A longitudinal study of voice onset time development in L2 Spanish stops

Charles Nagle

Iowa State University

Abstract

Recent longitudinal approaches to second language (L2) pronunciation development have prioritized developmental trajectories, highlighting individual variation in phonetic learning over time. Aligning with this research paradigm, the present study examined voice onset time (VOT) production in Spanish /b/ and /p/ over two semesters of elementary language instruction. Twenty-six native speakers of English who were novice learners of Spanish completed two L2 production tasks five times and an English production task once, designed to ascertain the frequency with which they prevoiced English voiced stops. Growth curve modeling revealed that linear and quadratic functions most accurately captured participants' L2 VOT development insofar as more gains occurred during the first half of the study. Speakers' propensity to prevoice in the L1 also predicted prevoicing in L2 Spanish /b/. However, individual results varied, including near-native learners and asymmetrical developers, individuals who improved their production of /p/ but not /b/. These results are interpreted within the frameworks of the Speech Learning Model and L2 Perceptual Assimilation Model.

Introduction

Contemporary models of cross-linguistic speech perception and production contend that second language phonetic learning remains possible throughout the lifespan. Both the Speech Learning Model (SLM; Flege 1995) and L2 Perceptual Assimilation Model (PAM-L2; Best and Tyler 2007) relate pronunciation development to perceived relationships between native and second language (L1 and L2) sounds. According to these models, L2 sounds that are similar but not identical to L1 categories are particularly problematic because listeners may not discern the subtle phonetic differences that render them distinct from their L1 counterpart. Learners' acquisition of such sounds depends on age of onset and other experiential variables such as quantity and quality of L2 experience, making L2 pronunciation development an incremental and ongoing process. Consequently, these models dovetail with a dynamic approach to SLA in which development is fractal, driven by the interaction of dynamic subsystems in a multidimensional state space (e.g., de Bot 2008; de Bot, Lowie, Thorne and Verspoor 2013).

The common thread uniting models of pronunciation development and more comprehensive approaches to SLA such as dynamic systems theory is time. As Ortega and Iberri-Shea observed, "It can be argued that many, if not all, fundamental problems about L2 learning that SLA researchers investigate are in part problems about 'time'" (2005, p. 26). Yet, despite the centrality of time to all aspects of SLA, including L2 phonetic learning, longitudinal pronunciation research is scarce, and truly multi-wave research is rarer still. In fact, many longitudinal studies examine development over two to three data points, preventing a more multifaceted investigation of the curve itself, including developmental rate and shape. For example, Riney and Flege (1999) assessed L1 Japanese speakers' production of the English /l/-/r/

contrast over a four-year window corresponding to participants' time at university. The authors identified a range of patterns, including speakers whose production seemed to get worse from the first to the fourth year. Yet, in the absence of denser multi-wave data, such broad trends remain challenging to explain.

To be sure, cross-sectional and semi-longitudinal designs have advanced our understanding of L2 pronunciation development. However, multi-wave longitudinal studies are uniquely positioned to shed light on inter-individual variation in phonetic learning, thereby contributing to L2 pronunciation models. Moreover, the incorporation of advanced statistical techniques such as growth curve modeling into mainstream SLA have made more complex analyses of group and individual trajectories possible. Drawing upon these methods, the present study contributes to the growing body of longitudinal work on L2 pronunciation development by examining L1 English-speakers' production of Spanish stop consonants over five data points distributed throughout learners' second and third semesters of general Spanish language instruction. This study therefore addressed inter-individual variation in rate and shape of L2 phonetic learning during the first few semesters of intensive communicative language coursework.

Longitudinal Perspectives on L2 Pronunciation Development

Over the past decade, longitudinal pronunciation research dealing with developmental trajectories has increased substantially, examining global constructs such as comprehensibility, fluency, and accentedness (Derwing, Munro and Thomson 2007; Derwing and Munro 2013) and L2 segments (Mora 2008; Munro and Derwing 2008; Chang 2012; Holliday 2015; Munro, Derwing and Thomson 2015; Schuhmann and Huffman 2015; Casillas 2016). Two foci can be identified within this body of work. First, Munro, Derwing, and colleagues' research has

endeavored to determine which pronunciation features tend to improve on their own without the

need for targeted instruction and to what extent those features develop. Consequently, this strand

has analyzed the assertion that L2 learners' pronunciation will improve rapidly over the first year

of more intensive L2 exposure before leveling off (Flege 1988). Munro and Derwing (2008)

studied L1 Mandarin and L1 Slavic language speakers' production of English vowels over a

yearlong period. Vowel intelligibility improved most significantly over the first half of the study,

though results varied by vowel target and L1 group. These results seem to corroborate the core

hypothesis of the SLM, namely that persistent perceptual links between similar sounds will limit

the amount of L2 phonetic accommodation that takes place, constraining development over time.

In a subsequent study examining the same speakers' fluency, comprehensibility, and

accentedness over a 7-year period, Derwing and Munro (2013) reported that whereas the L1

Slavic speakers' fluency and comprehensibility continued to improve over the course of the

study, their accentedness improved only during the 0-2 year period, and the L1 Mandarin

speakers' pronunciation did not improve at all. To explain these group differences, the authors

argued that the L1 Slavic language speakers exhibited a greater willingness to communicate in

the L2, which may have catalyzed their development in comparison to the L1 Mandarin group.

A second strand of longitudinal pronunciation research has investigated inter-individual

variation in the acquisition of L2 sounds, focusing on the acquisition of phonetic detail and

employing fine-grained acoustic analyses (Mora 2008; Chang 2012; Holliday 2015; Schuhmann

and Huffman 2015; Casillas 2016). Casillas (2016) studied ten native English speakers

participating in a domestic Spanish immersion program, collecting data on their perception and

production of Spanish stops on a weekly basis for seven weeks. Findings indicated that there was

an initial lag in development insofar as over the first few sessions learners produced little change

in voice onset time (VOT). However, after the second or third week, they began to adjust their production to fit the phonetic characteristics of Spanish, producing shorter VOT in Spanish voiceless stops and prevoicing Spanish voiced stops. Moreover, results suggested that trajectories varied a function of voicing and place of articulation.

Models of Cross-Linguistic Speech Perception and Production

Contemporary models of cross-linguistic speech perception and production such as the SLM or PAM-L2 relate L2 speech production to learners' discrimination of L2 sounds. Some L2 sounds pose little perceptual challenge because they are new, unlike any L1 sound. Conversely, similar sounds tend to be particularly difficult to perceive and produce because speakers assimilate them to a nearby L1 category, despite subtle yet significant phonetic differences between L1 and L2 phones. According to the SLM, if the learner manages to discriminate the L2 sound from its L1 attractor, then a new phonetic category will be established, resulting in a nativelike production. In this scenario, L1 and L2 categories may dissimilate over time, moving beyond the periphery of monolingual production in order to maximize the acoustic distance between them. On the other hand, if the learner does not distinguish the sounds, perceiving the L2 as an acceptable exemplar of the L1 category, then a new category will not be formed. Rather, the L2 sound will be processed via the L1 category, such that over time the sounds will assimilate, reflecting the composition of the shared cross-linguistic phonetic category. This process is known as equivalence classification. In this model, age of onset is a key determinant of category formation not because of the existence of a neurobiological critical period for language learning, but because at later ages of learning, perceptual attunement to the L1 has taken place. Consequently, L2 contrasts occurring in non-contrastive regions of the L1 phonetic space or those that make use of cues irrelevant to L1 contrasts become exceptionally difficult to

perceive and produce. Nevertheless, L2 phonetic learning remains possible throughout the

lifespan, even if the likelihood of achieving a nativelike L2 system decreases into adulthood.

Bearing many similarities to the SLM, PAM-L2 defines L2 perceptual ability in terms of

listeners' assimilation of L2 contrasts to L1 categories and their subsequent goodness-of-fit.

Discrimination is predicted to be poor in a single category assimilation because both members of

the L2 contrast are perceived as equally good or poor exemplars of an L1 category. On the other

hand, in category goodness, both L2 sounds are assimilated to a single L1 category but differ in

terms of their goodness-of-fit. If one sound is a better fit than the other, then discrimination of

the contrast may be good. According to Best and Tyler (2007), whether L2 listeners can learn to

discriminate these types of contrasts over time depends on factors such as functional load, the

relative frequency of minimal pairs in which a given L2 contrast occurs. Of the many points of

divergence between PAM-L2 and SLM, one of the most important relates to level of analysis.

According to the SLM, individuals compare phones at a "position-sensitive allophonic level"

(Flege 1995, p. 293). In contrast, PAM-L2 contends that listeners make cross-linguistic

comparisons at both the phonetic and phonological tiers, such that they may perceive two sounds

as phonetically distinct while nonetheless recognizing their phonological correspondence.

Although Best and Tyler's model focuses on L2 speech perception, because it draws upon the

direct realist approach in which the listener directly perceives speech gestures, learners'

production of a contrast (i.e., their ability to reproduce the gestural configuration and phasing

relations of L2 sounds) depends on their perception of the same information.

Drawing upon the principles of these models, specific predictions regarding

developmental trajectories and their relationship to L2 phonetic learning can be made. If

speakers perceive a L2 sound as equivalent to a nearby L1 category, blocking category

formation, then the phonetic properties of the L1 sound should influence its L2 counterpart, constraining development. Consequently, its trajectory should flatten over time, reaching an asymptote at an intermediate value well before nativelike production is achieved. Depending on the circumstances, this asymptote may be temporary since increased L2 input or targeted training could induce additional changes. If, on the other hand, a learner succeeds in establishing a new phonetic category, then no L1 anchoring effect should be observed. In other words, both developmental rate and shape (i.e., the point at which a learner's trajectory begins to flatten) provide information on the relative strength of L1-L2 relationships.

Voice onset time (VOT) and the production of L2 stops

VOT is an acoustic measure relating the release burst of the stop to the onset of voicing in the following segment. When voicing begins before the release burst, stops are said to have lead VOT, transcribed as a negative number. When voicing onset follows the release burst, stops are described in terms of lag. Stops produced with small positive VOT values are referred to as short-lag and stops with larger positive values as long-lag. In many languages in which two series of stops are contrastive, VOT is the primary phonetic cue distinguishing phonological voicing categories (Lisker and Abramson 1964). In Spanish, voiced stops are realized with voicing lead (i.e., with prevoicing), and voiceless stops are short-lag. Canonical VOT values for monolingual speakers range from 70-100 ms of prevoicing (i.e., -70 to -100 ms) for /b, d, g/ and 0-20 ms for /p, t, k/ (Williams 1977; Casteñada 1986; Rosner, López-Bascuas, García-Albea and Fahey 2000). In English, short-lag realizations in the 10-20 ms range predominate for voiced stops, although most native speakers produce at least some prevoiced tokens (Lisker and Abramson 1964, 1967; Smith 1978; Flege and Brown, Jr. 1982; Flege 1982). For /p, t, k/, VOT values of 60-90 ms are typical depending on place of articulation (Lisker and Abramson 1964;

Caramazza, Yeni-Komshian, Zurif and Carbone 1973; Flege 1987; Flege and Eefting 1988).

Consequently, VOT distributions for voiced and voiceless stops differ between the two

languages. In English, voiced stops are most often realized with short-lag VOT, whereas they are

produced with lead VOT in Spanish. Likewise, voiceless stops are long-lag in English but short-

lag in Spanish. According to the SLM, L2 learners will struggle to perceive and, consequently,

produce subtle phonetic differences such as these.

Supporting the SLM perspective, research has demonstrated that early learners are more

likely to produce nativelike stops (Caramazza et al. 1973; Flege 1987; Flege and Hillenbrand

1984; Flege and Eefting 1987a, 1987b, 1988; Flege 1991; Flege, Munro and MacKay 1995;

Stölten, Abrahamsson and Hyltenstam 2015). Flege (1995) has interpreted these findings as

evidence that early-onset SLA oftentimes results in the formation of a new phonetic category for

cross-linguistically similar sounds, including L1 and L2 stops. This new phonetic category then

enables the speaker to produce completely nativelike VOT in the L2. In contrast, adult learners

tend to produce L2 stops with VOT values typical of the L1 or with "compromise" values

intermediate to both languages (e.g., Flege 1987), which would suggest that they have simply

equated L2 stops with the corresponding L1 category or have perceived phonetic differences

between L1 and L2 stops without creating an independent L2 category. However, studies have

reported a small percentage of late learners whose attainment appears to be nativelike (e.g., Flege

and Eefting 1987b; Major 1992; Flege et al. 1995; Birdsong 2007; Stölten et al. 2015), which

underscores the fact that L2 phonetic learning in adolescence and adulthood varies considerably

across individuals as a function of their experience and personal characteristics.

In contrast to L2 speakers in immersion contexts, classroom foreign language learners

most frequently interact in the L2 with other learners and even then only on a limited basis and in

a formal context. Consequently, although one might expect classroom learners to exhibit little L2

phonetic development, studies have shown that their VOT production improves substantially

with increasing proficiency (Reeder 1998; Zampini 1998; Simon 2009; Mora 2008; Kissling

2013). For example, Reeder (1998) reported a stepwise reduction in the VOT of Spanish /p, t, k/

across four proficiency levels with the greatest shift between the level 2 intermediate and level 3

upper-division speakers (for /p/, 54 and 36 ms VOT, respectively). Zampini (1998) examined a

similar group of learners enrolled in an upper-division Spanish phonetics course. Even before

explicit instruction, participants produced English and Spanish /p/ with 60 and 25 ms VOT,

which suggests that they had already modified their L2 production considerably. Conversely,

with respect to Spanish /b/, they continued to produce short-lag variants even though they

exhibited a negative perceptual boundary in Spanish, which would indicate that they had become

sensitive to the lead/short-lag distinction.

Recent longitudinal studies on L2 stops have shown that although phonetic learning

occurs rapidly in the earliest stages of SLA, all L2 categories do not develop in tandem (Chang

2010, 2012; González López and Counselman 2013; Holliday 2015; Schuhmann and Huffman

2015; Casillas 2016). In most cases, cross-linguistic similarity appears to regulate both the

amount of L2 phonetic accommodation that occurs and L1 phonetic drift. Chang (2012) studied

English-speaking learners of Korean enrolled in an intensive six-week introductory language

course. Whereas VOT increased in Korean aspirated and English voiceless stops, parallel

changes were not evident in learners' production of the Korean fortis and English voiced

categories, arguably because learners perceived cross-linguistic phonetic differences between the

former but not the latter pair. Schuhmann and Huffman (2015) likewise found that novice L2

Spanish learners who received phonetics training improved their production of voiceless stops

more so than their production of voiced stops, and there was a stronger trend for English voiceless but not voiced VOT to drift downward toward the Spanish category.

In summary, research has shown that early learners produce more nativelike L2 stops, lending support to the notion that individuals are more likely to discern phonetic differences between similar sounds if the L2 is learned earlier in life. In adult SLA, L2 phonetic learning remains possible insofar as late learners produce more accurate L2 stops with experience. However, nativelike attainment is unlikely, possibly due to persistent perceptual links between L1 and L2 categories. Recent longitudinal studies have begun to explore inter-individual variation in the acquisition of L2 stops. The present study contributes to this line of work by examining group and individual developmental curves in the acquisition of L2 Spanish stops over two semesters of introductory language training.

The Present Study

This study examined English speakers' production of L2 Spanish /b/ and /p/ over two semesters of language instruction, taking into account their propensity to produce prevoiced /b/ in English. It was predicted that English speakers would improve their production of Spanish /p/ by reducing VOT over time. On the other hand, given that short-lag and prevoiced stops co-occur for phonologically voiced stops in English, it was expected that speakers would struggle to prevoice Spanish /b/. In other words, greater cross-linguistic overlap between Spanish and English voiced stops was predicted to constrain development of L2 prevoicing. If, as the SLM contends, cross-linguistic perceptual relationships guide production, causing linked categories to assimilate to one another, then the L1 should act as an anchor, causing development to level off before reaching a nativelike level. If, on the other hand, learners successfully establish a new category for L2 stops, then development should not slow. Participants in the present study were

late learners of Spanish who rarely interacted in the L2 outside of the language classroom.

Consequently, it was predicted that individual trajectories would slow and stabilize in most

cases, providing evidence of incomplete L2 category formation.

Method

Participants

Twenty-six L1 English speakers were recruited from multiple sections of a second-

semester communicative Spanish course at a midsized US university. Over the course of the

study, students were taught by a range of native and nonnative instructors, all of whom

completed a brief questionnaire to provide insight into any pronunciation instruction they may

have provided students as part of their regular lesson planning. Instructors reported no overt

attention to pronunciation in their teaching, except for one who discussed diphthongs and the

Spanish tap and trill.

Participants completed an adapted version of the Language Contact Profile (Freed,

Dewey, Segalowitz and Halter 2004) to provide data on the age at which they first began

learning Spanish, amount of previous experience with the language, and if they had received any

pronunciation training. They also rated their proficiency and reported on L2 use outside of the

classroom, which was minimal. The mean age of onset for the group was 14.38 years ($SD =$

4.11), and participants had on average 3.35 years of previous Spanish experience throughout

secondary school ($SD = 3.17$). Participants 13, 15, and 24 (henceforth, P13, P15, etc.) reported

experience with Japanese, Italian, and French, rating their proficiency in these languages as *very*

good, *good*, and *very good*[1]. Due to scheduling conflicts, five participants withdrew from the

study after the first two waves of data collection (i.e., after one semester), reducing the sample to

21 speakers. Six native speakers of Spanish participated to provide baseline VOT data in

Spanish. These L1 speakers were a representative sample of language instructors at the institution where data collection took place, but they did not teach participating students while the study was ongoing.

Target Items

Spanish target items were four fictitious characters contrasting /b/ and /p/ in word-initial position. The characters were minimal pairs in which the target stops occurred in stressed and unstressed syllables. Labial stops were selected for a number of reasons. First, /b/ and /p/ are cross-linguistically more similar in Spanish and English than /d/ and /t/, which differ in place of articulation (dental in Spanish v. alveolar in English). Secondly, voicing is easier to maintain for stops at more anterior points of articulation (Ohala and Riordan 1979), and research has shown that labial stops have shorter VOT values than their velar counterparts (cf., Cho and Ladefoged 1999). Therefore, it may be easier for L2 speakers to initiate prevoicing in /b/ before doing so in stops at other points of articulation and reduce VOT in /p/ given that it is closer to the L2 VOT target. To elicit VOT in English, four target words were chosen to match as closely as possible the vowel context, stress, and word length of the Spanish target characters (Table 1).

Lexical stress was taken into account because it has been shown to enhance both prevoicing of voiced stops in Spanish and aspiration of long-lag voiceless stops in English (e.g., Lisker and Abramson 1967; Casteñada, 1986; Simonet, Casillas and Díaz 2014). To that point, Simonet, Casillas and Díaz (2014) reported that monolingual English speakers produced /t/ with an average VOT of 71 ms in the unstressed context compared to 78 ms in the stressed context. The relationship between stress and the realization of prevoicing in Spanish /d/ was even more robust in that native Spanish speakers produced 50 and 69 ms of prevoicing on average (i.e., -50 and -69 ms VOT) in the unstressed and stressed environments. If English speakers are able to

acquire a Spanish mode of production over time, then they should produce more prevoicing in stressed syllables.

Table 1. Target items for the Spanish and English VOT tasks.

	/b/		/p/	
	Stressed	Unstressed	Stressed	Unstressed
Spanish	*Bafo*	*Bamuso*	*Pafo*	*Pamuso*
	/bafo/	/bamuso/	/pafo/	/pamuso/
	[ˈba.fo]	[ba.ˈmu.so]	[ˈpa.fo]	[pa.ˈmu.so]
English	*barks*	*bazookas*	*parks*	*pajamas*
	/barks/	/bazukas/	/parks/	/padʒamas/
	[paɪks]	[pə.ˈzu.kəz]	[pʰaɹks]	[pʰə.ˈdʒa.məz]

Tasks and Procedure

Participants completed two L2 production tasks, a picture task and a reading task, working with the set of word-initial stop target characters. On the picture task, students received images of a character, a verb, and an object or location, presented in that order, combining them to form a short sentence in Spanish (e.g., *Pafo corre en el parque*, 'Pafo runs in the park'). They received the name of the character to ensure that they produced it without hesitation but did not receive the other vocabulary items. On the reading task, students were presented with a sentence in Spanish that they read aloud. The sentences were similar to those they had formed on the picture task. These two tasks were hypothesized to differ with respect to the cognitive demands they placed on learners, thereby providing a more representative sample of their pronunciation under different speaking conditions[2]. There were ten sentences per target character per task or 80 total, 40 per task. Following the L2 production tasks, participants completed the English speaking task. English target words were embedded in sentences of similar length (e.g., 'Parks relax you because they are quiet'), and each sentence appeared five times.

Students took part in this study five times over the course of a calendar year, corresponding to their second and third semesters of college-level Spanish language instruction, as well as a final data point in their fourth semester. At each data collection session, participants completed the Spanish speaking tasks. They completed the full version of the language contact questionnaire at the first session and an abbreviated version at the third and fifth sessions to provide a snapshot of any Spanish-related activities they were engaging in at the time. L1 English VOT data was collected twice, at the second and fourth sessions. The L1 Spanish speakers participated in the experiment only once and did not record the English target items.

Participants completed all tasks on the computer using SuperLab software. They saw one set of images (picture task) or sentence (reading task) per screen, registered their recording, and advanced to the next trial by pressing any key. Recordings took place in a sound-attenuated booth using a dynamic, head-mounted microphone (Shure SM10A) connected to a laptop computer through a Shure X2u XRL to USB signal adapter. All recordings were made using Audacity software at a sampling rate of 44100 HZ and exported as WAV files. Character names were extracted from the recordings to facilitate acoustic analysis. Using Praat software (Boersma and Weenink 2012), VOT was annotated from the release burst of the stop to the onset of periodic vocal fold vibration. Prevoicing was coded as a negative VOT value. Thirty-nine of 18,560 tokens (0.2% of the total data) were excluded from analysis due to quality concerns, most often coughing or bumping the microphone. Once all files had been annotated in Praat, VOT was measured automatically using a script. To analyze students' English production for the L1 prevoicing predictor, VOT was analyzed in the ten tokens of English /b/ recorded at the second session. L1 prevoicing was defined as the proportion of prevoiced /b/ tokens an individual produced out of the ten trials.

Statistical approach: Growth curve modeling

Growth curve models are a subclass of mixed-effects models in which time is treated as a predictor regressed onto the outcome variable (Singer and Willet 2003; Cunnings and Finlayson 2015; Linck and Cunnings 2015). One feature of these models is the ability to include higher order polynomials (e.g., a quadratic or cubic term) to model curvature, making them an ideal analytical tool for longitudinal data sets. Growth curve models also have certain advantages over the more widely applied repeated measures analyses of variance (ANOVAs): They are robust in the face of missing data given that they estimate model parameters based on available data points, and they do not invoke some of the assumptions related to independence of observations and distribution of error.

Nested models were compared by performing a Chi-squared test on their deviance statistic. In order to do so, models were first fit using maximum likelihood estimation since restricted maximum likelihood deviance cannot be used to compare models with different fixed effects. If a more complex model was an improvement over its predecessor, then the additional parameters were taken to be significant (i.e., to have significantly improved model fit). Scholars have yet to determine exactly how to estimate p values from the t statistic that growth curve models generate. Consequently, t values of 1.96 or above were treated as statistically significant at $p < .05$ following the large-sample normal approximation (Mirman 2014). Actual p values may be slightly larger than reported here because of small-sample effects.

Following Murakami (2016), Table 2 summarizes models fit to the /b/ data. Models 1a-3b progress from an unconditional linear growth model with a random intercept (1a) to an unconditional cubic growth model with random intercepts and linear and quadratic slopes (3a), which models more complex quadratic and cubic trajectories and individual variation in rate of

change and curvature. Model 3a, the cubic model, did not significantly improve model fit ($\chi^2(1)$ = 0.66, p = .42). Consequently, 2b, the unconditional quadratic growth model, was taken as the reference model to which predictor variables were added. In addition to time, stress and L1 prevoicing (L1 pre) as primary predictors, age of learning (AoL) and previous experience (PE) were grand-mean centered[3] and incorporated into models 4a-4c as control predictors. Models 4a and 4b examined relationships between the intercept and stress and L1 prevoicing, addressing whether those predictors were related to performance. Model 4c incorporated a linear slope by L1 prevoicing parameter to determine if L1 prevoicing predicted rate of change in the development of prevoicing in L2 Spanish /b/. That is, L1 prevoicers could have produced lower overall VOT (i.e., more prevoicing) in Spanish /b/, acquired L2 prevoicing more rapidly due to their propensity to prevoice in their L1, or exhibited both characteristics. Models 4a and 4b improved fit, indicating that the stress and L1 prevoicing predictors were related to learners' L2 VOT production in Spanish /b/. However, model 4c did not improve fit ($\chi^2(1)$ = 0.04, p = .85), which suggests that L1 prevoicing was not related to rate of change. Therefore, model 4b was taken to be the final model. A similar strategy was applied to model VOT development in L2 /p/, excluding the L1 prevoicing predictor. Final models for L2 /b/ and /p/ are summarized in detail below.

Table 2. Taxonomy of Growth Curve Models Fit to VOT in L2 Spanish /b/

	Model description				Test against prior model	
Model	Fixed-effects	Random-effects (by subject)	AIC	Δ AIC	Statistic	p
1a	intercept, linear	intercept	39812			
1b	intercept, linear	intercept, linear	39629	-183	$\chi^2(2) = 186.80$	$< .001$
2a	intercept, linear, quadratic	intercept, linear	39603	-26	$\chi^2(1) = 27.87$	$< .001$
2b	intercept, linear, quadratic	intercept, linear, quadratic	39596	-7	$\chi^2(3) = 13.23$	.004
3a	intercept, linear, quadratic, cubic	intercept, linear, quadratic	39597	1	$\chi^2(1) = 0.66$	.42
4a	2b + stress, L1 pre, PE, AOL	2b	39546	-50	$\chi^2(4) = 57.71$	$< .001$
4b	**4a**	**2b + stress**	**39406**	**-140**	$\boldsymbol{\chi^2(4) = 147.92}$	**$< .001$**
4c	4a + L1 pre × linear	2b + stress	39408	2	$\chi^2(1) = 0.04$	.85

Notes. The *linear*, *quadratic*, and *cubic* parameters refer to the treatment of time/session, modeling VOT development as a straight line (linear) or with one (quadratic) or two turning points (cubic). When treated as a fixed effect, these terms examined group-level growth (i.e., development over time for this population of learners). As random effects, they explored whether individual trajectories varied randomly around the fixed effect. *Stress* was contrast-coded to obtain an ANOVA-style main effect, and *L1 pre(voicing)* was a continuous ratio. *PE = Previous Experience*. *AoL = Age of (L2) Learning*. Both variables were group mean-centered.

Results

To examine English speakers' VOT development over time, growth curve models were fit to learners' VOT in Spanish /b/ and /p/ separately using the lme4 package (version 1.1-7; Bates, Maechler, Bolker, and Walker 2014) of R (version 3.1.3; R Core Team 2015). Fixed effects included linear and quadratic session (i.e., time), stress, previous experience (PE), and age of learning (AoL). L1 prevoicing was also included in the model of VOT in L2 /b/. Stress was contrast coded (i.e., unstressed = -0.5, stressed = 0.5) to obtain an ANOVA-style main effect, and PE and AoL were grand-mean centered. L1 prevoicing was treated as a continuous fixed-effect predictor of VOT production in Spanish /b/. By-subjects random intercepts, linear and quadratic slopes, and stress were included to model inter-individual variation in those parameters. Table 3 reports the results of the final model (4b), and Figure 1 plots observed values as points and model-estimated trajectories as lines. Dotted and solid lines represent VOT production in unstressed (i.e., /bamuso/) and stressed (i.e., /bafo/) syllables.

Table 3. Growth curve model of VOT development in L2 Spanish /b/.

| | | | | | Random effects |
| | | *Fixed effects* | | | By Subject |
Parameters	Estimate	*SE*	*t*	*p*	*SD*
Intercept	8.14	3.61	2.26	.036	15.55
Linear slope	-25.74	5.58	-4.61	< .001	22.99
Quadratic slope	7.71	2.35	3.29	.003	8.56
Stress: Stressed	-8.10	3.43	-2.37	.026	15.67
L1 prevoicing	-45.57	9.72	-4.69	< .001	
PE	2.53	1.82	1.39	.177	
AoL	1.79	1.40	1.28	.212	

In order to produce more targetlike L2 stops, English speakers need to reduce VOT in L2 /b/ and /p/, prevoicing the former and producing short-lag stops for the latter. Therefore, a

decrease in VOT or a negative trajectory indicates improvement. Modeling revealed significant

main effects for linear (estimate = -25.74, *SE* = 3.61, *p* < .001) and quadratic (estimate = 7.71, *SE*

= 2.35, *p* = .003) slopes. The negative estimate for linear session indicates that participants'

average VOT production in L2 /b/ decreased at a rate of -26 ms VOT per semester of instruction,

whereas the positive estimate for the quadratic session parameter demonstrates that rate of

change diminished over the course of the study. In other words, VOT developed more rapidly at

the outset of the study and gradually leveled off over time. A significant main effect for stress

demonstrates that participants produced more prevoicing in the stressed condition (estimate = -

8.10, *SE* = 3.43, *p* = .026). The effect of L1 prevoicing was also significant (estimate = -45.57,

SE = 9.72, *p* < .001), which indicates that participants who prevoiced more often in English

produced more negative VOT values and were therefore more apt to prevoice in Spanish. Neither

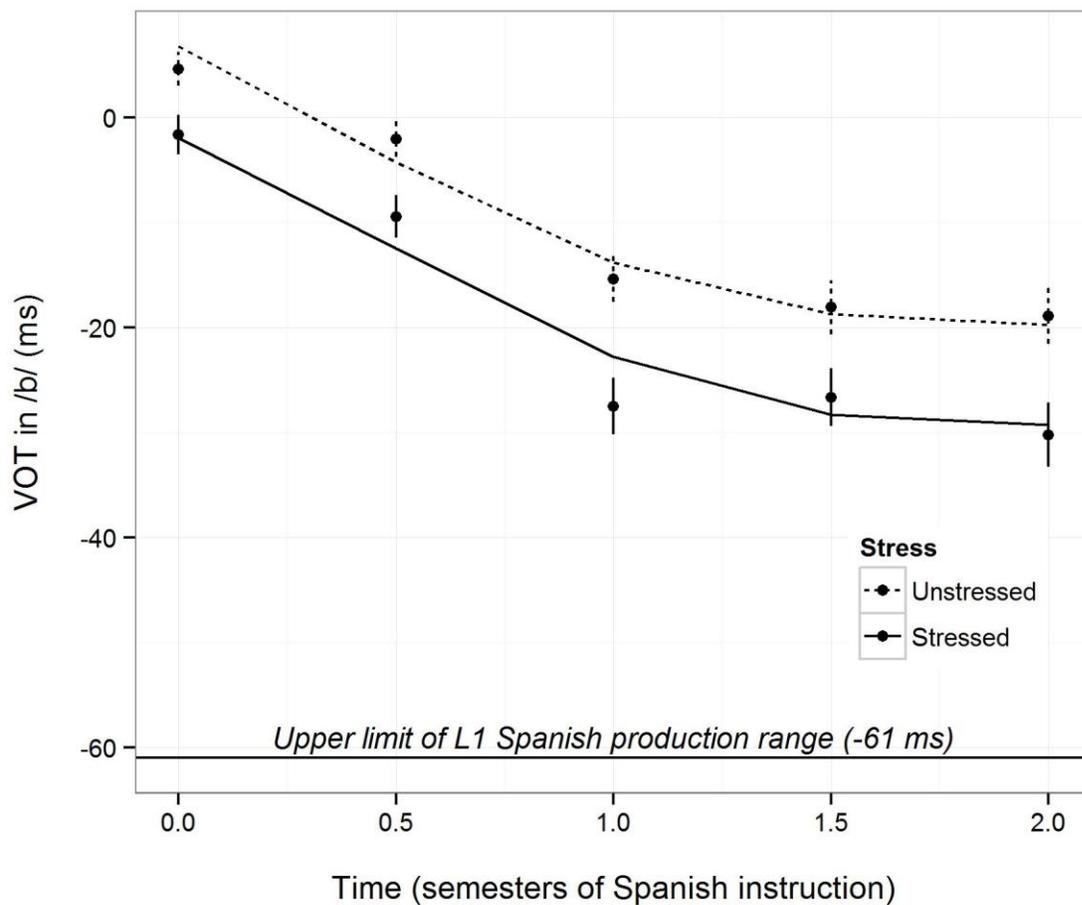PE nor AoL were related to participants' L2 VOT production.

Figure 1. VOT development in L2 Spanish /b/.

A similar strategy was employed to develop models for VOT in L2 Spanish /p/. Unlike

the model of L2 /b/, the introduction of a cubic function improved model fit for L2 /p/, ($\chi^2(1) =$

5.62, $p = .02$). However, Murakami (2016) suggested that a simpler model should be adopted

when the more complex model represents only a small improvement over its predecessor.

Therefore, the unconditional quadratic growth model was taken to be the reference model to

which predictors were added to avoid model over-specification. L1 prevoicing was not included

as a predictor in the models of L2 /p/. As reported in Table 4, the final model included stress,

AOL and PE as fixed effects. By-subject random effects included intercepts, linear and quadratic

session, and stress. Figure 2 plots VOT values in L2 Spanish /p/ over time.

Table 4. Growth curve model of VOT development in L2 Spanish /p/.

	Fixed effects				Random effects By Subject
Parameters	Estimate	SE	t	p	SD
Intercept	48.91	4.20	11.66	< .001	21.18
Linear slope	-18.65	5.21	-3.58	.001	25.74
Quadratic slope	6.31	1.97	3.20	.004	9.45
Stress: Stressed	0.48	1.47	0.33	.747	7.04
PE	3.64	2.59	1.41	.173	
AoL	3.13	2.00	1.56	.131	

The main effect for linear session was significant (estimate = -18.65, *SE* = 5.21, *p* = .001), which indicates that VOT in L2 /p/ decreased at a rate of -19 ms per semester of instruction. However, rate of change was not constant over the course of the study, as demonstrated by the quadratic session term (estimate = 6.31, *SE* = 1.97, *p* = .004). In contrast to the model of L2 /b/, in which stress emerged as a significant predictor of prevoicing, stress was not related to participants' VOT production in L2 /p/ (estimate = 0.48, *SE* = 1.47, *p* = .747), nor were PE and AoL.

In summary, models of VOT development in L2 Spanish /b/ and /p/ were similar insofar as learners produced more Spanish-like VOT over time and exhibited greater learning during the first half of the study as indexed by the statistically significant quadratic slopes. Participants' age of learning and previous experience were not related to their L2 VOT production, which is not surprising since these were late learners of Spanish with a comparable amount of L2 experience. The effect of stress varied by target phone. Stress enhanced prevoicing in L2 /b/ but was not related to participants' production of L2 /p/.
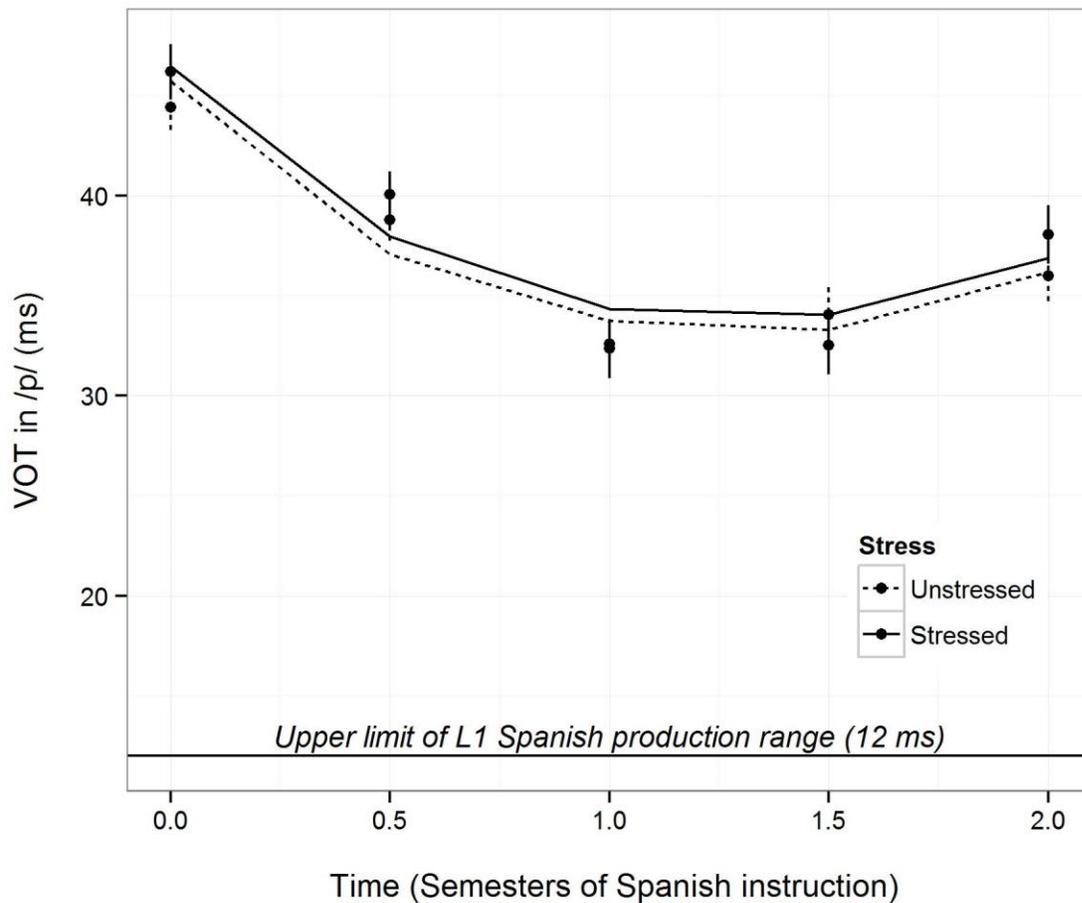
Figure 2. VOT development in L2 Spanish /p/.

The previously modeled VOT trajectories in L2 /b/ and /p/ represent a prototypical

individual, for whom L2 VOT develops quickly in the early stages of language learning and

gradually levels off. To examine individual variation in the acquisition of L2 stops, Figure 3

displays individual VOT plots for the 20 participants who completed all five waves of data

collection. Plots are grouped by L1 prevoicing category (frequent, infrequent, and non-

prevoicers) and participant number. Horizontal lines at 12 and -61 ms VOT represent the upper

limit of the native speakers' production, and shaded regions the native speaker range, which was
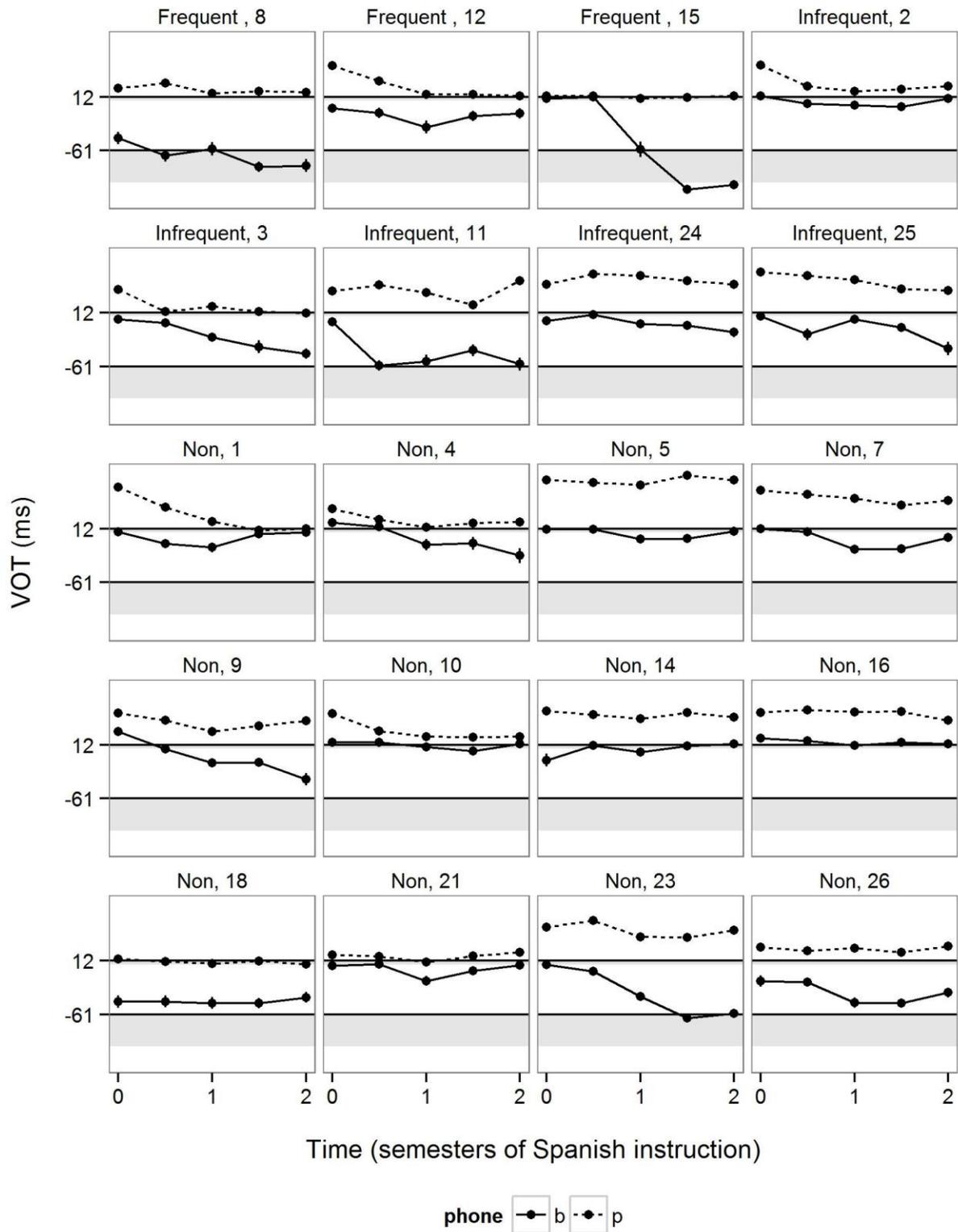
larger for /b/ than for /p/.

Figure 3. Individual developmental plots for frequent, infrequent, and non L1 prevoicers.

Inter-individual variation in developmental trajectories provides insight into the organization of learners' phonetic categories. If the formation of a new phonetic category facilitates nativelike production as the SLM claims, then it appears that both P8 and P15 had established a new category for Spanish /p/ even before the study since they both produced nativelike VOT from the first session. On the other hand, whether they did so for L2 /b/ is less obvious since both individuals were pervasive prevoicers in English, which possibly expedited their development of prevoicing in Spanish. In fact, the trajectory for P15 suggests that this may have been the case given that he demonstrated a precipitous shift from a short-lag production at the second session to categorical prevoicing at the third. In contrast to these two participants, whose production was already very targetlike from the outset, P3 (an infrequent prevoicer) made significant progress over the course of the study, achieving nativelike VOT in /p/ and near-native VOT in /b/ by the end of her third semester of Spanish instruction. Although in the absence of complementary perceptual data we cannot definitively claim that she had formed new phonetic categories, at the very least the production data suggest that some phonetic adaptation had taken place through her language coursework, especially compared to P5, P14, and P16 whose trajectories remained flat.

The performance of the asymmetrical developers (P1, P2, and P10), who reduced VOT in L2 /p/ but exhibited little prevoicing of L2 /b/, seems to indicate that they perceived differences between English and Spanish /p/ but not /b/. This explanation fits with the phonetic characteristics of English and Spanish since voiced stops are produced with lead VOT in both languages. That is, whereas in Spanish voiced stops are categorically prevoiced, prevoiced and short-lag variants co-occur in English. Detecting this type of subset relationship may have proven difficult for learners. On the other hand, distinguishing English and Spanish /p/ may have

been easier, at least from a purely phonetic perspective, since the VOT of English and Spanish

voiceless stops does not overlap. However, these conclusions must be interpreted with caution

since this study did not measure learners' perception of L2 stops.

Discussion

The present study investigated English speakers' VOT development in Spanish /b/ and /p/

over two semesters of language instruction, which did not include pronunciation training.

Speakers' propensity to prevoice English /b/ was taken into account to examine whether L1

prevoicing enhanced learners' acquisition of L2 prevoicing. Growth curve modeling

demonstrated that a quadratic function was the best representation of participants' L2 VOT

development over the two-semester period of observation insofar as most improvement took

place over the first half of the study. That is not to say, however, that a quadratic function would

appropriately represent the entire developmental curve. Rather, in the present study, the

quadratic model captured the fact that VOT development decelerated and stabilized after an

initial period of improvement and that this deceleration was statistically significant. This finding

aligns with previous longitudinal studies on vowel intelligibility (Munro and Derwing 2008) and

fluency, comprehensibility and foreign accent (Derwing and Munro 2013). With respect to the

acquisition of L2 stops, results indicate that L2 phonetic learning occurred rapidly within the first

few semesters of language instruction, corroborating previous research on the earliest stages of

language learning (Chang 2012; Holliday 2015; Schuhmann and Huffman 2015; Casillas 2016).

L1 prevoicing and stress emerged as significant predictors of participants' VOT

production in L2 Spanish /b/. Participants' tendency to prevoice English /b/ was associated with

greater levels of L2 prevoicing insofar as L1 prevoicers exhibited greater development of lead

VOT in Spanish /b/. Modeling furthermore revealed that lexical stress enhanced prevoicing but

did not affect learners' production of /p/, which suggests that participants had acquired a more

Spanish-like production pattern. Were they simply implementing stops as in English, lexical

stress should have increased VOT in /p/ since the long-lag category with which English voiceless

stops are implemented is subject to enhancement (Simonet et al. 2014). Although group

trajectories for both phones suggested an incremental decline towards more targetlike VOT

values in Spanish, individual results revealed at least four broad developmental patterns: (1)

asymmetrical developers, individuals who improved VOT production in one phone but not the

other; (2) symmetrical developers, individuals who improved VOT production in both phones;

(3) non-learners, individuals who did not seem to improve at all; (4) near-native learners,

individuals who produced nativelike VOT in at least one phone over the course of the study.

These group and individual patterns have implications for acquisitional models.

The SLM argues that accurate perceptual targets lead to accurate production such that

learners who establish a new L2 category for a similar L2 sound will produce it more accurately

than learners who perceive it as equivalent to a native category. In other words, perceptually

equating L1 and L2 sounds is predicted to limit the amount of L2 phonetic accommodation that

takes place, resulting in L2 phones whose characteristics are a composite of both languages. At

the group level, a "compromise" production was evident in both cases insofar as trajectories for

Spanish /b/ and /p/ began to level off halfway between values typical for monolingual English

and Spanish. By the end of their third semester of college-level Spanish language instruction,

participants produced Spanish /b/ with 30 ms of prevoicing and /p/ with 35 ms VOT on average.

Consequently, it appears that learners may have equated Spanish and English stop categories

while nevertheless perceiving some phonetic differences between them.

However, individual results only partially fit with this account. First, at least a few individuals produced nativelike VOT in Spanish /b/ and /p/, which suggests that they had quickly discerned VOT differences between English and Spanish stops. Moreover, many of the learners that did not achieve a nativelike production exhibited substantial improvement. Although this finding may seem straightforward, it is important to bear in mind that participants in this study were novice learners whose opportunities for L2 interaction were mostly limited to the language classroom. Consequently, results do not seem completely coherent with SLM and PAM-L2 principles, which would arguably predict less improvement in this learning scenario. On the other hand, the asymmetrical developmental patterns uncovered vis-à-vis voiced versus voiceless stops lends support to the SLM claim that perceived (dis)similarity between similar sounds drives phonetic learning. In particular, the high degree of cross-linguistic phonetic correspondence between English and Spanish voiced stops may explain why learners seemed to experience greater difficulties with /b/ than with /p/. That is, whereas English /b/ may be prevoiced, overlapping with Spanish /b/, Spanish /p/ is never aspirated (i.e., produced as a long-lag stop) like English /p/. This explanation intersects with Nathan (1987) and Simon (2009), both of whom argued that Spanish and Dutch learners of English acquired more accurate VOT in English /p/ but not /b/ due to the perceptual salience of aspiration (i.e., the greater cross-linguistic dissimilarity between L1 and L2 voiceless stops). PAM-L2 also offers a framework within which to interpret these results. According to this model, learners may have assimilated Spanish and English /p/ as two distinct phonetic categories linked by a common phonological representation, which arguably allowed them to approximate the phonetic characteristics of Spanish. On the other hand, participants may have assimilated Spanish and English /b/ as two instances of a

single phonetic (as well as phonological) category, which would explain why many struggled to prevoice Spanish /b/.

In conclusion, although learners in the present study had similar language learning histories, they nevertheless exhibited a high degree of variability in their acquisition of L2 Spanish stops. It therefore seems probable that multiple factors, including phonetic (dis)similarity, aerodynamics (e.g., Westbury and Keating 1986; Ohala 1997), and even individual differences in articulatory flexibility (Holliday 2015) and willingness to communicate (Derwing and Munro 2013), affect both the rate and shape of L2 phonetic learning. Future longitudinal work integrating perceptual measures into research on L2 pronunciation development will be needed to evaluate these claims.

Notes

1. P15 and P24 were excluded from modeling due to their previous experience with Italian and French. However, their data was included in individual analyses and plots.

2. A simple mixed-effects model was fit to the overall VOT data with task as both a fixed and by-subject random effect to determine whether participants' VOT production varied as a function of task. Task was not related to VOT production (estimate = 1.08, *SE* = 1.25, *p* = .40). Therefore, task was not introduced into the more complex developmental models discussed in this report.

3. Centering facilitates the interpretation of model parameters without changing the structure of the model. Grand mean centering subtracts the group average from individual values for a given variable. Therefore, individuals with positive values have a score above the average and individuals whose centered score is negative are below average on the target variable.

References

Bates, D., M. Maechler, B. Bolker, and **S. Walker.** 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.

Best, C. T. and **M. D. Tyler.** 2007. 'Nonnative and second-language speech perception.' In O.-S. Bohn and M. J. Munro (eds.): *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege*. John Benjamins: 13–34.

Birdsong, D. 2007. 'Nativelike pronunciation among late learners of French as a second language.' In O.-S. Bohn and M. J. Munro (eds.): *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege*. John Benjamins: 99–116.

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

Caramazza, A., G. H. Yeni-Komshian, E. B. Zurif and **E. Carbone.** 1973. 'The acquisition of a new phonological contrast: The case of stop consonants in French-English bilinguals.' *The Journal of the Acoustical Society of America* 54/2: 421–428.

Casillas, J. V. 2016. *Longitudinal development of fine-phonetic detail in late learners of Spanish* (Unpublished doctoral dissertation). University or Arizona.

Casteñada, V. M. L. 1986. 'El V.O.T. de las oclusivas sordas y sonoras españolas.' *Estudios de Fonética Experimental* 2: 92–110.

Chang, C. 2010. 'The implementation of laryngeal contrast in Korean as a second language. *Harvard Studies in Korean Linguistics* 13: 91–104.

Chang, C. 2012. 'Rapid and multifaceted effects of second-language learning on first-language speech production.' *Journal of Phonetics* 40/2: 249-268.

Cho, T. and **P. Ladefoged.** 1999. 'Variation and universals in VOT: evidence from 18 languages.' *Journal of Phonetics* 27: 207–229.

Cunnings, I. and **I. Finlayson.** 2015. 'Mixed effects modeling and longitudinal data analysis' in L. Plonsky (ed.): *Advancing Quantitative Methods in Second Language Research*. Routledge, pp. 159–181.

De Bot, K. 2008. 'Introduction: Second language development as a dynamic process.' *The Modern Language Journal* 92/2: 166–178.

De Bot, K., Lowie, W., Thorne, S. L. and **M. Verspoor.** 2013. 'Dynamic Systems Theory as a comprehensive theory of second language development.' In M. del P. García Mayo, M. J. Gutierrez Mangado, & M. Martínez Adrián (eds.): *Contemporary Approaches to Second Language Acquisition*. John Benjamins, pp. 199–221.

Derwing, T. M., **Munro, M. J.** and **R. I. Thomson.** 2008. 'A longitudinal study of ESL learners' fluency and comprehensibility development.' *Applied Linguistics* 29/3: 359–380.

Derwing, T. M. and **M. J. Munro.** 2013. 'The development of L2 oral language skills in two L1 groups: A 7-year study.' *Language Learning* 63/2: 163–185.

Flege, J. E. 1982. 'Laryngeal timing and phonation onset in utterance-initial English stops.' *Journal of Phonetics* 10: 177–192.

Flege, J. E. 1987. 'The production of "new" and "similar" phones in a foreign language: Evidence for the effect of equivalence classification.' *Journal of Phonetics* 15: 47–65.

Flege, J. E. 1988. 'Factors affecting degree of perceived foreign accent in English sentences.' *Journal of the Acoustical Society of America* 84/1: 70–79.

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.

Flege, J. E. 1995. 'Second language speech learning: Theory, findings, and problems.' In W. Strange (ed.): *Speech Perception and Linguistic Experience: Issues in Cross-language Research*. York Press, pp. 233–277.

Flege, J. E. and **W. S. Brown Jr.** 1982. 'The voicing contrast between English /p/ and /b/ as a function of stress and position-in-utterance.' *Journal of Phonetics* 10: 335–345.

Flege, J. E. and **J. Hillenbrand.** 1984. 'Limits on phonetic accuracy in foreign language speech production. *Journal of the Acoustical Society of America* 76/3: 708–721.

Flege, J. E. and **W. Eefting.** 1987a. 'Cross-language switching in stop consonant perception and production by Dutch speakers of English.' *Speech Communication* 6: 185–202.

Flege, J. E. and **W. Eefting.** 1987b. 'Production and perception of English stops by native Spanish speakers.' *Journal of Phonetics* 15: 67–83.

Flege, J. E. and **W. Eefting.** 1988. 'Imitation of a VOT continuum by native speakers of English and Spanish: Evidence for phonetic category formation.' *The Journal of the Acoustical Society of America* 83/2: 729–740.

Freed, B. F., D. P. Dewey, N. Segalowitz and **R. Halter.** 2004. 'The language contact profile.' *Studies in Second Language Acquisition* 26/2: 349–356.

González López, V. and **D. Counselman.** 2013. 'L2 acquisition and category formation of Spanish voiceless stops by monolingual English novice learners.' In J. Cabrelli Amaro, G. Lord, A. de Prada Perez and J. E. Aaron (eds.): *Selected Proceedings of the 16th Hispanic Linguistics Symposium*. Cascadilla Proceedings Project, pp. 118–127.

Holliday, J. J. 2015. 'A longitudinal study of the second language acquisition of a three-way stop contrast.' *Journal of Phonetics* 50: 1–14.

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

Lisker, L. and **A. Abramson.** 1964. 'A cross-language study of voicing in initial stops: Acoustical measurements.' *Word* 20: 384–422.

Lisker, L. and **A. Abramson.** 1967. 'Some effects of context on voice onset time in English stops.' *Language and Speech* 10/1: 1–28.

Lowie, W. 2011. 'Early L2 phonology: A dynamic approach' in M. Wrembel, M. Kul, and K. Dziubalska-Kołaczyk (eds.): *Achievements and Perspectives in SLA of Speech: New Sounds 2010*. Peter Lang, pp. 159–170.

Kissling, E. 2013. 'Teaching pronunciation: Is explicit phonetics instruction beneficial for FL learners?' *The Modern Language Journal* 97/3: 720–744.

Major, R. C. 1992. 'Losing English as a first language.' *The Modern Language Journal* 76/2: 190–208.

Mirman, D. 2014. *Growth Curve Analysis and Visualization Using R*. Taylor & Francis.

Mora, J. C. 2008. 'Learning context effects on the acquisition of a second language phonology' in C. Pérez-Vidal, M. Juan-Garau, and A. Bel (eds.): *A Portrait of the Young in the New Multilingual Spain*. Multilingual Matters, pp. 241–263.

Munro, M. J. and **T. M. Derwing.** 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.** and **R. I. Thomson.** 2015. 'Setting segmental priorities for English learners: Evidence from a longitudinal study.' *IRAL* 53/1: 39–60.

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

Nathan, G. S. 1987. 'On second-language acquisition of voiced stops.' *Journal of Phonetics* 15: 313–322.

Ohala, J. J. 1997. 'Aerodynamics of phonology' in *Proceedings of the Seoul International Conference on Linguistics*.

Ohala, J. J. and **C. Riordan, C.** 1979. 'Passive vocal tract enlargement during voiced stops' in J. J. Wolf & Klatt (eds.): *Speech Communication Papers*. Acoustical Society of America, pp. 89–92.

Ortega, L. and **G. Iberri-Shea.** 2005. 'Longitudinal research in second language acquisition: Recent trends and future directions' in *Annual Review of Applied Linguistics* 25: 26–45.

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/.

Reeder, J. T. 1998. 'English speakers' acquisition of voiceless stops and trills in L2 Spanish.' *Texas Papers in Foreign Language Education* 3/3: 101–118.

Riney, T. J. and **J. E. Flege.** 1999. 'Changes over time in global foreign accent and liquid identifiability and accuracy. *Studies in Second Language Acquisition* 20: 213–243.

Rosner, B. S., L. E. López-Bascuas, J. E. García-Albea and **R. P. Fahey.** 2000. 'Letter to the Editor Voice-onset times for Castilian Spanish initial stops.' *Journal of Phonetics* 28: 217–224.

Schuhmann, K. S. and **M. K. Huffman.** 2015. 'L1 drift and L2 category formation in second language learning.' In The Scottish Consortium for ICPhS 2015 (Ed.), Proceedings of the 18th International Congress of Phonetic Sciences. Glasgow, UK: University of Glasgow. ISBN 978-0-85261-941-4. Paper number 0850. Retrieved from http://www.icphs2015.info/pdfs/Papers/ICPHS0850.pdf.

Simon, E. 2009. 'Acquiring a new second language contrast: An analysis of the English laryngeal system of native speakers of Dutch.' *Second Language Research* 25/3: 377–408.

Simonet, M., J. V. Casillas and **Y. Díaz.** 2014. The effects of stress/accent on VOT depend on language (English, Spanish), consonant (/d/, /t/) and linguistic experience (monolinguals, bilinguals). In Speech Prosody 7: Proceedings of the 7th International Conference on Speech Prosody: Social and Linguistic Speech Prosody. Dublin, Trinity College. ISSN: 2333-2042.

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

Smith, B. L. 1978. 'Effects of place of articulation and vowel environment on "voiced" stop consonant production.' *Glossa* 12: 163–175.

Stölten, K., N. Abrahamsson and **K. Hyltenstam.** 2015. 'Effects of age and speaking rate on voice onset time.' *Studies in Second Language Acquisition* 37/1: 71–100.

Westbury, J. R. and **P. Keating.** 1986. 'On the naturalness of stop consonant voicing.' *Journal of Linguistics* 22/1: 145–166.

Williams, L. 1977. 'The voicing contrast in Spanish.' *Journal of Phonetics* 5: 169–184.

Zampini, M. L. 1998. 'The relationship between the production and perception of L2 Spanish stops.' *Texas Papers in Foreign Language Education* 3/3: 85–100.

Appendix A. Descriptive Spanish VOT Statistics for L1 Spanish Speakers and L2 Learners.

Native speakers of Spanish:

Speaker	/bafo/ M (SD)	/bamuso/ M (SD)	/pafo/ M (SD)	/pamuso/ M (SD)
1	-96 (22)	-68 (19)	9 (2)	9 (2)
2	-105 (18)	-68 (24)	9 (2)	12 (5)
3	-100 (29)	-62 (19)	7 (1)	8 (2)
4	-80 (22)	-63 (12)	8 (1)	9 (2)
5	-95 (32)	-70 (28)	10 (2)	9 (3)
6	-76 (31)	-61 (24)	10 (2)	12 (4)
Group	-91 (29)	-66 (21)	9 (2)	10 (3)

L2 Learners' Spanish VOT Productions:

	/bafo/		/bamuso/		/pafo/		/pamuso/	
	Picture M (SD)	Reading M (SD)	Picture M (SD)	Reading M (SD)	Picture M (SD)	Reading M (SD)	Picture M (SD)	Reading M (SD)
Wave 1	-3 (41)	1 (35)	2 (36)	7 (30)	50 (33)	42 (27)	48 (27)	42 (23)
2	-6 (43)	-10 (43)	-1 (36)	-2 (36)	44 (28)	37 (25)	41 (24)	36 (23)
3	-30 (58)	-28 (52)	-19 (45)	-19 (50)	35 (29)	28 (27)	35 (25)	28 (21)
4	-37 (60)	-24 (54)	-32 (57)	-16 (52)	34 (28)	29 (26)	35 (29)	30 (24)
5	-34 (63)	-34 (58)	-28 (58)	-22 (52)	40 (29)	33 (24)	35 (25)	33 (23)