An investigation of the production of ejectives by native (L1) and second (L2) language speakers of Q'eqchi' Mayan

Wendy Baker-Smemoe  
*Brigham Young University, wendy_baker@byu.edu*

Karl Olaw Christian Wagner  
*Brigham Young University*

Follow this and additional works at: [https://scholarsarchive.byu.edu/facpub](https://scholarsarchive.byu.edu/facpub)

Part of the Arts and Humanities Commons

**Original Publication Citation**


**BYU ScholarsArchive Citation**

Baker-Smemoe, Wendy and Christian Wagner, Karl Olaw, "An investigation of the production of ejectives by native (L1) and second (L2) language speakers of Q'eqchi' Mayan" (2013). *Faculty Publications*. 5902.  
[https://scholarsarchive.byu.edu/facpub/5902](https://scholarsarchive.byu.edu/facpub/5902)

This Peer-Reviewed Article is brought to you for free and open access by BYU ScholarsArchive. It has been accepted for inclusion in Faculty Publications by an authorized administrator of BYU ScholarsArchive. For more information, please contact ellen_amatangelo@byu.edu.
An investigation of the production of ejectives by native (L1) and second (L2) language speakers of Q’eqchi’ Mayan

Karl Olaw Christian Wagner⁎, Wendy Baker-Smemoe

Brigham Young University, A-41 ASB, Provo, UT 84602, USA

ARTICLE INFO

Article history:
Received 24 September 2012
Received in revised form 6 August 2013
Accepted 21 August 2013
Available online 23 September 2013

ABSTRACT

This study examines the production of voiceless unaspirated stops and ejective stops by native (L1) and second language (L2) speakers of Q’eqchi’ to determine (1) whether there were acoustic differences (in voice onset time and burst duration) between voiceless unaspirated stops and ejective stops and (2) whether L2 speakers were more accurate at producing phones that are similar to the native language (voiceless unaspirated stops) than those that are different (ejective stops). Acoustic analyses of the stops produced by the ten L1 and thirteen L2 Q’eqchi’ speakers revealed (1) that L1 speakers make a distinction in voice onset time and at times in burst duration in ejective versus voiceless unaspirated stops, (2) L2 speakers can produce a difference between ejective and voiceless unaspirated stops and (3) L2 speakers are more accurate at producing ejective than voiceless unaspirated stops. These results suggest that L2 speakers are actually less accurate at producing L2 phones that are similar versus less similar to native language phones. Results are discussed in light of current theories of L2 speech learning.

© 2013 Elsevier Ltd. All rights reserved.

1. Introduction

This study examines the production of voiceless unaspirated pulmonic egressive (hereafter “voiceless”) and glottalic egressive (hereafter “ejective”) stops in Q’eqchi’ Mayan as produced by native (L1) Q’eqchi’ speakers in Belize and Guatemala, as well as by second language (L2) learners of the Q’eqchi’ language, whose native language is English. Q’eqchi’ is part of the Mayan language family and is spoken by approximately 800,000 speakers located primarily in the departments of Alta Verapaz, Petén, Izabal, Baja Verapaz and Quiché, Guatemala and the southern portion of the Toledo District in Belize.

An analysis of L1 speakers’ and L2 learners’ production of Q’eqchi’ voiceless and ejective stops was pursued for two reasons. First, few acoustic analyses of Q’eqchi’ have been performed and little is known about how ejectives differ from the voiceless unaspirated stops in this language. Thus, an acoustic analysis of native speakers’ productions of these phones is needed. Second, Q’eqchi’ is a theoretically important language to study in terms of L2 acquisition because it provides a basis for testing L2 speech learning theories regarding the relative ease or difficulty of acquiring specific phones based on their similarity to the learners’ L1 phones.

In addition, this study also incorporates a unique class of L2 speakers. The L2 speakers in the study are native English speakers who each spent 22 months in Guatemala completely immersed in the culture and language of Q’eqchi’ for an average of 17 months. Serving as missionaries for the Church of Jesus Christ of Latter-day Saints (LDS church), they spent at least 12 hours a day communicating and interacting with the Q’eqchi’. Each L2 speaker thus had equally high levels of motivation in acquiring the language and learned the language in a very similar context. This group of English L1 speakers who have learned an indigenous language of Central America with a relatively small distribution and number of speakers is particularly unique and warrants further investigation.

2. Literature review

2.1. Second language acquisition

L2 acquisition studies often focus on foreign-accented speech and the underlying factors that determine why certain L2 consonants and vowels (hereafter “phones”) may be easier to acquire than others (Flege, Bohn, & Jang, 1997; Piske, Mackay, & Flege, 2001; So & Best, 2010). Many studies suggest that one of the most important factors is the learner’s L1 since it often affects the relative ease or difficulty of acquiring specific L2 phones. For
example, Aoyama, Flege, Guion, Akahane-Yamada, & Yamada (2004) examined Japanese learners' acquisition of English /i/ and /I/ and determined that English /i/ is easier for Japanese learners to perceive and produce, in part because it is less similar to Japanese /i/ than is English /I/. Such examples are replete in the literature, and most discussions suggest that a learner's L1 is a strong (although not conclusive) predictor of foreign accent in the L2 (i.e., Baker, Trofimovich, Flege, Mack, & Halter, 2008; Bohn & Best, 2012; Chakraborty & Goffman, 2011; Escudero & Williams, 2011; Halle, Best, & Levitt, 1999; Halle & Best, 2007).

Merely examining phonetic inventories of phones, however, is not enough to help with predicting the relative ease or difficulty of learning L2 phones. Even when phones are similar across the L1 and L2, L2 phones may still have "quite different patterns of temporal implementation" from their L1 counterparts (Flege & Port, 1981). In other words, even if two languages have similar phones (such as the voiceless alveolar stop /t/ in English and Q'eqchi') acoustic variations in these phones across languages can affect foreign accent. For example, even the degree to which vowels preceding nasal consonants are nasalized seems to vary from language to language (Clumeck, 1976). Baker (2010) demonstrated that, although Korean has stops in word-final position, native Korean speakers had difficulty producing English stops in this position, even after several years of speaking English. Moreover, recent studies have demonstrated that even one's own L1 dialect or the L2 dialect to which a speaker is exposed plays a role in the acquisition of L2 phones (i.e., Baker & Smith, 2010; O'Brien & Smith, 2010).

Moreover, other researchers have demonstrated that foreign accent on specific consonants and vowels can differ depending on where the consonant or vowel is located in a syllable. Perhaps the most studied example of this is English /i/ and /I/ as perceived and produced by native Japanese speakers. Because these two consonants vary considerably depending on their position in a syllable, learners of English are often more proficient in perceiving and producing these phones in one phonetic context than in others (i.e., Bradlow & Pisoni, 1999). Other researchers have found similar results with native French (Trofimovich, Gatbonton, & Segalowitz, 2007) and native Korean (Trofimovich, Baker, & Mack, 2001) speakers learning English. These differences in L2 pronunciation accuracy depending on the phonetic context is related to allophonic variations usually found across word contexts—variations that seem to affect the relationship between L1 and L2 phones (i.e., Strange et al., 2001).

Most models and theories regarding L2 speech learning also focus on how the relationship between L1 and L2 phones affect L2 speech learning, although these models also take into account other factors that influence L2 speech learning. All three have in common that they assume that L2 phones are perceived in terms of L1 phone categories. In some cases, the L2 phones are perceived as good or poor examples of these L1 categories and the matches between the L1 and L2 phones predicts how easily the L2 phones are learned. Three of these models are of interest to the current study and are described briefly below.

2.1.1. Speech Learning Model (SLM)

Put simply, the Speech Learning Model (SLM) proposed by Flege (1995, 2002, 2003) posits that the relative degree of similarity between L1 and L2 phones is a strong predictor of L2 speech learning, with more similar phones across the L1 and the L2 being more difficult to acquire. One of the reasons for this difficulty is that accurate production is based on accurate perception—thus if learners are unable to hear a difference between similar L1 and L2 phones (even though such differences exist), they will not be able to produce L2 phones accurately (Flege, 1995). The SLM posits that learners attune to auditory-acoustic differences between L1 and L2 phones. One of Flege's first studies on this topic was Flege (1987) where it was demonstrated that native English speakers were better able to learn French /y/ than French /u/. Flege (1987) proposed that because the native English speakers were better able to hear the difference between French /y/ and English vowels, they were better able to produce it. Since that time, the SLM has been modified (i.e., Flege, 2002, 2003), but its basic tenets remain the same and these findings have been robust (i.e., Aoyama et al., 2004; Baker, Trofimovich, Flege, Mack, & Halter, 2008).

2.1.2. Perceptual Assimilation Model-L2

Best's (1995) Perceptual Assimilation Model (PAM), expanded recently to include L2 speech learning development and called PAM-L2 (Best & Tyler, 2007), also bases L2 speech learning on the similarity between L1 and L2 phones, but posits that L1–L2 similarity determines the relative ease or difficulty of perceiving L2 contrasts (like English /i/ and /I/), rather than single L2 phones. PAM-L2 also posits that listeners attune to articulatory settings of L1 and L2 phones, not acoustic-auditory differences. Like the SLM, PAM-L2 proposes that L2 phones that are similar to L1 phones may be more difficult to learn. One of the best examples of this is Best, McRoberts, and Sithole (1987) who found that native English speakers could very accurately discriminate between different clicks, perhaps because of their dissimilarity to any English phones. This model demonstrates predictable ability in perceiving L2 contrasts based on whether both members of an L2 contrast are both perceptually assimilated to one L1 phonological category or to two different L1 phonological categories. Although there are several possible L1-L2 relationships, a few are of note to the present study. First is a non-assimilable contrast—where one or more of the phones in an L2 contrast differs so much from L1 categories that it is perceived as non-speech. PAM-L2 predicts that (at least in terms of perception) non-assimilable L2 phones will be the most easily discriminated and, if the assumption follows, easier to produce than other phones. A good example of a non-assimilable contrast is the contrast between two Zulu clicks (Best, McRoberts, LaFleur, & Silber-Isenstadt, 1995). Second, in two-category assimilation, the two phonemes in the L2 contrast assimilate to two different L1 phoneme categories. This may occur when a native English speaker hears the German vowels /i/ and /I/, since English has both of these categories. In the last two assimilation patterns, which are perhaps the most important for this study, both phones in the L2 contrast assimilate to the same L1 category. In single category (SC) assimilation, both phones are perceived as equally good members of the same L1 category—which may occur when native Korean speakers hear the same German contrast, since Korean has no /I/ vowel. These contrasts may be the most difficult to acquire. Finally, in category goodness assimilation, both phones in an L2 contrast assimilate to the same L1 category, but one of the L2 phones in the pair is a better exemplar of an L1 category than the other. This was the case for voiceless and ejective stops in Best et al. (1995). Like SLM, PAM-L2 would predict that the L2 phone in the pair that is the least like the L1 category would be the easiest to acquire.

2.1.3. Second Language Linguistic Perception

The Second Language Linguistic Perception (L2LP) model proposed by Escudero (2005), like the PAM-L2, posits that L2 phones are learned in contrasts, and that the relationship of these contrasts to native language phones predicts (though only in part) the ease of learning L2 phones. Similarly, in this model, when the L2 phones in an L2 contrast are similar to two different L1 phone categories (termed a “similar scenario” in this model) the L2 phone contrast will be relatively easier to learn. This is because, in a similar scenario, learners will be able to use their similar native language phone categories to perceive and produce them (Mayr & Escudero, 2010), such as when a native English speaker learners the distinction between German /i/ and /I/. By contrast, when the two L2 phones in the L2 contrast are both perceived as members of a single L1 phone category
(term a "new scenario"), learners must either create a new category ("category creation") or learn to split ("category split") an L1 category into two L2 categories (such as what happens with native Korean speakers learning English /l/ and /t/). It is also possible to have a "partially new scenario" where perception of the L2 contrast falls in between the new and similar scenarios, and ease of learning is assumed to fall between learning L2 contrasts in a new or similar scenario. Unlike PAM-L2, this model proposes that there may be several factors involved in how L2 phones are perceptually assimilated to L1 ones, including age of L2 acquisition, amount of experience, and other individual differences in L2 phone learning. Because of this, this model posits that different speakers will have different perception and production abilities based on their experience with the L2 (Mayr & Escudero, 2010). Importantly, also unlike the PAM-L2, the L2LP posits that learners attune to auditory-acoustic variations between L1 and L2 phones, not articulatory gestures.

As the review of these three models suggests, questions still remain about whether L2 phones that are similar to L1 phones are actually easier or harder to learn. Most of the studies examining the influence of cross-language similarity on L2 speech learning examine consonants and vowels that are very similar across the two languages, such as /l/ and /l/ in English (Escudero & Boersma, 2004), German (Bohn & Flege, 1990; Bohn & Flege, 1992) or English /r/ and /l/ (Aoyama et al., 2004). The reasons for this are obvious—these L2 contrasts are often the most difficult to learn. The problem with this is that defining similar and dissimilar is difficult, especially in languages where the phones are so similar across both languages. Moreover, most of these studies have been done on perception, not production, and sometimes these studies are carried out on naive listeners with no experience with the L2 (i.e., Best et al., 1987).

The goal of the present study is to examine the learning of L2 contrasts where one of sounds in the L2 contrast is very similar to the native language (i.e., the unaspirated [t]) and one where it is very different (i.e., the ejective [t].) The choice of examining ejectives was done in part because little research has been done on the acquisition of these phones (but see Best et al., 1995 and Werker & Tees, 1984 on the perception of ejectives by naive listeners).

The only known study of L2 acquisition of ejectives has involved the acquisition of Yucatec Maya ejectives by native Spanish speakers (Gonzalez-Poot, 2009). Gonzalez-Poot (2012) aimed to show that subjects whose native language lacks certain phonological features are still able to acquire phones with these features in a second language. In this case the Spanish speakers lacked the [constricted glottis] feature that characterizes the ejectives in Yucatec Mayan. In particular, Gonzalez-Poot (2012) compared the production and perception of ejectives in onset and coda positions using an auditory discrimination task as well as a forced-choice picture selection task. The results of his study found that the Spanish speakers, though not behaving quite native-like in all contexts, were able to produce and perceive ejectives to some degree of accuracy. The present study also examines the production of Mayan ejectives by L2 learners, but in addition seeks to acoustically analyze how the native speakers of Q'eqch'i' produce ejectives in comparison to the L2 Q'eqch'i' learners. The results of the analysis are also examined in light of the three models presented above.

2.3. Q'eqch'i' stops and ejectives

Q'eqch'i' is part of the K'ichean branch of the Mayan language family which covers a large portion of the Mayan languages spoken in Guatemala. It is spoken by approximately 800,000 speakers predominantly in the Alta Verapaz and Petén departments of Guatemala. However, as one of the geographically largest Mayan languages in Guatemala it can be found in neighboring departments as well as in southern Belize. The consonant inventory of Q'eqch'i' (based on Cho, 2004) is presented in Table 1.

In its series of standard stops, English contrasts between voiceless unaspirated and voiceless aspirated stops whereas the Q'eqch'i' series contrasts predominantly between voiceless unaspirated stops and ejectives. In addition, the bilabial stops in Q'eqch'i' contrast between a voiceless stop and an implosive. Voiceless unaspirated stops are produced by the blocking of the vocal tract so that all airflow ceases during the production of the consonant. In contrast, ejective consonants are distinguished from voiceless, non-glottalized stops by a closure of the glottis as well as

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Phonemic inventory of Q'eqch'i' consonants (IPA)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bilabial</td>
<td>Alveolar</td>
</tr>
<tr>
<td>Stops</td>
<td>p b</td>
</tr>
<tr>
<td>Nasals</td>
<td>m</td>
</tr>
<tr>
<td>Affricates</td>
<td>ts ts'</td>
</tr>
<tr>
<td>Fricatives</td>
<td>s</td>
</tr>
<tr>
<td>Trills</td>
<td>r</td>
</tr>
<tr>
<td>Laterals</td>
<td>l</td>
</tr>
<tr>
<td>Semi-vowels</td>
<td>w</td>
</tr>
</tbody>
</table>
a supralaryngeal closure during the production of the phone (Stevens & Hajek, 2004). The air between the two closures is compressed and thus leads to a strong and unique burst as the oral closure releases. The glottis remains closed until after the oral closure is released (Ladefoged & Maddieson, 1996). Though Q’eqchi’ and English both employ velar [k] and alveolar [t] voiceless plosives as phonemes, the phones are articulated in different manners between the languages. The Q’eqchi’ voiceless stops are considered short-lag, meaning that they are produced with little or no aspiration and a short VOT (Cho, 2004). In contrast, English voiceless stops are long-lag and usually incorporate aspiration to distinguish themselves from the voiceless stops in English (Olive, Greenwood, & Coleman, 1993). However, this distinction is greatest in initial word position and may be non-existent in other word positions (Lisker & Abramson, 1967). VOT may vary with place of articulation but at least one study has found that it is not a factor in the contrast between ejectives and other stops (Cho & Ladefoged, 1999). However, several other recent studies have found that ejectives and voiceless stops do in fact vary in terms of VOT (i.e., Gordon, Potter, Dawson, De Reuse, and Ladefoged, 2001; Stevens & Hajek, 2004). In fact, Gallagher (2012) found that even native speakers of Quechua were unable to perceptually distinguish between ejectives and voiceless stops if the VOT between them was not sufficiently distinct. Therefore, one goal of this study is to observe whether VOT is an important feature distinguishing between the voiceless stops and the ejectives in Q’eqchi’.

In Q’eqchi’, voiceless stops and ejectives occur in both the onset and coda syllable positions, thus occurring in word-initial, word-final and word-medial positions. The Q’eqchi’ stop series is again shown in Table 2. The consonants examined in this study are in bold. In Q’eqchi’ there are four voiceless unaspirated stops, /p, t, k, q/. The latter three have a contrasting ejective counterpart while /p/ contrasts with the implosive /ɓ/ rather than a voiced /ɓ/. Two voiced stops, /ɓ/ and /ɗ/, can appear in Q’eqchi’, occurring, however, exclusively in Spanish loanwords. The voiceless glottal, /ʔ/, also exists in conjunction with vowels. Since few if any previous studies detail the phonetic differences between the voiceless stops and ejectives as produced by L1 Q’eqchi’ speakers, it became a focus of this study to examine the differences between the productions of these two types of stops by the L1 speakers before comparing them with the L2 learners’ data.

Limited research has been performed on the nature of the ejective stops in Q’eqchi’. However, Campbell (1973) observed some interesting features of the ejective series of stops in Q’eqchi’, observations which challenge more general studies on such consonants. In one of these general studies on ejective consonants, Greenberg (1970) makes several generalizations about ejectives and injectives (i.e., implosives). He states that “injectives tend to have front articulation, ejectives to have back articulation” (Greenberg, 1970). However, in his studies on Mayan languages Campbell found that the Q’eqchi’ ejective stop series contains an implosive at the bilabial as well as the uvular position (Campbell, 1973). He further states that according to Greenberg’s implications, this would mean that since the uvular consonant is imploded, the other consonants of the series would be as well. In Q’eqchi’ the alveolar and velar ejective stops, as well as two affricates, all are produced as clear ejectives. Campbell suggests that a possible reason for the bilabial and uvular positions producing injectives has to do with the fact that they are located at the periphery of the oral cavity (Campbell, 1973). While the Campbell (1973) study observed some peculiarities in the production of some L1 and L2 speakers’ uvular ejectives, there is not enough data or evidence to make a conclusive statement on the certainty of uvular implosives in Q’eqchi’. Further study could reveal this claim by Campbell to be true. As it stands, the uvular ejective remains part of the ejective series.

3. Participants

A total of twenty-three participants were recruited to participate in this study. Both L1 and L2 Q’eqchi’ speakers underwent the same interview and recording procedures. However, only L2 speakers were asked to fill out a language questionnaire to determine language competency, as detailed below. The study was approved by the Brigham Young University ethics review board and participants were informed about the study and their rights, all signing a consent form prior to testing.

Ten native speakers of Q’eqchi’ were recruited in the Toledo district of Belize as well as the department of Alta Verapaz in Guatemala. All of the participants spoke Spanish as a second language and one participant from Belize also spoke English as well as the creole spoken in southern Belize. The participants lived in regions where the eastern dialect of Q’eqchi’ is predominant (Coban, 2004). All speakers were male, ranging in age from 19 to 45. The decision to only test male speakers was a limitation caused by the availability of L2 speakers, as discussed below.

Thirty L2 learners of Q’eqchi’ were asked to participate in this study, all of them native English-speaking participants. The participants were between the ages of 21 and 27. Each of the participants had previously lived in the department of Alta Verapaz in Guatemala among the indigenous population of Q’eqchi’ while serving a religious LDS mission. The duration of this immersion experience amongst the Q’eqchi’ was an average of 17 months, with several for as long as 21 months. During this time, they actively engaged in learning Q’eqchi’ and spoke it on a daily basis anywhere between four and eight hours a day with the L2 speakers. They would have been fewer years removed from their immersion experience, but with such a unique and small L2 learner group, this was not possible. The results, however, show that this was not a severe limitation as the L2 learners were still able to produce the foreign L2 phones quite adeptly.

The L2 speakers were asked in a questionnaire to indicate how active they were in using the language (reading, writing, speaking) as well as their current reading, writing, speaking and comprehension ability on a scale of 1 through 10. This information was used to determine that all the L2 speakers were still proficient enough in their L2 language skills to be of use to the study. Their current usage and self-ratings showed that they were qualified for inclusion in the research. On a 10 point scale the L2

<table>
<thead>
<tr>
<th>Table 2</th>
<th>Q’eqchi’ stops.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Voiceless</td>
<td>p</td>
</tr>
<tr>
<td>Ejective</td>
<td>–</td>
</tr>
<tr>
<td>Implosive</td>
<td>b</td>
</tr>
<tr>
<td>Voiced (Spanish loans)</td>
<td>b</td>
</tr>
</tbody>
</table>
learners rated (on average) their current speaking ability as a 7, the lowest rating being a 4 and the highest a 9. The lowest rated category was writing which was not a factor in the study.

4. Experiment 1: L2 learners’ perception of Q’eqchi’ stops

In order to determine the relative similarity of Q’eqchi’ and English stops, we first examined the L2 speakers’ perception of Q’eqchi’ stops in terms of English consonant categories. For this experiment, seven of the thirteen L2 speakers performed two tasks, a cross-language identification task and a similarity rating task (both described below).

4.1. Stimuli

For the stimuli, 33 CV syllables were taken from words produced by native Q’eqchi’ or native English speakers, depending on the language, and were excised from the word so that lexical familiarity would not influence perception. Syllables contained English /l/, /l/, /d/ before /a/, /u/, /u/, and /o/ and Q’eqchi’ /l/, /q/, /q/, /k/, and /q/ before /a/, /o/, /u/ and /l/.

4.2. Procedure

For the cross-language identification task, listeners heard a syllable (such as Q’eqchi’ /kl/) and were asked to mark to which English consonant it sounded most similar. The options were /pl/, /l/, /kl/, /bl/, /dl/, /gl/ and other. For the similarity rating task, participants were asked to rate how good of an example the English consonant they thought the token was on a scale from “1” (not a good example) to “5” (a very good example). We also asked 6 naive listeners to perform this task, but their results are quite similar to those of the participants in this study, and so are not shown here. Mayr and Escudero (2010) also found that both inexperienced and experienced listeners had similar perceptual assimilation patterns in their study.

4.3. Results

The results of this analysis are given in Table 3. Across the top of the table are the English consonant categories that could be chosen and on the left column are the actual consonants the participants heard. The percentage of times L2 Q’eqchi’ learners rated English and Q’eqchi’ tokens as members of different English consonant categories are given and the goodness of fit ratings of these consonants are given in parentheses.

The results of this perception task suggest the following. First, listeners perceived Q’eqchi’ /l/ (93%, 3.76 match) as more similar than Q’eqchi’ /l/ (62%, 2.92 match) to English /l/. Second, Q’eqchi’ /k/ (92%, 3.57 match) and Q’eqchi’ /k/ (100%, 3.14 match) are both seen as very good examples of English /k/. By contrast, Q’eqchi’ /q/ (85%, 2.88 match) and /q/ (86%, 2.47 match) are both worse matches to English /k/ than Q’eqchi’ /k/ or /k/.

Finally, both Q’eqchi’ /q/ and /q/ seem to be equally matched to English /k/.

These results suggest the following for predictions based on the three speech learning models discussed above. First, if learners are better able to acquire phones that are less similar to native consonant categories than those that are more similar, then the L2 Q’eqchi’ learners should be able to produce Q’eqchi’ ejectives more accurately than voiceless stops. In addition, they should be more accurate at producing uvular than velar or alveolar stops. These results suggest we can make predictions about (1) ejective versus voiceless stops (2) different places of articulation (alveolar, velar, and uvular) and (3) between uvular and velar stops. Based on these results, the three L2 speech learning models would predict the following.

4.3.1. SLM

The SLM predicts that the more similar the L2 phone is to an L1 phone, the more difficult it is to learn. Therefore, based on the results of the perception task above, the SLM would predict that (1) voiceless stops would be more difficult to learn than ejective stops (in general), (2) Q’eqchi’–English stops should be acquired more easily than velar or alveolar stops and (3) it should be more difficult for learners to acquire the ejective/voiceless distinction than the velar/uvular distinction. This is because the /kl/–/k/ and /q/–/q/ contrasts are both perceptually assimilated onto the same L1 phone category.

4.3.2. PAM-L2

In terms of PAM-L2, the perception study demonstrates an interesting combination: four Q’eqchi’ categories (/kl/, /k/, /q/, /q/) are all identified with a single English phoneme /k/. Comparing these categories to each other, two of the contrasts seem to follow a category goodness (CG) relationship since one of the members of the pair is a better match to English /k/ than the other member of the pair: /kl/–/q/ and /k/–/q/. By contrast, the pairs /kl/–/k/ and /q/–/q/ seem to follow a single category (SC) match since both members in these pairs are equally assimilated to the same L1 phone category.

Table 3

Cross-language identification results (similarity ratings on a scale from 1(not a good example) to 5 (a very good example)) are given in parentheses.

<table>
<thead>
<tr>
<th></th>
<th>Eng /p/</th>
<th>Eng /l/</th>
<th>Eng /k/</th>
<th>Eng /bl/</th>
<th>Eng /d/</th>
<th>Eng /g/</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eng /l/</td>
<td>100% (4.57)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eng /k/</td>
<td>100% (4.42)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eng /d/</td>
<td>10% (3.45)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q’eq /l/</td>
<td>93% (3.76)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q’eq /k/</td>
<td>5% (1.00) 62% (2.92) 33% (2.6)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q’eq /q/</td>
<td>92% (3.57) 100% (3.14) 8% (1.5)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q’eq /q/</td>
<td>5% (2.00) 5% (1.00) 85% (2.88) 5% (3.00)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q’eq /q/</td>
<td>10% (2.66) 86% (2.47) 4% (1.00)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Therefore, according to PAM-L2, it should be easier to learn and produce differences between velar-uvular contrasts than voiceless-ejective contrasts. Moreover, the ejective stops (including the alveolars) should be easier to learn than the voiceless Q’eqchi’ stops.

4.3.3. L2LP

Similar to the PAM-L2, the L2LP would predict that learning voiceless-ejective stop contrasts would be difficult. This is because learning these contrasts would involve a “new scenario” where two L2 phones are assimilated to the same L1 phone category. In other words, learning the Q’eqchi’ /t/ and /k/ contrasts would require learners to “split” L1 English categories /t/ and /k/ and would require two new phone categories to be created. However, it should be easier to learn the Q’eqchi’ /t’/ contrast than the /k’/ and /q’/ contrast. This is because the Q’eqchi’ /t’/ contrast is a partially new scenario, where Q’eqchi’ /t’/ is perceptually assimilated to several different L1 phones. By contrast, both phones in the Q’eqchi’ /k’/ and /q’/ contrasts were nearly always perceptually assimilated to English /k/, which makes both of these contrasts more like new scenarios, where most likely new categories will need to be formed. In addition, learners seem to display a “multiple assimilation pattern” where four L2 phones (Q’eqchi’ /k/, /k’, /lq/, /q’/ ) are perceptually assimilated to one L1 category, English /k/. The L2LP would predict that learning the L2 Q’eqchi’ phone contrasts /t/–/t’/ and /k’–k/ would be easier than learning /q’–lq’/ contrast since the latter pair would require two new categories. Moreover, like the PAM-L2, the L2LP would predict that it should be easier to learn the uvular-velar contrast, than the voiceless-ejective /k’–k/ and /q’–lq’/ contrasts since these two contrasts are completely neutralized (they are equally good members of the same L1 category) whereas the velars are less similar to English /k/ than the uvulars. In addition, L2LP would predict that different learners would have different assimilation patterns that would affect learning these L2 contrasts.

5. Experiment 2: Production

The second experiment of the study was conducted to fulfill the three purposes of this study: to examine (1) how native Q’eqchi’ speakers produce voiceless and ejective stops, (2) how well L2 Q’eqchi’ learners produce these same phones, and (3) whether the three L2 speech learning models accurately predict the production abilities of the L2 Q’eqchi’ learners. All participants (10 native Q’eqchi’ speakers and 13 L2 learners) took part in this experiment.

5.1. Stimuli

All the tokens used in this experiment were common vocabulary items found in Q’eqchi’. The selected words used as tokens did not fall under any specific restrictions as to word class or structure other than requiring the phones under investigation to occur in either word-initial or word-medial position. Each word contained either a voiceless stop (/t/, /k/ or /q/) or their ejective counterpart (/t’/, /k’/, or /q’/). The majority of the words consisted of one or two syllables. The words used were mostly nouns or verbs. Special considerations for syllables and stress were unnecessary since most syllables in Q’eqchi’ are simple CV(C) structures and stress in Q’eqchi’ always falls on the final syllable (Cho, 2004).

Participants were originally asked to produce 15 sentences as well as a list of 50 distinct words. Within the set of 50 words, 30 of the words contained the tokens required for analysis with the rest acting as distractors. These 30 tokens can be found in the tables below. Table 4 provides the list of each token in Q’eqchi’ containing the voiceless unaspirated stops in addition to a basic IPA transcription and English gloss. It also provides the corresponding information for the tokens containing the ejectives. The voiceless and ejective stops occurring in the 15 sentences were used in the speech perception task taken by a segment of the L2 Q’eqchi’ learner participants.

### Table 4

<table>
<thead>
<tr>
<th>Word initial</th>
<th>Word medial</th>
<th>Word initial</th>
<th>Word medial</th>
</tr>
</thead>
<tbody>
<tr>
<td>t'</td>
<td>/t'ane'k/</td>
<td>t'</td>
<td>/t'ane'k/</td>
</tr>
<tr>
<td>ti'</td>
<td>/t'ani/</td>
<td>te'</td>
<td>/t'e/</td>
</tr>
<tr>
<td>t'</td>
<td>/t'ani/</td>
<td>te'</td>
<td>/t'e/</td>
</tr>
<tr>
<td>q'</td>
<td>/q'eqchi/</td>
<td>q'</td>
<td>/q'eqchi/</td>
</tr>
<tr>
<td>q'anb'esink</td>
<td>/q'anbe'si/</td>
<td>q'</td>
<td>/q'eqchi/</td>
</tr>
<tr>
<td>q'alunk</td>
<td>/q'alunk/</td>
<td>q'</td>
<td>/q'eqchi/</td>
</tr>
</tbody>
</table>

*List of tokens—voiceless unaspirated stops and ejectives.*
5.2. Procedure

The L1 participants in the study were recorded in their local residences in Guatemala and Belize. The L2 participants in the study were either recorded at their residence or in a quiet location on the campus of Brigham Young University. All subjects were recorded producing the word on a portable Edirol (R-09) voice recorder. The recorder was held close to the participant as they read the list of tokens at a normal speaking rate. Participants were allowed to correct themselves and repeat any words they misread or mispronounced.

The acoustic analysis of the data was completed using Praat version 5.3.37 (Boersma & Weenink, 2013). Both qualitative observations of the waveform and spectrogram as well as quantitative measurements were taken. The types of measurements recorded in this study were motivated by an already present body of literature involving studies dealing with acoustic analysis of various languages (e.g., Ladefoged & Maddieson, 1996; Ladefoged, 2003). The most relevant parameters in studies on voiceless stops and ejectives in other languages have been VOT, burst intensity, burst duration, and inter-vocalic closure duration (Applebaum & Gordon, 2006; Hogan, 1976; Wysocki, 2004). VOT measurements were performed using the same criteria as Lisker and Abramson (1964).

For the current study VOT and burst duration were the key quantitative measurements. Another measurement that could have been taken was the intensity of the ejective bursts. However, this measure was not used because it seemed to vary greatly on how much the speaker emphasized the word and thus would most likely not have yielded accurate or useful data. Figs. 1 and 2 provide examples from the VOT and burst duration measurements taken from the data provided by the participants in word-initial and word-medial positions. These two measures provided enough information, identifying noticeable differences between L1 and L2 speakers in the production of the ejectives and voiceless unaspirated stops in Q’eqchi’.

The burst intensity of a stop can be an important factor in differentiating voiceless stops from ejectives (Applebaum & Gordon, 2006). However, due to the variation of ejective productions across languages other factors may need to be taken into account, such as the characteristics of the VOT, pitch

---

**Fig. 1.** Burst duration and VOT measurement for word-initial position. This figure provides the spectrogram and waveform for part of the word t'ane’k ’to fall’ produced by an L1 Q’eqchi’ speaker. The solid horizontal line below the waveform indicates the burst duration measurement. The VOT measurement ranges from the onset of the burst to the onset of the vowel.

**Fig. 2.** Burst duration and VOT measurement for word-medial position. This is part of the word saq’e ‘sun’ produced by an L1 Q’eqchi’ speaker. The same measurements were taken as indicated in Fig. 1.
or amplitude at voicing onset, and voice quality (Ham, 2008). The acoustic analysis of the data elicited by the native Q’eqchi’ speakers revealed that ejectives generally were released with forceful ejection leading to a strong burst and a longer VOT than the voiceless stops. Figs. 3 and 4 demonstrate these differences between speakers’ productions of ejectives and voiceless stops.

After all data, which included 30 tokens for each of the 23 speakers (a total of 690 tokens), were measured and organized, repeated measures ANOVAs were applied comparing each participant’s production of their ejective stops to the voiceless stop VOT and burst duration. In addition, subsequent and similar analyses were performed to determine differences in place of articulation and word position.

5.3. Ejective and voiceless unaspirated stop production of native Q’eqchi’ speakers

One purpose of this study was to analyze how native speakers of Q’eqchi’ produce voiceless stops and ejective stops. To accomplish this, we acoustically analyzed both the voiceless and ejective stops of the native speakers in two word positions: initial and medial. We measured both the duration of the burst and the VOT of the stop at three places of articulation: alveolar (/t/ and /t’/), velar (/k/ and /k’/) and uvular (/q/ and /q’/) (see Fig. 5).

The burst duration of the voiceless stops were on average 16.2 ms, while ejective bursts were on average 14.8 ms, meaning that the burst duration of voiceless stops was approximately 1.08 times longer than the duration of the ejective stops. Conversely, the VOT of the ejectives was on average 75.1 ms and for the voiceless 36.1 ms, meaning that the VOT of the ejectives was approximately 2.08 times longer than the VOT of the voiceless stops.

We performed two separate two-way (place of articulation and word position) repeated measures ANOVAs with each participant’s average burst duration for the voiceless and ejective stops as the dependent variable for the first ANOVA and VOT for the voiceless and ejective stops as the dependent variable for the second ANOVA. The results of these analyses determined that there was no difference in the duration of the burst for voiceless versus ejective stops (F(1,9)=.619, p<.435, η²=.012), and no effect of word position (F(1, 9)=1.33, p=.235, η²=.026) although there was an effect of place of articulation (F(1,2)=8.00, p=.001, η²=.242). Post-hoc analyses revealed that the alveolar burst duration was shorter than the uvular and velar burst durations. In addition, there was a place of articulation by word position interaction (F(1,2)=3.93, p=.02, η²=.136). Further analyses revealed that there was no difference in burst duration for alveolars and uvulars across word position, but there was an interaction in burst duration for velars in initial position versus medial position (see Fig. 5).

A similar analysis on the VOT for voiceless and ejective stops found that there was a significant difference between voiceless and ejective VOT (F(1,9)=119.65, p<.0001, η²=.705), but no effect of place of articulation (F(1,2)=2.82, p=.07, η²=.102), nor an effect of word position (F(1,2)=3.29, p=.07; η²=.062). In other words, ejective VOT was longer than voiceless VOT for every place of articulation and every word position (see Fig. 6).

In summary, the results of the first analysis demonstrates that native Q’eqchi’ speakers produce a difference in VOT between voiceless and ejective stops at all places of articulation and all word positions. By contrast, there was no difference in burst duration across alveolar and uvular voiceless and ejective stops at both word positions, but there was an interaction in the burst duration in velar initial versus medial positions. In addition, alveolar bursts were shorter than uvular and velar bursts.

Fig. 3. An ejective /k’/ in the word k’ehok, ‘to give’, produced by a native Q’eqchi’ speaker. It is characterized mainly by a strong burst and a long VOT (indicated by the dotted line on the waveform).

Fig. 4. A voiceless stop /k/ in the word katzkatz, ‘itch’, produced by a native Q’eqchi’ speaker. It is characterized by a relatively short VOT filled with aspiration (indicated as in Fig. 3).
5.4. Ejective and voiceless unaspirated stop production by L2 Q’eqchi’ learners

Another purpose of this study was to determine whether the L2 speakers also produced a difference between the ejective and voiceless stops in Q’eqchi’. We used the same statistical procedures used to answer the previous research question. The burst duration of the voiceless stops produced by the L2 Q’eqchi’ learners were on average 14.3 ms, while ejective bursts were on average 12.1 ms, meaning that the burst duration of plain stops were 1.18 times longer than the duration of the ejective stops. Conversely, the VOT of the ejectives was on average 78 ms, while the voiceless stops were on average 42 ms, meaning that the VOT of the ejectives was approximately 1.85 times longer than the VOT of the voiceless stops.

In this analysis, as with the native speakers, we ran two two-way repeated measures ANOVAs, one with burst duration and one with VOT for voiceless and ejective stops as the dependent variables. Place of articulation and word position were the within subjects variables for both analyses. These analyses determined that the L2 speakers did not produce a difference between burst duration for voiceless and ejective stops ($F(1,12)=2.27, p=.141, \eta^2_p=.034$) for any of the places of articulation ($F(1,2)=.658, p=.522, \eta^2_p=.02$), nor for any of the word positions ($F(1,1)=1.38, p=.244, \eta^2_p=.021$) (see Fig. 7). However, the L2 speakers did maintain a difference between the VOT for voiceless and ejective stops ($F(1,12)=97.98, p=.003, \eta^2_p=.777$), and the degree of difference between the two types of VOT was the same for each place of articulation ($F(1,2)=.069, p=.933, \eta^2_p=.002$) and each word position ($F(1,1)=.074, p=.787, \eta^2_p=.001$) (see Fig. 8).

In summary, it appears that both the native and L2 speakers of Q’eqchi’ produced a distinction in VOT for ejective versus voiceless unaspirated stops at all places of articulation and in all word positions, while neither group produced many differences between the burst durations of the two types of stops. Some differences, however, across the two groups seem to exist. First, there was less of a difference in VOT between L2 speakers’ ejectives...
and voiceless stops and between those of the L1 speakers’. Moreover, there was no difference in burst duration across places of articulation for the L2 speakers, while there was for the L1 speakers.

5.5. Comparisons of L1 and L2 productions of voiceless and ejective stops

The final purpose of this study was to examine whether learners were better able to acquire consonants that are similar versus dissimilar across the two languages. We chose to examine voiceless and ejective stops, since voiceless stops, although they do differ across the two languages, are somewhat similar across English and Q’eqchi’ (differing primarily in being long-lag aspirated and short-lag unaspirated respectively) whereas ejectives are very dissimilar across the two languages, since the English consonant inventory does not contain ejectives. For this analysis, we examined whether the native English speakers were able to produce voiceless or ejective stops similarly to native Q’eqchi’ speakers. The SLM would predict that the more similar the phone, the more difficult it is for L2 learners to acquire. That would mean that the native English speakers would be more accurate at producing ejectives than voiceless stops. Similarly, the PAM-L2 and L2LP would predict that learners should have more difficulty distinguishing between ejective-voiceless pairs (/k–/k’/ and /q–/q’) than between uvular-velar pairs (/k–/q’/ and /k’–/q’/) because of the assimilation patterns described above.

We first examined whether the two groups differed in their ability to produce a distinction between ejective and voiceless stops in both burst duration and VOT. We ran a repeated measures ANOVA with ejective and voiceless burst duration as the dependent variables with group (native versus L2 learners) as between and place of articulation (alveolar, velar, and uvular) and word position (medial and initial) as within subjects factors. The results of this analysis demonstrated no main difference between ejective and voiceless burst duration ($F(1,1) = 2.516$, $p = .115$, $\eta^2_p = .022$), or group ($F(1,1) = .087$, $p = .768$, $\eta^2_p = .001$) or word position ($F(1,1) = 2.50$, $p = .117$, $\eta^2_p = .021$), but there was an effect of place of articulation ($F(1,2) = 7.99$, $p = .001$, $\eta^2_p = .123$) and a group by place of articulation interaction ($F(1, 2) = 4.11$, $p = .01$, $\eta^2_p = .067$). Post-hoc analyses revealed that the native speakers maintained a distinction between alveolars and the other two places of articulation, while the L2 learners did not. No other interactions were significant. A similar analysis on VOT for voiceless versus ejective stops revealed a significant difference in VOT for ejectives versus voiceless stops ($F(1,1) = 204.81$, $p = .0001$, $\eta^2_p = .642$). All other main effects and interactions were not significant. These results suggest that the two groups (native speakers and L2 learners) differed in terms of burst duration, but did not differ in their production of VOT across ejective and voiceless stops.

The above analysis, however, did not determine whether the L2 Q’eqchi’ learners were better able to produce voiceless or ejective stops. Therefore, we next examined whether there was a difference in how the two groups (L2 Q’eqchi’ versus native Q’eqchi’ speakers) produced burst duration and VOT for voiceless stops versus ejective stops. To do this we first ran two ANOVAs, one with voiceless burst duration and one with voiceless VOT as the dependent variable. As independent variables, we examined group (native versus L2 learners), place of articulation (alveolar, velar, and uvular), and word position (initial and medial). The results for the burst duration analysis demonstrated that there was no main effect of group ($F(1,22) = .952$, $p = .331$, $\eta^2_p = .008$) or word position ($F(1,1) = .151$, $p = .6.01$, $\eta^2_p = .01$) but there was a main effect for place of articulation ($F(1,1) = .229$, $p = .116$, $\eta^2_p = .039$) or word position ($F(1,1) = .258$, $p = .613$, $\eta^2_p = .002$), but there was a group by word position interaction ($F(1,2) = 3.83$, $p = .05$, $\eta^2_p = .033$). All other interactions were not significant (all other $F$s < .18, all other $p$s > .17). Post-hoc analyses revealed that the L2 speakers produced longer VOTs than the native speakers for all of the places of articulation across all word positions.

We next examined whether the two groups differed in terms of their productions of ejectives, with the first analysis examining differences in burst duration and the second differences in VOT of ejective stops. The first analysis on ejective burst durations revealed that there was no main effect of group ($F(1,22) = 3.02$, $p = .076$, $\eta^2_p = .027$), but there was a main effect of place of articulation ($F(1,1) = 15.11$, $p = .0001$, $\eta^2_p = .211$) and word position ($F(1,1) = .781$, $p = .066$, $\eta^2_p = .065$). There was also a significant interaction of group and place of articulation ($F(1,2) = 4.408$, $p = .014$, $\eta^2_p = .065$). Post-hoc analyses revealed that the two groups differed in their production of burst duration for alveolars and uvulars, but not velars. No other interactions were significant (all other $F$s < 1.11, all other $p$s > .312). A similar analysis on voiceless stop VOT revealed that there was a significant main effect of group ($F(1,22) = 3.80$, $p = .05$, $\eta^2_p = .032$), place of articulation ($F(1,2) = 14.79$, $p = .0001$, $\eta^2_p = .206$) and word position ($F(1,1) = 8.199$, $p = .005$, $\eta^2_p = .067$). There were no significant interactions (all other $F$s < 2.17, all other $p$s > .118). These results, therefore, suggest that L2 speakers produced longer VOTs than the native speakers for all of the places of articulation across all word positions.

We also examined the VOT of the ejectives and found there was no main effect of group ($F(1,22) = .952$, $p = .331$, $\eta^2_p = .008$) or word position ($F(1,1) = 2.29$, $p = .116$, $\eta^2_p = .039$), nor word position ($F(1,1) = .258$, $p = .613$, $\eta^2_p = .002$), but there was a group by word position interaction ($F(1,2) = 3.83$, $p = .05$, $\eta^2_p = .033$). All other interactions were not significant (all other $F$s < 1.18, all other $p$s > .17). Post-hoc analyses revealed that the L2 speakers tended to produce longer VOTs for initial ejectives than the native speakers. (See Fig. 9 for a comparison between voiceless and ejective VOT for the native and L2 speakers.)

These results suggest that the two groups differed in terms of ejective burst duration only for the velar position. Moreover, the two groups only differed in their production of VOT in initial, not medial positions. Comparing the results of the analyses on the voiceless stops to the ejective ones suggest that the L2 learners were more accurate at producing ejective than voiceless stops.

![Fig. 8. VOT for both ejective and voiceless unaspirated stops at each place of articulation in milliseconds for the L2 Q’eqchi’ learners (nns).](image-url)
Because the perception tests also suggested that the L2 speakers may have difficulty learning the Q’eqchi’ /k’–/q’/ and /k’–/q’/ contrasts, we ran a series of analyses examining whether the two groups differed in their ability to produce a difference in burst duration and VOT for velars versus uvulars. We ran a series of repeated measures ANOVAs with velar and uvular burst duration and VOT as dependent variables across both voiceless and ejective stops, with word position as within and group as between-subject variables. The results of these analyses revealed that the native speakers produced a difference between ejective velar and uvular burst duration ($F(1.9) = 8.10$, $p = .01$) but did not produce a difference between voiceless burst duration nor VOT for either the voiceless or ejective uvular versus velar stops (all $F$‘s$<2.28$, all $p$‘s$>.148$). When we examined whether the L2 and native speakers differed in their production of velar and uvular stops, they only differed in their production of ejective /k’/–/q’/ burst duration ($F(1.22) = 7.58$, $p = .009$, $\eta^2 = .150$).

These results therefore suggest four main findings: (1) for place of articulation, alveolars seemed to be the most difficult for the L2 speakers to produce, then uvulars and then velars, based on the results of the burst duration analyses, (2) for word position, L2 learners seem to have more difficulty with word initial than word medial position (at least for ejective VOT), (3) L2 Q’eqchi’ learners’ production of voiceless stops differed more from the native speakers’ productions than for the ejective stops, and (4) even native speakers did not produce a difference in VOT and burst duration between uvular and velar stops. In the one case where they did, however, the L2 speakers did not produce a similar difference (for ejective (/k’–/q’/) burst duration).

6. Discussion

6.1. Voiceless unaspirated and ejective stops produced by native Q’eqchi’ speakers

One purpose of the study was to analyze and compare ejectives and voiceless stops in Q’eqchi’ as produced by Q’eqchi’ native speakers. The findings of the study demonstrated that voiceless and ejective stops did not differ in terms of burst duration across voiceless and ejective stops, but did differ in VOT at all places of articulation, with VOT being longer for the ejective stops. This finding contradicts what was found by the UCLA phonetics lab (Ladefoged & Maddieson, 1996) for native Q’eqchi’ speakers and what Cho and Ladefoged (1999) found for ejective versus voiceless stops in general. However, several recent studies have found significant VOT differences when comparing ejective and voiceless stops across Waima’a (Stevens & Hajek, 2004), Witsuwit’en (Wright, Hargus, and Davis, 2002), and other languages (Gordon et al., 2001). Thus, the results of this study add to these findings and suggest that, at least for the dialect of Q’eqchi’ examined in this study, there is a VOT distinction. One reason for this difference between our findings and those found in Cho and Ladefoged (1999) may be that our speakers are younger males, too young to have been part of Cho & Ladefoged’s study. Thus, these findings may indicate a recent change in the language, brought on perhaps by the interference or influence of Spanish. Further studies examining older versus younger Q’eqchi’ speakers, or monolingual versus bilingual speakers are needed to verify this hypothesis.

We also found that, in general, the native speakers maintain a difference in burst duration between alveolars and the other two places of articulation, while there was no distinction between uvular and velar burst duration for either ejective or voiceless stops. In addition, there was no distinction in VOT for the three places of articulation. Other studies have found a difference in VOT and burst duration based on place of articulation (i.e., Maddieson, 2001). Two reasons may explain why we did not see similar effects in our data. First, we did see a trend towards alveolars having the shortest VOT/burst duration, then uvulars, then velars. This pattern is typical across systems that have both velars and uvulars (Maddieson, 2001). It may be that with more speakers we would have obtained statistical significance. Second, it is also common to see differences in VOT or burst duration across some places of articulation and not others (Maddieson, 2001). We found this to be the case with the alveolars versus the more dorsal places of articulation, at least in terms of burst duration.

6.2. L2 Q’eqchi’ production

Another purpose of the study examined how L2 learners of Q’eqchi’ produced these same ejectives and voiceless stops as compared to native speakers. The results revealed that the L2 speakers did not produce a significant difference in burst duration for voiceless and ejective stops. The L2

---

**Fig. 9.** Comparisons of average VOT for L1 and L2 speakers for each place of articulation averaged across initial (i) and medial (m) word positions.
speakers did, however, maintain a difference in VOT between the voiceless and ejective stops, with the VOT longer for the ejective versus voiceless stops. Such findings are similar to studies that examine L2 learning of phones in other languages and mirror those of González-Poot (2012), who found that, at least in perception, learners of Q’eqchi’ distinguished between ejectives and voiceless stops accurately.

Despite being able to produce a VOT distinction between the two types of stops, the L2 Q’eqchi’ learners had difficulty with two specific features of the ejective/voiceless stop distinctions. First, the learners were less accurate at producing burst durations than VOT. The native Q’eqchi’ speakers maintained an ejective/voiceless distinction at least for the velar stops, but the L2 speakers did not produce a difference between ejectives and voiceless stops for any of the other places of articulation. Second, the L2 speakers’ difference in VOT of ejective versus voiceless stops was smaller than the difference found for the native Q’eqchi’ speakers. Further analyses determined that the reason for this was that the L2 Q’eqchi’ speakers produced the voiceless stops (unaspirated in Q’eqchi’) with a longer VOT, similar to English aspirated voiceless stops. This finding corroborates previous research of L2 learners acquiring stop laryngeal distinctions—they tend to produce L2 VOT much like what is found for similar phones in their native language (i.e., González-Lopez, 2012; Harada, 2007; Simon, 2009). By contrast, the L2 speakers produced the ejective stops similarly to the native speakers of Q’eqchi’. Simon (2009) also found that native Dutch speakers were able to produce long-lag stops in English more accurately than short-lag stops, even though Dutch has short-lag, not long-lag stops. Finally, the L2 Q’eqchi’ learners had difficulty producing a distinction between ejective and voiceless stops in the same way in all word positions. They appeared to be able to produce a greater distinction between the two types of stops in initial versus medial position. Again, previous research also demonstrates greater accuracy in some word contexts than others when L2 learners produce L2 phones (Bradlow & Pisoni, 1999; Trofimovich et al., 2007). In fact, a comparison of L2 speakers’ production of ejectives in initial position revealed that the L2 speakers tend to produce them much longer than the native speakers do.

These results may be explained in part by differences in perceptual saliency for some features of consonants than for others. Simon (2009) hypothesized that Dutch learners of English produced long-lag English stops more accurately than short-lag ones because long-lag stops are more perceptually salient than short-lag stops. Other researchers have also found that more perceptually salient aspects of the L2 are more easily acquired (Goldschneider & DeKeyser, 2001). Indeed, several recent studies have found similar results when learning L2 phones—perceptual saliency may play a key role (Baker, 2010; Lai, 2010). Most of this previous research seems to suggest that perceptual saliency may be related to word position (i.e., English long-lag versus short-lag stops in Simon, 2009). Such an assumption may explain why learners were able to produce VOT more accurately than burst duration, since VOT is considerably (in acoustic terms) longer than burst duration. Moreover, native Q’eqchi’ speakers seem to produce a longer VOT in medial versus initial position for ejective stops—suggesting that perhaps this feature may be more salient in word medial versus initial position in the input to which the learners would have been exposed. It may also be that the learners produced a much greater distinction in ejective VOT in initial position because in their native language, English, long-lag stops are produced with a greater distinction from short-lag stops in this position than in medial position (Lisker & Abramson, 1967).

6.3. Comparison of L2 speech learning

The results and their implications discussed above help to determine whether the L2 speakers were more accurate at producing voiceless or ejective stops in Q’eqchi’ and to examine these findings in light of the three models of L2 speech learning discussed in the introduction: the Speech Learning Model (SLM), the Perceptual Assimilation Model-L2 (PAM-L2) and the Second Language Linguistic Perception Model (L2LP). As stated above, each of these models predicts the level of difficulty of learning L2 phones based on their similarity to L1 phones. As such, these models must account for three findings of this study: (1) the L2 learners differed from the native Q’eqchi’ speakers in their production of voiceless stops more than they did in their production of ejective stops, (2) the L2 learners were more accurate in their production of the ejective/voiceless distinction than the uvular/velar distinction, and (3) the learners were most accurate in their production of velars, then uvulars, then alveolars. In other words, the L2 speakers were more accurate in producing ejective (least similar) than voiceless (most similar) stops. They were also more accurate at producing velars (similar) than uvulars (least similar), but were more accurate at producing uvulars (least similar) than alveolars (most similar). How these findings are explained by the three models is discussed below.

6.3.1. Speech Learning Model

Based on the results of this study, the SLM to some extent models L2 learners’ acquisition of voiceless and ejective stops in Q’eqchi’. To summarize, the SLM hypothesizes that L2 phones that are the most similar across the L1 and L2 are the most difficult to produce accurately. Moreover, it holds that perception precedes production in L2 learning, supporting the findings that perceptual saliency is an important aspect of L2 learning. The SLM thus accurately predicts the finding that L2 Q’eqchi’ learners differed from the native speakers more for the voiceless series of stops rather than the ejective series of stops. Since there are no ejectives in English, for the L2 speakers this would suggest that they would be able to learn these faster than the more similar voiceless stops.

Moreover, by extension, from this model one may predict that learners would have the most difficulty producing a difference between the ejective and voiceless stops for more similar than dissimilar phones. This also was supported by the data since the learners had the most difficulty in producing alveolar burst duration similarly to native Q’eqchi’ speakers. However, this prediction did not hold for the velar stops since learners were better at producing velars than either alveolar or uvular stops.

In addition, the finding that learners seemed to produce velars and uvular stops with the same VOT may also be explained by the SLM—since L2 learners perceptually assimilated both uvulars and velars to the same native language consonant (English /k/) the learners had difficulty separating these two L2 phonemes (Flege, 1995). However, the SLM inaccurately predicted that /kl–/kl/ and /ql–/ql/ would be difficult to distinguish since these L2 phones were all assimilated to the same L1 phone (English /k/). In fact, the learners were able to accurately maintain a distinction between these sound contrasts. Thus, the SLM can account for many, but not all, the findings related to L1–L2 similarity.

6.3.2. Perceptual Assimilation Model-L2

The results of this study can also be explained to some extent by PAM-L2. PAM-L2 also correctly predicted that learners would be more accurate at producing. Learners produced the voiceless stops with a longer VOT than the native Q’eqchi’ speakers, producing them more like English long-lag stops. However, PAM-L2 had more difficulty predicting the results of the velar and uvular ejective/voiceless distinctions. From the perception task, it was found the pairs /kl–/kl/ and /ql–/ql/ are SC categorizations—where both L2 phones are equally good matches to English /k/. PAM-L2 would predict that SC categorizations would be difficult to acquire. However, the opposite was the case.
Moreover, the results of the perceptual assimilation task also suggested that /k/-/q/ and /k/−/q/ are a CG categorization, where one phone in the pair (the uvulars) was a worse match to English /k/ than the other phone (the velars). Because of this PAM-L2 would have predicted that acquiring /k/-/q/ and /k/-/q/ distinctions would be easier than acquiring /k/-/k/ and /q/-/q/ distinctions. However, the opposite was true.

Moreover, PAM-L2 would predict that learners should be more accurate at distinguishing between the alveolar ejective/voiceless pair than the velar or uvular ejective/voiceless pairs since Q’eqchi’ /l/ was perceived to be a much worse match than Q’eqchi’ /l/ to English /l/. It would therefore be a CG categorization. However, learners seemed to have more difficulty acquiring alveolars than uvulars or velars. Thus, PAM-L2 was able to account for some, but not all, of the findings of this study.

### 6.3.3. Second language linguistic perception

The L2LP model would posit similar predictions to the PAM-L2 model. Like PAM-L2, L2LP incorrectly predicted that the uvular/velar distinction would be easier to acquire than the voiceless/ejective distinction for similar reasons as PAM-L2. It also incorrectly predicted that the alveolar voiceless/ejective pair would be easier to acquire than the uvular or velar voiceless/ejective pairs. It also incorrectly predicted that the alveolar ejective/voiceless contrast would be easier to learn than the voiceless/ejective distinctions for the velar and uvular places of articulation.

Not examined in this study, however, is the effect of other individual differences that the L2LP proposes affects L2 phonological acquisition, which is an important aspect of L2LP (Escudero, 2005). It is possible that, if learners’ productions were examined individually and compared to their perceptual assimilation of Q’eqchi’ stops with English stops, at least some of their productions could be explained by this model. Since we have the perception data from only some of the L2 learners, explaining all the findings of the study is not possible. However, Table 5 lists the individual perception and production data for the seven L2 learners who did participate in both the perceptual assimilation task and the production tasks described above. This table includes both their assimilation patterns and their production data. Under the perception (Perc) column is listed how many times the individual learner chose the most commonly chosen English consonant for each of the Q’eqchi’ sounds (which is listed in parentheses in each of these columns). Also shown is the average VOT (in milliseconds) for each of the participants for each of the Q’eqchi’ phones, listed under the Prod column. Native Q’eqchi’ speakers’ average VOT is given at the top of the table. Shaded cells in the table indicate when a speaker produced the L2 phone within 1 standard deviation of the average production for the native speakers of Q’eqchi’ (which was approximately 10 ms for each of the L2 phones). We assumed that those speakers who produced a VOT within this range produced the L2 phone most like the native speakers. We shaded these cells to indicate whether there were patterns that could be found in the data.

The participants are listed in the order from the least number of shaded cells (participants 1–3) to the most shaded cells (participants 4–7). These two groups also seem to follow two different patterns. In the first group, participants 1–3 are unable to maintain a distinction between ejective and voiceless stop VOT for at least 2 of the three ejective/voiceless pairs. Participant 1 is the most extreme example of this group: this participant falls within one standard deviation for each of the voiceless stops, but does not produce a VOT for the ejective stops that is much larger than the VOT for the voiceless stops. By contrast, in the second group, participants 4–7 maintain a distinction between voiceless and ejective stops. Some of these participants at times produce a much longer VOT for the voiceless stops than the average for the native Q’eqchi’ speakers, but seem to ‘make up for it’ by producing a much longer VOT for the ejectives than is average for the Q’eqchi’ speakers’ ejective VOT. In other words, they hyper-articulate the ejective VOT so that it is discernibly longer than their VOT for the voiceless counterpart.

What is not apparent is a discernible pattern between the participant’s perceptual assimilation of the Q’eqchi’ stops to English counterparts and whether the participant should be in either group 1 or group 2 (those that do not maintain a distinction between ejectives and voiceless stops and those that do). For example, both Participant 1 and Participant 7 perceptually assimilated Q’eqchi’ /l/ and /l/ completely with English /l/. However, while Participant 1 was not able to produce a distinction between the two Q’eqchi’ stops, Participant 7 was able to do so. Why this might be the case is explained below.

Thus, when examining the results of this study, it appears that each of the models can predict some, but not all, of the findings of this study. Two of the findings were especially difficult for all of the models to predict. The first was that the distinction between ejective/voiceless would be easier for the learners to acquire than the distinction between uvulars/velars. Because of the assimilation patterns with English and Q’eqchi’, the learners should have been better able to acquire the uvular/velar distinction than the ejective/voiceless distinction. This is because the Q’eqchi’ uvulars were perceptually assimilated to English /k/ as worse matches than the Q’eqchi’ velars. By contrast the ejective/voiceless pairs were perceived as equally good matches to English /k/. One reason for this finding might be that, although ejectives and voiceless stops were both seen as good matches to English /k/, there may have been greater perceptual saliency for the ejective/voiceless distinction than the uvular/velar distinction in aspects of the stops not examined in this study (i.e., most likely a more intense burst for ejectives). Indeed, even the native Q’eqchi’ speakers did not often make a distinction between uvular and velars in the burst duration and VOT—only the burst duration of the ejective uvular versus velar stops reached statistical significance. Thus, more than perceptual assimilation needs to be used to predict L2 learners’ ability to acquire L2 phones. It may be that other aspects of the pairs are salient, even if it is not possible to account for this based on the participants’ perceptual assimilation patterns. That may be why the relationship between perceptual assimilation and production is not clear in Table 5.

### Table 5

<table>
<thead>
<tr>
<th>NS Averages</th>
<th>Q’eq /l/</th>
<th>Q’eq /l/</th>
<th>Q’eq /k/</th>
<th>Q’eq /k/</th>
<th>Q’eq /q/</th>
<th>Q’eq /q/</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Perc (Eng /l/)</td>
<td>Prod</td>
<td>Perc (Eng /l/)</td>
<td>Prod</td>
<td>Perc (Eng /k/)</td>
<td>Prod</td>
</tr>
<tr>
<td>1</td>
<td>4/4</td>
<td>26</td>
<td>4/4</td>
<td>33</td>
<td>4/4</td>
<td>33</td>
</tr>
<tr>
<td>2</td>
<td>4/4</td>
<td>64</td>
<td>4/4</td>
<td>57</td>
<td>4/4</td>
<td>72</td>
</tr>
<tr>
<td>3</td>
<td>3/4</td>
<td>43</td>
<td>2/4</td>
<td>33</td>
<td>4/4</td>
<td>52</td>
</tr>
<tr>
<td>4</td>
<td>3/4</td>
<td>41</td>
<td>1/4</td>
<td>104</td>
<td>3/4</td>
<td>57</td>
</tr>
<tr>
<td>5</td>
<td>4/4</td>
<td>43</td>
<td>1/4</td>
<td>86</td>
<td>4/4</td>
<td>41</td>
</tr>
</tbody>
</table>

Each of the seven speakers’ perception patterns, Perc (i.e., number of times (out of 4) Q’eqchi’ consonants were perceived in terms of the most likely English consonant (Perc)) compared with their production accuracy, Prod (listed in ms).
The second finding was that difficult for the models to explain was that velars, but not alveolars, would be easier to acquire than uvulars. Since English has both alveolars and velars, but not uvulars and since the uvulars were not a good perceptual match with any existing English phone categories, all three models would predict that uvulars would be easier to acquire than both alveolars and velars. One reason for the difference between the acquisition of velars and alveolars may have been the perceptual assimilation of Q’eqchi’ velars versus alveolars onto English categories. The Q’eqchi’ to English assimilation pattern is unusual across studies on L2 learning because it represents a case where four L2 phones (Q’eqchi’ /k/, /q/, /ɪq/, /q’/) are all assimilated to a single L1 phone category (English /k/). It may be that the L2 learners focused intently on these L2 contrasts because of the crowded number of L2 phones in this limited L2 space. This focus may have helped them acquire these L2 phones more accurately than their alveolar counterparts.

In summary, these findings suggest that current models of L2 speech acquisition can account for the acquisition of languages whose phonological system varies greatly from the L1 system. However, the results of this study also suggest there may be areas where L2 speech learning models and theories need to be improved to account for all types of L1–L2 combinations. We hope that future studies examining both the perception and production of these phones will throw additional light on the predictive power of the three models discussed above.

7. Conclusion

The results of this study shed light on the production of voiceless and ejective stops by both native and L2 speakers in Q’eqchi’. These results demonstrate that there may be a larger VOT distinction between voiceless and ejective stops in Q’eqchi’ than has been reported earlier (Cho and Ladefoged, 1999). The findings also reveal that examining several native speakers’ productions of these phones is important to expand our understanding of language use, change, and maintenance in less commonly studied languages. In addition, this study demonstrates that L2 speakers of Q’eqchi’ for the most part are able to acquire ejectives accurately and to distinguish between voiceless and ejective stops. Perceptual saliency and the relationship to L1 phones seemed to predict the relative accuracy in acquiring these phones. Moreover, three L2 speech models, PAM-L2, SLM, and L2LP all, to some extent, predicted learning accuracy. These results demonstrate that examining less commonly taught languages provides important insights into L2 acquisition and speech learning.

Acknowledgments

We would like to acknowledge and thank the BYU ORCA office for the research grant making this project possible, as well as all the participants of the study, especially those native Q’eqchi’ speakers in Belize and Guatemala, for providing the necessary data to complete this research.

References

Coban, A.V. (2004). Gramática descriptiva del idioma Maya Q’eqchi’—X’tul’ab'al rix x’at’ub'1’ il xaltin'ub'al Q’eqchi’. Academia de las Lenguas Mayas de Guatemala.


