THE ACQUISITION OF RESYLLABIFICATION IN SPANISH BY ENGLISH
SPEAKERS

BY

DANIEL SCARPACE

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Linguistics
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2017

Urbana, Illinois

Doctoral Committee:

> Professor José Ignacio Hualde, Chair
> Professor Jennifer Cole
> Associate Professor Annie Tremblay, University of Kansas
> Associate Professor Duane Watson, Vanderbilt University
> Dr. Rebecca Foote

ABSTRACT

How do second language learners produce and recognize words in continuous speech when speech segmentation strategies differ between languages? This thesis explores phrases that have been affected by the process of resyllabification of consonants across word boundaries, where syllable and word boundaries are misaligned (e.g. *las alas* 'the wings' as [la.s#á.las] in Spanish). For production, learners must acquire a new gestural coordination pattern for producing these types of sequences across the word boundary, as well as avoid overmarking vowel-initial words with glottalization, which is the norm in some languages, such as English (Bissiri et al., 2011), but not Spanish (Hualde, 2005). For word recognition, models of speech segmentation that emphasize the consonant as a likely segment for the beginning of a word (such as the Possible Word Constraint (Norris et al., 1997) or the Syllable Onset Segmentation Heuristic (Dumay et al., 2001)) would correctly segment *salas* 'rooms' out of a string of sounds [la.sá.las], but would miss the possibility that *alas* is also a word. For languages with resyllabification, speakers must consider both segmenting the sequence into both V#CV and VC#V; that is, for /lasa/, both /s/ initial words and /a/ initial words must be activated equally in the lexicon. In English, where vowel-initial words are more likely to be accompanied by glottalization, for the same string /lasa/, /s/ initial words will be more strongly activated than /a/ initial words. Thus, English speakers learning Spanish must suppress their tendency to preclude the activation of lexical entries with different onsets (i.e. vowel initial words when preceded by a consonant).

Additionally, English-speaking learners of Spanish must learn an additional language-specific cue to word segmentation: the tap and trill contrast at the word boundary in Spanish. While both the tap and trill contrast between vowels word-medially (*pero* 'but' vs. *perro* 'dog'), at the word boundary, the tap is associated with a resyllabified /r/, as in *ver ocas* 'to

see geese', whereas the trill is obligatorily a word-initial segment, as in *ve rocas* 's/he sees rocks'.

Several tasks were designed to test both the production and perception of resyllabification. In all cases, native Spanish speakers were compared with learners of Spanish whose first language was English at various proficiency levels. Production data is given from a laboratory read-speech task in both Spanish and English designed to elicit VC#V sequences with three consonants: /n/ and /s/ to test overall resyllabification, and /r/ to test the allophonic distribution of this phoneme at the word boundary. The results show that second language learners struggle overall to resyllabify consonants: they favor marking vowel-initial words with glottal stops, although they do so less than they do in English. They do not produce the tap and the trill in the appropriate positions at the word boundary either, although this is complicated by the difficulty in producing these segments in any position.

Perception data is given in three different experiments: a word-monitoring task geared towards investigating /n/ and /s/ at the word boundary, a word-monitoring gating task investigating /r/, and finally a visual-world paradigm eye-tracking experiment investigating whether or not multiple lexical candidates are considered while listening to phrases where /n/, /s/, and /r/ straddle the word boundary. While ultimately perception results show that non-native speakers are able to successfully parse the intended word, even in resyllabification contexts, results show that they still have a bias towards consonant-initial words although this is lessened with increased proficiency. Additionally, native speakers appear to have this bias in perception as well, which is unreported in the literature. This finding calls into question the idea of complete resyllabification (Harris, 1983), wherein perception and production of V#CV and VC#V sequences are expected to be identical. Overall, this work adds descriptive contributions to word-juncture phenomena in Spanish and English and theoretical contributions to the ideas of lexical access and storage, phonological encoding, phonological universals in syllabification, and speech segmentation in a second language, as well as methodological concerns regarding eye tracking.

El mar y no tu telar. El mar y no el ejido, el mar y no su eco. Su cumbia y no su fría razón ando buscando. Su eco sensual malográndose oí. Oí el mar y no su cítara. Oh, Díos, ¿si con su sal forja cien aguas el mar y no tu telar, se asea la mariposa encubierta?

El marino tutelar. El marino elegido, el marino sueco. Sucumbía y no sufría razonando. Buscan dos huecos en su alma logran doce, oh. Y hoy el marino sucitará odios, ¿si con sus alforjas y enaguas, el marino tutelar se hace a la mar y posa en cubierta?

-Darío Lencini

Mairzy doats and dozy doats and liddle lamzy divey.
A kiddley divey too, wooden shoe?

If the words sound queer and funny to your ear, a little bit jumbled and jivey,
Sing "Mares eat oats and does eat oats and little lambs eat ivy."
-'Mairzy Doats', 1944

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Introduction and Motivation for the Study

Resyllabification, the process by which word-final consonants syllabify with the initial vowel of the following word (e.g. *las alas* [la.sá.las]), is a well-known phenomenon in many languages that has curiously not received much attention outside of theoretical phonology. In the traditional description, when preceding a vowel, a word-final coda consonant becomes the onset of the syllable of the next word, e.g. VC#V → V.C#V, where the syllable and word-boundaries do not align. In the most extreme view, VC#V and V#CV sequences are thus identical: the duration of each of the segments should be the same when comparing across the two conditions, as well as the gestural coordination between the consonant and the flanking vowels. This is the view held by traditional descriptions in Spanish (Hualde, 2005), many relating to processes in Spanish contingent on syllabic affiliation (Harris, 1983) and Optimality Theory (Colina, 2009): native speakers of Spanish should not be able to distinguish between intended productions of *la salas* ('you salt it') and *las alas* ('the wings') as after resyllabification has taken place, these phrases are homophonous. Whether or not resyllabification before or after a preplanning stage is uncertain: do speakers really produce the word-final /s/ of *las alas* as if it were a word-initial (and thus syllable-initial) /s/, sending the same gestural instructions to the articulators? It is not necessarily the case that resyllabification is complete: there may still be differences between V#CV and VC#V sequences. In Browman and Goldstein's Articulatory Phonology terms, a syllable-final consonant is coordinated with the offset of the previous vowel, whereas a syllable-initial consonant is coordinated with the onset of the following vowel (Browman and Goldstein, 1992). These timing differences should have recourses in consonantal duration: in the same

1

intervocalic context, a word-final /s/ might have a shorter duration than a word-initial /s/. Cross-linguistically, this seems to be the case (Shatzman and McQueen, 2006; Quené, 1993; Umeda, 1978; Spinelli et al., 2003), but there have been few studies (Estapa, 1989; Karn, 1996) investigating duration of word-final vs. word-initial consonants in Spanish, none conclusive, though it seems that differences in duration may be obscured in running speech rather than in isolated phrases. Some more recent work in Spanish (Hualde and Prieto, 2014; Torreira and Ernestus, 2012; Strycharczuk and Kohlberger, 2016) have also found for /s/ that word-final /s/ is more likely to be voiced, in addition to being shorter in duration. In Spanish, /n/ and /s/ (as well as /r/) are ideal consonants to investigate resyllabification because their presence contrast paradigmatically with person/number morphology in verbs, e.g. *dices* 'you say' vs. *dice* 's/he says', thus, it is relatively easy to construct minimal pairs, and verbal morphology is an important meaning-based point of grammar that native speakers and second language speakers alike should attend to. The consonants /n/ and /s/ constitute a type of a resyllabification where an acoustic-phonetic cue, duration, may correlate with word-position, whereas /r/ constitutes a type of resyllabification where an allophonic cue correlates with word-position: word-final /r/ in this context when resyllabified surfaces as the tap, [ɾ], whereas word-initial /r/ is always a trill, [r], e.g. the contrast between *ve rocas* [r] 's/he sees rocks' vs. *ver ocas* [ɾ] 'to see geese'.

In English, this phenomenon is known as internal open juncture (Nakatani and Dukes, 1977; Lehiste, 1960), or in the ESL literature as consonant-to-vowel linking. Because of only coronal consonants may end word s in Spanish, the consonants subject to resyllabification are limited to /n/, /s/, /r/, /d/, /l/, and in Peninsular Spanish, /θ/. English has no such phonotactic restriction: all consonants may appear word-finally, but there are many acoustic-phonetic cues to (original) syllable position: for example, for the voiceless stops /ptk/, in word-initial position they are accompanied by aspiration, but in other positions they are not. Thus, the phrases *keeps parking* and *keep sparking* are disambiguated by an aspirated [pʰ] in *parking* and an unaspirated [p] in *sparking*. Even in the VCV context there is never complete resyllabification across word boundaries in English: lie cold and like old are always in contrast Early work (Lehiste, 1960) on internal open juncture focused on describing these differences (and others, such as positional allophones of /l/ and /r/

in English), but as these consonants do not occur word-finally or do not have positional allophones in Spanish, the focus of this work will be on consonants that do not display positional allophones in English either: fricatives and nasals. For these consonants, there is the possibility that complete resyllabification occurs and phrases such as *see Mabel* and *seem able* are homophonous, just as they would be in Spanish. However, there is the possibility that resyllabification does not occur, and is instead blocked by the insertion of a glottal stop or the presence of glottalization of the word-initial vowel. In essence, this solution is simply another way to avoid vowel-initial syllables by epenthesizing a glottal stop into an otherwise empty onset, just as resyllabification avoids vowel-initial syllables by moving a consonant into the previously empty onset position. However, English and Spanish differ in the preference of one solution over the other: in Spanish, resyllabification is preferred, with glottalization being rare, if it ever occurs. In English, however, the picture is not clear: a number of studies report that glottalization happens sometimes in the VC#V context (from 25% to 100%), but it tends to be speaker specific (Umeda, 1978; Scobbie and Pouplier, 2010; Ladd and Schepman, 2003). This issue remains larger unexplored: previous research has tended to focus on other issues and has only commented on glottalization in this context as an aside. Even in the absence of glottalization if complete resyllabification has taken place: there may be other acoustic cues such as duration that distinguish between word-final and word-initial consonants that reflect different gestural organization of syllables.

While the preference of resyllabification over glottalization or vice versa may appear only to be a slight difference in the production of VCV sequences straddling the word boundary, the choice of either has implications for speech segmentation and lexical access. Models of speech segmentation that emphasize the consonant as a likely segment for the beginning of a word,such as the Possible Word Constraint or the Syllable Onset Segmentation Heuristic (Norris et al., 1997; Dumay et al., 2001), would correctly segment *salas* out of a string of sounds [la.sá.las], but would miss the possibility that *alas* is also a word. Theories of word-recognition have also focused on one word at a time (such as the Cohort model) beginning with the first segment as the word-initial segment also miss this generalization: when processing running speech and hearing a VCV sequence, Spanish speakers must consider both segmenting the sequence into both V#CV *and* VC#V; that is, for /lasa/, both /s/

initial words and /a/ initial words must be activated equally in the lexicon. In English, where vowel-initial words are more likely to be accompanied by glottalization, for the same string /lasa/, /s/ initial words will be more strongly activated than /a/ initial words. Most psycholinguistic models of speech recognition have tended to deal only with English, which is why resyllabification has not been given much thought in the literature. Although other models such as TRACE (Allopenna et al., 1998) and Shortlist (Norris, 1994) do not specifically deal with resyllabification, their mechanisms of continuous mapping can be applied to the data here.

Given these differences between English and Spanish in both production and processing, second-language speakers of Spanish whose native language is English have a difficult task in acquiring resyllabification: they must learn to suppress their consonant-initial bias in word segmentation, and their tendency to insert glottalization in the VC#V context. Most research in Second Language Phonology has focused on segments: either acquiring contrasts in the second language where the first has only one phoneme, or the phonetic realization of segments where the first and second languages differ. This study, in contrast, seeks to look at the acquisition of a speech segmentation strategy (for others, see White et al. (2010), also see Spinelli et al. (2003) and Tuinman et al. (2012) for lexical priming experiments) and its correlate in production. This study also compares the acquisition of a phonetic cue (duration, for /n/ and /s/) to the acquisition of a phonological cue (the contrast between [r] and [ɾ]) for marking the location of word boundaries: whereas the phonetic cue may be a cross-linguistic cue for word-initial segments, the tap and trill contrast is specific to Spanish and must be learned. These findings have implications for differing strategies of lexical access in different languages, and if it is possible to change strategies when acquiring a second language. If English speakers are unable to acquire resyllabification when learning Spanish, they will, in production, overmark vowel-initial words with glottalization, and in processing, be unwilling to accept vowels as the initial segment of words in the VCV context.

To accomplish these goals, the dissertation will be divided into several chapters: the remaining portion of this chapter (chapter I) will be a further exploration of the existing literature. Then, the general research questions of the study are outlined again, and predictions for the experiments will be made. First, the study needs to accurately describe

resyllabification from an acoustic point of view. Chapter II is a production study involving laboratory speech to see what the differences are between V#CV and VC#V sequences for (a) Native Mexican Spanish speakers, (b) L2 speakers who have not had explicit teaching about resyllabification, (c) L2 speakers who have had explicit teaching about resyllabification, and (d) advanced proficiency L2 speakers. This section will thus see if non-native speakers are able to acquire the production patterns of native Spanish speakers, resyllabifying word final consonants across the word boundary. Chapter III focuses on perception and processing, involving several experiments in which both native and non-native speakers must be sensitive to various types of resyllabification in Spanish. Chapter IV will involve a general discussion of all of the experiments and their implications for both L1 and L2 phonology and speech segmentation and lexical access in general.

1.2   Theoretical concerns regarding word-segmentation and resyllabification

It seems like the beginning of a syllable is a good place to hypothesize the beginning of a word. Why might this be? When is a word is found at the beginning of a sentence, or in isolation (as it is found in its entry in the mental lexicon), the beginning of the word is necessarily the beginning of a syllable. As explained, different languages have different processes through which listeners recognize words in a speech stream: through acquisition of a first language, speakers are attuned to systems used in their native language. This can be a drawback when learning a second language, due to differences in rhythm, phonotactics, or phonetic/phonological processes that occur between words (Cutler, 2001). Other possible difficulties may stem from orthography, because the word boundary is clearly delineated with a space in writing but not in speech, or psycholinguistic principles such as the Possible Word Constraint or the Syllable Onset Segmentation Heuristic, which state that when listening to running speech, listeners align syllable onsets with word-initial boundaries (Cutler and Norris, 1998).

These psycholinguistic findings mirror high-level constraints from the Optimality Theory (OT) literature. The first, ONSET, states that every syllable must have an onset, disallowing

5

syllabification like [las.á.las]. There are two possible repairs: languages like Spanish prefer to resyllabify the word-final consonant into the following syllable, [la.sá.las], but this violates the constraint ALIGN, which assigns a violation for the misalignment of syllable boundaries with some other boundary, here, the right edge of the word. However, languages like English and German opt for another solution, epenthesis of [ʔ] in [ən.ʔæ.pl̩] 'an apple', which violates the constraint DEP, but does not violate ALIGN[1]

1.3 Review of literature in production of word-segmentation cues in Spanish

1.3.1 Resyllabification of /n/ and /s/

In Spanish, resyllabification is known as *enlace* 'linking', *encadenamiento* 'interlocking', or *resilabificación*. Phonologically, Spanish is said to have complete resyllabification; that is, a word-final coda consonant fills the onset position of the following vowel-initial word (Hualde, 2005). In /n/-velarizing and /s/-aspirating dialects where allophonic rules that apply to coda consonants only (changing these coda segments to [ŋ] and [h], respectively), this resyllabification is said to apply after the consonant weakening, producing forms such as *los otros* [lo.ho.tɾos] 'the others'; however, this may be rare depending on the dialect: examining two /s/-aspirating dialects, Bybee (2000) reports maintence of [s] in the VC#V condition in 88% of tokens in Buenos Aires, compared to 18% in Cuba. In Ecuador, word-final /s/ may become voiced to [z] before vowel-initial words (Chappell, 2011). This is also the case in the Spanish as spoken in Catalonia, due to the influence of Catalan, which also voices word-final /s/ before vowel-initial words (Davidson, 2014). In these dialects of Spanish, the realization of /n/ or /s/ can thus be an allophonic cue to the word-affiliation of the consonant: e.g. *la salas* and *las alas* are kept distinct. Finally, in some dialects of Spanish in contact with

---

[1]It should be noted that in some sense the insertion of a glottal stop misaligns the left side of beginning of /æpl̩/ with the beginning of syllable [ʔæ]. However, this glottal stop may be realized as creaky voice on the initial vowel [æ], among other things that will be discussed in Chapter 2. Despite any slight misalignment, the glottal stop is a good indicator of a word boundary, and as English does not have a phonological glottal stop as part as its consonant inventory, although it may fill the onset slot syllabically, speakers may not judge it to be a full consonant in this position. Nevertheless, the previous word, [ən] is aligned with a syllable boundary on its right edge, unlike in the resyllabified variant, [ə.næ.pl̩].

languages that have a phonemic glottal stop /ʔ/, especially indigenous languages of Central America, a glottal stop or glottalization is possible in both the V#V context as well as the VC#V (resyllabified) context (Michnowicz and Kagan, 2016).

In all other dialects, however, there is the possibility that the output [la.sá.las] maps onto both *la salas* and *las alas*. For the phonological perspective on Spanish resyllabification see Harris (1983); Hualde (1991); Face (2002); Colina (2009); more recently, for whether or not resyllabification should be better viewed as ambisyllabicity in Spanish, see Strycharczuk and Kohlberger (2016), the remainder of this paper focuses on the phonetic reality of resyllabification. Some recent studies have nonetheless questioned whether there is some *phonetic* difference between word-initial and word-final-before-vowel consonants. Both Torreira and Ernestus (2012); Hualde and Prieto (2014) found in Peninsular Spanish that word-final /s/ is more likely to be voiced and shorter in duration than word-initial /s/, although some degree of overlap in duration between word-initial and word-final /s/ remains, allowing the possibility of ambiguity. These studies were extracted from speech corpora. These findings were replicated, also by Peninsular Spanish speakers, in a laboratory setting in Strycharczuk and Kohlberger (2016). Karn (1996) did a more controlled study of production in which 10 Mexican Spanish speakers produced both word-initial and resyllabified /s/ and found slight differences in duration. However, none of these studies have tested whether or not Spanish speakers perceive resyllabified /s/ is the same as word-onset /s/. In contrast, no studies have looked at resyllabification of /n/. Although duration was not measured, Torreira (2007) looked at pitch contour alignment during the sequence /la/ between 'canonical syllables' where the /la/ formed the beginning of a word, in *mi lana* 'my wool' and 'resyllabified consonant', where the /l/ originally belonged to the previous word but was resyllabified into an onset, in *el ama* 'the housekeeper'. The start of rising accents should begin at the beginning of the syllable (/la/), resulting in similar curvature coefficients for the pitch contour with respect to the beginning of the /l/ segment. This hypothesis was confirmed, indicating that despite word affiliation, the /l/ had been resyllabified and acted as the onset of the next syllable.

In the development of Latin into Modern Spanish, resyllabification underwent a number of changes. Latin, like Spanish, divided VCV sequences as V.CV within words and across

words as well. Evidence of resyllabification in Latin comes from Of the eight consonants that could end Latin words, only three (/n/, /l/, and /s/) remain word-final in modern Spanish. The other remaning consonants that are word-final in Spanish (/d/, /ɾ/, /θ/) arose through deletion of final vowels, e.g. AETATE > *edade* > *edad* (Penny, 2002, p.86). Old Spanish, which had a contrast between /s/ and /z/ (later merged in all modern dialects to /s/), additionally voiced word-final /s/ before vowel-initial words (i.e. resyllabified /s/), as do modern Catalan and Portuguese. In this way, speakers of Spanish could use the presence of [z] as a cue for aligning word boundaries, but with the loss of all intervocalic voiced fricatives in the language, phrases such as *ha sido* and *has ido* have become homophonous.

1.3.2 Special concerns of the Spanish rhotics

Table 1.1: Distribution of Spanish rhotics relevant to this paper

Word-internal [r]	Word-internal [ɾ]	
carro 'car'	*caro* 'expensive'	
perro 'dog'	*pero* 'but'	
Word-initial [r] (only)	Word-final before C or pause (variable)	Word-final before V ([ɾ] only)
rico 'rich'	*prestar dinero* 'to lend money'	*prestar atencion* 'to pay attention'
rojo 'red'	*ser capaz* 'to be capable'	*ser amigos* 'to be friends'

The tap [ɾ] and trill [r] contrast phonemically between vowels in Spanish. Some minimal pairs include *ca[ɾ]o* 'expensive' vs. *ca[r]o* 'car' and *pe[ɾ]o* 'but' vs. *pe[r]o* 'dog'. In other positions, however, there is neutralization of the contrast. As previously mentioned, word-initial <r>invariably surfaces as the trill [r], although this segment has various phonetic realizations in different Spanish-speaking regions. Following Hualde (2005), this paper adopts /r/ as the phoneme in words like *rata* 'rat'. Word-finally, however, [ɾ] and [r] are in free variation before a pause or a consonant, although the trill may be more emphatic. Before a vowel, the only word-final rhotic found is [ɾ]. For this reason, Hualde (2005) adopts /ɾ/ as the phoneme found word-finally in words like *mar* 'sea', with a optional strengthening rule that permits /ɾ/ → [r] before consonants and pauses. Strengthening to a trill is possible in this context only (i.e. not before a vowel-initial word) likely in order to keep phrases like *ver ocas* and *ve rocas* (which obligatorily begins with a trill) distinct. Note that due to this

8

optionality, a single word may surface with both a tap and a trill, despite the phonological contrast between these two segments elsewhere in words: se[ɾ] o no se[r] 'to be or not to be'. Importantly, in Spanish orthography, the tap/trill contrast is only represented word-internally between vowels: /r/ is written <rr> as in *perro*, and /ɾ/ is written <r> as in *pero*. Elsewhere, only <r> is used, even in the *ve rocas*/*ver ocas* pair that does contrast the tap and trill. This paper follows the aforementioned phonological analysis, although it should be noted that the tap/trill contrast is complex in Spanish and that this analysis may not capture the organization of the two sounds at any point during language acquisition for many English speakers learning Spanish. Indeed, the ability to properly discriminate between sounds that are otherwise non-contrastive at the word boundary is at the heart of the challenge facing second-language speakers in segmenting speech.

English speakers who learn Spanish are faced with a number of problems when acquiring the complete tap/trill contrast in Spanish. Previous research has mostly focused on the production of the tap and trill, but if perception precedes production, it is just as important to focus on Spanish learners' abilities to perceive the contrast. At the most basic level, learners must be able to hear the contrast between the trill, a segment that does not exist in English, and the tap, a segment that exists in English only as an allophone of /t/ and /d/. Rose (2010) performed a comprehensive set of experiments focusing on English learners of Spanish and their ability to discriminate various sounds of Spanish, including the tap and trill. To capture what might be the initial state in acquisition, Rose asked native speakers of English with little or no experience with Spanish to identify the sound present in words such as *caro*/*carro* as categories in English in a cross-language identification task. The Perceptual Assimilation Model (PAM) proposed by Best (1995) suggests that the success of discriminating non-native contrasts is related to the relationship of the sounds in the second language to existing categories in the first language. The trill /r/ was identified as /ɾ/ 96.2% of the time with a 2.3 goodness rating, whereas the tap /ɾ/ was identified as /ɹ/ 57.7% of the time with a 2.3 goodness rating and 30.4% as /d/ with a 2.8 goodness rating. This uncategorized vs. categorized contrast suggests that discrimination should be very good according to Best (1995), but poor according to Guion et al. (2000), due to the closeness in phonological space of the two rhotics. Rose then tested 75 native speakers of English (15

9

speakers in 5 different proficiency groups) in an AXB task for the minimal pairs *caro/carro* and *quería/querría* along with several other filler items in three blocks. Discrimination for the tap/trill contrast was already high for English speakers who had never studied Spanish (80.2%) and there was no statistical difference among the four groups who had studied Spanish (mean accuracy scores from 86.7% - 94.4%). Although these data suggest that English learners of Spanish can discriminate the two sounds at a nearly native-like level, the low number of stimuli used that were repeated three times might have made the task easier.

Furthermore, sound discrimination is not enough to acquire a contrast: learners must be able to correctly associate the sound that they heard with the correct word; this reflects correct lexical encoding of the contrast. When beginning to learn Spanish, they may or may not be assisted by Spanish orthography, which presents the contrast word internally between vowels as a contrast between the graphemes <r> and <rr>. On one hand, this may help learners tell that pero and perro are distinct words, but English orthography exhibits a poor grapheme-to-phoneme mapping, and English <r> and <rr> both map to /ɹ/, e.g. *arid* and *arrow*. As mentioned above, Spanish orthographic rules do not clue learners into the pronunciation of <r> in other positions, such as that the <r> in *rana* 'frog' must be pronounced as a trill. Initially, English speakers beginning to learn Spanish may be able to keep the two sounds distinct by first encoding /ɾ/ in words such as *caro* as /d/, as the tap is an allophone of /d/ in English. Thus *caro* and *carro* would be minimal pairs (/kado/ vs. /karo/), but not in the same way Spanish speakers encode them (/kaɾo/ vs. /karo/). However, not all taps in Spanish can be interpreted as allophones of English /d/ because they can occur in stressed as well as unstressed syllables, and at some point must encode them separately from /d/. Perhaps the greatest challenge presented to second-language speakers is the acquisition of allophony based on word context: in *ser o no ser* 'to be or not to be', the first <r> is necessarily a tap, but the second <r> is likely to be a trill; thus, the same word, *ser*, could appear in the same phrase with two different pronunciations. It follows that a proficient listener must pay attention to the /r/ - /ɾ/ contrast between vowels but ignore it entirely elsewhere.

Latin had no trill /r/; all written rhotics (<r>) represented the tap. The appearance of the trill /r/ came about in two ways. First, Latin had a singleton/geminate contrast

10

for word-internal consonants in between vowels. In the development of Latin into Spanish, geminate consonants simplified to singleton consonants (CUPPA > *copa* 'wine glass'), with the exception of the alveolar sonorants /nn/, /ll/, /rr/. /nn/ and /ll/ both palatalized into /ɲ/ and /ʎ/ respectively, while /rr/ became /r/. Thus, the singleton/geminate contrast in Latin is preserved as the tap/trill contrast in Spanish. Secondly, through fortition, initial /ɾ/ became /r/, matching the usual outcome of initial consonants being identical to their word-internal geminate counterparts (e.g. for /t/, TERRA > *tierra*, GUTTA > *gota*). All other positions retained the tap. Many word-final taps arose from the loss of a word-final vowel: MARE > *mar*. In similar situations where the loss of final /-e/ would have resulted in a word-final geminate, the final vowel was not lost (TURREM > *torre* 'tower', although *torre* may have resulted from paradigm leveling from the plural TURRES > *torres*, replacing former *tor* (Penny, 2002, p.82). Word-final /ɾ/ in Latin became word-internal through metathesis and retained as [ɾ] (SEMPER > siempre). Syllable-final /ɾ/ was maintained as [ɾ] word-internally as well: CORPUS > *cuerpo* 'body'. Where did the possibility of syllable-final [r] come from? The pronunciation of a trill may have been maintained when an intertonic vowel was lost following geminate -RR-, as in CARRICARE > *cargar* 'to load' (related to CARRU > *carro* 'cart'), leading to alternate pronunciations of <r> in syllable-final position. However, it is also possible that the free variation between the tap and trill in these positions is due to the lack of possibility of contrast, and the salience of the word-initial position (which only admits [r]).

1.4 Review of literature in production and perception of word-segmentation cues in English

English has several cues that distinguish minimal pairs in VCV sequences such as *lie cold* and *like old*. For all types of consonants, the word-final variant is shorter than the word-initial variant, e.g. the /m/ in *see Mabel* is longer than the /m/ in *seem able* (Lehiste 1960). For voiceless stop consonants /p, t, k/, the presence or absence of aspiration is an additional cue. The /k/ is aspirated in when word initial in *lie cold* but is unreleased in *like old*. Additionally, a glottal stop or laryngealization is frequent at the word juncture when

the following word begins with a vowel (Altenberg, 2005; Lehiste, 1960; Dilley et al., 1996; Ladd and Schepman, 2003). This is a particularly strong phonetic cue for native speakers for word segmentation Altenberg (2005); Bissiri et al. (2011), but when absent, native speakers are also sensitive to the alignment of the F0 contour rise. Norma Nelson is distinguished from Norman Elson in that the minimum F0 is aligned with the beginning of the stressed syllable in the last name. Because there is no resyllabification, the beginning of the stressed syllable is at the beginning of the consonant /n/ in *Norma Nelson* but at the beginning of the vowel /e/ in *Norman Elson*, both being the initial segment in the surname. When the F0 contour was manipulated in an experiment by Ladd and Schepman 2003, native English speakers gradiently changed their judgment of which name they were hearing.

According through analysis of metrical verse in Minkova (2003), Old English had [ʔ]-insertion before word-initial stressed vowels, which blocked resyllabification of word-final consonants across word-boundaries. This process was relaxed in Middle English, such that the insertion of a glottal stop became optional. In this period, many /n/-initial words lost their initial consonant due to the misanalysis of juncture due to frequent use with the indefinite article *a(n)*, as in ME *a noumpere > an umpire, a naddre > an adder, a naperon > an apron*, and *a nauger > an auger*. On the other hand, vowel-initial words could gain an initial /n/ in the reverse of the same process, as in *an eke-name > a nickname*, and *an ewt > a newt* (Bottke, 1943). In present day English, this process of [ʔ]-insertion may be optional, and may depend upon dialect.

1.5 Review of literature in production and perception of word-segmentation cues in other languages

1.5.1 French

By far the most closely related language to this study is French: French has resyllabification (called *enchaînement* in the French literature), which functions as it does in Spanish (Fougeron, 2007). French also has liaison, a latent consonant at the ends of some lexical items that resurfaces only before vowel-initial words, e.g. the final /ʁ/ in *dernier* 'last' is

not pronounced in *dernier train* 'last train', but is pronounced in *dernier oignon* 'last onion', making it homophonous with /ʁ/-initial words, such as *dernier rognon* 'last kidney' (Spinelli et al., 2003; Tremblay and Spinelli, 2014). French speakers thus must not only know that word-final consonants sometimes become onsets of the following word, but they must also be aware of liaison consonants that only occur in this resyllabification context. Spinelli et al. (2003) found that listeners were sensitive to the acoustic-phonetic cues that differentiated liaison consonants in *dernier oignon* from word-initial consonants in *dernier rognon*, and liaison did not block activation of vowel-initial words. Thus, French speakers have resolved the apparent difficulty in processing words where there is a syllable/word mismatch.

Several investigators have also looked at the ability of second-language speakers of French to parse liaison consonants. If the speech segmentation system of French speakers is specially adapted to deal with liaison and other types of resyllabification in French, it is thus necessary for non-native speakers to acquire this new system as they are learning the language. Dejean de la Batie and Bradley (1995); Shoemaker (2009); Tremblay (2011); Tremblay and Spinelli (2013); Gustafson and Bradlow (2016) all found that L2 speakers of French were able to successfully process resyllabified or liaison words, although to different extents, and were not always sensitive to acoustic-phonetic or distributional cues in the same way that native speakers were. These studies, however, did not attempt to deal with the role that glottalization may play in the production of these words.

1.5.2   Other languages

There are other languages that have phonological or phonetic cues to speech segmentation, although these are less well-studied. Finnish speakers, for example, exploit the fact that vowel harmony applies only within the domain of a word. If a Finnish speaker hears a word with several front vowels, but then hears a back vowel, the speaker knows that the back vowel must belong to a new word (Suomi et al., 1997).

Broselow (1984, 1988) informally reported on the difference between Egyptian Arabic and English syllabification across word boundaries and its implication for parsing by second-language speakers of Egyptian Arabic. Broselow finds that the imposition of a non-native

syllable structure can have ramifications on L2 processing and production. Whereas Egyptian Arabic has complete resyllabification across word boundaries (*fadda lilwalad* 'silver to the boy' = *faddal ilwalad* 'he preferred the boy'), English does not (Broselow argues for ambisyllabicity of some word-final consonants before vowels instead). She mentions that "the string /miʃana/ 'not I' will often be heard incorrectly as consisting of two non-existent words, *mi ʃana, and the typical pronunciation of this string by the English speaker will also be criticized as non-native by Egpytians" (Broselow, 1984, p.254). However, there is no evidence given beyond that English speakers often perceive word-final consonants as the beginning of the next word, and it is unclear what exactly in pronunciation marks the difference between a word-final ambisyllabic consonant and a word-final resyllabified consonant.

Vogel (1991) maintains that languages may differ in the domain where resyllabification may apply: Dutch is conservative this regard, only resyllabifying within a phonological word. Dutch has syllable-final devoicing, and in *roo*/d/+*ig* 'red-ish', the /d/ is resyllabified and remains voiced ([ro:d...]), whereas the suffix -*achtig* forms a separate phonological word, and when attached to *roo*/d/, the stem-final /d/ remains syllable final, and is devoiced to /t/ ([ro:t...]). Languages like Spanish and French do resyllablify across word boundaries, although they do not resyllabify across larger phonological domains, such as across a phonological phrase.

1.6   What is taught to L2 speakers of Spanish?

Unlike other aspects of linguistic structure such as morphology and syntax, Spanish pronunciation is often not taught in L2 classroom, especially at introductory levels when lexical items and phonemic categories begin to be formed. This may be due to the lack of training of teachers in phonology and the lack of materials in Spanish textbooks (Arteaga, 2000). If at all present, pronunciation instruction focuses on basic orthography-to-phoneme segmental issues, and not at all on suprasegmentals. Arteaga recommends that "students should be taught, for example, that the preferred open syllable structure of Spanish applies across word boundaries as in *los hombres hablan* [lo sóm bre sá βlan]" (Arteaga, 2000, p.343). This does not appear to happen in lower-level classrooms: if syllabification is taught at all, it is

limited to word-internal syllabification only, necessary in some respects for correct accent-mark placement. At a more advanced level, textbooks geared towards teaching phonology as well as improving English learners' Spanish pronunciation are available: in this section, I examine three popular texts, *Spanish pronunciation: Theory and practice* (Dalbor, 1997), *Sonidos en contexto* (Morgan, 2010), and *Fonética y fonología españolas* (Schwegler et al., 2010) for references to resyllabification as well as the treatment of /r/ at the word boundary.

Written entirely in English, Dalbor's *Spanish pronunciation: Theory and practice* spends a considerable amount of text on suprasegmentals prior to describing the consonants and vowels themselves: Chapter 5-9 discuss stress, intonation, and rhythm while Chapter 13 discusses syllables, and Chapters 18 and 29 discuss word-boundary phenomena. When discussing syllabification, Dalbor refers to the preference that Spanish has for open syllables and that *enlace* 'linking' exists and can cause problems as words are harder to pick out of the speech stream than in English, pointing out that phrases such as *el hijo/elijo*, *un aparte/una parte*, and *la sabes/las aves* are ambiguous. Dalbor also gives an example of a Spanish-speaking child suprised to learn that the word for grapes is *uvas*, and not **suvas* because she had only heard the word in a phrase like *las uvas*, *tus uvas*, *más uvas*, etc., with a resyllabified /s/. In production, Dalbor reports that English-speaking students mistakenly try to indicate word boundaries in Spanish through the use of a glottal stop when they should be linking consonants across word boundaries: "Such misguided efforts interrupt the smooth flow of Spanish and create an unpleasant choppy sound. Instead of an uttterance like *Los alumnos están en el aula* sounding like one long word in Spanish, [losalúmnosestáneneláula], it comes out *[losʔalumnosʔestanʔenʔelʔaula]" (Dalbor, 1997, p.268).

Regarding the distribution of /r/, Dalbor decribes the trill as being found intervocalically as in *carro* and beginning a word, as in *rico*, and that it can optionally occur word-finally before a pause. The tap is found elsewhere–crucially there is no mention of the obligatory presence of [ɾ] word-finally before a vowel (*ir a*), in fact, Dalbor claims "/r/ and /ɾ/ contrast in only one position in Spanish–intervocalically within a word" (Dalbor, 1997, p.258).

Fonética y fonología españolas also devotes a chapter to *encadenamiento*, and the preference that Spanish has for open syllables. It is explained by invoking the *grupo fónico* 'sound group', the combination of sounds made between pauses or breaths of air. The linking

of word-final consonants to word-initial vowels of *encadenamiento* applies the whole *grupo fónico* is produced as if it were one long word. The text also includes exercises to help students resyllabify consonants across word boundaries properly. Schwegler et al. focuses both on how the correct syllabification aids in production and perception: "Articular a la manera anglosajona *en + es + tos + a + ños* en lugar de *e + nes + to + sa + ños* es hablar con un acento extranjero. A menos que el estudiante no nativo supere la tentación de siempre dividir sílabas entre palabras, no podrá adquirir jamás una pronunciación española nativa o seminativa."[2] (Schwegler et al., 2010, p.100). Although in a subsequent chapter regarding vowel-vowel contact at the word boundary the authors introduce the idea that English marks vowel-initial words with a glottal stop, and that English learners of Spanish should not produce a glottal stop in, e.g. *vi a* [ʔ]*unos chicos*, unlike in Dalbor (1997) the connection to unresyllabified consonants is not explicitly made; that is, that the tendency for students is to produce *las alas* as [las.ʔá.las]. However, the authors are clear to state that the insertion of short pauses or glottal stops in the middle of phonological phrases is non-native sounding (Schwegler et al., 2010, p.137).

The text is, however, explicit in its treatment of the distribution of [r] and [ɾ] at the word boundary. Schwegler et al. are clear to note that the pronunciation of orthographic <r> varies considerably depending on its position within the word. For [r], students are given a list of words beginning with <r> (*Rosa, rojo, rico, ridículo*, etc.) which all obligatorily begin with the trill. The text then contrasts this with [ɾ], which can appear to be word-initial when resyllabified, but obligatorily is a tap: *prestar̲ atención* → *presta—*[ɾ]*atención* 'to pay attention' (Schwegler et al., 2010, p.281). Although word-final /ɾ/ can vary between [r] and [ɾ] before consonants and a pause, the authors point out that maintaining word-final /ɾ/ before a vowel (i.e. resyllabified /ɾ/) as a tap avoids confusion between *ver osas* [beɾósas] 'to see female bears' and *ve rosas* [berósas] 's/he sees roses' (Schwegler et al., 2010, p.285).

Terrell Morgan's *Sonidos en contexto* is a less technical text designed to cover the basics of Spanish pronunciation with more emphasis on communicative activities involving naming

---

[2] "Articulating in the English way *en + es + tos + a + ños* instead of *e + nes + to + sa + ños* is to speak with a foreign accent. Unless the non-native student gets past the tendency of always dividing syllables between words *[i.e. coinciding with words]*, s/he will never be able to acquire a native or semi-native Spanish pronunciation."

objects in pictures, finding evidence of phonological processes in signs, billboards, or advertisements, and reading passages. Resyllabification is succinctly explained through syllable diagrams of *con*, *eso*, and *con eso* together (Morgan, 2010, p.34), along with speaking activities designed to elicit resyllabification, such as counting the number of objects on the page (e.g. 'three stars': *tres estrellas*, 'two sheep' *dos ovejas*). There is no mention of what resyllabification should sound like, or how Spanish may differ from English in this respect, or any recommendations not to pause or introduce glottal stops in these positions.

With respect to /r/, Morgan presents the most basic facts: that although [r] and [ɾ] contrast intervocalically within words, there is neutralization in other positions. Word-initial <r> is [r], and word-final <r> belongs to the phoneme /ɾ/, and is thus is always [ɾ] (no mention is made of free variation between the tap and trill in this position). Among the many examples given for /ɾ/, only a scant few are word-final before a vowel, like *por eso* 'that's why' (Morgan, 2010, p.346) and none in the activities. Resyllabification is not brought up in this chapter.

To conclude, while all three texts do talk about resyllabification and stress the importance of word-initial <r> being a trill, not all of the texts clearly state how to avoid English-like productions. Only *Fonética y fonología españolas* is clear to note the perceptual importance of the tap/trill contrast at the word boundary.

CHAPTER 2

PRODUCTION

2.1 Overview

This section describes two production experiments: one in Spanish and one in English. These experiments involve read speech: participants read a series of sentences out loud presented to them on a computer screen. Each of the experiments were designed to elicit the production of resyllabified consonants /n, s, r/ across word boundaries. The goals and research questions of the experiments are as follows:

1. For productions by native speakers of Spanish, are there the high rates of resyllabification as hypothesized by the Spanish literature? Is resyllabification is complete, also as is hypothesized? If so, V#CV and VC#V sequences should be identical phonetically.

2. For productions by native English speakers in English, what is the rate of resyllabification for these speakers in English? Do speakers keep V#CV sequences distinct from VC#V sequences, through a higher rate of glottalization, as is hypothesized?

3. For productions by native English speakers in Spanish, to what extent are their rates of resyllabification/glottalization are similar to their same productions in the English experiment? It is hypothesized that they will be more similar to their English productions as compared to Spanish native speakers rates of glottalization/resyllabification. finally

4. For productions by native English speakers in Spanish, are there are asymmetries between their productions of /r/ and /ɾ/ word-medially (e.g. in *ca<u>r</u>o* / *ca<u>rr</u>o*) and at the word-boundary (e.g. *escrito<u>r</u> español* / *bebé <u>r</u>ápido*? It is hypothesized that even at the lower proficiency levels they may attempt to keep word-internal /r/ and /ɾ/

distinct word-medially, but at the word boundary it is likely that they do not know or produce this allophonic distinction.

The first section explains the data analysis for the experiments; the following sections describe the experimental procedures; finally the results are shown and discussed.

2.2 Background of analysis of production data

2.2.1 Glottalization

Glottalization of vowel-initial words has been investigated in several aforementioned studies (Umeda, 1978; Dilley et al., 1996; Garellek, 2012). Glottalization can appear in several forms, from a single glottal gesture to a change in voicing quality (creaky voice), all of which represent some form of glottal adduction that results in irregular vocal fold vibration. Rates of glottalization have been reported impressionistically or through visual inspection of the spectrogram or waveform (Umeda, 1978), or additionally investigated further (Dilley et al., 1996; Garellek, 2012). Dilley et al. (1996); Redi and Shattuck-Hufnagel (2001) give the follow criteria: (1) *aperiodicity*, irregularity in duration of glottal pulses from period to period, (2) *creak*, prolonged low f0 accompanied by damping of glottal pulses, (3) *diplophonia*, regular alternation in shape, duration, or amplitude of glottal periods, and (4) *glottal squeak*, a sudden shift to relatively high sustained f0 of low amplitude. Redi and Shattuck-Hufnagel (2001) found variability in the choice of manifestation of glottalization highly dependent upon speaker. Garellek (2012) used automatic extraction of the harmonics H1 and H2 to evaluate glottalization of vowels.

Given the number of tokens to be analyzed in this study, a number of automatically extracted acoustic cues related to glottalization will be analyzed. (1) Duration: it is hypothesized that longer vowels will include irregular voicing. (2) jitter: a higher amount of jitter likely to correlate with glottalization (Garellek, 2013), (3) Harmonic-Noise-Ratio is well correlated with modal vs. non-modal voicing, see Garellek (2013) for further information amd (4) voice breaks (failure to track f0) during the vowel. These acoustic measures should capture any of the four aforementioned types of glottalization. However, given that previous

studies did not use these measures in the postconsonantal context, as will be done here, the results may not be comparable.

2.2.2   Consonantal Duration

In tokens where there is no glottal stop or glottalization present, resyllabification may not be complete; that is, there still may be acoustic differences between V#CV and VC#V sequences, namely in the duration of the consonant. Word-final consonants may be shorter than word-initial consonants (Hualde and Prieto, 2014; Torreira and Ernestus, 2012; Strycharczuk and Kohlberger, 2016) as has been reported for /s/, but this pattern may not hold for /n/. The robustness of these patterns should be further examined for native speakers of Spanish. In Spanish, the duration of /s/ and /n/ are relatively straightforward to measure. The duration of /s/ is measured by marking the onset and offset of random energy in the waveform. The duration of /n/ is measured by inspecting the signal for the attenuation of energy: this change is more often than not abrupt due to complete consonantal closure.

2.2.3   Taps and Trills

The production of the Spanish tap, /ɾ/ is relatively stable across dialects. It is similar to the production of the English flap ([ɾ]) in words like *metal*. The sound is manifest on the spectrogram as a very short occlusion, around 20 ms. See Figure 2.1 for an example.

On the other hand, the Spanish trill /r/ has many dialectal variants (Hualde, 2005), but the most common and normative pronunciation is an alveolar trill, similar to the tap but with two or more occlusions. This segment is inherently longer than the tap, so a simple difference in duration will distinguish the two segments. See Figure 2.2 for an example.

What about productions of the American English-like approximant [ɹ]? Non-native speakers of Spanish may substitute [ɹ] for either /r/ or /ɾ/, resulting in a drop in F3 but no sudden drop in intensity characteristic of the tap or trill that would indicate an occlusion. See Figure 2.3 for an example. In previous studies (Johnson, 2008), non-native speakers produced [ɹ] for the word-initial /r/ at a rate of over 70%, and less than 30% for intervocalic /r/ even at the level of fourth-semester Spanish. As reported in Johnson (2008), low-proficiency
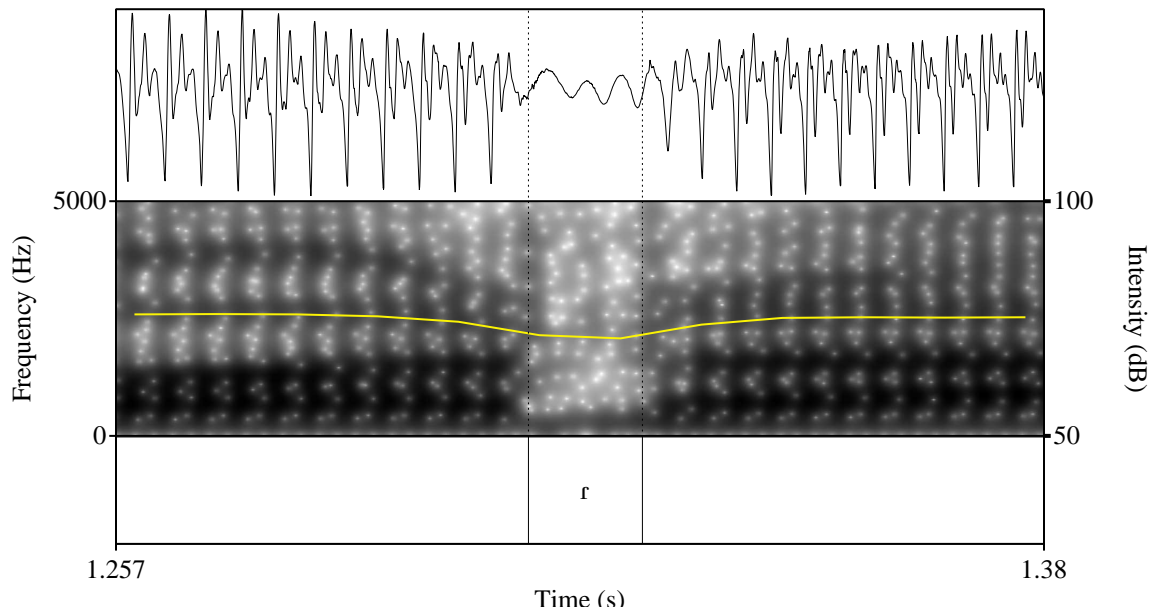
Figure 2.1: Example of a canonical intervocalic tap in Spanish. The word shown is *ahora* 'now'.
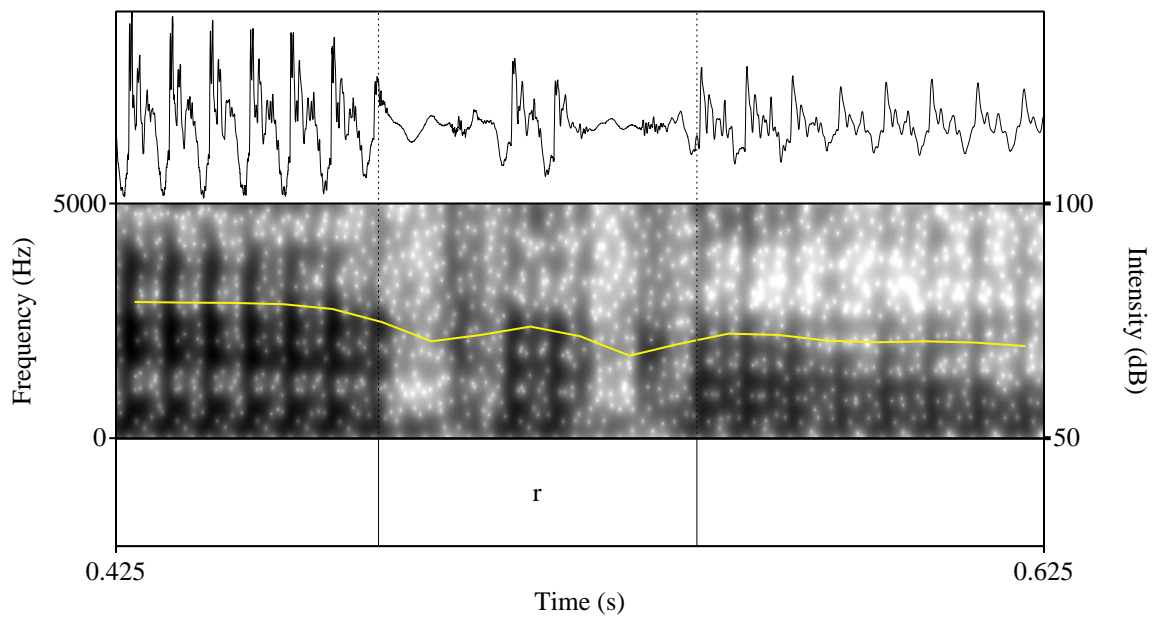
Figure 2.2: Example of a canonical intervocalic trill in Spanish. The word shown is *perros* 'dogs'.

21
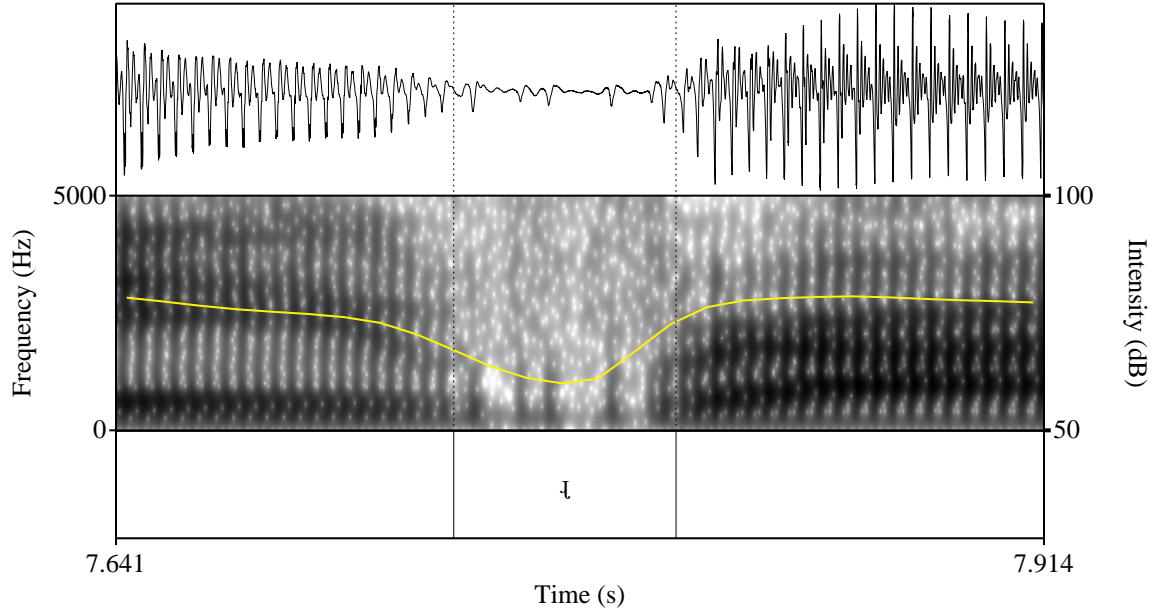
Figure 2.3: Example of an American English [ɹ] as an word-initial /r/ produced by a second-language learner of Spanish. Although there is a drop in amplitude, the consonant is very long, which does not suggest a tap or trill production. The word shown is *ratas* 'rats'.

speakers produce taps for both target taps and target trills at a rate of over 50% in word-medial position, but no data is reported for how often American [ɹ] is produced for target taps, nor is there any data for word-final <r>before a vowel-initial word.

2.3   Experiments

2.3.1   Methods

Participants

20 native Spanish speakers were recruited from Querétaro, Mexico, at the Universidad Autónoma de Querétaro. They were paid roughly $7 for their participation.

20 native English speakers were recruited from 300-level Spanish classes at the University of Illinois, Urbana-Champaign. They had all begun learning Spanish after the age of 12.

Table 2.1: List of stimuli for Experiment 1a (Spanish)

/s/		/n/	
V#CV	VC#V	V#CV	VC#V
la sala 'the room'	*las alas* 'the wings'	*si nombra* 'if s/he names'	*sin hombres* 'without men'
la santa 'the saint'	*las antas* 'the elk'	*si naces* 'if you are born'	*sin asas* 'without handles'
la cera 'the wax'	*las eras* 'the eras'	*si notas* 'if you notice'	*sin otros* 'without others'
la sortija 'the ring'	*las hormigas* 'the ants'	*si nacieras* 'if you were born'	*sin acero* 'without steel'
la sorpresa 'the surprise'	*las orejas* 'the ears'	*si nombraras* 'if you nominated'	*sin hombreras* 'without pads'
la sardina 'the sardine'	*las ardillas* 'the squirrels'	*si notaras* 'if you noticed'	*sin otoño* 'without autumn'

Materials

A series of two-word phrases containing V#CV or VC#V sequences were embedded in meaningful sentences (for example, for *las ardillas*: *las ardillas entraron en la cocina y comieron todos los nueces* 'the squirrels entered the kitchen and ate all of the walnuts'. The pivotal consonant was always /n/, /s/, or /r/ Although these two-word phrases were not minimal pairs, great care was taken to make sure the vowels surrounding the pivotal consonant were the same and lexical stress on each of the two words were identical in both the V#CV and VC#V conditions. For /s/, the two-word phrase was always an article-noun combination, either singular or plural, e.g. *la sala* 'the room' / *las alas* 'the wings'. These were often adapted from Karn (1996). For /n/, the first word was either *si* 'if' or *sin* 'without', followed by a vowel- or consonant-initial word as necessary. For /r/, the first word was always a verb in the infinitive (/r/ final) or in the past or present tense (vowel-final). A complete list of items is shown in Table 2.1. Additionally, the native English speakers read eight additional sentences containing minimal pairs demonstrating the word-internal contrast between /ɾ/ and /r/: *cero* 'zero' vs. *cerro* 'hill', and *ahora* 'now' vs. *ahorra* 'he saves'.

For Experiment 1b, in English, pairs of first and last names were used to elicit cross-word consonant resyllabification. The 26 names, taken from (Ladd and Schepman, 2003), were embedded in the first clause of a carrier phrase, for example: "It was Norman Elson, not Nora Elton." For 13 of the names, the first name ended in a vowel, while the last name began with a consonant (usually /n/). For the other 13, the first name ended in a consonant, and the last name began with a vowel: the pair formed a quasi-minimal pair, such as *Norma*

Table 2.2: List of stimuli for Experiment 1b (English)

V#CV	VC#V
Brenda Nevin	*Brendan Evans*
Kay Nevin	*Caine Evans*
Cara Norris	*Karen Orris*
Dee Norman	*Dean Ormond*
Ella Norwell	*Ellen Orwell*
Jay Neeson	*Jane Eason*
Jo Norval	*Joan Orville*
Joe Nellis	*Joan Ellis*
Kay Lashley	*Gail Ashley*
Laura Norton	*Lauren Orton*
Norma Nelson	*Norman Elson*
Ray Nutley	*Wayne Utley*
Shea Norton	*Shane Orton*

Nelson / Norman Elson. A complete list of the pairs of names is found in Table 2.2.

Procedure

Participants were seated in a sound-attenuated booth and performed the reading task at their own pace. They were presented one at a time with a meaningful sentence, as described in the materials section, on the screen, in a random order. They were told to silently read the sentence first to themselves, and then read it out loud. If they felt they had made an error in pronunciation, they could repeat the sentence starting from the beginning. The participant pressed a key on the keyboard to move onto the next sentence. Native English speakers recorded a total of 76 sentences in Spanish, and 26 sentences in English. Native Spanish speakers recorded a total of 68 sentences in Spanish.

Analysis

Each of the tokens were hand annotated using a custom annotation script in Praat. The target consonant and its flanking vowels were annotated for duration, and separately each token was classifying on different measures.
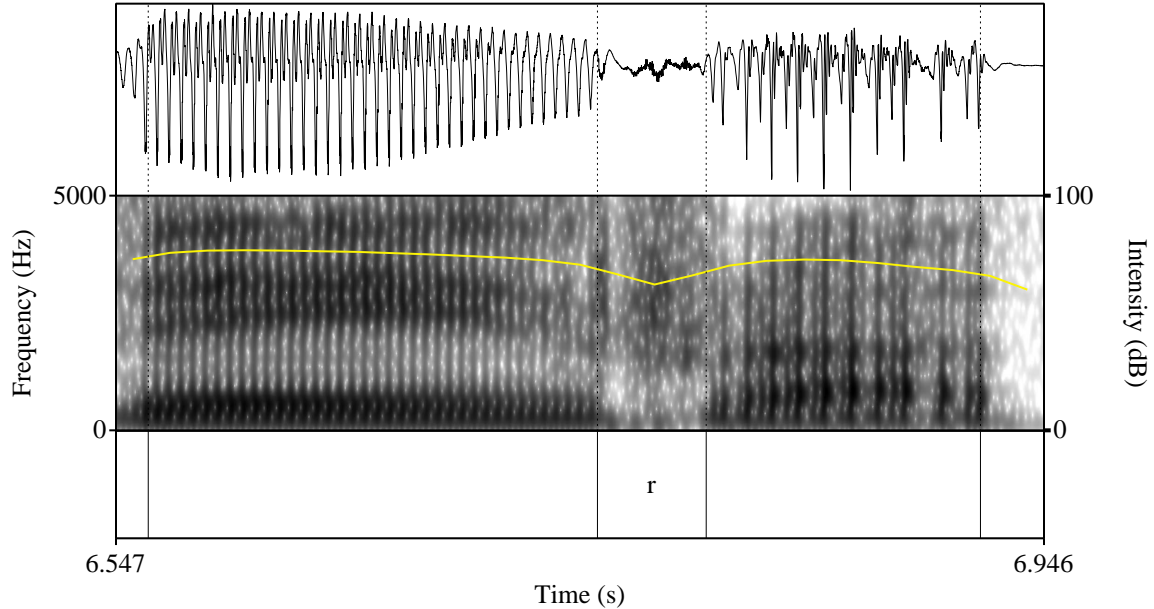
Figure 2.4: Example of a word-initial [r] produced by a second-language learner of Spanish. The trill is not the canonical variant with multiple occlusions, but is an acceptable 'assibilated' /r/. The word shown is *ratas* 'rats'.

1. For the analysis of the one-word stimuli, intended to measure the produced difference (or lack thereof) between word internal /ɾ/ and /r/, tokens were classified simply as 'tap', 'trill', or 'approximant', based on target-like acoustic realizations. To be classified as a 'tap', a single occlusion must be visibile in the spectrogram or waveform, with a drop in amplitude. The occlusion may not be complete, resulting in a more approximant-like production, but a short drop in amplitude must be present. For 'trill', as there are many variants of the phoneme /r/ across the Spanish-speaking world, a number of different possibilities were allowed. Most common were the canonical trill production with multiple occlusions or an assibiliated variant with evident frication. An example of a assibilated token, marked as 'trill' is shown in Figure 2.4. The classification of 'approximant' was saved for obviously English-like [ɹ] productions, with a lowering of F3 and little attenuation of amplitude of the waveform.

2. For the analysis of all of the two-word stimuli, any silence (with no voiced pulses or leakage) longer than 150 ms between the two words was considered to be a pause: the

25

token was disregarded from further analysis.

3. For the VC#V stimuli, the presence or absence of glottalization was marked by hand. This could include a full glottal stop in between the word-final consonant and the word-initial vowel, or simply creaky voice marked at the onset of the vowel, or any other variant. This was determined by carefully examining the waveform and spectrogram for irregular voicing periods. Two examples follow in Figures 2.5 and 2.6.

4. For the two-word /r/ stimuli (e.g. *ve rocas* vs *ver osos*, both the type of rhotic and the presence of glottalization were tagged, in the same way as previously mentioned.

Using a custom-written script in Praat, each stimulus was analyzed for voice quality to measure the degree of glottalization. This was done in the following way: the pitch contour was measured for the entire utterance (for females, a minimum pitch of 50 Hz and a maximum pitch of 400 Hz was set, for males, 50 Hz and 300 Hz, respectively). Then, data from Praat's Voice Report between the beginning and end of the second vowel were extracted (any token less than 50ms caused errors in the voice analysis (see the Praat manual for more on this issue) and were excluded) as follows:

1. *Degree of voice breaks* A percentage, indicated the total duration of the breaks between the voiced parts of the signal, divided by the duration of the entire vowel. A voice break is defined as the number of distances between consecutive pulses that are longer than 1.25 divided by the pitch floor. Normal modal voicing should have no voice breaks, although this measure is highly dependent on accurate segmentation (especially when the flanking context consists of voiceless sounds), and the duration of the vowel token. Creaky voice may not have any voice breaks either; this measure will only pick up cessation of vocal-fold vibration, which will indicate a full glottal stop.

2. *Fraction of locally unvoiced pitch frames* This measure tracks the number of unvoiced frames throughout the vowel, divided by the number of pitch frames. In practice, this seems like a similar measure of the degree of voice breaks, but it is less conservative: the beginning of an annotated vowel could have no voiced frames (indicating a longer glottal stop), but begin voicing somewhat later. In this case, there are unvoiced pitch

26

frames, but no break in voicing. Thus, the two measures capture somewhat different patterns of glottalization: in Figure 2.5, voicing begins after a short glottal stop. This token has a higher fraction of locally unvoiced pitch frames, but a degree of voice breaks of 0. In Figure 2.6, voicing has continued briefly after the release of the /n/, ceases, and then begins again to produce the full vowel. This token has both a break in voicing and a number of unvoiced frames.

3. *Jitter* The average absolute difference between consecutive pitch periods, divided by the average period. A higher jitter value indicates more glottalization.

4. *Harmonics-to-Noise Ratio (HNR)* A signal-to-noise ratio expressed in dB. Normal modal voicing for /a/ is about 20 dB, but this number depends on the vowel. Lower values indicate more noise, or higher glottalization. The vowels studied here were spread across the vowel space, so some overlap between glottalized and non-glottalized vowels is expected, but because the same vowels occurred in both the V#CV and VC#V contexts, overall, glottalized vowels should have a lower HNR. HNR can be measured between different frequencies: Garellek (2013) used HNR for frequencies of less than 500 Hz to focus on noise related to pulse-to-pulse irregularity, but this functionality is not yet available in Praat. Note also that HNR can only be measured over the voiced parts of a vowel, and as such, can only tell us about glottalization if there is any aperiodicity during the vowel itself. A short glottal stop prior to the vowel will not give a different HNR measure from that of a vowel not preceded by a glottal stop.

The voicing analysis is subject to scrutiny: because the duration of the word-initial vowel are by and large longer due to the inclusion of possible glottal stops in the analysis, they cannot be compared perfectly to the shorter vowels of the V#CV context. Moreover, the degree of voicing or lack thereof rests on accurate segmentation of the surrounding context, especially when flanked by /s/ or any type of /r/. Finally, some of the measures provided cannot always reflect the 'amount' of glottalization present on a vowel, but rather *how much* of the produced vowel was glottalized, which can be quite variable especially with L2 speech, which can be slower. Nevertheless, the data provided here is an attempt to objectively measure the presence of glottalization on vowels, in addition to the impressionistic data also
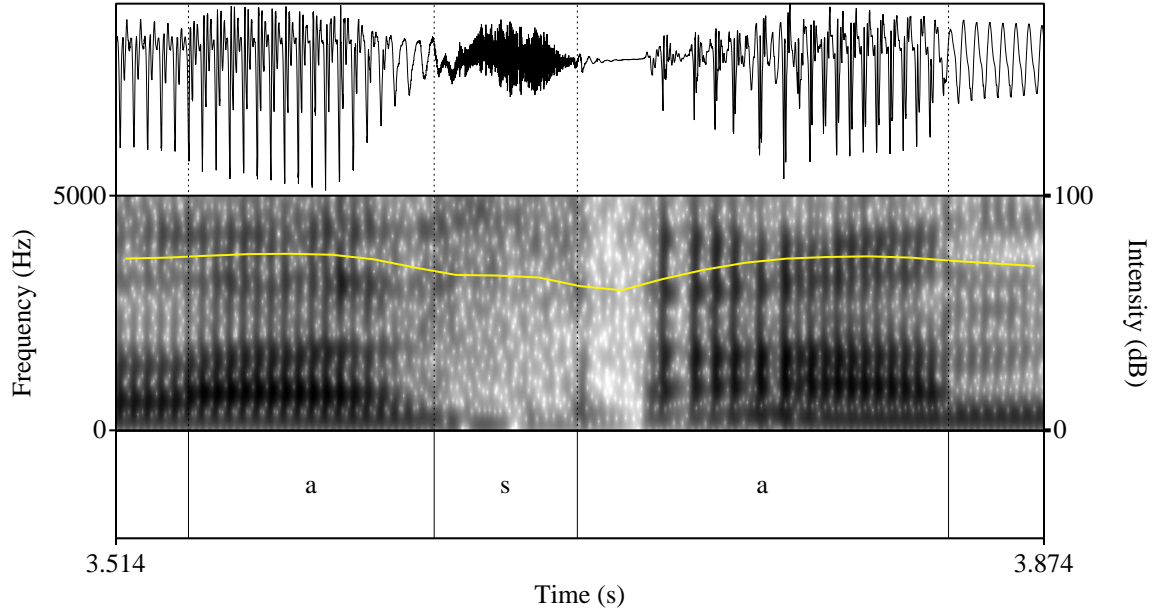
Figure 2.5: Example of a word-final /s/ that has not been resyllabified. The creaky voice on the word-initial /a/ is evident in both the waveform and spectrogram through the irregular pitch periods. This is a typical realization of a VC#V token produced by a second-language learner of Spanish. The words shown are *las antas* 'the elk'.

collected. Finally, note that in American English, and perhaps especially among younger speakers, entire words and phrases can be glottalized when they appear towards the ends of utterances. This is also possible towards the end of the recording session, which lasted around 30 minutes, or early in the morning or late in the afternoon. Thus, there may be several 'false positives': vowels in the V#CV context that were surprisingly glottalized, but may be due to glottalization over a larger domain.

Results

Spanish. The results from both native and non-native speakers are provided in the following tables and figures. First, to reiterate research questions (1) and (3), it was hypothesized that L2 speakers transfer their glottalization patterns into Spanish, a language in which it is not a common juncture phenomenon. It is clear that the L2 speakers of Spanish glottalize considerably more than the native speakers (overall, 62% vs. 6%), where glottalization is rare but possible. Looking at Figure 2.7, not every speaker shows this pattern. Some speakers
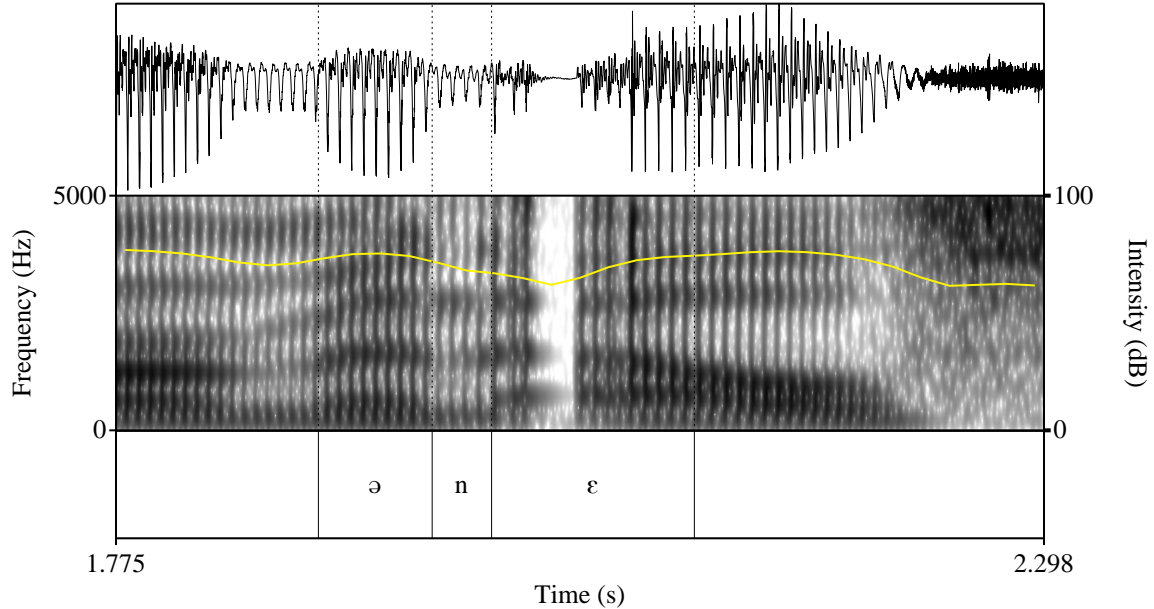
Figure 2.6: Example of a word-final /n/ that has not been resyllabified. A glottal stop appears before the vowel-initial word 'Elson'. This is a typical realization of a VC#V token in English. The words produced are *Norman Elson.*

have resyllabified nearly all consonants in this context, while others have resyllabified none of them.

With regard to the production of rhotics in Spanish, research question (4) hypothesized that English speakers will have more difficulty correctly making the word-boundary contrast between the tap and the trill as compared to the word-internal contrast. The results should be discussed for both the word-internal position and word-boundary positions separately. First, tokens of word-internal /ɾ/ and /r/ were not elicited from native speakers as to have a comparison, but since this is a strong lexical phonemic contrast, I will just compare the L2 speakers results with the literature on how these sounds are canonically produced **??**. L2 speakers produced these somewhat accurately: they were most accurate with producing the tap in words like *cero*, but were also likely to produce the tap in words which require the trill, like *cerro*. However, notably, they did not produce the trill in any tap words. English-like approximant realizations were also found.

For word-boundary rhotic productions, native speakers performed as expected. They maintained the contrast between taps and trills in this unusual context: word-initial /r/ was

29

Table 2.3: Spanish data: percentage of glottalization in VC#V tokens, native speakers

/n/	4.16%
/s/	4.16%
/r/	11.47%

Table 2.4: Spanish data: percentage of glottalization in VC#V tokens, L2 speakers

/n/	53.65%
/s/	56.62%
/r/	77.88%

produced as a trill 97% of the time, while word-final /r/ was resyllabified and produced as a tap 100% of the time. For L2 speakers, there was considerable variation. Of particular note, L2 speakers produced approximants in both word-initial and word-final contexts far more than either of the word-internal contexts. Like the word-internal condition, no trills were found where taps would be expected. However, trills were not produced word-initially as often as they were in the word-internal context (25% vs. 42%).

For the continuous measures of glottalization, these results largely match the impressionistic findings. For native speakers, the few cases of glottalization in the VC#V context in all four measures: a higher jitter, a lower HNR, more unvoiced frames, and a larger degree of voice breaks were found. For L2 speakers, which exhibited a much higher rate of glottalization, all four measures matched well with the impressionistic labelling, but to different extents. A generalized linear model was fit to see which measure best predicted the glottalization annotation (glottalized / non-glottalized). Using HNR alone accurately predicted
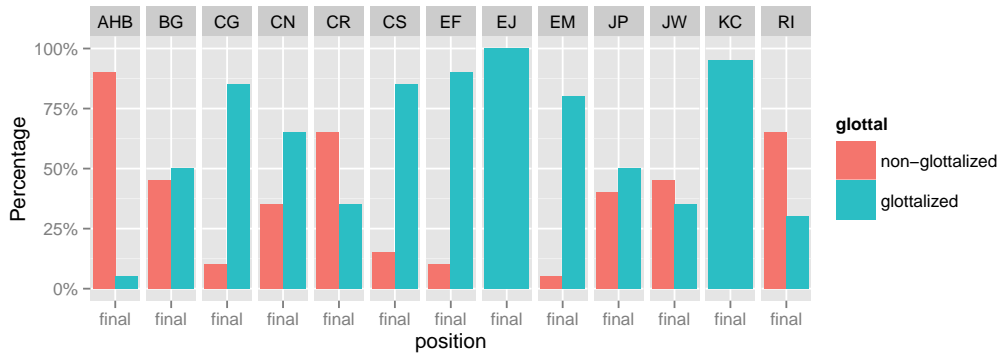
Figure 2.7: Percentage of glottalized tokens in VC#V contexts by each L2 speaker.
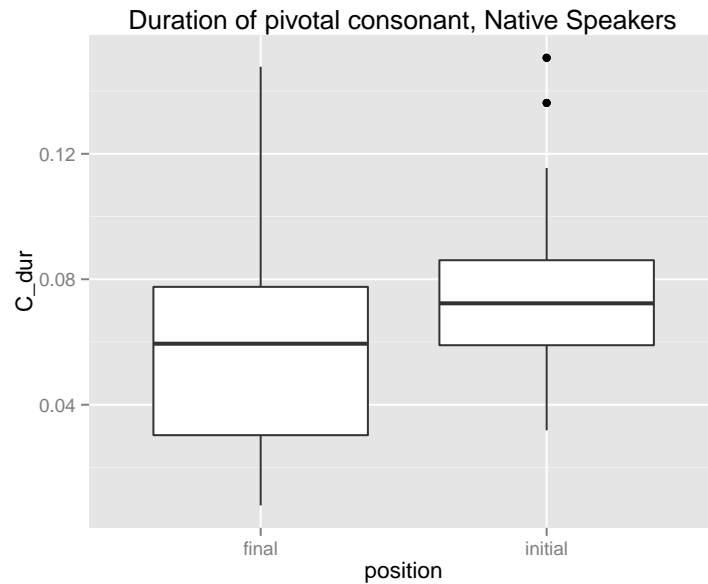
Figure 2.8: Duration (in ms) of the consonant in either the V#CV (initial) or VC#V (final) contexts. Successfully resyllabified consonants (in final position) are expected to be of similar duration to those of initial consonants.
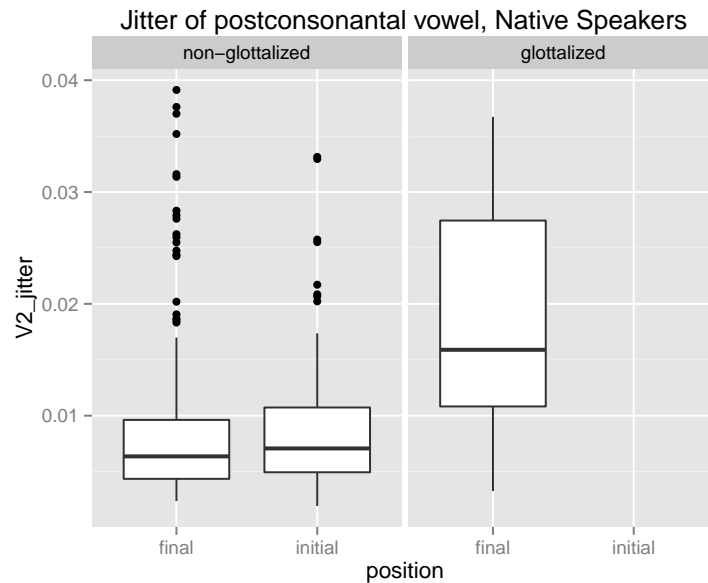
Figure 2.9: Jitter of the second vowel in either V#CV (initial) or VC#V (final) contexts. A higher jitter value indicates more glottalization.
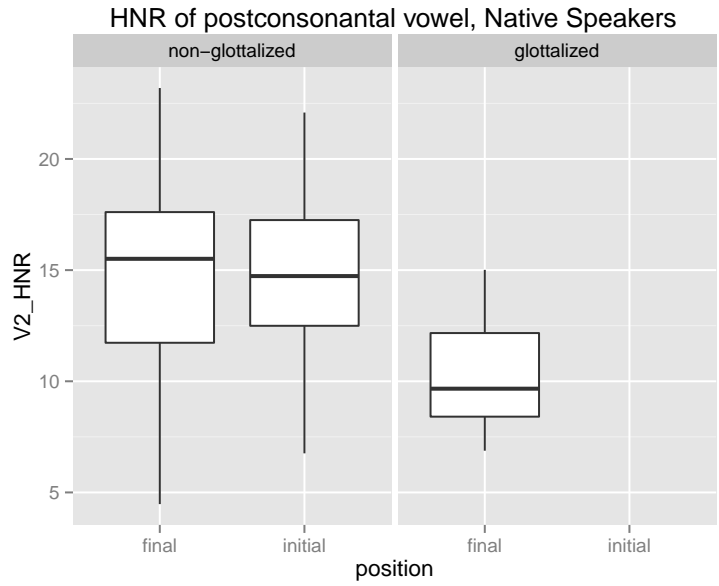
Figure 2.10: Jitter of the second vowel in either V#CV (initial) or VC#V (final) contexts. A higher jitter value indicates more glottalization.
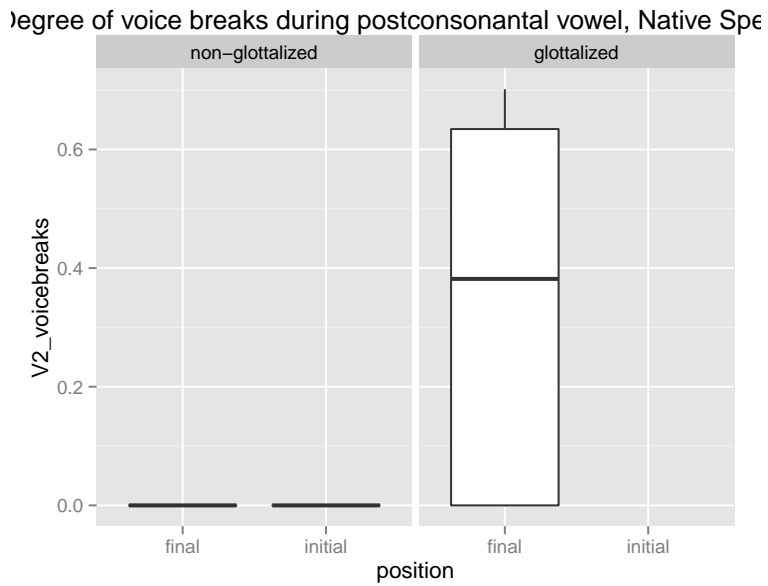
Figure 2.11: Jitter of the second vowel in either V#CV (initial) or VC#V (final) contexts. A higher jitter value indicates more glottalization.

Table 2.5: Classification of word-internal rhotic productions, L2 speakers

	approximant	tap	trill
r 'cero'	19.23%	80.76%	0%
rr 'cerro'	15.38%	42.30%	42.30%

Table 2.6: Classification of word-boundary rhotic productions, L2 speakers

	approximant	tap	trill
final 'ver osos'	41.34%	58.65%	0%
initial 've rocas'	28.43%	47.05%	24.51%

76.8% of the data, using jitter predicted 74.9% of the data, using degree of voice breaks also accurately predicted 83.77% of the data, while using percentage of unvoiced frames predicted 91.99% of the data accurately. Adding HNR and jitter to the unvoiced frames model improved the accuracy to 93.4%.

English It was hypothesized in research question (2) that English speakers prefer to mark VC#V sequences with glottalization. This was confirmed: both the impressionistic and voice data are provided for the utterances in English ('It was Norman Elson...'), showing that the speakers' rate of glottalization was much higher than it was for similar Spanish materials (92% vs. 62%). A few stray tokens in initial position were glottalized as well.

The measure of voice quality largely agree with the impressionisitc labelling. A higher jitter and HNR were found for glottalized tokens.

Discussion

The results from these read speech experiments largely confirm the hypotheses: Spanish speakers do not glottalize word-initial vowels followed by a consonant, although it is an available pronunciation. For Spanish speakers, resyllabification is the norm. To answer the second part of research question (1), which questioned whether or not resyllabification is complete in Spanish, the results here show that it is not strictly complete, as word-initial consonants are longer than word-final (resyllabified) consonants, although there is much overlap.

Regarding research question (L2 speakers of Spanish consistently produce glottalization in the VC#V context, as predicted. For these speakers, resyllabification has been blocked
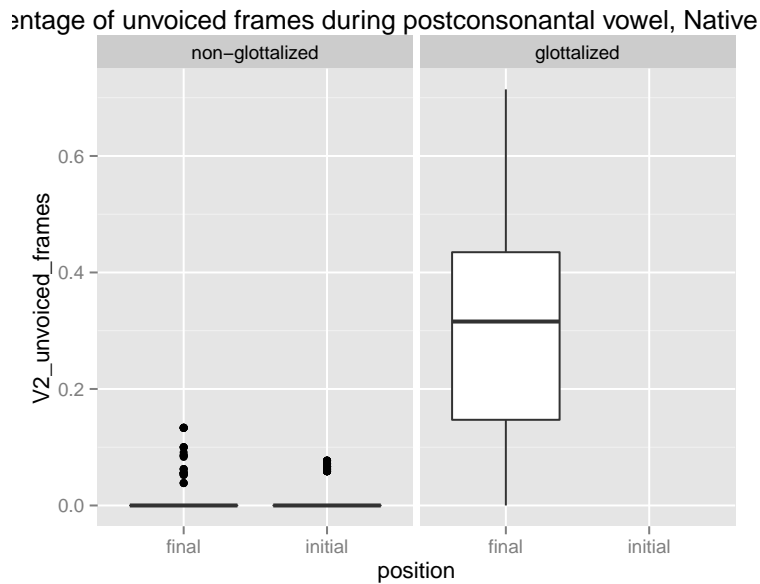
Figure 2.12: Percentage of unvoiced frames during the second vowel in either V#CV (initial) or VC#V (final) contexts. A higher jitter value indicates more glottalization.
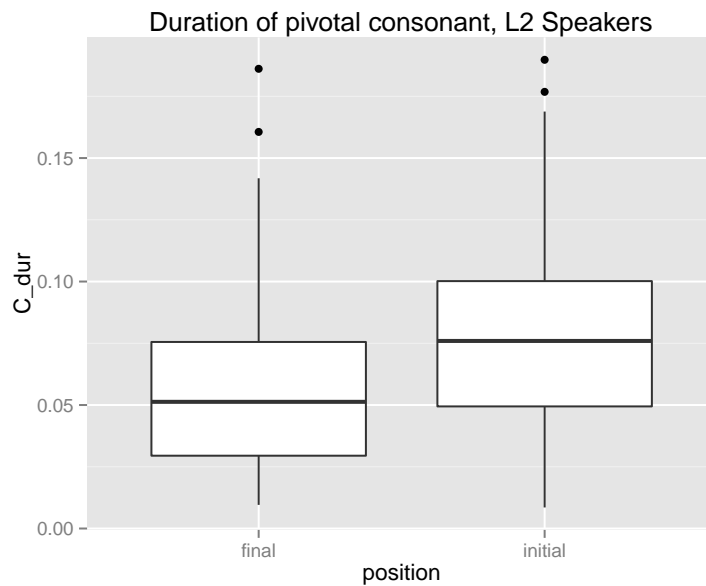
Figure 2.13: Duration (in ms) of the consonant in either the V#CV (initial) or VC#V (final) contexts. Successfully resyllabified consonants (in final position) are expected to be of similar duration to those of initial consonants.
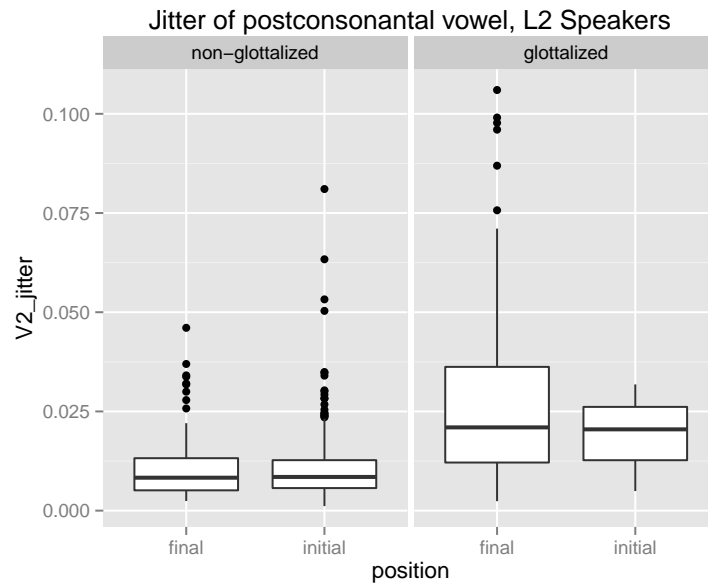
Figure 2.14: Jitter of the second vowel in either V#CV (initial) or VC#V (final) contexts. A higher jitter value indicates more glottalization.
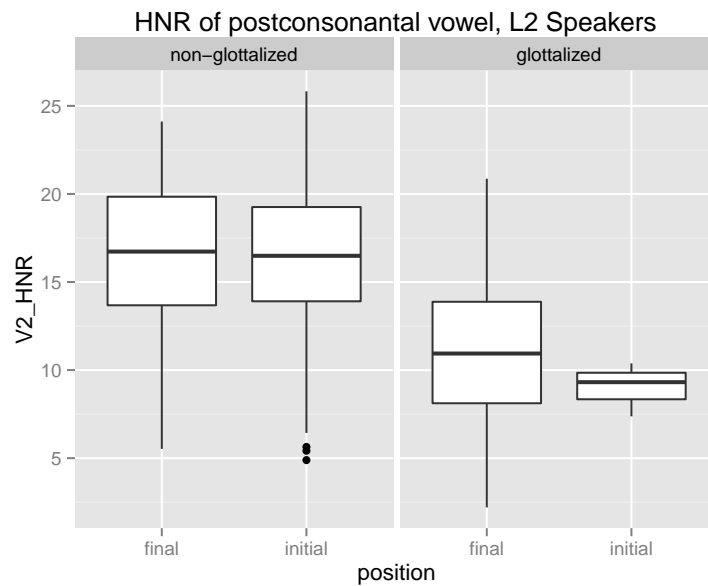
Figure 2.15: Jitter of the second vowel in either V#CV (initial) or VC#V (final) contexts. A higher jitter value indicates more glottalization.
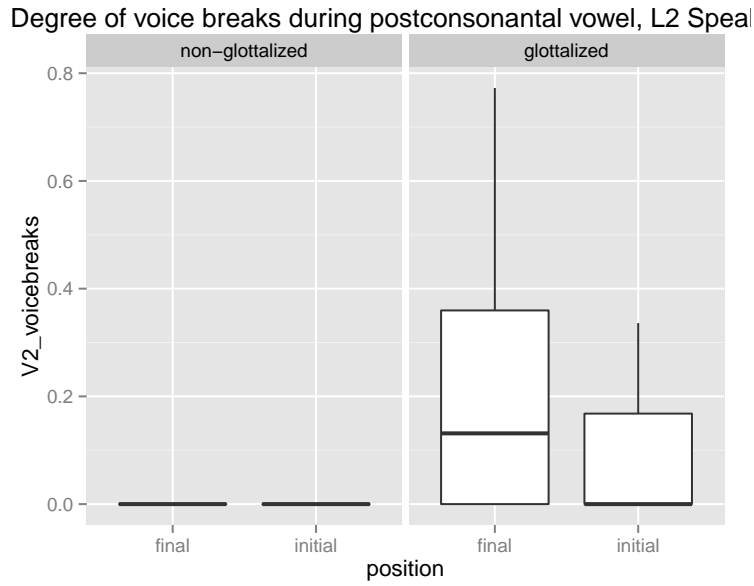
Figure 2.16: Degree of voice breaks during the second vowel in either V#CV (initial) or VC#V (final) contexts.
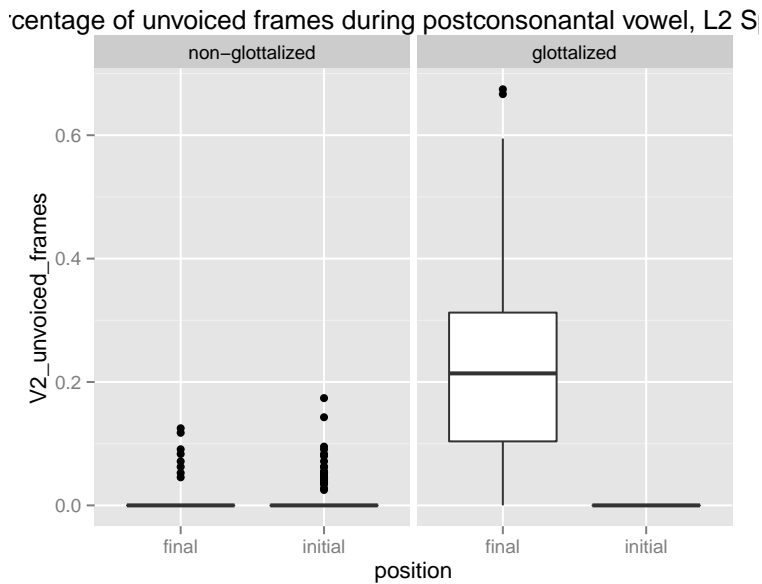
Figure 2.17: Percentage of unvoiced frames during the second vowel in either V#CV (initial) or VC#V (final) contexts.
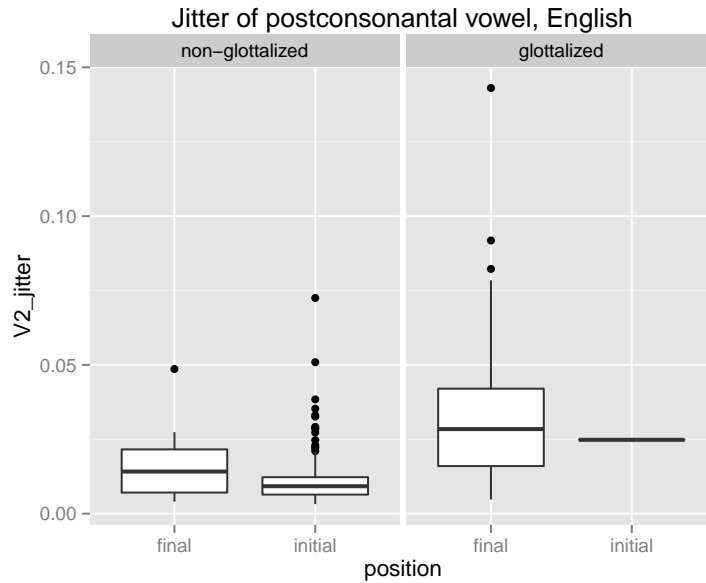
Figure 2.18: Jitter of the second vowel in either V#CV (initial) or VC#V (final) contexts. A higher jitter value indicates more glottalization.
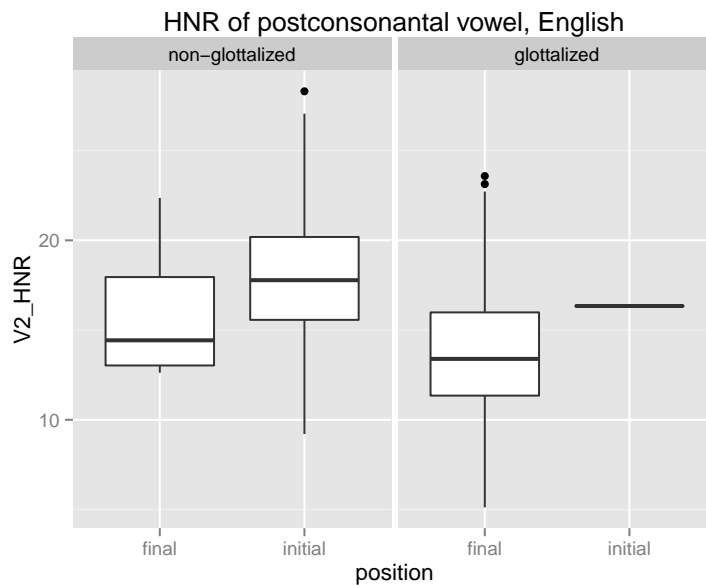
Figure 2.19: Jitter of the second vowel in either V#CV (initial) or VC#V (final) contexts. A higher jitter value indicates more glottalization.

Table 2.7: Classification of word-boundary rhotic productions, native speakers

	approximant	tap	trill
final 'ver osos'	0%	100%	0%
initial 've rocas'	0%	3.12%	96.87%

Table 2.8: English data: percentage of glottalization in VC#V tokens, native English speakers

glottalized	92.48%
non-glottalized	7.51%

by glottalization. The strategies that L2 speakers used to mark glottalization ranged from a full glottal stop to simple creaky voice on the word-initial vowel.

Regarding the analysis based on voice quality, none of the measures clearly corresponded to the types of glottalization described in Dilley et al. (1996). Both HNR and jitter correspond well to aperiodicity, but a full glottal stop was best measured by the percentage of unvoiced frames during the annotated vowel. Through careful examination of the tokens, a number of interesting patterns emerged, perhaps unique to this type of data. This was most clear in the V[ɾ]#V context produced by L2 speakers. Regardless of the type of rhotic produced, this context was often glottalized. When the tap was produced correctly, a short vowel necessarily followed the short occlusion. This was then followed by some glottalization and perhaps a full stop before the vocal folds began vibrating regularly again. An example is provided in Figure 2.20. This timing pattern shows that the speaker clearly marks the beginning of a vowel-initial word with glottalization, even if it means interrupting the vowel gesture.

The other striking pattern in the same V[ɾ]#V context was that this elicited a high rate of English-like approximants, even if the same speaker has been able to produce the tap in the word-internal context. Most often these approximants were followed by some degree of glottalization. There may be two reasons for this: (1) the sequence of tap followed by a glottal

Table 2.9: Mean values of voice quality measures in English by native English speakers

glottal	position	HNR	jitter	voicebreaks	unvoiced frames	C duration (ms)
non-glottalized	final	15.93	0.016	0	0.0056	0.0522
non-glottalized	initial	17.98	0.011	0	0.0019	0.0879
glottalized	final	13.73	0.032	0.229	0.2252	0.0533
glottalized	initial	16.34	0.024	0	0.4655	0.0487
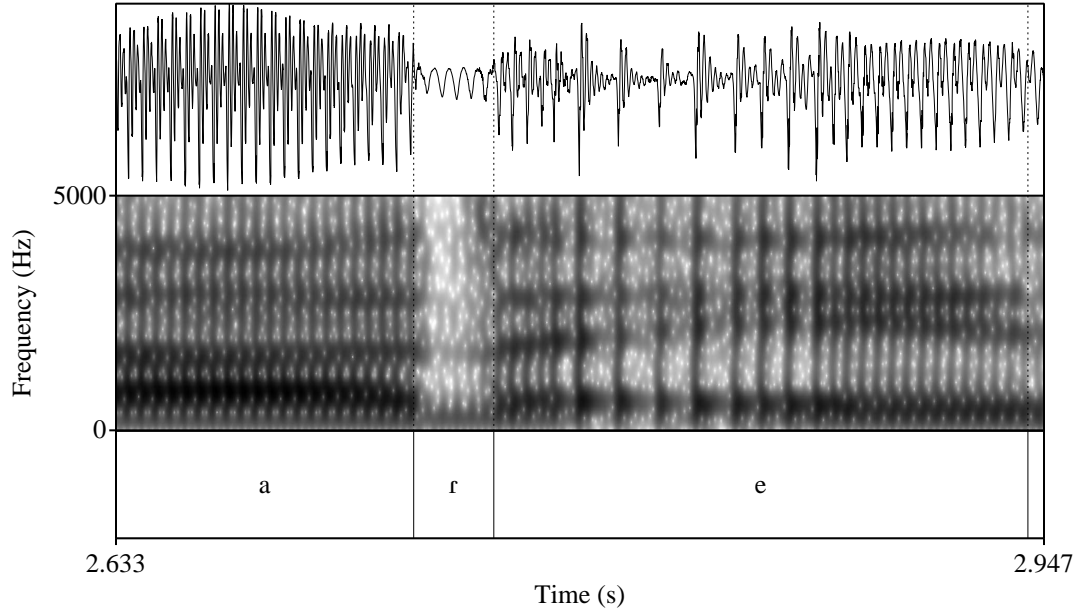
Figure 2.20: Example of a word-final [ɾ] that has not been resyllabified, but is produced correctly as a tap. A glottal stop appears before the vowel-initial word *estas* 'these'. This is produced by an L2 speaker of Spanish.

stop may be difficult to produce (hence the need for a longer vocalic element following the tap), so the speaker may revert to a more English-like production or (2) the speaker always produces an approximant in Spanish for the <r> that appears in coda position. Although there has been much written about the tap-trill contrast within words between vowels, there has been little if anything written about the production of the rhotics in other positions. The intent for this study was to look at word-initial <r>, but it is likely that coda <r> is often produced as an approximant due to its less salient position.

To return to research question (2), the amount of glottalization found in English, the hypothesize that American English speakers prefer to mark vowel-initial words with glottalization was overwhelmingly confirmed. Although the materials and experiment were identical to that of Ladd and Schepman (2003), the results from these speakers of American English do not agree regarding how often English speakers glottalize vowel-initial words. While that study found only around a third of its participants' tokens to be glottalized (and subsequently thrown out), nearly all of the produced tokens here showed some evidence of glottalization. The results are more in line with Umeda (1978), which also analyzed speakers of American

English, and found considerable variation among speakers. This study shows that glottalization to make word-initial vowels may be a difference between British and American English.

Comparing English with Spanish, research question (3), speakers tended to glottalize far more in English than in Spanish. Because the same speakers produced the materials in both languages, it was hypothesized that the reason some speakers glottalized less (and resyllabified more) in Spanish was that this was a more available pronunciation in their variety of English. This was not necessarily the case: although the two participants who resyllabified the most in English (25%) also had high rates of resyllabification in Spanish (40 - 50%), other participants who resyllabified around the same rate in Spanish exhibited little to no resyllabification in English.

Although the results for the word-internal rhotics were mixed (most tokens were produced as taps in both tap and trill words), because research question (4) has to do with the maintenance of a phonological contrast, we must also look at the distribution of taps, trills, and approximants for each speaker. A speaker may incorrectly encode the contrast between *cero* and *cerro* as between an approximant and a tap, but a speaker who produces a tap for both words has encoded these two words as homophones. In fact, for speakers who have difficulty producing the trill may be able to maintain the phonological distinctness of these two words by using a non-Spanish speech sound (the English-like approximant) in only one of these words.

Figure 2.21 examines this very question. Some speakers cannot produce any trills, and substitute taps. Other do not produce any Spanish sounds. However, a few speakers are in transition. Speaker CN, for example, only produces taps in tap words, but when the speaker cannot produce a trill, produces an approximant instead. The results here are in agreement with Rose (2010). There is no evidence of productions of trill where we would expect taps: the 'fuzzy lexical encoding' in Daidone and Darcy (2014) where speakers accept words with the trill where the word requires a tap does not seem to apply to production.

For the word-boundary rhotics, the hypothesis in research question (4) was that L2 speakers would have the most trouble in this condition due to lack of awareness. This is only partially confirmed. The speakers did have trouble, but they were not entirely inconsistent,

40
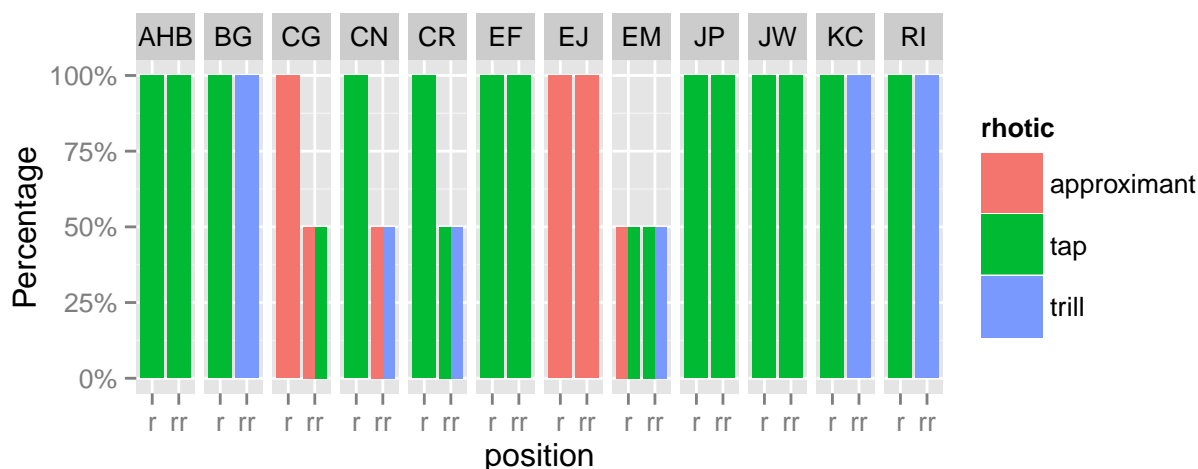
Figure 2.21: Percentage of each rhotic type produced in tap words, *cero*, 'r', and trill words, *cerro*, 'rr', divided by speaker.

nor did it completely deviate from native speaker norms. The trills produced were limited to word-initial contexts, which is consistent with the native speaker data. However, most of the productions in the word-initial context were taps. This may be due to the difficulty in producing this segment, however, L2 speakers produced trills in word-initial position around half as often as they did in word-internal position. It is possible that it is harder to produce the trill in word-initial position, but a more likely scenario is that L2 speakers simply have not been taught that word-initial <r> is produced as a trill. Recall that the distinction between the tap and the trill is made in orthography in the word-internal position (*caro* vs *carro*), but not at the word boundary (*ver ocas* vs *ve rocas*). Plausibly, L2 speakers substitute the simpler sound if they have not encoded these words as obligatorily having a trill.

In word-final position, as mentioned above, many more approximants than expected were found. It is likely that this is due to the presence of glottalization. If the L2 speakers produce glottalization, the rhotic segment is necessarily a coda of the final syllable of the word. It appears that a coda /r/ is produced often as an approximant for many of these speakers, and surfaces as such in these words preceding glottalization. Other speakers, however, are able to produce a tap, but still employ glottalization. Finally, many L2 speakers are able to produce native speaker-like taps without any glottalization, indicating resyllabification.

41

CHAPTER 3

PERCEPTION

3.1   Overview

3.2   Experiment 1

These experiments seek to answer the following research questions:

1. Are L2 learners able to learn a new parsing strategy for words? Does disambiguation occur only under acoustic conditions that match the L1 strategy? Native English speakers are not familiar with resyllabification processes in their native language, and are most likely to only parse words as vowel-initial if there is a pause or glottalization. Because Spanish does have resyllabification, learners may have a bias towards hearing these resyllabified consonants as word-initial consonants, as they would be in English.

2. Are L2 learners able to do so with acoustic cues from the target language? Despite resyllabification, there may still be some acoustic cues as to the lexical affiliation of the pivotal consonant which native speakers may be able to use.

3. Are some stimuli or different phonemes easier than others? Some words may be more familiar or frequent in the lexicon to L2 speakers, and they may use their knowledge of words in Spanish to drive their parsing mechanism. Also, there may be more consistent acoustic cues with some phonemes than others.

4. Do L2 speakers perform better / differently with a higher proficiency level? Beginning learners may perceive consonants as being word-initial due to their English parsing mechanism, but they may be able to suppress this system with more exposure and

proficiency in their second language. To investigate this, two groups of L2 speakers will be compared: a beginner group and an advanced group. Their individual proficiency scores will also be considered as a predictor of performance.

3.2.1 Methodology

Participants

Twenty-seven total participants spent roughly one hour to complete all of the experimental tasks (thirty minutes for native Spanish speakers). A control group was formed of nine native Spanish speakers ('NS') from various non /s/-aspirating dialects (as explained above). All native speakers had learned Spanish from birth, although three speakers were bilingual: two speakers also spoke Basque, one had moved to Galicia at a young age. The English native speakers were assessed in their Spanish proficiency through a written test, consisting of the vocabulary and cloze parts of the DELE (Diploma de Español como Lengua Extranjera) test, widely used in other studies of L2 learners of Spanish **??**. The results are reported in Table 3.1; learners with scores of 40-50 are considered to be advanced; 30-39, intermediate; 0-29, beginner. Nine native English speakers with high proficiency in Spanish (all were graduate students in the Department of Linguistics or the Department of Spanish, Italian, and Portuguese at the University of Illinois, Urbana-Champaign; eight had study-abroad experience) formed the Advanced L2 group, henceforth 'L2A'. All scored in either the intermediate or advanced ranges on the proficiency test. Nine native English speakers with low proficiency in Spanish were recruited from two Spanish classes: SPAN 141, a third-semester Spanish course focusing on grammar, and SPAN 228, a fourth-semester Spanish course focusing on writing. None had studied any other languages besides English. These speakers are considered to be beginning learners, hence forth 'L2B'. Seven out of the nine subjects scored in the beginner range on the proficiency test; the remaining two scored in the intermediate category.

43

Table 3.1: Biographical Information of Participants in the Experiments

	Mean	Minimum	Maximum	SD
Native Speakers (n = 4)				
Age	33.33	24	44	7.13
L2A (n = 9)				
Age	24.55	22	30	2.41
Age at beginning of Spanish study	11.55	6	16	3.27
Number of years of Spanish study	8.55	6	12	2.11
Number of years of Immersion	1.54	0	7	2.00
Proficiency Score (out of 50)	43.77	32	48	5.22
L2B (n = 9)				
Age	18.67	18	20	0.67
Age at beginning of Spanish study	12.33	6	16	2.54
Number of years of Spanish study	5.39	3.5	7	1.05
Number of years of Immersion	0	0	0	0
Proficiency Score (out of 50)	25.11	17	37	6.56

Materials

The minimal pairs for the stimuli are presented in Table 3.2 Minimal pairs were found using a combination of dictionaries; each pair differs only in the presence of a word-initial consonant, e.g. *eco* 'echo' (vowel initial) and *seco* 'dry' (consonant initial). The two target consonants in this study were /n/ and /s/, chosen because they are common verbal endings in Spanish and because they show no allophonic cues to onset/coda affiliation . They may, however, show differences in duration, and native Spanish speakers are able to attend to duration in both /n/ and /s/, although to different degrees of accuracy (Scarpace and Hualde, 2013). Furthermore, both /n/ and /s/ are phonotactically legal word-final codas and word-initial onsets. In Spanish, orthographic <h> is a silent letter, e.g. *hace* [áse]. Because words were to be presented visually, the presence of a silent letter is a possible confound; thus, the vowel-initial word in half of the 36 test pairs began with an orthographic <h>, while the other half did not. The target words were embedded in a phrase: *Escribe(n)(s)* X '/he/she/you (formal)/you (s)/they (n)/you (plural) (n)', depending on the target word. For example, if the target pair was nace / hace, the two phrases would be *escribe nace* and *escriben hace*; the two phrases being minimal pairs differentiated by the word-affiliation of the target consonant

Table 3.2: List of stimuli used in Experiments 2a and 2b

/n/			/s/	
/h/			**/h/**	
	C	V	C	V
	nace	hace	saga	haga
	nada 'nothing'	*hada* 'fairy'	*sé* 'I know'	*he* 'I have (AUX)'
	nado	hado	seno	heno
	nato	hato	soy	hoy
	Nilo	hilo	sube	hube
	nombre	hombre	sumo	humo
	nombro	hombro	suyo	huyo
	nube	hube	sorda	horda
	Nora	hora	soja	hoja
no /h/			**no /h/**	
	C	V	C	V
	nalga	alga	sala	ala
	narco	arco	salgo	algo
	nasa	asa	salta	alta
	nata	ata	sano	ano
	nave	ave	seco	eco
	nene	ene	sellos	ellos
	neuro	euro	sido	ido
	nido	ido	sodio	odio
	no	o	sola	ola

45
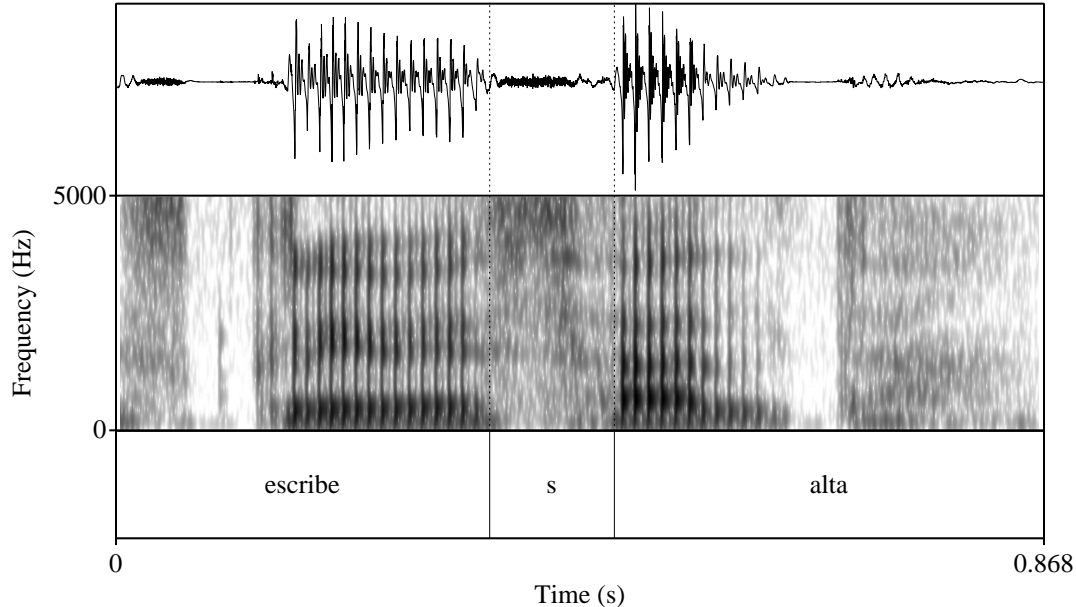
Figure 3.1: *escribe salta* 's/he writes *jump*'. Duration of /s/ is 116 ms.

/n/.

The stimuli for the experiments were recorded in a sound-attenuated booth in the Phonetics & Phonology Lab at the University of Illinois, Urbana-Champaign by a male native speaker of Colombian Spanish . The speaker was instructed to produce the forms in a casual register of speech, without introducing a pause between the two words. Additionally, the speaker was told to read the list again in a careful register of speech. All of these forms were analyzed after recording to ensure that a glottal stop or evidence of glottalization intervened between the two forms. Thus, each pair from the list in Table 3.2 formed part of a triplet: the phrase with a consonant initial word, the phrase with a vowel initial word, and the phrase with a vowel initial word with a glottal stop introduced between the two words, as in Figure 3.3. For participants of both language backgrounds, a glottal stop should be a positive signal of a word boundary to its left. Its presence in this experiment was also to ensure that native English speakers are aware that vowel-initial words are possible responses during the experiments, since it is hypothesized that these subjects will perceive resyllabified consonants in both (a) and (b) as belonging to consonant-initial words.

After recording, duration of /n/ and /s/ in the word-initial and resyllabified contexts were

46
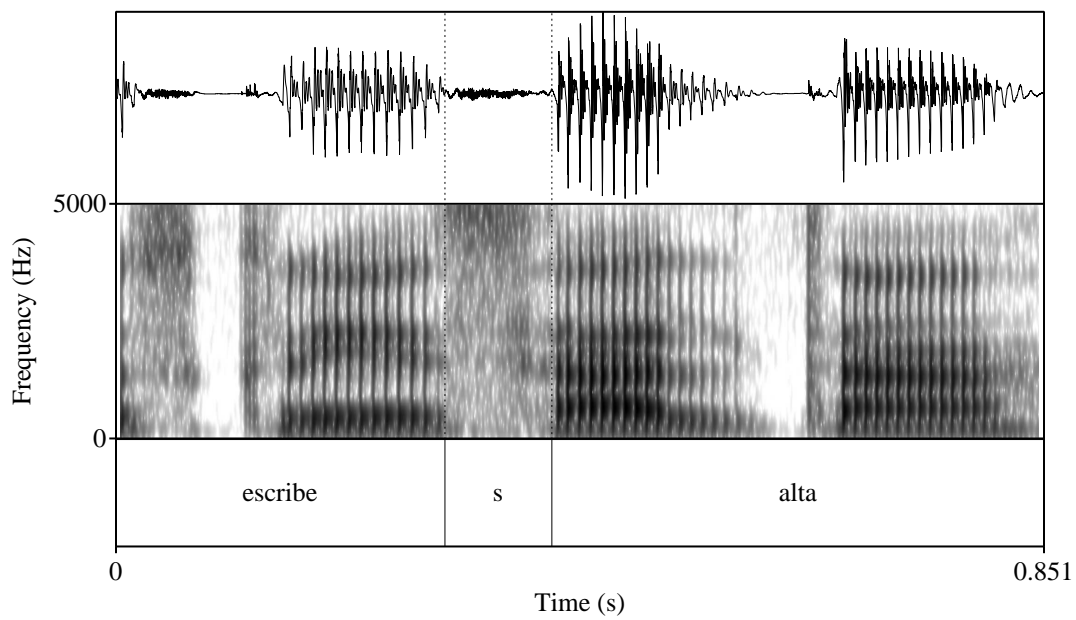
Figure 3.2: *escribes alta* 'you write *tall*'. Duration of /s/ is 96 ms.
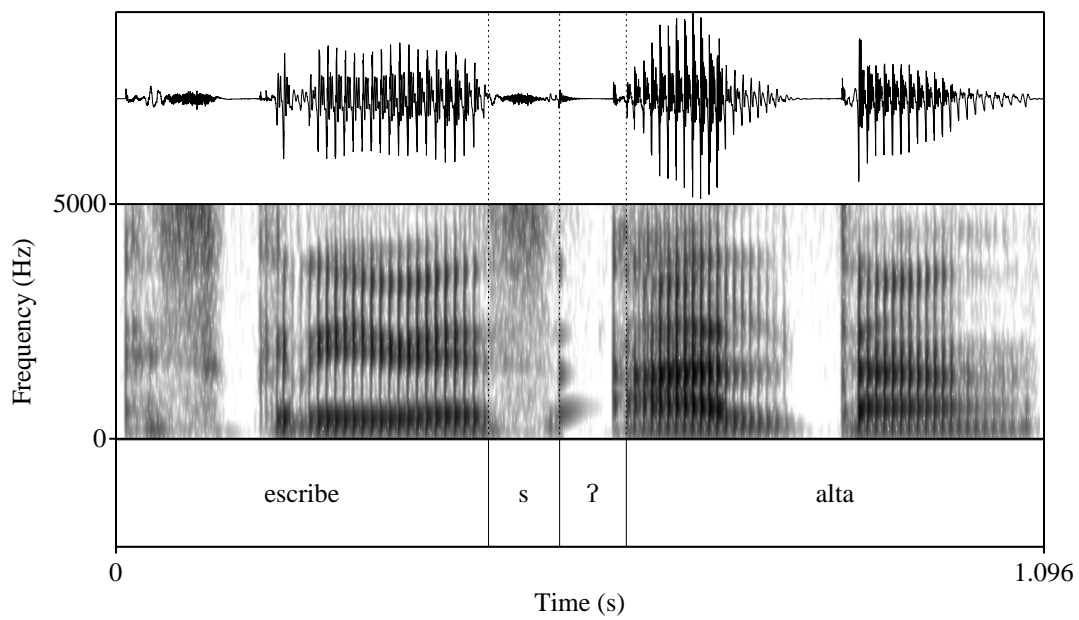
Figure 3.3: Duration of /s/ is 84 ms, duration of /ʔ/ is 79 ms

examined using Praat (Boersma and Weenink, 2015). The duration values are shown in Figure 3.4, as well as described in Table 3.3. For /n/, the duration of the segment in word-initial context was longer than in the resyllabified context, although some overlap exists. For /s/, durations for both word-initial /s/ and resyllabified /s/ are comparable, suggesting that resyllabification is closer to complete, and that greater confusion should exist for /s/ than for /n/. Additionally, the vowel previous to the pivotal consonant was measured, e.g. the final /e/ of the first word in *escribe seco*. Cross-linguistically, vowels in closed syllables tend to be shorter than vowels in open syllables. This still may hold to be true even if the coda consonant has been post-lexically resyllabified. Subjects therefore may pay attention to the length of the vowel as a cue to the lexical affiliation of the following consonant. Because voiced plosives in Spanish are spirantized between vowels (see the spectrograms above), there is not a clear delineation between the vowels and consonant in the VCV sequences (in these stimuli, always the *ibe*, [iβe], in *escribe*). Segmentation was done by finding a steady state in the formant values of the target vowel. The hypothesized distribution was not found: the length of the vowel appears to be even longer when the consonant is the final segment of a word.

Table 3.3: Duration of pivotal consonant in ms.

	/n/		/s/		Total	
	Mean (SD)	Range	Mean (SD)	Range	Mean (SD)	Range
initial	85.94 (10.60)	70.76 − 104.09	122.76 (16.14)	90.71 − 157.67	104.35 (23.01)	70.76 − 157.67
resyllabified	67.59 (23.97)	31.18 − 110.41	119.81 (16.01)	95.94 − 149.25	93.70 (33.24)	31.18 − 149.25

Three word lists were created for the experiment. Each word list contained only one of the items in each triplet (as in Figure 1) so as not to prime the other members of the triplet. Participants thus heard a total of 36 target stimuli, evenly balanced across the conditions. 72 filler items were also included to obscure the intention of the experiment, encompassing a range of other difficult to perceive contrasts for L2 speakers of Spanish.

3.2.2   Procedures

All participants completed a language background form prior to engaging in the experimental tasks. The first task was a word-monitoring experiment; the second was a forced-choice
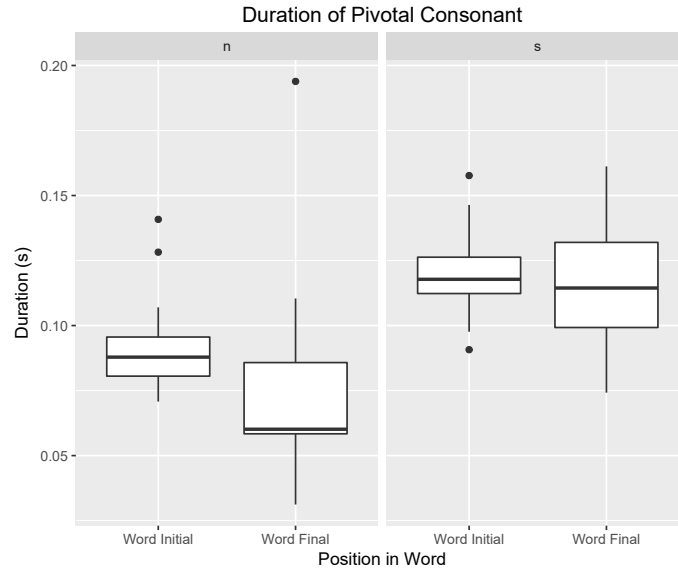
Figure 3.4: Duration of pivotal consonant in Experiment 1, subdivided by word position and consonant

task, using the same word list for both experiments. Both tasks were preceded by a short training session consisting of ten items (chosen from the filler items) and gave the participant feedback after each response. The experiments were designed and ran entirely in Matlab 2011a, using Psychophysics Toolbox extensions (Kleiner, Brainard, & Pelli, 2007); all subjects used the same laptop computer when engaging in the experiment and listened to all of the stimuli at a constant volume while wearing Shure SRH440 Professional Studio Headphones. Subjects spent roughly twenty minutes on Experiment 1. Finally, subjects filled out a word-familiarity questionnaire, in which they rated all of the experimental items from 1 to 3 based on how familiar they were with the given words. All non-native speakers then took a written proficiency test (as described above), testing their knowledge of Spanish vocabulary and grammar.

Word-Monitoring Task

In the word-monitoring task, participants saw a word on the screen and were told to press the spacebar as soon as they heard the word on the screen. If they did not heard the word on the screen, they were told to simply do nothing and wait for the next item to appear on the screen (4 seconds after the end of the stimulus). For example, the participant would

first see "seco", and then would hear [eskribes#ʔeko]. Because the glottal stop is a strong acoustic cue for the placement of a word boundary, clearly denoting that the second word is "eco", the participant should do nothing. The items were randomized within two blocks: in the first block, for each item, the visually-presented word was randomly selected between the two words in each pair presented in Table 3.2. In the second block, the subject heard the same items (in a randomized order), but were presented with the other member of the pair from Table 3.2. Subjects thus rated 36 words twice for a total of 72 responses. Responses and reaction time were automatically recorded.

3.2.3 Measures

In all of the data, the chosen measure is whether or not the aural word was identified as having an initial consonant. For example, if the aural stimulus was *escribes eco* and the presented stimulus on the screen was consonant-initial *seco*, if the participant pressed the spacebar (audio-visual match), the participant is recorded as having heard the aural stimulus as a word with an initial consonant. This measure was chosen over accuracy due to both the nature of the hypothesized resyllabification and of the research questions. If resyllabification results in ambiguity (i.e. duration is not a consistent cue), then the word that the speaker intends to produce in the recorded stimuli is irrelevant. What is more of interest is the acquisition of ambiguity in the resyllabification contexts. The lower proficiency second-language learners should thus be more biased to identify more words as being consonant-initial.

3.2.4 Results

The results for percentage of pivotal consonant identified as word initial are summarized Figures 3.5, and 3.6. Overall, stimuli with a glottal stop were successfully identified as matching a vowel-initial word or as mismatching a consonant-initial word over 93% of the time, indicating that all subjects are able to use the glottal stop as a positive word boundary signal.

Of the aural stimuli that did not include a glottal stop (Figure 3.7), beginning L2 learners had the highest percentage of words perceived as consonant-initial (nearly 65%), whereas
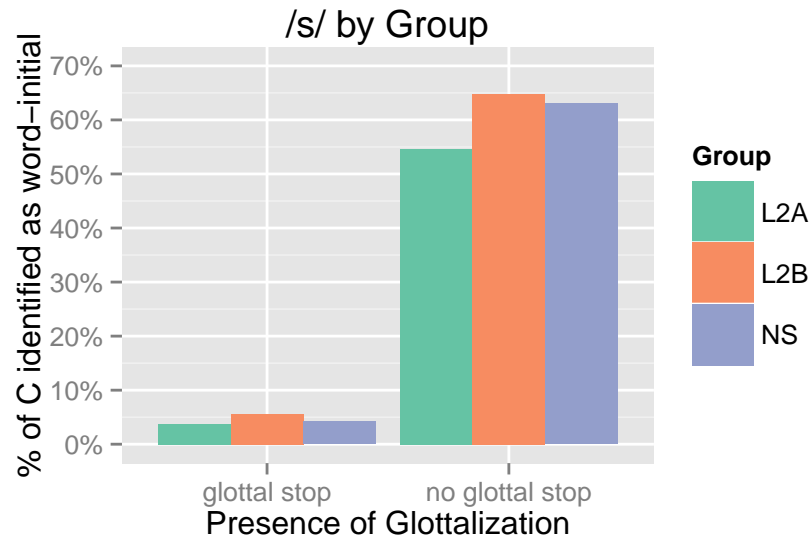
Figure 3.5: Word-monitoring task – Percentage of responses where the target word was identified as consonant initial.
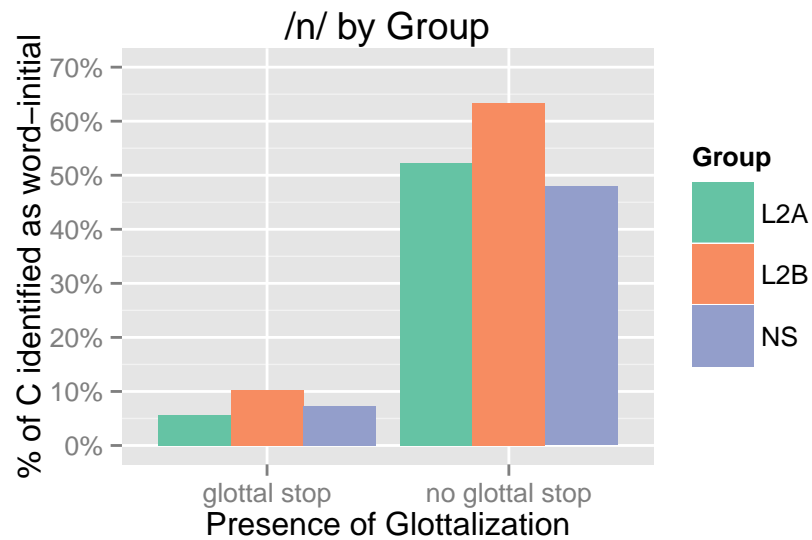
Figure 3.6: Word-monitoring task – Percentage of responses where the target word was identified as consonant initial.
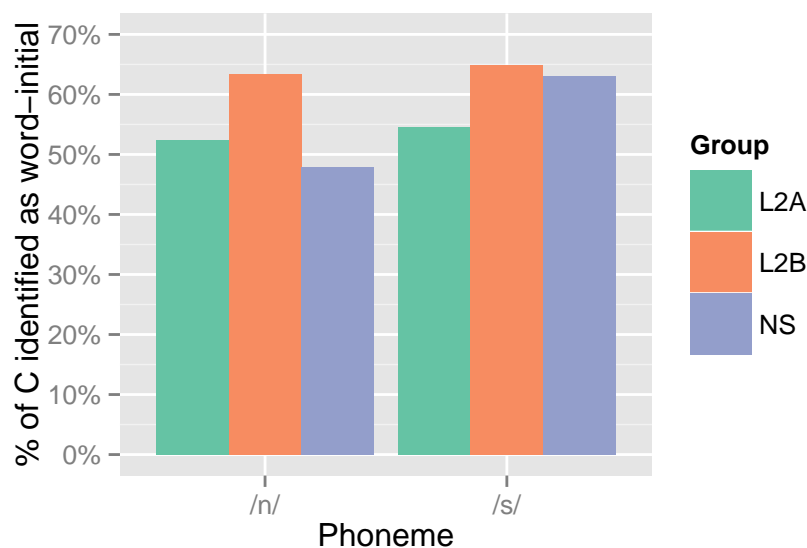
Figure 3.7: Word-monitoring task – Percentage of responses where the target word was identified as consonant initial.

native speakers perceived words as being consonant-initial only 55% of the time. Advanced learners performed similarly to native speakers, perceiving words as being consonant-initial 53% of the time. However, there was an asymmetry between the two target consonants /n/ and /s/: native speakers perceived /n/ as word-initial 48% of the time, more similarly to advanced L2 speakers, but perceived /s/ at a much higher rate, 63% of the time, similar to beginning L2 speakers. Both beginning and advanced L2 speakers performed similarly with both /n/ and /s/ stimuli.

More detailed differences between the subject groups' performances can be seen by correlating other measures with their responses. For the L2 speakers, as seen in Figure 3.8, there is a negative correlation between the selection of word affiliation of the pivotal consonant and proficiency, as measured by the DELE test (Pearson's r = -0.524, p = 0.02). This suggests that an improvement in proficiency will remove the bias towards hearing consonant initial words.

Despite categorical overlaps in duration between word-initial (V#CV) and resyllabified (VC#V) consonants, it was hypothesized that speakers may be able to use durational cues while listening to recorded phrases to place the pivotal consonant in the first or second word. In the following scatterplots (Figures 3.9a through 3.9f), each point represents one stimulus,
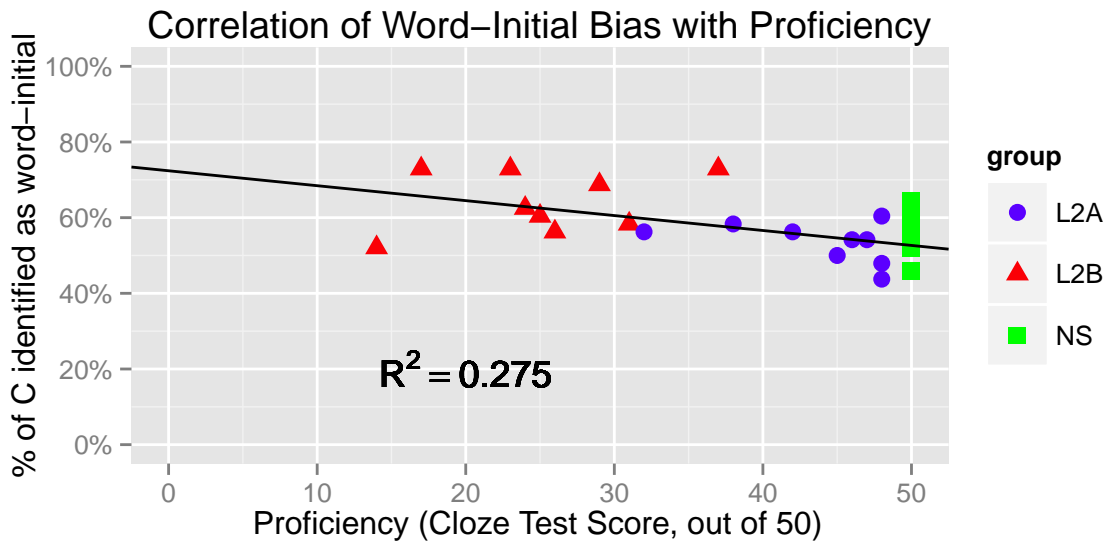
Figure 3.8: Each point represents a single subject, with their average response (percent identified word-initial consonant) correlated with their proficiency score (r = -0.524, p = 0.02).

with all of its subject ratings collapsed. Out of interest, the initial segment of the intended word is denoted by the markers on the plots: a filled black circle denotes the intended word was consonant initial (e.g. *seco*), while an orange triangle denotes that the intended word was vowel initial (e.g. *eco*). However, the linear regression line and the correlation coefficient data fit all of the data points shown in the graph.

Table 3.4: Pearson's correlation coefficient. None: 0.0-0.09, Small: 0.1-0.3, Medium: 0.3-0.5, Strong: 0.5-1.0. In parenthesis the p value is given.

	/n/	/s/	Overall
L2A	0.55 (p <0.001)	0.08 (p = 0.62)	0.30 (p = 0.011)
L2B	0.41 (p = 0.014)	0.33 (p = 0.49)	0.34 (p = 0.003)
NS	0.57 (p <0.001)	0.01 (p = 0.94)	0.53 (p <0.001)
All Groups	0.71 (p <0.001)	0.20 (p = 0.25)	0.54 (p <0.001)

The Pearson's correlation coefficients, which show the strength and direction of the correlation, and the relevant p values are presented in Table 3.4. Of particular interest is the interaction between group and phoneme. The asymmetry between /n/ and /s/ for native speakers seen in the raw percentage of consonants identified is strongly seen here. Native speakers are able to use the consonant duration cues the best in order to determine whether
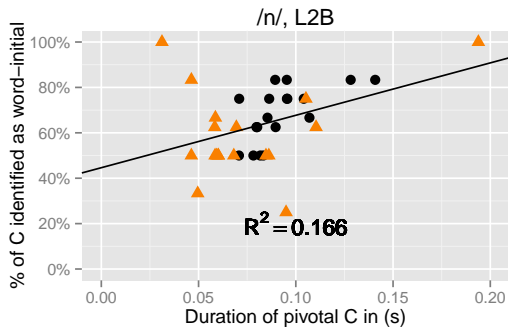
(a) Scatterplot depicting the correlation between duration of /n/ and how often the stimulus was identified as being consonant initial. The group was L2 beginners.
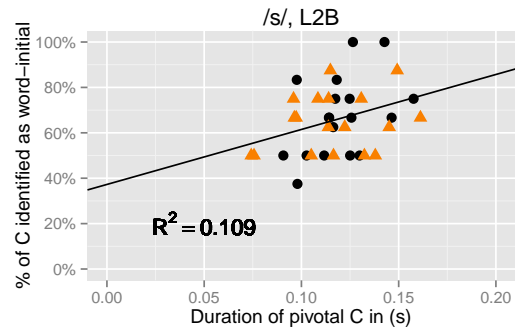
(b) Scatterplot depicting the correlation between duration of /s/ and how often the stimulus was identified as being consonant initial. The group was L2 beginners.
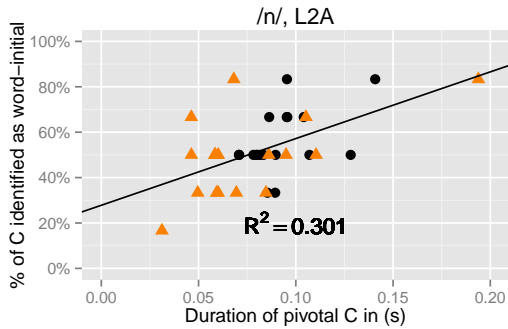
(c) Scatterplot depicting the correlation between duration of /n/ and how often the stimulus was identified as being consonant initial. The group was L2 advanced.
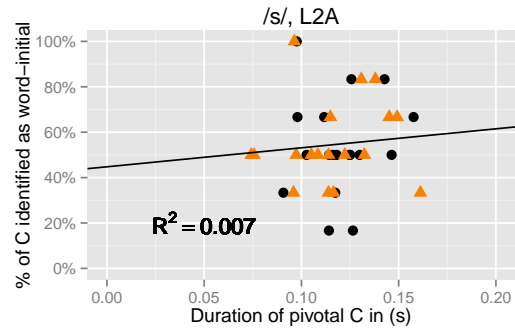
(d) Scatterplot depicting the correlation between duration of /s/ and how often the stimulus was identified as being consonant initial. The group was L2 advanced.
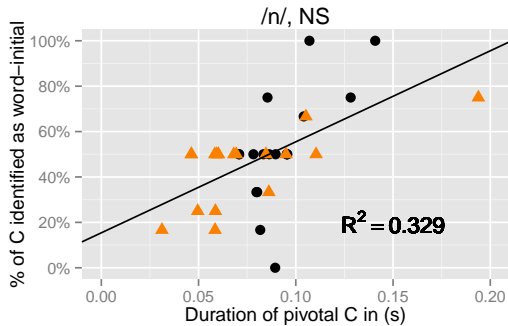
(e) Scatterplot depicting the correlation between duration of /n/ and how often the stimulus was identified as being consonant initial. The group was native speakers.
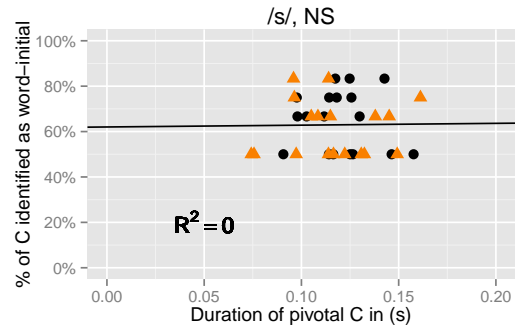
(f) Scatterplot depicting the correlation between duration of /s/ and how often the stimulus was identified as being consonant initial. The group was native speakers.

Figure 3.9: Scatterplots depicted the results from Experiment 1.

or not the consonant heard belongs to the second word (r = 0.57), followed by the advanced learners (r = 0.55), and then beginning learners (r = 0.41). However, for /s/, duration does not correlate as strongly with identification of word affiliation of the consonant; in fact, there is nearly zero correlation for native speakers and advanced learners. Beginning learners do use duration as a cue, however, as there is a medium correlation between duration and consonant position perceived. (r = 0.33)

3.2.5 Discussion

Overall, both the group of beginners and the group of native speakers performed differently than expected: although the beginning learners have the highest rate of perceiving consonants as word initial, they are able to use durational cues in perceiving that some words are vowel-initial even in the absence of glottalization. The native speakers, on the other hand, did not perceive a consonant as word-initial at chance, as was hypothesized by complete resyllabification in Spanish; rather, they also use acoustic cues to determine the lexical affiliation of the consonant, but differed in their ability to use that information based on the identity of the target consonant: they were unable to use duration cues for /s/ in the Word-Monitoring experiment. This is consistent with previous work on native Spanish speakers' inability to use durational cues associated with /s/ (Scarpace and Hualde, 2013), despite that /s/ and /n/ carry the same functional load in Spanish as verbal inflections. In the absence of the usage of acoustic cues, the speakers may resort to other psycholinguistic facts, such as a cross-linguistic tendency for CV syllables and for the word to align with the beginning of a CV syllable (the Possible Word Constraint). Advanced learners performed the most similarly to the hypothesized native speaker group: they identify consonant as word-initial only about half of the time: movement towards this value is seen to correlate with improvement in proficiency in both learner groups. Advanced learners are able to use acoustic cues, although, like the native speakers, they seem to be less able to do so with /s/.

   The results suggest that beginning learners have learned that resyllabification exists in Spanish. Even though ambiguous phrases such as the ones presented to the speakers occur in spoken language at a very low frequency (*la salas / las alas* are likely to be resolved

on the basis of context), and thus have few opportunities where misparsing of the word boundary alone results in meaning failure, beginning learners have heard enough tokens of common vowel initial words in resyllabification contexts (following a consonant final word) to correctly map these words onto other phonetic representations. This (and the advanced learners' success in acquiring resyllabification) suggests that learners have developed a new parsing strategy which must take place early in acquisition in order to successfully extract lexical entries from running speech. Alternatively, L2 speakers may be able to tap into similar processes that possibly occur in casual speech in English, such as ambisyllabicity across word boundaries, where a consonant may belong to both the first and second words, or resyllabification that may happen in casual speech with [ɾ]. These strategies, along with the strong cue of glottal stop as a word boundary, may coexist for the second language learner, and are activated appropriately depending on the amount of proficiency: as seen in this study, the beginning language learners are the most likely to perceive consonant-initial words among the groups. The advanced learners show results that are the most consistent with the hypothesis for native speakers—they appear to have learned the phonological resyllabification rule but are not sensitive to the phonetic cues or the peculiar asymmetry between /n/ and /s/ that affects the way the native speakers perceive the words.

3.3   Experiment 2

The research study described in this paper seeks to add to the current research on second-language acquisition of segmental contrasts by comparing a contrast that occurs transparently at the word-internal intervocalic position, and occurs very rarely in between words, but does not occur in other word positions. The tap/trill contrast in Spanish is not only an important phonemic contrast, but it also serves as a significant cue to word segmentation because the trill is the only rhotic to appear in word-initial position. (1) Are second-language learners of Spanish more sensitive to the tap/trill contrast in the word-internal condition than the word-boundary condition? (2) Do second-language learners become more native-like with greater proficiency? To investigate these questions, a word-monitoring task was devised that could test the sensitivity of learners to the phonological system of the rhotics

in Spanish. Because words were presented on the screen during the task, the question of orthography comes into play. However, because many students learn Spanish in a classroom context, it is not possible (and perhaps not advisable) to divorce the orthographic representations of the rhotics from the sounds themselves. The participants were asked to monitor for the first few syllables of a specific word represented on the screen; however, the relationship between what was presented on the screen and what was to be heard differs for each condition. For the word-internal condition, the task is more straightforward. If the word contained an intervocalic <r>, as in barones, the participant was to listen for an intervocalic tap [ɾ]. If the word contained an intervocalic <rr>, as in *barrotes*, the participant was to listen for an intervocalic trill [r]. The task was quite different for the word-boundary condition because the acoustic information preceding the word presented on the screen was important. If the word on the screen began with an <r>, as in *relojes*, then the participant was to listen for an intervocalic trill [r]. If the word on the screen did not begin with an ¡r¿ at all, but rather a vowel, as in elotes, then the participant was to listen for the first few syllables in that particular word—in each case, it would be preceded by a tap [ɾ]. This should be the most difficult condition based on orthography because the participant must reject a token such as *pedí*[r]*elo-* when monitoring for *elo*(*tes*) because a V[r]V sequence must be segmented as V#[r]V and never as V[r]#V. Thus, it is hypothesized that for learners of Spanish, the word-internal contrast should be the easiest, because it is both robust in the language and transparent in orthography, with no obvious differences between the conditions in which they are to monitor for a tap or a trill. The word-boundary conditions should be harder, because the orthographic representation is more opaque (in fact, the <r> represents a trill, whereas in the word-internal condition, the <r> represents a tap—within the context of the experiment, the participants must pay attention to word position carefully). Within the word-boundary conditions, it should be easier to assign <r> as a word-initial trill due to the overall salience of the trill as being associated with an <r>, and relatively difficult to reject the intervocalic trill at the word-boundary when monitoring for a vowel-initial word. All conditions should improve with improved proficiency.

3.3.1 Participants

Forty total participants spent roughly thirty minutes to complete all of the experimental tasks (fifteen minutes for native Spanish speakers). A control group was formed of twenty native Spanish speakers ('NS') from various dialects. Seven of the twenty native speakers were recruited and tested at the Colegio de México in Mexico City. All native speakers had learned Spanish from birth, although three speakers were bilingual: two speakers also spoke Catalan, and one had moved to Galicia at a young age. The twenty non-native speakers ('L2') were recruited both from the graduate and undergraduate student population at the University of Illinois, Urbana-Champaign. The undergraduate participants were enrolled in upper division coursework in Spanish at the time of testing. The English native speakers were assessed in their Spanish proficiency through a written test, consisting of the vocabulary and cloze parts of the DELE test, just as in the previous study.

3.3.2 Materials

A male Native Spanish speaker from Mexico recorded the stimuli for the experiment. The stimuli were verb-noun or verb-adjective pairs where a V[r $\sim$ ɾ]V sequence straddled the word boundary. The first word in these phrases were common verbs, either in the infinitive or conjugated in the imperfect (-ir verbs) or present tense (-ar/-er) because the lack of the final <r> in these verbs results in a conjugated form, e.g. *ver/ve*. The second word in each pair had the first two syllables in common, save for the target rhotic. For example, *pedi*[ɾ] *elotes* 'to ask for corn' (word final <r>: tap) vs. *pedí* [r]*elojes* 'I asked for clocks' (word-initial <r>: trill). Similar pairs to test the word-internal contrast were also constructed; for example, *veo ba*[ɾ]*ones* 'I see barons' vs. *veo ba*[r]*otes* 'I see bars'. The recordings were segmented in Praat and truncated at the disambiguation point, e.g. *pedi*[ɾ/r]*elo-*, such that the identity of the rhotic was the only cue to the identity of the second, truncated word. Sixteen pairs of each type were created and organized into four word lists, along with thirty-two filler items. The pairs were divided among each of the word lists such that any one subject did not see or hear both members of any particular pair during the experiment. For example, subject A might hear *pedir elo-* and see ELOTES, whereas subject B would hear *pedí relo-* and also

see ELOTES. Neither subjects A nor B would see the target word RELOJES.

3.3.3 Procedures

All participants completed a language background form prior to engaging in the experimental task, which was a word-monitoring experiment. The tasks were preceded by a short training session consisting of ten items (chosen from the filler items) and gave the participant feedback after each response. The experiments were designed and ran entirely in Matlab 2011a, using Psychophysics Toolbox extensions (Kleiner et al., 2007); all subjects used the same laptop computer when engaging in the experiment and listened to all of the stimuli at a constant volume while wearing Shure SRH440 Professional Studio Headphones. Subjects spent roughly ten minutes on the experiment. Finally, all non-native speakers then took a written proficiency test (as described above), testing their knowledge of Spanish vocabulary and grammar. The experiment was similar to the word-fragment priming task of (Soto-Faraco et al., 2001) and went as follows: the target word of the critical phrase was presented visually on a computer screen in front of the participant; for example, *elotes* or *relojes*. The truncated auditory stimulus (the two-word phrase) was then presented to the participant, and the participant was asked to press the space bar if the second word in the phrase could possibly be the one visually presented on the screen. If not, the participant would do nothing, and the next trial would begin after two seconds. For example, the participant might first see the word elotes on the screen, and then would hear *pedi*[ɾ] *elo-*. In this "match" condition, a press of the spacebar was marked as a correct response, because in the two-word condition, the tap must be a word-final segment, and the second word must begin with a vowel. The task was the same for the word-internal condition: a subject might have heard *veo ba*[ɾ]*o-* and decide if *barones* could have been the second word. Orthography is the main difference between the two conditions: in the word-internal condition, the phoneme-to-grapheme representation is transparent: a trill is written as <rr> and a tap is written as <r>; both could appear in the visually-presented word in the experiment. In the word-boundary condition, a trill would be represented by an initial <r>, but no written <r> is presented on the screen for a tap because that segment belongs instead to the preceding word

59

(not presented on the screen); instead, a vowel-initial word could correspond to a two-word phrase including a tap at the word boundary. Participants heard 32 target items (8 word-internal taps, 8 word-internal trills, 8 word-boundary taps, 8 word-boundary trills) as well as 32 filler items (exhibiting other segmental contrasts in Spanish such as voicing of stops) in a randomized order.

Table 3.5: Accuracy Results for the Word-Internal Condition

see:	<r>	<r>	<rr>	<rr>
	barones	*barones*	*barrotes*	*barrotes*
hear:	[ɾ]	[r]	[r]	[ɾ]
	veo ba[ɾ]o-	*veo ba[r]o-*	*veo ba[r]o-*	*veo ba[ɾ]o-*
	MATCH	NO MATCH	MATCH	NO MATCH
NS %	95.00%	86.58%	100.00%	86.58%
L2 %	91.25%	40.96%	98.68%	37.50%

Table 3.6: Accuracy Results for the Word-Boundary Condition

see:	<#V>	<#V>	<r>	<r>
	elotes	*elotes*	*relojes*	*relojes*
hear:	[ɾ]	[r]	[r]	[ɾ]
	pedi[ɾ]elo-	*pedi[r]elo-*	*pedi[r]elo-*	*pedi[ɾ]elo-*
	MATCH	NO MATCH	MATCH	NO MATCH
NS %	87.50%	72.50%	94.67%	84.70%
L2 %	85.00%	51.25%	92.21%	43.37%

3.3.4 Results

It is important to keep in mind what each condition represents when analyzing the data: Tables 3.5 and 3.6 provide examples of what each experimental condition was like for the participant, along with the mean accuracy data for each of the two speaker groups. For example, in Table 3.5, which describes the word-internal condition, the subject might have seen a word with an orthographic <rr> represented on the screen, such as *barrotes*. The participant was told to press the spacebar if the second word in the two-word phrase could have been *barrotes*. The participant then heard *veo barro-*, which included a trill. Because the target word was the same as the word in the auditory stimulus (although truncated), this condition was considered to be a "match", and a press of the spacebar was considered
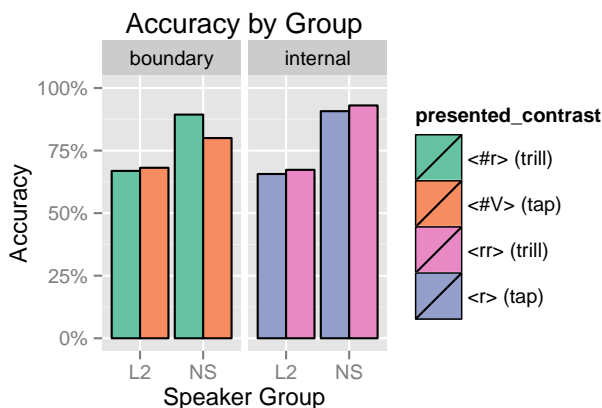
Figure 3.10: Overall accuracy in each condition by group.
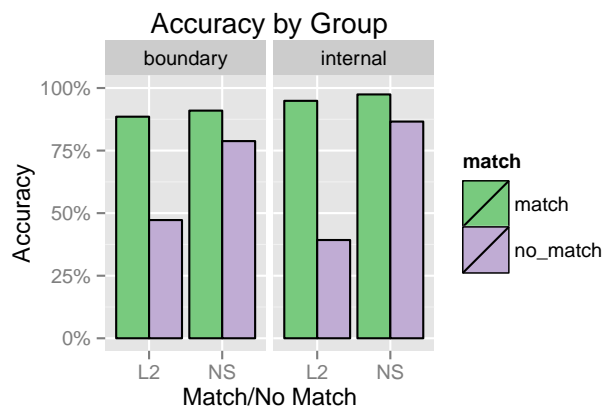
Figure 3.11: Overall accuracy sorted by match/no match condition.

to be an accurate response. In such trials, native speakers had a 100% hit rate, and second-language speakers had a 98.68% hit rate. The following column describes a nearly identical condition, where the participant was told to monitor for a word with orthographic <rr> but this time heard *veo ba*[ɾ]*o-*, which included a tap instead of a trill. In this "no match" condition, a press of the spacebar was considered to be an inaccurate response. Native speakers correctly rejected these auditory stimuli 86.58% of the time, but second-language speakers did so only 37.50% of the time.

The accuracy results are summarized in the following graphs, Figures 3.10 and 3.11. Native Speakers outperformed the L2 speakers in all conditions (overall, L2: 67% accurate, NS: 89%). However, when the accuracy data are separated by the match and non-match conditions (match: the spacebar press is a correct response), the two groups perform quite differently. The two groups have a higher rate of accuracy in the match condition than the non-match condition, suggesting that there is a bias towards pressing the spacebar. This difference between the two conditions is heightened in the L2 group: the L2 group is 91.7% accurate in the match condition, but 43.3% accurate in the non-match condition, whereas the NS group responded 94.2% and 82.7% respectively.

A d-prime analysis calculating biases revealed differences between the two speaker groups, see Figure 3.13. Within the L2 speakers, proficiency was highly correlated with d-prime (R^2 = 0.44).
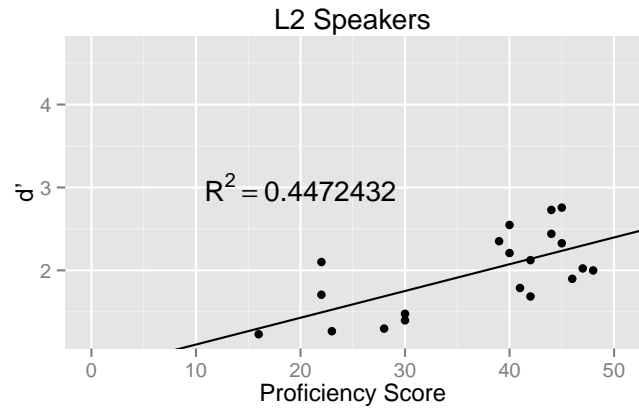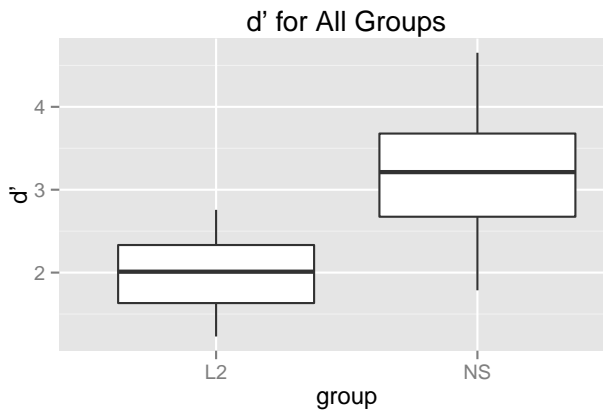
Figure 3.12: d-prime results for both groups.

Figure 3.13: Correlation scatterplot between d-prime and Spanish proficiency, L2 speakers.

Because only small differences were found between the groups and the conditions were found in the match condition, the rest of the data considered here will only focus on the non-match condition, thus the accuracy data represents percentage of correct rejections. The implications of choosing such an analysis can be found in the Discussion section. Figure 3.14 details the same information as in Figure 3.11 but only focuses on the non-match condition.

Figures 3.15 and 3.16 reveal correlations between proficiency and accuracy in each of the conditions described in Tables 3.5 and 3.6. These results are obscured by the mean data presented in the earlier bar graphs: where it is appears that there is no-to-little difference between the overall accuracy (of correct rejections) of the L2 speakers in the word-boundary and word-internal conditions, the R^2 values presented in these scatterplots tell a different story. First, Figure 3.15, which describes the word-internal condition: in the match conditions, all subjects perform close to ceiling. In the non-match condition, proficiency has a strong correlation with percentage of correct rejections, especially in rejecting [r] as a match for <r>. In Figure 3.16, which describes the word-boundary condition, many of the participants perform close to ceiling in the match conditions, but there is an exceptional amount of individual variety in the non-match conditions. Proficiency does not seem to help either: R^2 is 0 in the non-match/vowel-initial word condition, as even advanced learners perform equally poorly and native-like.

A series of linear mixed effects models with speaker and item as a random effect were
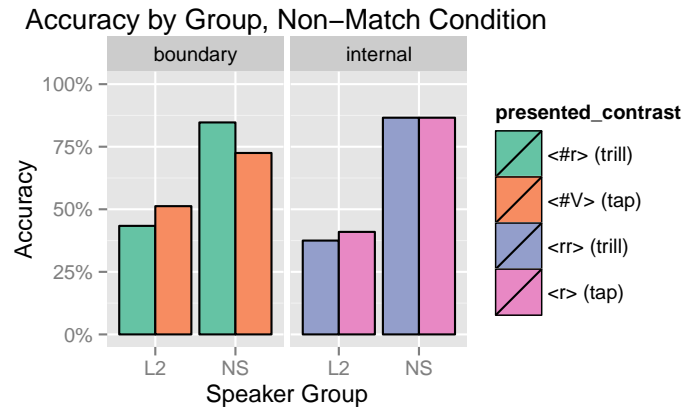
Figure 3.14: Percentage of Correct Rejections for each group divided by position of the contrast in the non-match condition: either at the word boundary (left) or word-interally (right). Each condition is further divided by the orthographic cue present in the visually-presented stimulus. For example, for <r> in the boundary condition, the word presented might be *relojes*. Because this is the non-match condition, the auditory stimulus heard would have involved a tap: *pedir elo-*. Thus, L2 speakers were below chance at correctly rejected *relojes* as the second word in *pedir elo-* .
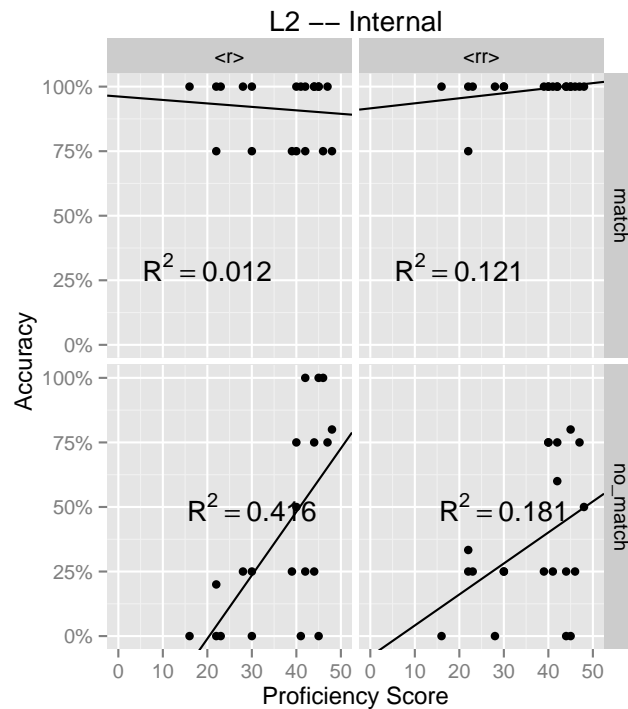
Figure 3.15: Scatterplots and correlations between proficiency scores and accuracy in each of the four conditions from Table 3.5, the word-internal condition. The "match" conditions represent hits and the "no match" conditions represent correct rejections. Each point on the plot represents a single L2 participant.
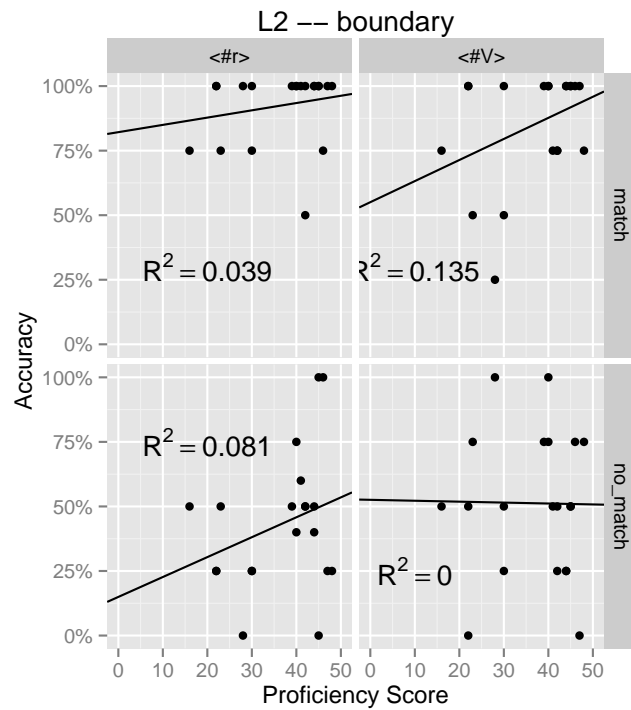
Figure 3.16: Scatterplots and correlations between proficiency scores and accuracy in each of the four conditions from Table 3.6, the word-boundary condition. The "match" conditions represent hits and the "no match" conditions represent correct rejections. Each point on the plot represents a single L2 participant.

created in order to determine if various fixed effects were significant. With data from both the native speakers and second-language learners combined, the best model associated accuracy with group and match condition ($p < 0.001$); there was a weak association with position of contrast ($p = 0.05$), but there was also a significant interaction between group and position of contrast ($p = 0.012$), and group and match condition ($p < 0.01$). Within the group of second-language speakers, the best model predictor of accuracy was an interaction between position of contrast and match condition ($p = 0.002$) and a three-way interaction between proficiency, position of contrast, and match condition ($p = 0.014$). There was also a weak association with proficiency alone ($p = 0.05$).

3.3.5 Discussion

There are a number of surprising findings in this study. First of all, the Second Language speakers perform notably worse in the non-match conditions; that is, when a correct response means rejecting a match between the visually and auditorily presented stimuli. There may be inherent bias in the experimental design to pressing the space bar, thereby over-accepting matches (the opposite response is to not press any button at all), or it may be the case that this high rate of false alarms is indicative of a lack of acquisition of the contrast, even in the basic word-internal context (e.g. *caro* vs. *carro*). This result neither confirms nor denies the results given in citetRose2010: recall that this previous study investigated learners' abilities to simply discriminate between minimal pairs, a much more explicit task. The task presented here requires participants to not only identify which sound is being heard but also to correctly assign it to the orthographic (or lexical) form represented on the screen. That is, even some of the lower proficiency speakers who have nonetheless had at least of year of Spanish study are not able to quickly reject [r] as a pronunciation of <r>. Simply put, both /r/ and /ɾ/ (as well as <r> and <rr>) belong to the same category, and the two sounds are in free variation for these lower proficiency speakers. A production task might bear this finding out; however, due to the difficulty in producing the two rhotics in Spanish, production of /r/ and /ɾ/ in this context as [ɹ] would not necessarily reflect one underlying category, but it may in fact reinforce it. Face (2006) found that in the word-

internal intervocalic position, there was a stage in which learners of Spanish were producing [ɾ] for both <r> and <rr> in a reading task, which supports this result. Although the tap/trill contrast in Spanish is an important distinction in Spanish, functionally, there are very few minimal pairs. For a beginning learner, the pairs *pero* 'but' / *perro* 'dog' and *caro* 'expensive' / *carro* 'car' are common, but it is likely that at an early stage of lexical encoding, all of these entries include the English /ɹ/ or some non-distinct rhotic. Because of the lack of other common minimal pairs, the contrast may be very slowly acquired, and these early entries may never be repaired. The lack of contrast may be reinforced by the variation between [r] and [ɾ] in syllable final position, which also has a lack of orthographic transparency. It is therefore of no great surprise that the word-boundary condition shows even greater variability among the higher proficiency speakers. This condition is also more difficult for the native speaker group, but even some of the highest proficiency second-language speakers have a lack of awareness that word-initial <r> is always a trill. The task is harder in this condition because of the necessity of paying attention to the preceding word context, but it might be additionally harder because /r/ presents an exception to the normal rules of resyllabification between words in Spanish. The usual rule is that /V#CV/ and /VC#V/ sequences are homophonous; that is, la salas 'you salt it' and las alas 'the wings' are identical, especially in fast speech. However, as mentioned above, *ve rocas* and *ver ocas* are never identical: the tap or trill is the cue to the placement of the word boundary. Pairs of phrases that are ambiguous are non-existent in the input of beginning learners, and the distribution of /r/ is never explicitly taught in classes. A follow-up production study would be illustrative to find out what realizations of /r/ second language speakers use in contexts other than the word-internal, intervocalic context. The amount of individual variation calls into question the usefulness of pooling second-language speakers together as a single group. Although overall scores (such as d-prime) improved with greater proficiency, in the most difficult conditions, proficiency was not correlated with accuracy. The individuals with high proficiency were mostly graduate students in the Department of Spanish, and some may have had explicit pronunciation instruction and may be metalinguistically aware of the tap/trill contrast at the word boundary. Any further research should explicitly ask participants after the experimental tasks have been completed if they are aware of the rule in question.

3.3.6 Conclusion

This study has shown that with second language speakers are able to acquire the tap/trill contrast in Spanish with varying levels of results. In the word-internal position (*pero/perro*), an increase in proficiency is well correlated with an increase in performance. However, in the word-boundary condition (*ve rocas/ve rocas*), only some second language speakers were able to correctly assign the tap or trill to the lexical item to which it belongs. Therefore, while there is an increase in accuracy as the proficiency level of the learner increases, it seems that contrasts at the word boundary are harder to acquire. This difficulty may be a result of the weak functional load that the tap/trill contrast holds (only between vowels, in free variation elsewhere except word-intially), or it may be due to the lack of transparency in the orthography (written as <r> in all cases). The word-boundary condition also has implications for word-segmentation strategies in the second language: there is weak evidence here that second-language learners are able to use the allophonic cue of the tap/trill to correctly place a word boundary. This study also shows that the lexical encoding of these rhotics is fairly weak for beginning learners of Spanish, as they are willing to accept either the tap or the trill as being in any of the presented words. Future research should couple perception tasks with production tasks in order to further understand the state of the acquisition of the rhotic system in second-language learners' Spanish.

CHAPTER 4

EYETRACKING

4.1   Why eye tracking?

While other experimental paradigms in processing recognition of spoken words, such as gating tasks and cross-modal priming (see Chapter 3 and Davis et al. (2002), among others) have shown to be successful, it is crucial to reduce task effects in any linguistic experiment. Using eye tracking with the visual world paradigm is a more intuitive task for participants: in essence, participants simply listen to words or phrases and click on a relevant object on the screen. They do not have to perform metalinguistic tasks such as deciding whether or not a given word exists within a given amount of time. Additionally, the dependent measure, eye gaze, is unconscious and automatic. The time course measures from eye tracking have also shown to be highly similar to the predictions made by the TRACE model (Allopenna et al., 1998) regarding lexical activation as the acoustic signal unfolds. However, as will be shown throughout this chapter, the measures analyzed in eye tracking are just as important to understanding how lexical activation occurs in the actual experiment.

4.2   Research Questions

1. For native speakers, in real time, are both consonant- and vowel-initial words activated equally in the VCV context across word boundaries? For example, given a string of phonemes /benobe/, which is consistent with the beginning of both *ve novelas* 's/he sees novels' (consonant-initial) and *ven ovejas* 'they see sheep' (vowel-initial), will proportions of eye-gazes to the images of the novels and sheep be equal? If so, the word recognition system of Spanish speakers is sensitive to resyllabification. If not, and

CHAPTER 4

EYETRACKING

4.1   Why eye tracking?

While other experimental paradigms in processing recognition of spoken words, such as gating tasks and cross-modal priming (see Chapter 3 and Davis et al. (2002), among others) have shown to be successful, it is crucial to reduce task effects in any linguistic experiment. Using eye tracking with the visual world paradigm is a more intuitive task for participants: in essence, participants simply listen to words or phrases and click on a relevant object on the screen. They do not have to perform metalinguistic tasks such as deciding whether or not a given word exists within a given amount of time. Additionally, the dependent measure, eye gaze, is unconscious and automatic. The time course measures from eye tracking have also shown to be highly similar to the predictions made by the TRACE model (Allopenna et al., 1998) regarding lexical activation as the acoustic signal unfolds. However, as will be shown throughout this chapter, the measures analyzed in eye tracking are just as important to understanding how lexical activation occurs in the actual experiment.

4.2   Research Questions

1. For native speakers, in real time, are both consonant- and vowel-initial words activated equally in the VCV context across word boundaries? For example, given a string of phonemes /benobe/, which is consistent with the beginning of both *ve novelas* 's/he sees novels' (consonant-initial) and *ven ovejas* 'they see sheep' (vowel-initial), will proportions of eye-gazes to the images of the novels and sheep be equal? If so, the word recognition system of Spanish speakers is sensitive to resyllabification. If not, and

looks to the consonant-initial word dominate, universal processing strategies privileging syllable-word boundary matches prevail, despite resyllabification. The hypothesis is that proportions of eye-gazes will be roughly equally to both consonant- and vowel-initial words.

2. For native speakers, in real time, are both /r/- and vowel-initial words activated equally in the V/r/C or V/ɾ/V context across word boundaries? For example, given a string of phonemes /pedirelo/, which is consistent with the beginning of *pedí relojes* 'I requested clocks' (/r/-initial word) **but not** *pedir elotes* 'to request corn' (vowel-initial word, following /r/), will proportions of eye-gazes to the corn and clocks be equal? If so, Spanish speakers do not use the allophonic distribution of the tap and trill phonemes for online speech segmentation. If not, despite the similarity of the tap and trill segments in this otherwise noncontrastive position, the word recognition system of Spanish speakers uses the distribution of the tap and trill phonemes to properly segment words. The hypothesis is that this allophonic distribution will show that the trill is a strong cue to word initiality.

3. For non-native speakers of Spanish, in real time, are both consonant- and vowel-initial words activated equally in the VCV context across word boundaries? If eye-gaze patterns match those of native Spanish speakers (who presumably do not activate consonant-initial words significantly more than vowel-initial words), the non-native speakers are sensitive to resyllabification in Spanish and are aware that syllable and word boundaries do not always coincide. If not, and gazes to consonant-initial words dominate, non-native speakers continue to use English (or universal) tendencies where syllable and word boundaries coincide and vowel-initial words generally begin with [ʔ]. The hypothesis is that advanced non-native speakers will have learned that resyllabification exists in Spanish and have fewer activations of consonant-initial words when a vowel-initial word is intended than when compared to lower proficiency speakers.

4. For non-native speakers, in real time, are both /r/- and vowel-initial words activated equally in the V/r/C or V/ɾ/V context across word boundaries? If so, non-native speakers do not use the allophonic distribution of the tap and trill phonemes for online

69

speech segmentation. If not, non-native speakers use the distribution of the tap and trill phonemes to properly segment words. Additionally, in order to confirm these results are due to properly processing the difference between the tap and trill phonemes overall, which is traditionally difficult for non-native speakers of Spanish, do they activate words like *perro* and *pera* equally in the word internal context? If so, they have not learned the distinction between these sounds and the word boundary results will be difficult. If not, they the word-boundary results will be more straightforward. The hypothesis is that advanced non-native speakers will have learned this word-boundary phonological distinction between the tap and the trill, while more beginning learners will not.

4.3   Methodology

4.3.1   Participants

9 native Spanish speakers and 28 non-native Spanish speakers (L2) participated in the study. They were graduate and undergraduate students at the University of Illinois. The L2 speakers were divided into two groups: 13 intermediate speakers and 15 advanced speakers, based on recruitment criteria (all advanced speakers were graduate students who had been studying Spanish for more than 5 years). All participants had normal or corrected-to-normal vision and did not report hearing impairment. None of the native speakers were from dialects where final /n/, /s/, or /r/ differ from the Spanish norm (as explained above). Students at the University of Illinois who were enrolled in upper division Spanish courses (SPAN 252, 200, etc.) received course credit for their participation.

4.3.2   Materials

Two lists of 48 two-word stimuli were created with various conditions. In all stimuli, the first word was some form of the verb *ver* 'to see', and the second word was a picturable noun. In the word-boundary conditions, Word 2 began with either /n/, /s/, or /r/ in the

consonant-initial condition, or with a vowel in the vowel-initial condition. For the word-internal condition, which tested the word-internal /r/ ˜/ɾ/ contrast, Word 2 contained a word-medial tap or trill in the second syllable, which was either unstressed (*carros* 'cars') or stressed (*garrafas* 'carrafes'). The two-word pairs formed target-competitor pairs. Examples are given in Table 4.1.

Table 4.1: Example Target/Competitor pairs for the experiment

Consonant	Word-Initial	Word-Final
/n/	*ve novelas* 's/he sees novels'	*ven ovejas* 'they see sheep'
/s/	*ve sardinas* 's/he sees sardines'	*ves ardillas* 'you see squirrels'
/r/	*ve* [r]*anitas* 's/he sees little frogs'	*ve*[ɾ] *anillos* 'to see rings'
/r/ (word-internal)	*ve go*[r]*inos* 's/he sees pigs'	*ve go*[ɾ]*ilas* 's/he sees gorillas'

Clipart images were found for all of the picturable items through Microsoft Office Clipart or through other online search engines. Each target and competitor were matched with two other distractors: a consonant-initial distractor and a vowel-initial distractor, for a four-word set that comprised each trial. The consonant-initial distractor began with a non-similar consonant from that of the target word. This was to ensure that upon hearing the beginning of the consonant, participants could quickly eliminate the distractors from activation. For /n/, words beginning with voiceless stops or fricatives were used. For /s/, words began with sonorants or stops. For /r/, words began with voiceless stops or fricative. Distractor items were also used as filler trials. Filler trials consisted of trials with a target image but no possible competitor (these items were not /s/, /n/, or /r/ initial), or quasi-minimal pairs in which the target and competitor differed only in gender, e.g. *ves votos* 'you see votes' vs. *ves botas* 'you see boots', similar to the pairings of the experimental stimuli where the pair is disambiguated by the end of the word. The 48 four-word sets were created such that each image appeared on the screen twice during the experiment, once as a target or competitor, and once as a distractor. The 48 four-word sets were organized into two lists, such that competitors for targets did not appear in a later trial as a target item, e.g. *novelas* and *ardillas* were target items in trials in List 1, but *ovejas* and *sardinas* were not, and vice versa for List 2.

The auditory stimuli were recorded by a male native speaker of Spanish from Central Mex-

ico. The stimuli were recorded in a sound-attenuated booth in the Phonetics and Phonology Lab at the University of Illinois, with an AKG c520 head-mounted condenser microphone at a sampling rate of 44.1 kHz using a Marantz PMD570 solid-state recorder connected to a Grace m101 preamplifier. The speaker read the phrases in isolation as naturally as possible, and was told to not introduce a pause between the two words. The recordings were then normalized for intensity.

Duration was measured for the pivotal consonant in each of the recordings used as the experimental auditory stimuli (8 with word-initial /n/, 8 with word-final /n/, 8 with word-initial /s/, 8 with word-final /s/, 8 with word-initial /r/, 8 with word-final /ɾ/, 8 with word-medial /r/, and 8 with word-medial /ɾ/. Segmentation was done in Praat by examining the waveform and spectrogram for the attenuation of amplitude for /n/, for the onset of random noise for /s/, and for the beginning of complete occlusions for /r/. Examples are given. Results are given in Table 4.2. Because differences in duration between word-initial and word-final consonants would be a potential cue for listeners, we submitted the measured durations to paired-samples t-tests within each phoneme. No differences were found between word-initial and word-final /n/, $t(7) = 1.77$, p=0.12, or for /s/, $t(7) = -0.30$, p=0.77, suggesting that resyllabification is complete for /n/ and /s/, but as expected, the durations between word-initial /r/ and word-final /ɾ/ did differ significantly, $t(7) = 20.18$, p < 0.0001. The duration between word-internal /r/ and word-internal /ɾ/ also differed significantly, $t(7)$ = 17.75, p < 0.0001, also as expected.

Table 4.2: Means and standard deviations of duration of target consonants for stimuli in eye tracking experiment

Phoneme	Position	Mean (SD) in ms
/n/	initial	67.05 (9.6)
/n/	final	62.61 (8.5)
/s/	initial	91.13 (10.3)
/s/	final	92.57 (9.1)
/r/	initial	61.02 (2.7)
/ɾ/	final	22.75 (5.6)
/r/	medial	67.22 (4.2)
/ɾ/	medial	22.75 (5.6)
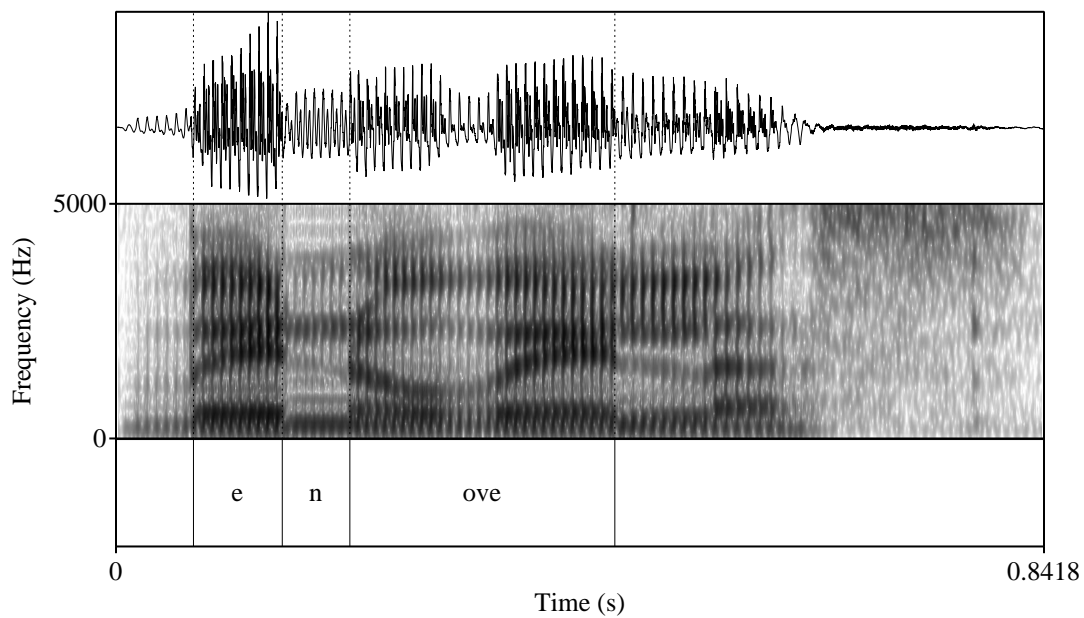
Figure 4.1: *ve novelas* 's/he sees novels'. Duration of /n/ is 61.31 ms.
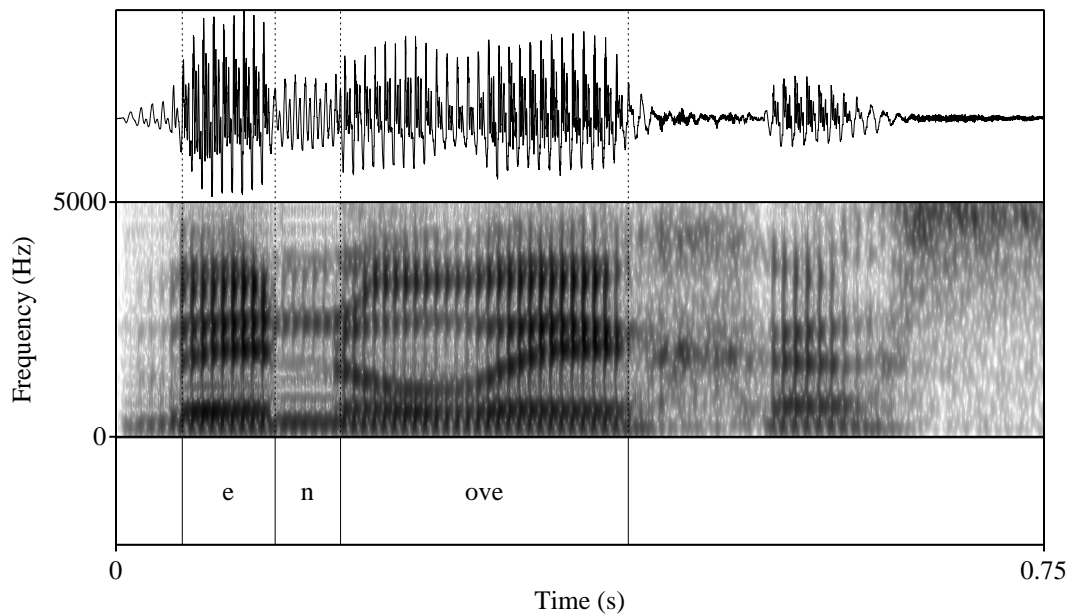
Figure 4.2: *ven ovejas* 'they see sheep'. Duration of /n/ is 52.85 ms.
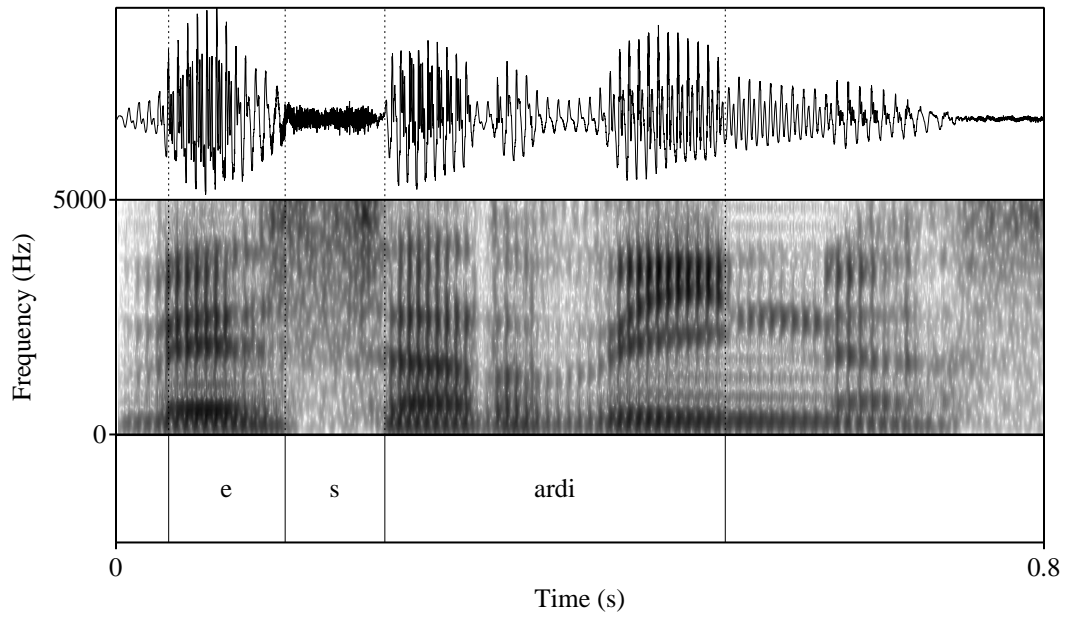
Figure 4.3: *ve sardinas* 's/he sees sardines'. Duration of /s/ is 85.93 ms.
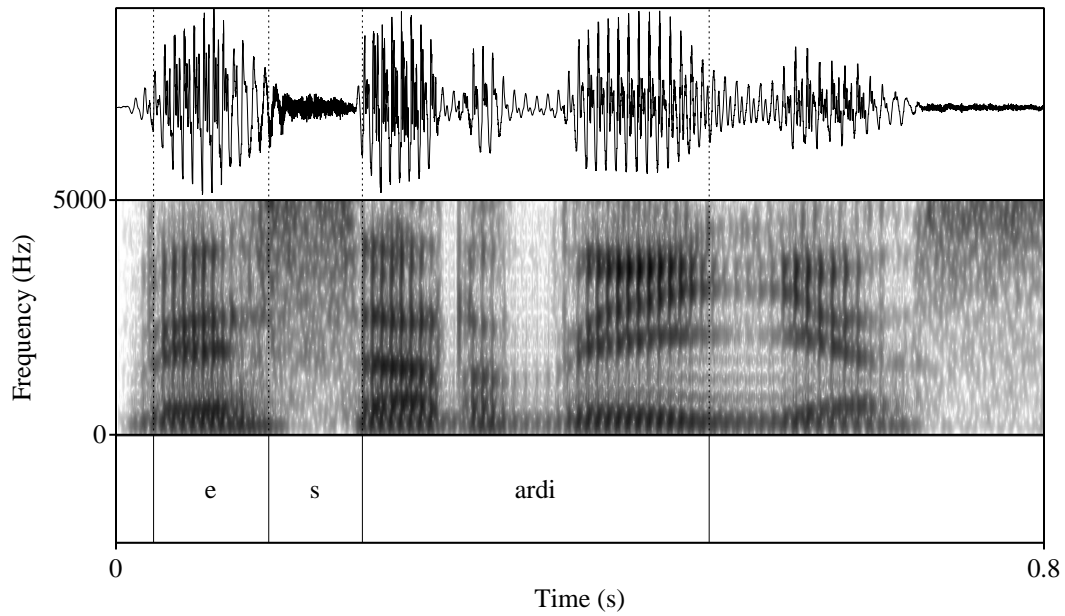
Figure 4.4: *ves ardillas* 'you see squirrels'. Duration of /s/ is 80.74 ms.
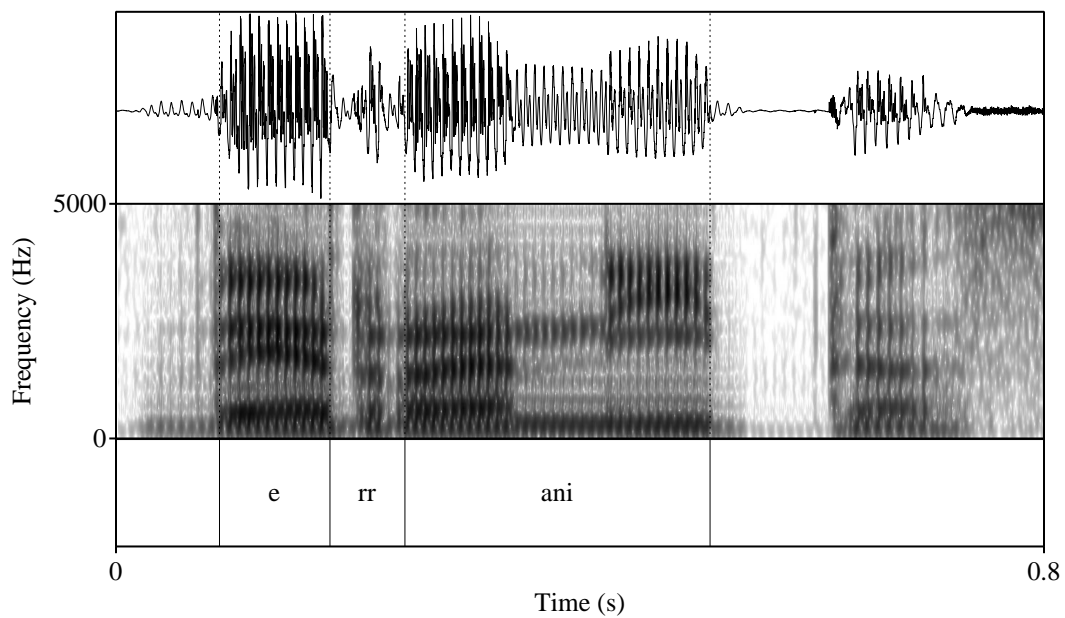
Figure 4.5: *ve ranitas* 's/he sees little frogs'. Duration of [r] is 64.63 ms.
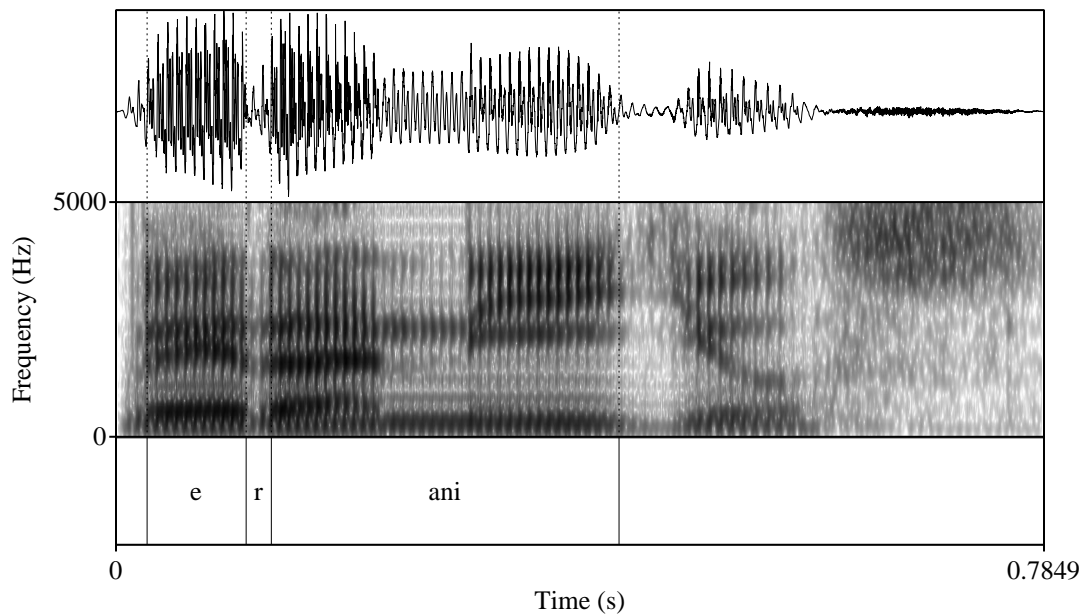
Figure 4.6: *ver anillos* 'to see rings'. Duration of [ɾ] is 21.31 ms.

4.3.3 Procedure

All testing took place in the Second Language and Bilingualism Lab at the University of Illinois, Urbana-Champaign. Participants were first trained on the 96 items in two phases: exposure and testing. In the exposure phase, the participant saw the image with its name in orthography by itself on the screen. S/he told the experimenter to continue to the next image when they were satisfied with the name of the object. Although most of the objects were well known to the participants, some objects were paired with less common names in Spanish, e.g. *aretes* for 'earrings' instead of *pendientes*. Each of the 96 images were shown in this way in random order. In the testing phase, the subject again saw the image, but was told to produce the name of it aloud. If the participant made any mistake, the experimenter corrected the participant verbally and noted which item the mistake was made with. Non-phonological mistakes were not corrected: for example, if the participant did not produce the proper VOT on voiceless stops or did not spirantize intervocalic voiced stops. Crucially, the experimenter did not correct any production of /r/ or /ɾ/. After the initial testing phase, the participant saw only the items that s/he had missed previously. This procedure continued until the participant had named all of the 96 items correctly. The participant then performed the testing phase for a second time, with the images in a different random order. This training portion took between 10 and 20 minutes.

The eye-tracking experiment itself was implemented through the Experiment Builder software provided by SR Research. The eye-movements were recorded with the desktop-mounted SR Research EyeLink 1000 with a sample rate of 1000 Hz. All participants began the experimental session by calibrating the software to follow the movements of their right eye. This was followed by a short practice session of four trials and then the main experiment of 48 trials. In each trial, the participants saw four images on the screen in a 2 x 2 grid for 2 seconds. An example is given in Figure 4.7 The images then disappeared, and a fixation cross appeared in the center of the screen. Participants needed to fixate in the general region of the cross for 500 milliseconds before it disappeared. Once the cross disappeared, the four images reappeared on the screen, and the auditory stimulus was heard through the headphones. The participants were told to click on the target image as soon as they heard the
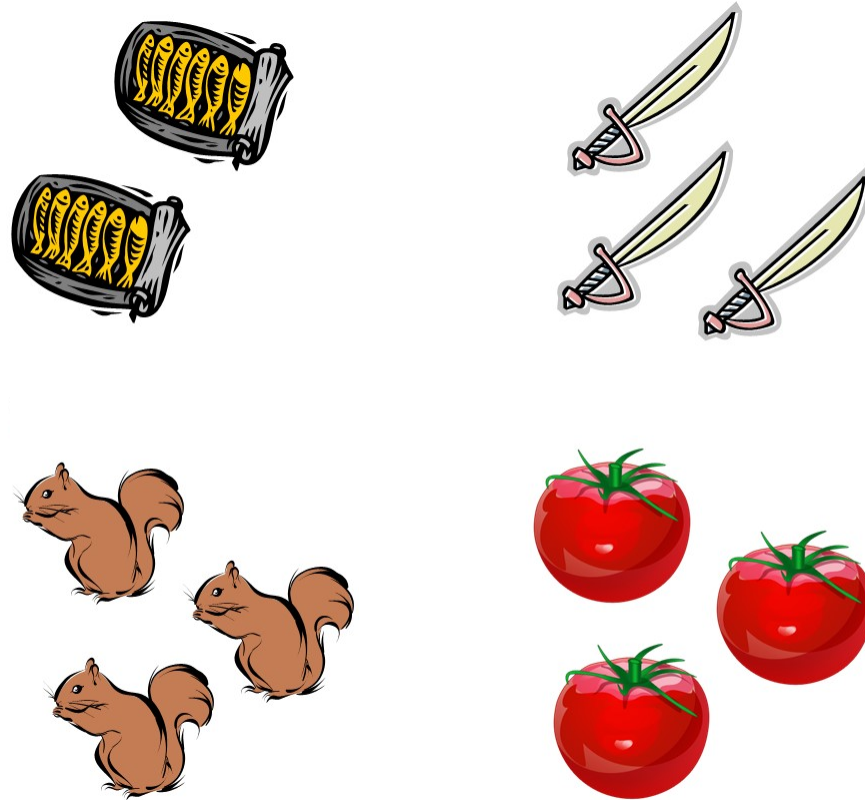
Figure 4.7: Sample display in eye-tracking experiment

target word in the two-word phrase. The participants had been told that they would hear a two-word phrase, the first of which would be some form of the verb *ver* 'to see', including possibly the infinitive itself. The eye movements were measured starting from the offset of the pivotal consonant (/n/, /s/, or /r/). The trial then ended, with a 1000 millisecond pause between trials. The 48 trials were presented in an entirely randomized order. Participants completed the experiment in roughly 10 minutes.

4.3.4   Analysis

The eye-movement data was collected with a data point (the x and y positions of the eye gaze) at every 1 ms (1000 Hz). Each of these data points was associated with a region on the screen corresponding to one of the four objects: the target, the competitor, and either of the two distractors.

One of the contributions of this chapter is to do a comparison of eye-tracking analysis techniques, using traditional methods, a new in-development packing for R entitled *eyetrackingR*, which includes a Growth Curve Analysis, and my own methodological contribution to the field charactizing fixation patterns on a trial-by-trial basis. First, I will motivate the necessity of such an endeavor. The details of the different methodologies will be expounded upon in the following sections.

In many eye-tracking studies Allopenna et al. (1998); Tremblay (2011); Tremblay and Spinelli (2013, 2014), the eye-gaze data is aggregated and averaged across time windows in each condition and the results show the proportion of looks to each of the four objects on the screen during each window. In others, proportions of fixations are measured at or relative to a particular time point McMurray et al. (2009). For the present study, this type of analysis is unsatisfactory, because the time course of eye movements is different for each individual trial. Averaging across trials, when participants may not always begin fixations right away, will obscure the actual patterns in each trial. As an example, a participant may quickly look at the target image, and then return to the fixation cross. This would indicate early activation of the target word. On another trial, a participant may look at the competitor image, but then mistakenly look at an unrelated image in the search for the target image,

before finally settling on the target. Under previous annotation schema, this would indicate activation of both the competitor item and the unrelated item. Finally, on another trial, a participant may simply wait longer to look at the target word, but looks at it without passing through any other regions. Both of these final two descriptions of trials will give the same result: 'late activation' of the target image, even though the patterns of eye movement are quite different. A more significant problem with averaging across different items comes from the fact that the crucial time points in the signal (onset and offset of pivotal consonant, disambiguation point) are not the same in each stimulus. Because the crucial consonant can be quite short, finding an 'average' onset point in order to create a time window of analysis is not accurate to the data. Additionally, participants may look at other objects after they have 'found' the target object – simply eliminating data after they have looked at the target image for the first time will obscure any competition effects – which is not relevant to the research questions. Finally, each of the studies above exhibit signficant downsampling of the eye-gaze data, usually at windows of 50 ms. The EyeLink 1000, which was used in this study, samples at a rate of 1000 Hz; downsampling at the aforementioned rate would effectively be sampling at 20 Hz. Because the consonants are so short and eye-movements are quick, this does not seem to be an effective use of the data.
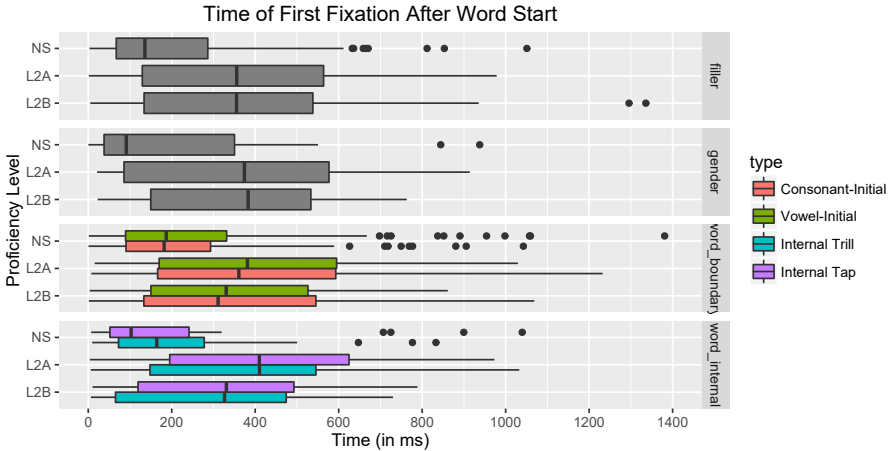
Figure 4.8: Time from the onset of the pivotal consonant to the first time a participant's eye gaze reached any region of interest, sorted by the three proficiency levels and word position. Data includes all pivotal consonants and all conditions.

In Figure 4.8, the time to fixate on any image on the screen is plotted, separated by

proficiency level. All of the conditions present in the experiment are shown, including filler items. These times identify the time at which the first region of interest that the participants' eye gaze enters after the onset of the target consonant. As mentioned above, this should represent the word that is being activated first, and this should happen early, roughly 200 ms after the onset of the acoustic material. The times represented do not indicate the time at which the correct (target) word was fixated at, but instead any image, which could have been one of the distractors. However, participants were free to look at the images *before* the onset of the word (or pivotal consonant): recall that the images were presented onscreen during the entirely of the auditory stimulus. It is impossible to separate out early fixations of this type. This methodology lessens the issue by only looking at which image was fixated at after the beginning of the word (or in the case of the word boundary condition, the target consonant, which could be word-initial or word-final). In some cases this time was immediately, as an image was already fixated at during the production; see the following section for more discussion on fixation patterns during trials. Thus, the data in Figure 4.8 is a rough index of speed in each condition by each participant group.

The motivation to examine first fixation time is to determine whether or not the time analyses reflect difficulty in processing the (semi)ambiguous phrases or simply reflect processing times inherent in the participant groups. One possible explanation for this is the relative difficulty of processing these temporarily ambiguous phrases. Participants may realize that it is not worth the effort to make an incorrect guess as most trials had some kind of ambiguity involved, reminiscent of the slower reaction time found by French speakers of liaison consonants (Spinelli et al., 2003), incurring a processing cost. However, in the syntactic processing literature, participants do not spend more time reading ambiguous sentences than unambiguous ones, excepting of course the time it takes to re-read garden path sentences (Clifton and Staub, 2008). Thus, if the learner groups are slower than native speakers to fixate on any image in the experimental conditions, it could be that the learner participants are slower in general to process their second language, or that they are doing something qualitatively differently; realizing the possible ambiguity in resyllabified contexts, they may be able to suppress word activation until they have heard more acoustic information. Instead, the data in Figure 4.8 suggests that participants performed identically in 'difficult' trials,

where there was ambiguity, and filler trials, both where there were no competitor items on the screen ('filler') and where there were competitor items, but unrelated to the experiment ('gender').

As native speakers performed the fastest in all conditions, all further results examining time differences between participant groups need to be considered carefully. The learner groups are slower because it is harder to process any trials in their second language; this is not an effect of the trials. It was suggested that a time-normalization algorithm be performed on the data in order to minimize these inherent differences between the native and learner groups. However, this is both impractical and unnecessary. There are several methodologies for time normalization; I will discuss time shifting and time scaling. In time-shifting normalization, the filler trials could be set as the 'baseline', and the first fixation times the filler trials could be subtracted from those in the experimental trials to create a 'difference' score. The main issue with this methodology is that it shifts only the time of the beginning of the trial, but the entire length of the trial has not been normalized. Learner trials tend to be longer overall, not just delayed at the beginning. Time scaling seeks to fix this problem by normalizing the duration of each trial so that the beginning and ending time points are identical for each trial. However, this will change the overall shape of the time course curves and distort them relative to each other proficiency group. For example, if an overall curve is compressed in time, it may appear that this group is faster to switch from one image to another. Both methodologies have a more severe drawback: by normalizing time, it becomes difficult to align the eye gaze data to the acoustic information being heard during the trials. Even if it is the case that the learner groups are slower to process words in their second language, it is *also* the case that by the time they do make their eye movements toward a particular image on the screen, they have heard more of the acoustic signal. In fact, while native speakers make their first fixations at an image earlier than do the learners, they do not make their first fixation at the target image consistently earlier than the learner groups. This can be seen in Figure 4.9. Learners may be making more usage of the acoustic signal as they have heard more of it, and are quicker to look at the correct item than native speakers. Thus, normalizing time would obscure one of the more informative aspects of the eye tracking methodology: the ability to align the acoustic signal with the eye gaze data.

81

However, the fact that some participants and participant groups *do not* immediately fixate at any image 200 ms after the onset of the pivotal consonant, regardless of reason, remains a challenge for the data analysis of eye tracking. The multiple methodologies for analyzing the data in this chapter seek to address these concerns.
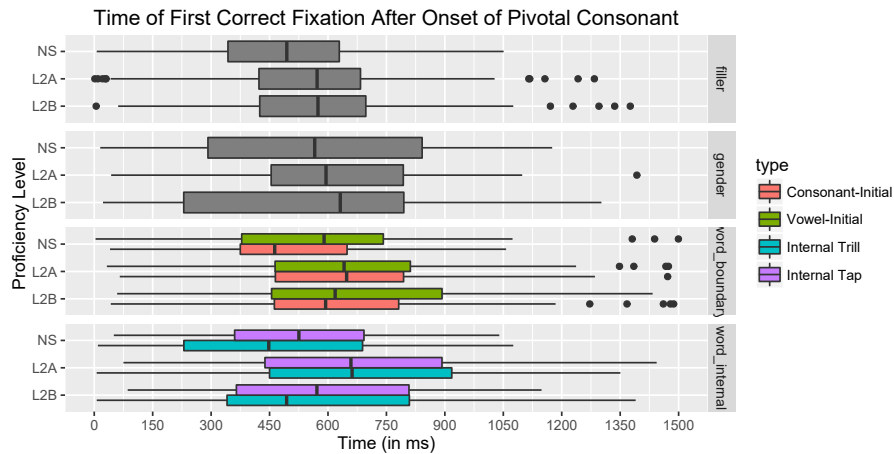
Figure 4.9: Time from the onset of the pivotal consonant to the first time a participant's eye gaze reached the correct region of interest, sorted by the three proficiency levels and word position. Data includes all pivotal consonants and all conditions.

Upon examination of each individual trial, a new analysis schema was developed, which can be used to better analyze descriptive eye-tracking data. Both the pattern of eye-movement (i.e. which images were fixated at and in which order) and the time relative to the important time points in the aural stimulus are measured. Although a video of the time course of eye movements can be created in the DataBuilder software provided with the EyeLink 1000 machine, a custom-written script producing a static version of this video was written to better examine the eye movement patterns. In these images, the trace of the eye movements is shown by a series of points, plotted on an X and Y axis corresponded to the gaze on the computer screen. The hue of the color of the point corresponds to the time during the trial on a gradient scale (beginning with red, decreasing in wavelength to purple as the trial progresses). Superimposed on the eye trace are blue, black, and red dots corresponding to the location of the eye fixation at the onset, offset, and disambiguation point in the aural stimulus, respectively. Because eye movements are delayed 200 ms following an incoming auditory stimulus, the time course of the eye movements is delayed by 200 ms. Thus, the

placement of the red dot is actually 200 ms following the disambiguation point in the audio. Additionally, the images of the target, competitor, and distractor words are plotted within the marked regions of interest, as they were in the actual trials. However, each of these images has a color overlay that was not present during the actual trials; these colors indicate target ('red'), competitor ('green'), and two different unrelated distractors ('blue' and 'purple'). These colors match the colors used for these categories elsewhere in the chapter. Note that the positions of each of these categories shifted in each trial. Similarly, each region is annotated in these images with text indicating the Spanish word of the object; no text was present in the actual trial.

Finally, a small icon of a computer mouse appears at the moment that the EyeLink host computer registered a mouse click. This is not necessarily the position of the mouse cursor at the time of the click but rather the spot in which the eyes were fixated at at the moment when the mouse was clicked and is a good approximation of the location of the cursor. Recall that the instructions for each trial was to click on the image of the word heard. We can consider that this is the time when the participant considered the trial 'over' – in many cases the participants' eyes wander considerably after this point before the actual trial cuts off, either to other unrelated images or the center of the screen. This is important to consider when comparing the methodologies for analyzing the eye-tracking data.

In Figure 4.10, the regions of interest for each of the four objects are shown: the target image is in the bottom right-hand corner (*obras*, 'works'), while the competitor is in the upper right-hand corner (*sobres*, 'envelopes'). At the beginning of the trial, the eye fixation is in the center of the screen, at the fixation cross. The cross then disappears, the images appear, and the auditory stimulus, *ves obras* 'you see works' is heard. Immediately, the participant (a native speaker of Spanish) looks towards *monos* 'monkeys', but at the beginning of the /s/ sound (the blue dot), the participant directs her gaze towards the /s/-initial word, *sobres*. The participant lingers on *sobres* after the offset of the /s/, but at the beginning of the disambiguating information between *sobres* and *obras*; that is, the vowel following the /r/, the participant shifts her gaze towards the target, *obras*.

In this example trace, it is clear that simply looking at the regions of fixation between the relevant time points (the onset of the pivotal consonant and some time after the disambigua-
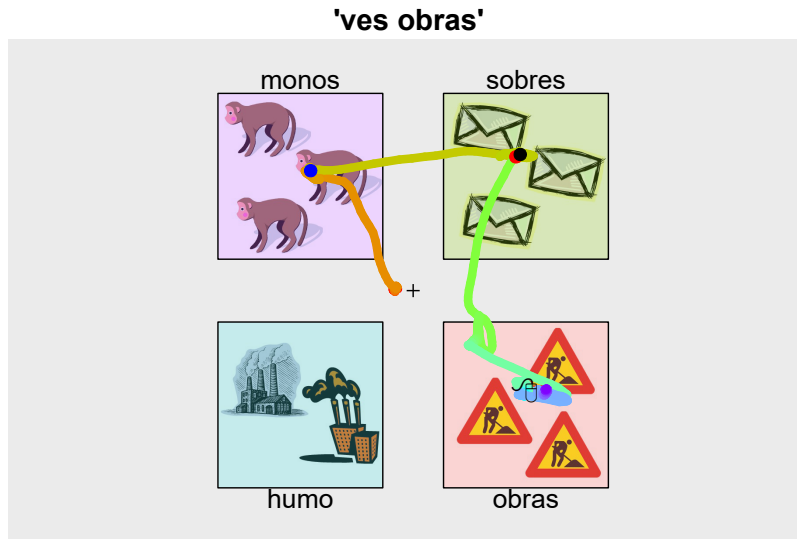
'ves obras'

Figure 4.10: Native speaker example eye-movement trace of *ves obras* 'you see works', a vowel-initial word in the resyllabification context. Details of the analysis are listed in the text.

tion point) will give misleading data. After the onset of the pivotal consonant there is not an immediate shift in fixation in a particular region of interest. The participants eyes must first leave the erroneous distractor item region ('*monos*'), where there eye-tracker picks up many frames, go between the two regions in the top portion of the screen (counted in the data as 'no fixation'), and then finally land in the region of the competitor item. It then takes a large number of frames to exit this region and enter the region of the target item. The data analysis that measures fixations to specific regions during periods of time discounts that this crucially relies on where the eye gazes were *before* the relevant cue in the acoustic signal. This token would be classified as 'Competitor -> Target'

The pattern in this example is clear: we want to characterize this behavior simply as looking at the competitor image before the target image. Thus, one measure in this study simply looks at the relative ordering at fixations between the target and competitor words. In the following example, Figure 4.11, the participant looks first at the target word, 'coristas', but then looks at 'corrilos', the competitor, briefly before looking back to the target word. This token would be classified as 'Target -> Competitor -> Target'. Other classifications included 'Competitor -> Target', and 'Only Target'. Any looks to the competitor indicates some knowledge of ambiguity in the phonological processor. More discussion of this coding

'ven coristas'

salones   naranjas

coristas   corrillos
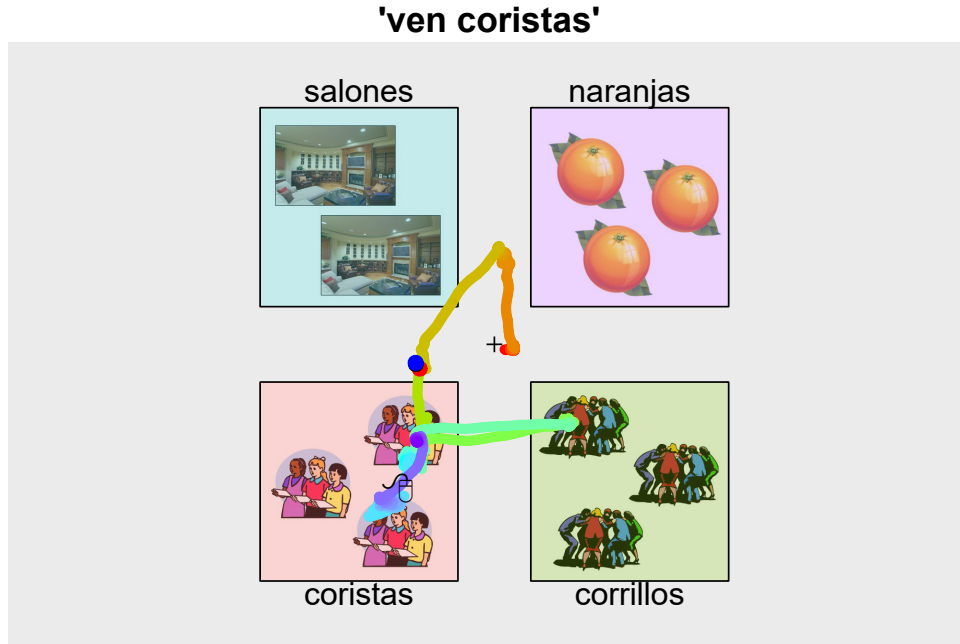
Figure 4.11: Native speaker example eye-movement trace

scheme and the analysis is detailed below with the results.

Finally, in Figure 4.12, we see the importance analyzing these images on a trial-by-trial basis and syncing them with the mouse-click information. On the surface, in this trial, there are looks to all four images–both the target (*aves*) and the competitor (*rabos*), as well as the two distractors. But clearly the looks to the two distractors are incidental–the look to the *sobres* happens before the pivotal consonant onset–the black and blue dots–and must be taken as a premature random eye-gaze and not as activation of the word *sobres*, and the looks to the *peras* occur only because the image is simply on the way from the previous fixation to the region with the target *aves* and do not reflect activation of this distractor word either. Finally, after the disambiguation period has ended (the red dot), the participant clicks the mouse button, still with eyes fixated in the target region. Only afterwards does the participant look at the competitor image (*rabos*). Why the participant does so is not clear: it is possible that the participant now considers the competitor word a possibility as it shares many phonemes in common with the word heard, but on the other hand, the participant subsequently looks back at another distractor image, demonstrating that perhaps the participant simply wanted look at the other images for no reason other
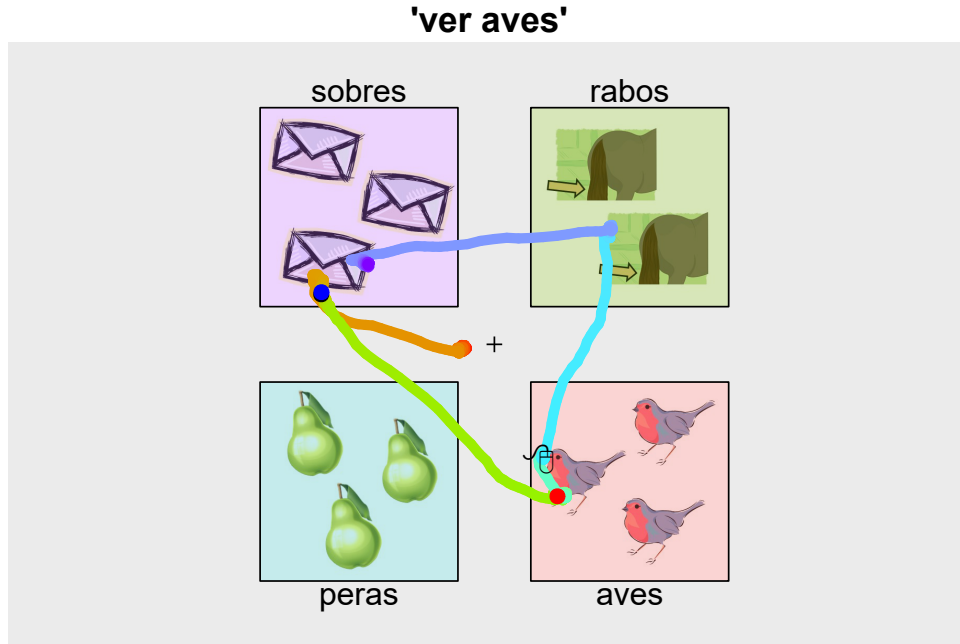
Figure 4.12: L2 Advanced example eye-movement trace

than that the trial was about to end. We cannot tell whether or not there is actual word activation after the mouse-click has been made, so we disregard all eye-fixations after the mouse-click occurs. Finally, we note another criticism of using the time-course of eye-fixation in traditional analyses (where data has been binned together and averaged): the time it takes to fixate at the target image in this trial has been delayed simply because of the amount of time (and distance) it takes to get from an incorrect fixation to the correct image; here, from *sobres* to *aves*. This time may have been much shorter had the participant looked at a different image or even the correct one earlier in the trial. In the traditional analysis of eye-tracking data, a delay in correct-item fixation is due to slower activation of the target word, potentially even activation of the incorrect word, but it is clear that this is not always the case. For this trial, the classification of 'Only Target' was chosen, reflecting the fact that the competitor item was not fixated at until much later in the trial, and that looks to the distractor items were incidental.

4.4 Results

The participants performed close to ceiling, at 98.2% accuracy, at clicking on the target word. All trials where the target-word was not clicked were excluded from further analysis (36 trials, 1.98%). The results for the eye movements are summarized in the following figures. The black vertical line in the figures represents the average point in the auditory stimuli where the target consonant ends; that is, 200 ms after this line we should start to see looks to either the target or the competitor.

4.4.1 Traditional analysis

As described above, in the traditional analysis used for eye-tracking data, each trial within a given condition is aggregated together. This is done as follows: First, the eye-gaze data started at the timepoint when the soundfile began playing (time = 0). Then, time was transformed for each trial, aligned with the *offset* of the pivotal consonant, becoming the new zero time. This normalizes any differences in length (time) before the hypothesized first movement of the eyes; in other words, if the /ben/ portion of *ve novelas* and *ve navajas* were for any reason different, this normalization technique will take this into account. Next, each sample is put into a 5 ms bin, and is coded for which interest area of the screen the eye was fixated at. At this point all trials are aggregated together, but within the same condition. A proportion of fixations to each of the interest areas is then calculated. For example, between time = 200 (200 ms after the offset of the pivotal consonant) and time = 205 (bin #40), if for all native speakers in the /n/-initial condition (e.g. *ve novelas*), there were 50 looks to the target item, 30 looks to the competitor item, 5 looks to each of the distractors and 10 fixations not in any region of interest, the proportions of fixations would be 45%, 27%, 4.5%, and 4.5% respectively and show up as such on the fixation proportion graphs in this section. Looks that are not at any region of interest are not shown on the graphs, but are still used in the calculations of the proportions.

For analysis, we are interested in the *target advantage* score; that is, the difference between the proportion of looks at the target item and looks at the competitor item within a particular window of time. The window chosen was between the dismabiguation point (the red vertical

line on the graphs) and 400 ms later. This score was calculated for each item.

Descriptive results

For /n/, for native speakers, there are more looks earlier to the target in the word-initial condition (*ve novelas)* than in the word-final condition (*ven ovejas*), where there is more competition between *novelas* and *ovejas*, suggesting that while native speakers are sensitive to resyllabification, there is still a preference for consonant initial words. We see the same pattern for L2 speakers, although even more pronounced, as L2 speakers begin looking at /n/-initial words even earlier.

For /s/, for native speakers, we see the same pattern as for /n/: more early looks to the consonant-initial stimulus. This is even more a striking case in the /s/-final condition, where looks to /s/-initial words predominate and persist longer than for /n/. For L2 speakers, participants tend to be more accurate, there are fewer looks to the competitors, although the divergence point between looks to the target and the competitor and distractors seems to happen much later than the other conditions.

For /r/, for native speakers, participants are quick and accurate in their eye movements, suggesting that the distribution of [r] and [ɾ] is used to disambiguate these phrases. This is especially the case for /r/-initial words, where the presence of the trill is a strong cue that the word is /r/-initial. Disambiguation of /r/-final words happens later, suggesting that these speakers entertain the possibility that a tap is a possible word-initial segment. For L2 speakers, the pattern is similar, although the divergence of eye movements happens much later, suggesting that although they are sensitive to this distribution, it is not yet reliable for these speakers.

Figure 4.13 shows a summary[1] of the fixations to word-boundary targets and competitors, divided by type (consonant or vowel initial targets), consonant (the pivotal consonant was an /r/ or an /n/ ~ /s/), and proficiency level.[2] Only the portion of the trial following the

---

[1]The raw data showing complete trial-long fixations to all four images in all conditions can be found in a series of figures at the end of this chapter in Section 4.6. These graphs also include the average consonant onset, offset, and disambiguation point, as well as the range of times when the trial ended; that is, when the participant clicked the mouse button.

[2]These graphs were created through *eyetrackingR* (Dink and Ferguson, 2016)(to be discussed further); there is thus some discrepancy between these graphs and the others of my own creation found in the Target

offset of the pivotal consonant is shown. A brief descriptive analysis follows. We are most interested in seeing when it is the fixations to the target and competitor diverge: is it before the overall disambiguating information, or after? For example if the participant hears *ve novelas*, and views *novelas* and *ovejas* as two words in competition, will they look at the target *novelas* before

Figure 4.14 shows a summary of the fixations to word-internal targets and competitors, divided by stress and whether the auditory stimulus included a tap or a trill. Unlike the word-boundary condition, the onset of the consonant occurs word-medially, and thus many of the fixations are already located on the target or competitor image by the beginning of the consonant (the vertical blue line). This is especially the case when the first syllable is stressed: for both native speakers and advanced L2 speakers, their fixations between targets and competitors have already diverged by the onset of the tap or trill. For beginners this is only the case for the condition where the first syllable is stressed and the target consonant is a tap. It is possible that there is disambiguating acoustic information earlier in the word; for example, on the stressed vowel previous to the tap or trill. In this stress condition, the otherwise disambiguating information is directly after the tap or trill: the quality of the following vowel is different. Thus, minimally, the fixations between target and onset should diverge after the offset of the consonant (vertical black line). This occurs for all speakers except for the beginning learners in the condition where the target is a trill. This suggests that these participants accept a trill, /r/ for a phonemic /ɾ/ and have not learned the distinction between the two even word internally, with a bias towards hearing a trill as a possible realization for the tap, but not vice versa.

For the stress condition where the second syllable of the word is stressed, we see a number of different patterns. Again, here the disambiguating acoustic information occurs further downstream of the word (at the position of the red vertical line). It is at this point that participants should switch their eye gaze towards the target image.

Advantage graphs and the raw graphs at the end of the chapter in how the proportions of fixations are calculated in the early parts of the trial when some fixations are not on any particular area of interest.
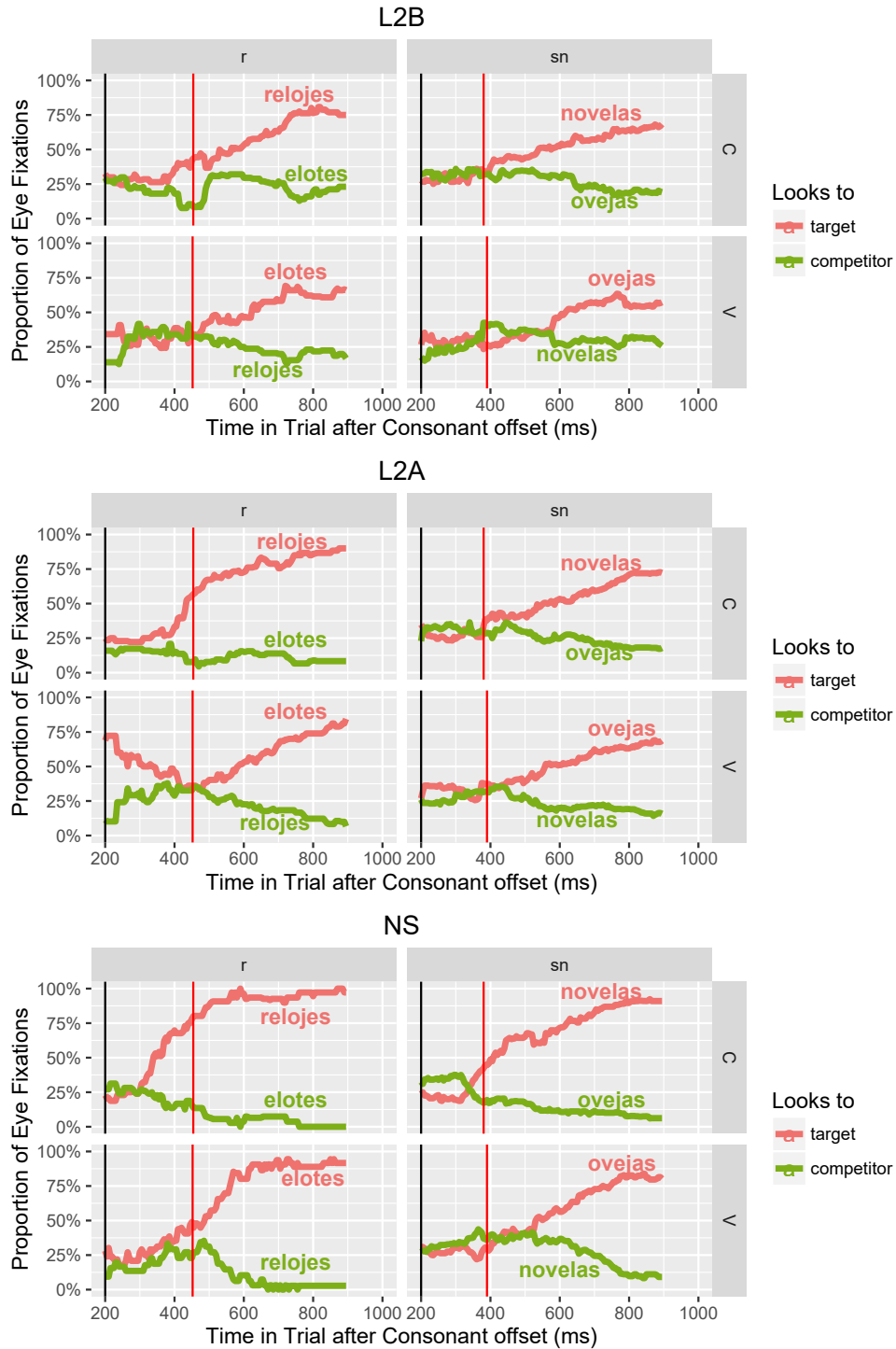
Figure 4.13: Summary of proportion of fixations to target and competitor words across time in the word boundary condition. The pivotal consonants are labeled. Example words for each condition are labeled. Time is only shown after the offset of the pivotal consonant. The red vertical line shows the average disambiguation point after the pivotal consonant.
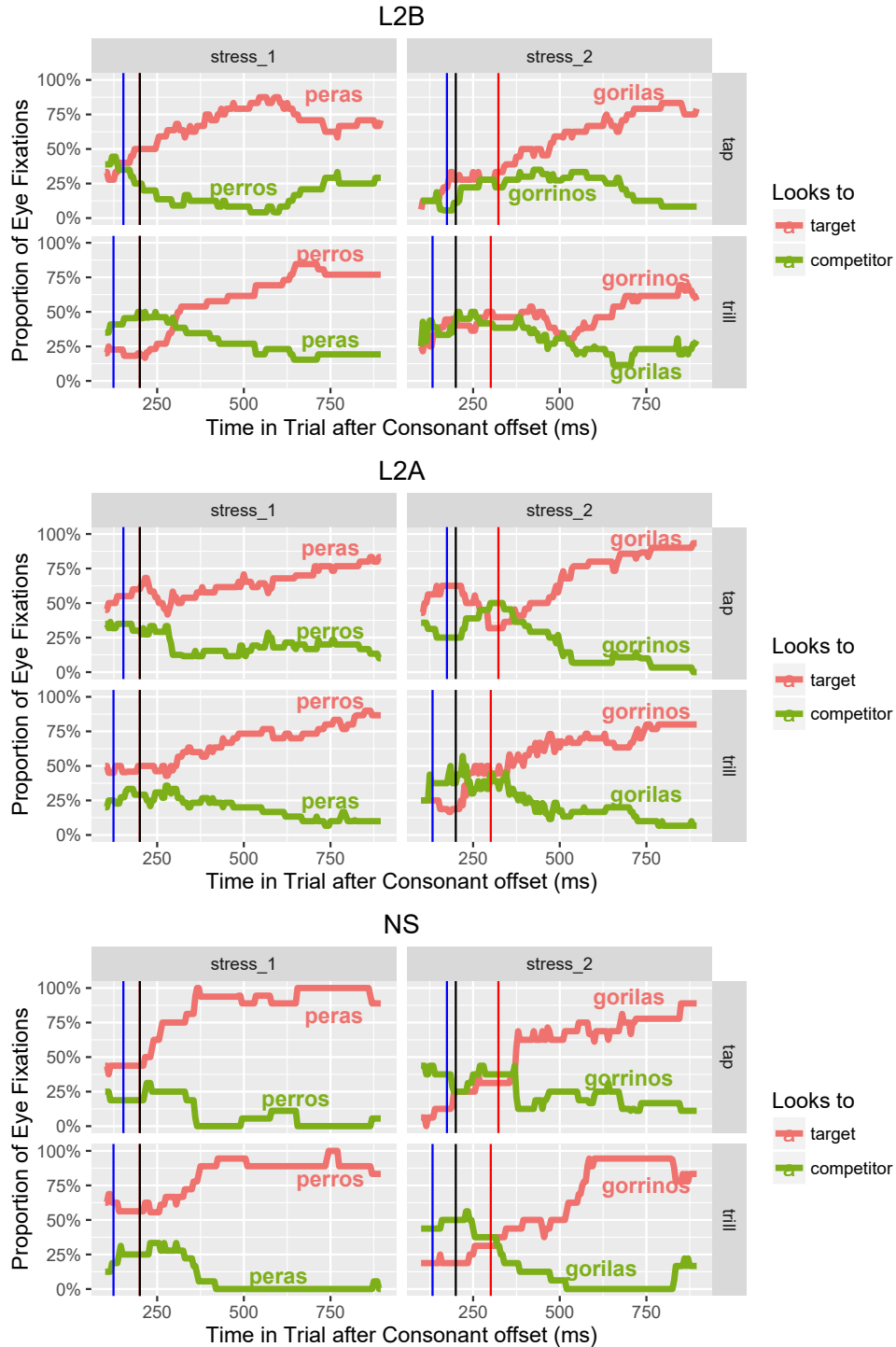
Figure 4.14: Summary of proportion of fixations to target and competitor words across time in the word internal condition, divided by the stress condition and which rhotic was heard. Example words for each condition are labeled. For the condition where the first syllable of the word was stressed ('stress 1'), the average onset of the rhotic is shown with the blue vertical line; its offset with the black line. For the condition where the second syllable was stressed, the red vertical line shows the average disambiguation point after the rhotic.

Statistical results

One traditional way to report statistics on the time course of eye fixations in eye tracking experiments is to divide the time course into particular critical windows, and devise a Target Advantage score for each condition (Tremblay and Spinelli, 2013). The Target Advantage is simply the difference between the proportion of fixations to the target image and the proportion of fixations to the competitor image. A positive number indicates that during a particular window there were more fixations to the target than to the competitor. The higher this number is (approaching 1.0), the less the participant is looking at the competitor image. In Figures 4.15 through 4.19, the Target Advantage curve is shown (a transformation from the graphs just shown); however, these graphs are just to illustrate the data; the actual input to the statistical model was simply the average during each particular window.

In all conditions except the word internal condition where stress fell on the initial syllable, four windows of analysis were chosen. These are labeled in the figures as 'a', 'b', 'c', and 'd', all of which are centered around the average disambiguation point in the trial, represented by the red vertical line. The duration from the beginning of window 'a' to the end of window 'b' was the duration of the time between the offset of the pivotal consonant and the disambiguation point. The windows 'c' and 'd' covered 400 ms after the disambiguation point. There was no theoretical basis for this 400 ms figure except that it appear that most fixations to the target words seemed to occur by this point. The division between 'a' and 'b' was exactly the midpoint; the same went for 'c' and 'd'. Thus, the duration of windows 'a' and 'b' were both equal, and the durations of both windows 'c' and 'd' were both 200 ms. For the word-internal condition where stress fell on the initial syllable (e.g. *veo perros*, the disambiguation point coincided with the consonantal offset. Thus, as shown in Figure 4.19, there were no 'a' and 'b' windows of analysis. For consistency, even though there were only two windows, they are referred to as windows 'c' and 'd'.

The Target Advantage scores in each window[3] were submitted to separate linear mixed effects models using the *lme4* package in R. For the word-boundary models, consonant (/r/, /s/, or /n/), type (consonant or vowel initial word), and proficiency (native speaker, L2

---

[3] A single LMEM including window as a factor interacting with all of the other variables was too large to be interpretable.

beginners, L2 advanced) were submitted as fixed effects, while participant and trial were submitted as random effects. The results are found in Tables 4.3 and Table 4.4. It is important to note that the results from these statistical tests only tell us during which window the Target Advantage scores are different from each other; it does not, for example tell us when exactly eye fixations began on either the target or competitor images. None of the factors were significant in windows 'a' and 'b', the time before the as this must be too early for the time course fixations to have diverged from each other. During window 'c', there are significant differences between 'type', between all three consonants, and the native speaker from the advanced L2 speaker groups. There is also an interaction between how the native speaker and L2 speaker groups behaved in the consonant vs. vowel initial conditions: the native speakers had a significantly higher Target Advantage score in the consonant-initial condition than the vowel-initial condition (overall L2 speakers showed no difference between the conditions). In window 'd', we again find significant differences between 'type', between all three consonants, and the native speaker from the L2 speaker groups. There is now additionally an interaction between how the native speaker and L2 speaker groups behaved in the /r/ vs. /s & n/ conditions: for native speakers, the Target Advantage scores are much higher in both the /n & s/ and /r/ conditions, whereas for L2 speakers, they are much lower in the /n & s/ condition. These results suggest that overall there is a significant difference between the L2 and native speaker groups (but no difference between the two L2 speaker groups), and that L2 speaker groups are able to make some usage of the phonological cues for /r/ that indicate placement of the word boundary (although descriptively it seems that only the advanced L2 speakers have a higher Target Advantage score in the trill-initial condition). For all groups, there was a higher Target Advantage score for the consonant-initial condition than the vowel-initial condition, indicating that all speakers use some phonological universals biasing syllable and word boundary coincidence, although it is interesting to note that there was *negative* Target Advantage score for nearly all groups during the first two windows of analysis. The significant differences between any of the conditions only arose after the disambiguation point had passed, contrary to expectations. However, this could be due to poor alignment of the acoustic signal to the aggregated trial analysis, or that these trials were too difficult to show divergences between conditions early in the signal. This issue will
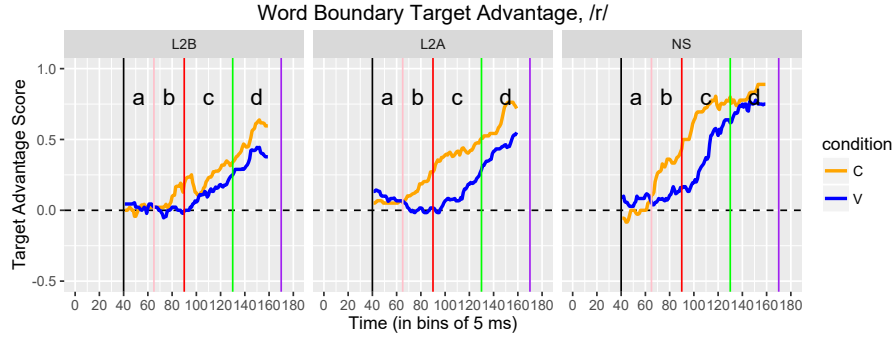
Figure 4.15: Target Advantage results for /r/ for all speakers. The x-axis represents bins of 5 ms, the y-axis represents proportions of fixations to the competitor subtracted from proportions of fixations to the target (target advantage) for trials where the target item was consonant initial (orange) and vowel initial (blue). The colored vertical lines show the four different windows used in the statistical analysis; for reference, the average offset of the target consonant is in black; the average disambiguation point is in red.

be taken up further in the discussion.

For the word internal-condition, only two fixed effects were included: consonant (tap or trill) and proficiency. Because the windows were inherently different, stress could not be used as a factor. For the condition where stress was placed on the second syllable of the word (e.g. *ver gorilas*, see Table 4.5. For the condition where stress was placed on the first syllable of the word (e.g. *ver perros*, see Table 4.6. Again, no significant differences were found until windows 'c' and 'd'. In both of these windows for both of the stress conditions, the only significant difference was between the two L2 speaker groups and the native speaker group. Target Advantage scores were significantly higher for both the tap and trill conditions compared to the L2 speaker groups. Importantly, no differences were found between the tap and trill for any speaker groups in any window of analysis. Therefore, the asymmetries found in the word-boundary conditions regarding /r/ are due to the phonological effects of the tap/trill contrast at the word boundary, and not due to difficulties in perceiving the difference between these segments in general (although, again, the L2 speaker groups had *lower* Target Advantage scores for both the tap and the trill word internally, and their fixations overall were very low compared to native speakers).
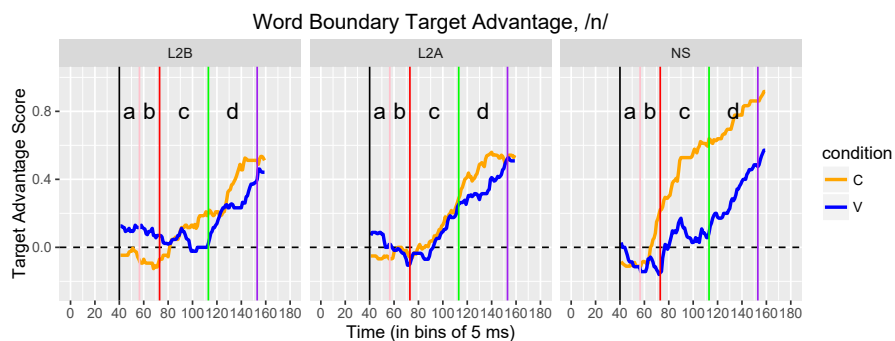
Figure 4.16: Target Advantage results for /n/ for all speakers. The x-axis represents bins of 5 ms, the y-axis represents proportions of fixations to the competitor subtracted from proportions of fixations to the target (target advantage) for trials where the target item was consonant initial (orange) and vowel initial (blue). The colored vertical lines show the four different windows used in the statistical analysis; for reference, the average offset of the target consonant is in black; the average disambiguation point is in red.
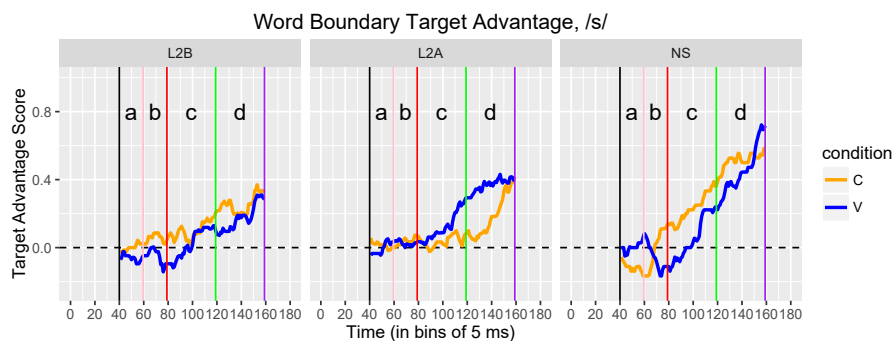
Figure 4.17: Target Advantage results for /s/ for all speakers. The x-axis represents bins of 5 ms, the y-axis represents proportions of fixations to the competitor subtracted from proportions of fixations to the target (target advantage) for trials where the target item was consonant initial (orange) and vowel initial (blue). The colored vertical lines show the four different windows used in the statistical analysis; for reference, the average offset of the target consonant is in black; the average disambiguation point is in red.
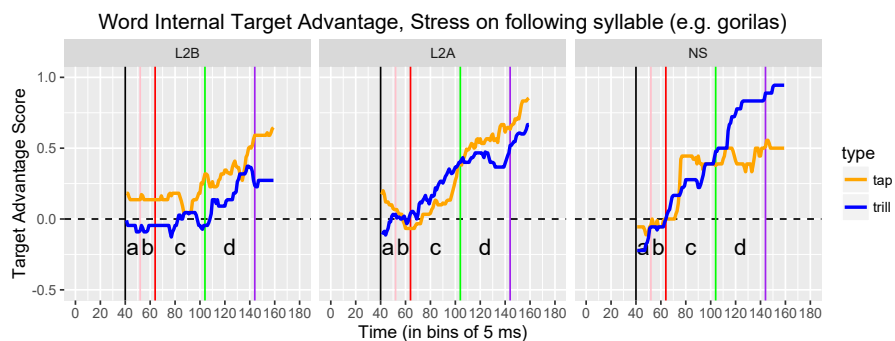
Figure 4.18: Target Advantage results for word-internal /r/ for all speakers. The x-axis represents bins of 5 ms, the y-axis represents proportions of fixations to the competitor subtracted from proportions of fixations to the target (target advantage) for trials where the target item contained a tap (orange) and contained a trill (blue). The colored vertical lines show the four different windows used in the statistical analysis; for reference, the average offset of the target consonant is in black; the average disambiguation point is in red.
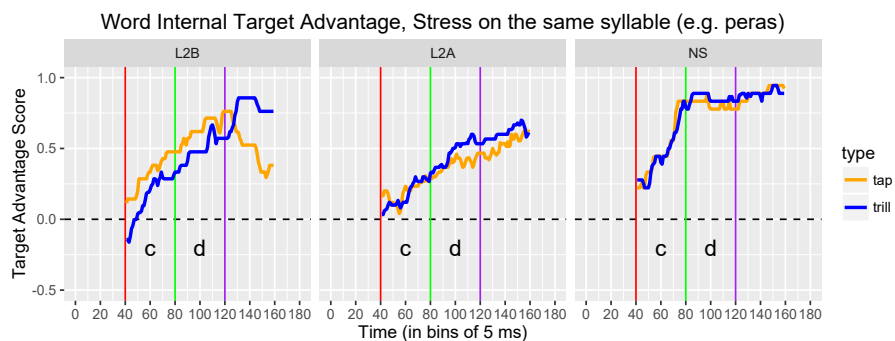
Figure 4.19: Target Advantage results for word-internal /r/ for all speakers. The x-axis represents bins of 5 ms, the y-axis represents proportions of fixations to the competitor subtracted from proportions of fixations to the target (target advantage) for trials where the target item contained a tap (orange) and contained a trill (blue). The colored vertical lines show the two different windows used in the statistical analysis; for reference, the average offset of the target consonant is in black; the average disambiguation point is in red.

Table 4.3: Linear mixed-effects models for word boundary condition on all participants' target advantage scores in windows of analysis 'a' and 'b'. Three-way interactions are not included as none of them reached t > 2.

	Variable	Estimate (SE)	t value
Window (a)	(Intercept)	0.0829 (0.031)	2.664
	type (C/V)	0.0282 (0.031)	0.906
	onset1 (n vs s)	0.0140 (0.038)	0.369
	onset2 (n & s vs r)	-0.0038 (0.022)	-0.177
	proficiency1 (L2A vs L2B)	0.0200 (0.024)	0.831
	proficiency2 (L2 vs NS)	-0.0086 (0.013)	-0.642
	type*onset1	-0.0542 (0.038)	-1.425
	type*onset2	-0.0134 (0.022)	-0.610
	type*proficiency1	0.0050 (0.024)	0.208
	type*proficiency2	0.0031 (0.013)	0.231
	onset1*proficiency1	-0.0359 (0.029)	-1.209
	onset2*proficiency1	0.0069 (0.017)	0.409
	onset1*proficiency2	0.0079 (0.016)	0.480
	onset2*proficiency2	0.0048 (0.009)	0.231
Window (b)	(Intercept)	0.0871 (0.027)	3.286
	type (C/V)	0.0866 (0.027)	-0.852
	onset1 (n vs s)	0.291 (0.032)	0.901
	onset2 (n & s vs r)	-0.0090 (0.019)	-0.479
	proficiency1 (L2A vs L2B)	0.0034 (0.025)	0.136
	proficiency2 (L2 vs NS)	0.0117 (0.014)	0.844
	type*onset1	-0.0555 (0.032)	-1.715
	type*onset2	-0.0050 (0.018)	-0.265
	type*proficiency1	0.0108 (0.025)	0.431
	type*proficiency2	-0.0012 (0.014)	-0.093
	onset1*proficiency1	-0.0285 (0.030)	-0.929
	onset2*proficiency1	-0.0137 (0.017)	-0.777
	onset1*proficiency2	0.0234 (0.017)	1.386
	onset2*proficiency2	-0.0052 (0.009)	-0.524

Table 4.4: Linear mixed-effects models for word boundary condition on all participants' target advantage scores in windows of analysis 'c' and 'd'. Three-way interactions are not included as none of them reached t > 2.

	Variable	Estimate (SE)	t value
Window (c)	(Intercept)	0.1902 (0.023)	8.071
	type (C/V)	-0.0694 (0.023)	-2.945
	onset1 (n vs s)	0.0970 (0.028)	3.400
	onset2 (n & s vs r)	-0.0364 (0.017)	-2.162
	proficiency1 (L2A vs L2B)	-0.0158 (0.026)	-0.606
	proficiency2 (L2 vs NS)	0.0669 (0.015)	4.449
	type*onset1	-0.0183 (0.028)	-0.643
	type*onset2	0.0335 (0.017)	1.992
	type*proficiency1	0.0086 (0.026)	0.333
	type*proficiency2	-0.0338 (0.015)	-2.245
	onset1*proficiency1	-0.0130 (0.031)	-0.415
	onset2*proficiency1	0.0062 (0.018)	0.334
	onset1*proficiency2	0.0200 (0.018)	1.101
	onset2*proficiency2	-0.0191 (0.010)	-1.779
Window (d)	(Intercept)	0.4512 (0.029)	15.122
	type (C/V)	-0.0659 (0.029)	-2.209
	onset1 (n vs s)	0.0879 (0.036)	2.406
	onset2 (n & s vs r)	-0.0634 (0.021)	-3.008
	proficiency1 (L2A vs L2B)	-0.0474 (0.027)	-1.695
	proficiency2 (L2 vs NS)	0.0777 (0.016)	4.810
	type*onset1	0.0246 (0.036)	-0.498
	type*onset2	0.0384 (0.021)	-1.718
	type*proficiency1	-0.0116 (0.027)	0.421
	type*proficiency2	-0.0126 (0.016)	0.425
	onset1*proficiency1	-0.0011 (0.034)	-0.653
	onset2*proficiency1	0.0066 (0.019)	-0.558
	onset1*proficiency2	0.0173 (0.019)	-0.393
	onset2*proficiency2	0.0026 (0.011)	2.047

Table 4.5: Linear mixed-effects models for word internal tap and trill (unstressed syllable on all participants' target advantage scores in different windows of analysis.

	Variable	Estimate (SE)	t value
Window (a)	(Intercept)	0.0487 (0.074)	0.659
	type (tap/trill)	0.1438 (0.074)	1.945
	proficiency1 (L2A vs. L2B)	0.0330 (0.085)	0.385
	proficiency2 (L2 vs NS)	-0.0251 (0.052)	-0.479
	type*proficiency1	0.0341 (0.085)	0.398
	type*proficiency2	0.0285 (0.052)	0.542
Window (b)	(Intercept)	0.0689 (0.099)	0.694
	type (tap/trill)	0.9995 (0.099)	1.005
	proficiency1 (L2A vs L2B)	0.0187 (0.094)	0.198
	proficiency2 (L2 vs NS)	-0.0112 (0.049)	-0.228
	type*proficiency1	0.0611 (0.094)	0.646
	type*proficiency2	-0.0061 (0.049)	-0.125
Window (c)	(Intercept)	0.1807 (0.066)	2.737
	type (tap/trill)	0.0022 (0.066)	0.034
	proficiency1 (L2A vs L2B)	-0.0215 (0.058)	-0.366
	proficiency2 (L2 vs NS)	0.0908 (0.036)	2.519
	type*proficiency1	0.00611 (0.058)	1.040
	type*proficiency2	-0.0157 (0.036)	-0.436
Window (d)	(Intercept)	0.4096 (0.101)	4.043
	type (tap/trill)	0.0078 (0.101)	0.084
	proficiency1 (L2A vs L2B)	-0.1107 (0.060)	-1.823
	proficiency2 (L2 vs NS)	0.0769 (0.035)	2.195
	type*proficiency1	0.0123 (0.060)	0.203
	type*proficiency2	-0.0638 (0.035)	-1.822

Table 4.6: Linear mixed-effects models for word internal tap and trill (stressed syllable on all participants' target advantage scores in different windows of analysis.)

	Variable	Estimate (SE)	t value
Window (c)	(Intercept)	0.298 (0.078)	3.828
	type (tap/trill)	0.0254 (0.078)	0.326
	proficiency1 (L2A vs L2B)	0.0303 (0.077)	0.392
	proficiency2 (L2 vs NS)	0.0937 (0.044)	2.101
	type*proficiency1	0.0280 (0.077)	0.363
	type*proficiency2	-0.0121 (0.044)	-0.271
Window (d)	(Intercept)	0.5890 (0.072)	8.141
	type (tap/trill)	-0.0053 (0.072)	-0.074
	proficiency1 (L2A vs L2B)	0.0734 (0.070)	1.042
	proficiency2 (L2 vs NS)	0.1184 (0.040)	2.910
	type*proficiency1	0.0591 (0.070)	0.838
	type*proficiency2	-0.0083 (0.040)	-0.205

4.4.2 Growth Curve Analysis

The following two analysis methods are non-traditional approaches that are attested in the literature and currently implemented in an R package entitled *eyetrackingR* (Dink and Ferguson, 2016), being developed by two Cognitive Psychology students at Northwestern University. This package is still in progress and available at http://www.eyetrackingR.com. Unfortunately, the R source codes are designed for visual-world experiments in which there are simply two pictures on the screen (two Areas of Interest), while this experiment includes four: a target, a competitor, and two unrelated distractors. As such, I have modified the original code to function properly for this particular experimental design. Any flaws in the coding or reasoning behind my modifications are mine alone and are not reflective of *eyetrackingR*.

Instead of evaluating the target-advantage scores in particular windows of interest, which reduces the complexity of a time course curve to a small set of data points, another methodology seeks to compare the shapes of the time course curves themselves by fitting a complex polynomial function to the time course data. The data is first transformed to empircal log odds to calculate how likely it was that participants were fixating at the target image. In *eyetrackingR*'s implementation, the analysis is only performed on the target image, and does not compare it to fixations on the competitor image or distractors, thus, the comparisons are made between conditions (e.g. when and how much the participant fixated at the target image in consonant vs. vowel initial conditions). Details of the methodology, called Growth Curve Analysis can be found in recent work, such as Barr (2008); Mirman et al. (2008); Mirman (2014).

The same procedure of cleaning and binning the data was performed as above. In the process of converting the data to empirical logit and creating overall curves to fit the data, several higher level (here, quartic) orthogonal polynomial values on time terms were created. For details on this process, see Chapter 3 of Mirman (2014). These linear mixed effects models (using the R package *lme4*) were fit to the data with several predictors: proficiency level, type (C or V), and consonant (r or s/n) as well as the four orthogonal polynomial time components. Then, these models were compared to each other using an ANOVA to

see which terms were most significant. Including all three predictors improved the model significantly ($\chi^2(6)$ =296.85, p < 0.001). The estimates and t-values of the best-fit model (Elog ~ level*type*rsn*(ot1 + ot2 + ot3 + ot4) + (1 | Soundfile) + (1 | Participant)) are shown in Table 4.7.
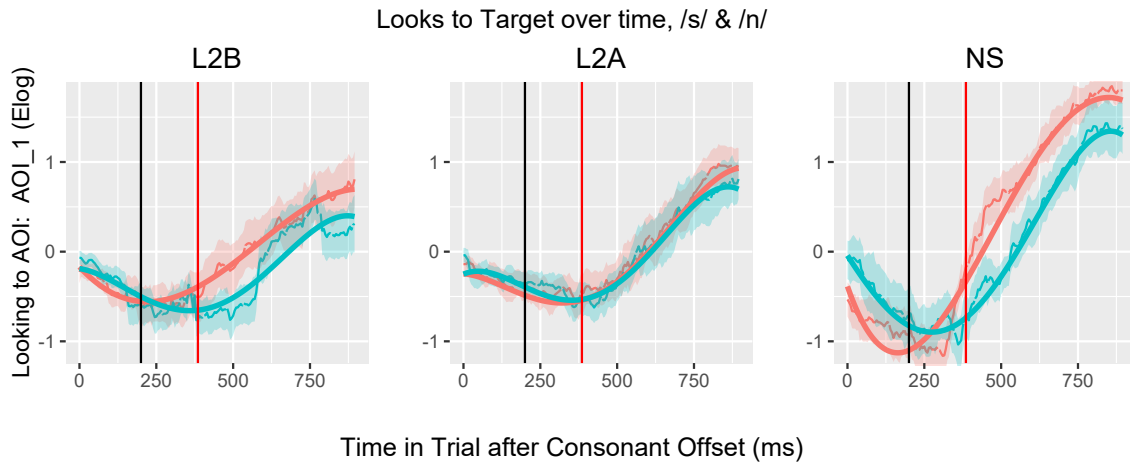
Figure 4.20: Empirical log odds (Elog) of fixations to the target image over time for trials with word-boundary /n/ and /s/, sorted by the three proficiency levels and consonant- vs. vowel-initial trials. Time is recorded from the offset of the pivotal consonant. The growth curve analysis model is overlayed on the transformed data. The black vertical bar represents the offset of /s/ or /n/, the red bar indicates the average disambiguation point later in the acoustic signal. On the plot, blue represents logit odds of looks in the vowel-initial target condition, red represents logit odds of looks in the consonant-initial target condition.

As can be seen in Figures 4.20 and 4.21, the shape of the curves differ significantly between the native speaker group and the two L2 speaker groups. For /r/, both native speakers and advanced L2 speakers fixate on the /r/-initial word earlier than they do the vowel-initial word; both do so before the disambiguation period ends, suggesting they are correctly fixating on the /r/-initial word only needing to hear the /r/ and not the disambiguating acoustic information. For /s/ and /n/, advanced L2 speakers treat these conditions identically; that is, they are just as likely to be looking at the target in both the consonant and vowel initial conditions. However, they are only more likely to be looking at the target image after the disambiguation period. For beginning L2 speakers, there is some evidence of consonant-initial bias for the /s/ and /n/ condition. However, for /r/, they are more likely to fixate at the target in the vowel-initial condition than at the target in the consonant-initial condition.

Looks to Target over time, /r/
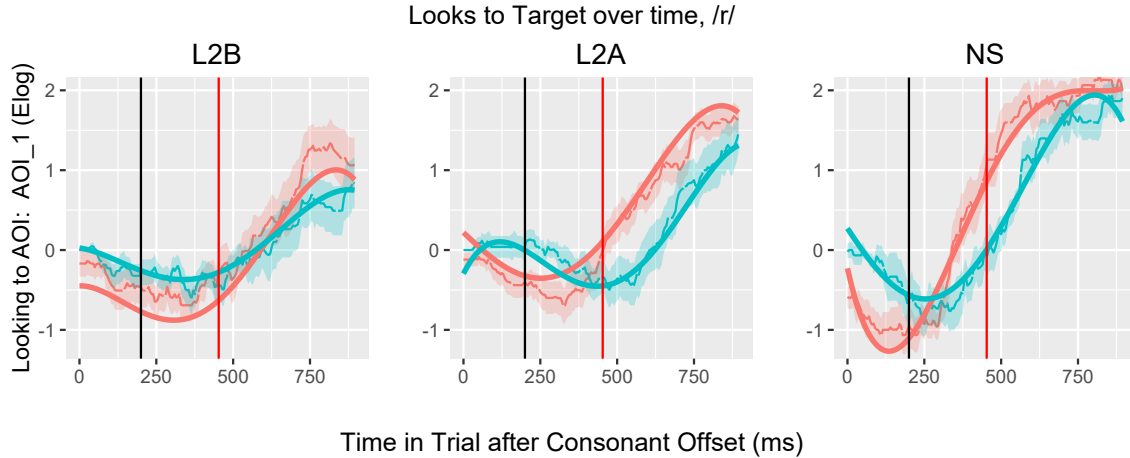
Figure 4.21: Empirical log odds (Elog) of fixations to the target image over time for trials with word-boundary /r/, sorted by the three proficiency levels and consonant- vs. vowel-initial trials. Time is recorded from the offset of the pivotal consonant. The growth curve analysis model is overlayed on the transformed data. The black vertical bar represents the offset of the tap or trill, the red bar indicates the average disambiguation point later in the acoustic signal. On the plot, blue represents logit odds of looks in the vowel-initial target condition, red represents logit odds of looks in the consonant-initial target condition.

4.4.3 Onset-Contingent Analysis

An additional analysis detailed in *eyetrackingR* is perhaps the most appropriate for the purposes of this experiment. The so-called Onset-Contingent Analysis (Fernald et al., 2008) asks the question: "How quickly does the participant switch to look at the correct image given that they initially gazed at the *incorrect* image, and vice versa?" Ultimately, participants should be faster to switch *to* the target image than *away from* it. The term 'Onset-Contingent' thus refers to the image that the participant is looking at the onset of the critical phoneme or word, as their fixation patterns are contingent on where they are fixating initially, a point I have discussed above. For the present experiment, we are most interested in the consonant-initial bias that beginning learners may have. These learners may take longer to switch *away* from a consonant-initial word when the target word is actually vowel-initial, than when they initially fixate at a vowel-initial word when the target was consonant initial. Advanced learners and native speakers should have suppressed this

Table 4.7: Linear mixed-effects models for word boundary condition on all participants' empirical logit odds (Elog) for fixating at the target image, for the Growth Curve Analysis

Variable	Estimate (SE)	t value
(Intercept)	0.1261 (0.073)	1.72
proficiency1 (L2A vs L2B)	-0.0795 (0.058)	-1.37
proficiency2 (L2 vs NS)	0.1028 (0.039)	2.61
onset (n & s vs r)	-0.1375 (0.005)	-2.65
type (C vs V)	-0.0819 (0.005)	-1.58
Linear	7.717 (0.063)	122.56
Quadratic	3.2979 (0.063)	52.37
Cubic	-1.755 (0.063)	-27.87
Quartic	-0.5836 (0.063)	-9.27
onset*proficiency1	0.0028 (0.005)	5.37
onset*proficiency2	-0.0259 (0.003)	-7.21
type*proficiency1	-0.0092 (0.005)	-1.74
type*proficiency2	-0.0167 (0.003)	-4.67
onset*type	-0.0056 (0.051)	-0.11
onset*type*proficiency1	-0.0243 (0.005)	-4.59
onset*type*proficiency2	0.0028 (0.004)	0.79

Table 4.8: Linear mixed-effects models for word internal condition on all participants' empirical logit odds (Elog) for fixating at the target image, for the Growth Curve Analysis

| Variable	Estimate (SE)	t value
(Intercept)	0.4777 (0.124)	3.82
stress (1st vs 2nd syllable)	-0.2584 (0.088)	-2.92
proficiency1 (L2A vs L2B)	-0.1201 (0.099)	-1.21
proficiency2 (L2 vs NS)	0.1330 (0.067)	1.97
type (tap vs trill)	-0.0379 (0.088)	-0.43
Linear	9.7657 (0.106)	91.55
Quadratic	-0.5314 (0.106)	-4.98
Cubic	-1.5838 (0.106)	-14.85
Quartic	0.5000 (0.106)	4.69
stress*proficiency1	-0.0725 (0.009)	-8.05
stress*proficiency2	-0.1286 (0.006)	-21.20
stress*type	-0.0220 (0.088)	-0.25
type*proficiency1	-0.0548 (0.009)	-6.04
type*proficiency2	0.0352 (0.006)	-5.80
onset*type*proficiency1	0.0395 (0.009)	4.39
onset*type*proficiency2	0.0258 (0.006)	4.26
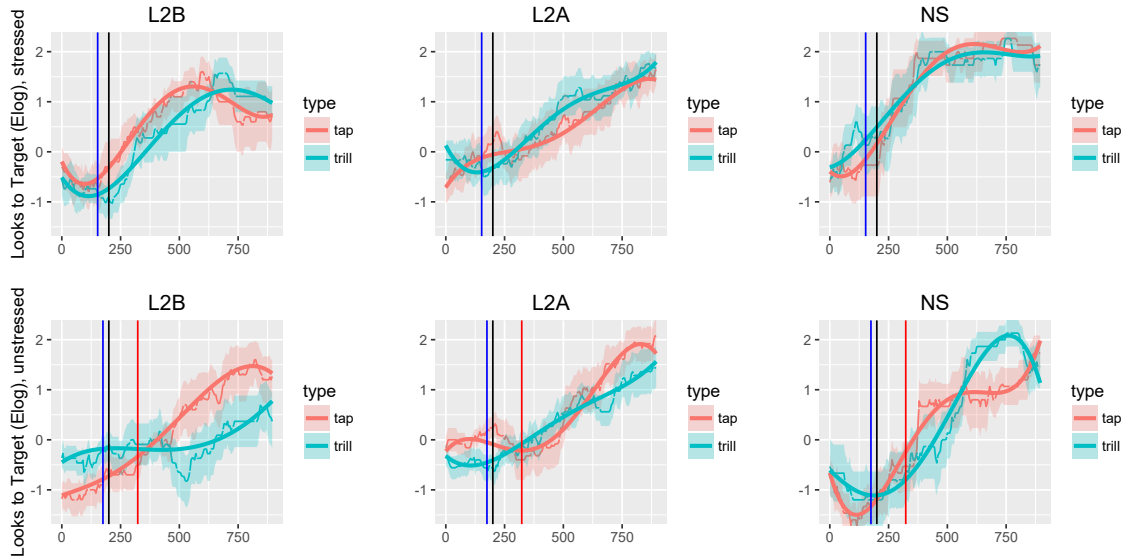
Figure 4.22: Empirical log odds (Elog) of fixations to the target image over time for trials with word-internal /r/, sorted by the three proficiency levels and tap vs. trill trials. Time is recorded from the offset of the pivotal consonant. The growth curve analysis model is overlayed on the transformed data. The blue vertical bar indicates the onset of the tap or trill, black vertical bar represents the offset of the tap or trill, and the red bar indicates the average disambiguation point later in the acoustic signal.

consonant-inital bias given that they have successfully acquired resyllabification, and should switch away from both consonant- and vowel-initial competitors at equal rates.

As mentioned above, however, this type of analysis is more transparent if there are only two possible images on the screen. In the experiment detailed in this chapter, there were four images on the screen. Thus, there are instead several possibilities of eye-gaze switching: (1) to the target, (2) to the competitor, or (3) to a distractor.

An additional problem encountered in this analysis not discussed in either (Dink and Ferguson, 2016) or (Fernald et al., 2008) is that of *when* to measure where the participant is originally looking. If one measures too early, the participant may not have diverted their gaze from the initial crosshair position and the analysis cannot be performed; too late, and the participant has already looked at multiple images. Ultimately, it is not the case that participants immediately switch their gaze to one of the four images on the screen as soon they hear any phonetic information, unlike has been argued for in the eye tracking literature. In 4.8, the time from the onset of the pivotal consonant to the time that the the participant first looked at an image is shown. Although this is an aggregation of all of the trials in the

experiment, and some consonants were longer than others, on average participants should have begun to look at an image 200 ms after the onset of the pivotal phoneme.

By using *eyetrackingR*'s default settings of checking the interest area fixated 300 ms after the onset of the trial, only 54.7% of the trials were analyzed (in all other cases, the participant had not yet fixated at any area of interest). This percentage is unacceptable for doing any serious analysis. One possible solution to normalize time such that the onset-contingent analysis starts at the first fixation of any region of interest. However, there are limitation to this approach: as detailed above, one loses the alignment with the acoustic-phonetic signal. If this is the case, instead of salvaging this type of analysis, a better approach is to simply analyze is to simply analyze the patterns of eye fixations. on a trial-by-trial basis, which abstracts away from time. This type of analysis is detailed below.

4.4.4 Trial-by-Trial Analysis

The purpose of this trial-by-trial analysis is to confirm that the traditional aggregated analysis above reflects what participants do on a trial-by-trial basis. In other words, we want to see the eye-gaze fixation patterns on each trial and characterize these patterns in terms of actual word-activation with no incidental spurious activation.

To begin with, trials with no correct fixations were eliminated (32 trials, 1.66% of the data). Additionally, each trial was coded for which image was clicked on at the end of the trial (as related above, this is an estimate of where the mouse was clicked, the actual data is the fixation of the eyes at the time of the mouse click). In 66 trials, the participant clicked on neither the target nor the competitor (3.4% of the data); these trials were also excluded from further analysis. Then, each trial was coded for when during the trial the participant first looked at the target image ('correct fixation'). This time was subtracted from the time of the offset of the pivotal consonant for each trial ('correct after offset'), this process is similar to the Onset Contingent analysis above, but without considering which image was first fixated at. Next, the time of the mouse-click was theorized to be the end of the trial, as on many occasions, after the mouse button was clicked, the participant spent the short time in between the click and the disappearance of the images on the screen looking at all

of the other images. We do not want to count these looks as potential word activations. All fixations after the mouse was click on each trial were removed. Then, the sequence of fixations to each of the four images on the screen were analyzed. Of particular interest is the order in which the target and competitor images were fixated at. This order was determined by a custom-built Python script, analyzing each individual trial. In all cases, fixations to the two distractor items were ignored: as mentioned above, the fixations on these images should not reflect actual word activation, but are a result of the participant mistakenly looking at the wrong image (see again Figure 4.10), which necessarily adds additional time before the target (or competitor) images are fixated at. The classifications (also discussed above, with relevant examples) for each trial were as follows:

1. *Only Target*: During the time from the offset of the pivotal consonant to the mouse click, the participant looked at the target image, but not the competitor image (looks to any distractor image were possible).

2. *Only Competitor*: During the time from the offset of the pivotal consonant to the mouse click, the participant looked at the competitor image, but not the target image (looks to any distractor image were possible).

3. *Competitor -> Target*: During the time from the offset of the pivotal consonant to the mouse click, the participant looked at the competitor image first, and then the target image.

4. *Target -> Competitor -> Target*: During the time from the offset of the pivotal consonant to the mouse click, the participant looked at the target image first, and then the competitor image, and the finally the target image again. This shows that the participant realized there was some ambiguity between the target and competitor words

5. *Target -> Competitor*: During the time from the offset of the pivotal consonant to the mouse click, the participant looked at the target image first, and then the competitor image, clicking on the competitor.

6. *Competitor -> Target -> Competitor*: During the time from the offset of the pivotal

106

consonant to the mouse click, the participant looked at the competitor image first, and then the target image, and the finally the competitor image again, clicking on it.

7. *Competitor -> Target -> Competitor -> Target*: During the time from the offset of the pivotal consonant to the mouse click, the participant looked at the competitor image first, and then the target image, and then the competitor image again, and finally looking at the target image and clicking on it.

To review, the trial-by-trial analysis helps determine the actual pattern of fixations on each individual trial, removing the element of time and its complications, but preserving the most important feature of eye tracking: the possible interpretations of the acoustic signal before the signal has finished. To compare with forced choice perception experiments (e.g. the perception data of Chapter 3), we cannot know if the participant was entertaining multiple possibilities while choosing which word was heard: the only result is the chosen word at the end of the trial. Measurements of reaction time or a Likert-scale confidence rating can be good proxies for how difficult the item was to decide, but this is not as transparent of a measure as tracing the possible interpretations as the trial unfolds. By characterizing each trial as an overall pattern, we can see how often it is that there is competition between the target and competitor words, and not simply an average of how early in the acoustic signal listeners over all trials look to particular images. The classifications involving more than one fixation at a particular image (e.g. an early fixation to the target, a look to the competitor, and then a return to the target) gives good evidence of lexical competition.

Figures 4.23 and 4.24 give the percentages of trials where each order of fixation pattern was found for word-boundary /n & s/ and /r/, respectively. Overwhelmingly, 80% or more of all trials involved just two or fewer looks at either target or competitor images (again, looks to distractors were not counted). We see that for all speaker groups, the most likely patterns were either 'Only Target' or 'Competitor -> Target', demonstrating a high level of ultimate accuracy. All groups looked at only the target image in more trials than any other pattern, although this was modulated by whether or not they heard a word-initial consonant. Both beginning learners and native speakers show a bias towards looking at the consonant-initial word: this can be seen in both their higher percentage of 'Only Target' trials for

consonant-initial words, as well as their higher percentage of 'Competitor -> Target' trials in the vowel-initial condition. Advanced learners, interestingly, show a slight tendency for the opposite effect.

For word-boundary /r/, both advanced learners and native speakers show a high level of 'Only Target' responses, suggesting that they are sensitive to the phonological distribution of the tap and trill at the word boundary. Advanced learners especially are accurate at detecting the word-initiality of the trill. Beginners, on the other hand, look at the competitor word first only slightly less than they look at the target item in both conditions, suggesting that they have not yet acquired this phonological rule. However, it is also possible that they simply cannot distinguish between the tap and trill in general, not just in the word boundary context. The results for the word internal condition are shown in Figure 4.25, and these results are very similar to that of word boundary /r/, except for one notable exception: both native speakers and advanced learners do not perform as well in this condition, looking at only the target word less than half of the time. There was not hypothesized to be as much lexical competition for these words, but that is not the case. One possible explanation is that because the target consonants here were word-internal, participants may have looked at the competitor word first simply by chance, because both the target and competitor words began with the same consonant (e.g. *gorilas* and *gorrinos*. Thus, the competitor words were activated earlier in the acoustic signal, and the participants need to recover from the fact that there was a later disambiguation point. For the statistical analysis, a generalized linear model with Poisson regression was used to model the count data. A χ^2 test was used to reveal the significant predictors in this analysis. Proficiency (χ^2 (1) = 20.18, p < 0.001) and consonant (χ^2 (2) = 17.92, p < 0.001) were both significant; however word-position (consonant or vowel initial) was not.

4.5   Discussion

This chapter sought to use eye tracking to better investigate the lexical activation of target and competitor words during the real-time processing of resyllabified consonants in Spanish by native speakers and two groups of L2 learners. Two types of consonants were investigated:

Figure 4.23: Time from the onset of the pivotal consonant to the first time a participant's eye gaze reached the target image, sorted by the three proficiency levels and word position. Data includes all pivotal consonants and all conditions.

/n/ and /s/, which are phonologically the same when resyllabified, and /r/, which surfaces as a tap when resyllabified. In addition, word-internal /r/ was investigated to ensure that if any differences were found between resyllabified and non-resyllabified /r/, they would be due to knowledge of the phonological behavior at the word boundary and not due to difficulty in perceiving these two similar segments in general.

Returning to the research questions at the beginning of the chapter, research question (1), whether or not native speakers are sensitive to resyllabification was only partially confirmed. Although there were plenty of vowel-initial word activations when consonant-initial words were intended, and vice versa, by any measure studied in this chapter, there still seems to be an overall bias towards consonant-initial targets, even though the acoustic-phonetic information did not contain salient durational cues to bias towards interpreting consonant-initial tokens over vowel-initial ones. This confirms recent accounts arguing for partial resyllabification in Spanish (Hualde and Prieto, 2014; Strycharczuk and Kohlberger, 2016). For research question (3), which asks the same question for L2 speakers, the hypothesis that L2 speakers would favor syllable-word alignments is partially confirmed again. Surprisingly, beginning L2 speakers perform more like native speakers in that they have more early looks to consonant-initial words, whereas advanced L2 speakers do not show a difference between

Figure 4.24: Time from the onset of the pivotal consonant to the first time a participant's eye gaze reached the target image, sorted by the three proficiency levels and word position. Data includes all pivotal consonants and all conditions.

the two word positions. Advanced L2 speakers seem to be more aware of potential resyllabification in a way that native speakers do not seem to deal with. This may be a due to a combination of two factors: L2 speakers in general are slower to process and recognize words in their second language, and that with increased exposure to the language they have had more chances for misparsing resyllabified words. In order to compensate for the slowness in word recognition, they have incorporated the possibility that syllable and word boundaries may not always align in Spanish. Native speakers of Spanish, however, already process words at a fast rate that some occasional initial misparsing is not a hindrance to overall processing and can be recovered from quickly. This is not unlike some other studies that have found that native speakers can be slower to recognized resyllabified words in their own language (Spinelli et al., 2003). That being said, even beginning L2 speakers must have learned something about resyllabification in Spanish as such a process happens frequently and they must have some mechanism to deal with it. Beginning learners in this eye tracking study do show evidence for competition between consonant-initial and vowel-initial words even in processing consonant-initial words. During such trials they look to the vowel-initial word *first* 24% of the time, when we might expect if they were so biased towards the consonant-initial word this number would be closer to 0%. A further study regarding processing resyllabified
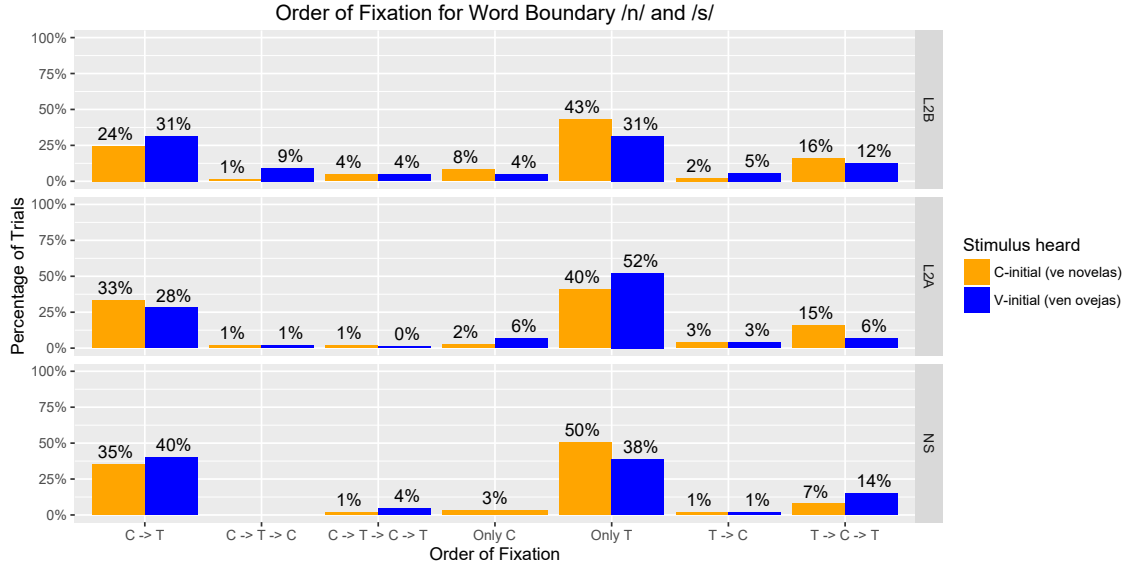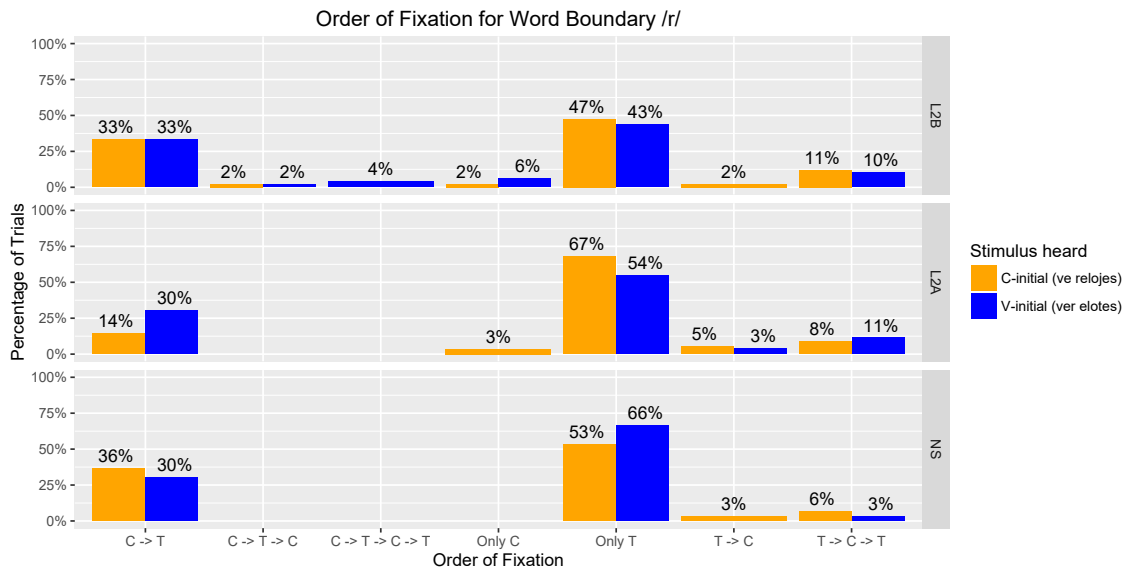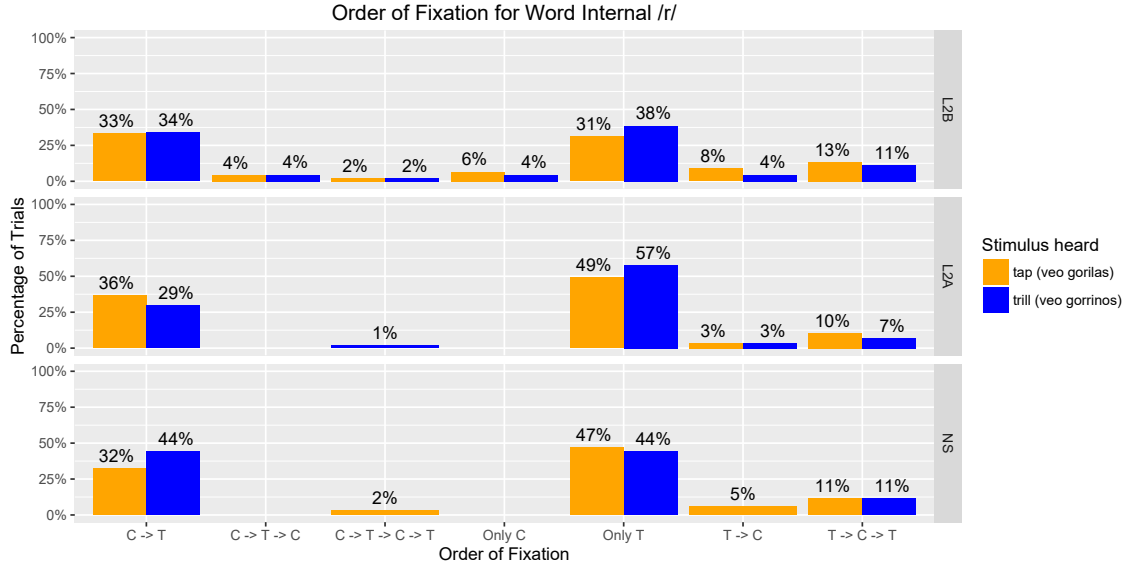
Figure 4.25: Time from the onset of the pivotal consonant to the first time a participant's eye gaze reached the target image, sorted by the three proficiency levels and word position. Data includes all pivotal consonants and all conditions.

words (however unnatural) in English would be necessary to confirm that such learning has occurred in Spanish.

For research questions (2) and (4), which deal with the allophonic distribution of the tap and trill at the word boundary and its significance in speech segmentation, the results are less clear. The Target Advantage and Growth Curve Analysis both showed that native speakers look at initial-trill words more often and sooner than they do vowel-initial words followed by taps, even though they should both be just as good of a cue to speech segmentation. Perhaps the longer duration and increased salience of the trill creates this asymmetry. However, the trial-by-trial analysis did not show that this was a particular good cue for segmentation. On only 53% of the trials did they only look at trill-initial words when it was the target. This was worse than advanced L2 speakers on the same trials, and perhaps more shockingly, worse than the /n/ and /s/ trials which were theoretically ambiguous. In fact, they were more accurate in only looking at the target word when they heard the tap (66% of trials). More speakers in the native speaker group are needed to confirm these results which seem to be at odds with the results from the other methodologies. The hypotheses for the non-native speaker groups were confirmed, however. The more advanced group was consistently able to use the allophonic cues for segmentation across all measures, and especially for the

word-initial trill. With regard to the word-internal tap and trill, the results are also mixed. The window analysis and Growth Curve Analysis showed little difference between fixations towards target taps and trills. However, the trial-by-trial analysis revealed that there was still significant lexical competition between words containing these two sounds word-medially. Even if participants ultimately arrive at the correct word, this much lexical competition was surprising, especially for the native speakers, who performed even worse than the word-boundary position, which is not a lexical contrast. It is possible there were difficulties due to the short nature of these sounds or the word-internal position more generally. Further research is warranted for this intriguing contrast in Spanish.

Another significant contribution in this chapter is the re-examining of the eye tracking methodology, especially concerning aggregating trials together for the purpose of creating time course curves. Comparing across the three methodologies, many of the results look quite similar in terms of the descriptive graphs. However, the type of information (statistical measures) they conveyed is quite different. The traditional window analysis looked at particular time windows and compared how much more likely the target image was fixated at. This was informative as far as how much more activation there was of a consonant-initial word than a vowel-initial one, but did not answer particularly well how early this happened, as there was not found to be any differences early on during the disambiguation period, which is when we would expect there to be phonological 'garden-pathing'. The Growth Curve Analysis, by fitting higher order polynomials to the data, gave more information about the overall shape of the time course curves, which is a better indication of earlier or later fixations. Finally, the trial-by-trial analysis was perhaps the most straightforward and transparent to understand, but eliminated the alignment with the acoustic signal. On the one hand, this solves the earlier problem of native speakers being faster to fixate on any region of interest, but on the other, the acoustic information the participants hear is important. As noted above in the discussion of the research questions, there were several discrepancies with this analysis and the other measures, but this data should not be discounted by any means, as the trial-by-trial fixation order patterns truly tell what participants were doing. The fact that there was a significant number of trials wherein the competitor was first fixated at and trials where the target was first fixated at is completely lost in both the window analysis and

the Growth Curve Analysis. A combination of measures is perhaps the best way to capture lexical activation as time unfold.

've rosas'
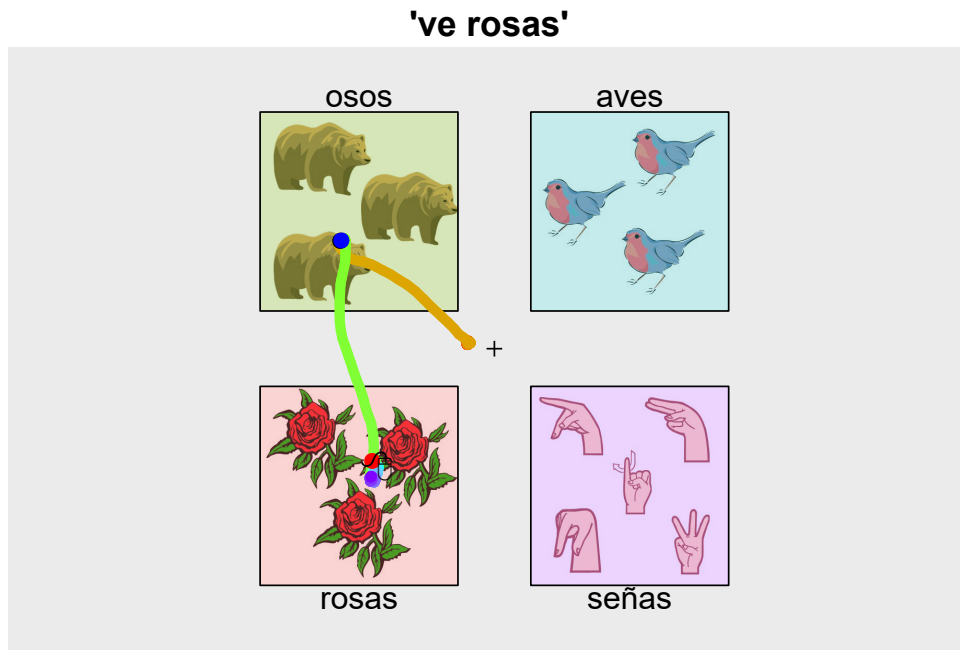
Figure 4.26: Beginning L2 participant eye trace. Note that the participant fixates at the competitor image first, but once she has heard the full /r/ fixates on the target word.

This being said, there is still room for further development in terms of eye tracking analysis. While the trial-by-trial patterns introduced the above section are fairly straightforward in terms of interpretation, the original eye fixation data is still rich. Consider the eye gaze traces developed for this thesis, for example, in Figures 4.10, 4.11 and 4.12, and here, Figures 4.26 and 4.27. To my knowledge, these types of images have not been created or analyzed before. Often a video of the trial is shown at conference presentations or other demonstrations of eye tracking technology, but these videos produced by SR-Research's DataViewer are treated as unanalyzable novelties. However, because this video has been turned into a static image, we are able to quickly analyze the eye fixation pattern. Perhaps the most striking innovation from this image is the overlay of the particular time events in the acoustic signal and how well they match up with a change in direction in the eye gaze trace. As was noted earlier, in all of the traditional and even more innovative eye tracking methodologies, the role of syncing the acoustic signal to events in the eye fixation time course is tenuous at best due to the aggregation of trials that have different critical time points in the trial. The blue,

black, and red vertical lines in the time course data in Figures 4.28, 4.34 etc. only represent an *average* timepoint of the consonant onset, offset, and disambiguiation point, respectively, across all of the relevant soundfiles. This was also the basis of the time windows for the statistical analysis. This gives us a good idea of *around* when participants should be looking where, but these graphs were not especially convincing, even calling into question the 200 ms 'rule' invoked for the time between the processing of a sound and the movement of the eyes. With such fine grain data at our disposal, we should be able to do a better job than simply binning together data between certain time points and hoping that variation 'washes out' in the end. The eye trace images leave no room for ambiguity. Here, the blue, black, and red dots on these graphs represent the precise moment in the acoustic signal 200 ms after the onset, offset, and disambiguation point of *this particular* trial. Looking at Figure 4.26 in particular, it is no coincidence that the blue and black dots seem to be overlapping: this is the time during which the participant hears the word-initial trill /r/. Well before the onset of this consonant, the participant has already starting looking at the V-initial competitor, *osos*, perhaps due entirely to chance or to coarticulation on the previous vowel letting her know that there is a rhotic upcoming. But it is precisely at the offset of this /r/ that her eyes begin to shift to the target word, *rosas*. It takes her some time to get there, during the next two phonemes /os/, but she arrives in the target region of interest before the disambiguation point (recall as well that the traditional fixation proportion-based analysis regards the area between regions of interest the same as looks to empty parts of the screen, which seems to miss the point when traveling between regions of interest).

When viewed as a two-dimensional image, these abrupt changes in X-Y direction stand out as important observable events. I have not been able to capture this behavior numerically as of yet, but it is clear that quite often these important time points in the acoustic signal create a sort of inflection point in these images. It is also difficult to tell from these images, but there also appears to be a change in velocity of the eye gaze traces; that is, the eye fixations may linger in a certain region or even in an exact point for some time, but after one of these consonantal offsets the eye gaze quickly moves to another region. I believe that such measures, once developed, will be able to give a far more accurate picture of lexical competition time-locked to the acoustic signal as each trial unfolds, utilizing the full power

of eye tracking. I hope that including enough images of these eye traces will give a strong impression that such a pursuit will be worthwhile.
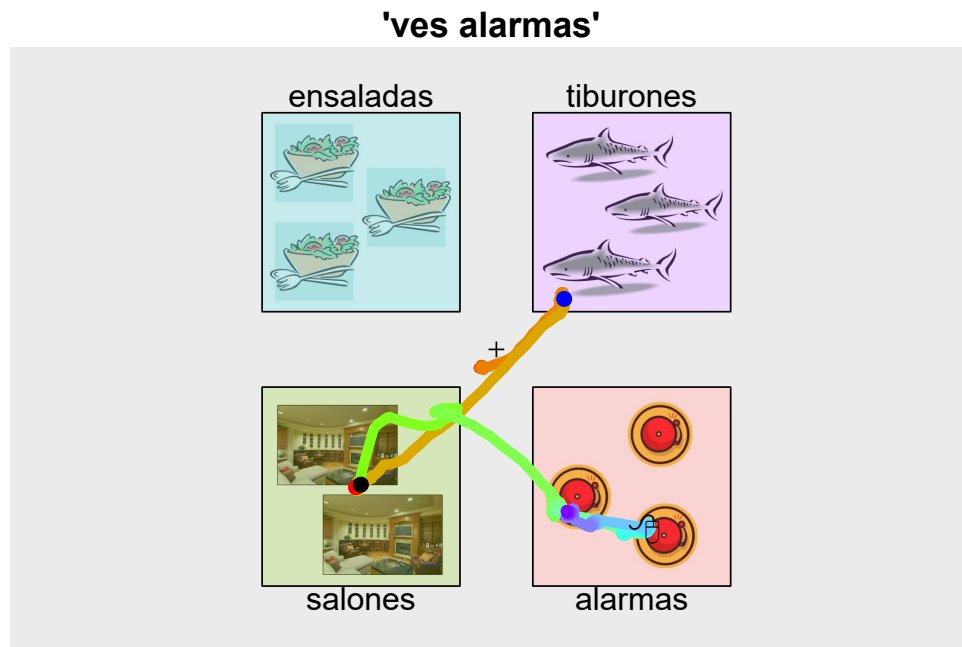
Figure 4.27: Native Speaker eye trace. Note that the participant erroneous fixates at one of the distractor items until the onset of the /s/ (blue dot), and then travels to the consonant-initial competitor until the offset of the /s/ (black dot) before switching to the target item at the end of the /l/ in *alarmas* (red dot).

4.6 Additional time-course figures

Figure 4.28: Results for word-boundary /r/ for native speakers. The x-axis represents bins of 5 ms, the y-axis represents proportions of fixations to the four objects on the screen. The top graph indicates /r/-initial target items; the bottom graph indicates vowel-initial target items (preceding a word-final [ɾ]). The blue vertical line shows the average beginning of the onset of the rhotic; black, the offset of the rhotic; and red, the end of the ambiguous period. The dashed black line indicates the average time the mouse was clicked during these trials for these participants.

Figure 4.29: Results for word-boundary /r/ for advanced L2 speakers. The x-axis represents bins of 5 ms, the y-axis represents proportions of fixations to the four objects on the screen. The top graph indicates /r/-initial target items; the bottom graph indicates vowel-initial target items (preceding a word-final [ɾ]). The blue vertical line shows the average beginning of the onset of the rhotic; black, the offset of the rhotic; and red, the end of the ambiguous period. The dashed black line indicates the average time the mouse was clicked during these trials for these participants.

Figure 4.30: Results for word-boundary /r/ for beginning L2 speakers. The x-axis represents bins of 5 ms, the y-axis represents proportions of fixations to the four objects on the screen. The top graph indicates /r/-initial target items; the bottom graph indicates vowel-initial target items (preceding a word-final [ɾ]). The blue vertical line shows the average beginning of the onset of the rhotic; black, the offset of the rhotic; and red, the end of the ambiguous period. The dashed black line indicates the average time the mouse was clicked during these trials for these participants.
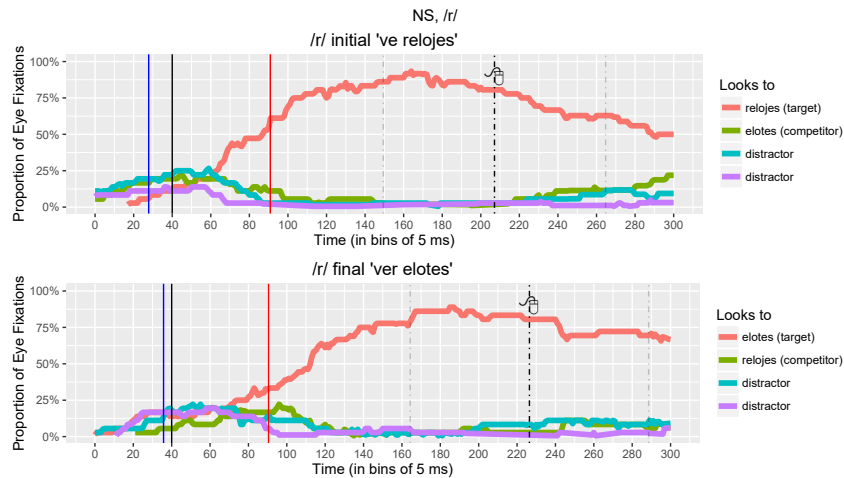
117

Figure 4.31: Results for word-boundary /s/ for native speakers. The x-axis represents bins of 5 ms, the y-axis represents proportions of fixations to the four objects on the screen. The top graph indicates /s/-initial target items; the bottom graph indicates vowel-initial target items. The blue vertical line shows the average beginning of the onset of /s/; black, the offset of /s/; and red, the end of the ambiguous period. The dashed black line indicates the average time the mouse was clicked during these trials for these participants.

Figure 4.32: Results for /n/ for native speakers. The x-axis represents bins of 5 ms, the y-axis represents proportions of fixations to the four objects on the screen. The top graph indicates /s/-initial target items; the bottom graph indicates vowel-initial target items. The blue vertical line shows the average beginning of the onset of /s/; black, the offset of /s/; and red, the end of the ambiguous period. The dashed black line indicates the average time the mouse was clicked during these trials for these participants.
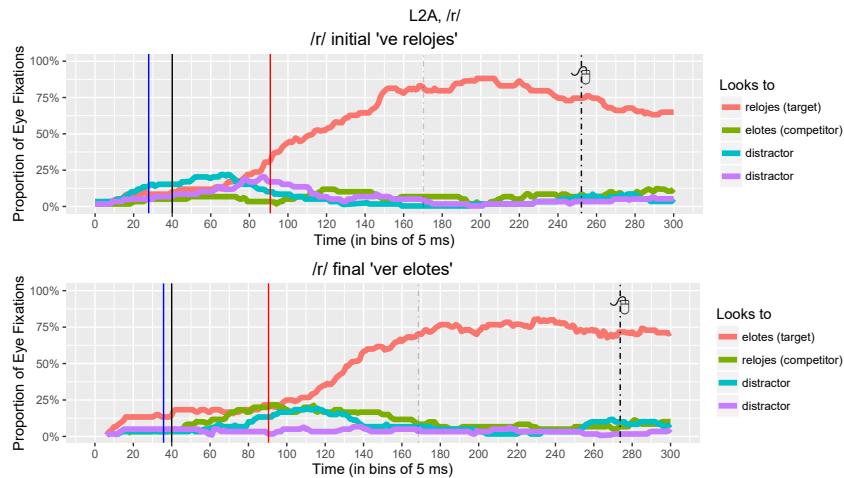
Figure 4.33: Results for /n/ for native speakers. The x-axis represents bins of 5 ms, the y-axis represents proportions of fixations to the four objects on the screen. The top graph indicates /s/-initial target items; the bottom graph indicates vowel-initial target items. The blue vertical line shows the average beginning of the onset of /s/; black, the offset of /s/; and red, the end of the ambiguous period. The dashed black line indicates the average time the mouse was clicked during these trials for these participants.
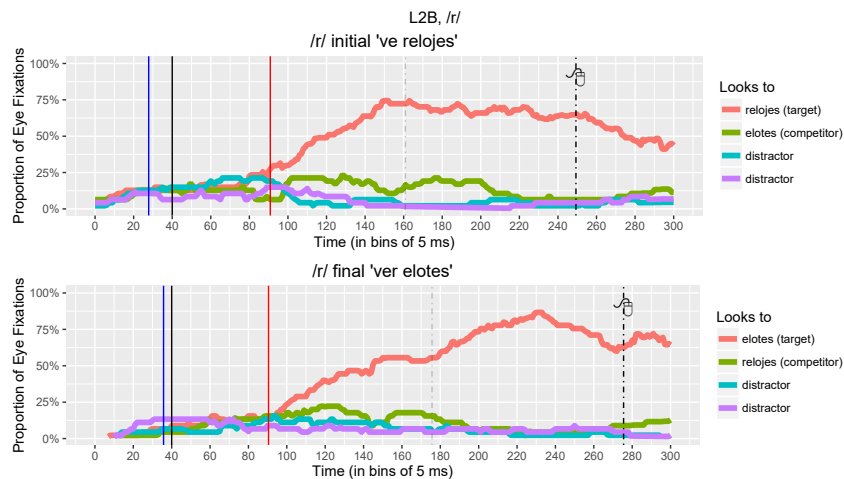

Figure 4.34: Results for /n/ for native speakers. The x-axis represents bins of 5 ms, the y-axis represents proportions of fixations to the four objects on the screen. The top graph indicates /n/-initial target items; the bottom graph indicates vowel-initial target items. The blue vertical line shows the average beginning of the onset of /n/; black, the offset of /n/; and red, the end of the ambiguous period. The dashed black line indicates the average time the mouse was clicked during these trials for these participants.
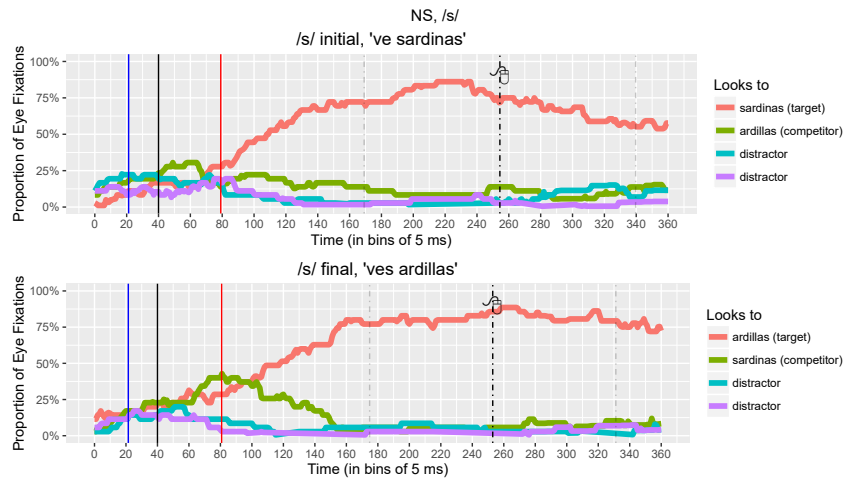
Figure 4.35: Results for /n/ for advanced L2 speakers. The x-axis represents bins of 5 ms, the y-axis represents proportions of fixations to the four objects on the screen. The top graph indicates /n/-initial target items; the bottom graph indicates vowel-initial target items. The blue vertical line shows the average beginning of the onset of /n/; black, the offset of /n/; and red, the end of the ambiguous period. The dashed black line indicates the average time the mouse was clicked during these trials for these participants.
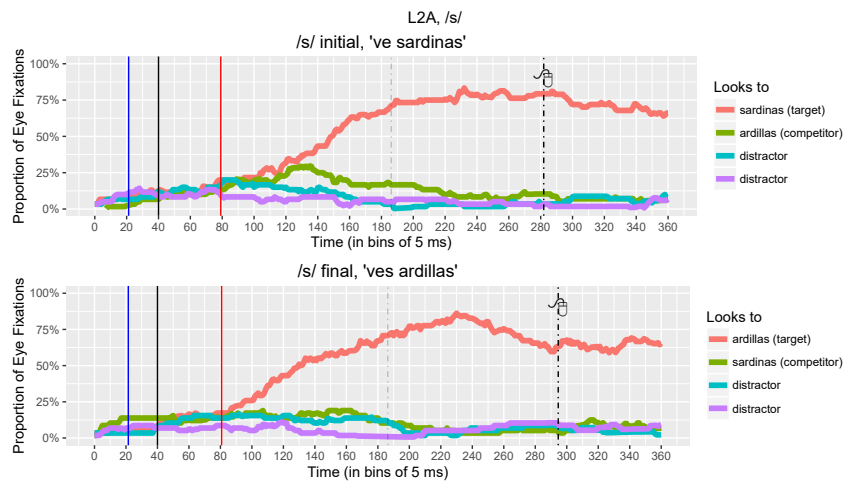

Figure 4.36: Results for /s/ for native speakers. The x-axis represents bins of 5 ms, the y-axis represents proportions of fixations to the four objects on the screen. The top graph indicates /n/-initial target items; the bottom graph indicates vowel-initial target items. The blue vertical line shows the average beginning of the onset of /n/; black, the offset of /n/; and red, the end of the ambiguous period. The dashed black line indicates the average time the mouse was clicked during these trials for these participants.
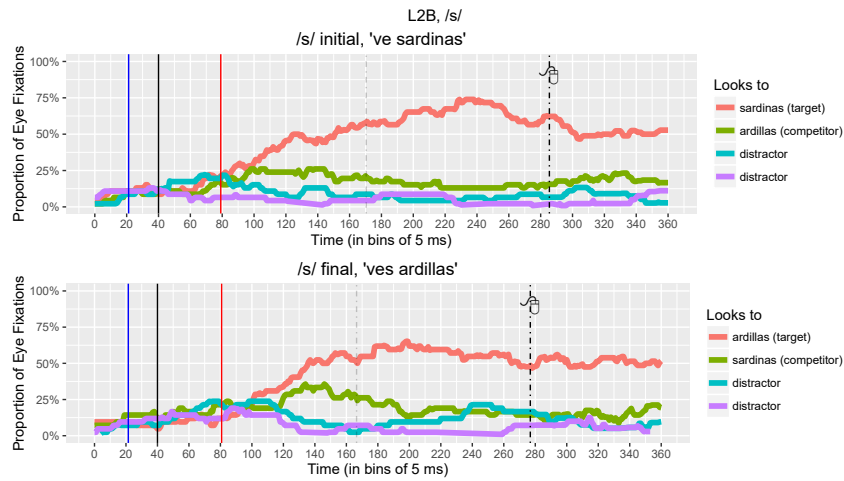
Figure 4.37: Results for word-internal /r/ and /ɾ/ for native speakers where the initial syllable is unstressed. The x-axis represents bins of 5 ms, the y-axis represents proportions of fixations to the four objects on the screen. The top graph indicates word-medial /ɾ/ target items; the bottom graph indicates word-medial /r/ target items. The blue vertical line shows the average beginning of the onset of /r/; black, the offset of /r/; and red, the end of the ambiguous period. The dashed black line indicates the average time the mouse was clicked during these trials for these participants.
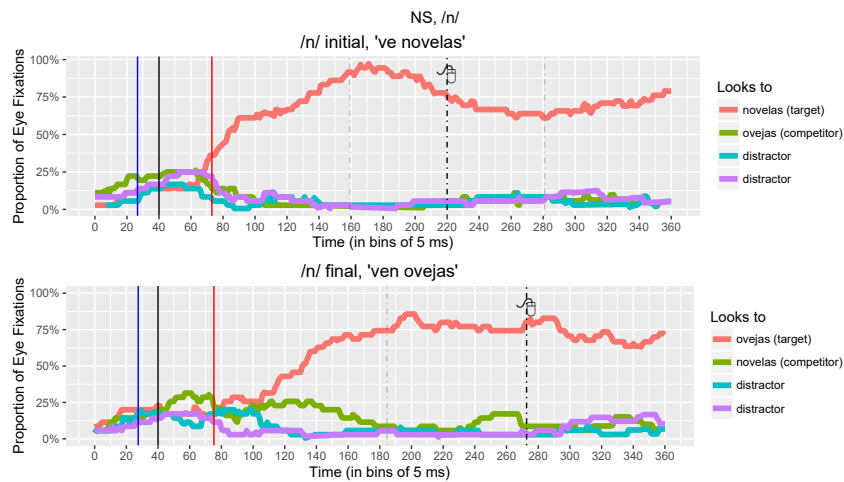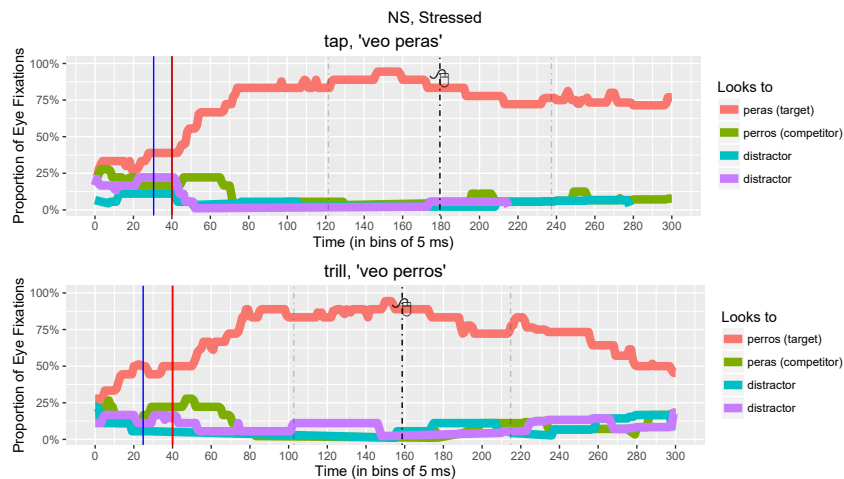

Figure 4.38: Results for word-internal /r/ and /ɾ/ for native speakers where the initial syllable is stressed. The x-axis represents bins of 5 ms, the y-axis represents proportions of fixations to the four objects on the screen. The top graph indicates word-medial /ɾ/ target items; the bottom graph indicates word-medial /r/ target items. The blue vertical line shows the average beginning of the onset of /r/; red, the offset of /r/ and also the end of the ambiguous period. The dashed black line indicates the average time the mouse was clicked during these trials for these participants.
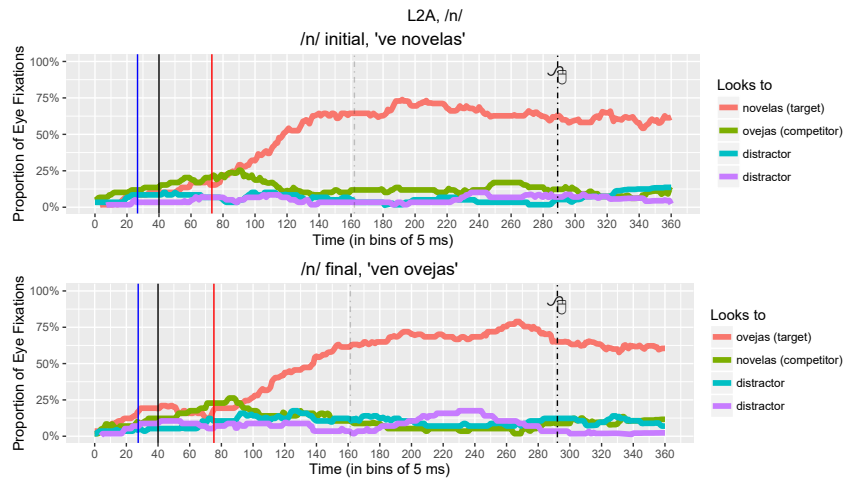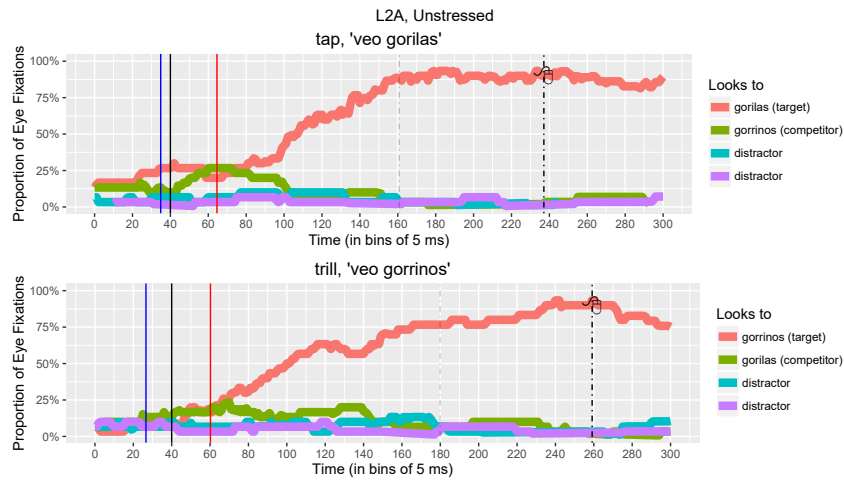
Figure 4.39: Results for word-internal /r/ and /ɾ/ for native speakers where the initial syllable is unstressed. The x-axis represents bins of 5 ms, the y-axis represents proportions of fixations to the four objects on the screen. The top graph indicates word-medial /ɾ/ target items; the bottom graph indicates word-medial /r/ target items. The blue vertical line shows the average beginning of the onset of /r/; black, the offset of /r/; and red, the end of the ambiguous period. The dashed black line indicates the average time the mouse was clicked during these trials for these participants.
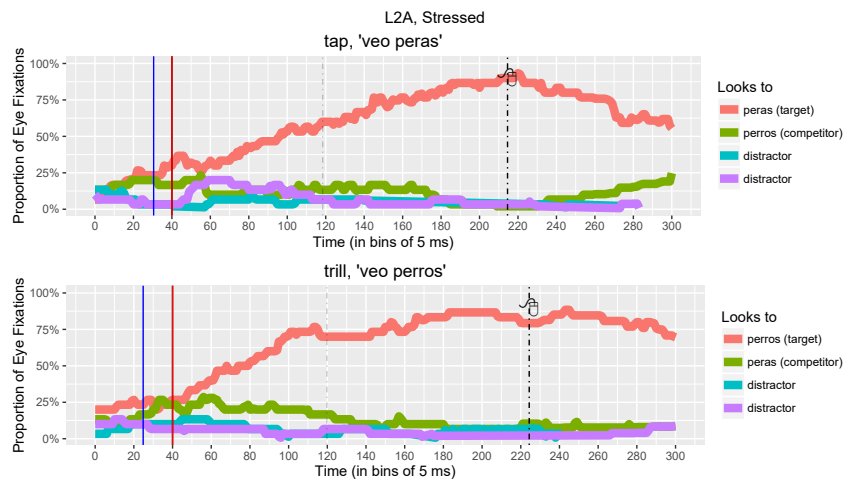
Figure 4.40: Results for word-internal /r/ and /ɾ/ for native speakers where the initial syllable is stressed. The x-axis represents bins of 5 ms, the y-axis represents proportions of fixations to the four objects on the screen. The top graph indicates word-medial /ɾ/ target items; the bottom graph indicates word-medial /r/ target items. The blue vertical line shows the average beginning of the onset of /r/; red, the offset of /r/ and also the end of the ambiguous period. The dashed black line indicates the average time the mouse was clicked during these trials for these participants.
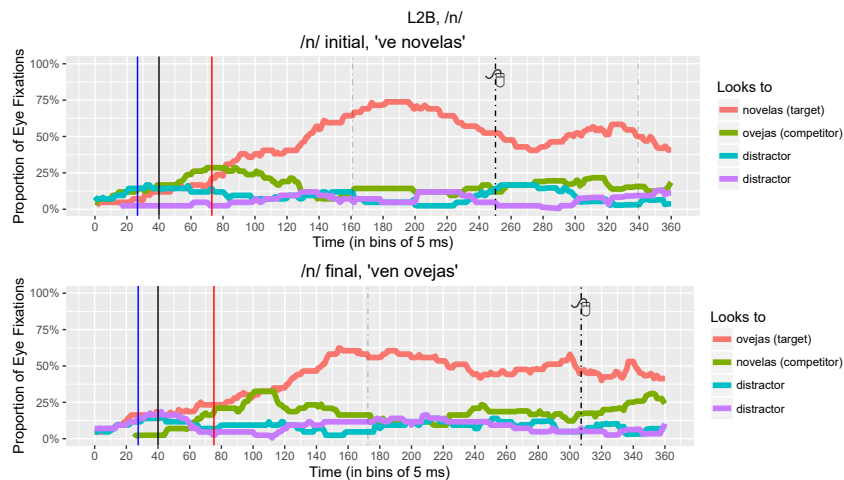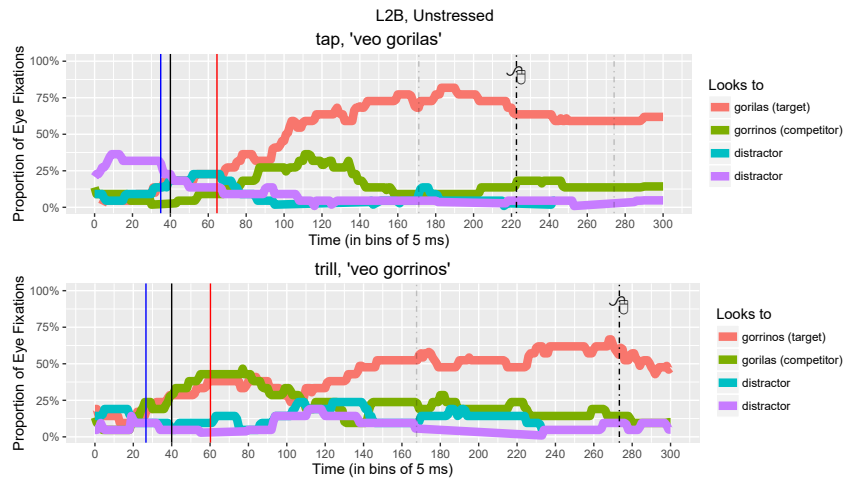
Figure 4.41: Results for word-internal /r/ and /ɾ/ for native speakers where the initial syllable is unstressed. The x-axis represents bins of 5 ms, the y-axis represents proportions of fixations to the four objects on the screen. The top graph indicates word-medial /ɾ/ target items; the bottom graph indicates word-medial /r/ target items. The blue vertical line shows the average beginning of the onset of /r/; black, the offset of /r/; and red, the end of the ambiguous period. The dashed black line indicates the average time the mouse was clicked during these trials for these participants.
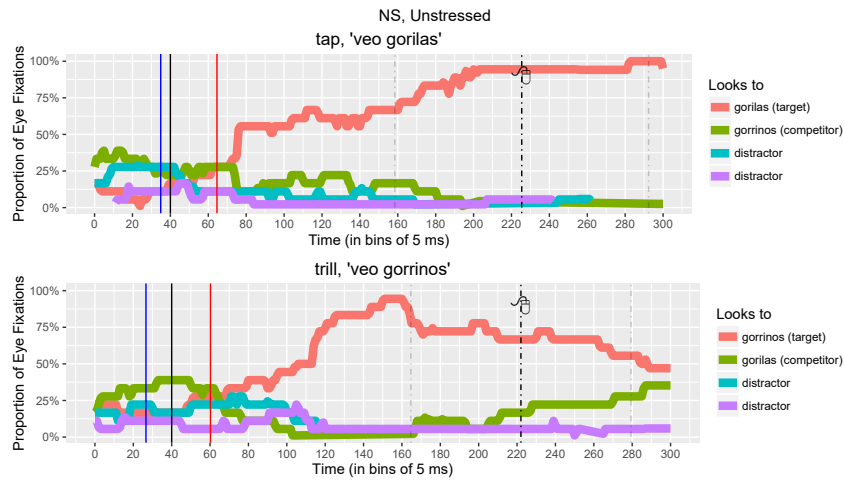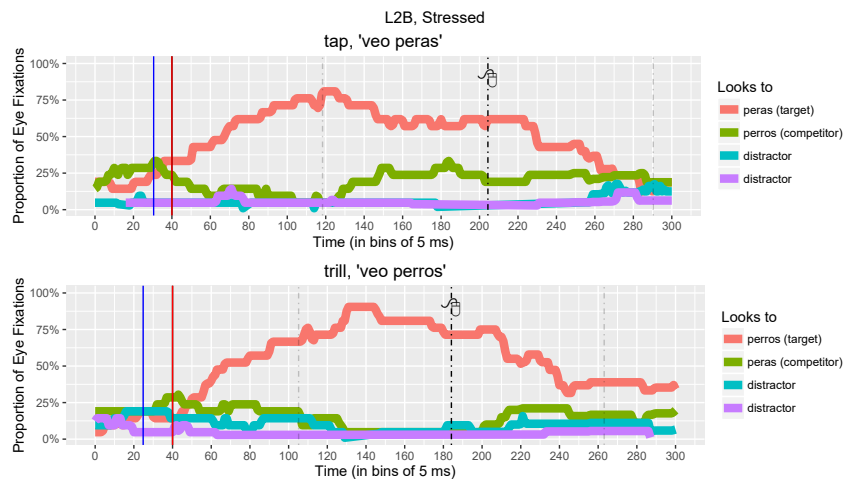

Figure 4.42: Results for word-internal /r/ and /ɾ/ for native speakers where the initial syllable is stressed. The x-axis represents bins of 5 ms, the y-axis represents proportions of fixations to the four objects on the screen. The top graph indicates word-medial /ɾ/ target items; the bottom graph indicates word-medial /r/ target items. The blue vertical line shows the average beginning of the onset of /r/; red, the offset of /r/ and also the end of the ambiguous period. The dashed black line indicates the average time the mouse was clicked during these trials for these participants.

CHAPTER 5

CONCLUSION

The research reported in this thesis has investigated the production and perception of resyllabification in Spanish by native and non-native speakers. To better understand the patterns present for non-native speakers, the thesis also took a brief look at production data in English. This final chapter summarizes the conclusions that can be made summing together the results from production and perception.

5.1 Evidence from Production

Chapter 2 described two production experiments in which native Spanish speakers and L2 speakers of Spanish whose first language was English read sentences, each of which contained a V#CV or VC#V sequence. For the sentences in Spanish, the goal of the study was to see (1) how often L2 speakers glottalize in the VC#V context and (2) if L2 speakers were able to produce the contrast between the tap and the trill at the word boundary. For (1), English speakers glottalized at a much higher rate than did native Spanish speakers, transferring their strong marking of vowel-initial words in English into Spanish. This was the case for any type of glottalization investigated, from simple creaky voice on the initial-vowel to a full glottal stop. For (2), although many L2 speakers were able to produce a trill in the word-internal context, they were not able to transfer over this ability to the word-initial context. With the additional complication of glottalization in the V[ɾ]#V context, L2 speakers have the most to learn with regard to the distribution of rhotics in Spanish. It is not unexpected that the production tasks proved difficult for L2 speakers: although they may be sensitive to resyllabification in perception, the difference in gestural timing needed to resyllabify in Spanish may be difficult to overcome. All of the speakers in this

124

experiment were undergraduates majors in Spanish whose proficiency may be considered intermediate. Future research should also include more advanced students to investigate whether native-like resyllabification and distribution of rhotics is attainable.

5.2 Evidence from Perception

Chapters 3 and 4 described several different perception experiments: one focusing on /n/ and /s/ as pivotal consonants in the V#CV and VC#V contexts to investigate the influence of acoustic-phonetic information. In general, the results from these studies show that native speakers and L2 speakers show sensitivity to resyllabification while listening to Spanish phrases; that is, when hearing *dices eco* with no pause between the words all groups are willing to accept *eco* as the possible second word. However, the groups differed in how often they were able to extract a vowel-initial word out of this context: beginning L2 speakers did so the least, and advanced L2 speakers did so the most (native speakers differed in their treatment of /n/ and /s/ stimuli); this correlated well with proficiency. L2 speakers also used acoustic cues differently than native speakers did: for /n/, length of the consonant seemed to correlate well with perception of it as a word-initial segment and speakers were sensitive to this. For /s/, however, length did not play as strong of a role, and only advanced L2 speakers and native speakers paid less attention to this cue.

For /r/, the experiment in Chapter 3 sought to investigate how well native and L2 speakers were sensitive to the fact that /r/ is a trill when word-initial, but a tap when word final in the VCV context. Both native speakers and L2 speakers were highly accurate at accepting the appropriate rhotic when it matched the lexical item on the screen (that is, when they heard a trill and and the word on the screen began with an <r>), but only L2 speakers were much less accurate in the mis-match condition. This suggests that they accept any kind of /r/ in these positions. This result may mask general perceptual problems with detecting the type of /r/ being produced (they had similar problems in the word-internal condition), or it may additionally be due to lack of awareness of the distribution.

For the eye tracking data in Chapter 4, the main result was that all groups were sensitive to resyllabified words and were able to extract them out of the two-word context, although

there was a bias towards hearing a consonant-initial word for all speakers, but especially for native speakers and beginning learners. For /r/, native speakers and advanced learners used the distribution of the tap and trill to clearly segment the words they heard, although this was strongest with the initial trill, /r/. The results did show some mixed results, especially among the different eye tracking measures. The eye tracking data in Chapter 4 is not at odds with the directly comparable data from the word-monitoring study of /r/. While native speakers and advanced L2 speakers seemed to do worse in the eye tracking experiment than in the word-monitoring experiment, they were still highly accurate by the end of the trial. Eye tracking simply gives us even more fine grained information about the lexical competition process as it unfolds. It is possible that these complications existed in the word-monitoring study as well, but were not observable. While the experiments in Chapter 3 and the eye tracking study in Chapter 4 were quite different, as Chapter 3's experiment involved entirely ambiguous phrases and the phrases in Chapter 4 were only temporarily ambiguous, the results were also not at odds, as in fact the beginning learners in both studies performed closer to native speakers in terms of their bias towards consonant-initial words than did the advanced learners. Finally, to return to eye tracking and lexical processing, it seems that L2 speakers, in general, were slower to look at any object on the screen, perhaps due to delays in lexical processing. Future work in eye tracking will take under consideration the difficulties in interpreting the eye gaze data, especially with the aggregated time course data. Further research comparing native and non-native speakers and inherent differences in processing times should still take advantage of the unique ability of eye tracking as a measure of online processing. This may be done by better classifying participants' results on a trial-by-trial basis and still examining the eye gaze data with alignment to the acoustic signal.

5.3   Final thoughts about lexical representation

How might vowel-initial words be stored in the mental lexicon? The evidence from these studies seems to show that despite the relative frequency of resyllabification in Spanish, there is still a bias towards consonant-initial words by native Spanish speakers. There is thus some cost or delay to activating vowel-initial words that superficially begin with a

consonant. Although all consonant-initial words are activated upon hearing the possible beginning of a word, the processor still strongly activates possible embedded vowel-initial words and is able to recover from a misparse quickly. How often this might happen in English, a language with less resyllabification than Spanish, is not known. The type of experiment presented in this thesis should be done with monolingual English speakers. In Spanish, however, vowel-initial words do not appear to be specially marked in the lexicon as being potentially consonant-initial.

Of note in this study was the lexical representation of the two rhotics in Spanish, which necessarily contrast in word internal position (*caro* vs *carro*). In other positions, the two sounds do not contrast, but are limited in their distribution. In particular, this study looked at the potential contrast between the tap and the trill at the word boundary (*ve rocas* vs *ver ocas*). It seems that <r>-initial words are encoded with an initial trill for native Spanish speakers, who are extremely sensitive to the presence of the trill to segment these words. For word-final /r/, it is less clear, as this segment sometimes does surface as a trill, but not in the context investigated in this paper. For L2 speakers, their representations of these sounds is much less fixed, which is unsurprising given the difficult nature of the production and perception of the sounds in question. Although the speakers produced trills only in word-initial position, which shows that these speakers were not confused about the distribution, even this was rare among the data investigated (taps still dominated this level of speaker). Moreover, many speakers were willing to accept a tap as a word-initial segment in the perception data. A more thorough investigation with multiple proficiency levels of speakers, especially in the production data, would shed more light in this area.

5.4 Summary and future directions

When considering the perception and production data together, it seems that perception precedes production in resyllabification. Although L2 speakers are able to extract resyllabified words out of the speech stream, they rarely produce words in this way. This is unsurprising, as Spanish learners must learn very early that resyllabification is a possibility if they are to process any native-speaker speech at all. However, while they clearly must be able to process

las uvas 'the grapes' as *las uvas* and not *la suvas* as some young children do, this has had few repercussions in producing native-like connected speech. While resyllabification does exist in English in some form (in collocations, or in very casual speech), it is rare, exemplified by the strong tendency for the speakers in this study to mark vowel-initial words with glottalization in both English and Spanish. Why might this be? As shown in Chapter 1, resyllabification is rarely taught to beginning or intermediate students of Spanish, and they likely receive no feedback from instructors. The emphasis in classroom has been on learning words based on their written form, where each word is necessarily separated by a space on the page. More practice with producing connected speech would prove beneficial to these speakers.

With regard to the rhotics, the pattern seems to be the opposite: although speakers never produce a trill in word-final position (in phrases like *ver osos*), they seem to be willing to accept them in this position at least some of the time. However, this story is more complicated given the issues with how these forms are stored in the lexicon and general problem in perceiving the difference between the tap and the trill in Spanish. More research is needed to tease apart the possibilities and find out whether or not L2 speakers really do "know" that the word-initial /r/ is always a trill.

Finally, this thesis intended to find a difference between /n/ and /s/ in resyllabification and exploited verbal paradigms (as verbs can end in /n/, /s/, or a vowel) to do. This proved somewhat unsuccessful and may have introduced complications in the data, as the stimuli that were vowel-initial could be considered somewhat strange: the usage of the second person singular in *ves ardillas* 'you see squirrels' may have been odd for participants. Similarly, in order to manipulate the distribution of the rhotics, inflected and infinitival forms had to be used. Thus, these experiments additionally tested the morphological knowledge of the participants. In the future, a simpler experimental paradigm only focusing on /s/ may yield more clean data: simply manipulating the singular/plural distinction (*la sardina* vs. *las ardillas*, *mi/mis*, *tu/tus*, etc.) as was done in the production experiments would avoid complications with processing morphology and test a simpler (but important) contrast.

REFERENCES

Allopenna, P. D., Magnuson, J. S., and Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of memory and language*, 38(4):419–439.

Altenberg, E. P. (2005). The perception of word boundaries in a second language. *Second Language Research*, 21(4):325–358.

Arteaga, D. L. (2000). Articulatory Phonetics in the First-Year Spanish Classroom. *The Modern Language Journal*, 84(3):339–354.

Barr, D. J. (2008). Analyzing 'visual world' eyetracking data using multilevel logistic regression. *Journal of Memory and Language*, 59(4):457 – 474. Special Issue: Emerging Data Analysis.

Best, C. T. (1995). A Direct Realist View of Cross-Language Speech Perception. *Speech perception and linguistic experience: Issues in cross-language research*, pages 171–204.

Bissiri, M. P., Lecumberri, M. L., Cooke, M., and Volín, J. (2011). The role of word-initial glottal stops in recognizing English words. In Cosi, P., De Mori, R., Di Fabbrizio, G., and Pieraccini, R., editors, *INTERSPEECH 2011, 12th Annual Conference of the International Speech Communication Association*, pages 165–168. ISCA Archive.

Boersma, P. and Weenink, D. (2015). Praat: doing phonetics by computer (version 6.0.08).

Bottke, K. G. (1943). Wrong division. *The Modern Language Journal*, 27(2):133–134.

Broselow, E. (1984). An Investigation of Transfer in Second Language Phonology. *International Review of Applied Linguistics in Language Teaching*, 22(4):253 – 269.

Broselow, E. (1988). Prosodic phonology and the acquisition of a second language. In Flynn, S. and O'Neil, W., editors, *Linguistic Theory in Second Language Acquisition*, volume 8 of *Studies in Theoretical Psycholinguistics*, pages 295–308. Springer Netherlands.

Browman, C. P. and Goldstein, L. (1992). Articulatory phonology: An overview. *Phonetica*, 49(3-4):155–180.

Chappell, W. (2011). The Intervocalic Voicing of/s/in Ecuadorian Spanish. In Michnowicz, J. and Dodsworth, R., editors, *Selected Proceedings of the 5th Workshop on Spanish Sociolinguistics*, pages 57–64, Somerville, MA. Cascadilla Proceedings Project.

Clifton, C. and Staub, A. (2008). Parallelism and Competition in Syntactic Ambiguity Resolution. *Language and Linguistics Compass*, (2):234–250.

Colina, S. (2009). *Spanish Phonology: A syllabic perspective.* Georgetown University Press.

Cutler, A. (2001). Listening to a second language through the ears of a first. *Interpreting*, 5:1–18.

Cutler, A. and Norris, D. (1998). The Role of Strong Syllables in Segmentation for Lexical Access. *Journal of Experimental Psychology: Human Perception and Performance*, 14(1):113–121.

Daidone, D. and Darcy, I. (2014). Quierro comprar una guitara: Lexical encoding of the tap and trill by L2 learners of Spanish. In et al., R. T. M., editor, *Selected Proceedings of the 2012 Second Language Research Forum*, pages 39–50. Cascadilla Proceedings Project, Somerville, MA.

Dalbor, J. B. (1997). *Spanish pronunciation: Theory and practice.* Holt, Rinehart and Winston, Fort Worth, TX.

Davidson, J. (2014). Intervocalic fricative voicing in the Spanish of Barcelona: Considerations for contact-induced sociophonetic innovation. In Potowski, K. and Bugel, T., editors, *Sociolinguistic Change Across the Spanish-speaking World: Case Studies in Honor of Anna Maria Escobar*. Peter Lang Pub Incorporated.

Davis, M. H., Marslen-Wilson, W. D., and Gaskell, M. G. (2002). Leading up the lexical garden path: Segmentation and ambiguity in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 28(1):218.

Dejean de la Batie, B. and Bradley, D. C. (1995). Resolving word boundaries in spoken French : Native and non-native strategies. *Applied Psycholinguistics*, 16:59–81.

Dilley, L., Shattuck-Hufnagel, S., and Ostendorf, M. (1996). Glottalization of word-initial vowels as a function of prosodic structure. *Journal of Phonetics*, 24(4):423–444.

Dink, J. W. and Ferguson, B. (2016). eyetrackingR: An R Library for Eye-tracking Data Analysis.

Dumay, N., Frauenfelder, U. H., and Content, A. (2001). The Role of the Syllable in Lexical Segmentation in French: Word-Spotting Data. *Brain and Language*, 81:144–161.

Estapa, R. (1989). La juntura interna abierta en español. *Universidade de Santiago de Compostela. Servizo de Publicacións e Intercambio Científico.*

Face, T. (2006). Intervocalic Rhotic Pronunciation by Adult Learners of Spanish as a Second Language. In Klee, C. A. and Face, T., editors, *Selected Proceedings of the 7th Conference on the Acquisition of Spanish and Portuguese as First and Second Languages*, pages 47–58. Cascadilla Proceedings Project, Somerville, MA.

Face, T. L. (2002). Re-examining Spanish 'Resyllabification'. In Satterfield, T., Tortora, C., and Cresti, D., editors, *Current Issues in Romance Languages: Selected Papers from the 29th Linguistic Symposium on Romance Languages (LSRL), Ann Arbor, 8-11 April 1999*, Amsterdam Studies in the Theory and History of Linguistic Science, pages 81–94. John Benjamins Publishing Company.

Fernald, A., Zangl, R., Portillo, L., A., and Marchman, V. A. (2008). Looking while listening: Using eye movements to monitor spoken language comprehension by infants and young children. In Sekerina, I., Fernández, E., and Clahsen, H., editors, *Developmental Psycholinguistics: On-line methods in children's language processing*, pages 97–135. John Benjamins, Amsterdam.

Fougeron, C. (2007). Word boundaries and contrast neutralization in the case of enchaînement in French. *Papers in Laboratory Phonology IX: Change in Phonology*, pages 609–642.

Garellek, M. (2012). Glottal stops before word-initial vowels in American English: Distribution and acoustic characteristics. *UCLA Working Papers in Phonetics*, (110):1–23.

Garellek, M. (2013). *Production and perception of glottal stops*. PhD thesis, University of California, Los Angeles.

Guion, S. G., Flege, J. E., Akahane-Yamada, R., and Pruitt, J. C. (2000). An investigation of current models of second language speech perception: the case of Japanese adults' perception of English consonants. *The Journal of the Acoustical Society of America*, 107(5):2711–2724.

Gustafson, E. and Bradlow, A. (2016). French Speech Segmentation in Liaison Contexts by L1 and L2 Listeners. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, 7:1–34.

Harris, J. (1983). *Syllable Structure and Stress in Spanish: a nonlinear analysis*. MIT Press, Cambridge, MA.

Hualde, J. and Prieto, P. (2014). Lenition of intervocalic alveolar fricatives in Catalan and Spanish. *Phonetica*, 71(2):109–127.

Hualde, J. I. (1991). On Spanish syllabification. In Campos, H. and Martínez-Gil, F., editors, *Current studies in Spanish linguistics*, pages 475–493. Georgetown University Press, Washington, DC.

Hualde, J. I. (2005). *The Sounds of Spanish*. Cambridge University Press, Cambridge.

Johnson, K. E. (2008). *Second Language Acquistion of the Spanish Multiple Vibrant Consonant*. PhD thesis, University of Arizona.

Karn, H. E. (1996). *Aspects of Phonetics and Phonology in a Text-to-Speech System for Spanish*. PhD thesis, Georgetown University.

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., Broussard, C., et al. (2007). What's new in Psychtoolbox-3. *Perception*, 36.

Ladd, D. R. and Schepman, A. (2003). "Sagging transitions" between high pitch accents in English: experimental evidence. *Journal of Phonetics*, 31(1):81–112.

Lehiste, I. (1960). An acoustic–phonetic study of internal open juncture. *Phonetica*.

McMurray, B., Tanenhaus, M. K., and Aslin, R. N. (2009). Within-category VOT affects recovery from "lexical" garden-paths: Evidence against phoneme-level inhibition. *Journal of memory and language*, 60(1):65–91.

Michnowicz, J. and Kagan, L. (2016). On glottal stops in Yucatan Spanish. In Sessarego, S. and Tejedo-Herrero, F., editors, *Spanish Language and Sociolinguistic Analysis*, Issues in Hispanic and Lusophone Linguistics, pages 217–240. John Benjamins Publishing Company.

Minkova, D. (2003). *Alliteration and sound change in early English (Cambridge Studies in Linguistics 101)*. Cambridge University Press, Cambridge.

Mirman, D. (2014). *Growth Curve Analysis and Visualization Using R.* Chapman & Hall/CRC The R Series. Taylor & Francis.

Mirman, D., Dixon, J., and Magnuson, J. S. (2008). Statistical and computational models of the visual world paradigm: Growth curves and individual differences. *Journal of Memory and Languages*, 59:474–494.

Morgan, T. (2010). *Sonidos en contexto.* New Haven, CT: Yale University Press.

Nakatani, L. and Dukes, K. (1977). Locus of segmental cues for word juncture. *Journal of the Acoustic Society of America*, 62(3):714–719.

Norris, D. (1994). Shortlist: a connectionist model of continuous speech recognition. *Cognition*, 52:189–234.

Norris, D., McQueen, J., Cutler, A., and Butterfield, S. (1997). The Possible-Word Constraint in the Segmentation of Continuous Speech. *Cognitive Psychology*, 34:191–243.

Penny, R. J. (2002). *A history of the Spanish language.* Cambridge University Press, Cambridge.

Quené, H. (1993). Segment durations and accent as cues to word segmentation in Dutch. *Journal of the Acoustic Society of America*, 94:2027–2035.

Redi, L. and Shattuck-Hufnagel, S. (2001). Variation in the realization of glottalization in normal speakers. *Journal of Phonetics*, 29:407–429.

Rose, M. (2010). Differences in discriminating L2 consonants: A comparison of Spanish taps and trills. In Prior, M. T., editor, *Selected proceedings of the 2008 Second Language Research Forum*, pages 181–196.

Scarpace, D. and Hualde, J. (2013). The production and perception of syntactic gemiantes in Spanish. In *International Conference on Phonetics and Phonology 2013, NINJAL*, Tokyo, Japan.

Schwegler, A., Kempff, J., and Ameal-Guerra, A. (2010). *Fonética y fonología españolas*. New Jersey: John Wiley & Sons, Inc.

Scobbie, J. and Pouplier, M. (2010). The role of syllable structure in external sandhi: An EPG study of vocalisation and retraction in word-final English/l/. *Journal of Phonetics*, 38:240–259.

Shatzman, K. and McQueen, J. (2006). The modulation of lexical competition by segment duration. *Psychonomic Bulletin and Review*, 13(6):966–971.

Shoemaker, E. M. (2009). *Acoustic Cues to Speech Segmentation in Spoken French: Native and Non-native Strategies*. PhD thesis, The University of Texas, Austin.

Soto-Faraco, S., Sebastián-Gallés, N., and Cutler, A. (2001). Segmental and suprasegmental mismatch in lexical access. *Journal of Memory and Language*, 45(3):412–432.

Spinelli, E., McQueen, J. M., and Cutler, A. (2003). Processing resyllabified words in French. *Journal of Memory and Language*, 48:233–254.

Strycharczuk, P. and Kohlberger, M. (2016). Resyllabification Reconsidered: On the Durational Properties of Word-Final /s/ in Spanish. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, 7:1–24.

Suomi, K., Mcqueen, J. M., and Cutler, A. (1997). Vowel Harmony and Speech Segmentation in Finnish. *Journal of Memory and Language*, 36(3):422–444.

Torreira, F. (2007). Tonal Realizations of Syllabic Affiliation in Spanish. *ICPhS XVI, Saarbrücken*, (August):1073–1076.

Torreira, F. and Ernestus, M. (2012). Weakening of Intervocalic / s / in the Nijmegen. *Phonetica*, 69:124–148.

Tremblay, A. (2011). Learning to parse liaison-initial words: An eye-tracking study. *Bilingualism: Language and Cognition*, 14(03):257–279.

Tremblay, A. and Spinelli, E. (2013). Language and Cognitive Processes Segmenting liaison-initial words : The role of predictive dependencies. *Language and Cognitive Processes*, 28(8):1093–1113.

Tremblay, A. and Spinelli, E. (2014). English Listeners' Use of Distributional and Acoustic-Phonetic Cues to Liaison in French: Evidence from Eye Movements. *Language and Speech*, 57(3):310–337.

Tuinman, A., Mitterer, H., and Cutler, A. (2012). Resolving ambiguity in familiar and unfamiliar casual speech. *Journal of Memory and Language*, 66:530–544.

Umeda, N. (1978). Occurrence of glottal stops in fluent speech. *The Journal of the Acoustical Society of America*, 64(July):88–94.

Vogel, I. (1991). Prosodic phonology: second language acquisition data as evidence in theoretical phonology. In Huebner, T. and Ferguson, C. A., editors, *Crosscurrents in second language acquisition and linguistic theories*, pages 47–65. John Benjamins Publishing.

White, L., Melhorn, J., and Mattys, S. (2010). Segmentation by lexical subtraction in Hungarian speakers of second-language English. *The Quarterly Journal of Experimental Psychology*, 63(3):544–554.