2017

Individual developmental trajectories in the L2 acquisition of Spanish spirantization

Charles Nagle
Iowa State University, cnagle@iastate.edu

Individual developmental trajectories in the L2 acquisition of Spanish spirantization

Charles Nagle

Iowa State University, USA

Abstract

In Spanish, voiced stops weaken to approximants and display variables degrees of lenition according to the context in which the stop occurs, making them a complex pronunciation feature. Accumulated findings from cross-sectional research on second language (L2) speakers suggests that many L2 learners struggle to produce the approximants even at the most advanced levels of study. The present study offers a new perspective on the approximants by studying individual learners' production of Spanish [β] over time and across phonetic contexts. Twenty-six English-speaking learners of L2 Spanish recorded two speaking tasks five times over a yearlong period corresponding to their second and third semesters of college-level language instruction. Mixed-effects models were fit to learners' C:V intensity ratio data to examine development, and stress and task type were included as substantive predictors. Although the group trajectory was flat, many learners displayed substantial change over time, including positive and negative trajectories.

Keywords: Spanish; approximants; growth curve modeling; longitudinal research; individual trajectories

1. Introduction

Over the past decade, scholars have underscored the need for more longitudinal studies of second language (L2) pronunciation development (Derwing & Munro, 2009; Derwing, 2010). Responding to these appeals, recent research has investigated inter-individual variation in L2 phonetic learning more intensively (Derwing & Munro, 2013; Derwing, Munro, & Thomson, 2008; Holliday, 2015; Munro & Derwing, 2008; Thomson, Derwing, & Munro, 2015). Examining both group and individual data, this growing body of work has demonstrated that even speakers with similar language learning histories and dispositions exhibit strikingly different developmental trajectories. Munro and Derwing (2008) investigated L1 Mandarin and L1 Slavic speakers' production of L2 English vowels over a yearlong period. For both groups, vowel intelligibility improved significantly, with most gains concentrated within the first six months of the study. However, trajectories varied by vowel target, and additional post-hoc analyses suggested that the lexical frequency of the word in which the vowel was embedded may have influenced speakers' performance over time. In a subsequent study on the L1 Slavic speakers' production of English /p/ over a 7-year period, Munro, Derwing, Thomson, and Elliot (2013) obtained similar results. Even though group /p/ intelligibility improved significantly over the first few recording sessions, individual patterns ranged from perfect intelligibility from the outset of the study to participants whose /p/ production seemed to become less intelligible over time. These results harken back to Flege (1991), who hypothesized that the ability to perceive and produce L2 stops may in fact be normally distributed in the population, though at the time most studies prioritized group means over individual cases.

Although the number of studies combining individual and longitudinal perspectives has increased over the past few years, more research in the area involving different target languages and learner populations is warranted. To date, most scholarship has focused on L2 English speakers who have relocated to a country in which English is spoken. Furthermore, studies have typically examined the acquisition of contrastive features such as novel L2 vowels. Conversely, the development of non-contrastive target sounds has received less attention. The present study therefore examined American English speakers' acquisition of the Spanish approximant allophones over a yearlong period, encompassing five data points distributed throughout learners' second and third semesters of communicative language coursework. Via the application of growth curve modeling to the production data, the present study sought to shed light on group and individual developmental patterns in the acquisition of a challenging non-contrastive L2 pronunciation feature.

1.1 Production of Medial Stops in Native English and Spanish

Before reviewing L2 studies, it is worthwhile to summarize research on native English and Spanish speakers' production of medial voiced stops, the target structure of this study. In English, underlying stops are occasionally realized as approximants in word-medial, intervocalic environments. Approximants tend to be longer than stops and exhibit greater intensity during the constriction, during which formant structure is at least partially preserved (Bouavichith & Davidson, 2013). Although studies have investigated a range of conditioning variables including vowel height and lexical frequency, stress has proven to be the most important predictor insofar as approximants are far more common in unstressed syllables. In other words, stress tends to disfavor stop consonant lenition in English (Bouavichith & Davidson, 2013; Lavoie, 2001; Ortega-Llebaria, 2004). Bouavichith and Davidson (2013) found that speakers produced /b/ as an approximant only 5% of the time in words like *libido* (/ləˈbido/), in which /b/ occurs in a stressed syllable, compared to nearly 50% of the time in words like *gazebo* (/gəˈzibo/), in which /b/ occurs in an unstressed syllable.

Compared to English, approximants are much more systematic in Spanish. By traditional accounts, underlying voiced stops weaken to approximants except after a homorganic consonant (nasals and /l/ for /d/)[1]. This process applies both within the word and across word boundaries (e.g., *la boda*: [la.βo.ða], 'the wedding'). However, the stop-approximant alternation is not categorical. Rather, research suggests that underlying stops surface with varying degrees of lenition even in environments where a stop is expected, such as after a nasal (Eddington, 2011). Moreover, the particular degree of lenition with which an underlying stop is produced depends on a number of contextual factors, including stress, the surrounding segments, and whether the stop is word-initial or -medial in sentential contexts (Cole, Hualde, & Iskarous, 1999; Colantoni & Marinescu, 2010; Lavoie, 2001; Ortega-Llebaria, 2004). In addition to these contextual sources of variation, spirantization varies widely across dialects, though approximants are strongly favored in intervocalic environments (e.g., Hualde, Simonet, & Nadeu, 2010).

The Spanish approximants pose a special challenge for English-speaking learners due to cross-linguistic mismatches in the phonetic implementation and orthographic representation of underlying stops. First, English-speaking learners must map the distribution of approximants, which is no easy task since their realization varies significantly as a function of dialect and phonological environment. Orthography further contributes to the difficulty of the learning task, especially as concerns the bilabial approximant, [β]. In Spanish, two graphemes ({b} and {v}) correspond to the same underlying phoneme (/b/) whose realization depends on the context in which it occurs ([b] or [β]). On the other hand, in English, the same two graphemes represent different phonemes, /b/ and /v/, whose realizations correspond to their underlying specification,

[b] and [v]. At the very least, these differences are likely to confuse learners in the early stages of learning since they might erroneously hypothesize that {b} represents [b] and {v} [β], drawing upon L1 phoneme-to-grapheme mappings. Finally, from a developmental perspective, errors are likely to persist since the approximants are non-contrastive. Learners face little communicative pressure to acquire the approximants because producing a stop where an approximant would be expected in native speech will not lead to errors in lexical items.

1.2 Second Language Acquisition of Spanish approximants

Studies on L2 learners' acquisition of Spanish spirantization have shown that only the most experienced L2 speakers, most of whom have spent an extended period abroad, consistently produce approximants in intervocalic position. In Face and Menke (2009), fourth semester learners, Spanish majors, and doctoral students read a short passage containing 50 tokens of /b d g/, which the authors categorized by manner: stop, fricative, approximant, or other. Whereas fourth semester learners and Spanish majors produced approximants at rates of 40% and 60%, doctoral students, who had studied abroad for 15 months and reported speaking Spanish for nearly 30 hours a week on average, supplied them 80% of the time. Alvord and Christiansen (2012) investigated a similar group of advanced L2 users who had spent two years as missionaries in a Spanish-speaking country, during which time they were paired with a native Spanish-speaking counterpart. Over half of the 34 participants demonstrated a high level of accuracy in their realization of intervocalic stops, producing approximants 80% of the time or more. Nevertheless, there was substantial between-subjects variation, ranging from one speaker who always produced approximants to a learner who produced them at a rate of only 22%. Taken together, these reports suggest that immersion catalyzes learners' acquisition of Spanish spirantization. However, repeated-measures designs including a pre- and posttest have shown that learners' production changes little after studying abroad (Díaz-Campos, 2004; Lord, 2010; Nagle, Morales-Front, Moorman, & Sanz, 2016). For example, Díaz-Campos (2004) found that study abroad participants were no more likely to produce [ð] and [ɣ] after studying in Spain for ten weeks than they were beforehand. Although Lord (2010) found that learners who had received targeted instruction on the approximants prior to studying abroad improved their production significantly more than their peers who had not received any training, the combined training and study abroad group achieved an accuracy level of only 28.7% at posttest, which underscores the difficulty of this particular Spanish pronunciation feature.

Beyond tracking learners' development of Spanish spirantization, two issues have marked the research agenda. First, studies have investigated whether learners produce varying levels of lenition according to the particular context in which the stop occurs. As with research on native speech, stress has received the most attention, and accumulated findings suggest that approximants are both more likely to occur in unstressed syllables and more likely to exhibit a higher degree of lenition in that context (Face & Menke, 2009; Shea & Curtin, 2011; Shively, 2008). In a detailed study on the subject, Shea and Curtin (2011) examined third and fifth semester learners' production of word-medial and -initial intervocalic voiced stops in stressed and unstressed contexts. Data was also collected from native Spanish speakers, who served as a comparison group. Whereas the third semester participants produced comparable C:CV intensity ratios irrespective of stress, position, and stop consonant place of articulation, fifth semester participants' production exhibited a more gradient pattern whose configuration approximated that of the native speakers. On the basis of these findings, Shea and Curtin argued that learners first implement approximants as a binary rule, acquiring the finer-grained phonetic variation present in native speech once they have accrued more L2 experience. In other words, the authors posited that intensity ratios should diverge over time as learners attune their production to the input.

Numerous studies have also addressed task type, demonstrating that learners produce more approximants on informal, semi-spontaneous speaking tasks such as an oral interview than on scripted tasks (Díaz-Campos 2006; Rogers & Alvord, 2014; Zampini, 1994). However, research in the area has typically crossed two features: whether tasks focus on form or meaning and whether they include orthography. It could be argued that scripted tasks combine a focus on form with orthography, whereas interview tasks are more communicative in nature and avoid orthographic effects. To that point, Alvord and Christiansen (2012) found that participants did not produce significantly different rates of approximantization on two scripted tasks, a word list and short story reading. Given that Spanish and English make use of the same graphemes but do not map them to underlying forms in the same way, it seems likely that taks involving orthography induce learners to produce more stop-like variants. That is, the presence of orthography may encourage learners to adhere to a more careful, hyperarticulated pronunciation, resulting in less spirantization of intervocalic stops. Conversely, when speakers are engaged in a more conversational style, rates of lenition tend to increase.

In summary, research has shown that most learners rarely produce approximants in Spanish, though they tend to supply them at higher rates in unstressed syllables and on certain informal speaking tasks. However, most studies in the area have been cross-sectional and have defined development in terms the suppliance of an approximant-like variant or lack thereof. In contrast, the present study tracked learners' production longitudinally and employed a continuous acoustic outcome measure to shed light on the gradual acquisition of this challenging L2 phonological alternation.

1.3 The present study

This study investigated English speakers' production of the Spanish approximant [β] on two tasks over five data points corresponding to learners' second and third semesters of college-level Spanish instruction. Native Spanish speakers also participated to provide L1 data. Mixed-effects models including time, stress, and task type were fit to the data to evaluate group-level development, and individual trajectories were also prioritized and analyzed. To capture subtle variation in speakers' production over time and across phonetic envrionments, degree of lenition was computed as a continuous intensity ratio (e.g., Eddington, 2011; Hualde, Simonet, & Nadeu, 2011; Ortega-Llebaria, 2004; Rogers and Alvord, 2014; Shea & Curtin, 2011). The following research questions were addressed:

1. How does learners' production of Spanish spirantization develop over two semesters of communicative Spanish language instruction?
2. Do stress and task type predict learners' production of Spanish spirantization?
3. What developmental patterns are evident at the individual level?

2. Method

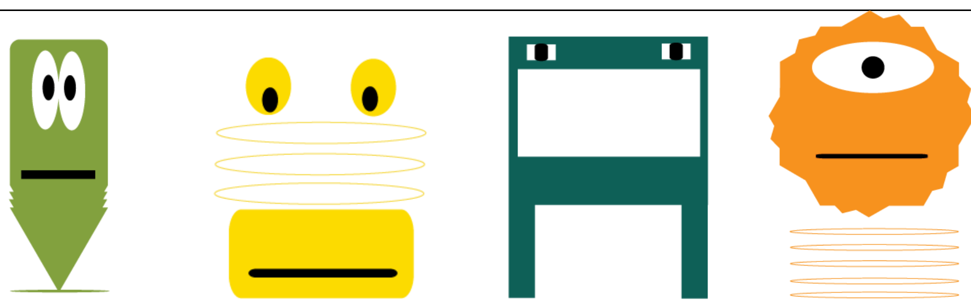
2.1. Participants

Twenty-six English-speaking learners of L2 Spanish participated during their second and third semesters of Spanish language coursework which did not include a programmatic focus on pronunciation. Participants were adult language learners (age of onset: *M* = 14.38 years, *SD* = 4.11), most of whom had taken Spanish in secondary school (previous experience: *M* = 3.35 years, *SD* = 3.17). Three participants (P13, P15, and P24) had additional language experience with Japanese, Italian, and French. This was not considered problematic because these languages do not exhibit phonological weakening of medial voiced stops. Between the second and third data points, P3 participated in a short-term study abroad program in Ecuador. Nevertheless, because his production improved prior to studying abroad, his data was retained in analyses. Due to scheduling conflicts, five participants withdrew from the study after the first semester (i.e., after the first two data points). Six native speakers of Spanish, two each from Argentina, Colombia, and Spain, participated to provide L1 comparison data. These individuals were instructors in the language program, but did not teach the L2 learners while the study was ongoing.

2.2 Spanish Target Items

As displayed in Figure 1, two nonce minimal pairs were created for the study, [lu.ˈβa.no] v. [lu.ˈpa.no] and [ru.βa.ˈlo.no] v. [ru.pa.ˈlo.no]. Regarding the phonetic properties of the target items, the intervocalic context was selected because it favors lenition in all dialects of Spanish (Hualde, Simonet, & Nadeu, 2011), making it the most stable environment across the many varieties that learners encounter as they study the language. Although Ortega-Llebaria (2004) reported that the vocalic environment did not affect native English and Spanish speakers' production of intervocalic /b/, Colantoni and Marinescu (2010) found that native speakers of Argentinian Spanish produced more lenited variants when /b/ was preceded by a back vowel, such as /u/. Consequently, /u/ was chosen as the preceding segment and /a/ as the following, the latter because /a/ is a phonologically neutral vowel. The target phone was transcribed as {b} to avoid potential transfer of /v/ from English, which may have inflated the L2 learners' level of lenition and obscured findings related to the approximant. The labial phone /b/ was examined because labial stops are more cross-linguistically similar than /d/, which is dental in Spanish and alveolar in English. Although it would have been informative to compare trajectories for /b/ and /g/, it was not feasible from a methodological standpoint to collect data for more than one stop given the scope of this longitudinal study. Given that previous research has documented robust effects for stress on speakers' lenition of voiced stops, the target characters also crossed this feature: /b/ occurs in a stressed syllable in the *Lubano-Lupano* target pair but in an unstressed syllable in *Rubalono-Rupalono*.

Nonce characters were created to achieve greater control over the phonetic composition of the target items. The target characters appeared in identical sentential contexts, embedded /b/ in phonetically identical environments (save lexical stress, which was purposefully crossed as described), and had the same lexical weight (zero) since participants had no experience with the characters before the study. Furthermore, previous research involving fictitious characters has shown them to be an effective and even entertaining means of eliciting data (Ortega-Llebaria, 2004).



Orthography	Lubano	Lupano	Rubalono	Rupalono
Phonology	/lubano/	/lupano/	/rubalono/	/rupalono/
Phonetics	[lu.ˈβa.no]	[lu.ˈpa.no]	[ru.βa.ˈlo.no]	[ru.pa.ˈlo.no]

Figure 1. Target Items.

2.3 Tasks and Procedure

Participants first completed a training module to learn the characters and the other vocabulary required for the study. They received a character packet consisting of a print-out of each of the four characters and had five minutes to review each character and memorize its name. Participants reported that this amount of time was sufficient. After receiving the character packet, they completed a computerized module containing all of the vocabulary relevant to the study. In addition to the characters, vocabulary items included ten verbs and 15 nouns drawn from students' introductory Spanish textbook. Each vocabulary item was presented over headphones and was accompanied by an image with the term printed below it. At the conclusion of the computerized vocabulary training, participants took a quiz, advancing to the speaking tasks only after they received a score of 100%. Most participants demonstrated mastery of the vocabulary on their first attempt, and only one individual had to complete the computerized portion three times.

Participants completed two speaking tasks: a picture task, on which they received images of a target character, a verb, and a noun, combining them to form a short sentence in Spanish, and a reading task, on which they received short sentences in Spanish and read them aloud. On both tasks, participants received the name of the character to prevent pausing and hesitations. Because both tasks supplied learners with the printed name of the target character, they differed only in terms of cognitive effort or focus. Whereas the picture task encouraged learners to focus on meaning to create a descriptive sentence using the images provided, the reading task was more form-focused. Both tasks were programmed using SuperLab software, but neither was timed. Rather, participants advanced at their own pace by pressing the space bar once they had recorded each sentence. At each data collection session, participants recorded 40 sentences per task, ten per target character. Recordings were carried out in a sound-attenuated booth using a dynamic, head-mounted microphone (Shure SM10A), sampled at a rate of 44100 HZ in Audacity, and exported directly to a laptop computer.

In addition to the speaking tasks, participants completed an adapted version of the Language Contact Profile (Freed, Dewey, Segalowitz, & Halter, 2004) at the first, third and fifth data collection sessions (Appendix A). Participants were asked to report on their language learning histories, including the age at which they began learning Spanish, their amount of previous Spanish coursework, whether they had learned or studied any additional languages, and whether they had traveled or lived abroad for an extended period. They furthermore estimated how often they communicated with other Spanish speakers outside of the classroom on a 5-point scale (0: *Never* to 4: *Daily*) and if they participated in any Spanish-related extracurricular activities (e.g., Spanish club, Spanish coffee hour, etc.).

2.4 Coding

To obtain a continuous outcome measure more amenable to tracking subtle shifts in production over time, degree of lenition was operationalized as an intensity ratio (Eddington 2011; Hualde, Simonet, & Nadeu, 2011; Shea & Curtin, 2011; Rogers & Alvord, 2014). Although the data set included instances of medial /p/, only the /b/ data was analyzed in the present study. Medial /b/ and the following vowel were annotated by hand using Praat acoustic analysis software (Boersma & Weenink, 2012). The stop was defined by two acoustic landmarks: onset, signaled by a decrease in intensity, evident as declining amplitude on the waveform and less robust formant structure on the spectrogram (particularly in F3 and F4), and offset, demarcated by the release burst, if present, or increasing amplitude and the reappearance of a more complete formant structure corresponding to the following vowel. The offset of the vowel was labeled at the point at which formant structure began to diminish and intensity dropped. Figure 2 displays a prototypical token for *Rubalono*. Sixty of 18,560 tokens (less than 1% of the total data set) had to be excluded from analysis due to coughing, laughing, or bumping the microphone. Praat scripts were employed to extract the minimum intensity of the consonant and the maximum intensity of the following vowel, which were subsequently used to compute C:V intensity ratios.

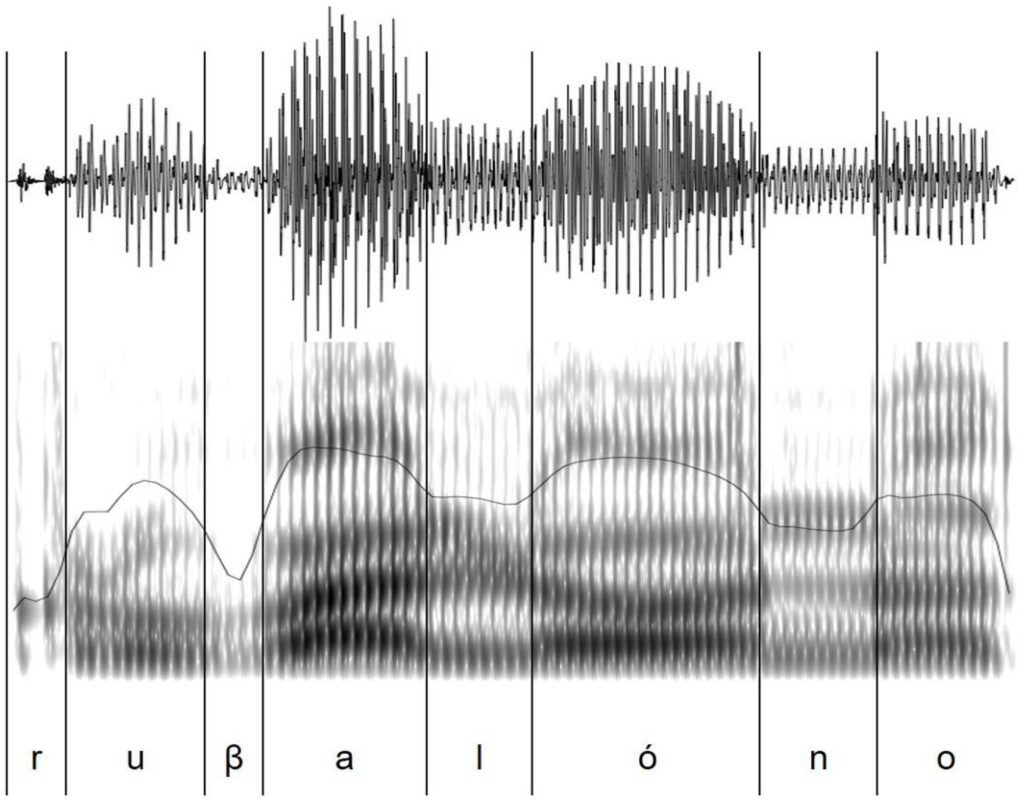


Figure 2. The phonetic characteristics of an approximant with intensity curve overlay.

2.5 Summary of Growth Curve Modeling

Because mixed-effects models have been described extensively elsewhere (Cunnings & Finlayson, 2015; Curran, Obeidat, & Lisardo, 2010; Linck & Cunnings, 2015; Murakami, 2016; Singer & Willet, 2003), only a brief overview is presented here. Growth curve models are a subclass of mixed-effects models that estimate between- and within-individual change over time through fixed and random effects. Fixed effects represent a population parameter, and random effects allow observed values to vary randomly around that estimate. One of the methodological advantages of growth curve modeling is the ability to model curvature by including higher order polynomials (e.g., quadratic and cubic slopes). Consequently, strictly linear growth need not be assumed. Furthemore, whereas complete cases are needed for analyses of variance, growth curve models estimate parameters based on available data, making them an ideal analytical tool for longitudinal data sets.

Scholars have advocated for a number of approaches to modeling depending on the nature of the research and theoretical assumptions. In confirmatory modeling, random effects are determined a priori based on the specific predictions to be tested. On the other hand, in exploratory, data-driven modeling, fixed and random effects are successively incorporated and retained only if they improve model fit. Consequently, model structure is cumulative in this approach such that each model is more complex than its predecessor. The Akaike Information Criterion (AIC) provides a rough index of model fit insofar as a decrease in AIC from one model to the next indicates an improvement. However, conducting a chi-square test on the deviance statistics of nested models fit with maximum likelihood estimation is a more robust approach. If the test is significant, then the more complex model is a better fit, and the additional parameters it contains are considered to be statistically significant. Additionally, scholars have argued that the large-sample normal approximation may be adopted to determine the significance of individual fixed effects (e.g., Linck & Cunnings, 2015). According to this criterion, absolute t values > 1.96 are treated as statistically significant at $p < 0.05$.

One additional issue that bears mentioning is centering, which facilitates the interpretation of fixed effects without altering the structure of the model. In mixed-effects modeling, an intercept is calculated for each model by setting all additional parameters to zero. This is problematic for certain predictors, such as age of onset, which is typically defined as the age at which an individual began learning the L2. Setting age of onset to zero is not coherent at a theoretical level because doing so would functionally define learners as native speakers.

Consequently, this predictor can be grand-mean centered by subtracting the mean age of onset from each individual's reported value. In the centered predictor, zero denotes an individual whose age of onset is equal to the group mean, and positive and negative values indicate ages greater than or less than the mean, respectively.

3. Results

In the present study, an exploratory, bottom-up approach to model building was adopted, following procedures described in Singer and Willett (2003). All modeling was carried out in R (version 3.1.3; R Core Team 2015) using the lme4 package (version 1.1-7; Bates, Maechler, Bolker, and Walker 2014). Stress and task were contrast-coded (unstressed = -0.5, stressed = 0.5; reading = -0.5, picture = 0.5), and age of onset (AO) and previous experience (PE) were grand-mean centered. Data for the L2 learners is presented first, including a comprehensive description of the modeling procedure and summary of the final model fit.

3.1 L2 Spanish Speakers

Table 1 summarizes the taxonomy of models fit to the learner data. Null and unconditional linear growth models (Models 1 and 2) were first fit before incorporating the stress, task, AO and PE predictor variables (Models 3, 4, and 5). According to the null model (Model 1), which generalizes across individuals and sessions (i.e., time), learners' grand-mean C:V intensity ratio production was 0.629 ($p < .001$). To gain insight into what portion of the variance lay between individuals, the intraclass correlation coefficient (ICC) was calculated by dividing between-subjects variance by total variance (cf. Singer & Willett, 2003). The ICC for the null model was 0.22, which suggests that approximately 22% of the variance in C:V intensity ratio production lay between individuals or that proportionately more variance lay within individuals. The unconditional linear growth model (Model 2) incorporated session as a continuous predictor of C:V intensity ratio to evaluate if learners' pronunciation of medial /b/ changed over time. The linear slope term did not reach statistical significance (estimate = -0.0007, SE = 0.0096, p = 0.94), indicating that the true change trajectory for the group was not statistically different from zero. Consequently, although the slope parameter was retained since it was of central interest to the present study, subsequent models focused on predicting intra-individual variation in participants' production irrespective of time. Contrast-coded stress and task predictors were integrated into Models 3 and 4. Comparing the time, stress, and task within-subjects factors, stress was by far the strongest predictor as evidenced by the proportional reduction in within-subjects variance. For time, from Model 1 to 2, within-subjects variance decreased from 0.0105 to 0.0096 (0.0009 units), which represents a proportional decrease of 0.0009 / 0.0105 = 8%. Approximately 8% of variance in C:V intensity ratio production was associated with linear time or session. In contrast, variance decreased by 0.0016 units from Models 2 to 3 (0.0096 – 0.0080 = 0.0016), which indicates that 17% of the variance (0.0016 / 0.0096 = .17) was associated with stress. On the other hand, there was virtually no reduction in within-subjects variance from Model 3 to 4, which suggests that less than 1% of variance was associated with task. Finally, grand-mean centered control predictors for age of onset (AO) and previous experience (PE) were incorporated into Model 5. Neither predictor was significant, and the between-subjects level 2 variance components for the intercept and slope changed very little (for the intercept, 0.0026 to 0.0024 or a decrease of 0.0002 units; for the slope, 0.0016 to 0.0017 or an increase of 0.0001 units).

Table 1. Taxonomy of mixed-effects models fit to the L2 learners' C:V intensity data

		Model 1	Model 2	Model 3	Model 4	Model 5
Fixed Effects						
	Intercept	0.6291***	0.6299***	0.6301***	0.6302***	0.6318***
		(0.0121)	(0.0117)	(0.0117)	(0.0117)	(0.0113)
	Stress			-0.0818***	-0.0810***	-0.0810***
				(0.0160)	(0.0028)	(0.0028)
	Task				-0.0059*	-0.0059*
					(0.0029)	(0.0029)
	AO					-0.0024
						(0.0054)
	PE					-0.0065
						(0.0071)
	Linear slope		-0.0006	-0.0007	-0.0007	-0.0007
			(0.0093)	(0.0093)	(0.0093)	(0.0093)
Variance Components						
Level 1: Within	Within-person	0.0105	0.0096	0.0080	0.0080	0.0080
		(0.1023)	(0.0982)	(0.0893)	(0.0893)	(0.0893)
Level 2: Between	Intercept	0.0029	0.0026	0.0026	0.0026	0.0024
		(0.0536)	(0.0510)	(0.0510)	(0.0510)	(0.0488)
	Slope		0.0016	0.0016	0.0016	0.0017
			(0.0406)	(0.0406)	(0.0406)	0.0407)
Goodness-of-fit	Deviance	-6692	-6957	-7690	-7693	-7695
	AIC	-6686	-6945	-7676	-7678	-7657
	BIC	-6669	-6907	-7632	-7628	-7612

Notes. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$. "Variance components" refers to the random effects structure of the models. All models included by-subject random intercepts and slopes, save the null model which did not take time/session into account. Model building grew in complexity insofar as fixed effects were integrated into each successive model, and models were compared to one another via chi-square testing.

As described, the final model included fixed effects for session (i.e., time), stress, task, AO and PE, and by-subject random intercepts and slopes. This model, refit with restricted maximum likelihood estimation, is summarized in Table 2. As a group, learners' C:V intensity ratio did not change over time, as indexed by the fact that the coefficient for the corresponding fixed effect did not reach statistical significance (estimate = -0.0007, $SE = 0.0096$, $p = 0.94$). According with the model comparisons presented above, stress was a stronger predictor than task (for stress, estimate = -0.0810, $SE = 0.0028$, $p < 0.001$; for task, estimate = -0.0059, $SE = 0.0029$, $p = 0.04$), which accords with the model comparisons presented above. On average, learners produced intensity ratios of 0.672 and 0.591 in unstressed and stressed syllables, respectively, which demonstrates that they produced a greater degree of lenition in the unstressed context. The negative coefficient for task suggests that the reading task was associated with lower intensity ratios (i.e., less lenition) than the picture task. However, these results must be interpreted with caution since task was only marginally significant. Neither AO nor PE, both of which were included as grand-mean centered control predictors, were associated with learners' intensity ratio production.

Table 2. Final mixed-effects model of L2 learners' C:V intensity ratio in intervocalic Spanish /b/.

					Random effects
		Fixed effects			By Subject
Parameters	Estimate	*SE*	*t*	*p*	*SD*
Intercept	0.6318	0.0113	51.99	< 0.001	0.0528
Linear Slope	-0.0007	0.0096	-0.07	0.944	0.0418
Stress: Stressed	-0.0810	0.0028	-28.41	< 0.001	
Task: Reading	-0.0059	0.0029	-2.05	0.040	
AO	-0.0024	0.0054	-0.41	0.688	
PE	-0.0065	0.0071	-0.84	0.411	

To bring the individual learner into sharper focus, Table 3 ranks participants by change score, the intensity ratio difference between their productions at the first and last data points of the study. Figure 3 also plots individual trajectories as a function of stress. Solid and dashed lines at ratios of 0.722 and 0.754 represent single native speaker intensity minima in stressed and unstressed syllables rather than native speaker group averages (i.e., 0.748 and 0.809). As is evident, even though modeling indicates that as a group participants did not produce significantly greater levels of lenition of medial Spanish /b/ over the course of the study, individual trajectories varied substantially. Of the 20 individuals who completed all five data sessions, nine improved their production (overall change score > 0.050), six exhibited a relatively flat trajectory (overall change score < ± 0.050), and five appeared to get worse (overall change score < -0.050). Learner 3 improved the most over the course of the study, achieving a gain of 0.191 units, and learner 16 exhibited a similar amount of change but in the opposite direction (-0.193). Additional models were fit to the group of learners whose trajectory was positive to determine if the observed development was statistically significant. Modeling indicates that C:V intensity ratio increased at a stastitically significant rate of 0.033 (SE = 0.007, p = 0.002) per semester of instruction for this group of learners.

Table 3. Descriptive statistics for L2 learner C:V intensity ratios at Time 1 and Time 5 with learners grouped by overall trajectory: positive (change > 0.05), flat (change < ± 0.05), and negative (change < -0.05).

		Time 1			Time 5		
Learner	Change	*M* (*SD*)	Min.	Max.	*M* (*SD*)	Min.	Max.
Positive Trajectory (9)							
3	0.191	0.438 (0.147)	0.229	0.662	0.628 (0.038)	0.527	0.707
7	0.082	0.558 (0.139)	0.265	0.821	0.640 (0.080)	0.447	0.745
26	0.071	0.630 (0.065)	0.534	0.865	0.701 (0.078)	0.523	0.862
9	0.063	0.610 (0.112)	0.444	0.960	0.673 (0.033)	0.631	0.770
10	0.058	0.658 (0.078)	0.506	0.859	0.716 (0.112)	0.326	0.893
15	0.057	0.661 (0.029)	0.586	0.711	0.718 (0.015)	0.690	0.749
14	0.057	0.493 (0.071)	0.270	0.635	0.550 (0.083)	0.351	0.697
18	0.056	0.671 (0.093)	0.450	0.935	0.728 (0.072)	0.388	0.808
11	0.055	0.621 (0.093)	0.342	0.723	0.675 (0.062)	0.393	0.767
Flat Trajectory (6)							
2	0.042	0.642 (0.058)	0.518	0.733	0.684 (0.040)	0.513	0.774
25	0.024	0.624 (0.119)	0.338	0.768	0.648 (0.083)	0.426	0.752
21	0.009	0.672 (0.059)	0.423	0.759	0.681 (0.054)	0.543	0.764
23	0.001	0.668 (0.083)	0.341	0.834	0.670 (0.085)	0.388	0.802
5	-0.011	0.588 (0.134)	0.356	0.863	0.577 (0.137)	0.302	0.747
24	-0.037	0.655 (0.069)	0.416	0.762	0.618 (0.075)	0.352	0.729
Negative Trajectory (5)							
8	-0.081	0.640 (0.066)	0.506	0.764	0.559 (0.163)	0.265	0.815
12	-0.081	0.644 (0.069)	0.362	0.818	0.563 (0.110)	0.258	0.748
1	-0.094	0.617 (0.127)	0.340	0.746	0.523 (0.182)	0.271	0.772
4	-0.124	0.691 (0.104)	0.444	0.899	0.567 (0.117)	0.303	0.741
16	-0.193	0.652 (0.120)	0.344	0.819	0.459 (0.167)	0.260	0.751

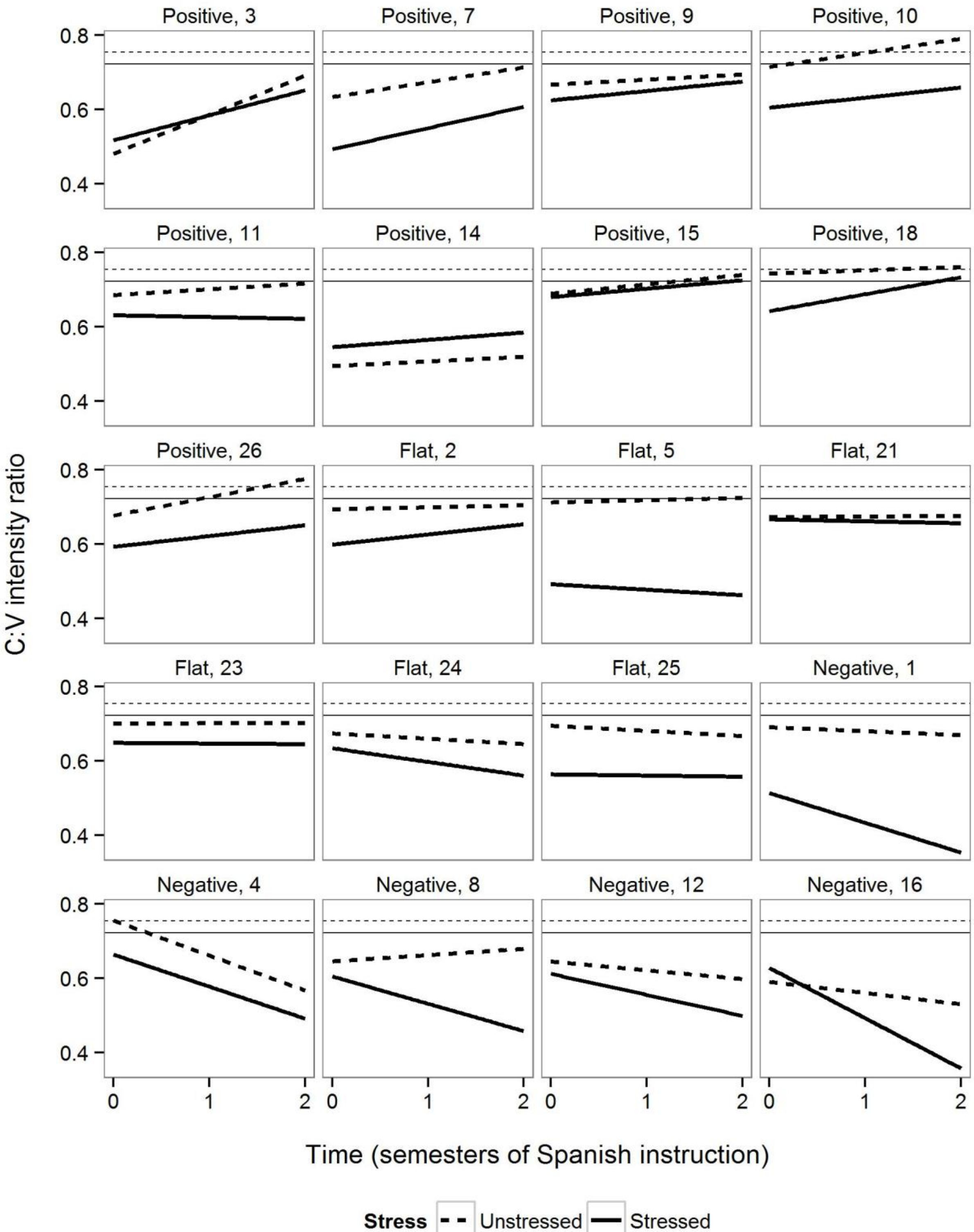


Figure 3. Individual linear C:V intensity ratio trajectories as a function of stress. Each box represents an individual L2 speaker, and speakers are grouped by their overall trajectory:

positive, flat, or negative. Increasing intensity ratios indicate higher levels of lenition, and the solid and dashed lines represent native speaker minima at 0.722 (stressed) and 0.754 (unstressed).

3.2 Native Spanish Speakers

Table 4 presents descriptive statistics for the native speaker group, and Table 5 summarizes the three models fit, a null model and two subsequent predictor models incorporating stress and task as contrast-coded fixed effects. The random effects structure of each model contained random intercepts by-subject. The ICC for the native speaker null model was 0.11, indicating that 11% of the variance lay between individuals. Native speakers' grand-mean C:V intensity ratio production was 0.778 (SE = 0.009, $p < 0.001$). Whereas including stress as a fixed effect significantly improved model fit ($\chi^2(1) = 80.80$, $p < 0.001$), including task did not ($\chi^2(1) = 0.67$, $p = 0.41$). The proportional reduction in within-subjects variance from Model 1, the null model, to Model 2 suggests that approximately 28% of variance in native speaker C:V intensity ratio was associated with stress, and the negative coefficient demonstrates that native speakers produced C:V intensity ratios that were 0.061 units lower on average when /b/ occurred in a stressed syllable. Average intensity ratios for /b/ in stressed and unstressed syllables were 0.748 and 0.809, though individual production patterns varied (Figure 3).

Table 4. Descriptive statistics for L1 Spanish C:V intensity ratios.

	Stressed			Unstressed		
Speaker	*M* (*SD*)	Min.	Max.	*M* (*SD*)	Min.	Max.
1	0.757 (0.088)	0.590	0.913	0.856 (0.044)	0.723	0.957
2	0.773 (0.046)	0.681	0.846	0.793 (0.046)	0.676	0.876
3	0.733 (0.032)	0.649	0.800	0.824 (0.023)	0.762	0.859
4	0.722 (0.051)	0.625	0.822	0.754 (0.020)	0.715	0.803
5	0.767 (0.051)	0.652	0.837	0.828 (0.023)	0.785	0.861
6	0.735 (0.053)	0.634	0.815	0.798 (0.032)	0.719	0.840

Table 5. Taxonomy of mixed-effects models fit to the native speaker C:V intensity data

		Model 1	Model 2	Model 3
Fixed Effects				
	Intercept	0.7783***	0.7782***	0.7782***
		(0.0090)	(0.0091)	(0.0091)
	Stress: Stressed		-0.0610***	-0.0610***
			(0.0062)	(0.0062)
	Task: Reading			0.0051
				(0.0062)
Variance Components				
Level 1: Within	Within-person	0.0032	0.0023	0.0023
		(0.0568)	(0.0477)	(0.0476)
Level 2: Between	Intercept	0.0004	0.0004	0.0004
		(0.0209)	(0.0209)	(0.0209)
Goodness-of-fit	Deviance	-676	-757	-757
	AIC	-670	-749	-748
	BIC	-660	-735	-730

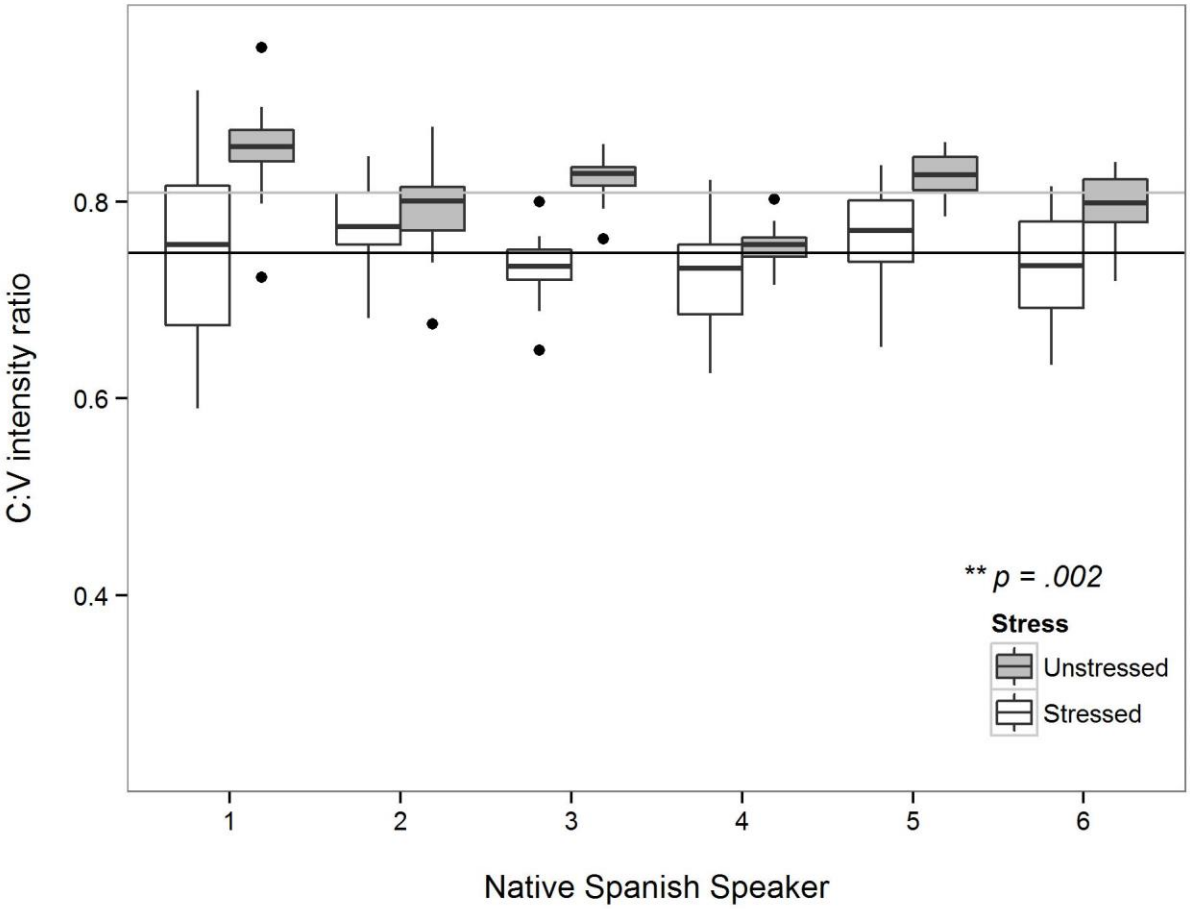


Figure 4. Boxplot of individual L1 Spanish speakers' C:V intensity ratio as a function of stress. Lines represent model-estimated means at 0.748 (stressed context) and 0.809 (unstressed context).

Summarizing results, with respect to how variance was partitioned (i.e., between v. within individuals), more variance lay within individuals for both learners and native speakers. However, the between-subjects variance in the learner data (22%) exceeded that of the native speakers (11%). Although learners exhibited little group-level development of C:V intensity ratios over time, at least nine individuals displayed a positive overall trajectory. Moreover, by the end of the study, four learners produced intensity ratios that exceeded native speaker thresholds. Learners 10 and 26 produced nativelike intensity ratios for unstressed /b/ (i.e., when /b/ occurred in an unstressed syllable), learner 15 did so for stressed /b/, and learner 18 did so in both cases. As regards the stress and task predictor variables, the former was significantly associated with degree of lenition for both groups. In contrast, task was only a weak predictor of C:V intensity ratio for the L2 learners and was not related to native speakers' production.

4. Discussion

This study tracked L1 English learners' acquisition of the Spanish approximant [β] in word-internal stressed and unstressed environments on two tasks over two semesters of language instruction, taking into account both group and individual patterns of development. Growth curve modeling of the C:V intensity ratio data demonstrated that as a group learners' production of intervocalic /b/ did not change significantly over their second and third semesters of college-level Spanish language instruction. This finding accords with previous reports documenting low rates of approximantization at the novice and intermediate levels (e.g., Zampini, 1994). However, one of the primary aims of this study was to illuminate individual trajectories in L2 phonetic learning. Although the group trajectory was flat, there were many individuals who produced higher, more targetlike intensity ratios over time, which suggests that they had begun to acquire more nativelike patterns of L2 voiced stop consonant spirantization. Of the 20 learners who supplied data at all five collection sessions, nine exhibited a positive trajectory, five a negative trajectory, and six a flat trajectory. In other words, approximately 70% of the sample exhibited some form of change, directionality aside. Moreover, among the positive trajectory learners, four individuals produced nativelike intensity ratios in at least one of the phonetic contexts under consideration (i.e., stressed v. unstressed syllable) by the conclusion of the study.

In addition to tracking development over time, this study investigated two variables that have been shown to affect L2 speakers' production of the Spanish approximant allophones.

Stress was a far more powerful predictor than task type as evidenced by model fit indices and cross-model comparisons with respect to the proportional reduction in within-individual variances. Given that the explanatory power of task was neglible, accounting for only 1% of intra-individual variance in learner intensity ratios, discussion will focus on stress. Coinciding with previous reports (Face & Menke, 2009; Shea & Cutrin, 2011; Shively, 2008), most learners produced greater intensity ratios (i.e., greater levels of lenition) when /b/ occurred in an unstressed syllable. However, two additional patterns were evident. At least three individuals (learners 9, 15, and 21) produced comparable intensity ratios in both stressed and unstressed syllables, and two individuals (learners 8 and 18) exhibited differential trajectories. For example, learner 8 produced more stop-like variants over time in stressed syllables while producing more approximant-like tokens in unstressed syllables. It could be the case that participants who did not distinguish stops occurring in stressed and unstressed syllables had simply not accrued enough L2 experience to develop contextually-defined levels of lenition (Shea & Curtin, 2011). Yet, this explanation seems unlikely since participants had similar amounts of previous experience and were enrolled in the same communicative language program. Alternatively, it could be that these novice learners were simply engaging in different forms of hypothesis testing, attempting to reconcile highly variable L2 phonetic input with conflicting L1 and L2 orthographic representations. This would explain both the diversity of developmental trajectories attested and learners' partially idiosyncratic production patterns with respect to stress. This perspective aligns with Chang (2010), who related English speakers' variable acquisition of the Korean three-way stop contrast to individual differences in cue weighting and explicit production strategies, among other factors.

In contrast to previous longitudinal work, which has typically focused on the acquisition of high functional load contrasts or features[2], the present study targeted a non-contrastive allophonic alternation in Spanish, which could be characterized as having zero functional load. Research has demonstrated that developmental trajectories for contrastive features tend to be positive, reflecting the fact that these target structures improve rapidly over the first few months of more intensive L2 contact (e.g., Munro & Derwing, 2008). On the other hand, the present study suggests that learning trajectories for non-contrastive pronunciation features are qualitatively different. Because L2 speakers face little communicative pressure to acquire this type of target sound, acquisition may depend on a broader array of factors, including individual differences in motivation, willingness to communicate, and other personal characteristics. In that regard, Alvord and Christiansen (2012) and Rogers and Alvord (2014) reported that factors such as motivation and language use predicted advanced Spanish learners' spirantization of /b, d, g/. Although the present study did not directly address this issue, it bears mentioning that the individual who improved the most (learner 3) displayed a highly integrative motivational profile, remarking that he wanted to live in a Spanish-speaking country and integrate into the culture. He also reported frequently interacting in Spanish through his extracurricular activities, which included volunteering at an English as a second language after school program and performing with a salsa dance group. Consequently, this learner could be characterized as very willing to communicate in the target language, a trait which has been linked to speakers' global pronunciation skills (e.g., Derwing & Munro, 2013). Given accumulating evidence that individual differences play a role in both global and granular aspects of L2 pronunciation development, future work investigating these variables in greater detail seems warranted. Taken together, results from previous longitudinal work and the current study underscore the need for a multifactorial approach integrating experiential, linguistic, and individual predictors to account for multiple sources of variability in L2 phonetic learning.

5. Conclusion

This study examined novice L2 Spanish learners' production of word-internal Spanish /b/ over two semesters of college-level Spanish language instruction to shed light on the acquisition of Spanish spirantization, an allophonic alternation that has proven challenging for L2 speakers. Although the aggregated group trajectory was flat, over half of the participants exhibited either a positive or negative trend. In comparison to contrastive features such as L2 vowels, which appear to undergo some degree of automatic develpoment, the acquisition of non-contrastive features may depend on individual characteristics, including a high level of motivation and willingness to communicate in the target language. From a methodological perspective, this study employed growth curve modeling (i.e., mixed-effects models) to arrive at a more multifaceted and robust exploration of within- and between-subjects variation in L2 pronunciation development over time.

6. Notes

1. Barlow (2003) has, however, advocated for a fortition approach according to which approximants are the underlying segment and stops are derived. To support this account, she draws upon distributional and language acquisition data. For example, children learning Spanish as their native language first distinguish voiced and voiceless stops in terms of continuancy and substitute /l/ for stops, which would suggest that they have analyzed spirants as the basic unit.
2. Functional load refers to the communicative value of a given L2 contrast, as determined by multiple factors such as the frequency of minimal pairs in which the contrast participates and

the density of the surrounding phonological neighborhood. For a representative study, see Munro and Derwing (2006).

7. References

Alvord, S. M., & Christiansen, D. E. (2012). Factors influencing the acquisition of Spanish voiced stop spirantization during an extended stay abroad. *Studies in Hispanic and Lusophone Linguistics, 5*(2), 239–276.

Barlow, J. A. (2003). The stop-spirant alternation in Spanish: Converging evidence for a fortition account. *Southwest Journal of Linguistics, 22*(1), 51-86.

Bates, D., Maechler, M. Bolker, B. & Walker, S. (2014). *_lme4: Linear mixed-effects models using Eigen and S4_. R package version 1.1.-7*. Retrieved from http://CRAN.R-project.org/package=lme4.

Boersma, P., & Weenink, D. (2012). Praat: doing phonetics by computer (Version 5.3.38). Retrieved from http://www.praat.org/

Bouavichith, D., & Davidson, L. (2013). Segmental and Prosodic Effects on Intervocalic Voiced Stop Reduction in Connected Speech. *Phonetica*, *70*(3), 182–206. http://doi.org/10.1159/000355635

Chang, C. (2010). The implementation of laryngeal contrast in Korean as a second language. *Harvard Studies in Korean Linguistics, 13*, 91-104.

Cole, J., Hualde, J. I., & Iskarous, K. (1999). Effects of prosodic and segmental context on /g/ deletion in Spanish. In O. Fujimura, B. D. Joseph, & B. Palek (Eds.), *Proceedings of the Fourth Linguistics and Phonetics Conference* (pp. 575–589). Prague: The Karolinum Press.

Colantoni, L., & Marinescu, I. (2010). The scope of stop weakening in Argentine Spanish. In M. Ortega-Llebaria (Ed.), *Selected Proceedings of the Fourth Conference on Laboratory*

Approaches to Spanish Phonology (pp. 100–114). Somerville, MA: Cascadilla Proceedings Project. Retrieved from http://www.lingref.com/cpp/lasp/4/paper2371.pdf

Cunnings, I., & Finlayson, I. (2015). Mixed effects modeling and longitudinal data analysis. In L. Plonsky (Ed.), *Advancing quantitative methods in second language research* (pp. 159–181). New York: Routledge.

Curran, P. J., Obeidat, K., & Losardo, D. (2010). Twelve frequently asked questions about growth curve modeling. *Journal of Cognition and Development*, *11*(2), 121–136. http://doi.org/10.1080/15248371003699969

Derwing, T. M. (2010). Utopian goals for pronunciation teaching. In J. Levis and K. LeVelle (Eds.), *Proceedings of the First Pronunciation in Second Language Learning and Teaching Conference* (pp. 24–37). Iowa State University.

Derwing, T. M., & Munro, M. J. (2013). The development of L2 oral language skills in two L1 groups: A 7-year study. *Language Learning*, *63*(2), 163–185.

Derwing, T. M., Munro, M. J., & Thomson, R. I. (2007). A longitudinal study of ESL learners' fluency and comprehensibility development. *Applied Linguistics*, *29*(3), 359–380.

Díaz-Campos, M. (2004). Context of learning in the acquisition of Spanish second language phonology. *Studies in Second Language Acquisition*, *26*(02), 249–273.

Díaz-Campos, M. (2006). The effect of style in second language phonology: An analysis of segmental acquisition in study abroad and regular-classroom students. In C. A. Klee & T. L. Face (Eds.), *Selected proceedings of the 7th Conference on the Acquisition of Spanish and Portuguese as First and Second Languages* (pp. 26–39). Somerville, MA: Cascadilla Proceedings Project.

Eddington, D. (2011). What are the contextual phonetic variants of in colloquial Spanish? *Probus*, *23*(1), 1–19.

Face, T. L., & Menke, M. R. (2009). Acquisition of the Spanish voiced spirants by second language learners. In J. Collentine, M. García, B. Lafford, & F. Marcos Marín (Eds.), *Selected Proceedings of the Eleventh Hispanic Linguistics Symposium* (pp. 39–52). Somerville, MA: Cascadilla Proceedings Project. Retrieved from http://www.lingref.com/cpp/hls/11/paper2201.pdf

Flege, J. E. (1991). Age of learning affects the authenticity of voice-onset time (VOT) in stop consonants produced in a second language. *Journal of the Acoustical Society of America*, *89*(1), 395-411.

Freed, B. F., Dewey, D. P., Segalowitz, N., & Halter, R. (2004). The language contact profile. *Studies in Second Language Acquisition*, *26*(02), 349–356. http://doi.org/10.1017/S027226310426209X

Hualde, J. I., Simonet, M., & Nadeu, M. (2011). Consonant lenition and phonological recategorization. *Laboratory Phonology*, *2*(2), 301–329.

Lavoie, L. M. (2001). *Consonant Strength: Phonological Patterns and Phonetic Manifestations*. New York: Routledge.

Linck, J. A., & Cunnings, I. (2015). The utility and application of mixed-effects models in second language research. *Language Learning*, *65*(S1), 185–207.

Lord, G. (2010). The combined effects of immersion and instruction on second language pronunciation. *Foreign Language Annals*, *43*(3), 488–503.

Munro, M. J., & Derwing, T. M. (2006). The functional load principle in ESL pronunciation instruction: An exploratory study. *System*, *34*(4), 520–531.

Munro, M. J., & Derwing, T. M. (2008). Segmental acquisition in adult ESL learners: A longitudinal study of vowel production. *Language Learning*, *58*(3), 479–502.

Munro, M. J., & Derwing, T. M. (2009). Putting accent in its place: Rethining obstacles to communication. *Language Teaching*, *42*(4), 476–490.

Munro, M. J., Derwing, T. M., & Thomson, R. I. (2015). Setting segmental priorities for English learners: Evidence from a longitudinal study. *IRAL*, *53*(1), 39–60.

Munro, M. J., Derwing, T. M., Thomson, R. I., & Elliot, D. (2013). *Naturalistic L2 segment development: Implications for pedagogy*. Paper presented at the Pronunciation in Second Language Learning and Teaching Conference, Ames, IA.

Murakami, A. (2016). Modeling systematicity and individuality in nonlinear second language development: The case of English grammatical morphemes. *Language Learning*. http://doi.org/10.1111/lang.12166

Nagle, C., Morales-Front, A., Moorman, C., & Sanz, C. (2016). Pronunciation development in the study abroad context: Methodological and programmatic considerations. In D. M. Velliaris (Ed.), *Handbook of Research on Study Abroad Programs and Outbound Mobility* (pp. 673-695). Hershey, PA: IGI Global.

R Core Team. (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Rogers, B. M. A., & Alvord, S. M. (2014). The gradience of spirantization: Factors affecting L2 production of intervocalic Spanish [β δ γ]. *Spanish in Context*, *11*(3). 402–424.

Shea, C. E., & Curtin, S. (2011). Experience, representations and the production of second language allophones. *Second Language Research*, *27*(2), 229–250. http://doi.org/10.1177/0267658310375753

Shively, R. L. (2008). L2 acquisition of [β], [ð], and [ɣ] in Spanish: Impact of experience, linguistic environment and learner variables. *Southwest Journal of Linguistics*, *27*(2), 79–114.

Singer, J. D., & Willet, J. B. (2003). *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*. New York: Oxford University Press.

Zampini, M. L. (1994). The role of native language transfer and task formality in the acquisition of Spanish spirantization. *Hispania*, *77*(3), 470–481.

Appendix A

Background Questionnaire

Part 1: Background Information
1. Sex?
2. Age?
3. Country of birth?
4. What language(s) did you grow up speaking?
5. What language(s) do you speak at home? If more than one, please list with whom you speak each language.
6. At what age did you begin learning Spanish?
7. Have you ever been to a Spanish-speaking country for more than one week? If yes, for each trip, please specify where, why (vacation, study abroad, etc.), and for how long.
8. Other than the experience listed in (7), have you ever lived in a situation in which you were exposed to a language other than your native language(s) (e.g., living in a multilingual community, exposure through family members, etc.)? If yes, please provide details.
9. Using the table provided, please rate your language ability in each of the languages that you know. Use the following ratings: 0 – Poor; 1 – Good; 2 – Very good; 3 – Native/nativelike.
10. Did you study Spanish in elementary, middle, or high school? If no, put a 0 in each of the boxes. If yes, for each level, list how long in years or semesters and indicate whether you primarily had native or non-native language instructors.
11. Have you ever practiced or received any training related to your pronunciation in Spanish? If yes, please list the type (class, a friend, an audio CD, etc.), focus (the Spanish 'r' sound, the rolling 'r,' etc.), and the length of the practice.
12. What year are you in school?
13. What is your major?

Part 2: Spanish Language Use
1. On average, how often do you communicate with native or fluent Spanish speakers?
   0: Never 1: A few times a semester 2: Monthly 3: Weekly 4: Daily
2. Using the scale listed above, please rate how often you speak with each of the following individuals in Spanish:
   a. My instructor outside of class:
   b. Friends who are native or fluent Spanish speakers:
   c. Classmates:
   d. Strangers whom I thought were Spanish speakers:
3. Using the same scale, estimate how much time you spend on average doing each of the following activities:
   a. Watching television or movies in Spanish:
   b. Reading newspapers and magazines in Spanish:
   c. Reading novels or short stories in Spanish:
   d. Listening to music in Spanish:
4. Please list any other activities that you commonly do (or have done) using Spanish.