2, A-Adv-M2, B-Beg-F3
/r/ <= /d/ <= /t/ <= /ɾ/	4	A-Beg-M2, A-Adv-F1, B-Beg-F1, B-Beg-F2
/r/ <= /ɾ/ <= /t/ <= /d/	3	A-Int-F1, A-Adv-F1, B-Adv-F5
/d/ <= /ɾ/ <= /r/ <= /t/	2	A-Int-M1, A-Adv-M2
/r/ <= /t/ <= /ɾ/ <= /d/	2	A-Beg-F2, A-Adv-F1
/d/ <= /ɾ/ <= /t/ <= /r/	2	A-Int-M2, B-Adv-F6
/ɾ/ <= /r/ <= /t/ <= /d/	2	B-Adv-F2, B-Adv-F4
/ɾ/ <= /t/ <= /d/ <= /r/	1	B-Adv-F4
/d/ <= /t/ <= /r/ <= /ɾ/	1	A-Adv-F2
/ɾ/ <= /t/ <= /r/ <= /d/	1	B-Adv-F4
/ɾ/ <= /d/ <= /t/ <= /r/	1	B-Adv-F4
/r/ <= /t/ <= /d/ <= /ɾ/	1	A-Adv-F1
/d/ <= /r/ <= /t/ <= /ɾ/	1	B-Beg-F4

Table 6.24 shows the individual ranking patterns of accuracy scores when scores for /t/ are adjusted with the D_{35} criterion. Grayed-out cells indicate a change from the rankings in Table 6.23 which were based on the D_{41} criterion. For example, participant A-Int-F1 no longer had equal scores for /d/ and /t/. Thus, there were only 7 rankings

found with the use of the D_{35} criterion that satisfied the observed ordering of scores (from lowest to highest) such that /r <= ɾ <= d <= t/. A permutation test found this ordering significant (m=.30;p<.05); however, the effect was slightly less robust than the 8 rankings found with the D_{41} criterion. The second most common ordering (from lowest to highest: /r <= d <= ɾ <= t/) was evidenced 6 times and was not significant.

Table 6.24 Accuracy score rankings for all learner participants in which word-medial /t/ scores reflect the 35ms (D_{35}) criterion. The grayed-out cells indicate a change in patterning from Table 6.23 in which rankings reflect the application of the 41ms criterion.

Ranking (18 total)	Occurrences (47 total)	Experienced level participants
/r/ <= /ɾ/ <= /d/ <= /t/	7	A-Int-M1, A-Int-F3, A-Adv-M1, A-Adv-M2, A-Adv-F1, B-Adv-F1, B-Adv-F5
/r/ <= /d/ <= /ɾ/<= /t/	6	A-Beg-M1, A-Int-M1, A-Int-F2, A-Adv-M2, A-Adv-F1, B-Beg-F3
/ɾ/ <= /d/ <= /r/ <= /t/	5	A-Beg-F1, A-Int-M1, A-Adv-M2, B-Adv-F3, B-Adv-F4
/ɾ/ <= /r/ <= /d/ <= /t/	4	A-Int-M1, A-Int-F3, A-Adv-M2, B-Adv-F4
/d/ <= /r/ <= /ɾ/ <= /t/	4	A-Int-M1, A-Int-F2, A-Adv-M2, B-Beg-F3
/r/ <= /d/ <= /t/ <= /ɾ/	3	A-Beg-M2, A-Adv-F1, B-Beg-F2
/r/ <= /ɾ/ <= /t/ <= /d/	3	A-Int-F1, A-Adv-F1, B-Adv-F5
/d/ <= /ɾ/ <= /r/ <= /t/	2	A-Int-M1, A-Adv-M2
/r/ <= /t/ <= /ɾ/ <= /d/	2	A-Beg-F2, A-Adv-F1
/d/ <= /ɾ/<= /t/ <= /r/	1	A-Int-M2
/ɾ/ <= /r/ <= /t/ <= /d/	2	B-Adv-F2, B-Adv-F4
/ɾ/<= /t/ <= /d/ <= /r/	1	B-Adv-F4
/d/ <= /t/ <= /r/ <= /ɾ/	1	A-Adv-F2
/ɾ/<= /t/ <= /r/ <= /d/	1	B-Adv-F4
/ɾ/ <= /d/ <= /t/ <= /r/	1	B-Adv-F4
/r/ <= /t/ <= /d/ <= /ɾ/	2	A-Adv-F1, B-Beg-F1
/d/ <= /r/ <= /t/ <= /ɾ/	1	B-Beg-F4
/d/ <= /t/ <= /ɾ/<= /r/	1	B-Adv-F6

Table 6.25 presents the mean adjusted overall accuracy scores by level in which scores for /t/ reflect the VOT duration criterion. The number in parentheses shows the average overall score for /t/ without the VOT criterion. For example, all intermediate

level participants averaged 67% accuracy for /t/ with the VOT criterion; 97% accuracy without any duration criterion. In effect, all participant levels demonstrated the /r < d < ɾ < t/ ranking both with and without the VOT duration criterion; however, the /r < ɾ < d < t/ ranking was not significant when either VOT duration criterion was included.

Table 6.25 Mean adjusted overall accuracy scores by level when scores for /t/ include the D41 and D35 criteria.

Level	n	Mean adjusted overall accuracy score					
		/d/	/t/ with no VOT criterion	/t/ with D_{41} criterion	/t/ with D_{35} criterion	/r/	/ɾ/
Beginning	8	27%	93%	67%	59%	14%	59%
Intermediate	5	25%	97%	73%	65%	20%	27%
Advanced	10	64%	98%	91%	87%	56%	67%
Experienced	11	84%	100%	100%	99%	86%	90%

Table 6.26 shows the mean adjusted overall accuracy scores by gender in which scores for /t/ include both VOT duration criteria. For example, mean adjusted overall accuracy for the learner males (i.e., all six males from the beginning, intermediate, and advanced levels across both experiments) was 76% for /t/ when using the D_{41} criterion in comparison to 96% accuracy without the VOT criterion. The score for learner males was lowest with the application of the D_{35} criterion (63%). The observed ranking for both male and females learner groups did not change with the inclusion of either VOT criterion. The ranking for males was /d < r < ɾ < t/ and /r < d < ɾ < t/ for females.

Table 6.26 Mean adjusted overall accuracy scores by gender when /t/ scores include the VOT criterion. The number is parentheses reflect scores without the VOT criterion.

Level	n	Mean adjusted overall accuracy score					
		/d/	/t/ with no VOT criterion	/t/ with D_{41} criterion	/t/ with D_{35} criterion	/r/	/ɾ/
Learner Males	6	12%	95%	76%	63%	17%	45%
Learner Females	17	53%	96%	81%	75%	40%	59%
Experienced Males	4	75%	100%	100%	99%	75%	91%
Experienced Females	7	88%	100%	100%	99%	96%	89%

In sum, scores for /t/ were twice reanalyzed using two different VOT duration criteria. The first criterion (D_{41}) was motivated to account for all VOT variability in the native speaker productions, resulting in a threshold criterion of 41ms (i.e. representing the longest native speaker VOT). The second criterion (D_{35}) was principled on the normal distribution of the native speaker VOT durations. Following a statistically conventional approach, the threshold criterion was established at 35ms, representing two standard deviations above the native speaker mean. The two criteria account for native speaker variability in different ways, the distinction when comparing target scores was not meaningful. The application of the D_{41} criterion did not affect any of the experienced level participant scores; however, the application of the D_{35} criterion reduced scores for 2 experienced level participants. For the learner group participants, raw scores for /t/ were lowered for 14 out of 23 participants when using the D_{41} criterion while raw scores for 18 out of 23 participants were lowered with the D_{35} criterion. A reanalysis of all medial contrast scores when using either VOT criterion demonstrated the significant ranking of /r <= ɾ <= d <= t/; however, the second ranking of /r <= d <= ɾ <= t/ was no longer found to be significant when using either VOT criterion. An analysis of rankings by level and gender with either VOT criterion also yielded the same ranking patterns as were found without the VOT criterion. In all analyses, all level groups demonstrated the ranking /r < d < ɾ < t/. Similarly, both re-analyses of mean scores by gender for the learner groups demonstrated the same rankings: /d < r < ɾ < t/ for males and /r < d < ɾ < t/ for females. Section 9.1 presents a discussion of these results and the use of both criteria where it addresses Question 4 of this study.

7. Trills results: L2 realizations

The primary purpose of this chapter is to describe the range of variants produced by L2 participants in both Experiment A and B. Word-initial and word-medial /r/ forms are presented concurrently due to the similarity of forms across word positions.[43] In the current chapter, seven variant types of /r/ are identified based on the observed combinations of auditory and acoustic features.

The secondary purpose of this chapter is to compare productions of /r/ scored "1" (accurate) by word position and vowel height. In effect, two of this study's questions are addressed. Question 2 asked: *Does trill accuracy emerge with higher scores word-medially over word-initial trill scores?* Question 5 asked: *Is there an effect of vowel height on trill productions scored as accurate?*

This chapter is organized as follows: Section 0 describes 7 variant types that were identified for all L2 productions of word-initial and word-medial /r/, and presents the acoustic measures for accurate productions. In Section 7.2.1, Question 2 of this study is addressed and productions are compared by word position. Section 7.2.2 addresses Question 5 of this study by comparing the vowel heights of "accurate" realizations.

7.1 Description: L2 productions of /r/

7.1.1 Word-medial and word-initial /r/: Variant types

A total of 1260 /r/ productions were auditorily and acoustically analyzed; 630 realizations for word-initial /r/ and 630 for word-medial /r/. For each variant categorization, the

[43] The comparison of word-medial /r/ with word-medial /d, t, ɾ/ can be found in Section 6.2.1.

production was noted as having given the impression of an absence of an obstruction, a single obstruction, or two or more successive obstructions in the oral tract. The acoustic features for each type included the number of visible stripes on the spectrogram and, if possible, the duration of attenuation (e.g., measurement of the segment representative of an approximant [ɹ] was not assumed reliable). The combination of these auditory and acoustic features led to the identification of seven variant types for word-medial and word-initial /r/ tokens. A similar IPA symbol is also associated with each type in order to aid the reader in associating a type with a set of features. The types and features are shown in Table 7.1.

Table 7.1 Realizations of L2 /r/ by variant type. Frequency is based on 1260 L2 productions.

/r/ variant type	Similar IPA sound (Aud./ Acou.)	n	Frequency	Auditory impression: number of obstructions	Visible number of stripes	Visible attenuation of spectral signal 50>ms	Score
Type 1	[r]	347	27.5%	2+	2+	yes	"1"
Type 2	[ɾ]	403	32.0%	1	1	no	"0"
Type 3	[ɹ]	361	28.7%	0	0	na	"0"
Type 4	[ɾ]/[ɹ]	55	4.4%	1	0	na	"0"
Type 5	[ɾ]/[r]	34	2.7%	1	0+	yes/no	"0"
Type 6	[ɹ]/[z]	35	2.8%	0	0	yes/no	"0"
Type 7	[r]/[ɹ, ɾ]	25	2.0%	2+	0-1	no	"0"

Type 1 (≈[r]) variants had been scored as "1" (accurate) and demonstrated auditory and acoustic features consistent with the native speaker baselines, accounting for 28% of all forms. The auditory analysis gave the impression of two or more rapid successive obstructions in the oral tract. Upon visual inspection of the spectrogram, two or more visible stripes were present. The duration of these segments was greater than

50ms. A sample waveform and spectrogram of a Type 1 /r/ form is presented in Figure 7.1.

Type 2 (≈[ɾ]) forms were characterized by the auditory impression of a single oral tract obstruction and the visible presence of a single stripe in the spectrogram. All Type 2 forms were scored as "0" (inaccurate) and were identified in 32% of all word-initial and word-medial /r/ productions. Figure 7.2 shows a realization categorized as a Type 2 variant. The majority of these forms would have been scored "1" (accurate) had the target been word-medial /ɾ/.

Type 3 (≈[ɹ]) forms were associated with 29% of all realizations and shared auditory and acoustic features with L1 English participant productions of L1 English target <r> (e.g., *starry* [staɹi]). Auditorily, these productions did not give any impression of an oral tract obstruction. Acoustically, the spectrogram and waveform demonstrated little, if any, attenuation in the segment. The visual signal demonstrated a continuous formant structure throughout the target and a dip in F3 which is characteristic of the English approximant [ɹ] (Hashi et al. 2003); however, a dip in F2 may have been present instead of a dip in F3, suggesting a raised tongue-tip during production (Hagiwara 1995). All Type 3 forms were scored as "0" (inaccurate). A sample waveform and spectrogram of a Type 3 form is shown in Figure 7.3.

Type 4 (≈[ɾ]; ≈[ɹ]) realizations demonstrated a discrepancy between the auditory and acoustic features required to receive a score of "1" (accurate). Auditorily, productions gave the impression of a single obstruction in the oral tract; however, the acoustic signal did not demonstrate any stripes. The waveform and spectrogram revealed a continuous voiced formant structure with a slight dip in the intensity trace. In many instantiations, the formant structure revealed a drop in F3, consistent with the English approximant [ɹ]. Type 4 forms represented 4.4% of all productions and were all scored as "0" (inaccurate). Figure 7.4 shows a sample waveform and spectrogram of a Type 4 variant.

Fewer than 3% of all word-initial and word-medial productions were classified as Type 5 (≈[ɾ]; ≈[r]) forms. These instantiations gave the impression of a single obstruction in the oral tract; however, a visual inspection of the waveform and spectrogram revealed frication with as many as two stripes in the spectrogram. These forms gave the auditory impression of [ɾ] but visually conformed to a [r] with or without frication. Thus, these instantiations received a score of "0" (inaccurate). Figure 7.5 demonstrates a waveform and spectrogram of a Type 5 variant.

Type 6 (≈[ɹ]; ≈[z̺]) realizations were associated with approximately 3% of all productions. The productions did not auditorily suggest a complete obstruction in the oral tract. Spectrally, the realizations were consistent with Type 3 (≈[ɹ]) forms; however, frication was noted across the higher frequencies in the spectrogram. Thus, many of these instantiations auditorily and acoustically fell between an approximant [ɹ] and a retroflex apicoalveolar fricative [z̺] as shown in Figure 7.6. All Type 6 forms were scored as "0" (inaccurate).

Type 7 (≈[r]; ≈[ɹ, ɾ]) variants included 2% of all productions and demonstrated a discrepancy between the auditory and acoustic analyses required to received a score of "1" (accurate). The auditory impression reflected two or more successive obstructions in the oral tract; however, the visual inspection did not reveal two or more stripes. Despite the appearance of a single stripe, the intensity trace contained two dips within the segment, suggesting amplitude modulation. All Type 7 instantiation were scored "0" (inaccurate). Figure 7.7 shows a sample waveform and spectrogram of a Type 7 form.

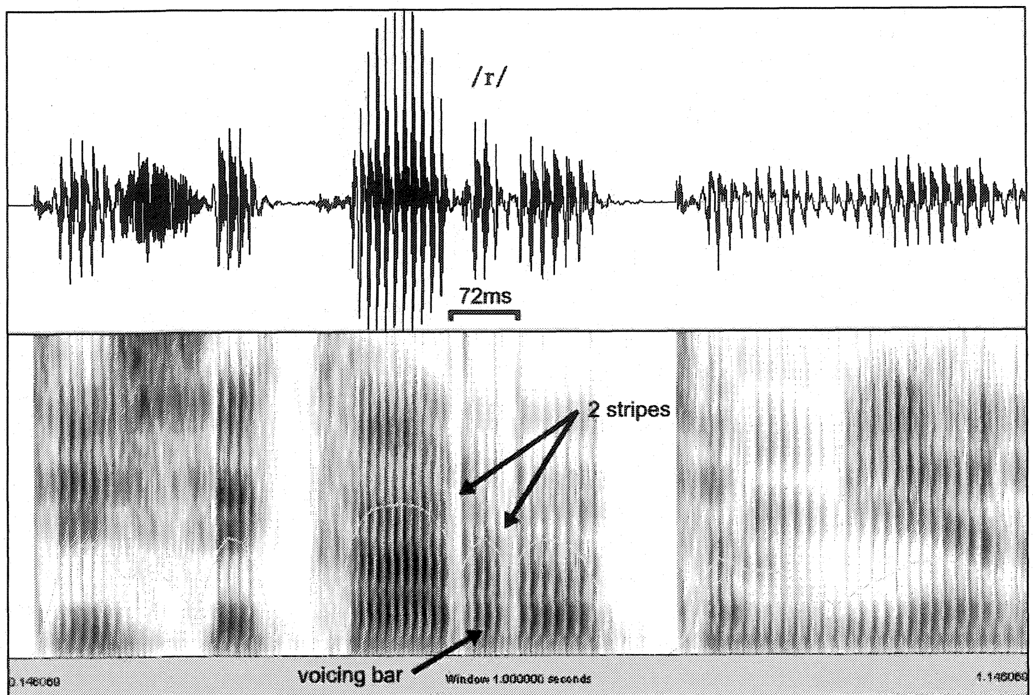

Figure 7.1 Waveform and spectrogram showing an L2 realization of word-medial /r/ designated as a Type 1 variant. Token [dise karo tambyen] in *Dice carro también,* Session 2, participant A-Int-M2, scored as "1" (accurate).

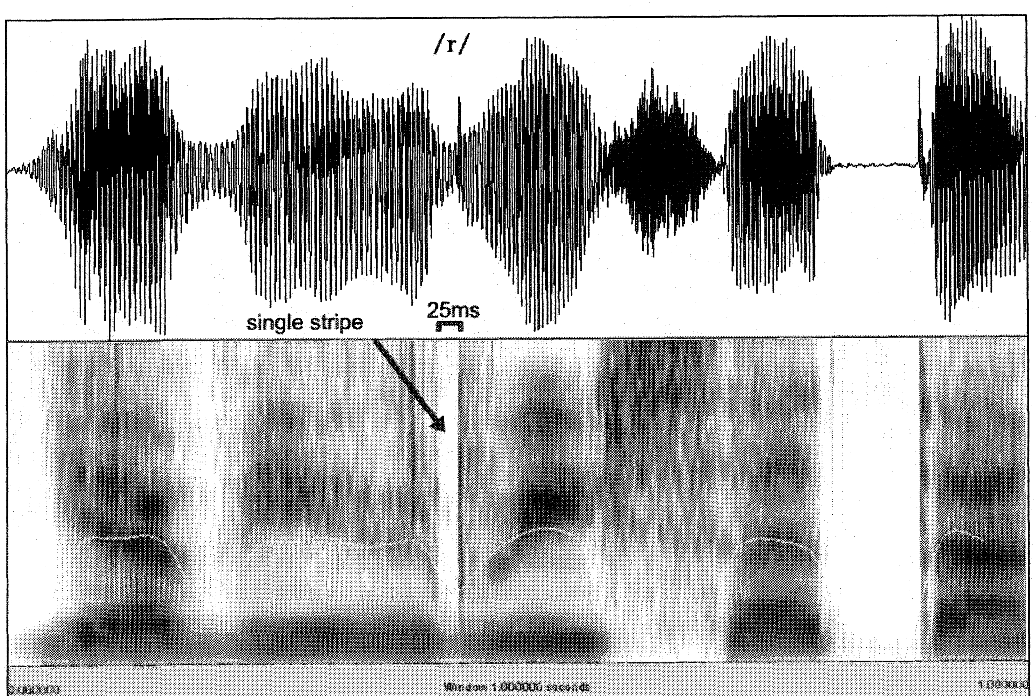

Figure 7.2 Waveform and spectrogram showing an L2 realization of word-initial /r/ designated as a Type 2 variant. Token [ya βi ɾisa ta] in *Ya vi risa también*, Session 1, participant B-Adv-F3, scored as "0" (accurate).

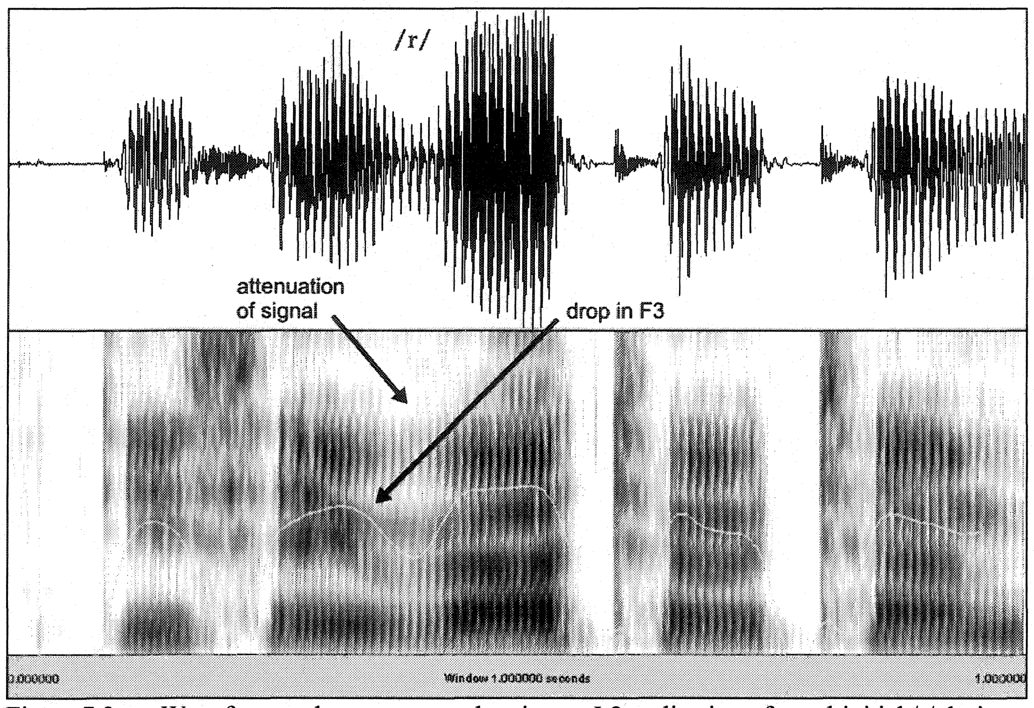

Figure 7.3 Waveform and spectrogram showing an L2 realization of word-initial /r/ designated as a Type 3 variant. Token [dise ɹato ta] in *Dice rato también*, Session 3, participant A-Adv-M1, scored as "0" (inaccurate).

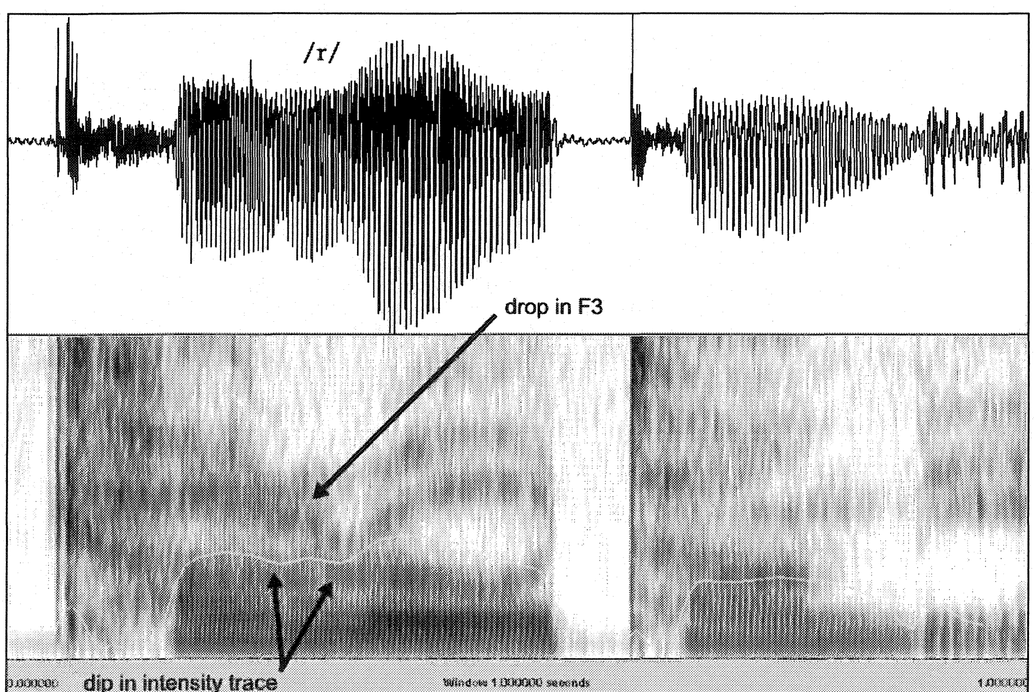

Figure 7.4 Waveform and spectrogram showing an L2 realization of word-medial /r/ designated as a Type 4 variant. Token aurally [kaɾo tambyen], visually [kaɹo tambyen], in *Ya vi carro también*, Session 1, participant B-Beg-F3, scored "0" (inaccurate).

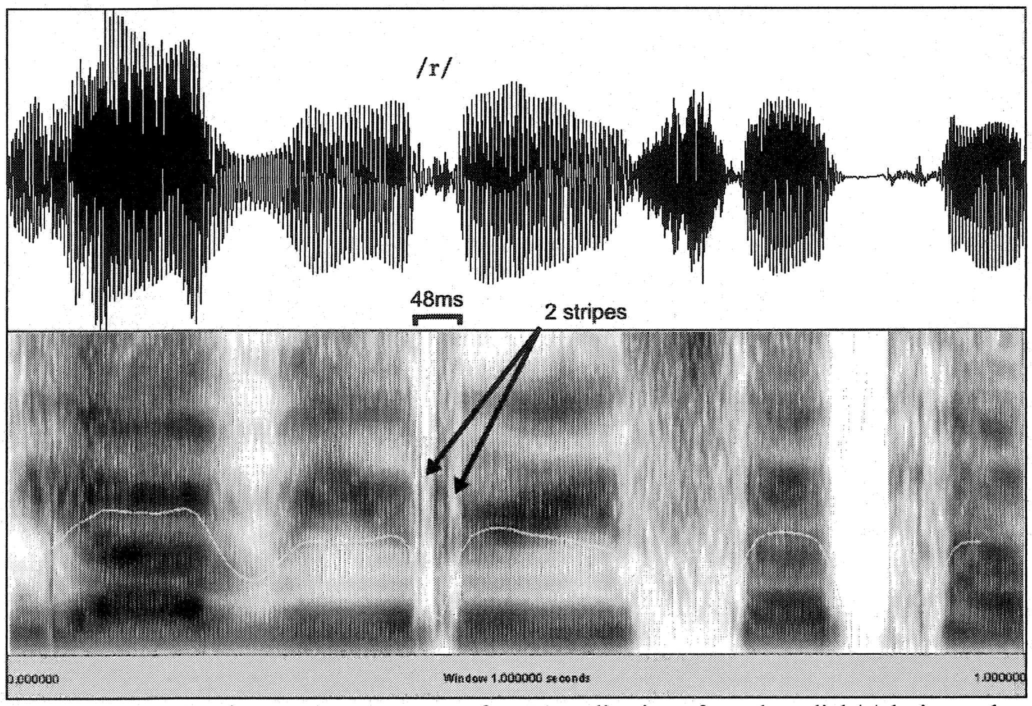

Figure 7.5 Waveform and spectrogram of an L2 realization of word-medial /r/ designated as a Type 5 variant. Token aurally [ya βi ɾisa ta], visually [ya βi ɾisa ta], in *Ya vi risa también*, Session 2, participant B-Adv-F6, scored "0" (inaccurate).

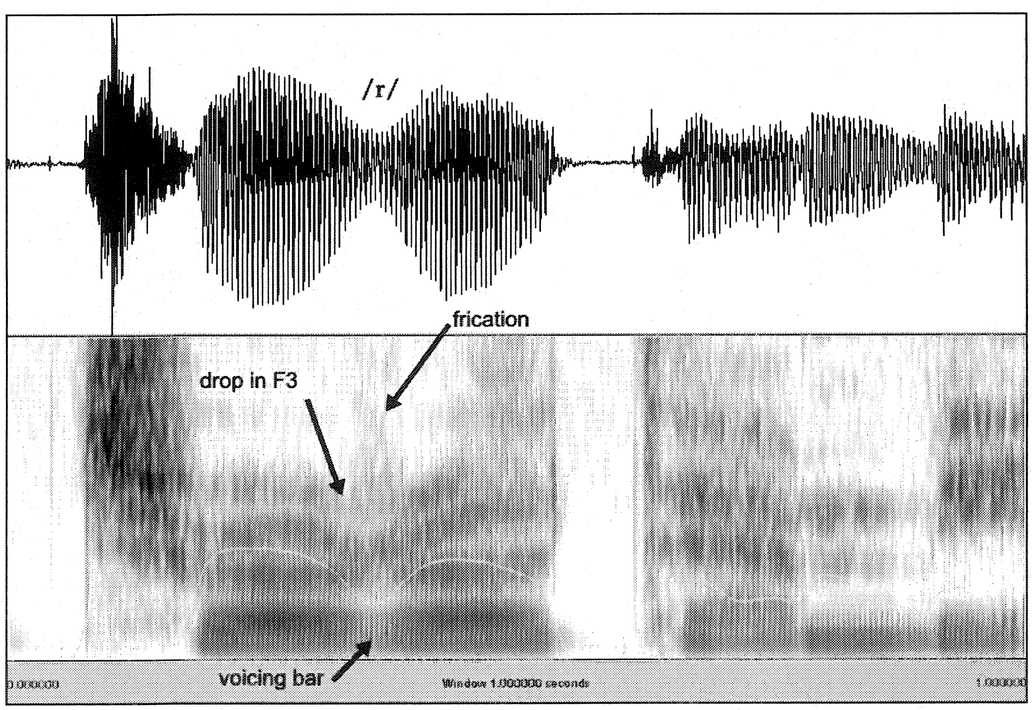

Figure 7.6 Waveform and spectrogram of an L2 realization of word-medial /r/ designated as a Type 6 variant. Token [tʃuɹo tambye] in *Dice churro también*, Session 1, participant A-Int-F1, scored "0" (inaccurate).

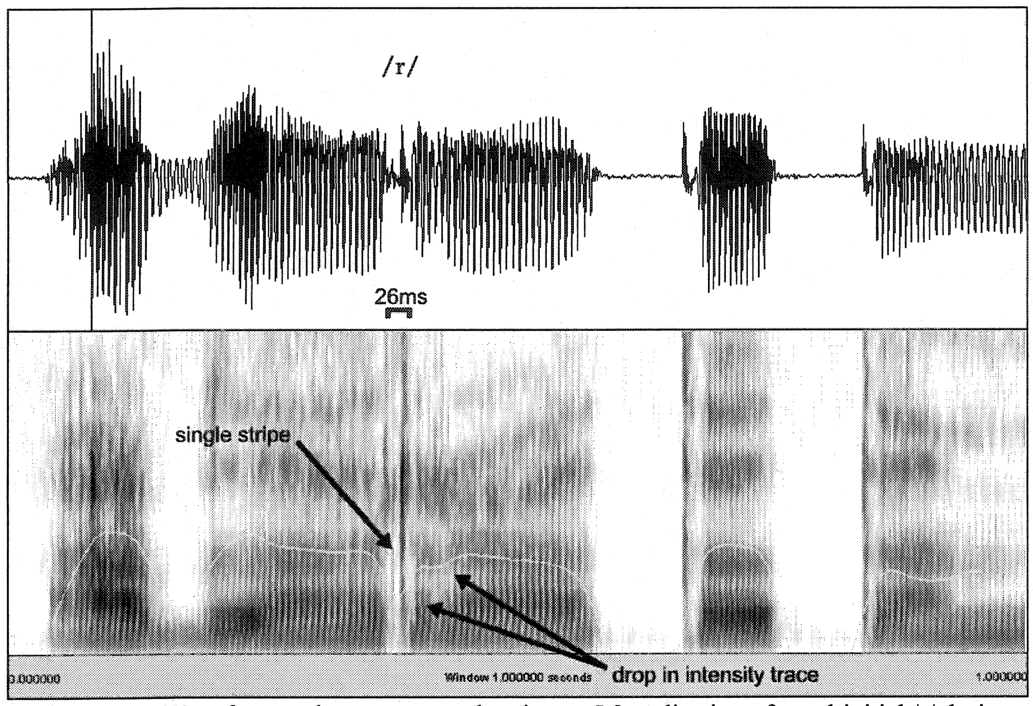

Figure 7.7 Waveform and spectrogram showing an L2 realization of word-initial /r/ designated as a Type 7 variant. Token aurally [aβla rato ta], visually [aβla ɾato ta], in *Habla rato también*, participant B-ExpL-F1, scored "0" (inaccurate).

Table 7.2 presents the frequency of word-initial /r/ variants by subject level. For example, at the advanced level for both experiments there were eight participants. Averaging across the frequencies of each type resulted in an overall level frequency for each type. Thus, the Type 2 variant was most common at the advanced level (41%) followed by the Type 3 variant (22%). Between levels, Type 1 variants accounted for a greater proportion of all variants at the experienced level while Type 3 variants accounted for a greater proportion of the realizations at the beginning and intermediate levels. Regarding an absence of variants, the intermediate level did not evidence any Type 5 or 7 variants; however, all other levels demonstrated all seven variant types. Appendix Q presents each participant's distribution of variants for word-initial /r/.

Table 7.2 Variant types of word-initial /r/ by subject level. Standard deviation is shown in parentheses. Only Type 1 forms were scored as "accurate."

Level	n	word-initial /r/ variant type						
		1 (≈[r])	2 (≈[ɾ])	3 (≈[ɹ])	4 (≈[ɾ]; ≈[ɹ])	5 (≈[ɾ]; ≈[r])	6 (≈[ɹ]; ≈[z̞])	7 (≈[r]; ≈[ɹ, ɾ])
Beginning	8	8% (17%)	19% (15%)	45% (36%)	9% (10%)	7% (14%)	8% (15%)	4% (8%)
Intermediate	5	0% (0%)	19% (34%)	76% (36%)	2% (3%)	0% (0%)	2% (3%)	0% (0%)
Advanced	10	20% (30%)	41% (36%)	22% (35%)	4% (8%)	5% (9%)	6% (14%)	2% (4%)
Experienced	11	59% (12%)	27% (13%)	2% (61%)	1% (6%)	3% (0%)	3% (6%)	5% (2%)

Table 7.3 shows the distribution of variant types for /r/ in word-medial position. In contrast to the distribution of types in word-initial position for beginning level participants, Type 2 (≈[ɾ]) variants accounted for the greatest proportion of realizations instead of Type 3 (≈[ɹ]) forms (60% of Type 2 (≈[ɾ]) vs. 45% of Type 3 (≈[ɹ]) forms). Within each level, the proportion of Type 1 (≈[r]) variants was higher word-medially than in word-initial position. For example, Type 1 (≈[r]) accounted for 20% of all realizations at the advanced level word-initially, but 51% word-medially. Regarding the absence of

variants, Types 3-6 were not evidenced word-medially at all by the experienced level participants. Type 6 (≈[ɹ]; ≈[z̧]) forms were not identified in any of the advance level participants while the Type 7 (≈[r]; ≈[ɹ, ɾ]) variant was not found in the intermediate level participant realizations. The distribution of word-medial /r/ types for each participant is located in Appendix R.

Table 7.3 Variant types of word-medial /r/ by subject level. Standard deviation is shown in parentheses. Only Type 1 forms were scored as "accurate."

Level	n	word-medial /r/ variant type						
		1 (≈[r])	2 (≈[ɾ])	3 (≈[ɹ])	4 (≈[ɾ]; ≈[ɹ])	5 (≈[ɾ]; ≈[r])	6 (≈[ɹ]; ≈[z̧])	7 (≈[r]; ≈[ɹ, ɾ])
Beginning	8	13% (23%)	60% (24%)	13% (16%)	8% (6%)	3% (6%)	3% (6%)	1% (2%)
Intermediate	5	18% (40%)	30% (37%)	45% (47%)	3% (3%)	3% (7%)	1% (2%)	0% (0%)
Advanced	10	51% (41%)	30% (33%)	10% (32%)	4% (10%)	2% (3%)	0% (0%)	2% (4%)
Experienced	11	82% (30%)	15% (30%)	0% (0%)	0% (0%)	0% (0%)	0% (0%)	3% (5%)

Table 7.4 Variant types of word-initial /r/ by gender. Standard deviation is shown in parentheses. Only Type 1 forms were scored as "accurate."

Gender	n	word-initial /r/ variant type						
		1 (≈[r])	2 (≈[ɾ])	3 (≈[ɹ])	4 (≈[ɾ]; ≈[ɹ])	5 (≈[ɾ]; ≈[r])	6 (≈[ɹ]; ≈[z̧])	7 (≈[r]; ≈[ɹ, ɾ])
Learner Males	6	0% (0%)	24% (29%)	70% (31%)	4% (6%)	0% (0%)	1% (3%)	0% (0%)
Learner Females	17	16% (26%)	30% (32%)	32% (38%)	6% (9%)	6% (12%)	7% (14%)	3% (6%)
Experienced Males	4	42% (37%)	36% (36%)	6% (11%)	3% (6%)	3% (6%)	6% (11%)	6% (6%)
Experienced Females	7	68% (34%)	22% (31%)	0% (0%)	0% (0%)	3% (5%)	2% (4%)	5% (6%)

Table 7.4 shows the variant types of word-initial /r/ by gender. Within levels, both the learner males and females demonstrated the greatest proportion of Type 3 (≈[ɹ]) variants while both experienced gender groups resulted in the largest proportion of Type 1 (≈[r]) variants. Between levels, both male groups showed higher frequencies of Type 3 (≈[ɹ]) forms than the frequency of Type 3 (≈[ɹ]) forms for the female counterparts. In contrast, the female groups showed higher frequencies of Type 1 (≈[r]) variants than Type 1 frequency for the male counterparts.

Table 7.5 sets out variant type distributions for word-medial targets by gender groups. For the learner groups, Type 2 (≈[ɾ])variants were more common in word-medial position in comparison with Type 2 (≈[ɾ]) frequencies in word-initial position. Thus, while Type 3 (≈[ɹ]) forms were most common for both learner gender groups in word-initial position, Type 2 (≈[ɾ]) forms were most common in word-medial position. For the experienced gender groups, frequencies of Type 2 (≈[ɾ]) forms were less common in word-medial position (25% and 10%) than frequencies of Type 2 (≈[ɾ]) variants in word-initial position (36% and 22%).

Table 7.5 Variant types of word-medial /r/ by gender. Standard deviation is shown in parentheses. Only Type 1 forms were scored as "accurate."

Gender	n	word-medial /r/ variant type						
		1 (≈[r])	2 (≈[ɾ])	3 (≈[ɹ])	4 (≈[ɾ]; ≈[ɹ])	5 (≈[ɾ]; ≈[r])	6 (≈[ɹ]; ≈[z̠])	7 (≈[r]; ≈[ɹ, ɾ])
Learner Males	6	15% (36%)	47% (37%)	27% (39%)	9% (13%)	2% (6%)	0% (0%)	0% (0%)
Learner Females	17	36% (39%)	38% (32%)	16% (32%)	4% (5%)	3% (4%)	1% (4%)	2% (3%)
Experienced Males	4	72% (48%)	25% (50%)	0% (0%)	0% (0%)	0% (0%)	0% (0%)	3% (6%)
Experienced Females	7	87% (13%)	10% (13%)	0% (0%)	0% (0%)	0% (0%)	0% (0%)	3% (5%)

In sum, this section has presented seven variant types that have been established to correspond to all L2 productions of word-medial and word-initial /r/. With increasing

Spanish level, Type 1 (≈[r]) variants accounted for an increased proportion of all realized forms. In addition, more Type 1 (≈[r]) variants were evidenced at the experienced level than at the learner levels. Regarding gender, males showed similar trends to females for /r/ realizations word-initially in that both groups most frequently produced the Type 3 (≈[ɹ]) variant, followed by the Type 2 (≈[ɾ]) form; however, while some learner females produced Type 1 (≈[r]) forms, none were produced by learner males. Regarding word-medial /r/, learner males evidenced more Type 3 (≈[ɹ]) forms after Type 2 (≈[ɾ]) forms while females showed more Type 1 (≈[r]) forms after Type 2 (≈[ɾ]) forms.

7.1.2 Measurement results of /r/ productions scored "accurate"

This section presents the acoustic measurement results of word-initial and word-medial /r/ tokens that were scored "1" (accurate). The measures presented here correspond to the Type 1 (≈[r]) variants described in Section 7.1.1. For both word-initial and word-medial categories, only those subjects that received a score of "1" for at least one /r/ production in the category are included in the results. For example, participant A-Adv-F1 received a score of "1" (accurate) for 18 out of 27 word-medial /r/ targets but none for word-initial /r/ targets. Thus, her average durations are only included below for the word-medial /r/ category.

Table 7.6 presents the average measurement results for word-medial trills across participant levels and gender. For example, seven participants at the advanced level demonstrated productions that were scored "1" (accurate). Averaging across each average duration yielded an overall duration of 73ms, 1.97 degree of voicing value, and 2.31 stripes on average. Appendix S presents the mean measures for all productions of word-medial /r/ tokens scored as accurate by participant. Table 7.7 similarly shows the mean measures for word-initial productions that were scored "1" (accurate). For example, overall trill duration calculated across 10 participants from the experienced level was 74ms, 1.18 degree of voicing value, and 2.29 number of stripes. Appendix T shows the mean measures for all accurately produced word-initial /r/ tokens by participant.

Table 7.6 Mean measures for word-medial /r/ productions scored "1" by level. Standard deviation is shown in parentheses. The absence of standard deviation is indicated by "na" since standard deviation requires more than one sample.

Level	n	Mean measurements of word-medial /r/		
		Overall duration in ms	Degree of voicing	Number of stripes
Beginning	3	62 (9)	1.56 (.51)	2.26 (.45)
Intermediate	1	75 (na)	2.00 (na)	2.08 (na)
Advanced	7	73 (16)	1.97 (.05)	2.31 (.26)
Experienced	10	73 (12)	1.92 (.19)	2.30 (.38)
Native Speakers	11	76 (18)	1.7 (.6)	2.2 (.8)

Table 7.7 Mean measures for word-initial /r/ productions scored "1" by level. Standard deviation is shown in parentheses. The absence of standard deviation is indicated by "na" since standard deviation requires more than one sample.

Level	n	Mean measurements of word-initial /r/		
		Overall duration (ms)	Degree of voicing	Number of stripes
Beginning	3	64 (4)	1.45 (.43)	2.29 (.26)
Intermediate	0	na	na	na
Advanced	5	71 (15)	1.95 (.12)	2.28 (.36)
Experienced	10	74 (11)	1.81 (.30)	2.29 (.40)
Native Speakers	11	88 (23)	1.6 (.7)	2.5 (1.2)

Table 7.8 and Table 7.9 show the average measures by gender groups for word-medial and word-initial /r/, respectively. For word-medial measures, averaging across 10 learner females yielded 70ms duration, 1.85 degree of voicing value, and 2.30 number of stripes. Experienced males demonstrated the longest overall duration (77ms); females the shortest (70ms). Regarding word-initial measures, averaging across eight learner females yielded a mean duration of 68ms while experienced males demonstrated the longest

overall durations of 76ms. No results are provided for learner males as no male participants from the beginning, intermediate, or advanced groups evidenced any word-initial realizations that were scored "1" (accurate).

Table 7.8 Mean measures by gender for word-medial /r/ productions scored "1". Standard deviation is shown in parentheses. The term "na" refers to not applicable since standard deviation requires more than one sample.

Gender	n	Mean measurements of word-medial /r/		
		Overall duration in ms	Degree of voicing	Number of stripes
Learner Males	1	75 (na)	2.00 (na)	2.08 (na)
Leaner Females	10	70 (15)	1.85 (.32)	2.30 (.30)
Experienced Males	3	77 (5)	1.96 (.06)	2.27 (.28)
Experienced Females	7	72 (14)	1.90 (.22)	2.31 (.44)

Table 7.9 Mean measures by gender for word-initial /r/ productions scored "1". Standard deviation is shown in parentheses. The term "na" refers to not applicable since there were no learner males with tokens scored as accurate.

Gender	n	Mean measurements of word-initial /r/		
		Overall duration in ms	Degree of voicing	Number of stripes
Learner Males	0	na	na	na
Leaner Females	8	68 (12)	1.76 (.36)	2.28 (.31)
Experienced Males	3	76 (2)	1.83 (.29)	2.14 (.25)
Experienced Females	7	73 (14)	1.79 (.32)	2.35 (.45)

In sum, an analysis of the measures for productions scored "1" (accurate) revealed a range of measures for duration, degree of voicing, and number of stripes. In the current study, these measures are presented to quantitatively describe the acoustic features of L2 productions; however, parametric inferential comparisons are not provided due to the small sample size between groups and the extreme variance between group scores. In the next section, accuracy scores for word-initial vs. word-medial /r/ are compared in order to address the effect of word position on /r/ accuracy.

7.2 Comparison of word-medial with word-initial /r/ accuracy scores

7.2.1 Rankings of adjusted overall accuracy scores for L2 /r/

This section addresses Question 2 of this study: *Does trill accuracy emerge with higher scores word-medially over word-initial trill scores?* Thus, this section compares the word-initial scores for /r/ with the word-medial accuracy scores for /r/ by participant, level, and gender. Recall that adjusted overall accuracy scores were calculated over 27 attempted targets for participants in Experiment A. For the beginning and advanced levels in Experiment B, scores were based on 18 attempted targets. The experienced participant scores in Experiment B reflect a total of 9 attempted targets. Appendix D presents the individual scores by session alongside each participant's overall raw accuracy. In the following tables, "learner" group refers to the beginning, intermediate, and advanced Spanish levels. The "experienced" group refers to the experienced Spanish level participants.

Table 7.10 compares the word-initial /r/ and word-medial /r/ adjusted overall accuracy scores for each participant in the learner group. For example, participant A-Adv-F2 revealed an adjusted overall accuracy score of 20% for word-initial /r/ and 96% accuracy for word-medial /r/. Effectively, 11 out of 23 participants yielded the ranking such that word-initial scores were lower than word-medial scores. A permutation test on this ranking yielded a significant pattern ($m=.49$; $p<.001$).

Table 7.10 Comparison of word-initial vs. word-medial adjusted accuracy scores for all beginning, intermediate, and advanced levels.

	Participant	Adjusted overall target accuracy	
		initial /r/	medial /r/
Experiment A	A-Beg-M1	0%	0%
	A-Beg-M2	0%	0%
	A-Beg-F1	53%	75%
	A-Beg-F2	0%	0%
	A-Int-M1	0%	0%
	A-Int-M2	0%	100%
	A-Int-F1	0%	0%
	A-Int-F2	0%	0%
	A-Int-F3	0%	0%
	A-Adv-M1	0%	0%
	A-Adv-M2	0%	0%
	A-Adv-F1	0%	75%
	A-Adv-F2	20%	96%
Experiment B	B-Beg-F1	7%	19%
	B-Beg-F2	0%	0%
	B-Beg-F3	0%	0%
	B-Beg-F4	13%	19%
	B-Adv-F1	0%	0%
	B-Adv-F2	39%	94%
	B-Adv-F3	71%	87%
	B-Adv-F4	97%	100%
	B-Adv-F5	0%	25%
	B-Adv-F6	6%	88%

In Table 7.11, the adjusted overall scores are presented for experienced level participants. For example, participant B-ExpL-M1 received a score of 26% for word-initial /r/ and 100% for word-medial /r/. Out of 11 participants, 7 revealed the ranking of word-initial /r/ scores lower than scores for word-medial /r/. A permutation test found this ranking to be significant at the .001 alpha level (m=.64).

Table 7.11 Adjusted overall word-medial target for experienced level participants.

	Participant	Adjusted overall target accuracy	
		initial /r/	medial /r/
Experiment B	B-ExpL-M1	26%	100%
	B-ExpL-M2	90%	100%
	B-ExpL-M3	0%	0%
	B-ExpL-M4	78%	100%
	B-ExpL-F1	78%	87%
	B-ExpL-F2	90%	100%
	B-ExpL-F3	100%	100%
	B-ExpL-F4	100%	100%
	B-ExpL-F5	100%	100%
	B-ExpL-F6	26%	87%
	B-ExpL-F7	26%	74%

Table 7.12 shows the mean adjusted accuracy scores by level when averaging across participant overall scores. For instance, the four advanced level participants in Experiment A yielded a mean overall accuracy of 5% and 43% for word-initial and word-medial /r/, respectively. Within levels, scores for word-medial /r/ were systematically higher than scores for word-initial /r/. Between levels, scores for word-initial /r/ were highest for the experienced group (65%) and lowest for the intermediate group (0%); however, the beginning group demonstrated very low scores as well (7%). Regarding word-medial scores, there was a robust trend between participant level and mean accuracy. Scores for medial /r/ were lowest for the beginning level (12%) and highest for participants at the experienced level (86%).

Table 7.12 Mean adjusted overall accuracy scores by level. Standard deviation is shown in parentheses.

Level	n	Mean adjusted overall accuracy score	
		initial /r/	medial /r/
Beginning	8	7% (17%)	12% (24%)
Intermediate	5	0% (0%)	20% (45%)
Advanced	10	25% (35%)	57% (44%)
Experienced	11	65% (38%)	86% (30%)

Table 7.13 shows the adjusted overall accuracy scores for word-initial and word-medial /r/ by gender groups. For example, there were six learner males and averaging across their overall accuracy scores yielded a mean of 0% for word-initial /r/ and 17% for word-medial /r/. Within each level, scores for word-initial /r/ were lower than scores for word-medial /r/. Between levels, scores for females were consistently higher than their male counterparts.

Table 7.13 Mean adjusted overall accuracy scores for /r/ by word position and gender. Standard deviation is shown in parentheses.

Gender	n	Mean adjusted overall accuracy score	
		initial /r/	medial /r/
Learner Males	6	0% (0%)	17% (41%)
Learner Females	17	18% (29%)	40% (42%)
Experienced Males	4	48% (43%)	75% (50%)
Experienced Females	7	74% (34%)	93% (10%)

In sum, this section has presented the results of the investigation of word-initial vs. word-medial adjusted overall accuracy scores by participant, level, and gender. Inferential analysis of rankings for participants revealed a significant ranking order such that word-initial accuracy scores were lower than word-medial scores. This same relationship was revealed by groupings for level and gender. Regarding participant experience, there was a robust trend for increased scores alongside increased participant Spanish level. For gender, females demonstrated higher scores than males on average.

7.2.2 The role of vowel height in productions of /r/ scored "accurate"

The purpose of this section is to address Question 5 of this study: *Is there an effect of vowel height on trills that were scored "1" (accurate)?* This question is based on research that found that the aerodynamic requirements for trill production were greater in a high vowel environment (Solé 2002). Thus, raw accuracy for both word-medial and word-initial productions is combined in order to provide more meaningful comparison of scores through sufficient sample sizes. The data presented here reflect only those participants that scored "1" (accurate) for at least one word-initial /r/ or one word-medial /r/ production.

Table 7.14 presents participant overall raw /r/ accuracy scores by vowel environment. In the design for Experiment A, vowel height was varied word-medially but not word-initially. Therefore, scores reported for Experiment A participants are for word-medial targets only. For example, participant A-Adv-F1 attempted 27 word-medial /r/ targets across three sessions; 9 attempted targets for each of the three vowel height categories. Hence, her score of 78% for high vowel environment accuracy reflects 7/9 realizations scored "1" (accurate). In Experiment B, vowel height was varied in both word positions. Therefore, vowel category scores represent both word-initial and word-medial targets. For instance, participant B-Beg-F1 was recorded in 2 sessions in Experiment B. Across both sessions, she attempted 6 word-initial and 6 word-medial targets for each vowel environment. Thus, her score of 8% for accuracy in a high vowel environment reflects 1 out of 12 productions scored “1” (accurate). Across all L2 participants, 18 out of 23 subjects revealed the ranking such that scores for targets in a low vowel environment were higher than scores for /r/ in a high vowel environment.

A permutation test, as described in Section 4.5, revealed a significant effect of vowel height for L2 scores ($q=-20.8$; $p<.001$). This effect reflects significant differences between the high and mid vowel environment scores ($q=-17.0$; $p<.001$). No difference was found between low and mid vowel environment scores, although the average accuracy across all L2 participants was slightly higher in low vs. mid vowel environments (70% vs. 66%, respectively). A second permutation test included the native speakers, in order to test whether the effect was also present in native speaker production. There was also a significant effect of vowel height ($q=-17.62$; $p<.001$) for native speakers. High vowel scores and mid vowel scores differed maximally($q=-12.62$; $p<.01$); however, no significant effect was found between the mid and low vowel environment scores. A third permutation test included only the native speaker scores and did not show a significant effect (with $\alpha = .05$) between scores and any of the three vowel heights; however, averaging across native speaker scores for each vowel height category did reveal a trend such that scores were higher with lower vowel height.

Table 7.14 Overall raw accuracy score for word-initial and word-medial /r/ by vowel environment.

	Participant	Overall raw accuracy score		
		High vowel	Mid vowel	Low vowel
Exp. A	A-Beg-F1	78%	56%	67%
	A-Beg-F2	0%	0%	11%
	A-Int-M2	100%	78%	89%
	A-Adv-F1	33%	89%	78%
	A-Adv-F2	89%	78%	89%
Experiment B	B-Beg-F1	8%	8%	25%
	B-Beg-F4	0%	33%	17%
	B-Adv-F1	0%	0%	17%
	B-Adv-F2	33%	67%	83%
	B-Adv-F3	50%	75%	83%
	B-Adv-F4	75%	100%	100%
	B-Adv-F5	0%	25%	17%
	B-Adv-F6	25%	75%	50%
	B-ExpL-F1	67%	67%	83%
	B-ExpL-F2	50%	100%	100%
	B-ExpL-F3	100%	100%	100%
	B-ExpL-F4	83%	100%	100%
	B-ExpL-F5	100%	100%	100%
	B-ExpL-F6	67%	83%	83%
	B-ExpL-F7	67%	83%	67%
	B-ExpL-M1	33%	67%	83%
	B-ExpL-M2	67%	100%	100%
	B-ExpL-M3	0%	0%	33%
	B-ExpL-M4	50%	100%	100%
Native speakers	NS-F1	83%	100%	100%
	NS-F2	100%	83%	67%
	NS-F3	17%	83%	100%
	NS-F4	83%	100%	100%
	NS-F5	100%	100%	100%
	NS-F6	67%	100%	100%
	NS-M1	100%	100%	100%
	NS-M2	100%	100%	100%
	NS-M3	100%	50%	100%
	NS-M4	100%	67%	100%
	NS-M5	100%	100%	100%

Table 7.15 presents the raw accuracy scores by participant level. For example, across all beginning level participants there were four subjects who demonstrated any /r/ accuracy in any of the vowel environments. Averaging across their mean scores for each vowel category yielded an overall level accuracy of 22%, 24%, and 30% for /r/ in high, mid, and low environments, respectively. Within each level, scores for /r/ were lowest in high vowel environments, higher in mid vowel environments, and then highest in low vowel environments. Between levels, scores in each vowel category increased with increased Spanish level (with the exception of the 1 participant from the intermediate level).

Table 7.15 Overall raw accuracy score for word-initial and word-medial /r/ by vowel environment. Standard deviation is shown in parentheses.

Level	n	Overall raw accuracy score		
		High vowel	Mid vowel	Low vowel
Beginning	4	22% (38%)	24% (25%)	30% (25%)
Intermediate	1	100% (na)	78% (na)	89% (na)
Advanced	8	38% (32%)	64% (34%)	65% (33%)
Experienced	11	62% (29%)	82% (30%)	86% (21%)
Native Speakers	11	86% (26%)	89% (17%)	97% (10%)

Vowel environment accuracy scores by gender are presented in Table 7.16. For instance, averaging across males at the experienced level yielded a mean accuracy of 38% for high vowel environment, 67% for mid vowel environment, and 79% for low vowel environment. Within each level, environment scores were ranked (from lowest to highest) such that high vowel < mid vowel < low vowel, with the exception of native speaker males who demonstrated lowest trill emergence in mid vowel environments. Between groups, no gender patterns were observed. The experienced female group showed higher scores than males in each category; however, the native speaker males demonstrated higher trill emergence in the high and low categories in comparison to female emergence scores.

Table 7.16 Overall raw accuracy /r/ scores for each vowel environment by gender group. Standard deviation is shown in parentheses.

Gender	n	Mean overall raw accuracy score		
		High vowel	Mid vowel	Low vowel
Learner Males	1	100% (na)	78% (na)	89% (na)
Learner Females	12	33% (33%)	50% (36%)	53% (34%)
Experienced Males	4	38% (28%)	67% (47%)	79% (32%)
Experienced Females	7	76% (19%)	90% (13%)	90% (13%)
Native speaker Males	6	100% (0%)	83% (24%)	100% (0%)
Native speaker Females	5	75% (31%)	94% (9%)	94% (14%)

In sum, the inferential analysis of scores for each participant revealed a significant effect of vowel height with a majority of the effect due to the high vowel environment. A descriptive analysis of vowel category scores across level and gender also revealed (for L2 speakers) a strong trend between vowel categories such that scores for /r/ in high vowel environments were lowest, followed by mid vowel scores, and then low vowel scores which were highest. The effect of vowel height was also found when native speakers were included alongside the L2 groups; however, no significant effect of vowel height was found when native speakers were analyzed alone.

8. Taps results: L2 realizations

The primary purpose of this chapter is to describe the range of auditory and acoustic productions of L2 word-medial /ɾ/ and onset cluster /ɾ/. Variant types were established based on the combination of auditory and acoustic features that had been noted for each production. The variants for both word-medial and onset cluster /ɾ/ are presented together, as the same forms were found for both categories; however, example waveforms and spectrograms reflect each type in both word positions. Following the description of variant types, the acoustic measures for productions scored "1" (accurate) are then presented. The comparison of word-medial /ɾ/ with word-medial /d, t, r/ can be found in Section 6.2.1.

The secondary purpose of this chapter is to address the two remaining questions of this study. Question 3 asked: *Does tap accuracy emerge with higher scores word-medially over scores of taps in onset clusters?* Thus, a comparison is made of both /ɾ/ scores for each participant. Question 6 asked: *Is the emergence of the svarabhakti vowel limited to highly experienced L2 speakers of Spanish?* That is, are experienced L2 learners most likely to evidence the svarabhakti vowel in their productions over less experienced L2 learners?

The organization of this chapter is as follows: Section 8.1 describes the L2 /ɾ/ productions. Section 8.1.1 establishes the five variant types which were established for /ɾ/ in word-medial and onset cluster positions. In Section 8.1.2, the overall duration for word-medial productions scored "1" (accurate) are detailed while the measures for onset cluster /ɾ/ tokens scored as accurate are presented in Section 8.1.3. Section 8.2 provides a comparison of /ɾ/ accuracy scores. In Section 8.2.1, Question 3 of this study addresses the analysis of accuracy scores by word position. Section 8.2.2 relates to Question 6 of this study by comparing the emergence of svarabhakti scores across participant levels.

8.1 Description: L2 productions of /ɾ/

8.1.1 Description of L2 /ɾ/: Variants types

A total of 1260 productions of /ɾ/ were analyzed across Experiment A and B; 630 for onset cluster /ɾ/ and 630 for word-medial /ɾ/. For each production, detection of the number of successive oral tract obstructions was the primary focus for the auditory analysis, while the number of visible stripes and stripe duration was targeted in the inspection of the spectrogram. The various combinations of these notations led to the establishment of five variant types for /ɾ/ productions which are presented in Table 8.1. For each variant type, a corresponding IPA symbol is provided to more easily allow the reader to associate a type with its auditory and/or acoustic features.

Table 8.1 Variants of word-medial and onset cluster /ɾ/. Frequency is based on 1260 productions.

/ɾ/ variant	Similar IPA sound (Aud./ Acou.)	n	Frequency	Auditory impression - number of obstructions	Visible number of stripes	Visible attenuation of spectral signal <30ms	Score
Type 1	[ɾ]	537	42.6%	1	1	yes	"1"
Type 2	[ɹ]	498	39.5%	0	0	no	"0"
Type 3	[ɾ]/[ɹ]	120	9.5%	1	0	no	"0"
Type 4	[r]/[z̧,ɾ,r]	61	4.8%	2+	0+	yes/no	"0"
Type 5	[ɾ]/[d,ð]	44	3.5%	1	1	no	"0"

Type 1 (≈[ɾ]) variants were scored "1" (accurate) and represented the auditory feature of an impression of a single oral tract obstruction and, acoustically, a single visible stripe in the spectrogram (with duration < 30ms). These instantiations accounted for 42.6% of all /ɾ/ realizations produced across all four levels of L2 participants. Figure 8.1 shows a sample waveform and spectrogram for a word-medial Type 1 variant while Figure 8.2 and Figure 8.3 demonstrate productions in onset cluster position from a 1s and 250ms window, respectively.

Type 2 (≈[ɹ]) forms accounted for 39.5% of all productions and were scored "0" (inaccurate). The auditory analysis revealed lack of oral tract obstruction. From a visual inspection, no stripes or measurable attenuation were present in the spectrogram. Auditorily and acoustically these instantiations were similar to productions of the L1 English retroflex approximant [ɹ]. A word-medial Type 2 variant is shown in Figure 8.4. Figure 8.5 demonstrates a Type 2 onset cluster realization from a 1s analysis window and the same production is shown in Figure 8.6 from a 250ms window.

Type 3 (≈[ɾ]; ≈[ɹ]) forms accounted for 9.5% of all realizations and gave the auditory impression of a single obstruction in the oral tract; however, the visual inspection did not reveal a stripe or any measurable attenuation. Thus, Type 3 instantiations were scored "0" (inaccurate). Figure 8.7 presents a sample waveform and spectrogram of a word-medial Type 3 instantiation from a 1s window. A Type 3 instantiation in onset cluster is shown in Figure 8.8 and Figure 8.9.

Type 4 (≈[r]; ≈[ʐ,ɾ,r]) instantiations made up 4.8% of all /ɾ/ productions and were scored "0" (inaccurate). The auditory analysis gave the impression of two or more rapid obstructions in the oral tract. The visual inspection yielded a range from attenuation with frication to one or more stripes in the spectrogram. In some instantiations, attenuation was less than 30ms duration, and in others, greater than 30ms duration. Thus, the production auditorily gave a broad impression of a trill, while the acoustic signal was more characteristic of an apicoalveolar fricative retroflex, alveolar tap, or alveolar trill. In Figure 8.10, a waveform and spectrogram show a Type 4 production found in word-medial position. An instantiation produced in onset cluster is shown from a 1s window in Figure 8.11 and from a 250ms window in Figure 8.12.

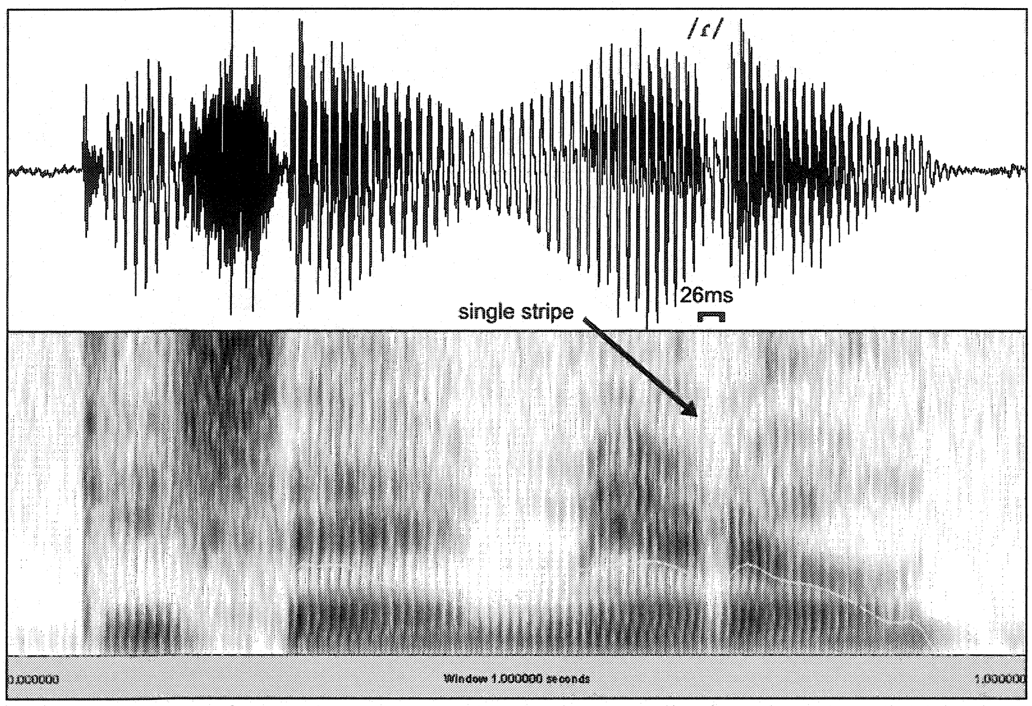

Figure 8.1 Waveform and spectrogram showing an L2 realization of word-medial /ɾ/ designated as a Type 1 variant. Token [dise miɾo] in *Dice miro también,* Session 2, participant A-Beg-M1, scored as "1" (accurate).

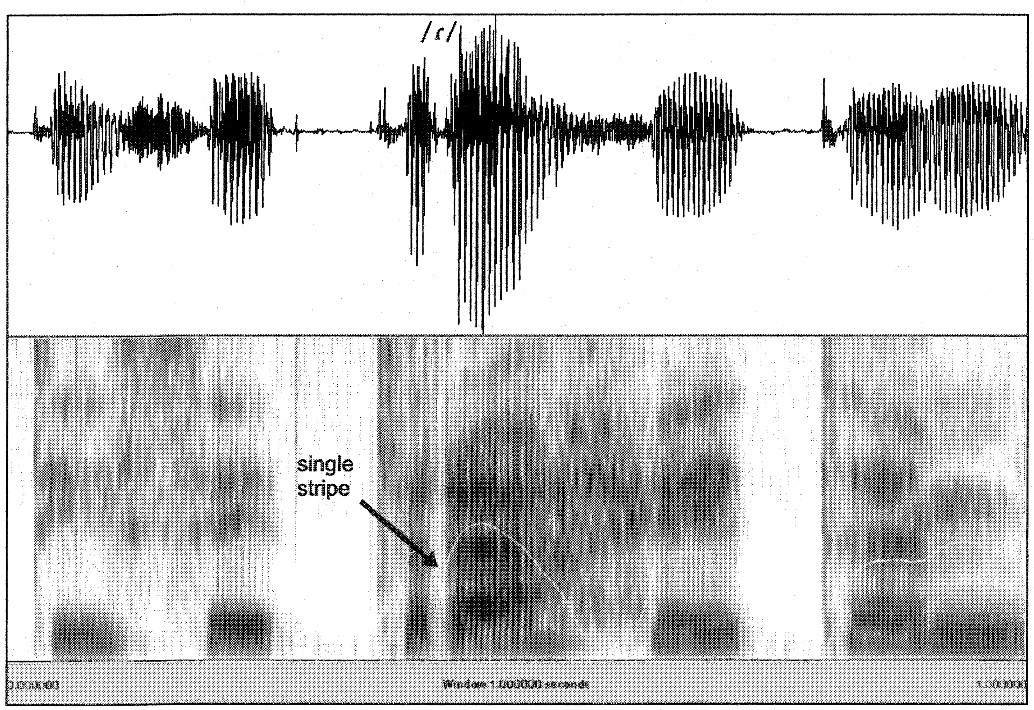

Figure 8.2 Waveform and spectrogram showing (from a 1s window) an L2 realization of onset cluster /ɾ/ designated as a Type 1 variant. Token [dise tɾaxe tam] in *Dice traje también*, Session 2, participant A-Adv-F2, scored as "1" (accurate).

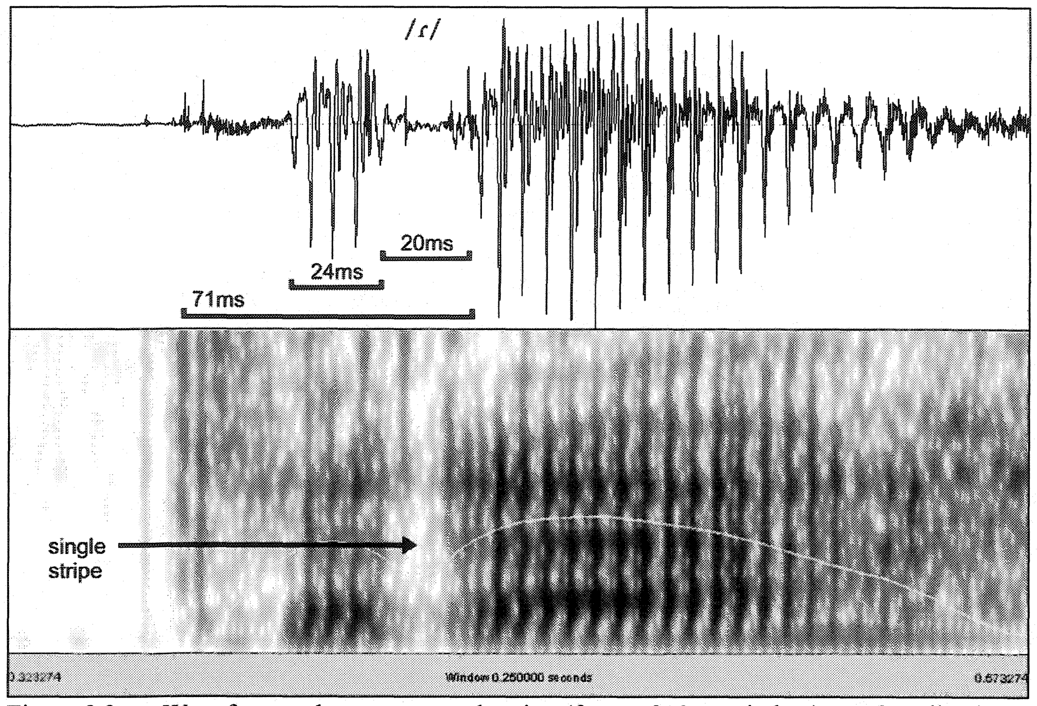

Figure 8.3 Waveform and spectrogram showing (from a 250ms window) an L2 realization of onset cluster /ɾ/ designated as a Type 1 variant. Token [dise tɾaxe tam] in *Dice traje también*, Session 2, participant A-Adv-F2, scored as "1" (accurate).

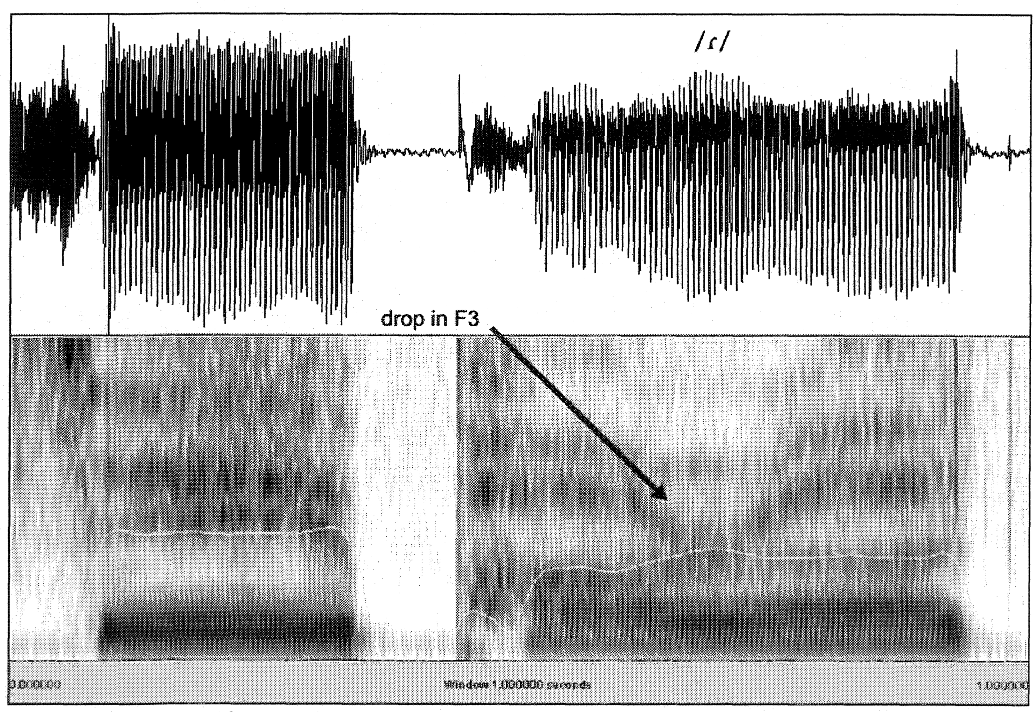

Figure 8.4 Waveform and spectrogram showing (from a 1s window) an L2 realization of word-medial /ɾ/ designated as a Type 2 variant. Token [se paɹa] in *Dice para también*, Session 1, participant B-Beg-F3, scored "0" (inaccurate).

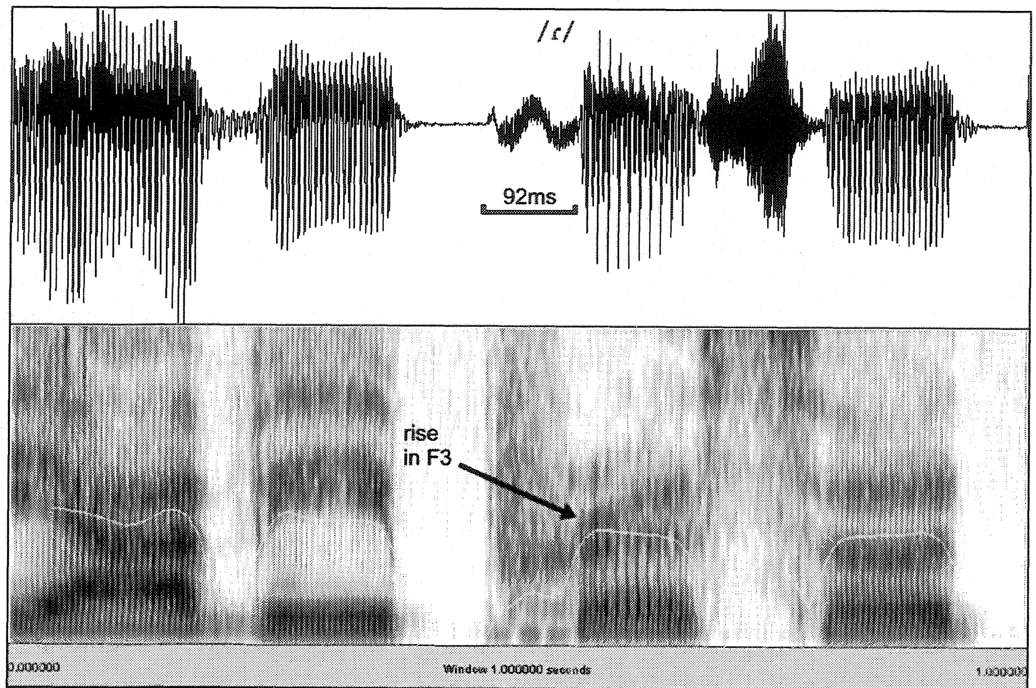

Figure 8.5 Waveform and spectrogram showing (from a 1s window) an L2 realization of onset cluster /ɾ/ designated as a Type 2 variant. Token [ya βi pɹɛso] in *Ya vi preso también*, Session 1, participant B-Beg-F2, scored "0" (inaccurate).

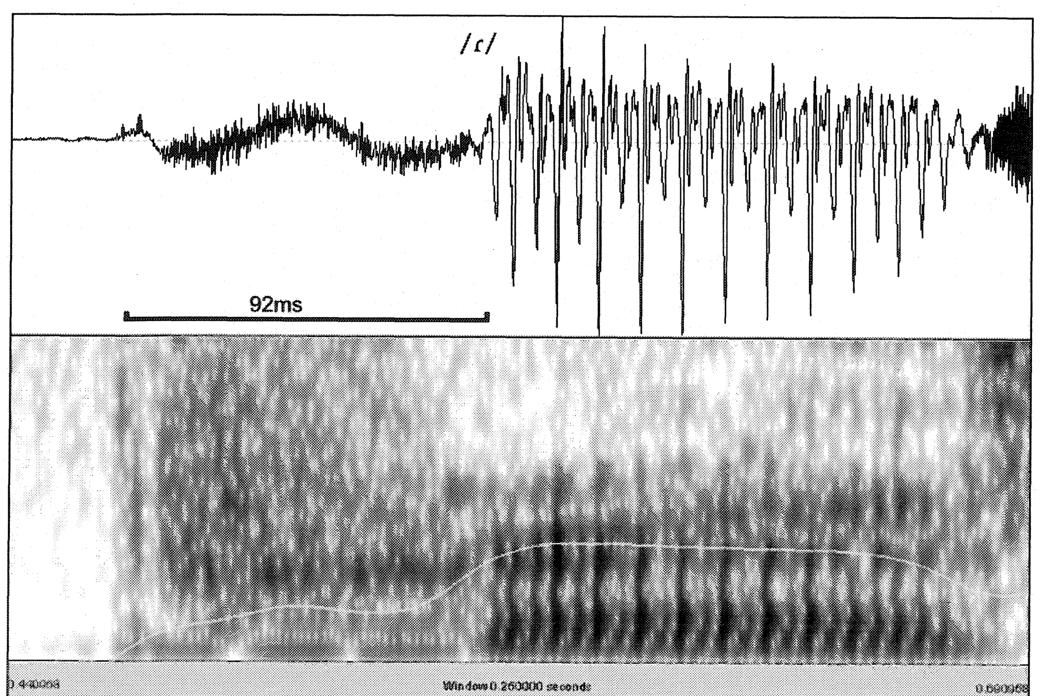

Figure 8.6 Waveform and spectrogram showing (from a 250ms window) an L2 realization of onset cluster /ɾ/ designated as a Type 2 variant. Token [pɹɛ] in *Ya vi preso también*, Session 1, participant B-Beg-F2, scored "0" (inaccurate).

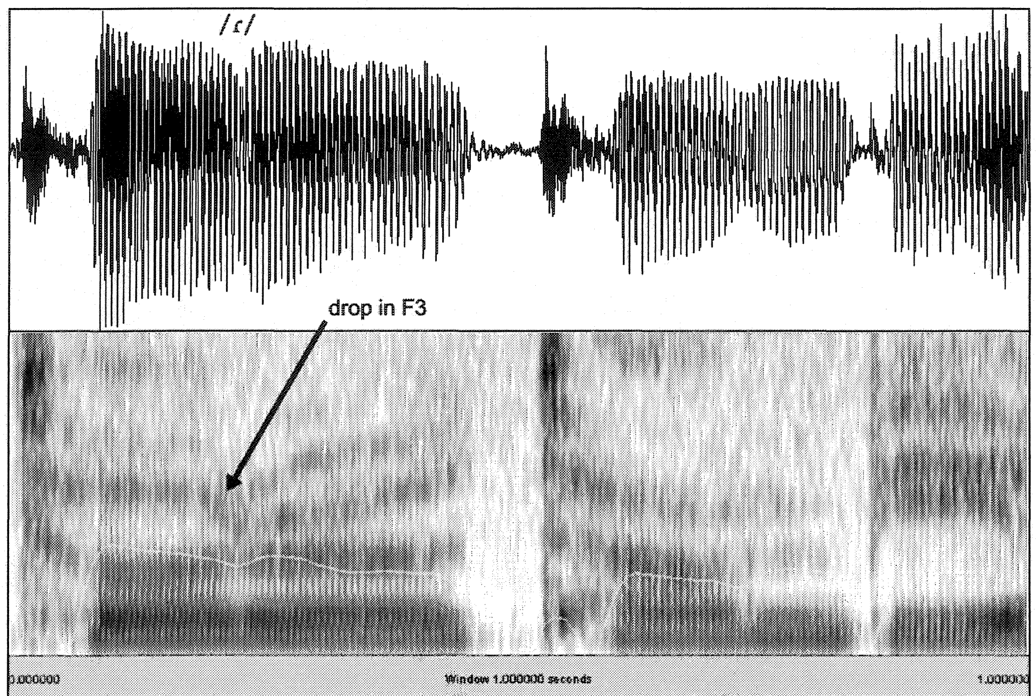

Figure 8.7 Waveform and spectrogram showing an L2 realization of word-medial /ɾ/ designated as a Type 3 variant. Token aurally [toɾo tambye], visually [toɹo tambye] in *Dice toro también*, Session 3, participant A-Beg-F2, scored "0" (inaccurate).

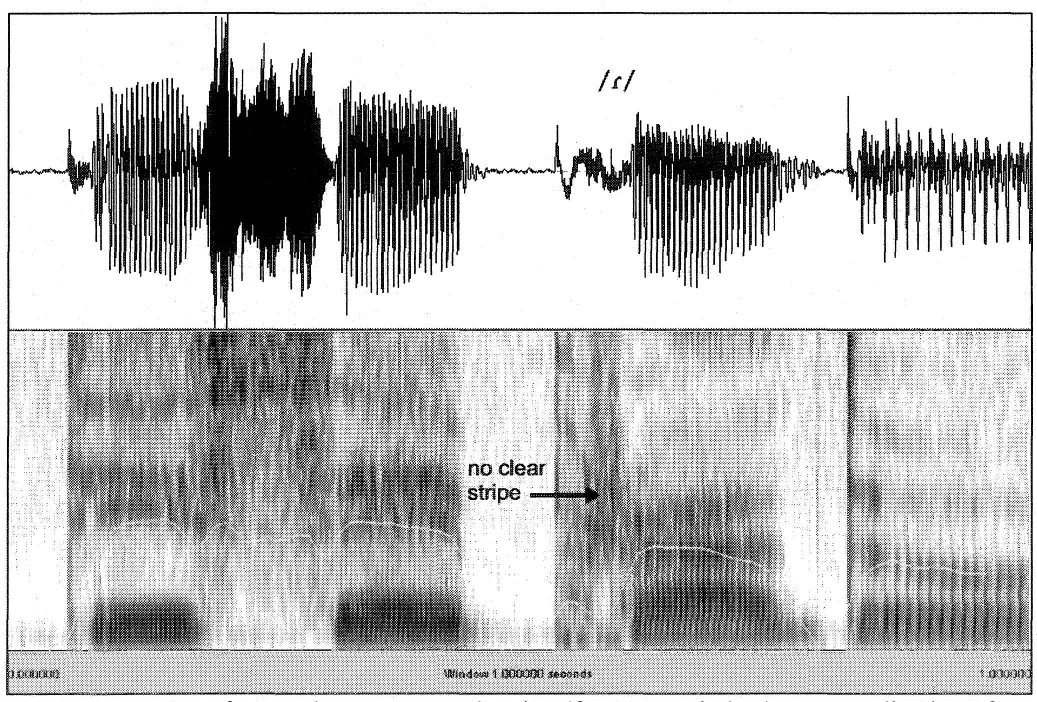

Figure 8.8 Waveform and spectrogram showing (from a 1s window) an L2 realization of onset cluster /ɾ/ designated as a Type 3 variant. Token aurally [dise pɾado], visually [dise pɹado] in *Dice prado también*, Session 2, participant B-Beg-F2, scored "0" (inaccurate).

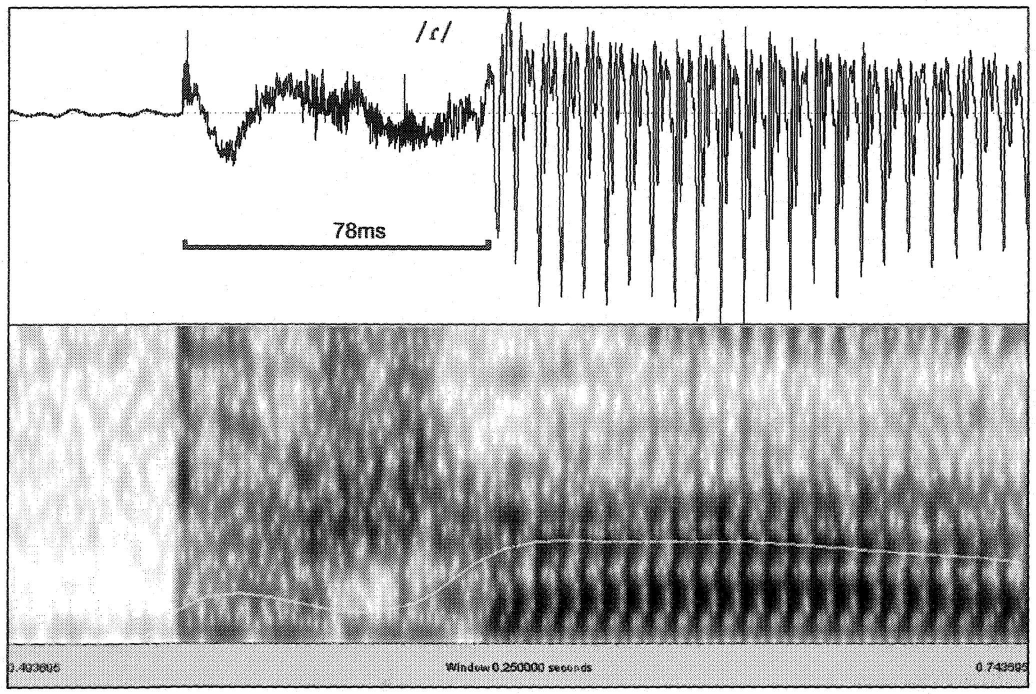

Figure 8.9 Waveform and spectrogram showing (from a 250ms window) an L2 realization of onset cluster /ɾ/ designated as a Type 3 variant. Token aurally [pɾa], visually [pɹa] in *Dice prado también*, Session 2, participant B-Beg-F2, scored "0" (inaccurate).

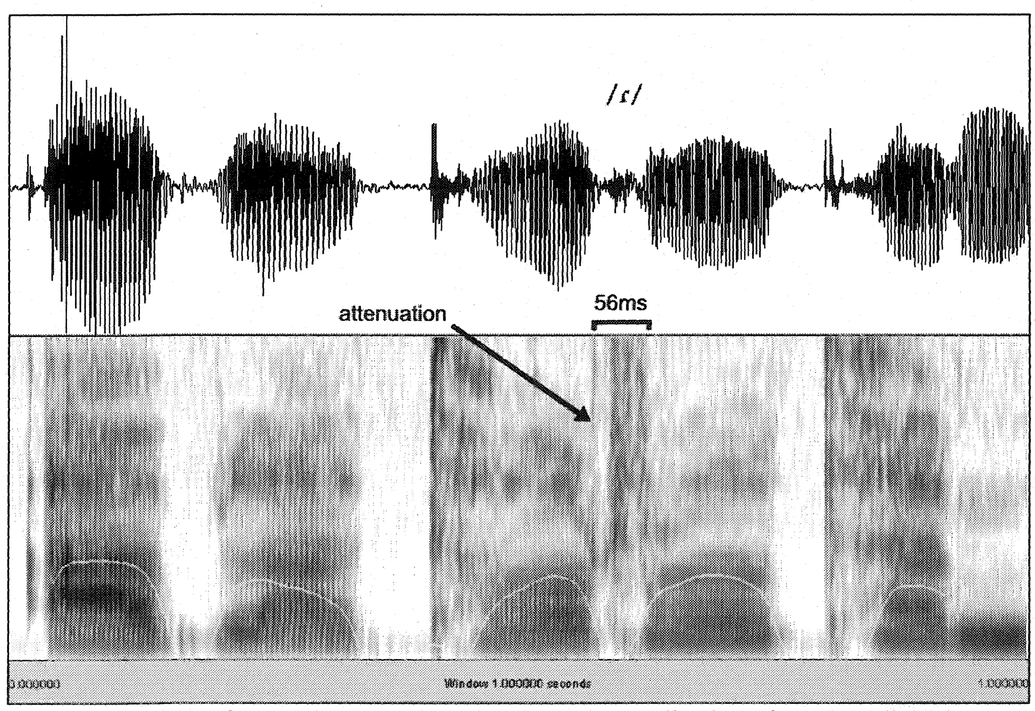

Figure 8.10 Waveform and spectrogram showing an L2 realization of word-medial /ɾ/ designated as a Type 4 variant. Token aurally [aβla toɾo] in *Habla toro también*, Session 2, participant B-Beg-F1, scored "0" (inaccurate).

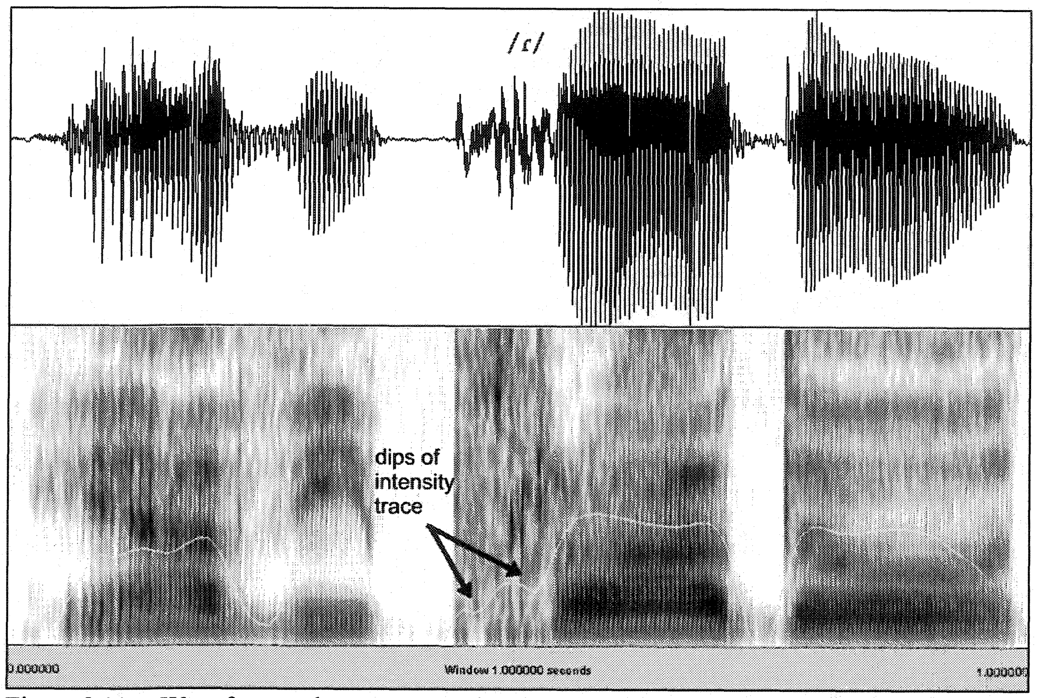

Figure 8.11 Waveform and spectrogram showing (from a 1s window) an L2 realization of onset cluster /ɾ/ designated as a Type 4 variant. Token aurally [ya βi prado] in *Ya vi prado también*, Session 1, participant B-Beg-F4, scored “0” (inaccurate).

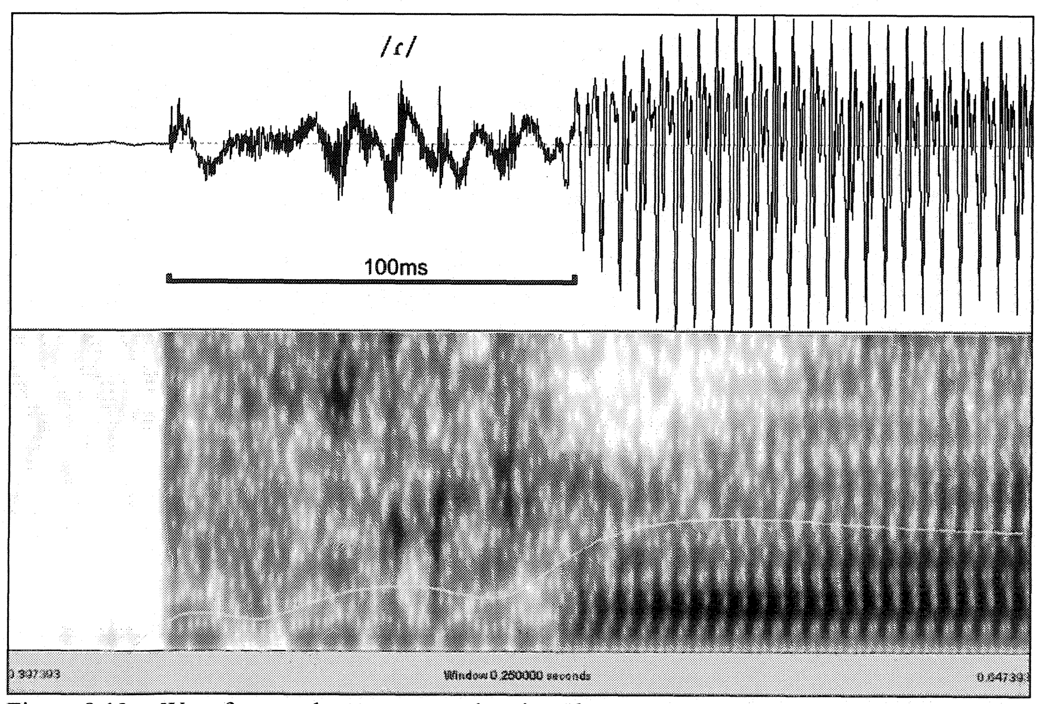

Figure 8.12 Waveform and spectrogram showing (from a 250ms window) an L2 realization of onset cluster /ɾ/ designated as a Type 4 variant. Token aurally [ya βi prado] in *Ya vi prado también*, Session 1, participant B-Beg-F4, scored “0” (inaccurate).

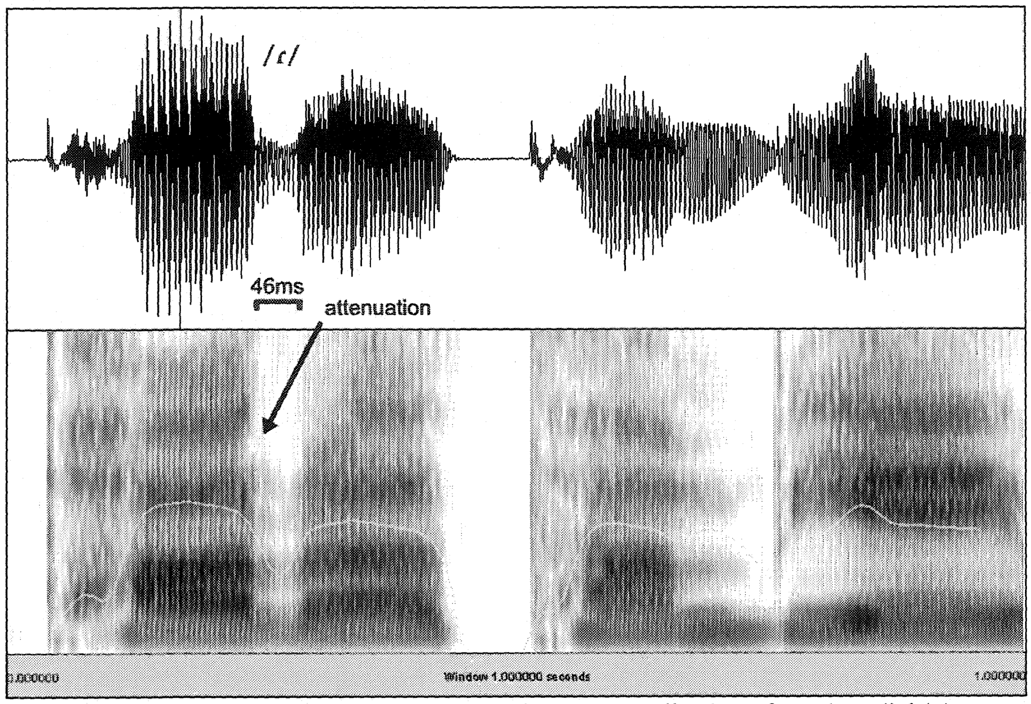

Figure 8.13 Waveform and spectrogram showing an L2 realization of word-medial /ɾ/ designated as a Type 5 variant. Token aurally [paɾa también], visually [paða tambyen] in *Habla para también*, Session 1, participant B-Adv-F6, scored "0" (inaccurate).

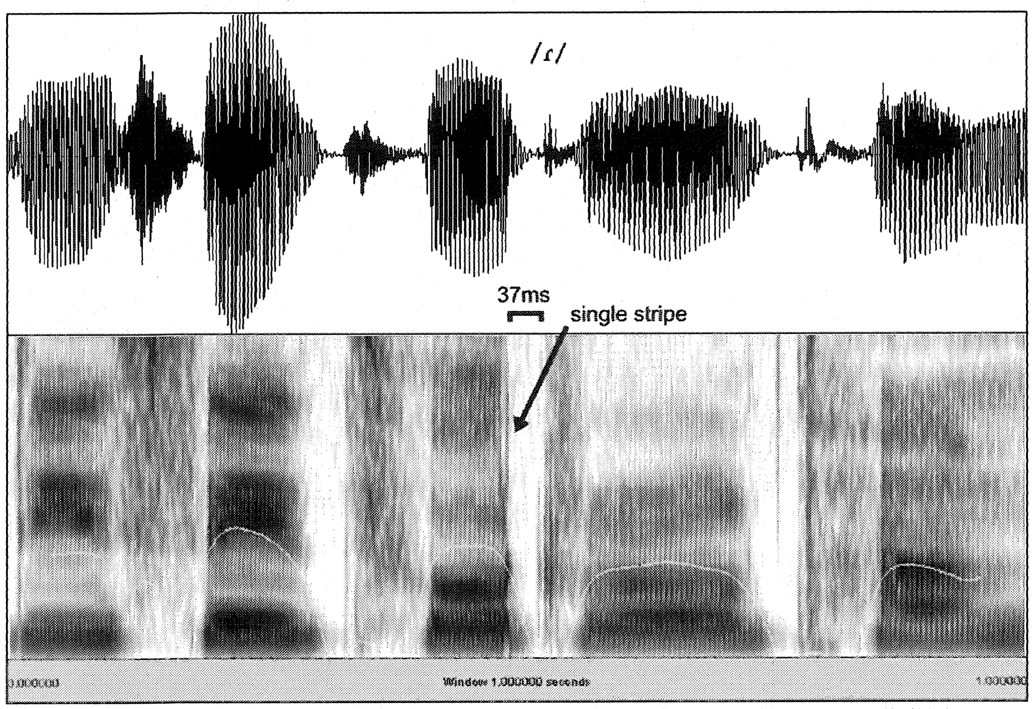

Figure 8.14 Waveform and spectrogram showing an L2 realization of word-medial /ɾ/ designated as a Type 5 variant. Token aurally [dise toɾo ta], visually [dise todo ta] in *Dice toro también*, Session 2, participant B-Adv-F6, scored "0" (inaccurate).

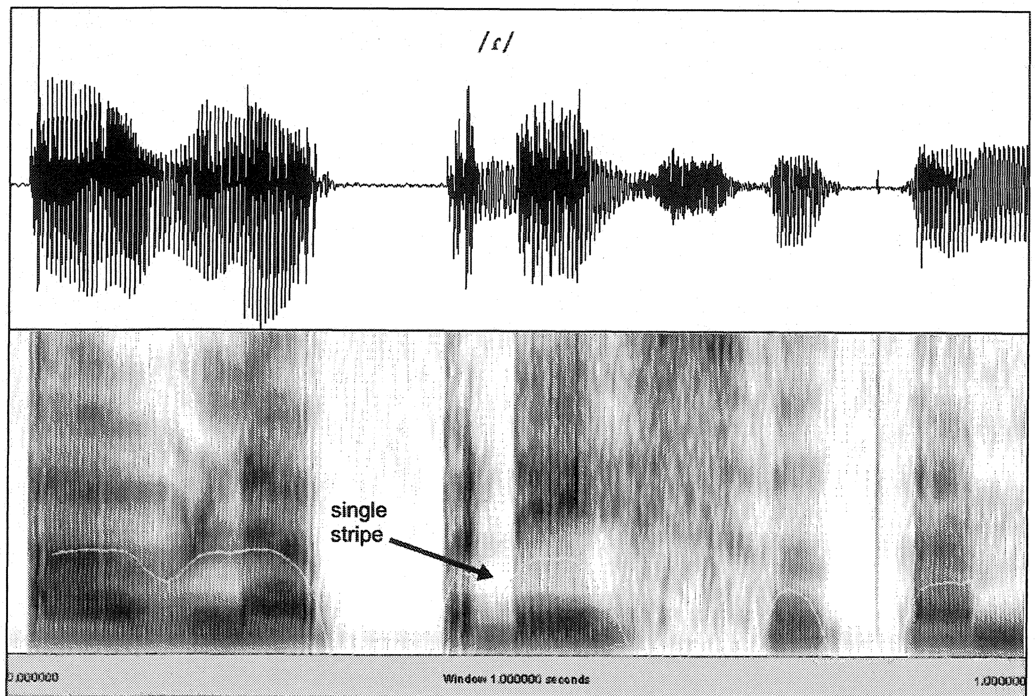

Figure 8.15 Waveform and spectrogram showing (from a 1s window) an L2 realization of onset cluster /ɾ/ designated as a Type 5 variant. Token aurally [aβla pɾeso tam], visually [aβla pɾeso tam] in *Habla preso también*, Session 2, participant B-Adv-F2, scored “0” (inaccurate).

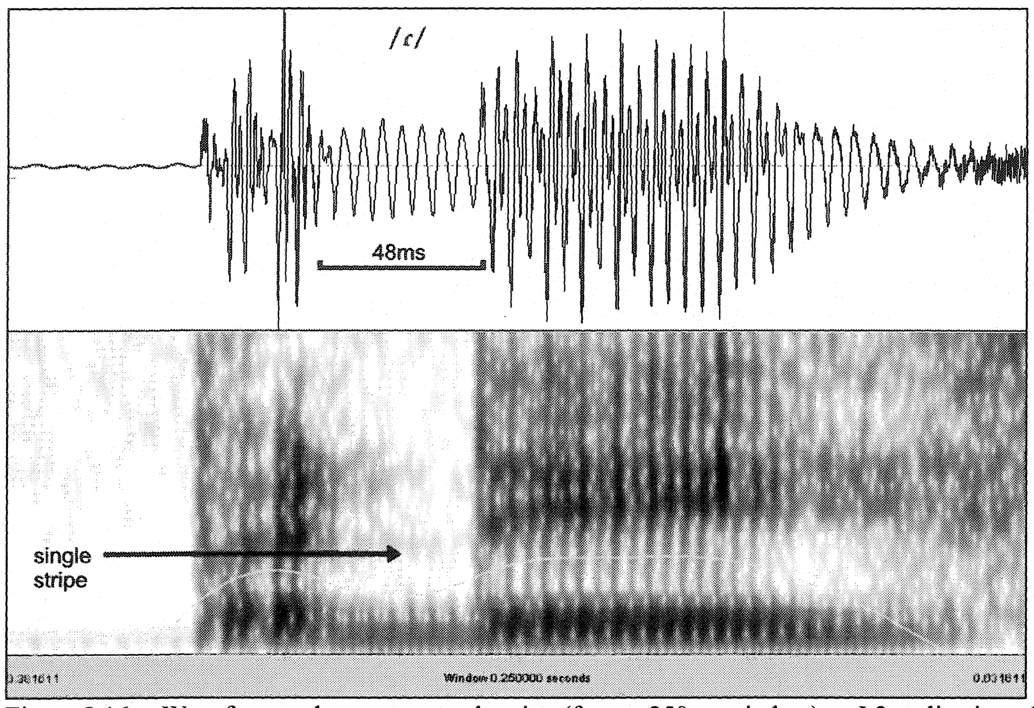

Figure 8.16 Waveform and spectrogram showing (from a 250ms window) an L2 realization of onset cluster /ɾ/ designated as a Type 5 variant. Token aurally [aβla pɾeso tam], visually [aβla pᵊdeso tam] in *Habla preso también*, Session 2, participant B-Adv-F2, scored "0" (inaccurate).

Type 5 (≈[ɾ]; ≈[d,ð]) forms accounted for 3.5% of all L2 productions. The instantiations gave the auditory impression of a single closure in the oral tract; however, the acoustic features where characteristic of voiced stop closure with release burst or a spirantized approximate. In addition, all Type 5 forms demonstrated a duration longer than 30ms. All Type 5 variants were scored as "0" (inaccurate).

In Table 8.2, the variant types for word-medial /ɾ/ are presented by Spanish level. At the beginning, advanced, and experienced levels, the Type 1 (≈[ɾ]) variant was most common; however, at the intermediate level, the Type 2 variant was most frequently observed, accounting for 72% of participant productions. All levels demonstrated all variant types, with the exception of the intermediate level which did not show any Type 4 (≈[r]; ≈[z̦,ɾ,r]) forms. Between levels, slight patterning emerged if one excludes the intermediate level results. The proportion of Type 1 (≈[ɾ]) variants increased from the beginning, advanced, to experienced levels (55% → 63% → 86%), while the proportion of Type 2 – 4 variants decreased (e.g., for Type 4 instantiations: 9% → 6% → 3%). Appendix U shows the frequency of each variant type found in word-medial /ɾ/ productions of each participant.

Table 8.3 shows the proportion of variant types realized for onset cluster /ɾ/ for each level. All levels demonstrated all variant types, with the exception of the intermediate level which did not reveal any Type 4 (≈[r]; ≈[z̦,ɾ,r]) or Type 5 (≈[ɾ]; ≈[d,ð]) forms. Type 1 (≈[ɾ]) forms accounted for a majority of the productions at the beginning, advanced, and experienced levels. Between levels, the proportion of Type 1 (≈[ɾ]) forms increased from the beginning, advanced, to experienced levels (10% → 45% → 69%). Type 2 (≈[ɹ]) and Type 4 (≈[r]; ≈[z̦,ɾ,r]) variants decreased in frequency from the beginning, advanced, to experienced levels. Appendix V presents variant type frequencies of onset cluster /ɾ/ for each participant.

Table 8.4 sets out the distribution of variant types for word-medial /ɾ/ by gender group. For example, averaging across all learner males reveals that the Type 2 (≈[ɹ]) variant was most frequent (46%) while the Type 4 (≈[r]; ≈[z̦,ɾ,r]) variant was not evidenced at all (0%). For the learner groups, Type 1 (≈[ɾ]) forms were more frequent for the females than the males (56% vs. 42%). For the experienced levels, males and females equally evidenced Type 1 (≈[ɾ]) variants (86% vs. 86%). The experienced females demonstrated all five variant types; however, the experienced males did not evidence any Type 2 (≈[ɹ]) or Type 3 (≈[ɾ]; ≈[ɹ]) forms.

Table 8.2 Frequency of variant types for word-medial /ɾ/ by participant. Standard deviation is shown in parentheses.

Level	n	word-medial /ɾ/ variant type				
		1 (≈[ɾ])	2 (≈[ɹ])	3 (≈[ɾ]; ≈[ɹ])	4 (≈[r]; ≈[ʐ,ɾ,r])	5 (≈[ɾ]; ≈[d,ð])
Beginning	5	55% (36%)	21% (22%)	12% (10%)	9% (22%)	4% (4%)
Intermediate	4	25% (33%)	72% (35%)	2% (3%)	0% (0%)	1% (2%)
Advanced	10	63% (34%)	17% (32%)	7% (10%)	6% (18%)	7% (9%)
Experienced	11	86% (13%)	3% (10%)	3% (5%)	3% (5%)	5% (8%)

Table 8.3 Frequency of variant types for onset cluster /ɾ/ by participant. Standard deviation is shown in parentheses.

Level	n	onset cluster /ɾ/ variant type				
		1 (≈[ɾ])	2 (≈[ɹ])	3 (≈[ɾ]; ≈[ɹ])	4 (≈[r]; ≈[ʐ,ɾ,r])	5 (≈[ɾ]; ≈[d,ð])
Beginning	5	10% (16%)	68% (33%)	11% (11%)	11% (19%)	1% (2%)
Intermediate	4	4% (8%)	88% (21%)	8% (13%)	0% (0%)	0% (0%)
Advanced	10	45% (29%)	26% (33%)	17% (12%)	4% (7%)	9% (21%)
Experienced	11	69% (34%)	11% (22%)	15% (15%)	3% (10%)	2% (4%)

Table 8.4 Frequency of variant types for word-medial /ɾ/ by gender group. Standard deviation is shown in parentheses.

Gender	n	word-medial /ɾ/ variant type				
		1 (≈[ɾ])	2 (≈[ɹ])	3 (≈[ɾ]; ≈[ɹ])	4 (≈[r]; ≈[ʐ,ɾ,r])	5 (≈[ɾ]; ≈[d,ð])
Learner Males	6	42% (39%)	46% (43%)	10% (11%)	0% (0%)	2% (3%)
Learner Females	17	56% (35%)	25% (33%)	7% (9%)	7% (20%)	5% (7%)
Experienced Males	4	86% (11%)	0% (0%)	0% (0%)	3% (6%)	11% (9%)
Experienced Females	7	86% (15%)	5% (13%)	5% (6%)	3% (5%)	2% (4%)

Table 8.5 shows the proportion of each variant type by level for onset cluster /ɾ/ productions. Both female groups revealed all five variant forms; however, the learner males did not demonstrate any Type 4 (≈[r]; ≈[ʐ,ɾ,r]) instantiations and the experienced males did not reveal any Type 4 or Type 5 (≈[ɾ]; ≈[d,ð]) instantiations. For both learner groups, the Type 2 (≈[ɹ]) variant was most common; 78% for males and 46% for females. At the experienced level, the Type 1 (≈[ɾ]) form was most common for males (72%) and females (67%).

Table 8.5 Frequency of variant types for onset cluster /ɾ/ by gender group. Standard deviation is shown in parentheses.

Gender	n	onset cluster /ɾ/ variant type				
		1 (≈[ɾ])	2 (≈[ɹ])	3 (≈[ɾ]; ≈[ɹ])	4 (≈[r]; ≈[z̥,ɾ,r])	5 (≈[ɾ]; ≈[d,ð])
Learner Males	6	8% (11%)	78% (25%)	13% (13%)	0% (0%)	1% (3%)
Learner Females	17	29% (31%)	46% (41%)	13% (12%)	7% (14%)	5% (16%)
Experienced Males	4	72% (41%)	17% (33%)	11% (9%)	0% (0%)	0% (0%)
Experienced Females	7	67% (33%)	8% (14%)	17% (18%)	5% (13%)	3% (5%)

In sum, this section has identified five variant types for word-medial and onset cluster /ɾ/ productions. The frequency of each variant type has been presented for each participant, Spanish level, and gender group. Between Spanish levels, the distribution of variant types showed a strong trend to become more concentrated around Type 1 (≈[ɾ]) forms. For gender groups, learner females most frequently evidenced Type 1 forms for word-medial /ɾ/ while learner males most frequently revealed Type 2 (≈[ɹ]) forms; however, for onset cluster /ɾ/, both male and female learners most frequently demonstrated Type 2 forms. For the experienced gender groups, both males and females most frequently revealed Type 1 (≈[ɾ]) forms for /ɾ/ in both word positions.

8.1.2 Measurement results of word-medial /ɾ/ productions scored "accurate"

This section presents the overall duration of word-medial /ɾ/ that were scored "1" (accurate). Table 8.6 presents the overall word-medial /ɾ/ durations by level. For example, all eight participants at the beginning level yielded an overall average duration of 23.4ms. Between levels, there were no apparent differences in overall /ɾ/ durations.

shows the overall durational averages for each participant.

In Table 8.7, the mean overall duration for word-medial /ɾ/ is presented by gender group. Learner females yielded an overall duration average of 22.8ms (2.1 stdev); 24.1ms (0.7 stdev) for learner males. In the experienced groups, overall tap duration was similar between groups (23.4ms vs. 23.7ms). Between groups, there were no apparent differences in measures.

Table 8.6 Mean overall duration of word-medial /ɾ/ by level.

Level	n	Overall mean duration for word-medial /ɾ/	
		duration in ms	stdev
Beginning	8	23.4	0.9
Intermediate	3	25.6	2.2
Advanced	9	21.9	1.8
Experienced	11	23.6	1.4
Native Speakers	11	23.0	5.4

Table 8.7 Mean overall duration of word-medial /ɾ/ by gender group.

Gender	n	Mean adjusted overall accuracy score	
		duration in ms	stdev
Learner Males	4	24.1	0.7
Learner Females	16	22.8	2.1
Experienced Males	4	23.4	1.0
Experienced Females	7	23.7	1.7

In sum, this section has presented the overall durations of word-medial /ɾ/ scored "1" (accurate) by participant, level, and gender. A presentation of the overall duration indicated no differences between levels or genders; however, the results presented here may be used to further describe, quantitatively, the overall duration of taps across all participants and groupings.

8.1.3 Measurement results of onset cluster /ɾ/ productions scored "accurate"

In this section, the measures for onset duration, svarabhakti vowel duration, and stripe to vowel duration are presented for all /ɾ/ in onset cluster productions. Table 8.8 shows overall participant measures by level for onset cluster /ɾ/. Between levels, the onset duration demonstrated an inverse relation with Spanish level; average onset duration decreased with increased Spanish level. Svarabhakti vowel duration and stripe to vowel durations did not suggest any trends between levels. From the advanced to the experienced level, the standard deviation for onset duration decreased from 21.1 to 6.2, demonstrating less duration variability at the experienced level. Appendix X shows the overall average measure for each participant that had at least 1 token scored as accurate.

Table 8.8 Mean overall duration of onset cluster /ɾ/ by level. The first number in the parentheses corresponds to the number of participants calculated in the mean; the second number the standard deviation: *mean (n; stdev)*. The term "na" refers to values of standard deviation which are not calculable due to means based on a single sample.

Level	n	onset duration in ms	svarabhakti vowel duration in ms	stripe to vowel duration in ms
Beginning	4	78 (4; 5.8)	12 (3; 7.5)	26 (4; 10.3)
Intermediate	1	74 (1; na)	6 (1; na)	19 (1; na)
Advanced	9	71 (9; 21.1)	11 (7; 9)	22 (9; 6.6)
Experienced	10	62 (10; 6.2)	18 (10; 7.2)	24 (10; 5.8)
Native Speakers	11	58 (11; 11.5)	21 (11; 7.8)	18 (11; 5.7)

Table 8.9 sets out the measures for onset cluster /ɾ/ for each gender group. The learner males revealed the longest onset durations on average (85ms) while the shortest

onset duration was found for experienced females (63ms). The learner males demonstrated a greater range in onset durations values (20.8 stdev) vs. the learner females (11.5 stdev); however, the opposite was true at the experienced level (1.9 stdev for males vs. 8.1 for females). The learner males revealed the shortest svarabhakti vowel segments (10ms) while the experienced males yielded the longest durations for svarabhakti vowel segments (21ms).

Thus, between Spanish levels, the overall onset duration decreased with increased Spanish level. Between gender groups, females at the learner level showed shorter onset durations than learner males. Regarding the svarabhakti vowel duration, both genders at the experienced level showed longer segments than the same gender group at the learner level.

Table 8.9 Mean overall duration of onset cluster /ɾ/ by gender group. The first number in the parentheses corresponds to the number of participants calculated in the mean; the second number the standard deviation: *mean (n; stdev).*

Gender	n	onset duration in ms	svarabhakti vowel duration in ms	stripe to vowel duration in ms
Learner Males	3	85 (3; 20.8)	10 (2; 4.9)	21 (3; 2.5)
Learner Females	11	70 (11; 15.5)	11(9; 8.7)	23 (11; 8.4)
Experienced Males	4	62 (4; 1.9)	21 (4; 3.7)	22 (4; 5.4)
Experienced Females	6	63 (6; 8.1)	17 (6; 8.8)	26 (6; 6.2)

8.2 Comparison of word-medial vs. onset cluster /ɾ/ accuracy scores

8.2.1 Rankings of adjusted overall accuracy scores for /ɾ/

Table 8.10 details the adjusted overall accuracy scores of onset cluster /ɾ/ and word-medial /ɾ/ for each participant at the beginning, intermediate, and advanced Spanish levels. For example, participant A-Beg-F2 was recorded in 2 sessions in Experiment A and received an adjusted overall accuracy of 0% and 32% for onset cluster and word-medial /ɾ/, respectively. Across all 23 learners, 15 revealed the ranking (from lowest to highest) of onset cluster accuracy < word-medial accuracy. A permutation test yielded a significant result on this ranking (m=.65; p<.01). Only two participants showed rankings (from lowest to highest) such that onset cluster scores > word-medial scores.

The scores of onset cluster /ɾ/ and word-medial /ɾ/ are presented for all experienced level participants in Table 8.11. For instance, participant B-ExpL-F6 received an adjusted overall accuracy score of 95% for onset cluster /ɾ/; 100% for word-medial /ɾ/. Six experienced level participants evidenced the ranking such that onset cluster /ɾ/ scores were lower than word-medial /ɾ/ scores; however, a permutation test did not find this ranking significant. Three experienced level participants revealed the reverse ranking such that scores for word-medial /ɾ/ were lower than scores for /ɾ/ in onset clusters.

Table 8.12 presents the mean adjusted overall scores for /ɾ/ by participant level. Within each level, scores for onset cluster /ɾ/ were consistently lower than scores for word-medial /ɾ/ with the exception of participants at the advanced level in Experiment A. Between levels, the experienced level showed the highest scores for both word positions. With the exception of the intermediate level, scores for each word position increased from the beginning, advanced, to experienced levels (i.e., scores for word-medial /ɾ/ were 59%, 67%, and 90% for the beginning, advanced, and intermediate levels, respectively).

Table 8.10 Comparison of adjusted overall accuracy scores for onset cluster vs. word-medial /ɾ/ by participant.

	Participant	Adjusted overall target accuracy	
		Onset cluster /ɾ/	Word-medial /ɾ/
Experiment A	A-Beg-M1	5%	88%
	A-Beg-M2	41%	80%
	A-Beg-F1	9%	8%
	A-Beg-F2	0%	32%
	A-Int-M1	0%	0%
	A-Int-M2	23%	80%
	A-Int-F1	0%	48%
	A-Int-F2	0%	8%
	A-Int-F3	0%	0%
	A-Adv-M1	32%	24%
	A-Adv-M2	0%	0%
	A-Adv-F1	100%	100%
	A-Adv-F2	100%	100%
Experiment B	B-Beg-F1	0%	84%
	B-Beg-F2	27%	72%
	B-Beg-F3	0%	6%
	B-Beg-F4	54%	100%
	B-Adv-F1	41%	66%
	B-Adv-F2	41%	78%
	B-Adv-F3	75%	42%
	B-Adv-F4	34%	90%
	B-Adv-F5	34%	96%
	B-Adv-F6	68%	72%

Table 8.11 Comparison of adjusted overall accuracy scores for onset cluster vs. word-medial /ɾ/ by experienced level participant.

	Participant	Adjusted overall target accuracy	
		Onset cluster /ɾ/	Word-medial /ɾ/
Experiment B	B-ExpL-M1	14%	84%
	B-ExpL-M2	100%	100%
	B-ExpL-M3	100%	84%
	B-ExpL-M4	100%	96%
	B-ExpL-F1	81%	96%
	B-ExpL-F2	95%	96%
	B-ExpL-F3	68%	84%
	B-ExpL-F4	0%	60%
	B-ExpL-F5	100%	100%
	B-ExpL-F6	95%	100%
	B-ExpL-F7	100%	96%

Table 8.12 Mean adjusted overall /ɾ/ accuracy scores by level

Level	n	Mean adjusted overall /ɾ/ accuracy	
		Onset cluster /ɾ/	Word-medial /ɾ/
Beginning	8	17% (21%)	59% (38%)
Intermediate	4	5% (10%)	27% (35%)
Advanced	10	52% (32%)	67% (34%)
Experienced	11	77% (37%)	90% (12%)

Table 8.13 shows scores for the male and female groups at the learner and experienced levels. For all gender groups, scores for /ɾ/ in onset cluster were lower than scores for /ɾ/ in word-medial position. Between levels, the scores at the experienced level were higher than scores for the learner counterparts (e.g., experienced level male scores

were 78% and 91%, compared with 17% and 45% for learner level males). Within the learner groups, both scores for female learners were higher than scores for male learners. In contrast, males at the experienced level scored 1% point higher in both word positions in comparison to experienced level females.

Table 8.13 Mean adjusted overall accuracy scores for /ɾ/ by word position and gender. Standard deviation is shown in parentheses.

Gender	n	Mean adjusted overall /ɾ/ accuracy	
		Onset cluster /ɾ/	Word-medial /ɾ/
Learner Males	6	17% (18%)	45% (42%)
Learner Females	17	34% (35%)	59% (36%)
Experienced Males	4	78% (43%)	91% (8%)
Experienced Females	7	77% (36%)	90% (14%)

Thus, participant scores demonstrated a consistent ranking such that scores for /ɾ/ in onset position were lower than scores for /ɾ/ in word-medial position; however, this ranking was not notable when analyzing scores only at the experienced level. Between levels and gender groups, a descriptive analysis revealed the same robust relationship. With the exception of scores at the intermediate level, scores revealed a strong trend to increase with increased Spanish level.

8.2.2 The role of the svarabhakti vowel in onset cluster /ɾ/

Table 8.14 shows the average svarabhakti vowel emergence (SVE) score for each level. Across all 8 participants at the beginning level, mean SVE was 2%. Between levels, mean SVE score was equal between the beginning and intermediate levels. Above the intermediate level, SVE scores increased with level (1% → 21% → 59%). The experienced group received the highest SVE score: 59% across all 11 participants with a 34% standard deviation. Individual scores for SVE are presented in Appendix Y.

Table 8.14 Mean adjusted overall svarabhakti vowel emergence (SVE) by level. Standard deviation is shown in parentheses.

Level	n	Mean SVE scores	
Beginning	8	1%	(2%)
Intermediate	4	1%	(2%)
Advanced	10	21%	(30%)
Experienced	11	59%	(34%)
Native Speakers	11	83%	(19%)

Table 8.15 sets out the mean SVE score for each gender group. Learner males received the lowest mean SVE score (1% with a 2% standard deviation). Females at the learner level received higher scores than the males (13% vs. 1%), but with greater variability between participants (25% stdev for females vs. 2% stdev for males). For the experienced groups, the males showed higher scores than females (69% vs. 52%); however, range of scores was comparable between male and females (40% stdev for males vs. 31% stdev for females).

Table 8.15 Mean adjusted overall svarabhakti vowel emergence (SVE) by gender. Standard deviation is shown in parentheses.

Level	n	Mean adjusted overall SVE	
Learner Males	6	1%	(2%)
Learner Females	17	13%	(25%)
Experienced Males	4	69%	(40%)
Experienced Females	7	52%	(31%)

In order to test for an effect of level on SVE scores, a chi square test was performed on the binary probability that a participant evidenced the lowest native speaker SVE score. That is, a threshold SVE score of 44% was established based on the fact that 44% was the lowest SVE score received by any native Spanish speaker. Thus, a participant received a score of "0" if he or she demonstrated a SVE score lower than 44%; a score of "1" if he or she had a SVE score greater than or equal to 44%.

Table 8.16 sets out the number of participants at each level that received a binary SVE score of "0" (non-native-like SVE score) and "1" (native-like SVE score). For example, at the beginning level there were no participants that evidenced any SVE scores greater than 44%. That is, there were eight occurrences of "0" and no occurrences of "1"; however, at the experienced level there were three scores of "0" and eight scores of "1." A chi-square test of independence revealed a significant effect of level on native-like svarabhakti vowel emergence score ($X^2 = 29.8$; $p<.001$).[44] In order to test where the significance was found, four post-hoc multiple pair-wise comparisons were performed between each level and the native speaker level. A Bonferroni correction on the alpha level required significance to be lower than .0125. The pair-wise comparisons found no

[44] Because of the low count numbers in some of the cells, it might be argued that a chi-square test is inappropriate for this comparison; however, a Fisher's Exact Test for independence of count data was also carried out (with 5 Levels of experience), resulting in a p-value of 1.804e-07 which also suggests evidence of dependence between accuracy scores and experience level. Special thanks to Marios Pavlides and Paul Sampson in the Department of Statistics for their help with these tests.

significant difference between the experience level participants and the native speakers; however, a significant difference was found between each learner level and the native speaker level. Comparing the native speaker level to the beginning, intermediate, and advanced levels, the results were $X^2 = 19$, $p<.001$; $X^2 = 19$, $p<.001$; and $X^2 = 14.2$, $p<.001$, respectively.

Table 8.16 Number of participants at each level who demonstrated either did not demonstrate native-like SVE ("0") and number that did show native-like SVE ("1").

Level	n	Binary SVE score	
		"0"	"1"
Beginning	8	8	0
Intermediate	5	5	0
Advanced	10	8	2
Experienced	11	3	8
Native Speakers	11	0	11

In sum, this section has found that only the experienced level participants demonstrated svarabhakti vowel emergence similar to native Spanish speakers. These results are discussed in Section 9.1. The following chapter addresses all the results of this study and a wide range of their implications for the study of interlanguage and phonology.

9. Discussion

The organization of this chapter is as follows: Section 9.1 addresses this study's six questions based on the results from Experiment A and Experiment B. Section 9.2 addresses several patterns of L2 variants and their implications for interlanguage development. In Section 9.3, the role of differences in Spanish level and gender are addressed in light of various results. Section 9.4 discusses the results and their implications for phonological perspectives on L1 Spanish. In Section 9.5, this study's methods of using a two-tiered assessment of determining pronunciation accuracy are evaluated. Section 9.6 sets out the implication of the results for pronunciation instruction. Suggestions and ideas for future study are presented in Section 9.7. This investigation is summarized in Section 9.8.

As in previous chapters, the "learner level" refers to all participants from the beginning, intermediate, and advanced Spanish levels. The "experienced level" only refers to participants from the experienced Spanish level (i.e., non-native instructors of Spanish).

9.1 Addressing this study's questions

Question 1: Is there a relative degree of difficulty of word-medial /d, t, r, ɾ/?

Robust patterns were found suggesting differences in difficulty. Subjects' scores revealed a clearly most difficult and least difficult phone ([r] and [t], respectively); however, not all subjects and groups adhered to the same internal ranking of scores. The data presented in Section 6.2.1 revealed two significant rankings (from lowest to highest): /r <= d <= ɾ <= t/ and /r <= ɾ <= d <= t/. Thus, it was just as likely to observe rankings in which scores for /d/ were lower than or equal to scores for /ɾ/ as it was to find scores for

/ɾ/ lower than or equal to scores for /d/. Although both rankings were significant for scores across sessions, both rankings were only supported by 9 out of 23 participants, emphasizing the existence of inter-subject variation of accuracy scores. Thus, degree of difficulty may be generalized across learners; however, different individual development patterns should not be ignored. Averaging across levels revealed the direction (from lowest to highest) of /r < d < ɾ < t/ for all learner levels. In the cross-gender comparison, scores for the female learners again supported the ranking (from lowest to highest) of /r < d < ɾ < t/; however, scores for the male learners revealed a third ranking (from lowest to highest) of /d < r < ɾ < t/.

The experienced level participants did not demonstrate any significant ranking of scores; however, averaging across all experienced participants demonstrated a trend in the score rankings (from lowest to highest) of /d <= r <= ɾ <= t/. Averaging across all experienced male scores revealed the same score for both /d/ and /r/ (75% adjusted accuracy), revealing the ranking (from lowest to highest) of /r = d <= ɾ <= t/. Averaging across female scores at the experienced level showed a third ranking for experienced level participants of /d <= r <= ɾ <= t/.

In regards to this study's hypotheses, the first hypothesis for this question was intended to investigate Lado's (1957) claim that splitting of allophones is the most difficult process in L2 phonological acquisition, predicting that scores for /d, t, ɾ/ would be lower than scores for /r/. The second hypothesis for this question predicted that scores for /r/ would be lowest based on the observation that a new phoneme requiring novel gestural skills would be most difficult. Although individual patterns supported both hypotheses, the most common rankings across all speakers supported the second hypothesis. Thus, this study shows strongest support for the hypothesis that the learning of a sound that requires additional gestural skills (e.g., /r/) is more difficult than the learning of sounds that do not require new gestural skills (e.g., /d, t, ɾ/). One may also note that the results also suggest that the acquisition patterns of segments in L2 are similar to L1 acquisition in which trills are acquired relatively late in Spanish (Bosch 1983, Gonzalez 1989, Stoel 1974).

While Lado's claim was not categorically supported, some participants did evidence lower scores for /d/ and /ɾ/ than scores for /r/, suggesting for those participants that the reorganization of existing phones (e.g., /d,ɾ/) was more difficult than learning a new phoneme with novel gesturing (e.g., /r/). Additionally, when participants were grouped by gender, the male learners demonstrated lowest scores for /d/; however, with only two male subjects for each learner level, one cannot be certain that similar results would be found among a larger population. One could also argue that the sound categories studied in this investigation have not completely tested the difficulty of the allophonic split. That is, the analysis of L2 /d/ did not address the production of [d], but only one allophone of /d/, namely [ð]. Thus, a future study on the assessment of negative transfer may find differing results on the acquisition of /d/ when all allophones and their word-positions are analyzed.

Question 2: Does trill accuracy emerge with higher scores word-medially over word-initial trill scores?

Results in Section 7.2.1 revealed that participants were significantly more likely to achieve higher accuracy scores for word-medial /r/ than for word-initial /r/. Although only 11 out of 23 learner level participants across both experiments were able to produce /r/, the word-medial /r/ scores for all 11 participants were higher than their corresponding word-initial /r/ scores, demonstrating a significant and robust effect of word-position. Patterning of scores by gender and level also supported the effect of word position on /r/ scores even though female scores were higher than male scores, suggesting a main effect of word position with gender interaction. In addition, the /r/ accuracy scores suggest that the emergence of word-initial /r/ accuracy requires the emergence of /r/ accuracy word-medially as demonstrated by all participants in this study. This effect of word position may have been due in part to spelling, since word-initial trills are indicated by a single grapheme <r>.

One possible reason for the effect of word position found in Experiment A may have been that all word-initial /r/ targets were preceded by the mid vowel [e] and were consequently more difficult to produce than word-medial /r/ targets which also followed high and low vowels. In effect, the low vowel environment might have boosted word-medial scores more than the high vowel environment might have hindered word-medial /r/ accuracy; however, in Experiment B vowel height was varied for both positions and the same robust effect of word position on /r/ scores was observed. In Experiment B, all participants that evidenced any /r/ accuracy in either position showed higher scores for /r/ word-medially than scores for /r/ word-initially.

In sum, the findings suggest that there are word-position effects on segmental production in L2 acquisition. Just as Zampini (1998) found that learners were more likely to spirantize /b/ word-medially than between word boundaries (i.e., word-initially), the results of this study also found an effect of word position on /r/ accuracy scores. Still, because orthography is confounded with word-position for /r/ in Spanish, the relative influence of orthography and word-position is difficult to determine.

Question 3: Does /ɾ/ accuracy emerge with higher scores word-medially over scores of /ɾ/ in onset clusters?

Similar to effects found for trills, scores for tap accuracy significantly differed by word-position as shown in Section 8.2.1. Across all learner participants, word-medial /ɾ/ accuracy scores were significantly higher than onset cluster /ɾ/ scores. Average scores calculated across levels and gender groups also suggested an effect of syllable position on accuracy scores. For the experienced level participant scores, the effect of syllable position was not significant, although a trend for higher scores for /ɾ/ word-medially than in onset cluster position emerged when results were averaged across gender and level.

Despite the robust effect for syllable position, several participants demonstrated different patterns of rankings. Out of 23 learner participants, 2 showed higher scores for onset cluster /ɾ/ than scores for word-medial /ɾ/. Additionally, two participants at the advanced level evidenced 100% accuracy for /ɾ/ in both positions. At the experienced level, 3 out of 11 participants showed higher scores for /ɾ/ in onset cluster than scores for /ɾ/ word-medially.

In sum, the null hypothesis was rejected for this question, supporting the prediction that scores for word-medial taps will be systematically greater than scores for taps in onset clusters. These results suggest an effect of syllable position on the acquisition of /ɾ/ for L2 learners; however, because an effect for word position was not found for the experienced level participant scores, the results suggest that the word-position effect may only apply during the earlier stages of interlanguage development.

Question 4: Is the relative ranking of scores for /t/ affected by length of voice onset time as a criterion for /t/ accuracy?

As shown in Section 6.2.2, two separate approaches for establishing a VOT criterion resulted in reduced scores for /t/ for some participants; however, both approaches yielded similarly significant results. Each criterion separately resulted in the significant ranking of /r <= ɾ <= d <= t/ while the second ranking of /r <= d <= ɾ <= t/ was no longer found to be significant.[45] Although individual variation for score rankings was observed, a reanalysis (using either criterion) of gender and level group scores still supported the ranking (from lowest to highest) of /r <= d <= ɾ <= t/.

Both approaches to establishing a VOT criterion (i.e., determining what VOT range should be considered native-like) accounted for variability in different ways. In the first approach, the VOT duration criterion was based on all native speaker variability. Thus, 41ms was the longest evidenced native speaker VOT and was selected as a threshold requirement for an L2 production to be scored as "accurate." In effect, 14 out

---

[45] For example, in the case of the 41ms duration criterion and the score ranking order /r <= ɾ <= d <= t/, scores for participant A-Beg-M2 no longer satisfied the ranking after the inclusion of the 41ms VOT criterion. Similarly, for the score ranking order /r <= d <= ɾ <= t/, score rankings for three participants (A-Beg-M2, B-Beg-F1, and B-Beg-F2) no longer adhered to the ordering; however, the permutation still found the /r <= ɾ <= d <= t/ ordering significant, but not /r <= d <= ɾ <= t/.

of 23 learner level participants had scores reduced for /t/ while no scores were reduced for the experienced level participants. In a second approach, the VOT duration criterion was statistically based to account for the normal distribution of native speaker VOT durations, treating productions on the tail above the mean as aberrations and possible speech errors. Thus, two standard deviations about the mean (i.e., 35ms) was selected as the threshold criterion for an L2 production to be scored as "accurate." Using the second more conservative approach reduced scores for /t/ for 18 out of 23 learner participants and 2 experienced participants. In comparison to the first approach, results were not significantly different using the second method although the significance of the rankings was slightly more robust with the first approach.

One might ask whether it would *ever* matter which approach to use since neither approach led to a meaningful difference in the analysis of the results; however, a different distribution of native speaker durations might have dramatically altered the comparison of both approaches. For example, had a single native speaker produced even one token with an extreme VOT duration (e.g., 60ms), then the comparison of both approaches would have been quite different as almost no scores for /t/ would have been reduced with a 60ms threshold. Additionally, if VOT durations were more compressed or even bi-modal, a normal distribution could not have been assumed and the second approach would not have been appropriate. In the current study, neither approach led to any additional insight into the results which may be due to the fact that perceptually, there is no noticeable difference between 41ms and 35ms (Lehiste 1970). A future study might want to use a just noticeable difference criterion instead.[46] Still, either approach in the current study could be argued for and are founded on different principles of variability (i.e., what should count as native-like).

The inclusion of a VOT criterion also has implications for the acquisition of VOT for word-medial /t/. At the experienced level, no participant scores for /t/ were affected by the inclusion of the 41ms VOT criterion and only 2 participants had scores reduced with the 35ms criterion. In effect, all experienced level participants were producing

---

[46] Lehiste (1970) suggests that a just noticeable difference duration would be between 10 and 40ms.

word-medial /t/ with VOT durations within the complete native speaker range. In contrast, scores for /t/ were most dramatically altered with both approaches for the beginning level participants. The effect of the VOT criteria on scores for /t/ was less dramatic with the increase of Spanish level. Intuitively, these results support the observation that increased Spanish experience leads to more native-like timing patterns of VOT for word-medial /t/. While this observation may not seem surprising, it is nonetheless important as previous studies of L2 productions of VOT have been limited to /t/ in word-initial position, suggesting that their results might be generalized to released stops in other positions.

In sum, the inclusion of either VOT criterion in the scoring of /t/ accuracy did affect accuracy scores. Without either duration criterion, two rankings were identified; however, two separate criteria and their analyses resulted in only one significant ranking (from lowest to highest) of /r <= d <= ɾ <= t/. Importantly, both criteria account for variability in different ways although neither analysis led to a meaningful difference in the results.

Question 5: Is there an effect of vowel height on trill productions scored as "accurate?"

There was a significant effect of preceding vowel height on /r/ raw accuracy scores as shown by the data in Section 7.2.2. Although slight variation existed between speakers, inferential analysis found that scores for /r/ were significantly lower in high vowel environments. This effect was significant between the high vowel environment and both lower vowel environments. Thus, participants were significantly less likely to receive a score of "1" (accurate) for /r/ in high vowel environments than for /r/ in mid or low vowel environments. In effect, there was no significant statistical difference between scores for /r/ in mid and low vowel environments; however, there was a robust trend for /r/ scores to be higher in low vs. mid vowel environments. The effect of vowel height was also supported by native speakers' productions of /r/. Averaging scores within each level also supported the effect of vowel height on /r/ scores. For all levels, scores for /r/ were lowest in the high vowel environment and highest in the low vowel environment. Averaging scores across gender levels also supported this trend. The finding of lower scores for /r/ in high vowel environments is possibly explained by the fact that [r] has additional aerodynamic requirements in the context of a high vowel (Solé 2002).

Studies of palatograms and motor memory might also help to explain the effect of vowel height on trill accuracy. Gili (1921) noted that realizations of [ɾ] and [r] are produced farther forward on the alveolar ridge in the context of a high vowel. While place of articulation may be confounded with aerodynamic requirements, one might also predict that learners are more accustomed to producing [r] in mid and low vowel environments, assuming that all three environments are equally frequent in the language. If this is indeed the case, then learners may succeed in producing the trill more often when targeting a particular place of articulation that accommodates [r] in mid and low environments, but not in high vowel environments. Still, this line of reasoning does not account for the fact that native speaker [r] productions are affected by vowel height, suggesting that aerodynamics is a factor beyond any potential interaction with tongue tip placement.

In sum, scores for trills were higher in the context of lower vowels as predicted by the additional aerodynamic requirements required for trills in high vowel environments (Solé 2002). Thus, these results suggest that vowel height may be an important influence in the L2 acquisition of some segments but not in others.

Question 6: Is the emergence of the svarabhakti vowel limited to highly experienced L2 speakers of Spanish?

Across all L2 participant levels, only the experienced level demonstrated a native-like distribution of svarabhakti vowel emergence (SVE). As described in Section 8.2.2, several advanced level participants also demonstrated native-like SVE frequencies although the advanced level as a group did not evidence SVE similar to the native

Spanish speaker group. There was some SVE at the beginning and intermediate levels; however, none of the frequencies of SVE were comparable to native Spanish speakers.

The SVE scores suggest that the production of the svarabhakti vowel is a feature that is very late to emerge in the acquisition process of L1 English speakers of L2 Spanish. Future research might investigate whether speakers of L1 backgrounds other than English also require a lot of experience in L2 Spanish before evidencing the epenthetic vowel. For L2 English speakers of L2 Spanish, the relatively late emergence of the element might be due to the requirement for the learner to acquire both the features of the initial consonant as well as /ɾ/. That is, one question is whether both the first element in the cluster and the tap must be appropriately acquired for native-like SVE.

The results of this study have several implications for our understanding the underlying nature of the svarabhakti vowel. The fact that the native speaker productions showed an extreme range in the emergence of the vowel (44%-100%) suggests that the element is not strictly phonetic in nature. That is, one could posit that the voicing of the glottis before the contact of the [ɾ] serves to aid in the aerodynamics or articulatory timing of the tap occlusion; however, if this were indeed the motivation for its emergence, we would expect it to be found more often across all native speakers. It's possible (although impossible to verify) that the emergence of the svarabhakti vowel was historically systematic, suggesting that it could have originally been motivated on a general phonetic constraint which became phonologized over time. Thus, a possible phonological motivation for the vowel might be a segment or syllable timing constraint in the Spanish grammar which motivates the tap segment or onset to occupy a specific duration relative to a larger prosodic domain. This argument would also account for why the element is also found following a tap in utterance final position (e.g., *mar* [maɾᵊ] 'sea') and between a tap and following consonant (e.g., carne [kaɾᵊnɛ] 'meat') (Gili 1921) .[47] Perhaps a relatively short initial consonant (e.g., [p]) alongside the relatively short

[47] Two additional arguments include the notion of prevoicing of a tap (similar to the prevoicing of utterance initial Spanish [b, d, g]) or anticipatory voicing assimilation of the following nucleic vowel; however, neither motivation would account for why the svarabhakti vowel also emerges following a tap in Spanish.

duration of [ɾ] triggers the emergence of the epenthetic vowel. Additionally, this hypothetical constraint might be used to aid in perception. That is, in order for a listener to more effectively identify the consonant followed by the following tap (i.e., the onset), the epenthetic vowel may serve as a form of phonetic boundary marker. This phonological argument is supported by a recent study which found that the epenthetic vowel was masked (i.e., it was not perceptually salient) when following a [ɾ]; however, the epenthetic vowel was salient when followed by a non-tap segment (Widdison 2004). A need remains for research to conclusively determine the underlying nature of the svarabhakti element (e.g., native speaker perception and well-formedness judgments of onsets both with and without the element). Nevertheless, the fact that the native speakers, as well as learners, evidenced productions of tap occlusions without the element suggests that the svarabhakti vowel and tap are somehow separately motivated.

9.2 L2 variant patterns of particular interest

This study has classified L2 variants based on a set of features extracted from native Spanish speaker productions. One must be careful in the discussion of variants to realize that there may be other differences between variants in the same category that were not noted in this study's analyses. Ideally, one would be able to identify all the auditory and acoustic features for each L2 production; however, practicalities of research do not always allow one to extract all possible detail. Nevertheless, in this section I address several forms and their patterns for a variety of participants, addressing interlanguage and future study.

The realization of a uvular place of articulation for /r/ has several implications for interlanguage. Participant B-Beg-F4 produced what sounded like uvular trills for three /r/ targets. Interestingly, these three realizations occurred on the same word list and only for the word-initial /r/ targets (*rico, remo,* and *rato*); however, the impression of the uvular place of articulation was not present for the word-medial /r/ targets on the same list (*churro, perro,* and *carro*).[48] The realization of a uvular trill for an apical trill target in interlanguage supports the notion that patterns of interlanguage reflect patterns of L1 acquisition since children have been known to produce the uvular trill when attempting the apical trill in L1 (Wright, pc). In addition, a few Caribbean dialects of Spanish do produce /r/ with a uvular variant (Widdison 1998). Widdison has suggested that the uvular variant is a manifestation of an acoustic constraint and it has been suggested that this process is a result of an over-exertion of the retraction and elevation of the tongue dorsum which acts as a secondary articulation in the apical trill. Thus, one explanation for the L2 uvular variant might be that it occurred because the learner was overgeneralizing and hyperarticulating a newly learned secondary articulation of the tongue dorsum. A second, and perhaps more plausible explanation for participant B-Beg-F4's uvular variant is that she had studied French for four years, suggesting that her interlanguage is a conglomerate of all previous L2 experience.

Another implication of participant B-Beg-F4 is that learners demonstrate different variants for /r/ in different positions. Learners may be processing the same phone as two different sounds in their interlanguage when that phone occurs in two different word positions. Just as L1 English speaker learners are assumed to categorize /d/ and /t/ as different sounds, word-initial /r/ may be categorized differently from word-medial /r/, at least at the earlier stages of interlanguage. This distinction is also supported by different scores for the same target in different word positions. The question then remains as to why students would categorize /r/ differently based on word-position? One possibility is that early interlanguage categorization relies on orthography (i.e., <r> for word-initial trill and <rr> for word-medial trill). Alternatively, the establishment of underlying categories may start with prosodic dependencies such as word-boundaries or syllable types.

One limitation of this study is that the set of variant categories distinguished do not allow one to identify native-like L2 productions that may reflect a dialect of Spanish

[48] Suggesting that production patterns change from list to list, just as one would expect changes in productions from one task (e.g., free elicitation) to the next (e.g., paragraph reading).

not represented by the sampled native speakers. For example, participant B-Adv-F2 received a score of "0" for 6 out of 9 productions of onset cluster /ɾ/ due to closure durations that were longer than 30ms duration.[49] The <30ms duration criterion was used because the 11 native speakers in the study all demonstrated similar closure durations for almost all onset cluster /ɾ/ productions; however, participant B-Adv-F2 had studied abroad in northern Argentina, a dialect of Spanish not represented by the native speakers in this study. Thus, it's not clear if her productions of /ɾ/ in onset clusters were representative of an Argentinean Spanish dialect that she may have acquired, or if her productions were characteristic of interlanguage development. Participant A-Adv-M1 also demonstrated a long tap closure in several instantiations of /ɾ/ in onset clusters; however, the long closure duration was not dominant among his productions. Thus, the long duration in his realizations supports the possibility that participant B-Adv-F2 was demonstrating interlanguage forms; however, the fact that she evidenced near native accuracy in all other targets, as well as the consistency of her onset cluster /ɾ/ productions suggests that her /ɾ/ productions might have been characteristic of a dialect not sampled in this study.

The study of interlanguage variants also has an interesting implication for the process of overgeneralization. Whereas one might use the term overgeneralization to describe the process of categorically using one form for another, the results of this study suggest that overgeneralization is best understood as a probablistic rule. For example, participant B-Adv-F3 realized trills for both the /ɾ/ and /r/ targets. Out of 18 word-medial /r/ targets across two sessions, she produced 11 realizations that were categorized as a Type 1 (≈[r]) trill variant (i.e., scored as an "accurate"). Out of 18 word-medial /ɾ/ productions, 10 were categorized as a Type 4 (≈[r]; ≈[z̧,ɾ,r]) tap variant (i.e., as the voiced multiple vibrant); however, 7 were categorized as Type 1(≈[ɾ]) tap variants (i.e., as an "accurate" single tap), and the remaining productions were realized as Type 2 (≈[ɹ])

[49] Her overall accuracy scores were 41% and 78% for onset cluster and word-medial /ɾ/, respectively. For all 18 /ɾ/ in onset clusters, the average closure to vowel duration was 36ms with a 7.3ms standard deviation (49ms longest duration).

variants. In effect, not all of her productions of /ɾ/ were realized as a trill, suggesting that overgeneralization is a probabilistic rather than categorical process. Thus, analyses of interlanguage processes should analyze the occurrence of overgeneralization as a token by token effect.

The classification of all productions into variant types has implications for performance errors. So-called "slips of the tongue" are often assumed to be aberrations that can be ignored; however, one contribution of this study is that no production has been discarded, but rather, all productions were classified as some variant type. For instance, participant A-Adv-F2 evidenced a spirantized [t] in which the release burst was replaced with frication. Auditory re-inspection of the acoustic signal might suggest that the production was merely a slip of the tongue; however, the participant produced similar realizations for 3 out 9 /t/ targets across two different tokens in the same session. Thus, the consistency of these non-native-like productions suggests that the variants were not mere slips of the tongue, but rather reflective of an interlanguage system.

The interlanguage variants also suggest that interlanguage change is based on sub-phonemic features, and not forms categorically (e.g., L1 Spanish speakers do not acquire aspiration for L2 English stops overnight, but rather acquire it gradually). That is, the variants in this study suggest that L1 forms may be transferred in a gradient manner by their features and that changes occur at the feature level rather than at the categorical level, supporting the phenomenon of *compromised* forms. The occurrence of compromised productions has been well established in the L2 acquisition of voice onset time (Flege and Efting 1986, Flege 1991); however, compromised forms have yet to be formally studied for categories other than stop consonants. In the current study, several participants evidenced seemingly compromised forms for /r/ and /ɾ/. For example, participant B-Adv-F2 produced one token of word-initial /r/ that gave the impression of a single obstruction in the oral tract alongside r-coloring. Visually, the production revealed two white stripes in the upper half of the spectrogram (consistent with the multiple vibrant) alongside an F2-F3 pinch which might account for the impression of the r-coloring in the auditory analysis. Thus, the token shared features of L1 English [ɹ] and

L2 Spanish [r]. In another example, participant B-Adv-F5 produced approximants for /r/ targets; however, visual inspection revealed slight frication in the spectrogram. Interestingly, these slightly fricated [ɹ] productions were only found for her L2 targets and not for her English [ɹ] realizations. Hence, the slight frication suggests a compromise between the features of L1 English [ɹ] and the features of L2 Spanish [r].

The identification of compromised forms also has implications for the identification of how and when forms are labeled as "transfer" vs. "developmental" processes.[50] For example, Major (1986) transcribed L1 English speaker productions of L2 Spanish /r/ and /ɾ/. In his study, any /r/ or /ɾ/ target realized as [ɹ] would have been labeled as a "transfer" process; however, as discussed in the previous paragraph, participant B-Adv-F5 demonstrated that the auditory impression of [ɹ] may not reveal features that are otherwise present in a visual inspection. The implication for Major is that there may not be a clear distinction between a "transfer" and "developmental" process. At the very least, many productions identified as "transfer" may in fact be developmental in nature. This evidence also supports the use of acoustic inspection for interlanguage forms as auditory inspection alone may not sufficiently reveal a production's nature. A more radical implication of the current study's findings might be to argue that no interlanguage form is truly in a "transfer" state. That is, at any stage of interlanguage (with possibly the exception of fossilization in ultimate attainment), a form must be developing, although the schedule of development at the beginning may be slow enough to give the impression of an exact L1 form (i.e., transfer). This reasoning would suggest that the term "transfer" implies a static state *prior* to any development. Thus, a better approach might be to classify all interlanguage productions as developmental in nature. In effect, one could think of all forms as lying along a continuum of

---

[50] Actually, these "processes" are often referred to as "errors" which unfairly suggests a judgment that learners are doing something wrong (i.e., that they should not be involved in the development process, but rather should miraculously go from L1 to L2 forms overnight).

interlanguage development, much in the same way that the acquisition of voice onset time exists along a temporal continuum.[51]

In sum, the study of interlanguage variants has several implications for interlanguage development. The production of uvular variants for /r/ suggests that interlanguage development is characteristic of L1 development, given similar uvular productions by L1 Spanish learners. Different variants for the same segment in different word-positions suggest that the early stages of interlanguage categories may be affected by prosodic structure. A limitation of studying variants based on native speaker features is that not all native speaker dialects may be represented; however, a benefit of categorizing all productions is that production errors may actually reveal patterns of interlanguage development. Additionally, the study of variants suggests that interlanguage forms may develop on a feature, rather than on a categorical, level.

9.3 Addressing Spanish experience and gender

The categorization of participants between beginning, intermediate, advanced, and experienced levels may falsely imply a linear snapshot or progression between L2 stages of development. That is, one might predict that students designated as beginning level may eventually improve their accuracy scores so they come to resemble scores of students at the advanced levels; however, this assumption may not be valid. Many of the advanced learners began Spanish in high school and were proficient enough in their language placement to begin university level Spanish courses beyond the first-year level. In contrast, the beginning learners were taking Spanish for the first time (i.e., potentially starting Spanish at a later age) and started with the first-year university Spanish program.

[51] To exemplify the problem of classifying interlanguage forms as transfer processes, we may hypothesize a scenario in which a student produces [ɾ] for all word-initial and word-medial /r/ targets. English has [ɾ] word-medially but not word-initially (with the exception of a phrase like *at Ed's* as pointed out by Major (1986)). So, is it a transfer process word-medially, yet a developmental process word-initially? A less problematic approach would be to treat all interlanguage productions along some continuum of development.

Thus, as shown from the language background survey, participant age of learning was confounded with Spanish experience.

Another factor potentially confounding with Spanish level is that the advanced learners were more likely to continue studying Spanish (e.g., studying abroad) than the beginning students, implying different levels of motivation that may affect pronunciation accuracy. Similarly, the advanced students may have had a Spanish major or minor whereas the beginning students could potentially have been taking Spanish because of a graduation requirement. An informal survey by a Spanish instructor (Fox, pc) found that more than three-fourths of the Spanish undergraduate majors at the University of Washington had entered into the major at the second-year level of Spanish, supporting inherent differences between subjects' motivation to learn Spanish.

Regarding anomalous Spanish backgrounds, two experienced level participants are of particular interest. Participant B-ExpL-F2 differed from the other experienced participants in that she began studying Spanish at the age of 5 and also lived for the longest period of time (8 years) in a Spanish speaking country. Thus, based on her language background, one would expect her productions and accuracy to be most similar to the native Spanish speakers. On the other end of the language experience continuum, participant B-ExpL-F7 was the only experienced level speaker not to have lived abroad and consequently one might expect her productions and accuracy to resemble patterns shown by the advanced level students (i.e., the next lower level). The accuracy scores for /r/ support the prediction that B-ExpL-F7 would not have native-like accuracy. She received scores of 26% for word-initial /r/ and 74% for word-medial /r/ (vs. participant B-ExpL-F2's scores of 90% and 100% for word-initial and word-medial /r/, respectively). In contrast, B-ExpL-F2 showed 67% accuracy for /d/ vs. B-ExpL-F7 who received 100% accuracy for /d/. Thus, these results do not support the idea that length of residence will necessarily lead to native-like pronunciation for all segments.

In regards to gender differences, comparison of accuracy scores suggested that females generally used more native-like productions of segments than their male counterparts, in the corresponding learner level. For all seven sound categories in this study, females at the learner level demonstrated higher average accuracy than average scores for males. The one exception is the overall raw accuracy score for /t/ in which both groups averaged 96% accuracy; however, when VOT was included as a criterion, the average score for males was reduced to 74% accuracy while the average scores for females was 80%, favoring females in this category, as well.

Gender differences were also observed in regards to the number of variants produced by the participants. For example, in the analysis of the five variants realized for /ɾ/ in onset cluster, both the learner females and the experienced females produced all five variant forms; however, the Type 4 (≈[r]; ≈[z̧,ɾ,r]) variant was not demonstrated by the learner males and the Type 4 and Type 5 (≈[ɾ]; ≈[d,ð]) variants were not revealed by the experienced males. For the learner groups, both genders evidenced the Type 2 (≈[ɹ]) variant most frequently; however, the males were more consistent with its use (78% vs. 46% for females). In addition, the experienced males demonstrated a higher frequency of accurate Type 1 (≈[ɾ]) forms (72%) vs. the females (67%). These results suggest that males may be more consistent (even though their variants may not be native-like) in their productions than females. This observation also held true for the learner levels for all sound categories in this study. In the case of word-medial /d, t, r, ɾ/, word-initial /r/, and onset cluster /ɾ/, females at the learner level demonstrated at least 1 more variant than the males. The difference between the number of productions between the males and females at the experienced level was not consistent for all sound categories, suggesting that the range of variants may be less affected with increased Spanish experience.

Accounting for apparent gender differences in accuracy scores and range of variation is difficult, given the fact that there were low numbers of males who volunteered for this study. Thus, the analysis of gender differences is limited to descriptions of trends, as more meaningful inferential analyses are not assumed reliable. Nevertheless, there are several possible explanations for the observed gender differences of scores and range of variants. First, the observed differences may be a result of individual differences given the low numbers of males compared to females (e.g., at the learner level, 6 males vs. 17 females). In effect, the 6 males may not be representative of the general population of males from the beginning level. Consequently, we might expect similar patterns for males and females if more males had participated in the study. Second, if one assumes that the observed differences *are* representative of results that would be found in the general population at the beginning level, then one would want to explain why females show higher accuracy as well as more variant forms than male scores and number of variant forms. These differences might be a reflection of different learning styles. That is, females may begin by producing a wider range of features (i.e., variants) before homing in on the native-like form(s) later on. In contrast, males may concentrate their attempts across fewer variants (even, if the variants are not native-like), revealing a greater consistency of forms. Thus, males may shift a more narrow distribution towards the correct target whereas females may only narrow their distribution only after discovering the native-like variants. Third, the differences may merely be a result of subject differences. That is, the females, due to their increased numbers, may represent a greater portion of the students with the highest aptitudes for acquiring pronunciation.

In sum, the results of this study have suggested effects of Spanish experience as well as gender on both accuracy scores and range of variants; however, the low number of observations per independent grouping, especially in the case of gender, does not allow for inferential comparisons of results. In addition, one must be careful to account for confounding factors such age of learning and motivation when assessing differences in target language experience. In the case of gender, the results of this study have suggested, at least for the learner level participants, that females show higher accuracy scores than males for all target sounds, even though females also showed a greater number of variants than their male counterparts. Regarding gender, future studies of interlanguage development should attempt to obtain balanced numbers of males and female participants, and whenever possible, large enough samples of each.

9.4 Interlanguage and L1 phonology

Importantly, the results of the current study suggest that prosodic structure (e.g., both word and syllable position) influences how and when a category will be produced. Assuming that interlanguage is a system that conforms to the rules of Language, the linguist's model of underlying sound categories, whether through a set of distinctive features (Chomsky and Halle 1968), geometric relations of features (Clements 1985), gestural modeling (Browman and Goldstein 1986), or probability distributions over representations of the phonetic space (Pierrehumbert 2001), needs to be able to account for effects beyond the segmental level, allowing for both robust trends across learners and individual patterns of variation. Regardless of one's theoretical perspective, the results of the current study suggest that interlanguage is a system shared by many learners and affected by target language experience and gender groupings.

The analysis of the variants in this study have implications for the initial state of interlanguage phonological development. This study has assumed the Full Transfer/Full Access hypothesis proposed by Schwartz and Sprouse (1996) which argues that the initial state of the L2 (i.e., interlanguage) is the L1. That is, the entire L1 system forms the basis of beginning interlanguage system. New interlanguage forms are created based on input and experience. In the current, L1 English forms were most frequent at the lower Spanish levels (e.g., the realization of [ɹ] for /r/ and aspirated [t]) while more experienced learners produced fewer L1 English forms, suggesting an initial transfer of L1 features followed later by interlanguage and target language forms. Importantly, the results demonstrate the rapidity at which interlanguage forms (i.e., variants not common to L1 English) emerged. For example, even after 1 week of Spanish instruction all 8 beginning level participants evidenced 2 or more variants out of the 7 identified types for /r/, and one participant realized 6 out of 7 types. Thus, even though the L1 features are assumed to be the starting point for interlanguage, the development of features occurs upon first exposure to the L2. In addition to the rapidity at which interlanguage develops, the current study's results suggest that different sound categories follow different schedules of development. Otherwise, we might have expected accuracy scores to be relatively the same for each participant for all categories.

The analysis of the native speaker productions of word-final /r/ also have implications for Bonet and Mascaro's (1997) claim that the tap is preferred by native speakers word-finally. In effect, the authors must make this claim in order to support their analysis that the [ɾ] and [r] realization is dependent on sonority structure; however, the acoustic data in Section 5.1.5 do not support the claim that word-final <r> is realized word-finally as a [ɾ].[52] It might be argued that identification of word-final <r> as [ɾ] would be better assessed by native speakers. Still, given the acoustic dissimilarities between word-medial /ɾ/ and word-final <r>, Bonet and Mascaro's claim is not supported.

9.5 Implications for determining accuracy

It must be recognized that the pronunciation analyzed in this study is a reflection of speech produced in a formal task. Participants were reading from word lists in a sound attenuated booth alongside an unfamiliar researcher. One might argue that laboratory speech is not a representative sample of "regular" L2 speech; however, in a typical classroom or even real world situation, pronunciation may not be the center of the subject's attention. Thus, one benefit of laboratory speech is that the participant is able to focus solely on his or her pronunciation and is given the best possible opportunity to pronounce L2 sounds. Furthermore, laboratory speech may have the benefit of limiting potential influences on pronunciation such as distracting cognitive processes (e.g., worrying about syntax) and potential sociolinguistic influences (e.g., frequently changing forms based on who is listening). A potential detriment of laboratory speech may be the sociolinguistic effect of nervousness that can accompany self-monitoring in a test lab setting.

---

[52] The mean duration of word-final <r> segments was 74ms (16.0 stdev).

While a two-tiered method of pronunciation accuracy assessment has more stringent accuracy requirements than either auditory or acoustic measures alone, the two tiered method of accuracy assessment may on occasion result in an unfair scoring of a production. For each sound category in this study, there was one or more variant types that revealed a discrepancy between the auditory and acoustic features. That is, based solely on the auditory signal, a production might have given the impression of a native-like production; however, the acoustic features may not have been consistent with native speaker acoustic baselines (or vice-versa). For example, participant B-ExpL-F6 produced what sounded like [ɾ] for the target word *miro* 'I see'; however, the waveform and spectrogram showed no clear stripe and in fact, produced visual signals almost identical to an approximated /d/. Thus, in the two-tiered assessment, she was awarded a score of "0" (inaccurate) for this production based on the visual inspection.

The overall study design also provides an additional benefit in the analysis of native speaker forms. While a two tiered method of accuracy assessment may lead to multiple categorizations of variants in the L2 productions, it also has the outcome of identifying native speaker variation that might not emerge from a single-tier analysis. Additionally, by assigning scores to native speakers, we can end up with a range of native-like accuracy that might avoid a ceiling effect of scores (i.e., having all native speakers at 100%). By having a range of scores for native speakers alongside a high number of participants (>30) parametric inferential analysis can provide additional insight into score distributions. Still, it's important to point out that the categorization of variants is dependent on the features established by the researcher.

In regards to the acoustic criteria, one might ask if the acoustic features correspond to actual perceptual cues. A joint perception-production study of interlanguage might begin to answer that question; however, it's not clear that the establishment of acoustic correlates for perception is absolutely necessary or even appropriate. Knightly (2000) found that native speakers had difficulty assessing segmental accuracy due to the influence of incorrectly produced surrounding sounds.

Thus, one shouldn't necessarily assume or require a production/perception correspondence in the selection of features.

Another potential benefit of a two tiered analysis is that accuracy assessment may indeed be more reliable. Given that there are two levels of inspection, multiple researchers may find greater reliability when assigning accuracy scores. Alternatively, the reliability between variant types may be expected to be lower if there are more opportunities for disagreement. Further investigation should test the extent of additional reliability gained by a two tiered analysis.

9.6 Recommendations for instructors

The lack of formal pronunciation instruction in texts and the classroom has been widely discussed (Arteaga 2000, Chela-Flores 2001, Wipf 1985, Wong 1985, Yule and Macdonald 1995) and, unfortunately, pronunciation is given little priority in the classroom by university level instructors (Harlow and Muyskens 1994). This may be partly due to publishers who are afraid that textbooks will be less likely to be adopted if a text contains too much depth in an area (e.g., phonetics) that may seem especially complex to the unfamiliar. The rise of the Communicative method of instruction (Terrell 1989) over the last 15 years has also de-emphasized pronunciation instruction in favor of other goals. Furthermore, there has been an unfounded belief among instructors that pronunciation is a skill that cannot be learned (Wong 1985). Nevertheless, recent investigations have found a positive correlation between formal instruction and pronunciation accuracy (Gonzalez-Bueno 1997, Munson 2001). Thus, formal pronunciation instruction, either through its absence or methods, can influence how and in which environments certain sounds are acquired. Based on the results of the current study, there are a number of recommendations that can be made for teachers in the instruction of L2 Spanish pronunciation.

First, teachers should be sure to emphasize the orthography to sound correspondence, and in particular, the correspondence of the grapheme <r> to the sounds of [ɾ] and [r]. As shown by the trill results in Section 7.2.1, students across all levels of Spanish were more likely to realize the multiple vibrant word-medially (when it is represented by two graphemes) than in word-initial position (where it is represented by a single grapheme). The lower scores for /r/ word-initially suggest that students may be unaware of the different orthographic representations of a trill and its word positions. In addition, the fact that some students were producing trills in place of taps as well as taps in place of trills suggests that there may be confusion regarding when and where <r> represents a tap vs. trill. Thus, instructors may ease the acquisition of taps and trills by placing greater emphasis on where to produce them.

Second, instructors may ease the acquisition of the trill by encouraging students to first practice trills outside the context of a high vowel environment. Section 7.2.2 found that students were more likely to produce a multiple vibrant trill in the context of a low or mid-vowel environment. That is, producing a trill in a high vowel environment appears to be more difficult which is also supported by the fact that there are more aerodynamic requirements for the production of a trill in a high vs. low vowel environment (Solé 2002). Thus, formal instruction materials that contain practice tokens may be best designed to first present students with 'easier' tokens (e.g., trills in low vowel environments) and then later have more advanced tasks to include more difficult tokens (e.g., trills in high vowel environments). By presenting easier tasks first, students may more quickly experience success and feel a greater sense of confidence in their pronunciation.

Third, instructors may take advantage of the acoustic representations of native speech forms in this thesis (e.g., waveforms and spectrograms) that visually illustrate differences between English and Spanish sounds. Both Gonzalez-Bueno (1997) and Munson (2001) used waveform and spectrograms to demonstrate differences between Spanish and English voice onset time. Both of their studies found that learners were more likely to improve their L2 timing patterns after exposure to instruction that included waveform and spectrogram samples. In the current study, there was a robust trend for lower level learners to have longer (i.e., less native-like) VOT. Teachers can incorporate materials such as the native speaker waveforms and spectrograms (see Section 5.1) and compare them to corresponding non-native waveforms and spectrograms (see Section 5.2). These materials may aid learners by giving them a visual representation of typical differences between the learner's L1 and the L2 target.

Fourth, the results of this study should encourage teachers to emphasize that variation exists within speakers and not just across dialects. Importantly, students shouldn't be expected to attain a level of pronunciation that is limited to a single precise form with a required set of features. As shown in Section 5.1, native speakers evidenced a range of features in their productions for any given target. For example, in the case of /d/ (Section 5.1.1), productions were realized as [ð] or as an approximated version [ð̞] (without frication). Nevertheless, textbooks (e.g., Van Patten et al. 1996) and teachers can falsely give the impression that there is only one 'correct' way of realizing a particular sound, or even that a specific dialect will always realize a target with the same features. Teachers can educate their students by making students aware of intra-speaker variation. In turn, students may not feel like they have to attain a precise and categorical form for any given target. Additionally, students may begin to pay more attention to native speaker variation when they are aware that it exists and that variation is in itself native-like.

The methodology for assessing accuracy in this study may provide instructors and researchers with clearly defined criteria for evaluating productions. Without a strategy of evaluation, pronunciation development cannot be clearly assessed. If evaluators are aware of the features that should be present in L2 productions (e.g., both duration and spectral features as in the case of the spirant target [ð]), then some errors might be a result of a missing feature in the realization rather than an incorrect representation (or its absence) altogether. Furthermore, if students are made aware of the appropriate auditory and acoustic features necessary for native-like productions, they may become better able to accurately produce L2 targets. It should be noted, however, that a potential limitation of applying auditory-acoustic accuracy assessment to L2 productions in a standard curriculum is the length of time required to evaluate even a single participant's speech. The analysis of a five minute recording which includes 63 Spanish target words and 18 English targets, can take several hours depending on the complexity of the productions (e.g., a /r/ realized as [ɹ] will have no durational measurements vs. a /r/ realized as a trill will have several measurable features). Thus, while linguistic research can and should take advantage of instrumental studies, the typical language instructor (who may have no background in phonetics) may not have the time or ability to incorporate phonetic analysis into pronunciation assessment.

9.7 Future study

The results of the current study argue that syllable structure has an effect on the realization of /ɾ/ and that word position has an effect on the realization of /r/. Thus, future studies might test whether the effects of these environments also hold true for other segments. In addition, future investigations should also be sure to control for additional prosodic effects such as syntactical boundaries. For example, Egido and Cooper (1980) found that the application of the flapping rule in English by native speakers was affected by a following embedded clause and by a following syntactic deletion site; however, the application of the flapping rule was not affected by a following NP or VP. Thus, future studies of L2 segment productions may want to obtain native speaker productions in a range of syntactic, word, and phonotactic environments to assure that L2 productions are appropriately compared within and across environments.

The current study has attempted to determine the extent to which L2 speakers differ from native speakers; however, a future question might ask *What are the acoustic features that contribute most to foreign accent?* It has long been argued (Delattre 1963) that obtaining acoustic measurements can help determine those features that contribute most to foreign accent. Thus, examination of spectral and temporal patterns in native speaker segments (or even L2 productions thought to be native-like) would allow for the creation of natural stimuli which, in turn, would allow for the identification of individual perceptual cues.

Regarding perception, researchers might want to determine whether L2 learners are using the same cues in identifying and discriminating between target contrasts as L1 Spanish speakers. If L2 learners are consistently using the same cues in the identification of L2 segments, additional support would be found for the systematicity of interlanguage development from L1 English to L2 Spanish. Furthermore, even when cues are established, not all cues equivalently assist in the identification of segments (Wright 2001), thus studies could ask whether learners weight cues to the same degree.

Future studies into L2 acquisition may also benefit from the collection of gestural configurations in the productions of L2 segments. Thus, in addition to collecting acoustic information, knowledge of motoric control and articulator placement could be obtained through x-ray technology. For example, it has been shown that some British English speakers produce taps with the tongue tip and others use the blade of their tongue (Bladon and Nolan 1977). For the current study, one could hypothesize that a "good" vs. "poor" triller may correlate with the portion of the learner's tongue that is being used to attempt the trill target. Assuming that native speakers are using the same portion of the tongue tip to produce a trill (which should first be determined), one could potentially predict good vs. poor trillers. Thus, if a learner normally produces a tap with the tongue tip, he or she may more quickly learn the trill whereas the learner that normally uses the tongue blade would have to learn a new manner *and* alter his or her articulator. A future study could use x-ray technology to test this hypothesis.

Future studies may want to replicate the analysis methods using different elicitation tasks; however, investigators should be careful when comparing cross-study results as interlanguage productions have been shown to vary based on the nature and formality of the elicitation task (Tarone 1983). Still, not all acoustic features are affected by elicitation task. Ladefoged et al. (1976) were not able to find any F1 and F2 differences between vowels produced in stressed mono-syllabic words across different speech styles, suggesting that some spectral features may be particularly resistant to variation across speech styles.

Future studies into the acquisition of trills might investigate differences in durations of trill productions. Carballo and Mendoza (2000) found longer durational measurements in trill occlusions and overall segment length for "poorer" trillers of L1 Spanish. L2 learners may reveal a similar trend in their productions which would support similar acquisition patterns in trills for both L1 and L2 learners of Spanish. Such investigation may show that "better" L2 trillers have shorter trills and shorter closures; however, methods should be clearly defined on the assessment of "poor" versus "good" trillers. An investigation similar to the current study could further test the relational difficulty of accurately producing /Cɾ/ and /ɾ/ by controlling for word-position. This study compared word-initial tap /#Cɾ/ with word-medial tap/VɾV/; however, because the results of the trill scores varied based on word-position, controlling for word-position with taps (e.g., *mi.ro* 'I look' vs. *mi.cro* 'bus') may result in stronger claims about the difficulty of a tap in a complex onset for L2 learners.

9.8 Summary

This study has addressed several questions relating to the relative degree of difficulty of word-medial /d,t,r,ɾ/, word-initial /r/ vs. word-medial /r/, and word-medial /ɾ/ vs. onset cluster /ɾ/. This study has defined accuracy based on the auditory and acoustic features found in native Spanish speaker productions. In order to account for the variation in the native speaker productions, L2 participant scores for each target were adjusted accordingly. The results found that accuracy scores followed two significant rankings (from lowest to highest) of /r <= d <= ɾ <= t/ and /r <= ɾ <= d <= t/; however, only the second ranking was found to be significant when scores for /t/ included a voice onset time criterion. These results suggest that /r/ is most difficult for learners and that /t/ is least difficult. Regarding /r/, there was a strong effect of word position and vowel height on accuracy scores. For productions of /ɾ/, there was a strong effect of syllable position on accuracy scores.

REFERENCES

Abramson, Arthur S., and Lisker, Leigh. 1973. Voice timing perception in Spanish word-initial stops. *Journal of Phonetics* 1:1-8.

Allison, Kathy Oppelt. 1993. A Computer-Assisted Comparative Acoustic and Perceptual Analysis of Spanish and English Intonation and its Application for Foreign Language Instruction, Applied Linguistics and Foreign Language Pedagogy, Pennsylvania State University: Ph.D. Dissertation.

Archibald, John. 1998. Second language phonology, phonetics, and typology. *Studies in Second Language Acquisition* 20:189-211.

Arslan, Levent M., and Hansen, John H. L. 1996. A study of temporal features and frequency characteristics on American English foreign accent. *Journal of the Acoustical Society of America* 102:28-40.

Arteaga, Deborah L. 2000. Articulatory phonetics in the first-year Spanish classroom. *Modern Language Journal* 84:339-354.

Asher, James J., and García, Ramiro. 1969. The optimal age to learn a foreign language. *Modern Language Journal* 53:334-342.

Beardsmore, Hugo Baetens. 1979. The recognition and tolerance level of bilingual speech. *Working Papers in Bilingualism* 19:116-128.

Beebe, L. 1980. Sociolinguistic variation and style shifting in second language acquisition. *Language Learning* 30:433-447.

Bergen, John J. 1974. A practical framework for teaching pronunciation in beginning Spanish courses. *Hispania* 57:479-483.

Best, Catherine T. 1995. A direct realist view of cross-language speech perception. In *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, ed. Winifred Strange, 171-204. Baltimore: York Press.

Birdsong, David. 2004. Second Language Acquisition and Ultimate Attainment. In *Handbook of Applied Linguistics*, eds. A. Davies and C. Elder, 82-105. London: Blackwell.

Bladon, R. A. W., and Nolan, F. J. 1977. A video-fluorographic investigation of tip and blade alveolars in English. *Journal of Phonetics* 5:185-193.

Bloch, Bernard. 1941. Phonemic overlapping. *American Speech* 16:278-284.

Boersma, Paul, and Hayes, Bruce. 2001. Empirical tests of the Gradual Learning Algorithm. *Linguistic Inquiry* 32:45-86.

Boersma, Paul, and Weenink, David. 2003. Praat: doing phonetics by computer (Version 4.0.1) [Computer program]. *Retrieved from* http://www.praat.org/.

Bonet, Eulália, and Mascaró, Joan. 1997. On the representation of contrasting rhotics. In *Issues in the Phonology and Morphology of the Major Iberian Languages*, ed. Fernando Martínez-Gil and Alfonso Morales-Front, 559-593. Washington, DC: GUP.

Bongaerts, Theo. 1999. Ultimate attainment in L2 pronunciation: The case of very advanced late L2 learners. In *Second language acquisition and the critical period hypothesis*, ed. D. Birdsong, 133-159. Mahwah: Lawrence Erlbaum.

Bosch, L. 1983. El desarrollo fonológico infantil: una prueba para su evaluación. *Anuario de Psicología* 28:87-114.

Boyce, Suzanne, and Epsy-Wilson, Carol Y. 1997. Coarticulatory stability in American English /r/. *Journal of the Acoustical Society of America* 101:3741-3753.

Bradley, Travis G., and Schmeiser, Ben. 2002. On the phonetic reality of Spanish /tap/ in complex onsets. Ms.

Browman, C.P., and Goldstein, L. 1986. Towards an articulatory phonology. In *Phonology Yearbook 3*, eds. C. Ewan and J. Anderson, 219-252. Cambridge: Cambridge University Press.

Carballo, Gloria, Mendoza, Elvira, and Valencia-Naranjo, Nieves. 1997. Interobserver agreement of perceived intelligibility of /r/ in children. *Perceptual and Motor Skills* 84:1099-1104.

Carballo, Gloria, and Mendoza, Elvira. 2000. Acoustic characteristics of trill productions by groups of Spanish children. *Clinical Linguistics & Phonetics* 14:3587-3601.

Cárdenas, D. N. 1958. The geographic distribution of the assibilated R, RR in Spanish America. *Orbis* 7:407-144.

Carlisle, Robert S. 1997. The modification of onsets in a markedness relationship: Testing the Interlanguage Structural Conformity Hypothesis. *Language Learning* 47:327-361.

Celce-Murcia, Marianne. 1977. Phonological factors in vocabulary acquisition: A case study of a two year old French-English bilingual. *Working Papers in Bilingualism* 13:27-41.

Cenoz, Jasone, and Garcia Lecumberri, Maria Luisa. 1999. The acquisition of English pronunciation: learners' views. *International Journal of Applied Linguistics* 9:3-17.

Chela-Flores, Bertha. 2001. Pronunciation and language learning: An integrative approach. *International Review of Applied Linguistics* 39:85-101.

Chomsky, N., and Halle, M. 1968. *The Sound Pattern of English*. New York: Harper & Row.

Clements, G. N. 1985. The geometry of phonological features. In *Phonology Yearbook 2*, 223-250.

Clements, G. N. 1990. The role of sonority cycle in core syllabification. In *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech*, eds. John Kingston and Mary E. Beckman, 283-333. Cambridge: CUP.

Cochrane, R. 1977. The acquisition of /r/ and /l/ by Japanese children and adults learning English as a second language. *Journal of Multilingual and Multicultural Development* 1:331-360.

Corder, S. P. 1967. The significance of learner's errors. *International Review of Applied Linguistics* 5:161-170.

Corder, S. P. 1978. Language-learner language. In *Understanding Second and Foreign Language Learning: Issues and Approaches*, ed. Jack C. Richards, 71-93. Rowley, MA: Newbury House.

Dalby, Jonathan, and Kewley-Port, Diane. 1999. Explicit pronunciation training using automatic speech recognition technology. *CALICO Journal* 16:425-445.

de Jong, Kenneth. 1998. Stress-related variation in the articulation of coda alveolar stops: Flapping revisited. *Journal of Phonetics* 26:283-310.

Delattre, Pierre. 1963. Research techniques for phonetic comparison of languages. *International Review of Applied Linguistics* 1:85-97.

Dickerson, Lonna J. 1975. The learner's interlanguage as a system of variable rules. *TESOL Quarterly* 9:401-407.

Dinnsen, D.A., and Eckman, F. R. 1975. A functional explanation of some phonological typologies. In *Functionalism*, ed. R. Grossman et al. Chicago: Chicago Linguistic Society.

Dunkel, H. 1948. *Second Language Learning*. Boston: Ginn & Company.

Eckman, Fred R. 1977. Markedness and the Contrastive Analysis Hypothesis. *Language Learning* 27:315-330.

Eckman, Fred R., and Iverson, Gregory K. 1997. Structure preservation in interlanguage phonology. In *Focus on Phonological Acquisition*, eds. S.J. Hannahs and M. Young-Scholten. Philadelphia: John Benjamins.

Egido, Carmen, and Cooper, William E. 1980. Blocking of alveolar flapping in speech production: the role of syntactic boundaries and deletion sites. *Journal of Phonetics* 8:175-184.

Einstein, Miriam. 1982. A study of social variation in adult second language acquisition. *Language Learning* 32:367-391.

Elliot, Raymond. 1995a. Field independence/dependence, hemispheric specialization, and attitude in relation to pronunciation accuracy in Spanish as a foreign language. *Modern Language Journal* 79:356-371.

Elliot, Raymond. 1995b. Foreign language phonology: Field independence, attitude, and the success of formal instruction in Spanish pronunciation. *The Modern Language Journal* 79:530-542.

Elliot, Raymond. 1997. On the teaching and acquisition of pronunciation within a communicative approach. *Hispania* 80:96-108.

Farhady, Hossein. 1982. Measures of language proficiency from the learner's perspective. *TESOL Quarterly* 16:43-59.

Flege, J.E. 1980. Phonetic approximation in second language acquisition. *Language Learning* 30:117-134.

Flege, J.E., and Fletcher, Kathryn L. 1992. Talker and listener effects on degree of perceived foreign accent. *Journal of the Acoustical Society of America* 91:370-389.

Flege, J.E., Munro, M. J., and MacKay, Ian R. 1995a. Factors affecting strength of perceived foreign accent in a second language. *Journal of the Acoustical Society of America* 97:3125-3134.

Flege, J.E., Takagi, Naoyuki, and Mann, Virginia. 1995b. Japanese adults can learn to produce English /r/ and /l/ accurately. *Language and Speech* 38:25-55.

Flege, J.E. 1998. The role of subject and phonetic variables in second-language learning. *The Chicago Linguistic Society* 34:213-232.

Flege, James E. 1995. Second language speech learning: Theory, findings and problems. In *Speech Perception and Linguistic Experience: Issues in Cross Language Research*, ed. Winifred Strange, 233-277. Baltimore: York Press.

Flege, James E., Frieda, Elaina M., and Nozawa, Takeshi. 1997. Amount of native-language (L1) use affects the pronunciation of an L2. *Journal of Phonetics* 25:169-186.

Flege, James Emil. 1986. A critical period for learning to pronounce foreign languages? *Applied Linguistics* 8:162-177.

Flege, James Emil, and Efting, Wieke. 1986. Linguistic and developmental effects on the production and perception of stop consonants. *Phonetica* 43:155-171.

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

Flege, James Emil, Frieda, Elaina M., Walley, Amanda C., and Randazza, Lauren A. 1998. Lexical factors and segmental accuracy in second language speech production. *Studies in Second Language Acquisition* 20:155-187.

Gili, Samuel. 1921. La r simple en la pronunciación española. *Revista de Filología Española* 8:271-280.

Gonzalez, M. J. 1989. Análisis del desarrolo fonológico en sujetos malagueños. *Infancia y Aprendizaje* 48:7-24.

Gonzalez-Bueno, Mañuela. 1997. The effects of formal instruction on the acquisition of Spanish stop consonants. In *Contemporary perspectives on the acquisition of Spanish*, ed. William R. Glass, 57-75. Somerville, MA: Cascadilla Press.

Good, Phillip. 2000. *Permutation tests: A practical guide to resampling methods for testing hypotheses*. New York: Springer.

Gradman, H. 1971. Limitations of contrastive analysis predictions. *Working Papers in Linguistics* 3:11-15.

Greenberg, J. 1978. Some generalizations concerning initial and final consonant clusters. In *Universals of Human Language*, eds. J. Greenberg, C. Ferguson and E. Moravcsik, 243-279. Stanford, CA: Stanford University Press.

Guión, Susan G., Flege, James E., and Loftin, Jonathan D. 2000. The effect of L1 use on pronunciation in Quichua-Spanish bilinguals. *Journal of Phonetics* 28:27-42.

Guiora, Alexander Z., Brannon, Robert C. L., and Dull, Cecelia Y. 1972. Empathy and second language learning. *Language Learning* 22:111-130.

Hagiwara, Robert. 1995. Acoustic realizations of American /r/ as produced by women and men. *UCLA Working Papers in Phonetics* 90:1-187.

Hammond, R., and Flege, James E. 1989. The acquisition of second language phonological systems in a communicative framework - the role of attitudes and experience. *Romance Languages Annual* 1:671-676.

Harlow, Linda L., and Muyskens, Judith. 1994. Priorities for intermediate-level language instruction. *Modern Language Journal* 8:131-154.

Harris, James W. 1969. *Spanish Phonology*. Cambridge: MIT Press.

Harris, James W. 1983. *Syllable Structure and Stress in Spanish*. Cambridge, MA: MIT Press.

Hashi, Michiko, Honda, Kiyoshi, and Westbury, John R. 2003. Time-varying acoustic and articulatory characteristics of American English [upside-down-r]: a cross-speaker study. *Journal of Phonetics* 31:3-22.

Haugen, Einar. 1938. Notes on voiced T in American English. *Dialect Notes* 6:627-634.

Hill, J. 1970. Foreign accents, language acquisition and cerebral dominance revisited. *Language Learning* 20:237-248.

Horna, Josephine Eva. 1998. An Investigation into the Acoustics of American English Flaps, with a Secondary Emphasis on Spanish Flaps in Fluent Speech, Linguistics, New York University: Dissertation.

Jimenez, B. 1987. Acquisition of Spanish consonants in children aged 3-5 years, 7 months. Language, Speech, and Hearing Services in Schools:344-356.

Jongman, Allard, and Blumstein, Sheila E. 1985. Acoustic properties for dental and alveolar stop consonants: a cross-language study. *Journal of Phonetics* 13:235-251.

Kelm, Orlando. 1987. An acoustic study on the differences of contrastive emphasis between native and non-native Spanish speakers. *Hispania* 70:627-633.

Kerswill, Paul, and Wright, Susan. 1990. The validity of phonetic transcription: Limitations of a sociolinguistic research tool. *Language Variation and Change* 2:255-275.

Kim, Duk-Young, and Jung, Yuntaek. 1998. Adult ESL Korean speakers' interlanguage acquisition of consonant clusters. *SECOL Review* 22:150-169.

Knightly, Leah Mirasol. 2000. Benefits of overhearing Spanish: Implications for second language pronunciation, Psychology, University of California, Los Angeles: Dissertation.

Krashen, Stephen D., Long, Michael H., and Scarcella, Robin C. 1982. Age, rate, and eventual attainment in second language acquisition. In *Child-adult differences in second language acquisition*, eds. Stephen D. Krashen and et al., 161-172. Rowley, MA: Newbury House.

Kuhl, Patricia, and Iverson, Paul. 1995. Linguistic Experience and the "Perceptual Magnet Effect". In *Speech Perception and Linguistic Experience*, ed. Winifred Strange, 121-154. Baltimore: York Press.

Labov, William. 1966. *The Social Stratification of English in New York City*. Washington, D.C.: Center for Applied Linguistics.

Labov, William. 1972. *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press.

Ladefoged, Peter, Kameny, Iris, and Brackenridge, William. 1976. Acoustic effects of style of speech. *Journal of the Acoustical Society of America* 59:228-231.

Ladefoged, Peter, and Maddieson, Ian. 1996. *The Sounds of the World's Languages*. Oxford: Blackwell.

Lado, R. 1957. *Linguistics Across Cultures: Applied Linguistics for Language Teachers*. Ann Arbor: University of Michigan Press.

Larsen-Freeman, Diane, and Long, Michael H. 1991. *Introduction to Second Language Acquisition Research*. New York: Longman.

Lefkowitz, Natalie, and Hedgecock, John. 2002. Sound barriers: influences of social prestige, peer pressure and teacher (dis)approval on FL oral performance. *Language Teaching Research* 6:223-244.

Lehiste, Ilse. 1970. *Suprasegmentals*. Cambridge, MA: M.I.T. Press.

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

Lin, Yuh-Huey. 2001. Syllable simplification strategies: a stylistic perspective. *Language Learning* 51:681-718.

Lipski, John. 1990. Spanish taps and trills: Phonological structure of an isolated opposition. *Folia Linguistica* 24:153-174.

Lisker, Leigh, and Abramson, Arthur S. 1964. A cross-language study of voicing in initial stops: Acoustical measurements. *Word* 20:384-422.

Macken, Marlys A. 1980. The acquisition of stop systems: a cross-linguistic perspective. In *Child Phonology: Perception and Production*, eds. G. Yeni-Komshian, J.F. Kavanagh and C. Ferguson, 143-168. NY: Academic Press, Inc.

Macken, Marlys A., and Ferguson, Charles A. 1987. Phonological universals in language acquisition. In *Interlanguage Phonology: The Acquisition of a Second Language Sound System*, eds. Georgette Ioup and Steven Weinberger, 3-22. Cambridge, MA: Newbury House.

Magloire, Joe-umlaut-l, and Green, Kerry P. 1999. A cross-language comparison of speaking rate effects on the production of voice onset time in English and Spanish. *Phonetica* 56:158-185.

Major, Roy C. 1986. The Ontogeny Model: evidence from L2 acquisition of Spanish r. *Language Learning* 36:453-504.

Major, Roy C. 1992. Losing English as a first language. *The Modern Language Journal* 76:190-208.

Malmberg, Bertil. 1965. *Estudios de fonética hispánica*. Madrid: Consejo Superior de Investigaciones Científicas.

Manrique, Ana María Borzone de, and Massone, María Ignacia. 1981. Acoustic analysis and perception of Spanish fricative consonants. *Journal of the Acoustical Society of America* 69.

Massone, M. I. 1988. Estudio acústico y perceptivo de las consonantes nasales y líquidas de español. *Estudios de Fonética Experimental* III:13-34.

Mateus, Maria Helena, and d'Andrade, Ernesto. 2002. *The Phonology of Portuguese*. Oxford: Oxford University Press.

McGowan, Richard S., Nittrouer, Susan, and Manning, Carol J. 2004. Development of [approximant-<r>] in young, Midwestern, American children. *Journal of the Acoustical Society of America* 115:871-884.

Medina-Rivera, Antonio. 1999. Variación fonológico y estilística en el español de Puerto Rico. *Hispania* 82:529-541.

Mendez, Antonio. 1982. Production of American English and Spanish vowels. *Language and Speech* 25:191-197.

Monnot, Michel, and Freeman, Michel. 1972. A comparison of Spanish single-tap /r/ with American /t/ and /d/ in post-stress intervocalic position. In *Papers in Linguistics and Phonetics in Memory of Pierre Delattre*, ed. A. Valdman, 409-416. The Hague: Mouton.

Moreno de Alba, J. 1972. Frecuencias de la asibilación de /r/ y /rr/ en México. *Nueva Revista de Filología Española* 21:363-370.

Mota, C. de la. 1990. A study of [r] in spontaneous speech. *Proceedings of the XII International Congress of Phonetic Sciences* 4.

Munson, Jeremy C. 2001. Formal pronunciation instruction for the beginning adult learner of Spanish: Acoustic measures and criteria of accurate L2 phonology, Department of Linguistics, University of Washington: unpublished Masters thesis.

Navarro Tomás, T. 1918. *Manual de Pronunciación Española*. Madrid: Centro de Estudios Históricos.

Nemser, W. 1971. Approximative systems of foreign language learners. *International Review of Applied Linguistics* 9:115-123.

Neufield, Gerald G. 1980. On the adult's ability to acquire phonology. *TESOL Quarterly* 14:285-298.

Nuñez Cedeño, Rafael A. 1987. Intervocalic /d/ rhotacism in Dominican Spanish: A non linear analysis. *Hispania* 70:363-368.

Nuñez Cedeño, Rafael A. 1989. La /R/, único fonema vibrante del español: Datos del Caribe. *Anuario de Lingüística Hispánica* 5:153-171.

Oyama, Susan. 1982. A sensitive period for the acquisition of a nonnative phonological system. In *Child-adult differences in second language acquisition*, eds. Stephen D. Krashen, Michael H. Long and Robin C. Scarcella, 20-39. Rowley, MA: Newbury House Publishers, Inc.

Patkowski, Mark S. 1990. Age and accent in a second language: A reply to James Emil Flege. *Applied Linguistics* 11:73-89.

Patkowski, Mark S. 2003. Laterality effects in multilinguals during speech production under the concurrent task paradigm: Another test of the age of acquisition hypothesis. *International Review of Applied Linguistics* 41:175-200.

Pennington, Martha C., and Richards, Jack C. 1986. Pronunciation Revisited. *TESOL Quarterly* 20:207-225.

Pierrehumbert, Janet B. 2001. Stochastic phonology. *Glot International* 5:195-207.

Piske, Thorsten, MacKay, Ian R., and Flege, James E. 2001. Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics* 29:191-215.

Price, Patti Jo. 1981. A cross-linguistic study of Flaps in Japanese and in American English, University of Pennsylvania: Unpublished doctoral dissertation.

Purcell, E., and Suter, R. 1980. Predictors of pronunciation accuracy: Reexamination. *Language Learning* 30:271-287.

Quilis, Antonio. 1970. El elemento esvarabático en los grupos [pr, br, tr...]. In *Phonetique et Linguistique Romaines: Melanges offerts a M. Georges Straka*, 99-104. Lyon-Strasbourg: Societe de Linguistique Roman.

Quilis, Antonio, and Carril, Ramón B. 1971. Análisis acústico de [r-hatchek] en algunas zonas de hispanoamérica. *Revista de Filología Española* 54:271-316.

Quilis, Antonio, and Fernández, J. 1973. *Curso de fonética y fonología españolas*. Madrid: Instituto Miguel de Cervantes.

Quilis, Antonio. 1981. *Fonética Acústica de la Lengua Española*. Madrid: Editorial Gredos.

Quilis, Antonio. 1993. *Tratado de Fonología y Fonética Española*. Madrid: Editorial Gredos, S.A.

Recasens, D. 1987. An acoustic analysis of V-to-C and V-to-V coarticulatory effects in Catalan and Spanish VCV sequences. *Journal of Phonetics* 15:299-312.

Recasens, D. 1991. On the production characteristics of apicoalveolar taps and trills. *Journal of Phonetics* 19:267-280.

Reeder, Jeffrey T. 1997. Mimephonic ability and phonological performance in adult learners of Spanish. In *Contemporary perspectives on the acquisition of Spanish*, ed. William R. Glass, 77-90. Somerville, MA: Cascadilla Press.

Schacter, J. 1974. An error in error analysis. *Language Learning* 24:205-214.

Schumman, John H. 1978. Social and psychological factors in second language acquisition. In *Understanding Second and Foreign Language Learning: Issues and Approaches*, ed. Jack C. Richards, 163-178. Rowley, MA: Newbury House.

Schwartz, Bonnie, and Sprouse, Rex. 1996. L2 cognitive states and the Full Transfer/Full Access model. *Second Language Research* 12:40-72.

Scovel. 1969. Foreign accent, language acquisition, and cerebral dominance. *Language Learning* 19:245-253.

Selinker, Larry. 1969. Language Transfer. *General Linguistics* 9:67-92.

Selinker, Larry. 1972. Interlanguage. *International Review of Applied Linguistics* 10:209-231.

Sharf, Donald J. 1960. Distinctiveness of 'Voiced T' words. *American Speech* 35:105-109.

Simoes, Antonio R. M. 1996. Phonetics in Second Language Acquisition: An acoustic study of fluency in Adult Learners of Spanish. *Hispania* 79:87-95.

Smit, Ute. 2002. The interaction of motivation and achievement in advanced EFL pronunciation learners. *International Review of Applied Linguistics* 40:89-116.

Snow, Catherine E., and Hoefnagel-Höhle, Marian. 1982a. Age differences in the pronunciation of foreign sounds. In *Child-Adult Differences in Second Language Acquisition*, eds. Stephen D. Krashen, Michael H. Long and Robin C. Scarcella, 84-92. Rowley, MA: Newbury House Publishers, Inc.

Snow, Catherine E., and Hoefnagel-Höhle, Marian. 1982b. The Critical Period for language acquisition: Evidence from second language learning. In *Child-Adult Differences in Second Language Acquisition*, eds. Stephen D. Krashen, Michael H. Long and Robin C. Scarcella, 93-111. Rowley, MA: Newbury House Publishers, Inc.

Solé, María-Josep. 2002. Aerodynamic characteristics of trills and phonological patterning. *Journal of Phonetics* 30:655-688.

Southwood, and Flege, J.E. 1999. Scaling foreign accent: direct magnitude estimation versus interval scaling. *Clinical Linguistics & Phonetics* 13:335-349.

Stevens, John J. 2000. On the labiodental pronunciation of Spanish /b/ among teachers of Spanish as a second language. *Hispania* 83:139-149.

Stockwell, Robert P., and Bowen, J. Donald. 1965. *The Sounds of English and Spanish*. Chicago: University of Chicago Press.

Stoel, Caroline. 1974. The Acquisition of Liquids in Spanish, Linguistics, Stanford University: Unpublished doctoral dissertation.

Stoel-Gammon, Carol. 1985. Phonetic Inventories, 15-24 months: A longitudinal study. *Journal of Speech and Hearing Research* 28:505-512.

Stokes, Jeffrey D. 2001. Factors in the Acquisition of Spanish Pronunciation. *Review of Applied Linguistics* 131-132:63-84.

Suter, Richard W. 1976. Predictors of pronunciation accuracy in second language learning. *Language Learning* 26:233-253.

Tarone, Elaine E. 1978. The phonology of interlanguage. In *Understanding Second and Foreign Language Learning: Issues and Approaches*, ed. Jack C. Richards, 15-33. Rowley, MA: Newbury House.

Tarone, Elaine E. 1980. Some influences on the syllable structure of interlanguage phonology. *International Review of Applied Linguistics* 18:139-152.

Tarone, Elaine E. 1983. On the variability of interlanguage systems. *Applied Linguistics* 4:142-163.

Terrell, Tracy. 1989. Teaching Spanish pronunciation in a communicative approach, ed. Peter C. et al. Bjarkman, 196-214.

Torreblanca, Maximo. 1984. The asibilación de 'r' y 'rr' en la lengua española. *Hispania* 67:614-616.

Umeda, Noriko. 1975. Vowel duration in American English. *Journal of the Acoustical Society of America* 58:434-445.

Van Patten, B., Lee, J., and Ballman, T. 1996. *¿Sabías que... ?: Beginning Spanish*. New York: McGraw-Hill.

Vanderweide, T. 1994. Government phonology and principles of L1 syllabification, Unpublished master's thesis, University of Calgary.

Warsi, Jilani S. 2002. Effects of visual instruction on second language productive phonology, Linguistics, Boston University: Dissertation.

Weinberger, Steven H. 1987. The realization of syllable context on linguistic simplification. In *Interlanguage Phonology: The Acquisition of a Second Language Sound System*, eds. Georgette Ioup and Steven Weinberger, 401-417. Cambridge, MA: Newbury House.

Widdison, K. 1998. Phonetics and phonology: Phonetic motivation for variation in Spanish trills. *Orbis* 40:51-61.

Widdison, K. 2004. The perceptual awareness of vowel fragments appearing in Spanish Cr and CC environments.

Williams, Lee. 1977. The voicing contrast in Spanish. *Journal of Phonetics* 5:169-184.

Wipf, Joseph A. 1985. Towards improving second language pronunciation. *Die Unterrichspraxis* 18:55-63.

Wong, Rita. 1985. Does pronunciation teaching have a place in the communicative classroom? In *Georgetown University Roundtable on Languages and Linguistics*, 226-236. Washington, DC: Georgetown University Press.

Wright, Richard. 2001. Perceptual cues in contrast maintenance. In *The Role of Speech Perception in Phonology*, eds. Johnson Keith and Elizabeth Hume.

Yeni-Komshian, G.H., Flege, J.E., and Liu, S. 2000. Pronunciation proficiency in the first and second languages of Korean-English bilinguals. *Bilingualism, Language and Cognition* 3:131-149.

Young, Richard. 1988. Variation and the interlanguage hypothesis. *Studies in Second Language Acquisition* 10:281-302.

Yule, George, and Macdonald, Doris. 1995. The different effects of pronunciation teaching. *International Review of Applied Linguistics in Language Teaching* 33:345-350.

Zampini, Mary. 1994. The role of native language transfer and task formality in the acquisition of Spanish spirantization. *Hispania*:470-481.

Zampini, Mary. 1998. L2 Spanish spirantization: A prosodic analysis and pedagogical implications. *Hispanic Linguistics* 10:154-188.

Zlotchew, Clark M. 1974. The transformation of the multiple vibrant to the fricative velar in the Spanish of Puerto Rico. *Orbis* 23:81-84.

Zue, Victor W., and Laferriere, Martha. 1979. Acoustic study of medial /t,d/ in American English. *Journal of the Acoustical Society of America* 66:1039-1050.

Appendix A: Participant language background survey

1	Gender: *male* *female*					
2	Age: ___________					
3a	Is English your native language?		*yes*	*no*		
3b	Is Spanish your native language?		*yes*	*no*		
4	Do you speak or understand any languages other than English?		*yes*	*no*		
	if **yes,** which language(s)?	1	___________			
		2	___________			
		3	___________			
5	Have you ever studied any foreign language before?		*yes*	*no*		
	if **yes,** which language(s)?	1	___________		how long?	___________
		2	___________		how long?	___________
		3	___________		how long?	___________
6	Have you studied Spanish at UW before at the...					
	100 level?	1	*yes*	*no*	how many quarters?	_____
	200 level?	2	*yes*	*no*	how many quarters?	_____
	300 level?	3	*yes*	*no*	how many quarters?	_____
7	Have you ever lived in other countries?		*yes*	*no*		
	if **yes,** where?	1	___________		how long?	___________
		2	___________		how long?	___________
		3	___________		how long?	___________
8	Have you ever traveled for vacation to a Spanish speaking country?		*yes*	*no*		
	if **yes,** which one(s)?	1	___________		how long?	___________
		2	___________		how long?	___________
		3	___________		how long?	___________
9	Have you ever lived in a Spanish speaking community?		*yes*	*no*		
	if **yes,**		how long?	___________		
10	Did you grow up listening to or speaking Spanish on a regular basis?		*yes*	*no*		
	if **yes,**		please explain briefly:			

Appendix B: Sample Spanish word list

Dice ojo también.
Habla azul también.
Ya vi pescado también.
Ya vi sacar también.
Dice subir también.
Dice prado también.
Habla rato también.
Habla toro también.
Ya vi para también.
Dice remo también.
Habla carro también.
Dice churro también.
Dice prisa también.
Dice pido también.
Habla beber también.
Dice bota también.
Ya vi dado también.
Habla dato también.
Dice pito también.
Dice perro también.
Habla miro también.
Ya vi risa también.
Dice todo también.
Habla preso también.
Habla deseo también.
Dice papel también.
Ya vi gato también.

Appendix C: Sample English word list

He'll see cousin today.

He'll say listen today.

He saw timid today.

He'll see needy today.

He'll see payroll today.

He'll see raider today.

He saw eddy today.

He saw starry today.

He'll say aware today.

He'll say eighty today.

He'll say teary today.

He saw gray today.

He saw duty today.

He'll say haughty today.

He'll see robber today.

He saw afar today.

He'll see unclear today.

He'll see relay today.

He'll say body today.

He saw treaty today.

He'll say draw today.

He saw seven today.

He'll see sample today.

He'll say student today.

He'll see certain today.

He saw yellow today.

He'll say sixty today.

Appendix D: Results of individual native speaker productions of word-medial /d/. Standard deviation is shown in parentheses.

Native Speaker	n	Mean Duration (ms)	Percent "spirantized" productions	Percent "approximated" productions
NS-F1	9	47 (10)	100%	0%
NS-F2	9	43 (11)	100%	0%
NS-F3*	9	44 (14)	100%	0%
NS-F4	9	53 (5)	100%	0%
NS-F5	7	53 (8)	78%	22%
NS-F6	5	45 (2)	56%	44%
NS-M1	2	54 (11)	22%	78%
NS-M2	8	42 (9)	89%	11%
NS-M3	6	58 (4)	67%	33%
NS-M4	9	58 (12)	100%	0%
NS-M5	9	-- --	0%	100%
Average (across all tokens)	73	49 (11)	73%	26%

* Participant NS_F3 had one spirant production that measured 30ms in duration.

Appendix E: Results of individual native speaker productions of word-medial /t/. Standard deviation is shown in parentheses.

Native Speaker	n	Stop closure duration (SCD) (ms)		Voice onset time (VOT) (ms)		SCD:VOT (VOT = 1)	
NS-F1	9	70	(12)	25	(6)	2.9	(.4)
NS-F2	9	108	(12)	15	(3)	7.5	(2.4)
NS-F3	9	94	(13)	17	(3)	5.7	(1.4)
NS-F4	9	76	(8)	21	(6)	3.9	(1.2)
NS-F5	9	99	(12)	19	(3)	5.5	(.9)
NS-F6	9	91	(17)	21	(4)	4.5	(1.3)
NS-M1	7	50	(11)	29	(6)	1.8	(.6)
NS-M2	9	79	(7)	28	(5)	2.9	(.7)
NS-M3	9	72	(9)	26	(3)	2.8	(.5)
NS-M4	9	90	(12)	25	(8)	4.0	(1.3)
NS-M5	9	84	(13)	24	(6)	3.7	(1.3)
Average (across all tokens)	97	84	(18)	23	(6)	4.2	(1.9)

Appendix F: Results of individual native speaker productions of word-medial /r/. Standard deviation is shown in parentheses.

Native Speaker	n	Overall duration (ms)		Degree of voicing		Number of stripes	
NS-F1	9	75	(9)	1.3	(0.7)	2.7	(1.0)
NS-F2	9	104	(20)	1.8	(0.7)	2.6	(1.5)
NS-F3	9	70	(15)	1.4	(0.9)	1.4	(1.1)
NS-F4	9	73	(17)	1.7	(0.7)	2.1	(0.9)
NS-F5	9	64	(15)	2.0	(0.0)	2.1	(0.3)
NS-F6	9	74	(13)	1.9	(0.3)	2.2	(0.4)
NS-M1	9	63	(4)	2.0	(0.0)	2.0	(0.0)
NS-M2	9	79	(12)	2.0	(0.0)	2.2	(0.7)
NS-M3	9	76	(15)	2.0	(0.0)	1.9	(0.8)
NS-M4	9	77	(12)	1.1	(0.9)	2.2	(0.8)
NS-M5	9	78	(22)	1.8	(0.4)	2.2	(0.4)
Average (across all tokens)	99	76	(17)	1.7	(0.6)	2.2	(0.8)

Appendix G: Results of individual native speaker productions of word-initial /r/. Standard deviation is shown in parentheses.

Native Speaker	n	Overall duration (ms)		Degree of voicing		Number of stripes	
NS-F1	9	76	(19)	1.6	(0.5)	3.0	(0.7)
NS-F2	9	114	(31)	1.6	(0.7)	3.6	(1.8)
NS-F3	9	93	(18)	1.3	(0.7)	1.9	(1.5)
NS-F4	9	110	(15)	1.3	(0.9)	3.4	(0.5)
NS-F5	9	90	(7)	1.6	(0.5)	2.9	(0.3)
NS-F6	9	80	(23)	1.4	(0.9)	2.1	(1.4)
NS-M1	9	73	(25)	2.0	(0.0)	2.3	(0.7)
NS-M2	9	84	(17)	1.9	(0.3)	2.6	(0.7)
NS-M3	9	93	(24)	1.9	(0.3)	1.7	(1.1)
NS-M4	9	80	(14)	1.0	(1.0)	1.4	(1.1)
NS-M5	9	77	(15)	1.8	(0.4)	2.2	(0.7)
Average (across all tokens)	99	88	(23)	1.6	(0.7)	2.5	(1.2)

Appendix H: Results of individual native speaker productions of word-medial /ɾ/. Standard deviation is shown in parentheses.

Native Speaker	n	Overall duration (ms)	
NS-F1	9	21	(6.1)
NS-F2	9	19	(4.0)
NS-F3	9	21	(3.3)
NS-F4	9	23	(4.3)
NS-F5	8	26	(6.0)
NS-F6	9	28	(4.6)
NS-M1	9	19	(2.0)
NS-M2	9	18	(2.0)
NS-M3	9	27	(3.0)
NS-M4	9	28	(2.9)
NS-M5	9	26	(6.0)
Total	98	23	(5.4)

Appendix I: Results of individual native speaker productions of onset cluster /ɾ/. Standard deviation is shown in parentheses.

Native Speaker	n	Burst to vowel duration (ms)		n	Svarabhakti duration (ms)		n	tap closure to vowel duration (ms)	
NS-F1	9	62	(17.9)	8	13	(2.3)	8	20	(9.0)
NS-F2	9	59	(9.0)	9	17	(5.8)	9	15	(4.1)
NS-F3	9	54	(10.3)	9	22	(6.5)	9	14	(3.7)
NS-F4	9	56	(9.9)	9	24	(9.1)	9	18	(5.5)
NS-F5	9	59	(7.9)	9	27	(9.3)	9	19	(3.9)
NS-F6	9	48	(14.1)	7	16	(3.7)	8	22	(4.0)
NS-M1	9	61	(18.9)	8	24	(6.9)	8	19	(3.3)
NS-M2	9	58	(22.5)	5	28	(10.9)	5	17	(2.4)
NS-M3	9	50	(7.2)	7	22	(5.5)	7	16	(3.8)
NS-M4	9	60	(6.0)	8	22	(3.5)	8	27	(9.0)
NS-M5	9	33	(14.8)	4	15	(2.7)	4	17	(1.0)
Average (across all tokens)	99	55	(15.2)	83	21	(7.8)	84	19	(6.1)

Appendix J: All participant scores by session, overall raw accuracy (Raw acc.) and overall adjusted accuracy (Adjusted acc.)

Participant	Session	medial /d/	medial /t/	medial /r/	initial /r/	medial /ɾ/	onset cl. /ɾ/	SVE [ᵊɾ]
A-Beg-M1	S1	33%	89%	0%	0%	89%	0%	0%
	S2	44%	100%	0%	0%	78%	0%	0%
	S3	0%	100%	0%	0%	78%	11%	11%
	Raw acc.	26%	96%	0%	0%	81%	4%	4%
	Adjusted acc.	**26%**	**98%**	**0%**	**0%**	**88%**	**5%**	--
A-Beg-M2	S1	0%	67%	0%	0%	67%	0%	0%
	S2	0%	89%	0%	0%	89%	100%	0%
	S3	11%	89%	0%	0%	67%	0%	0%
	Raw accuracy	4%	81%	0%	0%	74%	33%	0%
	Adjusted acc.	**4%**	**83%**	**0%**	**0%**	**80%**	**41%**	--
A-Beg-F1	S1	33%	89%	11%	0%	11%	0%	0%
	S2	56%	89%	89%	44%	11%	11%	0%
	S3	78%	100%	100%	100%	0%	11%	11%
	Raw accuracy	56%	93%	67%	48%	7%	7%	4%
	Adjusted acc.	**56%**	**94%**	**75%**	**53%**	**8%**	**9%**	--
A-Beg-F2	S1	67%	100%	0%	0%	33%	0%	0%
	S2	67%	100%	0%	0%	22%	0%	0%
	S3	67%	100%	0%	0%	33%	0%	0%
	Raw accuracy	67%	100%	0%	0%	30%	0%	0%
	Adjusted acc.	**67%**	**100%**	**0%**	**0%**	**32%**	**0%**	--
A-Int-M1	S1	0%	100%	0%	0%	0%	0%	0%
	S2	0%	100%	0%	0%	0%	0%	0%
	S3	0%	100%	0%	0%	0%	0%	0%
	Raw accuracy	0%	100%	0%	0%	0%	0%	0%
	Adjusted acc.	**0%**	**100%**	**0%**	**0%**	**0%**	**0%**	--
A-Int-M2	S1	11%	100%	100%	0%	78%	11%	0%
	S2	0%	100%	78%	0%	78%	11%	11%
	S3	22%	100%	89%	0%	67%	33%	0%
	Raw accuracy	11%	100%	89%	0%	74%	19%	4%
	Adjusted acc.	**11%**	**100%**	**100%**	**0%**	**80%**	**23%**	--
A-Int-F1	S1	100%	100%	0%	0%	44%	0%	0%
	S2	100%	100%	0%	0%	44%	0%	0%
	S3	100%	100%	0%	0%	44%	0%	0%
	Raw accuracy	100%	100%	0%	0%	44%	0%	0%
	Adjusted acc.	**100%**	**100%**	**0%**	**0%**	**48%**	**0%**	--

Appendix E: (continued)

Participant	Session	medial /d/	medial /t/	medial /r/	initial /r/	medial /ɾ/	onset cl. /ɾ/	SVE [ᵊɾ]
A-Int-F2	S1	0%	67%	0%	0%	0%	0%	0%
	S2	0%	89%	0%	0%	11%	0%	0%
	S3	0%	100%	0%	0%	11%	0%	0%
	Raw accuracy	0%	85%	0%	0%	7%	0%	0%
	Adjusted acc.	**0%**	**87%**	**0%**	**0%**	**8%**	**0%**	--
A-Int-F3	S1	22%	100%	0%	0%	0%	0%	0%
	S2	22%	100%	0%	0%	0%	0%	0%
	S3	0%	100%	0%	0%	0%	0%	0%
	Raw accuracy	15%	100%	0%	0%	0%	0%	0%
	Adjusted acc.	**15%**	**100%**	**0%**	**0%**	**0%**	**0%**	--
A-Adv-M1	S1	22%	89%	0%	0%	33%	22%	0%
	S2	33%	89%	0%	0%	22%	33%	0%
	S3	44%	100%	0%	0%	11%	22%	0%
	Raw accuracy	33%	93%	0%	0%	22%	26%	0%
	Adjusted acc.	**34%**	**94%**	**0%**	**0%**	**24%**	**32%**	--
A-Adv-M2	S1	0%	100%	0%	0%	0%	0%	0%
	S2	0%	100%	0%	0%	0%	0%	0%
	S3	0%	100%	0%	0%	0%	0%	0%
	Raw accuracy	0%	100%	0%	0%	0%	0%	0%
	Adjusted acc.	**0%**	**100%**	**0%**	**0%**	**0%**	**0%**	--
A-Adv-F1	S1	100%	100%	56%	0%	100%	100%	11%
	S2	100%	100%	56%	0%	100%	67%	44%
	S3	100%	100%	89%	0%	100%	100%	56%
	Raw accuracy	100%	100%	67%	0%	100%	89%	37%
	Adjusted acc.	**100%**	**100%**	**75%**	**0%**	**100%**	**100%**	--
A-Adv-F2	S1	78%	67%	56%	0%	100%	89%	89%
	S2	78%	100%	100%	33%	100%	100%	89%
	S3	67%	100%	100%	22%	100%	89%	89%
	Raw accuracy	74%	89%	85%	19%	100%	93%	89%
	Adjusted acc.	**75%**	**91%**	**96%**	**20%**	**100%**	**100%**	--
B-Beg-F1	T1	11%	100%	0%	0%	100%	0%	0%
	T2	67%	100%	33%	11%	56%	0%	0%
	Raw accuracy	39%	100%	17%	6%	78%	0%	0%
	Adjusted acc.	**39%**	**100%**	**19%**	**7%**	**84%**	**0%**	--
B-Beg-F2	T1	0%	67%	0%	0%	56%	22%	0%
	T2	22%	89%	0%	0%	78%	22%	0%
	Raw accuracy	11%	78%	0%	0%	67%	22%	0%
	Adjusted acc.	**11%**	**79%**	**0%**	**0%**	**72%**	**27%**	--

Appendix E: (continued)

Participant	Session	medial /d/	medial /t/	medial /r/	initial /r/	medial /ɾ/	onset cl. /ɾ/	SVE [ᵊɾ]
B-Beg-F3	T1	0%	100%	0%	0%	0%	0%	0%
	T2	0%	89%	0%	0%	11%	0%	0%
	Raw accuracy	0%	94%	0%	0%	6%	0%	0%
	Adjusted acc.	**0%**	**95%**	**0%**	**0%**	**6%**	**0%**	--
B-Beg-F4	T1	11%	100%	0%	11%	100%	11%	0%
	T2	11%	89%	33%	11%	100%	78%	0%
	Raw accuracy	11%	94%	17%	11%	100%	44%	0%
	Adjusted acc.	**11%**	**96%**	**19%**	**13%**	**100%**	**54%**	--
B-Adv-F1	T1	44%	100%	0%	0%	44%	22%	0%
	T2	89%	100%	0%	0%	78%	44%	0%
	Raw accuracy	67%	100%	0%	0%	61%	33%	0%
	Adjusted acc.	**67%**	**100%**	**0%**	**0%**	**66%**	**41%**	--
B-Adv-F2	T1	100%	100%	100%	44%	67%	44%	44%
	T2	100%	100%	67%	22%	78%	22%	11%
	Raw accuracy	100%	100%	83%	33%	72%	33%	28%
	Adjusted acc.	**100%**	**100%**	**94%**	**39%**	**78%**	**41%**	--
B-Adv-F3	T1	89%	100%	67%	33%	56%	56%	56%
	T2	44%	89%	89%	89%	22%	67%	44%
	Raw accuracy	66%	94%	78%	61%	39%	61%	50%
	Adjusted acc.	**67%**	**96%**	**87%**	**71%**	**42%**	**75%**	--
B-Adv-F4	T1	100%	100%	100%	89%	89%	44%	0%
	T2	100%	100%	100%	78%	78%	11%	0%
	Raw accuracy	100%	100%	100%	83%	83%	28%	0%
	Adjusted acc.	**100%**	**100%**	**100%**	**97%**	**90%**	**34%**	--
B-Adv-F5	T1	100%	100%	11%	0%	78%	33%	0%
	T2	100%	100%	33%	0%	100%	22%	11%
	Raw accuracy	100%	100%	22%	0%	89%	28%	6%
	Adjusted acc.	**100%**	**100%**	**25%**	**0%**	**96%**	**34%**	--
B-Adv-F6	T1	0%	100%	78%	0%	89%	56%	0%
	T2	0%	100%	78%	11%	44%	56%	0%
	Raw accuracy	0%	100%	78%	6%	67%	56%	0%
	Adjusted acc.	**0%**	**100%**	**88%**	**6%**	**72%**	**68%**	--

Appendix E: (continued)

Participant	Session	medial /d/	medial /t/	medial /r/	initial /r/	medial /ɾ/	onset cl. /ɾ/	SVE [ᵊɾ]
B-ExpL-M1	Raw accuracy Adjusted acc.	100% **100%**	100% **100%**	89% **100%**	22% **26%**	78% **84%**	11% **14%**	11% --
B-ExpL-M2	Raw accuracy Adjusted acc.	100% **100%**	100% **100%**	100% **100%**	78% **90%**	100% **100%**	100% **100%**	100% --
B-ExpL-M3	Raw accuracy Adjusted acc.	0% **0%**	100% **100%**	0% **0%**	0% **0%**	78% **84%**	89% **100%**	89% --
B-ExpL-M4	Raw accuracy Adjusted acc.	100% **100%**	100% **100%**	100% **100%**	67% **78%**	89% **96%**	89% **100%**	78% --
B-ExpL-F1	Raw accuracy Adjusted acc.	89% **90%**	100% **100%**	78% **87%**	67% **78%**	89% **96%**	67% **81%**	67% --
B-ExpL-F2	Raw accuracy Adjusted acc.	67% **67%**	100% **100%**	89% **100%**	78% **90%**	89% **96%**	78% **95%**	78% --
B-ExpL-F3	Raw accuracy Adjusted acc.	100% **100%**	100% **100%**	100% **100%**	100% **100%**	78% **84%**	56% **68%**	56% --
B-ExpL-F4	Raw accuracy Adjusted acc.	89% **90%**	100% **100%**	100% **100%**	89% **100%**	56% **60%**	0% **0%**	0% --
B-ExpL-F5	Raw accuracy Adjusted acc.	89% **90%**	100% **100%**	100% **100%**	100% **100%**	100% **100%**	89% **100%**	89% --
B-ExpL-F6	Raw accuracy Adjusted acc.	89% **90%**	100% **100%**	78% **87%**	22% **26%**	100% **100%**	78% **95%**	22% --
B-ExpL-F7	Raw accuracy Adjusted acc.	100% **100%**	100% **100%**	66% **74%**	22% **26%**	89% **96%**	100% **100%**	56% --

Appendix K: Frequency of word-medial /d/ variants for each L2 participant

	Participant	/d/ variant type							
		1 (≈[ð])	2 (≈[ð̞])	3 (≈[d])	4 (≈[ɾ])	5 (≈[t])	6 (≈[θ])	7 (≈[ɾ]; ≈[ð̞])	8 (≈[ð]; ≈[d])
Experiment A	A-Beg-M1	22%	0%	44%	26%	0%	0%	7%	0%
	A-Beg-M2	4%	0%	26%	67%	4%	0%	0%	0%
	A-Beg-F1	26%	30%	4%	0%	33%	0%	7%	0%
	A-Beg-F2	63%	4%	19%	0%	7%	0%	4%	4%
	A-Int-M1	0%	0%	100%	0%	0%	0%	0%	0%
	A-Int-M2	11%	0%	22%	63%	0%	0%	4%	0%
	A-Int-F1	81%	19%	0%	0%	0%	0%	0%	0%
	A-Int-F2	0%	0%	59%	41%	0%	0%	0%	0%
	A-Int-F3	15%	0%	74%	4%	4%	0%	4%	0%
	A-Adv-M1	0%	33%	4%	44%	0%	0%	19%	0%
	A-Adv-M2	0%	0%	59%	30%	0%	0%	11%	0%
	A-Adv-F1	96%	4%	0%	0%	0%	0%	0%	0%
	A-Adv-F2	37%	37%	0%	19%	0%	0%	7%	0%
Experiment B	B-Beg-F1	39%	6%	22%	6%	11%	6%	6%	6%
	B-Beg-F2	11%	0%	44%	33%	0%	0%	11%	0%
	B-Beg-F3	0%	0%	78%	17%	0%	0%	6%	0%
	B-Beg-F4	17%	0%	72%	0%	0%	0%	0%	11%
	B-Adv-F1	33%	33%	0%	28%	0%	0%	6%	0%
	B-Adv-F2	17%	83%	0%	0%	0%	0%	0%	0%
	B-Adv-F3	78%	0%	6%	0%	0%	11%	6%	0%
	B-Adv-F4	67%	33%	0%	0%	0%	0%	0%	0%
	B-Adv-F5	78%	22%	0%	0%	0%	0%	0%	0%
	B-Adv-F6	0%	0%	6%	67%	0%	0%	22%	6%
	B-ExpL-M1	44%	56%	0%	0%	0%	0%	0%	0%
	B-ExpL-M2	44%	56%	0%	0%	0%	0%	0%	0%
	B-ExpL-M3	0%	0%	0%	44%	0%	0%	56%	0%
	B-ExpL-M4	33%	67%	0%	0%	0%	0%	0%	0%
	B-ExpL-F1	56%	33%	0%	0%	0%	11%	0%	0%
	B-ExpL-F2	56%	11%	0%	0%	0%	22%	0%	11%
	B-ExpL-F3	56%	44%	0%	0%	0%	0%	0%	0%
	B-ExpL-F4	100%	0%	0%	0%	0%	0%	0%	0%
	B-ExpL-F5	56%	44%	0%	0%	0%	0%	0%	0%
	B-ExpL-F6	56%	33%	0%	0%	0%	11%	0%	0%
	B-ExpL-F7	22%	78%	0%	0%	0%	0%	0%	0%

Appendix L: Frequency of "spirantized" vs. "approximated" instantiations for word-medial /d/ productions scored "1" (accurate) by participant.

	Subjects with any /d/ productions scored "1" (accurate)	n	Number of tokens scored "1" accurate	Number of tokens scored "1" (accurate) with measurable duration	Frequency of Type 1 (≈[ð]) "spirantized" productions	Frequency of Type 2 (≈[ð̞]) "approximated" productions
Experiment A	A-Beg-M1	27	7	7	100%	0%
	A-Beg-M2	27	2	2	100%	0%
	A-Beg-F1	27	15	7	47%	53%
	A-Beg-F2	27	18	17	94%	6%
	A-Int-M2	27	3	3	100%	0%
	A-Int-F1	27	27	22	81%	19%
	A-Int-F3	27	4	4	100%	0%
	A-Adv-F1	27	27	26	96%	4%
	A-Adv-F2	27	20	10	50%	50%
Experiment B	B-Beg-F1	18	8	7	88%	13%
	B-Beg-F2	18	2	2	100%	0%
	B-Beg-F4	18	3	3	100%	0%
	B-Adv-F1	18	12	7	58%	42%
	B-Adv-F2	18	18	3	17%	83%
	B-Adv-F3	18	12	12	100%	0%
	B-Adv-F4	18	18	12	67%	33%
	B-Adv-F5	18	18	14	78%	22%
	B-ExpL-M1	9	9	4	44%	56%
	B-ExpL-M2	9	9	4	44%	56%
	B-ExpL-M4	9	9	3	33%	67%
	B-ExpL-F1	9	8	5	63%	38%
	B-ExpL-F2	9	6	5	83%	17%
	B-ExpL-F3	9	9	5	56%	44%
	B-ExpL-F4	9	9	9	100%	0%
	B-ExpL-F5	9	9	5	56%	44%
	B-ExpL-F6	9	8	5	63%	38%
	B-ExpL-F7	9	9	2	22%	78%

Appendix M: Mean overall duration for spirantized productions by participant.

Subjects with any /d/ productions scored as "1" (accurate)	Number of tokens scored "1" (accurate) with measurable duration	Percent accurate productions realized as a spirant	Mean overall spirant duration (ms)
A-Beg-M1	7	100%	46
A-Beg-M2	1	100%	40
A-Beg-F1	7	47%	54
A-Beg-F2	17	94%	54
A-Int-M2	3	100%	57
A-Int-F1	22	81%	68
A-Int-F3	4	100%	45
A-Adv-F1	26	96%	59
A-Adv-F2	10	50%	50
B-Beg-F1	7	88%	44
B-Beg-F2	2	100%	60
B-Beg-F4	3	100%	43
B-Adv-F1	7	58%	43
B-Adv-F2	3	17%	52
B-Adv-F3	12	100%	70
B-Adv-F4	12	67%	56
B-Adv-F5	14	78%	50
B-ExpL-F1	5	63%	45
B-ExpL-F2	5	83%	49
B-ExpL-F3	5	56%	65
B-ExpL-F4	9	100%	59
B-ExpL-F5	5	56%	46
B-ExpL-F6	5	63%	47
B-ExpL-F7	2	22%	40
B-ExpL-M1	4	44%	50
B-ExpL-M2	4	44%	52
B-ExpL-M4	3	33%	36

Appendix N: Frequency of word-medial /t/ variants by participant.

	Participant	/t/ variant type			
		1 (≈[t])	2 (≈[ɾ])	3 (≈[d])	4 (≈[θ])
Experiment A	A-Beg-M1	96%	0%	0%	4%
	A-Beg-M2	81%	15%	0%	4%
	A-Beg-F1	93%	0%	4%	4%
	A-Beg-F2	100%	0%	0%	0%
	A-Int-M1	100%	0%	0%	0%
	A-Int-M2	100%	0%	0%	0%
	A-Int-F1	100%	0%	0%	0%
	A-Int-F2	85%	4%	7%	4%
	A-Int-F3	100%	0%	0%	0%
	A-Adv-M1	93%	0%	0%	7%
	A-Adv-M2	100%	0%	0%	0%
	A-Adv-F1	100%	0%	0%	0%
	A-Adv-F2	89%	0%	0%	11%
Experiment B	B-Beg-F1	100%	0%	0%	0%
	B-Beg-F2	78%	22%	0%	0%
	B-Beg-F3	94%	0%	6%	0%
	B-Beg-F4	94%	0%	6%	0%
	B-Adv-F1	100%	0%	0%	0%
	B-Adv-F2	100%	0%	0%	0%
	B-Adv-F3	94%	0%	0%	6%
	B-Adv-F4	100%	0%	0%	0%
	B-Adv-F5	100%	0%	0%	0%
	B-Adv-F6	100%	0%	0%	0%
	B-ExpL-M1	100%	0%	0%	0%
	B-ExpL-M2	100%	0%	0%	0%
	B-ExpL-M3	100%	0%	0%	0%
	B-ExpL-M4	100%	0%	0%	0%
	B-ExpL-F1	100%	0%	0%	0%
	B-ExpL-F2	100%	0%	0%	0%
	B-ExpL-F3	100%	0%	0%	0%
	B-ExpL-F4	100%	0%	0%	0%
	B-ExpL-F5	100%	0%	0%	0%
	B-ExpL-F6	100%	0%	0%	0%
	B-ExpL-F7	100%	0%	0%	0%

Appendix O: Mean stop closure duration (SCD) and voice onset time (VOT) for Type 1 (≈[t]) L2 word-medial productions scored "1" (accurate) by participant.

	Participant	n	SCD (ms)		VOT (ms)		SCD:VOT (VOT = 1)	
			mean	stdev	mean	stdev	mean	stdev
Experiment A	A-Beg-M1	26	77	13	20	4	3.9	1.1
	A-Beg-M2	22	62	14	38	9	1.7	0.5
	A-Beg-F1	25	100	27	31	12	3.6	1.7
	A-Beg-F2	27	66	14	53	16	1.4	0.6
	A-Int-M1	27	75	10	43	17	2.2	1.2
	A-Int-M2	27	63	14	26	7	2.7	1.4
	A-Int-F1	27	110	16	22	7	5.4	1.9
	A-Int-F2	23	57	14	45	9	1.3	0.4
	A-Int-F3	27	101	16	29	12	4.1	1.8
	A-Adv-M1	25	59	10	29	7	2.2	0.8
	A-Adv-M2	27	56	10	43	10	1.4	0.4
	A-Adv-F1	27	87	13	22	5	4.2	1.3
	A-Adv-F2	24	84	17	26	7	3.4	0.9
Experiment B	B-Beg-F1	18	66	16	46	17	1.7	1
	B-Beg-F2	14	72	11	39	14	2.1	1
	B-Beg-F3	17	74	11	30	9	2.7	0.9
	B-Beg-F4	17	67	11	25	4	2.8	0.6
	B-Adv-F1	18	62	14	20	6	3.5	1.8
	B-Adv-F2	18	86	18	26	30	4.7	1.8
	B-Adv-F3	17	92	12	19	5	5.2	1.4
	B-Adv-F4	18	97	10	22	5	4.6	1
	B-Adv-F5	18	100	12	17	5	6.4	2.1
	B-Adv-F6	18	76	16	32	8	2.6	1.2
	B-ExpL-M1	9	84	10	28	3	3.1	0.5
	B-ExpL-M2	9	92	10	26	5	3.7	0.7
	B-ExpL-M3	9	68	14	22	7	3.4	1.3
	B-ExpL-M4	9	64	7	17	4	4.1	1
	B-ExpL-F1	9	95	10	14	3	6.9	1.5
	B-ExpL-F2	9	85	10	16	3	5.4	1.3
	B-ExpL-F3	9	104	10	27	5	3.9	0.9
	B-ExpL-F4	9	90	11	19	8	5.3	2.1
	B-ExpL-F5	9	69	10	22	8	3.6	1.4
	B-ExpL-F6	9	102	8	18	5	5.8	1.4
	B-ExpL-F7	9	69	10	24	3	2.9	0.6

Appendix P: Adjusted overall word-medial target accuracy for experienced level participants.

	Participant	Adjusted overall target accuracy			
		/d/	/t/	/r/	/ɾ/
Experiment B	B-ExpL-M1	100%	100%	100%	84%
	B-ExpL-M2	100%	100%	100%	100%
	B-ExpL-M3	0%	100%	0%	84%
	B-ExpL-M4	100%	100%	100%	96%
	B-ExpL-F1	90%	100%	87%	96%
	B-ExpL-F2	67%	100%	100%	96%
	B-ExpL-F3	100%	100%	100%	84%
	B-ExpL-F4	90%	100%	100%	60%
	B-ExpL-F5	90%	100%	100%	100%
	B-ExpL-F6	90%	100%	88%	100%
	B-ExpL-F7	100%	100%	74%	96%

Appendix Q: Frequency of word-initial /r/ variant types for each participant.

	Participant	word-initial /r/ variant type						
		1 (≈[r])	2 (≈[ɾ])	3 (≈[ɹ])	4 (≈[ɾ]; ≈[ɹ])	5 (≈[ɾ]; ≈[r])	6 (≈[ɹ]; ≈[ʐ])	7 (≈[r]; ≈[ɹ, ɾ])
Experiment A	A-Beg-M1	0%	41%	41%	19%	0%	0%	0%
	A-Beg-M2	0%	93%	7%	0%	0%	0%	0%
	A-Beg-F1	67%	19%	4%	4%	0%	4%	4%
	A-Beg-F2	0%	59%	33%	7%	0%	0%	0%
	A-Int-M1	0%	74%	11%	0%	15%	0%	0%
	A-Int-M2	89%	7%	0%	4%	0%	0%	0%
	A-Int-F1	0%	67%	22%	7%	0%	4%	0%
	A-Int-F2	0%	4%	93%	4%	0%	0%	0%
	A-Int-F3	0%	0%	100%	0%	0%	0%	0%
	A-Adv-M1	0%	67%	4%	30%	0%	0%	0%
	A-Adv-M2	0%	0%	100%	0%	0%	0%	0%
	A-Adv-F1	67%	33%	0%	0%	0%	0%	0%
	A-Adv-F2	85%	7%	0%	0%	0%	0%	7%
Experiment B	B-Beg-F1	17%	44%	0%	11%	6%	17%	6%
	B-Beg-F2	0%	72%	0%	11%	17%	0%	0%
	B-Beg-F3	0%	72%	17%	11%	0%	0%	0%
	B-Beg-F4	17%	78%	0%	0%	6%	0%	0%
	B-Adv-F1	0%	94%	0%	0%	6%	0%	0%
	B-Adv-F2	83%	11%	0%	0%	6%	0%	0%
	B-Adv-F3	78%	11%	0%	0%	0%	0%	11%
	B-Adv-F4	100%	0%	0%	0%	0%	0%	0%
	B-Adv-F5	22%	61%	0%	11%	0%	0%	6%
	B-Adv-F6	78%	17%	0%	0%	6%	0%	0%
	B-ExpL-M1	89%	0%	0%	0%	0%	0%	11%
	B-ExpL-M2	100%	0%	0%	0%	0%	0%	0%
	B-ExpL-M3	0%	100%	0%	0%	0%	0%	0%
	B-ExpL-M4	100%	0%	0%	0%	0%	0%	0%
	B-ExpL-F1	78%	11%	0%	0%	0%	0%	11%
	B-ExpL-F2	89%	0%	0%	0%	0%	0%	11%
	B-ExpL-F3	100%	0%	0%	0%	0%	0%	0%
	B-ExpL-F4	100%	0%	0%	0%	0%	0%	0%
	B-ExpL-F5	100%	0%	0%	0%	0%	0%	0%
	B-ExpL-F6	78%	22%	0%	0%	0%	0%	0%
	B-ExpL-F7	67%	33%	0%	0%	0%	0%	0%

Appendix R: Frequency of word-medial /r/ variant types for each participant.

	Participant	word-medial /r/ variant type						
		1 (≈[r])	2 (≈[ɾ])	3 (≈[ɹ])	4 (≈[ɾ]; ≈[ɹ])	5 (≈[ɾ]; ≈[r])	6 (≈[ɹ]; ≈[z̞])	7 (≈[r]; ≈[ɹ, ɾ])
Experiment A	A-Beg-M1	0%	15%	74%	4%	0%	7%	0%
	A-Beg-M2	0%	19%	67%	15%	0%	0%	0%
	A-Beg-F1	48%	4%	33%	0%	0%	7%	7%
	A-Beg-F2	0%	15%	70%	7%	0%	7%	0%
	A-Int-M1	0%	0%	100%	0%	0%	0%	0%
	A-Int-M2	0%	78%	15%	7%	0%	0%	0%
	A-Int-F1	0%	19%	74%	4%	0%	4%	0%
	A-Int-F2	0%	0%	93%	0%	0%	7%	0%
	A-Int-F3	0%	0%	100%	0%	0%	0%	0%
	A-Adv-M1	0%	33%	67%	0%	0%	0%	0%
	A-Adv-M2	0%	0%	100%	0%	0%	0%	0%
	A-Adv-F1	0%	100%	0%	0%	0%	0%	0%
	A-Adv-F2	19%	81%	0%	0%	0%	0%	0%
Experiment B	B-Beg-F1	6%	17%	11%	22%	0%	44%	0%
	B-Beg-F2	0%	50%	11%	22%	17%	0%	0%
	B-Beg-F3	0%	6%	94%	0%	0%	0%	0%
	B-Beg-F4	11%	28%	0%	0%	39%	0%	22%
	B-Adv-F1	0%	72%	11%	17%	0%	0%	0%
	B-Adv-F2	33%	22%	0%	0%	28%	11%	6%
	B-Adv-F3	61%	28%	0%	0%	0%	0%	11%
	B-Adv-F4	83%	6%	0%	0%	11%	0%	0%
	B-Adv-F5	0%	0%	33%	22%	0%	44%	0%
	B-Adv-F6	6%	67%	6%	6%	11%	0%	6%
	B-ExpL-M1	22%	22%	22%	0%	0%	22%	11%
	B-ExpL-M2	78%	11%	0%	0%	0%	0%	11%
	B-ExpL-M3	0%	89%	0%	11%	0%	0%	0%
	B-ExpL-M4	67%	22%	0%	0%	11%	0%	0%
	B-ExpL-F1	67%	11%	0%	0%	11%	0%	11%
	B-ExpL-F2	78%	11%	0%	0%	0%	0%	11%
	B-ExpL-F3	100%	0%	0%	0%	0%	0%	0%
	B-ExpL-F4	89%	0%	0%	0%	0%	11%	0%
	B-ExpL-F5	100%	0%	0%	0%	0%	0%	0%
	B-ExpL-F6	22%	78%	0%	0%	0%	0%	0%
	B-ExpL-F7	22%	56%	0%	0%	11%	0%	11%

Appendix S: Measurements of word-medial /r/ realizations scored “1” (accurate) by participant. Degree of voicing values were noted as “0”=voiceless, “1”=partially voiced, or “2”=completely voiced.

	Participant	n	Mean measurements of Type 1 (≈[r]) word-medial /r/		
			Overall duration in ms	Degree of voicing	Number of stripes
Exp. A	A-Beg-F1	18	72	2	2.78
	A-Int-M2	24	75	2	2.08
	A-Adv-F1	18	77	1.94	2.28
	A-Adv-F2	23	66	2	2.13
Experiment B	B-Adv-F2	15	63	1.87	2.33
	B-Adv-F3	14	58	2	2.14
	B-Adv-F4	18	86	2	2.61
	B-Adv-F5	4	60	2	2
	B-Adv-F6	14	101	2	2.71
	B-Beg-F1	3	57	1.67	2
	B-Beg-F4	3	57	1	2
	B-ExpL-F1	7	72	2	2.29
	B-ExpL-F2	8	67	1.5	2
	B-ExpL-F3	9	82	2.11	2.44
	B-ExpL-F4	9	99	1.67	3.22
	B-ExpL-F5	9	65	2	2.22
	B-ExpL-F6	7	57	2	2
	B-ExpL-F7	6	62	2	2
	B-ExpL-M1	8	74	2	2.25
	B-ExpL-M2	9	83	1.89	2.56
	B-ExpL-M4	9	74	2	2

Appendix T: Measurements of word-initial /r/ realizations scored "1" (accurate) by participant. Degree of voicing values reflect "0"=voiceless, "1"=partially voiced, or "2"=completely voiced.

	Participant	n	Mean measurements of Type 1 (≈[r]) word-initial /r/		
			Overall duration in ms	Degree of voicing	Number of stripes
A	A-Beg-F1	13	65	1.85	2.38
A	A-Adv-F2	5	56	2	2
Experiment B	B-Adv-F2	6	83	2	2.33
Experiment B	B-Adv-F3	11	61	2	2.18
Experiment B	B-Adv-F4	15	89	1.73	2.87
Experiment B	B-Adv-F6	1	64	2	2
Experiment B	B-Beg-F1	1	60	1	2
Experiment B	B-Beg-F4	2	68	1.5	2.5
Experiment B	B-ExpL-F1	6	66	2	2
Experiment B	B-ExpL-F2	7	71	1.14	2
Experiment B	B-ExpL-F3	9	86	1.67	2.78
Experiment B	B-ExpL-F4	8	91	1.75	3
Experiment B	B-ExpL-F5	9	76	2	2.67
Experiment B	B-ExpL-F6	2	51	2	2
Experiment B	B-ExpL-F7	2	68	2	2
Experiment B	B-ExpL-M1	2	77	1.5	2
Experiment B	B-ExpL-M2	7	76	2	2.43
Experiment B	B-ExpL-M4	6	74	2	2

Appendix U: Frequency of variant types for word-medial /ɾ/ by participant.

	Participant	word-medial /ɾ/ variant type				
		1 (≈[ɾ])	2 (≈[ɹ])	3 (≈[ɾ]; ≈[ɹ])	4 (≈[r]; ≈[ʐ,ɾ,r])	5 (≈[ɾ]; ≈[d,ð])
Experiment A	A-Beg-M1	81%	7%	11%	0%	0%
	A-Beg-M2	74%	7%	15%	0%	4%
	A-Beg-F1	7%	19%	7%	63%	4%
	A-Beg-F2	30%	56%	15%	0%	0%
	A-Int-M1	0%	100%	0%	0%	0%
	A-Int-M2	74%	22%	4%	0%	0%
	A-Int-F1	44%	48%	7%	0%	0%
	A-Int-F2	7%	89%	0%	0%	4%
	A-Int-F3	0%	100%	0%	0%	0%
	A-Adv-M1	22%	41%	30%	0%	7%
	A-Adv-M2	0%	100%	0%	0%	0%
	A-Adv-F1	100%	0%	0%	0%	0%
	A-Adv-F2	100%	0%	0%	0%	0%
Experiment B	B-Beg-F1	78%	6%	6%	6%	6%
	B-Beg-F2	67%	17%	6%	0%	11%
	B-Beg-F3	6%	56%	33%	0%	6%
	B-Beg-F4	100%	0%	0%	0%	0%
	B-Adv-F1	61%	17%	17%	0%	6%
	B-Adv-F2	72%	0%	6%	0%	22%
	B-Adv-F3	39%	6%	0%	56%	0%
	B-Adv-F4	83%	6%	0%	0%	11%
	B-Adv-F5	89%	0%	11%	0%	0%
	B-Adv-F6	67%	6%	6%	0%	22%
	B-ExpL-M1	78%	0%	0%	11%	11%
	B-ExpL-M2	100%	0%	0%	0%	0%
	B-ExpL-M3	78%	0%	0%	0%	22%
	B-ExpL-M4	89%	0%	0%	0%	11%
	B-ExpL-F1	89%	0%	0%	11%	0%
	B-ExpL-F2	89%	0%	11%	0%	0%
	B-ExpL-F3	78%	0%	11%	0%	11%
	B-ExpL-F4	56%	33%	11%	0%	0%
	B-ExpL-F5	100%	0%	0%	0%	0%
	B-ExpL-F6	100%	0%	0%	0%	0%
	B-ExpL-F7	89%	0%	0%	11%	0%

Appendix V: Frequency of variant types for onset cluster /ɾ/ by participant.

	Participant	onset cluster /ɾ/ variant type 1 (≈[ɾ])	2 (≈[ɹ])	3 (≈[ɾ]; ≈[ɹ])	4 (≈[r]; ≈[ʐ,ɾ,ɹ])	5 (≈[ɾ]; ≈[d,ð])
Experiment A	A-Beg-M1	4%	74%	22%	0%	0%
	A-Beg-M2	0%	96%	4%	0%	0%
	A-Beg-F1	7%	37%	4%	52%	0%
	A-Beg-F2	0%	100%	0%	0%	0%
	A-Int-M1	0%	100%	0%	0%	0%
	A-Int-M2	19%	52%	30%	0%	0%
	A-Int-F1	0%	89%	11%	0%	0%
	A-Int-F2	0%	100%	0%	0%	0%
	A-Int-F3	0%	100%	0%	0%	0%
	A-Adv-M1	26%	44%	22%	0%	7%
	A-Adv-M2	0%	100%	0%	0%	0%
	A-Adv-F1	89%	0%	11%	0%	0%
	A-Adv-F2	93%	0%	7%	0%	0%
Experiment B	B-Beg-F1	0%	89%	11%	0%	0%
	B-Beg-F2	22%	39%	33%	6%	0%
	B-Beg-F3	0%	94%	6%	0%	0%
	B-Beg-F4	44%	17%	6%	28%	6%
	B-Adv-F1	33%	33%	28%	6%	0%
	B-Adv-F2	33%	0%	0%	0%	67%
	B-Adv-F3	61%	0%	11%	22%	6%
	B-Adv-F4	28%	39%	33%	0%	0%
	B-Adv-F5	28%	44%	28%	0%	0%
	B-Adv-F6	56%	0%	28%	11%	6%
	B-ExpL-M1	11%	67%	22%	0%	0%
	B-ExpL-M2	100%	0%	0%	0%	0%
	B-ExpL-M3	89%	0%	11%	0%	0%
	B-ExpL-M4	89%	0%	11%	0%	0%
	B-ExpL-F1	67%	22%	11%	0%	0%
	B-ExpL-F2	78%	0%	22%	0%	0%
	B-ExpL-F3	56%	0%	11%	33%	0%
	B-ExpL-F4	0%	33%	56%	0%	11%
	B-ExpL-F5	89%	0%	11%	0%	0%
	B-ExpL-F6	78%	0%	11%	0%	11%
	B-ExpL-F7	100%	0%	0%	0%	0%

Appendix W: Measurements of accurately scored word-initial /r/ realizations by participant.

	Participant	n	Mean measurement of Type 1 (≈[ɾ]) word-medial /ɾ/	
			Overall duration (ms)	stdev
Experiment A	A-Beg-M1	22	24	3.2
	A-Beg-M2	20	25	3.2
	A-Beg-F1	2	23	4.2
	A-Beg-F2	8	22	3.8
	A-Int-M2	20	25	3.0
	A-Int-F1	12	24	3.2
	A-Int-F2	2	28	1.4
	A-Adv-M1	6	23	3.6
	A-Adv-F1	27	24	3.7
	A-Adv-F2	27	21	3.7
Experiment B	B-Beg-F1	14	22	3.3
	B-Beg-F2	12	24	3.7
	B-Beg-F3	1	24	---
	B-Beg-F4	18	24	4.1
	B-Adv-F1	11	18	4.0
	B-Adv-F2	13	24	3.6
	B-Adv-F3	7	21	2.5
	B-Adv-F4	15	22	3.3
	B-Adv-F5	16	20	4.6
	B-Adv-F6	12	23	3.5
	B-ExpL-M1	7	23	1.6
	B-ExpL-M2	9	22	2.8
	B-ExpL-M3	7	25	2.9
	B-ExpL-M4	8	23	4.5
	B-ExpL-F1	8	27	3.1
	B-ExpL-F2	8	23	3.7
	B-ExpL-F3	7	24	2.8
	B-ExpL-F4	5	24	2.3
	B-ExpL-F5	9	22	3.2
	B-ExpL-F6	9	22	2.5
	B-ExpL-F7	8	24	4.0

Appendix X: Measurements of onset cluster /ɾ/ realizations scored "accurate" for each participant. The first number in the parentheses corresponds to the number of tokens calculated in the mean; the second number the standard deviation: *mean (n; stdev)*.

	Participant	Mean measurements for Type 1 (≈[ɾ]) onset cluster /ɾ/		
		onset duration in ms	svarabhakti vowel duration in ms	stripe to vowel duration in ms
Experiment A	A-Beg-M1	72 (1; na)	13 (1; na)	21 (1; na)
	A-Beg-F1	77 (2; 7.1)	19 (1; na)	14 (2; 4.2)
	A-Int-M2	74 (5; 8.6)	6 (2; 8.5)	19 (5; 9.7)
	A-Adv-M1	109 (6; 12.8)	0 (na;na)	24 (4; 5.3)
	A-Adv-F1	72 (24; 10.1)	15 (15; 6.9)	14 (19; 3.1)
	A-Adv-F2	50 (25; 12.2)	18 (25; 5.2)	15 (24; 3.3)
Experiment A	B-Beg-F2	86 (4; 14.6)	4 (4; 8)	33 (3; 14.6)
	B-Beg-F4	77 (7; 13.1)	0 (8; 0)	36 (8; 12.2)
	B-Adv-F1	57 (6; 9.7)	2 (6; 5.3)	17 (5; 7.8)
	B-Adv-F2	66 (6; 9.7)	24 (6; 15.4)	30 (6; 4.1)
	B-Adv-F3	64 (11; 14.9)	14 (11; 8.2)	24 (11; 11.1)
	B-Adv-F4	99 (5; 14.3)	0 (5; 0)	26 (5; 24.3)
	B-Adv-F5	46 (5; 4.1)	2 (5; 5.4)	17 (5; 3.8)
	B-Adv-F6	74 (10; 11.9)	2 (7; 5.3)	32 (8; 16.4)
	B-ExpL-M1	63 (1; na)	25 (1; na)	15 (1; na)
	B-ExpL-M2	61 (9; 7.9)	19 (9; 5.2)	24 (9; 5.7)
	B-ExpL-M3	63 (8; 14.3)	23 (8; 10.4)	22 (8; 9.8)
	B-ExpL-M4	59 (8; 8.6)	17 (8; 4)	28 (7; 7.6)
	B-ExpL-F1	62 (6; 10.5)	15 (6; 4.1)	30 (6; 4.9)
	B-ExpL-F2	64 (7; 16.2)	18 (7; 2.4)	26 (7; 7.7)
	B-ExpL-F3	78 (5; 14.8)	27 (5; 7.3)	34 (5; 16.7)
	B-ExpL-F5	58 (8; 6.6)	26 (8; 5.7)	16 (8; 5.3)
	B-ExpL-F6	59 (6; 14)	7 (6; 7.8)	23 (5; 3.6)
	B-ExpL-F7	55 (9; 9.7)	7 (9; 7.1)	24 (9; 4.6)

Appendix Y: Svarabhakti vowel emergence (SVE) scores by participant

	Participant	SVE score
Experiment A	A-Beg-M1	4%
	A-Beg-M2	0%
	A-Beg-F1	4%
	A-Beg-F2	0%
	A-Int-M1	0%
	A-Int-M2	4%
	A-Int-F1	0%
	A-Int-F2	0%
	A-Int-F3	0%
	A-Adv-M1	0%
	A-Adv-M2	0%
	A-Adv-F1	37%
	A-Adv-F2	89%
Experiment B	B-Beg-F1	0%
	B-Beg-F2	0%
	B-Beg-F3	0%
	B-Beg-F4	0%
	B-Adv-F1	0%
	B-Adv-F2	28%
	B-Adv-F3	50%
	B-Adv-F4	0%
	B-Adv-F5	6%
	B-Adv-F6	0%
	B-ExpL-M1	11%
	B-ExpL-M2	100%
	B-ExpL-M3	89%
	B-ExpL-M4	78%
	B-ExpL-F1	67%
	B-ExpL-F2	78%
	B-ExpL-F3	56%
	B-ExpL-F4	0%
	B-ExpL-F5	89%
	B-ExpL-F6	22%
	B-ExpL-F7	56%

VITA

EDUCATION

2005 **University of Washington**, Department of Linguistics, Seattle, WA. PhD, March 2005, relating to acoustic phonetics, second language acquisition, and Spanish phonology. Dissertation title: "The relative degree of difficulty of L2 Spanish /d, t/, trill and tap by L1 English speakers: Auditory and acoustic methods of defining pronunciation accuracy."

2001 **University of Washington**, Department of Linguistics, Seattle, WA. MA Romance Linguistics, December 2001. Thesis title: "Formal pronunciation instruction for the beginning adult learner of Spanish: Acoustic measures and criteria of accurate L2 phonology."

1996 **US Peace Corps Training Center**, Cochabamba, Bolivia. Three-month instruction on culture, language, and small business development as applied specifically to Bolivia.

1991 - 1995 **Denison University**, Granville, OH. BA, *cum laude*. Double major in Spanish and Economics.

1993 - 1994 **Universidad de Salamanca**, Salamanca, Spain. Full-year study abroad. All lectures and course-work in Spanish.

POSITIONS HELD

2002 - 2004 **Staff Associate**, Seattle, WA. Language Learning Center, UW. Assisted language instructors with language technology needs. Projects included audio digitization, multimedia reformatting, digital video editing, video-conferencing, multi-language vocabulary software programming, website design, issues with Unicode, and online courseware development.

2003 - 2004 **Software designer/programmer**, Seattle, WA. Vowel Overlap Indication Software in 3D, Dr. Alicia Beckford-Wassink. 3D and 2D modeling of spectral and temporal features of vowels. Implementation of normalization algorithms to allow cross-speaker comparisons. Created GUI and data structures from the ground up. Addressed cross-platform issues (Mac vs. PC).

2004, 2005	**Perl instructor**, Seattle, WA. Department of Linguistics, UW. "Beginning Perl programming for Linguists," Winter quarter, LING 480. Designed and developed course curriculum with a focus on fundamental programming and text-processing. Re-taught course in Winter quarter 2005 under LING 270.
2003	**Data analyst**, Seattle, WA. Grant funded by the Center for Mind, Body, and Learning (now the Institute for Learning and Brain Sciences), working with Dr. Julia Herschensohn. Transcribed the speech of child second language learners of Spanish. Created descriptive reports and performed inferential tests using statistical software.
2002	**C++ programmer**, Seattle, WA. Grant funded by the Center for Mind, Body, and Learning (now the Institute for Learning and Brain Sciences), working with Dr. Alicia Beckford-Wassink. Created software that was used in Jamaica for a language perception experiment.
2001 - 2002	**Web-based Spanish instructor**, Seattle, WA. Department of Spanish, UW. "Web-Assisted Spanish," SPAN 110. Four quarters of web-based curriculum.
1999 - 2001	**Spanish instructor**, Seattle, WA. Department of Spanish, UW. "Beginning Spanish," SPAN 101, 102, 103, & 110. Ten quarters of first-year Spanish curriculum.
1998 - 1999	**Benefits representative,** Seattle, WA. Merrill Lynch. Helped Spanish speaking clients with retirement benefit inquiries. Managed client accounts and prepared spreadsheet reports.
1996 - 1998	**Peace Corps volunteer,** Monteagudo, Bolivia. US Peace Corps. Worked alongside an indigenous grassroots organization. Managed communal food store project with UNICEF funds, resulting in the sustainability of seven stores. Trained individuals, instructed basic accounting, coordinated meetings, and wrote project synthesis in Spanish and English for UNICEF.

PUBLICATIONS

2004 — Acoustic contrasts of Spanish /d,t,r/ and tap. *Proceedings from the 20th Northwest Linguistics Conference.*

2003 — POSSE: Parsing of Spanish Syllable Experiment. *University of Washington Working Papers in Linguistic,* 22.

2003 — An optimality theoretic account of Spanish primary nominal stress. *Proceedings from the 19th Northwest Linguistics Conference.*

2001 — Identifying accurate Spanish pronunciation through waveform and spectrogram analyses. *Proceedings from the Western Conference on Linguistics*, 14.

CONFERENCE PRESENTATIONS

2005 — The relative degree of difficulty of L2 Spanish /d,t,r/ and tap. 79th Annual Meeting of the Linguistic Society of America, San Francisco, CA. January 6, 2005.

2004 — Acoustic contrasts of Spanish /d,t,r/ and tap. 20th Northwest Linguistics Conference, Seattle, WA. May 1, 2004.

2004 — Combining online and classroom Spanish instruction: Pedagogical and Technological challenges. Washington Association of Foreign Language Teachers, Tacoma, WA. March 22, 2003.

2003 — An optimality theoretic account of Spanish primary nominal stress. 19th Northwest Linguistics Conference, Victoria, BC, Canada. March 1, 2003.

2002 — An empirical study of the role of formal phonetic instruction in adult L2 learners of Spanish. SCMLA 59th Annual Meeting, Austin, TX. November 2, 2002.

2001 — Identifying accurate Spanish pronunciation through waveform and spectrogram analyses. Western Conference on Linguistics, Seattle, WA. October 28, 2001.

2001 Formal pronunciation instruction for beginning adult learners of Spanish. Washington Association of Foreign Language Teachers, Pasco, WA. October 12, 2001.

INVITED PRESENTATIONS

2004 Degree of difficulty of Spanish /d,t,r,∡/. UW Linguistics colloquium series, Seattle, WA. April 9, 2004.

2004 Second language pronunciation. Linguistics 101, Pronunciation Fundamentals, UW, Seattle, WA. April 1, 2004.

2003 An overview of computational linguistics. Linguistics 200, Introduction to Linguistics, UW, Seattle, WA. May 30, 2003.

GRANTS AND HONORS

2004 Nominated by the Department of Linguistics faculty for the UW Graduate School Dissertation Fellowship.

2004 $34,000 in funding through the Student Technology Fee to replace fifteen outdated PCs in the UW Language Learning Center. (Proposal #2004-60)

2003 $31,000 grant through the Student Technology Fee for desktop and multi-point video-conferencing equipment in the UW Language Learning Center. (Proposal #2003-17)

PROFESSIONAL SERVICE

2003 Participated on Department of Linguistics faculty search committee for a computational linguist. Screened applicants and interviewed candidates.

2002 - 2004 Elected “web-master” for the organization Linguistic Students at the University of Washington (LSUW). Email moderator for the two university wide lists.

2001 Reviewed abstracts for the West Coast Conference of Linguistics.

2000	Participated on a text book selection committee for the Department of Spanish.
1999 - 2000	Student representative to the UW Graduate & Professional School Senate.

PROFESSIONAL MEMBERSHIPS

2004 - present	Linguistic Society of America (LSA).

COMMUNITY SERVICE

2002 - present	Volunteer for King County Superior Court's Community Accountability Board. Helping local juveniles take responsibility for misdemeanor offenses.
2000 - 2004	Yearly participation in University of Washington World Languages Day, presentations on language to high school students.

COMPUTER SKILLS

- **Programming**: Perl, Matlab, C++ (experience); C#, C (coursework); PHP (practice)
- **Signal Processing**: Praat, SoundEdit, Goldwave, Peak
- **Operating systems**: Windows, Mac OS (Unix), Command line interface
- **Web development**: HTML, Dreamweaver, Flash (ActionScript), CGI, Courseware (Moodle), Content Management Systems (Mambo server), DBCS/SBCS
- **General software**: MS Word, Excel, Publisher, Power Point, Adobe Acrobat & Photoshop, SPSS statistical software

LANGUAGE SKILLS

- **Spanish** (near-native fluency)
- **Finnish, French, German, Japanese, Latin and Russian** (formal academic study)
- **Guaraní** (familiarity)

PERSONAL

- **Musician and singer-songwriter**: Engineered album in 2001.
- **Traveler**: Recent trips include Guatemala, Aruba, Costa Rica, and Peru.
- **Wood-worker**: Design and creation of handmade dovetailed boxes.