The Acquisition of Vocalic Sequences
by English-Speaking L2 Learners of Spanish

Bethany MacLeod
First Generals Paper
January 10, 2008
Advisor: Laura Colantoni
Table of Contents

<table>
<thead>
<tr>
<th></th>
<th>Introduction</th>
<th>page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>Vocalic Sequences</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>2.1. Diphthongs &amp; Hiatus</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>2.2. English</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>2.3. Spanish</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>Literature Review</td>
<td>6</td>
</tr>
<tr>
<td>4</td>
<td>Experiment</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>4.1. Hypotheses</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>4.2. Participant Groups</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>4.3. Stimuli</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>4.4. Procedure</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>4.5. Data Analysis</td>
<td>18</td>
</tr>
<tr>
<td>5</td>
<td>Hypotheses Evaluation &amp; Discussion</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>5.1. Hypothesis 1: The Diphthong/Hiatus Contrast</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>5.2. Hypothesis 2: The Effect of Phonological Transfer</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td>5.3. Hypothesis 3: Phonetic Transfer &amp; Glides</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td>5.4. General Discussion</td>
<td>35</td>
</tr>
<tr>
<td>6</td>
<td>Conclusions &amp; Future Work</td>
<td>39</td>
</tr>
</tbody>
</table>

References | 43   |

Appendix A - Stimuli | 46   |
Appendix B - Passage  | 47   |
Appendix C - Mean Syllable Durations & Normalizing Factors | 48   |
Introduction

Investigations of the production of Spanish vowels by native English speakers have shown that learners tend to substitute English vowels for Spanish vowels (Morrison 2003, 2006), to reduce Spanish vowels in unstressed syllables (Menke & Face 2008), and to use a smaller proportion of the acoustic vowel space than native speakers (Menke & Face 2008, Reeder 1998). While these studies provide important insights into the process of acquiring the vocalic inventory of Spanish, they have left unaddressed the acquisition of sequences of vowels. Spanish includes a contrast between two vowels in sequence being syllabified as a diphthong (monosyllabic glide+vowel sequence) or as a hiatus (disyllabic sequence of two vowels) (Harris 1969, Hualde 1991, 1997, 1999, 2004, Hualde & Prieto 2002, Roca 1997, Aguilar 1999, etc).

The present study aims to fill this gap by examining the production of diphthongs and hiatus by beginning and intermediate learners of Spanish and comparing them to a native Spanish control group. The purpose of the experiment is threefold. First, to determine in what ways the production of these two sequence types (diphthongs and hiatus) differs between the learner groups and the native speakers in terms of two phonetic correlates of the contrast (discussed in §2.1). Second, to track the path of development of the sequences as proficiency with Spanish increases. Third, to determine whether transfer from English affects the realization of the Spanish sequences in terms of the syllabification of diphthongs as well as the acoustic realization of the glides.

The general hypotheses that are tested are (a) that the distinction between a diphthong and a hiatus (as defined by two acoustic correlates of the contrast to be discussed in §2.1) will increase as proficiency with Spanish increases and that (b) transfer from English will affect the realization of the vocalic sequences, but that as proficiency with Spanish increases, the effect of the transfer will be lessened. These hypotheses are tested through a delayed repetition task of Spanish tokens controlled for sequence type, glide and vowel quality, and presence/absence and place of articulation of a preceding consonant.

The results of the experiment suggest that the distinction between a diphthong and a hiatus does change with proficiency, but the direction of the change was unexpected. Further acoustic analysis determines that the learners are overshooting the durations of the vowel portions of the sequences, resulting in a somewhat exaggerated contrast. Second, the results indicate that distinguishing between phonetic and phonological transfer is crucial for understanding exactly what is being transferred between the L1 and the L2. Lastly, the study suggests that the salience of duration and intensity, combined with a general lack of articulatory control in reproducing the Spanish segments, is responsible for the differences between the
productions of diphthongs and hiatus by the learner groups and the native speakers. These findings are significant because they further support previous research (Colantoni & Steele 2008) that not only are perceptual factors at play during second language acquisition, but also general articulatory concerns, as well as demonstrate the need for clarity regarding the type of transfer under investigation.

The remainder of this paper will be as follows. §2 provides necessary background information on vocalic sequences in English and in Spanish. §3 situates the current study in context by discussing the previous research surrounding this topic. §4 presents details on the experiment that was conducted including the hypotheses (§4.1), information on the participants (§4.2), specifics on how the stimuli were chosen (§4.3), the experimental procedure (§4.4), and notes on how the data were analyzed (§4.5). §5 evaluates the hypotheses in light of the results of the experiment and discusses the implications, while §6 summarizes the conclusions and points out areas for future research on this topic.

2 Vocalic Sequences

To investigate the production of vocalic sequences by English speaking L2 learners of Spanish, it is important to be clear about the nature and distribution of these sequences in both English and Spanish. §2.1 discusses the notions of two different kinds of vocalic sequence: the diphthong and the hiatus. §2.2 describes where these sequences occur in English and §2.3 looks at their distribution in Spanish. The two languages will be discussed in terms of their use of diphthongs and hiatus both with and without a preceding consonant.

2.1 Diphthongs and Hiatus

Sequences of two vowels are very common in Spanish. They have been a point of interest in the literature due to the fact that there exists a contrast between the realization of a sequence of two vowels as a diphthong (a single syllable) and as a hiatus (two syllables) when at least one of the vowels is high (Harris 1969, Hualde 1991, 1997, 1999, 2004, Hualde & Prieto 2002, Roca 1997, Aguilar 1999, etc.). One of the challenges of studying these vocalic sequences is that while there is a phonological definition of the two categories, a phonetic definition is less precise. The phonological definitions classify a diphthong as a monosyllabic glide-vowel or vowel-glide sequence such as [je], [wa] or [oj] and identify a hiatus as a bisyllabic sequence of two vowels such as [i.e], [u.a] or [o.i]. A general rule of syllabification proposes that a vocalic sequence will be realized as a hiatus when the high vowel is stressed and will be realized as a
diphthong when the high vowel is unstressed (Borzone de Manrique 1979, Quilis 1993). A phonetic definition of these categories is tied up in what it means to be a glide or a vowel, but Lehiste & Peterson (1961) propose that a diphthong is “a vocalic syllable nucleus containing two target positions”. A hiatus, following this definition, would also contain two target positions, but would consist of two syllable nuclei.

In order to try and determine whether there are acoustic correlates of a diphthong versus a hiatus that can be measured and quantified in the speech of native speakers, previous experimental studies have investigated in what ways the production of these sequences corresponds to the intuitions of native speakers. The results of these studies have indicated that there are three main phonetic differences between a diphthong and a hiatus. First, a hiatus tends to be longer than a diphthong (Aguilar 1999, Hualde & Prieto 2002, MacLeod 2007). Second, a hiatus has a faster rate of change of the second formant (F2) in the transition than does a diphthong (Aguilar 1999, MacLeod 2007). Lehiste & Peterson (1961: 273) define the rate of change as “the frequency range in cycles per second through which the formant moves in a given time interval.” Third, the proportion of the sequence occupied by the transition is greater in a diphthong than in a hiatus in native speech (Lehiste & Peterson 1961, MacLeod 2007). The studies just mentioned have shown that there is a strong correlation between the general rule of syllabification of vocalic sequences discussed earlier and these three acoustic properties of the sequences. As a result, these three properties can be used to compare the productions of diphthongs and hiatus across different dialects or proficiencies of Spanish, across varying rates of speech, and in different speech styles or registers. The two acoustic correlates that will be included in the present study are duration and %-transition. F2-slope was not included due to recent research calling into question its reliability as a correlate of the diphthong/hiatus contrast (Colantoni & Limanni 2008).

Acoustic studies of Spanish diphthongs have also found that the formant values of the high vowels /i, u/ appearing in a rising diphthong are variable depending on the position of the vowel. Borzone de Manrique (1976) examined positional variants of these two vowels in three contexts including in an initial rising diphthong (as in *hielo* /ielo/ [ˈje.lo] ‘ice’ or *huelga* /uelga/ [ˈwel.ya] ‘strike’), and in a rising diphthong following a consonant (as in *viejo* /viexo/ [ˈbje.xo] ‘old’ or *suave* /suabe/ [ˈswa.ße] ‘smooth’). She found that when /i/ appeared in a word-initial diphthong it had a lower mean F1 and a somewhat higher mean F2 than when it appeared in a

---

1 Other research (eg. Garrido 2007) has considered vocalic sequences in which neither vowel is high, as in a word such as *pelear* /pelear/ ‘to fight’, variably realized as [pe.le.ˈar] or [pe.ˈlar] depending on dialect, but these sequences will not be included in the present study.

2 Lehiste & Peterson (1961: 277)
diphthong with a preceding consonant. In contrast, when /u/ appeared in a word-initial diphthong it had lower mean F1 and F2 than when it appeared in a diphthong with a preceding consonant. In addition to examining the formant values, Borzone de Manrique also calculated the F2 rate of change and compared it across the two vowels and contexts. She found that there was a faster rate of change in the diphthongs in initial position than there was in the diphthongs following a consonant for both /i/ and /u/.

2.2 English

Phonologically, English is typically described as having falling diphthongs, such as in *time* ['tajm] or *boy* ['boj], but not rising diphthongs. However, phonetic sequences of a glide plus a vowel (GV sequences) are found in English, such as [wi] in *twist* ['twist] or [je] in *yell* ['jel], that have similar spectral characteristics to the rising diphthongs found in Spanish. True hiatus do not seem to be found in English since a glide segment occurs between two vowels that would otherwise be in hiatus such as in *piano* [pi.'jæ.no] or *duet* [du.'wet].

GV sequences containing the front glide [j] and the back glide [w] without a preceding consonant (V contexts) are common in English such as in *yes* ['jes], *you* ['ju], *what* ['wat] and *when* ['wen]. However, when there is a consonant preceding these GV sequences (giving CGV sequences) the distribution of the front glides is severely restricted. Front CGV sequences are only found in English when the GV sequence is [ju], such as in *beauty* ['bju.ri] or *cute* ['kjut]. There are no other front CGV sequences in English. Tokens such as *['tjen] or *['pjast] are not possible. On the other hand, back CGV sequences do exist and are found in words such as *dwindle* ['dwin.dəl], *swam* ['swæm] and *quick* ['kwik], but not when the consonant is a labial such as in *['pwm] or *['fwal] (Davis & Hammond 1995, Ohala & Kawasaki-Fukumori 1997). Examples of English tokens containing GV sequences are provided in (1) below.

<table>
<thead>
<tr>
<th>Glide-Vowel Sequences</th>
<th>Front</th>
<th>Back</th>
</tr>
</thead>
<tbody>
<tr>
<td>V Contexts</td>
<td></td>
<td></td>
</tr>
<tr>
<td>yawning ['jæ.nɪŋ]</td>
<td></td>
<td>waiting ['weɪ.jiŋ]</td>
</tr>
<tr>
<td>yogurt ['jʊ.ɡəɹt]</td>
<td></td>
<td>winter ['wɪn.təɹ]</td>
</tr>
<tr>
<td>CV Contexts</td>
<td>none</td>
<td>twister ['twɪs.təɹ]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>sweden ['swi.dən]</td>
</tr>
</tbody>
</table>

Examples of English words containing GV sequences for /w/ and /j/ with and without a preceding consonant

---

3 In loanwords, such as *fjord* ['fɪəd], other sequences are found, although they are rare (Major 1987).
While homorganic sequences of a consonant and a back glide such as [pw], [bw] or [mw] are not found in onset position in English words, they do occur across syllable boundaries in compounds and across word boundaries in running speech. For example, the compound *ribwich [ˈrɪb.wɪtʃ] and the phrase *Up with cheese! [ʌp wɪtʃ] contain a sequence of a labial consonant followed by the labiovelar glide. Even though the two elements in the sequence would not be considered to be a branching onset, they do occur adjacent to each other and need to be produced in linear order by English speakers. Similarly, although front CGV sequences are limited to those where the GV sequence is [ju] as mentioned above, sequences with a vowel other than /u/ are found in compounds such as *leap-year [ˈlip.juː] and in phrases such as lock your bike [lɑk jʊ ɔɹ bʌjk]. Additionally, certain sequences that are illicit in the onset of a stressed syllable can be found as the onset of an unstressed syllable. For example, the word *[tʃæn] is not possible, yet we find a remarkably similar syllable word-finally in the word *Christian [ˈkɹɪs.tʃən]. When produced carefully, the last syllable in Christian contains the sequence [tʃ] which is normally considered illicit in onset position. However, in faster or more casual speech, the sequence [tʃ] is often realized as [tʃ] as a result of affrication (Lawrence 2000).

In sum, there is a dichotomy between the phonological and the phonetic distributions of CGV sequences in English. While phonologically, their distributions are quite limited, phonetically these sequences do occur in compounds and across word boundaries as shown above.

2.3 Spanish

Phonological descriptions of the vocalic elements of Spanish generally posit that the inventory contains both rising and falling diphthongs (as in siete ['sjɛ.te] ‘seven’ and seis ['sejs] ‘six’, respectively) as well as hiatus (as in río ['ri.o] ‘river’ and púa ['pu.a] ‘thorn, barb’) (Harris 1969). As mentioned above, English is usually analyzed as having GV sequences rather than true rising diphthongs. In contrast, many analyses of Spanish propose that the glides [j] and [w] are allophones of the high vowels /i/ and /u/ respectively and that the glide variants surface when the high vowels are adjacent to a stressed vowel (Roca 1991, Harris & Kaisse 1999) such as in piara /piara/ ['pja.ɾa] ‘herd’ or cual /kual/ ['kwal] ‘which’. In these cases, the glide is syllabified as part of the nucleus, rather than part of the onset, as in English\(^4\) (English: Gick 1999).

\(^4\) This analysis is not without criticism. Mendez (1998) discusses morphological resyllabification as a diagnostic for syllabic structure and argues that Spanish prevocalic glides are syllabified as part of the onset, rather than the nucleus.
The distribution of vocalic sequences in Spanish is less restricted than that in English with Spanish containing diphthongs and hiatus in both CV (with a preceding consonant) and V (without a preceding consonant) contexts. The only real restriction in the distribution is with hiatus containing the back vowel /u/ in V contexts. A word such as *[o.ˈu.a] is not found in Spanish. Examples of Spanish tokens in the other environments are shown in the table in (2) and (3) below, with the relevant sequences bolded.

### (2) Diphthongs

<table>
<thead>
<tr>
<th></th>
<th>Front</th>
<th>Back</th>
</tr>
</thead>
<tbody>
<tr>
<td>V Contexts</td>
<td></td>
<td></td>
</tr>
<tr>
<td>hierro</td>
<td>[je.ro] 'iron'</td>
<td>hueso ['we.so] 'bone'</td>
</tr>
<tr>
<td>hiedra</td>
<td>[je.ðra] 'ivy'</td>
<td>huevo ['we.ðo] 'egg'</td>
</tr>
<tr>
<td>CV Contexts</td>
<td></td>
<td></td>
</tr>
<tr>
<td>viaje</td>
<td>['bja.xe] 'trip'</td>
<td>juego ['xwe.yo] 'game'</td>
</tr>
<tr>
<td>tienda</td>
<td>['tjen.da] 'store'</td>
<td>cuando ['kwan.do] 'when'</td>
</tr>
</tbody>
</table>

Examples of Spanish words containing front and back diphthongs with and without a preceding consonant

### (3) Hiatus

<table>
<thead>
<tr>
<th></th>
<th>Front</th>
<th>Back</th>
</tr>
</thead>
<tbody>
<tr>
<td>V Contexts</td>
<td></td>
<td>none</td>
</tr>
<tr>
<td>oía</td>
<td>[o.ˈi.a] 'hear.3sg.past'</td>
<td></td>
</tr>
<tr>
<td>caía</td>
<td>[ka.ˈi.a] 'fall.3sg.past'</td>
<td></td>
</tr>
<tr>
<td>CV Contexts</td>
<td></td>
<td></td>
</tr>
<tr>
<td>día</td>
<td>['di.a] 'day'</td>
<td>sittúa ['si.ˈu.a] 'place.3sg.past'</td>
</tr>
<tr>
<td>tenía</td>
<td>[te.ˈni.a] 'have.3sg.past'</td>
<td>ría ['ru.a] 'street'</td>
</tr>
</tbody>
</table>

Examples of Spanish words containing front and back hiatus with and without a preceding consonant

## 3 Literature Review

This study examines the acquisition of the diphthong/hiatus contrast in three respects. First, the overall implementation of the contrast (in terms of the acoustic correlates duration and %-transition) is compared across the three speaker groups (beginning learners, intermediate learners, and native speakers) to find out whether the learners are able to produce the contrast in a native-like way. Second, the effect of transfer of the L1 phonological constraint against homorganic CGV sequences is investigated. Third, the glides of the two languages are evaluated in terms of their relative intensity in order to determine whether phonetic transfer of the English norms for glide production is occurring. The purpose of this section is to provide a context for this research within the literature as regards each of the three viewpoints mentioned above.
Expanding on the first respect – the overall realization of the diphthong/hiatus contrast – many studies have looked at the contrast in native speakers, some of which were referred to in §2.1, but the purpose of those studies was to investigate the so-called exceptional hiatus (Hualde 1999, Face & Alvord 2004, Cabré & Prieto 2006) and to establish phonetic correlates of the phonological contrast between diphthongs and hiatus (Aguilar 1999, Hualde & Prieto 2002). None has investigated how the diphthong/hiatus contrast is produced and acquired by L2 learners of Spanish.

While no previous work has studied L2 vocalic sequences, earlier research has considered how individual Spanish vowels are produced by learners (Reeder 1998, Morrison 2003, Morrison 2006, Vokic 2005, Menke & Face 2008). Morrison (2003) tested native English speakers who had only studied Spanish in a classroom setting for fewer than 5 years and were classified as relatively inexperienced. The participants read 5 sentences, each containing one of the 5 Spanish vowels, 3 times in random order. Morrison found that the learners substituted the English vowels /i/, /e/, /æ/, /o/ and /u/ for Spanish /i/, /e/, /a/, /o/ and /u/ respectively, where the nature of the phonemes in question was defined via F1 and F2 norms as determined through collection of native speaker data. Menke & Face (2008) conducted a study that included three groups of L2 Spanish speakers: 20 fourth semester university students, 20 Spanish majors who were graduating shortly, and 13 PhD students in Spanish. The participants read a short story aloud and were recorded. The researchers found a general expansion of the F1/F2 vowel space as experience increased. Specifically, the front vowels became more fronted and the back vowels were produced farther back as the level of proficiency of the learners rose. They also found that unstressed vowels were reduced in that they were realized as more central by each group of learners. No native control group was included, so while the development of the Spanish vowels can be seen as proficiency increases, Menke & Face were unable to determine whether the learners of the highest level tested were converging on the native speaker norms or not. A control group of 5 native Spanish speakers was included in Reeder (1998), however, and their productions of Spanish vowels were compared to those by L2 learners of Spanish at four different levels of proficiency. 70 native English speakers participated in the study: 20 level-1 or beginning learners enrolled in their first Spanish class, 20 level-2 or intermediate learners enrolled in a third semester Spanish class, 20 level-3 learners taking an upper year undergraduate or masters level Spanish class, and 10 level-4 speakers who were employed as Spanish language teachers. The methodology included both a sentence reading task and a picture naming task. Acoustic analysis of the collected data determined that both the accuracy and precision of the production within the F1/F2 space.
improved as experience increased and approached the formant values found in the native speaker control group.

As mentioned above, the second way in which the L2 production of Spanish vocalic sequences is investigated in the present study relates to the effect of transfer of the L1 phonological constraint banning homorganic CGV sequences. Transfer is an important notion in L2 acquisition research, referring to the interference of the L1 during acquisition of the L2. In terms of phonological acquisition, this interference can take many forms such as segmental substitutions, changes in prosody, and alterations to syllable structure stemming from L1 phonotactic constraints (Vokic 2005, Major 2008). The effects of transfer are often discussed in relation to loanword phonology, but are also very relevant in a study of L2 acquisition.

Although transfer is a very important concept in the second language literature, it is often left unclear exactly what is being transferred from the L1 to the L2. It is essential at this stage to clarify the difference between phonological and phonetic transfer, as this distinction will figure prominently in the following discussion. Markham (1997) defines phonological acquisition of a second language as “the process of establishing a system of abstract productional or perceptual constructs (categories) used in opposition and showing functional equivalence to the language norm.” (Markham 1997: 16) If this defines phonological acquisition, then the notion of phonological transfer can be said to be the transfer of the existing L1 system of abstract production or perceptual constructs onto use of the L2. In other words, phonological transfer involves using the phonological status of a segment or a contrast in the L1 during production or perception of a segment or contrast deemed equivalent in the L2, or transferring syllable structure or phonotactic constraints from the L1 for use during production or perception of the L2. An example of the outcome of phonological transfer is vowel epenthesis between two consonants in an English onset cluster in order to avoid violating syllable structure constraints in Japanese which disallow CCV syllables (Major 1987). The transferred entity in this case is the syllable structure constraint. On the other hand, Markham defines phonetic acquisition of a second language as “the process of establishing surface performance in production and perception showing functional equivalence to the phonetic language norm”. If this is the definition of phonetic acquisition, then phonetic transfer can be said to be the transfer of the existing L1 surface performance norms onto use of the L2. In other words, phonetic transfer encompasses transfer of the acoustic realization of a particular segment with no reference to its phonemic status, participation in an opposition, or position within a domain (syllable, word, phrase, etc) in either the L1 or the L2. An example of the outcome of phonetic transfer is the realization of the Spanish dental stops /t, d/ as alveolar by
English speakers resulting from the transfer of the English norm for production of /t, d/ as alveolar.

A wide range of studies have found that second language learners will frequently use their native language settings, including segmental and syllabification norms, when producing sounds in the L2. Flege & Davidian (1984) examined the production of word-final obstruents by a group of 12 native speakers of Spanish who had arrived in the United States at an average age of 20 and had lived there for 7 years on average. They all spoke English fluently, but with an obvious accent. Word-final obstruents are somewhat rare in Spanish and are realized as fricatives (Harris 1969). The researchers used a picture-naming task to elicit 10 tokens ending in a stop. The Spanish speakers were found to fricativize word-final stops significantly more often than the English, Chinese, and Polish controls as a result of a transfer process from Spanish. McAllister, Flege & Piske (2002) tested the ability to learn the Swedish vocalic quantity contrast by non-native speakers with three different first languages: Estonian, English, and Spanish. 20 native speakers of each language participated; they had all begun learning Swedish as adults, had lived in Sweden for at least 10 years, and reported using Swedish on a daily basis. The participants produced 40 different tokens aloud by means of a picture naming task since the quantity distinction is reflected in Swedish orthography. A durational analysis of the vowels suggested that the Estonian speakers were most successful at acquiring the length contrast and that the English speakers were more successful than the Spanish speakers. The researchers conclude that this is due to the fact that duration plays a more important role in Estonian than it does in English and in turn is more important in English than in Spanish.

Transfer of syllable structure is evident in the production of English sC clusters by native speakers of Brazilian Portuguese where the target cluster is often realized via prosthesis in which an epenthetic [i] appears before the cluster (Cardoso 2008). Similarly, Abrahamsson (1999) found in his longitudinal study of a native speaker of Spanish acquiring Swedish that epenthesis of [e] preceding an sC cluster occurs in both elicited and conversational speech. Broselow & Finer (1991) examined the acquisition of English syllable coda clusters by 24 native Korean and 8 native Japanese speakers from the perspective of sonority in order to evaluate the relative effect of transfer and markedness. They defined a more marked coda cluster as one in which the consonants are closer in sonority. They tested the production of codas in which the degree of markedness varied and found that the learners seemed to arrive at a sonority setting that is intermediate to the L1 and the L2, indicating that transfer and markedness were interacting during acquisition.

In a larger domain, Broselow (1984) discovered transfer of L1 syllable structure at the word juncture in her study of the errors found in the speech of English speakers acquiring
Egyptian Arabic. Broselow concluded that differences in syllable structure between the two languages cause the learners to alter the word boundaries when producing phrases in Egyptian Arabic.

The above examples all show the negative effects of transfer in that they cause the learner to make errors in the L2. However, there are also instances where the outcome of transfer is positive. Major (1987) tested production of English onset clusters by 5 native speakers of Japanese who had an intermediate level of proficiency with English (as measured by their scoring between 400 and 450 on the TOEFL test) and who had been living in the United States for 2-4 months. There were three tasks: a word list, a reading passage, and a short conversation. Major found that the Japanese speakers were relatively proficient at producing fricative+s clusters, even though Japanese does not have underlying consonant clusters. He attributed this to the fact that Japanese has surface clusters, resulting from the devoicing and deletion of /i/ and /u/ between two voiceless obstruents. The Japanese example given by Major is /sukiaki/ which becomes [şuкиaki] as a result of devoicing of /u/ which then becomes [skiaki] due to deletion of the voiceless vowel. Transfer of this pair of rules gives a native-like realization of an English word such as *spy /spaj/ which becomes [şupaj] due to insertion necessary to avoid syllable structure violations, which then becomes [şupaj] from devoicing and then [spaj] from deletion, giving the English pronunciation of the word. Of course, this transfer is not always positive, since it applies to other English words such as *city giving the non-native-like realization [sti] (Major 1987).

The effect of transfer can change over time, as noted by Major (1987b), Hansen (2004) and Zampini (2008), and we can see changes in the behaviour of learners as they gain experience with the L2. Hansen (2004) examined the developmental sequence associated with the acquisition of English L2 syllable codas by two native speakers of Vietnamese who had arrived in the United States 6 months before the onset of the study. Data were collected through a series of interviews spanning a one-year time period. It was found that the acquisition of codas was affected by transfer from the L1, developmental processes, and grammatical conditioning. Production of codas progressed by first using the L1 norms and allowing segments in the coda that are possible in the L1, next by gradually introducing new segments into the coda. The order in which these segments were introduced was largely affected by developmental and markedness constraints.

However, not all deviations from L2 norms can be attributed to transfer. Flege (1980) found that native speakers of Arabic learning English did have some effect of phonetic norms of Arabic, but that they were also approximating the English norms. He attributes this approximation to the gradual learning of the articulatory control needed to produce L2 sounds.
accurately. This shows that in addition to dealing with interference effects from the L1, learners must also develop the necessary motor controls to correctly produce L2 segments. Colantoni & Steele (2007) uncovered a similar result in their study of the acquisition of French /ʁ/ by intermediate and advanced English-speaking learners of French in that the learners, lacking the specific articulatory control necessary to produce both manner and voicing, focused on one of the parameters at a time.

As mentioned above, the third respect in which the diphthong/hiatus contrast is investigated in this study concerns the physical characteristic of relative intensity (RI) of the glides in English and Spanish. Relative intensity here is defined as the difference between the intensity of the glide and the intensity of the following vowel. The intensity of glides is generally lower than that of vowels, resulting in a negative RI for glides (Balakrishnan et al. 1996). Although it is normally assumed that the front and back glides of English (/j/ and /w/) are very similar to the front and back glides of Spanish ([j] and [w]), previous research into the RI of the glides in these two languages has determined that in fact they are quite different. MacLeod (2008) found that the Costa Rican Spanish labiovelar glide [w] had a significantly higher RI than the Canadian English /w/ and that the mean RI value for the English /w/ was actually comparable to the value of the voiced velar stop /g/ in Spanish. This suggests that the glides in Spanish are more vocalic than their English counterparts, a result that lends support to the phonemic analysis of Spanish glides as allophones of high vowels (Roca 1991, Harris & Kaisse 1999). The effect of the difference in RI between the English and Spanish glides was that native speakers of Costa Rican Spanish perceived the English back glide /w/ as the sequence [gw] and produced it as such. This shows that differences in RI can have significant effects on the perception and production of L2 sounds and could certainly contribute to foreign accent in the speech of L2 learners.

4 Experiment

An experiment was designed and conducted in order to examine two aspects of the acquisition of the diphthong/hiatus contrast by English-speaking learners of Spanish. The first aspect reflects the manner in which learners of Spanish produce the contrast as compared to the way native speakers of Spanish produce it and how the learners’ production of the contrast changes with increased experience with the language. As discussed in §2.2, previous research has shown that, in the speech of native speakers of Spanish, a hiatus tends to be longer than a diphthong and that the proportion of the sequence that is occupied by the transition tends to be smaller in a hiatus than in a diphthong. This study will measure the durations of the first vowel
(or glide, in the case of diphthongs), the transition, and the second vowel in order to discover how beginning learners and intermediate learners of Spanish produce diphthongs and hiatus and compare these values to those of native speakers.

There are three important and intertwined factors that are predicted to affect the way in which the diphthong/hiatus contrast is produced by learners. The first factor, already alluded to, is experience. By testing two groups of learners, beginner and intermediate, this study aims to observe how production of diphthongs and hiatus changes with increased experience with Spanish. The second factor is the environment in which the vocalic sequence is found. In this study, the two environments that were tested were those in which the sequence occurred with a consonant in the onset and those in which the sequence occurred without an onset. The third factor is the quality of the first element in the sequence. The first element is one of the following: the front vowel /i/, the front glide [j], the back vowel /u/, or the back glide [w]. As discussed in §2.1, consonant-glide-vowel sequences in which the consonant and glide share place of articulation are not found in stressed, onset position in English, but they are found in this position in Spanish such as in bueno /bueno/ ['bwe.no] ‘good’. This asymmetry is predicted to affect the production of Spanish CGV sequences in that those that are homorganic will be realized with the GV portion in hiatus, giving CV.V, whereas those that are non-homorganic will be realized in a more native-like way.

The second aspect of the acquisition of the diphthong/hiatus contrast by English-speaking learners of Spanish that is examined here focuses on the glide portion of diphthongs. As noted in §3, previous work has shown that while the initial sounds in the Spanish word huevo /uebo/ ['we.ßo] ‘egg’ and the English word wade ['wejd] are described as the labiovelar glide and are both transcribed using the same IPA symbol, there are significant phonetic differences in their realization with respect to relative intensity (MacLeod 2008). The present study measures the relative intensity of both the front glide [j] and the back glide [w] in Spanish words as produced by the two groups of learners and compares it to the values found in the speech of native speakers of Spanish. This comparison is furthered by measuring the relative intensities of the glides in the production of English words by native English speakers in order to determine to what extent are the learners using English relative intensity values in their production of Spanish glides and how this production changes as the learner gains knowledge of the second language.

The experiment addresses several specific hypotheses, which are presented in §4.1. Details regarding the participants who took part in the study are found in §4.2. The stimuli

---

5 In the case of diphthongs, these environments are clear as in [tjén.da]. In the case of hiatus, an onset means that the consonant is in the onset of the first syllable or vowel in the sequence as in [dí.a].
used in the experiment included both English and Spanish tokens and information concerning how they were selected is provided in §4.3 while the actual procedure of the experiment is explained in §4.4.

4.1 Hypotheses

There are three hypotheses that were tested in this study. The first concerns the general production of the diphthong/hiatus contrast by native speakers and learners of Spanish as is found in (4) below.

(4) The mean difference between a diphthong and a hiatus, in terms of duration and %-transition, will increase as proficiency with Spanish increases. This implies that the mean difference in duration and %-transition between the two sequence types will be smallest in the beginning learners, greater in the intermediate learners, and largest in the native speakers.

The second hypothesis addresses the effect of transfer of the phonological syllabification constraint against homorganic CGV sequences on the production of Spanish diphthongs.

(5) Phonological transfer of the English constraint against homorganic CGV sequences will result in the diphthongs in which the consonant and glide share place of articulation being produced in a less native-like way in that the learners will produce the glide portion of the sequence with a longer duration than the NSs, producing a sequence of a consonant and a vowel, rather than a consonant and a glide in order to avoid violating the constraint. The CV diphthongs in which the consonant and the glide do not share place of articulation will be realized in a more native-like manner than those where they do. Further to this, it is predicted that the effect of the transfer of this constraint will weaken as proficiency with Spanish increases. As a consequence, we expect to find a negative correlation between proficiency in the IL group and duration of the glide segment. As proficiency increases, the duration of the glide segment in CV diphthongs where the consonant and glide share place of articulation will decrease and approach the native Spanish norms.
The third and final hypothesis refers to phonetic transfer of the acoustic realization of the glides.

(6) Beginning learners will use the English realization of both front and back glides (characterized by a lower relative intensity than the Spanish counterparts) as a result of phonetic transfer while the intermediate learners will produce glides with more native-like relative intensities as a result of their greater proficiency with Spanish. Within the intermediate learners we will see a positive correlation between relative intensity and proficiency in that as proficiency increases, so does relative intensity, changing the RI values from the relatively low English norms up to the relatively high Spanish norms.

4.2 Participants

In order to test the predictions specified in the preceding section, an experiment was conducted that collected data from four groups of participants: beginning learners of Spanish, intermediate learners of Spanish, native speakers of Spanish, and native speakers of English. All of the participants were recruited through posters at the University of Toronto and the majority are students at the university. All who took part in the study were 18 years of age or older and had either completed university education or were in the process of completing it. The relevant characteristics of the participants are summarized in (7). Both the beginning and intermediate learners of Spanish were native speakers of North American English. The 8 beginning learners were all enrolled in their first Spanish class and took part in the experiment after having completed 3 weeks of class (approximately 15 hours of instruction). The 10 intermediate learners were enrolled in a 3rd or 4th year Spanish class and have an average of 4.5 years of experience with the language. 7 of these 10 learners had only received classroom instruction in Spanish, while the remaining 3 had spent time in a Spanish-speaking country. Specifically, participant NA spent 5 weeks learning Spanish in Mexico, HE spent one summer in Peru, and JO lived in Ecuador for 5 years.

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Average Age</th>
<th>Average Years Experience</th>
<th>Type of Exposure to Spanish</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beginning Learners</td>
<td>8</td>
<td>21</td>
<td>0</td>
<td>15 hours in 1st Spanish class</td>
</tr>
<tr>
<td>Intermediate Learners</td>
<td>10</td>
<td>22</td>
<td>4.5</td>
<td>mostly classroom, 3 participants with experience abroad</td>
</tr>
<tr>
<td>Native Spanish</td>
<td>5</td>
<td>29</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Native English</td>
<td>10</td>
<td>21</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Since the intermediate group includes a greater variety of experiences with Spanish than the beginners, members of this group were recorded reading a passage entitled *El Viento del Norte y el Sol* (provided in Appendix B). In order to assign a proficiency value to each speaker, the readings by the learners were presented to two native speaker judges\(^6\) of Spanish, along with the same reading by the 5 native speaker participants, in random order. The native speaker judges assigned a proficiency value from 1 (very foreign sounding) to 5 (native sounding) to each recording\(^7\), following the methodology of Colantoni & Steele (2007). The experience and mean proficiency for each intermediate learner is shown in (8). While there is a range of proficiencies from 1 to 4, eight of the ten speakers in the IL group have a proficiency rating of less than 3. These values can be compared to the results of Colantoni & Steele (2007). In that study, it was found that an intermediate group of participants had a mean proficiency rating of 1.9 (ranging from 1.5-2.8), and that an advanced group had a mean rating of 3.8 (ranging from 2.8-4.5). Following these guidelines, 8 of the 10 participants in the IL group would qualify as intermediate in that they have a proficiency rating of less than 3.0, and 2 would qualify as advanced since their rating was higher than 3.0. Furthermore, the mean proficiency rating of the lower eight participants was 1.8, while the mean proficiency of the upper two was 3.8.

\[
\begin{array}{|c|c|c|c|c|}
\hline
\text{Participant} & \text{Age} & \text{Average Years Experience} & \text{Type of Exposure to Spanish} & \text{Proficiency} \\
\hline
\text{MA} & 20 & 5 & \text{classroom} & 1.0 \\
\text{PA} & 40 & 3 & \text{classroom} & 1.0 \\
\text{DA} & 20 & 3 & \text{classroom} & 1.5 \\
\text{AV} & 20 & 5 & \text{classroom} & 2.0 \\
\text{HE} & 19 & 3 & \text{classroom, and 1 summer in Peru} & 2.0 \\
\text{JN} & 18 & 7 & \text{classroom} & 2.0 \\
\text{SA} & 22 & 3 & \text{classroom} & 2.0 \\
\text{YJ} & 24 & 4 & \text{classroom} & 2.5 \\
\text{NA} & 21 & 2 & \text{classroom, and 5 weeks in Mexico} & 3.5 \\
\text{JO} & 19 & 7 & \text{classroom, and Ecuador 5 years} & 4.0 \\
\hline
\end{array}
\]

Summary of Intermediate Group Participant Information

A group of five native speakers of Spanish, all from Mexico, served as a control group for the study. This particular variety of Spanish was included primarily due to availability of speakers. The native speakers were limited to one dialect since previous research has shown...
that different varieties may utilize different acoustic or articulatory parameters to realize the contrast (MacLeod 2007).

4.3 Stimuli

As mentioned earlier, this study examined production of both Spanish and English tokens. As such, there are two sets of stimuli. The Spanish stimuli contained 90 tokens and were controlled for sequence type (diphthong or hiatus\(^8\)), quality of the first vocoid in the sequence (front, i.e. /i/ or [j], or back, i.e. /u/ or [w]) and environment of the sequence (without a preceding consonant (V contexts), or with a preceding consonant (CV contexts)). Examples of the Spanish stimuli are shown in (9) and (10) below. The target sequences are bolded.

As can be seen in (10), there are no stimuli containing a hiatus in which the first element is the back vowel /u/ without a preceding consonant. This is because there are no such words in Spanish and the stimuli in this study were limited to include only actual words and no nonse words in order to avoid altered or unnatural pronunciation.

The English stimuli included 30 disyllabic words containing either a /w/ or a /j/ in the first syllable and were controlled for the environment of the glide (without a preceding consonant (V contexts), or with a preceding consonant (CV contexts)), except for /j/ which

---

\(^8\) As predicted by the general syllabification rule (Quilis 1993) discussed in §2.1.
does not occur with a preceding consonant in English. Examples of the English stimuli are shown in (11).

<table>
<thead>
<tr>
<th>Glide-Vowel Sequences</th>
<th>Front</th>
<th>Back</th>
</tr>
</thead>
<tbody>
<tr>
<td>V Contexts</td>
<td>yawning [ˈjaː.nɪŋ]</td>
<td>waiting [ˈweɪ.tɪŋ]</td>
</tr>
<tr>
<td>yogurt [ˈjoʊ.ɡəɹt]</td>
<td>winter [ˈwɪn.təɹ]</td>
<td></td>
</tr>
<tr>
<td>CV Contexts</td>
<td>none</td>
<td>twister [ˈtwɪ.stəɹ]</td>
</tr>
<tr>
<td></td>
<td>sweden [ˈswɪ.dən]</td>
<td></td>
</tr>
</tbody>
</table>

Examples of English stimuli containing front and back glides with and without a preceding consonant

All of the stimuli are provided in Appendix A.

4.4 Procedure

The task that the participants performed was a delayed repetition task. The stimuli, as described above, were presented aurally\(^9\) followed by a 0.5 second tone. After the tone had finished, the participants repeated the word that they heard. The use of the delayed repetition task was chosen to ensure that the participants were not simply repeating the acoustic characteristics of the stimulus, but rather were making use of a more abstract level of knowledge (Werker & Logan 1985). Without the delay, there was a risk that the participants would be mimicking the voice of the native Spanish speaker who recorded the stimuli (Munro 2008). Any task that depended on orthographic representation of the tokens would be introducing two confounds: first, in that both glides and vowels are represented by the same graphemes in Spanish and second, the realization of a VV-sequence as a hiatus is cued by the presence of an accent on the high vowel. Other tasks avoiding orthographic representation, such as picture naming or a map task, presuppose a level of lexical knowledge of Spanish that the BLs certainly lacked. In order to keep the task variable constant across the three proficiency groups, it was necessary to choose a task that even the lowest level could perform. The delayed repetition task has been used successfully in the literature, including by Flege, Munro & MacKay (1995) and Tremblay (2008).

The Spanish stimuli were presented first and the English stimuli second with each group of stimuli repeated twice by each participant. The native Spanish speakers and intermediate

---

9 The Spanish stimuli were read aloud by a 29 year old female native speaker from Venezuela and the English stimuli were read aloud by a 33 year old female native speaker from Ontario.
learners also read the short reading passage, *El Viento del Norte y el Sol*, provided in Appendix B, as discussed in §4.2, in order to provide a measure of their proficiency with Spanish.

All recordings were made in the sound-attenuated booth in the University of Toronto phonetics lab using a Sound Devices 722 digital recorder and Audio-Technica AT831b lavaliere microphone with a sampling frequency of 44,100 kHz and 24-bit resolution on a mono channel.

4.5 Acoustic Analysis

Once the sound files had been recorded, an acoustic analysis was conducted in order to generate the data necessary to address the hypotheses as presented in §4.1. The two measurements that were needed were duration and intensity. The duration was measured by manually marking the beginning and the ending of the vocalic sequence using cues from the spectrogram and the waveform. Where the sequence occurred following another vocalic element (as in oía [o.'i.a] ‘hear.IMPERFECT’), the beginning of the sequence was marked by isolating the steady state of F2 in the first element. Where the sequence occurred word-initially or following a consonant, the start of the sequence was marked at the first regular vocal pulse, at the zero crossing, and the end of the sequence was marked at the offset of F2 (Chitoran 2002, Limanni 2008).

The segmentation of the sequences is exemplified in the spectrogram for the word *tiara* /tía/ ['tja.ɾa] ‘tiara’ in (8) below. The entire sequence is found between the bars marked 1 and 4. The first element, in this case the glide [j], is marked between bars 1 and 2, while the second element, the vowel /a/, is marked between bars 3 and 4. The first and second elements were isolated by marking the beginning and ending of a steady state in the second formant. The transition appears between bars 2 and 3. The durations of the each section of the sequence (first element, transition, and second element) as well as the total duration of the sequence were recorded. These durational measurements were taken for all of the stimuli including both diphthongs and hiatus in Spanish as well as all of the English stimuli. The proportion of the sequence that is occupied by the transition (%-transition) was calculated by dividing the duration of the transition by the duration of the entire sequence.

In addition, in the case of Spanish diphthongs and in the English stimuli, the intensity of the first element (the glide) was measured along with the intensity of the following vowel in order to calculate the relative intensity. The intensity line is the curving line marked on the right hand side of the spectrogram below. Intensity was measured during the lowest point in production of the glide and at the highest point of the following vowel. The relative intensity
was then calculated by taking the difference between these two vowels, normally resulting in a negative value (Balakrishnan et al. 1996).

5 Hypotheses Evaluation & Discussion

This section considers the results of the study from the perspective of each of the hypotheses in order to determine whether the hypotheses were confirmed or not and what conclusions can be drawn.

5.1 Hypothesis 1: The Diphthong/Hiatus Contrast

The first hypothesis for this study concerns the phonetic manifestation of the diphthong/hiatus contrast. The hypothesis is restated in (13) below.

(13) The mean difference between a diphthong and a hiatus, in terms of duration and %-transition, will increase as proficiency with Spanish increases. This implies that the mean difference in duration and %-transition between the two sequence types will be smallest in the beginning learners, greater in the intermediate learners, and largest in the native speakers.
As was discussed in §2.1, two of the acoustic characteristics of hiatus are that they tend to be longer than diphthongs and that they tend to have a smaller %-transition than diphthongs. The results of the experiment will be evaluated first in terms of how the diphthong/hiatus contrast is expressed through duration and next through %-transition.

The extent to which comparing raw durations is useful is limited, particularly when comparing durations of speech produced by disparate groups such as beginning learners and native speakers since it is to be expected that beginning learners would speak more slowly than native speakers due to their inexperience with the language. In order to better facilitate cross-group comparison of durations, the speech rates were normalized across all 23 speakers. This was accomplished by calculating the average syllable duration for each speaker and comparing them. A one-way ANOVA found that there were significant differences in average syllable duration between the participants. In order to normalize the speech rates, the participant with the slowest speech rate was identified (GP1-1, speech rate = 347ms/syllable) and all of the durational measurements for the other 22 participants were multiplied by a factor appropriate to each to bring their mean syllable durations into line with GP1-1. The mean syllable durations and normalizing factors for each participant can be found in Appendix C.

The first hypothesis, presented in (13) above, predicts that the difference between a diphthong and a hiatus will increase as proficiency with Spanish increases. If diphthongs and hiatus were more distinct in the speech of NSs than of learners, we would expect that the difference in mean duration between the two sequence types would be greater in the speech of NSs than in the speech of learners, thereby creating a more noticeable separation between the two. Specifically, we expect that the BLs will produce diphthongs and hiatus with the smallest difference in duration, the ILs will produce the two sequence types with a larger difference, and the NSs will have the greatest difference. As we can see in the chart in (14) below, however, this situation was not found.

<table>
<thead>
<tr>
<th></th>
<th>Mean Diphthong Duration</th>
<th>Mean Hiatus Duration</th>
<th>Raw Difference</th>
<th>Proportional Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>BLs</td>
<td>306</td>
<td>449</td>
<td>143</td>
<td>47%</td>
</tr>
<tr>
<td>IIs</td>
<td>309</td>
<td>452</td>
<td>143</td>
<td>47%</td>
</tr>
<tr>
<td>NSs</td>
<td>311</td>
<td>426</td>
<td>115</td>
<td>38%</td>
</tr>
</tbody>
</table>

Mean durations in ms. and proportional difference in duration by proficiency group for diphthongs and hiatus.

The raw difference in normalized duration between a diphthong and a hiatus is actually smaller in the speech of the NSs (115ms) than it is in the speech of either the BLs (143ms) or the IIs (143ms). Additionally, we can calculate the proportional difference in duration between the
two sequence types within each speaker group, shown in the final column of the chart in (14). Here we find that there is a larger proportional difference in duration between diphthongs and hiatus in the speech of BLs and ILs than in NSs. The beginning and intermediate learners produce hiatus that are 47% longer than diphthongs, on average, while the hiatus produced by the native speakers are only 38% longer than their diphthongs. Looking at the values in the chart in (14) in graph format in (15), we find that the BL and IL groups are producing the contrast in a remarkably similar way, while the NS group produces diphthongs similarly to the other two groups, but produces hiatus with a lower mean duration. The gap between the distributions of the two sequences types in the speech of the BLs and the ILs is larger than that found in the speech of the native speakers. Ultimately then, it seems that, in terms of duration, the BLs and ILs are producing the contrast with more of a difference between the two sequence types than are the NSs, which is the opposite of what was predicted.

While native English speakers do not have experience from their L1 with a contrast between diphthongs and hiatus, it seems that the difference between these two sequences is quite salient, possibly due to the fact that one of the main manifestations of the contrast is the durational difference and duration is one of the acoustic correlates of stress in English (Fry 1955, Beckman 1986). Since stress is used contrastively in English, native speakers would need to be aware of it. This awareness of stress seems to make the diphthong/hiatus contrast salient, even to beginning learners. The wide gap between the durations of the two sequence types found in the speech of the learners could reflect the sensitivity to changes in stress, coupled with a general lack of precision in producing that change in Spanish. In a sense then, the learners are exaggerating the contrast by overshooting the target durations of hiatus and, as was seen in the
graph in (15), there is a larger gap between the mean durations of diphthongs and hiatus in the speech of BLs and ILs than that in the speech of NSs.

To further compare the production of the diphthong/hiatus contrast by the three proficiency groups, we can examine the composition of the sequences in terms of the mean normalized durations of the first element in the sequence (a glide in the case of a diphthong and a vowel in the case of a hiatus), the second element (a vowel), and the transition between the two. Although the mean durations of diphthongs across proficiency groups were fairly stable as we saw in (15), if we focus on the duration of the first element in the sequence, as shown in (16), we see that both the BLs and the ILs produce glides that are somewhat shorter than the NSs, while the first vowel in a hiatus tends to be somewhat longer in the speech of both learner groups than in the NSs’.

If we compare the mean durations of the second element in the sequence as in (17), we see that both learner groups consistently produce longer vowels in both diphthongs and hiatus than do the NSs. Taken together, these two results suggest that the BLs and ILs produce Spanish vowels with a greater duration than the native speakers, even when speech rate is controlled for. This lends further support to the idea that the learners lack precision in producing these vocalic sequences as discussed earlier.
In addition to investigating the durations of the first and second elements in the sequence, we can also look at the duration of the transition between the two. If we compare this duration by proficiency group and sequence type, we see in (18) that as proficiency increases, the duration of the transition increases (except for a slight deviation in the IL group’s production of hiatus).

As discussed in §4.2, the experiences of the participants in the BL group with respect to Spanish are very similar, and all of the native speakers are from Mexico. However, in order to evaluate the variability found within these groups across speakers, we include a plot of the normalized mean durations of diphthongs and hiatus by speaker in (19).
In the first 8 participants on the left, corresponding to the BL group, we see that most show fairly similar behaviour in terms of their production of the diphthong/hiatus contrast, with the exception of BL1 who produces a much smaller durational difference and BL3 who produces a much larger durational difference than the other beginners. The NS group is quite stable, particularly in terms of their hiatus durations. This falls into line with the finding in MacLeod (2007) that there is a greater variability associated with the production of diphthongs than with hiatus. The members of the BL and IL groups show variability in their durations of both sequence types and it is clear that all but one of the learners produces hiatus with greater duration than any of the native speakers. The hypothesis presented in (13) above predicted that the distinctness between diphthongs and hiatus would increase as proficiency with Spanish increased. While the BL and NS groups are fairly homogeneous with respect to their experience with Spanish, the IL group is more variable. In order to control for this variation, the IL group was also coded for a proficiency measure. Using this proficiency measure, we
can determine to what extent the durational difference within the IL group is correlated with proficiency. While there was no obvious trend or significant correlation between proficiency and the durational difference between diphthongs and hiatus in the front V contexts, the graph in (20) below shows that there is a negative correlation between proficiency and the durational difference between diphthongs and hiatus within the IL group for front sequences in the CV context. A Pearson product-moment correlation coefficient was computed and the correlation was found to be significant ($R = -0.859$, $N = 6$, $p = 0.029$).

\[(20)\]

Similarly, the negative correlation illustrated in (21) for the back sequences in CV contexts was also found to be significant ($R = -0.934$, $N = 6$, $p = 0.006$).

\[(21)\]
Putting these results together, it seems that the BLs have a larger gap between diphthongs and hiatus in terms of duration, but that as proficiency increases in the IL group, the proportional difference in duration between the two sequence types decreases and approaches the NS values. The hypothesis related to duration in (13) predicted that the BLs would have the least difference in duration between diphthongs and hiatus and that as proficiency increased in the IL group, the difference would increase and approach the NS values. As we have seen, the difference in duration between the two sequences does change as proficiency increases, but that the direction of the change is the opposite of what was predicted. The BLs start out with a fairly large gap between the distribution of the two sequences and with increased experience with the language, the gap is lessened.

The second variable on which the production of vocalic sequences was compared was %-transition. The prediction was that the BLs would have a fairly small difference in %-transition between diphthongs and hiatus and that as proficiency increased in the IL group, the difference would increase and approach the NS values. As regards this variable, the graph in (22) below illustrates that the NSs consistently devote more of the sequence to the transition, in both diphthongs and hiatus, than do the learners. Additionally, we can see a distinct pattern of increasing %-transition in both sequence types starting with the BLs and finishing with the NSs, which of course mirrors the pattern seen in the graph of the normalized duration of the transition.

As discussed in §2.1, one of the phonetic characteristic of hiatus is that they tend to have a lower %-transition than diphthongs. If the BLs realize vocalic sequences with a lower %-

![Graph](image-url)
transition than the NS, then this suggests that they are producing them as somewhat more hiatus-like than the NSs are and as we can see from (22), the ILs fall in the middle. However, it is important to note that the difference in %'-transition between the BLs and the NSs is only 4% and the difference between the ILs and the NSs is only 2%. This is very likely not a practically significant value and as such, does not constitute evidence that the learners are producing diphthongs in a more hiatus-like way than the NSs. As was discussed earlier regarding the durational difference, it was found that the BLs and ILs produced Spanish vowels with a longer duration than the NSs did, even when speech rate was controlled for. This means that there is less of the sequence remaining that can be occupied by the transition. It seems then, that the smaller %'-transition found in the speech of both learner groups is likely a side effect of the longer durations found in their vowels.

Examining %'-transition strictly in terms of the contrast, we hypothesized that the NSs would produce these sequences with a larger difference between the two than the learners would. However, if we compare the proportional difference in %'-transition across proficiency groups, we find that they are virtually identical as shown in the chart in (23).

<table>
<thead>
<tr>
<th>Proficiency</th>
<th>Mean Diphthong %'-Transition</th>
<th>Mean Hiatus %'-Transition</th>
<th>Difference</th>
<th>Proportional Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1BL</td>
<td>0.3286</td>
<td>0.2731</td>
<td>-0.0555</td>
<td>-17%</td>
</tr>
<tr>
<td>2AL</td>
<td>0.3459</td>
<td>0.2893</td>
<td>-0.0566</td>
<td>-16%</td>
</tr>
<tr>
<td>3NS</td>
<td>0.3717</td>
<td>0.3087</td>
<td>-0.0630</td>
<td>-17%</td>
</tr>
</tbody>
</table>

Mean %'-transition of diphthongs and hiatus, mean difference in %'-transition, and proportional difference between the two sequences by Proficiency Group

So the difference between the realizations of diphthongs and hiatus in terms of %'-transition in the speech of the BLs and the ILs conforms to the native speaker norms. In short, all three speaker groups produce the contrast between diphthongs and hiatus in the same way with respect to %'-transition. As noted above, the prediction regarding the diphthong/hiatus contrast was that as proficiency increased, the difference in %'-transition would increase as well. Given that there is no disparity between the learner groups and the native speaker group, it is unsurprising that no significant correlation was found between %'-transition and the proficiency measure within the intermediate group either.

In sum, we hypothesized that the proportional difference in duration and %'-transition would increase as proficiency increased both across the proficiency groups (BL, IL, and NS) and within the less homogeneous IL group. As discussed above, it was found that all 3 groups
produced diphthongs with very similar mean durations, but that the BL and IL groups produced hiatus which were longer than those in the NS group. This results in a greater proportional difference in duration between diphthongs and hiatus in the speech of the BLs and the ILs than in the speech of the NSs, which is the opposite of what was predicted. Examining the relationship between proficiency and the durational difference within the IL group, it was found that there is a significant negative correlation between the two variables in that as proficiency increases, the durational difference between diphthongs and hiatus decreases\(^{10}\). The production of these sequences was also examined in terms of their composition with respect to the proportion of the sequence that is occupied by each element: the first (a glide or a vowel), the second (a vowel), and the transition between the two. This analysis found that for hiatus, the BLs and the ILs were very similar in their behaviour in that they produced both vowels in the sequence with longer duration and produced the transition with shorter duration than the NSs. For diphthongs, again the BLs and ILs were quite similar, producing the glide portion of the sequence shorter, the vowel portion longer, and the transition portion shorter than the NSs. This suggests that the learners are realizing Spanish vowels (but not glides) with longer duration than the NSs, even when speech rate is controlled for. However, it seems that that this characteristic of the learners’ speech could be due to the general salience of durational differences stemming from the L1 experience with contrastive stress as well as an inability to reproduce the durations accurately which results in the overshooting of the durational targets for hiatus and a somewhat exaggerated diphthong/hiatus contrast in the speech of both learner groups. In terms of %-transition, while the proportion of the sequence occupied by the transition did increase across the proficiency groups, as mentioned above the difference was very small (only 4% between the NS and BL groups) and as such is likely not of practical significance. In addition, the proportional difference in %-transition was almost identical across the proficiency groups and within the IL group there was no correlation between proficiency and %-transition.

5.2 Hypothesis 2: The Effect of Phonological Transfer

As discussed in §2.2, there is a phonological constraint against consonant + glide sequences in an onset where both elements in the sequence have the same place of articulation. This constraint bans the sequences [pw], [bw], [mw], [tj], [dj], [nj], [lj], and [rj] in onset position in English, disallowing words such as *[pwɪl] or *[ljɑn] (Davis & Hammond 1995, Ohala & Kawasaki-Fukumori 1997). The second hypothesis predicts that this constraint will be

\(^{10}\) This result was found for both sequences in front and back CV contexts, but not for the front V context.
transferred to production of Spanish diphthongs, resulting in an asymmetry between CV diphthongs in which the consonant and glide share place of articulation and those where they do not. The hypothesis is reproduced in (24) below.

(24) The phonological constraint against CG sequences in an onset in which the consonant and the glide share place of articulation will be transferred to production of Spanish CV diphthongs in the speech of the beginning and intermediate learners. This transfer will result in the diphthongs in which the consonant and glide share place of articulation being produced in a less native-like way in that the learners will produce the glide portion of the sequence with a longer duration than the NSs, producing a sequence of a consonant and a vowel, rather than a consonant and a glide in order to avoid violating the constraint. The CV diphthongs in which the consonant and the glide do not share place of articulation will be realized in a more native-like manner than those where they do. Further to this, it is predicted that the effect of the transfer of this constraint will weaken as proficiency with Spanish increases. As a consequence, we expect to find a negative correlation between proficiency in the IL group and duration of the glide segment. As proficiency increases, the duration of the glide segment in CV diphthongs where the consonant and glide share place of articulation will decrease and approach the native Spanish norms.

To test this hypothesis, each diphthong token with a preceding consonant was coded for place of articulation of the consonant. Diphthongs containing a front glide were coded as either ‘tj’, indicating that the place of articulation was coronal (/t/, /d/, /ɾ/, or /l/), or as ‘other-j’, indicating that the place of articulation was not coronal. Diphthongs containing a back glide were coded as either ‘pw’, indicating that the place of articulation was labial (/p/, /b/, or /m/), or as ‘other-w’, indicating that it was not labial. The mean normalized duration of the glide was then compared across these four consonant-glide types and across the three proficiency groups. If the hypothesis in (24) were confirmed, we would expect to see the duration of the glide being much longer in the ‘tj’ and ‘pw’ sequences than in the other sequences in the speech of the BLs as compared to the NSs. As is shown in the chart in (25) and the plot in (26) below, while the duration of the glide in the ‘tj’ sequences is longer than in the other sequences, there is very little difference in the durations of the glides between the BLs and the NSs in the ‘tj’ sequences (3ms difference) and there is certainly no greater difference within the ‘tj’ sequences than in the ‘other-j’ sequences as was predicted.
With respect to the ‘pw’ sequences, the chart in (25) does show that the BLs produced the back glide with a preceding labial consonant with a longer duration (mean 69ms) than the NSs (mean 53ms), however, the difference between the groups is only 16ms and is likely not a practically significant one.

![Table showing mean duration in ms. of glide in CG sequence by sequence type and proficiency group.]

The results of this analysis have shown that the hypothesis in (24) was not confirmed. It seems then, that even though the ‘pw’ and ‘tj’ sequences are not found in onset position in English, the learners, even the BLs with very little experience with Spanish, are able to produce them in a fairly native-like way. There are a couple of possible explanations for the BLs’ precocious performance with these diphthongs, which will be discussed in §5.
5.3 Hypothesis 3: Phonetic Transfer & Glides

The specific acoustic characteristic of the glides that was investigated here was the relative intensity. As explained in §4.5, the relative intensity was calculated by taking the difference between the intensity of the glide and the intensity of the following vowel. The hypothesis made regarding the glides is provided below.

(27) Beginning learners will use the English realization of both front and back glides (characterized by a lower relative intensity than the Spanish counterparts) as a result of phonetic transfer while the intermediate learners will produce glides with more native-like relative intensities as a result of their greater proficiency with Spanish. Within the intermediate learners we will see a positive correlation between relative intensity and proficiency in that as proficiency increases, so does relative intensity, changing the RI values from the relatively low English norms up to the relatively high Spanish norms.

In order to evaluate the first part of this hypothesis – whether the BLs are using the English norms for glides – the first step is to examine the relative intensities of each glide type and compare the productions of the beginning learners in Spanish to the English norms.

(28) As can be seen in (28) above, for both glide types and both contexts\(^\text{11}\), the relative intensity of the glide is much higher during production of a Spanish glide by BLs than it is during production of the English glide. In addition, a statistical analysis using one-way ANOVA found a significant difference between the relative intensities of the English and Spanish glides.

\(^{11}\text{Again, we cannot compare the production of Spanish and English glides in front CV contexts since this sequence type is not found in English.}\)
in each of the environments tested (back CV: $F(1, 254) = 66.655, p<0.001$; back V: $F(1, 177) = 98.222, p<0.001$; front V: $F(1, 183) = 182.248, p<0.001$). This indicates that, contrary to what was predicted by hypothesis three, the beginning learners are not using the English realization of [j] and [w] in their production of Spanish glides, but rather are realizing the Spanish glides with relative intensities that are much higher than what is found in English.

If the BLs are not using the English norms, then it is possible that they are producing the glides within the native Spanish norms. In the chart examining the mean RI of both glide types and contexts across the three proficiency groups shown and the English tokens in (29) below, the most striking aspect is that the BLs are producing both glides in both contexts with a somewhat higher RI than that found in the speech of NSs. This is exactly opposite of what was predicted by hypothesis three since it was expected that the BLs would use the English RI values when producing the glides which would result in a much lower RI than the NSs. Not only are the BLs producing the Spanish glides with a relative intensity that is quite different from the English norms, they are also producing these glides with a relative intensity that is higher than what the NSs are using. This result will be discussed further below after investigating each glide type and context separately.

(29)

If we examine the production of the Spanish glides across the three proficiency groups and across glide types and contexts, we find an asymmetry in the production of front and back glides. Although in each glide type and context the BLs have a somewhat higher RI than the NSs, in the case of [w], the BLs are producing the glide within the Spanish norms since a one-way ANOVA found no statistically significant difference in RI in either the CV or V contexts between BLs and NSs. The fact that these beginning learners with very little exposure to Spanish are able to produce the back Spanish glides within the norms of the native speakers
suggests that the difference in relative intensity between the English and Spanish glides is quite salient to the learners. This finding falls in line with the results of MacLeod (2008), as discussed in §3, in which it was found, first, that there exists a significant difference in relative intensity between English and Spanish glides and second, that the difference between the RI of the English and Spanish labiovelar glides was salient to Costa Rican Spanish listeners. However, in the case of [j], a one-way ANOVA found a statistically significant difference in RI between BLs and NSs in both contexts (front CV\textsuperscript{12}: $F(1, 255) = 9.998, p = 0.002$; front V: $F(1, 123) = 31.072, p<0.001$). In contrast, there were no significant differences in RI between the ILs and NSs in either context or glide type, indicating that the IL group, on average, produces Spanish glides within the Spanish norms.

Although it was found that the IL group falls within the Spanish norms overall, as we have discussed earlier, this group is not particularly homogeneous with respect to the proficiencies of its members. We can further focus in on the IL group and make use of the proficiency measure collected through the passage reading task in order to determine whether a change in relative intensity correlates with an increase in proficiency as was predicted by the second part of hypothesis in (27).

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{figure.png}
\caption{Mean Relative Intensity by Proficiency of Intermediate Learners}
\end{figure}

The line graph in (30) above shows a strong linear correlation between relative intensity and proficiency. As proficiency increases, relative intensity decreases. A Pearson product-moment correlation coefficient was computed and the correlation was found to be highly significant ($R = -0.268$, $N = 584$, $p < 0.0001$). The second part of the hypothesis in (27) predicted that within the IL group, increases in relative intensity would be correlated with an increase in proficiency as was predicted by the second part of hypothesis in (27).

\textsuperscript{12} While the difference in RI in the front CV contexts was found to be statistically significant, this difference is likely not a practically significant difference since the mean RI in this context for the BLs was -0.23dB and the same for NSs was -1.74dB giving a difference of only -1.51dB.
proficiency resulting in a positive correlation between the two variables. As we have seen, while there is a strong correlation, the relationship between RI and proficiency is, in fact, negative. The expected outcome was that the BLs would produce the Spanish glides using the English norms, causing the RI to be much lower than the NS values, and that we would see the RI values increasing as proficiency increased in the IL group converging on the native Spanish relative intensities. As shown above, the direction of the change is the opposite of what was predicted. The BL group did not use the low English RI values and instead overshot the Spanish norms. From this starting point, the members of the IL group lower the relative intensities of their Spanish glides as proficiency increased, bringing the average value of the IL group into line with the NS group.

To further examine the production of Spanish glides across the proficiency groups, we can conduct an analysis looking at each participant individually. The chart in (31) below shows the mean RI values for each participant by context and glide type. In each of the four environments we see a general negative trend, with the steepest found in the front V contexts. While there is certainly variability, the overall movement is from higher RI values to lower as proficiency increases.
Summarizing the findings of this study in terms of hypothesis three, we have found that the beginning learners do not simply transfer the norms of English glide production with respect to relative intensity to their production of Spanish glides. In fact, the BLs are producing the Spanish back glide in a fairly native-like way in terms of relative intensity in both the CV and V contexts, disconfirming the prediction of hypothesis 3 for back glides. In contrast, the BLs are not producing the front glide in a native-like way or in an English-like way in either context, but particularly in the V contexts in that they tended to overshoot that value by producing these glides with a higher RI than the native speakers of Spanish. From this high RI starting point, the intermediate learners steadily lower their RI values as proficiency increases causing the average relative intensity of the IL group to be within the NS norms for both glide types and contexts. Ultimately, we have found that the English norms are not transferred in the speech of the BLs, but that there is a steady change in RI as proficiency increases.

5.4 General Discussion

Two over-arching predictions were made regarding the production of Spanish vocalic sequences by learners of Spanish. First, it was predicted that the learners would not produce the diphthong/hiatus contrast in a native-like way due to the fact that there is no contrast between diphthongs and hiatus in English and so it was presupposed that native English speakers with very little experience with Spanish would have difficulty producing the difference between the two sequence types. Second, it was expected that the beginning learners of Spanish would exhibit transfer from English in the form of a) phonological transfer of the English constraint against CGV sequences in which the consonant and glide have the same place of articulation and b) phonetic transfer of English-like relative intensity values in production of the Spanish glides. Furthermore, it was predicted that the effects of these transfers would be less evident in the speech of the ILs than in the BLs.

Addressing the first prediction, that the difference between a diphthong and a hiatus would increase as proficiency with Spanish increased, the results of the experiment showed that the BLs and ILs produced diphthongs and hiatus with the greatest difference between the two sequence types and that the NSs had a smaller difference, which is the opposite of what was predicted. When the relationship between proficiency and the difference between a diphthong and a hiatus was probed, it was found that there is a significant negative correlation between

13 In terms of the two acoustic correlates of the contrast, duration and %-transition.
the two variables in that as proficiency increases, the durational difference between diphthongs and hiatus decreases\textsuperscript{14}. The upshot of this result is that rather than producing a reduced contrast, the BLs and ILs produced it with a somewhat exaggerated difference in terms of duration when compared to the NSs. When the composition of the sequences (durations of first element, transition, and second element) was further investigated, it was found that in both diphthongs and hiatus, the two learner groups were producing the vowel components with longer durations than were the native speakers and that this was causing the expanded difference between a diphthong and a hiatus since diphthongs contain only one vowel element, while hiatus contain two. Ultimately, this result suggests that the learners are able to hear the difference between a diphthong and a hiatus, and are able to reproduce it in such a way that the distributions of the two sequence types do not overlap. However, the reproduction of the contrast is not completely native-like due to the overly long durations of the Spanish vowels by both learner groups.

The second prediction concerned transfer. As discussed in §3, it is important that we distinguish between phonological and phonetic transfer. The first prediction was that the learners would transfer the phonological constraint against homorganic CGV sequences resulting in the two vocalic elements in such sequences being realized in hiatus (CV.V). As was seen in §5.2, there was no discernible difference between the realization of homorganic CGV sequences and non-homorganic CGV sequences. This suggests that the phonological transfer of the above constraint did not occur. The learners seemed to have no particular difficulty in producing sequences that are essentially illicit in their L1. As promised in §5.2, here we will discuss several possible explanations for why this might be, although the true reason for this result could be a combination of several factors. First, as noted in §2.2, while homorganic CGV sequences are not found in stressed onset position in English, they do occur heterosyllabically across syllable boundaries, as in upwards [\textipa{up\,\,wə\,\,\,ðz}], as well as in compounds and phrases. Since the English speakers were found to have very little to no trouble producing homorganic CGV sequences, it is possible that their experience with these sequences was made available for production of the Spanish sequences through phonetic transfer. As explained in §3, phonetic transfer is transfer of the acoustic realization of a segment with no regard for its phonemic status, participation in an opposition, or position within a domain (syllable, word, phrase, etc) in either the L1 or the L2. As regards the homorganic CGV sequences, although English speakers have no experience with them in

\textsuperscript{14} As discussed, there was no significant correlation found between %-transition and proficiency.
stressed onset position, they do have experience with them across syllable boundaries\textsuperscript{15}. Since phonetic transfer is not interested in syllable boundaries, it is possible that the very fact that homorganic CGV sequences occur linearly in English, even with the intervening syllable boundary, could be providing a guideline for the learners in producing these sequences in onset position in Spanish. This phonetic transfer would be an instance of the positive effect of transfer in that it aided the learners in producing sequences that are phonologically illicit in their L1.

A second possibility is that there are robust cues in the input that signal the learners to acquire a syllabification that is not consistent with their L1 grammars, namely the homorganic CGV sequences in stressed onset position. Steele (2001) showed that 9 English-speaking novice L2 learners of French in the initial stages of acquisition after fewer than 30 hours of exposure to French were able to restructure their interlanguage grammars to accommodate the syllabification of word-final obstruent-liquid clusters. Steele argued that this rapid acquisition of a structure which is illicit in the L1 was due to the learners having reset the parameter related to syllabification as a result of the presence of appropriate and robust cues. The beginning learners in the present study had been exposed to Spanish for no more than 15 hours and yet were able to produce the homorganic CGV syllabification which is illicit in English. Perhaps their quick acquisition of this syllabification is due to the presence of cues in the input allowing them to adjust their interlanguage grammars as well. Determining whether such cues exist and are used by the learners will be left to future research.

A third and less interesting possibility is that perhaps producing homorganic CGV sequences is simply not particularly difficult and so even though the learners do not have experience with such sequences in onset position in their L1, they readily produce them in the L2. A more thorough examination of the articulatory requirements of producing these sequences would be needed in order to determine if this possibility is correct. However, in addition, it would be necessary to determine why these sequences are absent from English and furthermore to address the cross-linguistic tendency for languages to avoid homorganic consonant clusters in the onset (Maddieson & Ladefoged 1995). If these sequences involve a relatively low effort, then it would be surprising that so few languages would incorporate them.

The second prediction relating to transfer was that the BLs would use the acoustic realization of English glides in terms of relative intensity during production of the Spanish glides as a result of phonetic transfer. As discussed in §5.3, not only did the BLs not use the English relative intensity values, which are much lower than the native values for the Spanish

\textsuperscript{15} And to a lesser extent in the onset of unstressed syllables (as in Christian [ˈkrɪs.tjən]) and in loanwords (as in pueblo [ˈpwe.blo] ‘town’)
glides, they actually produced the Spanish glides with RI values higher than the NS norms. In a similar vein to the argument in §5.1, intensity, like duration, is an acoustic correlate of stress and as a result, it is likely a very salient characteristic of segments to English speakers (Fry 1955, Beckman 1986, Jacewicz 2005). As noted in §5.3, while the BLs produced the Spanish glide [w] within the NS range in both contexts, they did not produce [j] within the NS range in either context, but rather produced [j] with a mean relative intensity that was higher than the native Spanish values, giving a hyperarticulated output. One possible explanation for this overshoot is that the learners lack the necessary precision to produce the glide with the appropriate Spanish relative intensity. Since all of the diphthongs that were tested were stressed and the vast majority were word-initial (i.e. prominent positions), hyperarticulation is not unexpected (Lindblom 1989). However, hyperarticulation does not necessarily affect each parameter of a given production equally. As we have seen, the relative intensity of the Spanish front glides were hyperarticulated, but the durations of these glides were not. The mean normalized duration of [j] varied only by 5ms between the BLs (mean 99ms) and the NSs (mean 104ms). A similar result was found in Colantoni & Steele (2007) in their investigation of the acquisition of French /ʁ/. In that study, a group of intermediate learners showed an exchange of accuracy with respect to manner versus voicing, producing /ʁ/ with target-like manner, but insufficient voicing. Colantoni & Steele conclude that the learners lacked the necessary articulatory control to produce both parameters in a target-like way simultaneously and as a result only focused on the parameter that was most salient to them. In the present study, the BLs produced the Spanish glide [j] with a hyperarticulated RI, but a target-like duration. Perhaps this is also due to their lack of control needed to produce all of the parameters pertaining to the L2 segments in a native-like way.

The hypothesis regarding the production of Spanish glides was that the BLs would produce them with English-like RI values as the result of phonetic transfer. As we have seen, this transfer did not occur, likely due to the salience of the difference in RI between the glides of the two languages. Further to this hypothesis, it was predicted that the effect of the phonetic transfer would be less in the speech of the ILs than in the speech of the BLs. While the transfer did not occur, a significant correlation between proficiency and relative intensity was still found. As proficiency increased, the relative intensity of the glides decreased. If the reason for the overshoot of the RI values in the speech of the BLs is lack of articulatory control, then it is conceivable that as the L2 speakers gain experience with Spanish they also gradually gain the necessary control to produce the glides with native-like intensities. Overall, this is what was found in Colantoni & Steele (2007) in that the advanced learners produced /ʁ/
with duration and percentage-voiced that were closer to the native control group than did the intermediate learners.

While the majority of this analysis has focused on normalized durations, the raw durations may also play a role. As can be seen in the chart in (32), the BLs consistently produced VV-sequences that are longer than both the NSs and the ILs, while the ILs produced sequences that were longer than the NSs. If we focus on the front sequences, we see that, in fact, the mean duration of a front diphthong in the speech of the BLs is actually longer than the mean duration of a front hiatus in the speech of the NSs, which is significant since hiatus are generally longer than diphthongs.

(32) Mean Duration in ms. of Front and Back Diphthongs and Hiatus by Proficiency Group

<table>
<thead>
<tr>
<th>Proficiency Group</th>
<th>back glide [w]</th>
<th>front glide [j]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Diphthong</td>
<td>Hiatus</td>
</tr>
<tr>
<td>Beginning Learners</td>
<td>248</td>
<td>392</td>
</tr>
<tr>
<td>Intermediate Learners</td>
<td>221</td>
<td>345</td>
</tr>
<tr>
<td>Native Spanish</td>
<td>190</td>
<td>264</td>
</tr>
</tbody>
</table>

Even though the BLs’ diphthongs had a composition (in terms of the normalized durations of the glide, transition, and vowel) that was not vastly different from the NS values, their extreme duration would in all probability cause them to be perceived as hiatus. In addition, the BLs produced front glides with a higher RI than the native speakers. In the case of the CV contexts, the RI was barely below zero\(^{16}\). A relative intensity nearing zero suggests that both segments in the sequence are vowels since the intensities of the two elements are almost identical. These vowel-like relative intensities combined with the long duration of the sequences would likely mean that the sequences would be perceived as hiatus by native speakers. A perception test in which native speakers judge the learners’ productions as diphthong or hiatus would shed light on the degree to which these characteristics of the BLs speech affect categorization of the segments.

6 Conclusions and Future Work

From the results of the experiment described in this paper, we can generate four main conclusions. First, as regards the investigation of how the two learner groups would produce the diphthong/hiatus contrast, it was found that both groups are able to realize the contrast in a fairly native-like way with respect to the two parameters measured: duration and %-transition.

\(^{16}\) The mean RI of front CV diphthongs by beginning learners was -0.23dB.
In terms of %-transition, we saw that the three proficiency groups performed remarkably similarly. In terms of duration, it was found that both the BLs and the ILs produced the contrast with a somewhat larger difference in duration due to the fact that they realized vowels with longer durations than the NSs. As discussed earlier, it is perhaps the case that the difference in duration is salient to English speakers due to the fact that it is one of the acoustic correlates of stress. However, since the learners produced overly long vowel duration, it seems that they lack the articulatory control needed to generate native-like vowel durations.

The second main conclusion stems from one of the central themes of the paper: transfer. The second hypothesis predicted that phonological transfer of the constraint against homorganic CGV sequences would occur and the third hypothesis predicted that phonetic transfer of the English relative intensity values for glides would occur. As pointed out in the discussion, the distinction between phonological and phonetic transfer proved to be very important. Although phonological transfer of the L1 constraint against homorganic CGV sequences was found not to have occurred, it seems that positive phonetic transfer did take place, allowing both learner groups to produce these sequences with little difficulty and no significant differences from the NS speakers. In the literature, whether a given study is investigating phonological or phonetic transfer is infrequently stated. Although the occurrence of phonetic transfer regarding homorganic CGV sequences was unexpected, acknowledging the distinction between the two types of transfer allowed for a reasonable conclusion regarding the learners’ quick acquisition of an illicit syllabification. One of the contributions of the present study then, is the result that being clear about what is being transferred is very important in any study examining second language acquisition.

The third main conclusion of this study pertains to the notions of saliency and articulatory control. As we saw in §5.3, phonetic transfer of the English RI values for glides did not occur. It was suggested that this was due to the fact that intensity, being a correlate of stress, would be salient to English speakers. This same notion was put forth in relation to the results of the durational contrast between diphthongs and hiatus, where both learner groups produced a robust contrast between the two sequence types. In terms of both measurements (relative intensity and duration), the BLs realized segments that were higher than that of the NSs. The relative intensities of the front glide [j] were higher than the native control group and the durations of the Spanish vowels in both diphthongs and hiatus were overly long. So while intensity and duration seem to serve as salient cues for the English listeners, their actual implementation was not native-like. As suggested earlier, one possible reason for the discrepancy between the BL and NS groups was that the learners lacked the necessary articulatory control to produce these parameters in a native-like way. It is not uncommon for
beginning learners of an L2 to have difficulty producing an L2 segment accurately, even when
the cues for its production are salient. As mentioned earlier, Colantoni & Steele (2007) found
that intermediate learners of French lacked the required control to produce both native-like
manner and voicing of /ʁ/, but advanced learners produced the rhotic with more native-like
values of duration and %-voice. Flege (1980) found in his study of Arabic-speaking learners of
English that most of the learners were able to generalize the durational contrast found in Arabic
between /t-d/ and /k-g/ to the English contrast between /p-b/, but that one speaker in particular
exaggerated the size of the voice contrast. Flege notes that an exaggerated phonetic correlate
of a feature could be due to the fact that it is particularly salient or relatively easy to articulate.
If the result of such an exaggeration is that the segments produced are not native-like with
respect to the exaggerated parameter, then it is reasonable to conclude that as proficiency with
the L2 increases, learners of the language will seek to reduce the degree of exaggeration and
conform to the native norms.

The final main conclusion concerns development. As the overriding purpose of this
study was to investigate the acquisition of the diphthong/hiatus contrast by English-speaking
learners of Spanish, changes in the production of these sequences that are related to increases in
proficiency are of particular consequence. Although the direction of the change in the
durational difference between a diphthong and a hiatus and the direction of the change in
relative intensity of the Spanish glides were the opposite of what was predicted, there were
clear developmental paths revealed for these two measures. First, as regards the
implementation of the diphthong/hiatus contrast in terms of duration, it was found that the BLs
produced the contrast with a somewhat exaggerated difference, but that as proficiency
increased within the IL group, the proportional difference in duration between the two sequence
types decreased. Second, with respect to the relative intensity of the glides, it was found that
the BLs overshot the native speaker norms for front glides, but that as proficiency increased in
the IL group, the RI values decreased. In this way, we have seen that the path that
development takes is that the English speakers originated with higher than native-like values of
duration and RI and as their experience with Spanish increased, the values lowered to conform
to the native control values.

As noted throughout the preceding paper, there are several areas that provide obvious
opportunities for future research on the acquisition of vocalic sequences in Spanish. First, in the
discussion regarding the fact that phonological transfer of the constraint against homorganic
CGV sequences did not seem to occur since the BLs showed very native-like glide durations in
this sequences it was pointed out that this quick acquisition of these sequences could be
stemming from the presence of robust cues to the new syllabification. In order to determine
what such cues would be and whether or not the BLs are making use of them constitutes an interesting and important extension to the present study.

Second, as was discussed in §5.4, the BLs produced diphthongs with very long raw durations and it is likely the case that even though the composition (in terms of normalized duration of the first element, transition, and second element) is not largely different from that of native speakers, these sequences would be perceived as hiatus. This could be tested in further study by presenting the sequences to native judges who give their intuitions on the syllabification.

Third, an interesting extension to this study would be to collect production data from learners during a reading task, rather than a repetition task. It is likely that the orthographic representation of the Spanish tokens tested here would affect the production of the words in such a way that the learners would produce the sequences with less of a robust contrast between the two. This effect would be due to the fact that both the glide in a diphthong and the first vowel in a hiatus are represented by the same grapheme in Spanish. For example, both the front glide in *tiara* /tiara/ [ˈtja.ɾa] ‘tiara’ and the high, front vowel /i/ in *tía* /tia/ [ˈti.a] ‘aunt’ are represented by the grapheme <i>. It is true that the hiatus is signalled by the presence of the accent on the high vowel, but the fact that in both cases the vocalic sequence is represented by the orthographic sequence <ia> is likely to affect the production of these sequences, particularly in the speech of beginning learners, leading to a flattening of the diphthong/hiatu contrast.

Lastly, as was noted in §5.4, a further investigation into the status of homorganic CGV sequences in English would be useful in pinpointing the exact type of experience that English speakers have with them. A complete understanding of the nature of these sequences would help to elucidate the effects of phonological and phonetic transfer and their consequences for the acquisition of the diphthong/hiatu contrast.
References


Reeder, J. T. (1998). An Acoustic Description of the Acquisition of Spanish Phonetic Detail by Adult English Speakers. Department of Spanish and Portuguese. Austin, TX, University of Texas at Austin. PhD.


Vokic, G. (2005). The Role of Structural Position in the Acquisition of L2 Phonology, Purdue University. PhD.


<table>
<thead>
<tr>
<th>V-initial</th>
<th>W-initial</th>
<th>C-w-initial</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Spanish Stimuli</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diphthongs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spanish Stimuli</td>
<td></td>
<td></td>
</tr>
<tr>
<td>English Stimuli</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>V Contexts</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CV Contexts</td>
<td></td>
<td></td>
</tr>
<tr>
<td>V Contexts</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Spanish and English Stimuli</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Appendix A – Stimuli</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix B – Reading Passage

*El viento del norte y el sol*

El viento del norte y el sol discutían acerca de cuál de los dos sería el más fuerte, cuando, de repente, pasó un viajero envuelto en una amplia capa. Al verlo, convinieron en que el primero que consiguiera quitarle la capa sería el más fuerte. El viento del norte comenzó a soplar con mucha furia, pero, cuanto más soplaba, más se aferraba el viajante a su capa, hasta que el viento norte desistió. El sol brilló entonces con todo su esplendor, e inmediatamente, el viajante arrojó su capa. Así, el viento norte tuvo que reconocer la superioridad del sol.
Appendix C – Mean Syllable Durations & Normalizing Factors