John Benjamins Publishing Company

This is a contribution from *Variation and Evolution. Aspects of language contact and contrast across the Spanish-speaking world*.
Edited by Sandro Sessarego, Juan J. Colomina-Almiñana and Adrián Rodríguez-Riccelli.
© 2020. John Benjamins Publishing Company

CHAPTER 6

Did you say *peso* or *beso*?

The perception of prevoicing by L2 Spanish learners

Matthew Pollock
Indiana University

The Perceptual Assimilation Model for L2 speakers and the Speech Learning Model make predictions about the difficulty of acquisition based on pre-existing boundaries in learners' L1s. This study focuses on differences between voice onset time in English and Spanish stops, especially related to perceptual cues. Participants – 10 Spanish native speakers and 131 L1 English learners of Spanish at various levels – categorized 120 stimuli containing Spanish minimal pairs beginning with voiced and voiceless stops and the distractor /r/. Classifications varied based on acoustic manipulations of VOT, the original phone, and proficiency level. While VOT is an important determiner in perceptual boundaries, and learners can acquire L2 distinctions (although often not achieving native-like patterns), additional acoustic differences affect sound identification.

Keywords: perception, prevoicing, VOT, experimental phonetics, acquisition, L2 learning, PAM L2, SLM

1. Introduction

In Lisker and Abramson's (1964) seminal analysis of voice onset time (VOT), they determined that duration between the burst of the stop and the start of the voicing served as a phonological boundary between stop consonants. When considering results from eleven languages with two, three or four stop categories, they found a difference between monolingual English and Spanish stop tendencies.[1] English has long-lag VOT for voiceless stops and short-lag for voiced stops, while Spanish has short-lag VOT in voiceless stops and lead VOT (or prevoicing) in voiced ones (Figure 1). Other researchers have since reanalyzed differences in voice onset time

[1] Several studies have examined Spanish-English heritage and bilingual speakers' VOT systems (Amengual, 2012; Bullock & Toribio, 2009), but these are not considered in the current analysis, as findings are often variable across authors and specific speech communities.

https://doi.org/10.1075/ihll.29.06pol

across languages, including in Spanish and English, with similar results (Casillas & Simonet, 2018; Cho & Ladefoged, 1999; Flege & Eefting, 1986; García, Diehl, & Champlin, 2009; Llanos, Dmitrieva, Shultz, & Francis, 2013; Olson, 2013; Zampini, 1998). The current study set out to determine whether L2 learners of Spanish that are L1 English speakers have perceptual difficulty assimilating this distinction in stop production, given the overlap in VOT patterns of voiced English stops and voiceless Spanish ones.

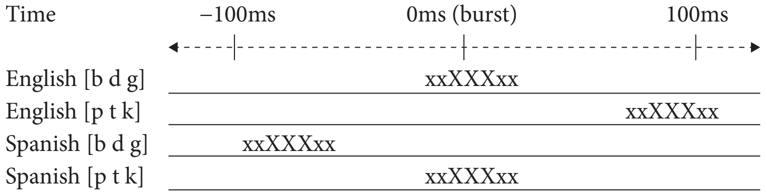


Figure 1. Distribution of English and Spanish voiced and voiceless stops (in milliseconds)

As Pallier et al. (1997) and Antoniou et al. (2012) argue, late-stage L2 learners are influenced by L1 listening tendencies when distinguishing between L2 sounds. Hunnicutt and Morris (2016) show, with the exception of some speakers in the American south, that prevoicing is rare in voiced English stops. In order to develop a new voicing boundary for the stops of their L2, there must be a shift in native English speaking Spanish learners' L1 boundaries. Strange and Schafer (2008), referring to infant acquisition of voicing, describe successful discrimination of English [b-p^h] by both native and non-native 6-month old children. Comparatively, the discrimination of Spanish [b-p] was not consistently achieved, indicating that, even before the acquisition of L1-specific boundaries at 10-12 months, listeners could not easily distinguish between the unaspirated voiced and voiceless stops.

The current study analyzed the perception of voicing by English-speaking L2 Spanish learners at a large Midwestern university, comparing them to the native and non-native graduate-level instructors of Spanish who provide much of their classroom input. It was expected that students at an advanced proficiency level would approximate the tendencies of native speakers, having developed acoustic boundaries more akin to natives than students at a basic or intermediate level. Using an online survey tool, participants categorized stimuli with digitally-manipulated stops as beginning with a voiced or voiceless sound.

2. Review of relevant literature

Although VOT serves as the primary acoustic cue in differentiating voiced and voiceless stops in Spanish; namely, the difference between lead VOT (or prevoicing) and short-lag VOT; there are secondary cues that influence speaker perceptions. Linguistic factors such as F1 or F0 (Benkí, 2005; Llanos et al., 2013), as well as social factors such as context or speaker experience (e.g., Newlin-Lukowicz, 2014) influence identification of voicing. Additionally, non-native perceptual tasks performed by naïve monolingual listeners often yield different results than bilinguals, indicating a fundamental reorganization of speaker acoustic awareness with the acquisition of additional languages (Caramazza et al., 1973; Casillas & Simonet, 2018; García-Sierra et al., 2009; Mack, 1989). These additional variables play a central role in perception, requiring consideration of (i) classroom perceptual acquisition, (ii) monolingual stop voicing in English and Spanish, (iii) bilingual acoustic spaces, and (iv) acquisitional models.

2.1 Perceptual acquisition in the language classroom

Research has shown that L1 English-speaking Spanish learners are able to perceive phonetic features in a manner approximating native-like norms (e.g., Knouse, 2012; Ringer-Hilfinger, 2012; Schoonmaker-Gates, 2017). In a meta-analysis of study abroad research, Solon and Long (2019) argue that while production of regional variants may only increase slightly after abroad experience, perceptual awareness increases both after institutional instruction in a home university as well as in immersion and study abroad environments. Schoonmaker-Gates (2017) found that explicit instruction of Castilian voiceless interdental fricative [θ] was necessary to increase student awareness of the variant – however, Solon and Long (2019) have shown that this is not necessarily always the case for acquisition of regional accent features.

Other social factors have been found to play a role in acquisition as well. Using a perception test, Schoonmaker-Gates (2012) showed that learners with higher Spanish competence were more accurate in their identification of speakers as either "native" or "non-native," indicating that learner awareness of L2 categorical distinctions increased with proficiency. This, she argued, resulted from cognitive processing – lower proficiency learners devoted more resources to deriving meaning from input, whereas advanced learners had additional resources to devote to noticing phonetic cues. Schoonmaker-Gates (2013) found similar results with respect to language exposure: learners with greater familiarity of Spanish dialects and more Spanish-speakers in their personal social networks were better able to differentiate between "native" and "non-native" speech.

The role of perception with respect to VOT has been examined before in the Spanish language classroom. Zampini (1998) focused on /p/ versus /b/, finding a lack of correlation between perception and production by learners over the course of a semester. Although there were indications of acquisition, some learners improved production and others improved perception, but none significantly improved both, leading her to claim that these are independent processes that can proceed at different rates. Language proficiency, if defined by production tendencies alone, may therefore not adequately reflect perceptual ability.

2.2 The acoustic space of monolingual Spanish and English speakers

The role of VOT in the description of stop perception has evolved over the past half century. Early classifications, such as that of Lisker and Abramson (1964, p. 399), described voiced and voiceless English stops as distinguished by aspiration and voicing. They recorded speakers and provided descriptive differences between sounds. For example, there is a distinction of around 60ms of VOT duration between voiced and voiceless stops in English. Alternatively, in Spanish, voiced stops are produced with a greater distinction on average. Unlike English, which has short- and long-lag VOT to distinguish stops, Spanish voiced stops has around 110ms prevoicing, and voiceless stops has a short-lag VOT of around 15ms. Many other accounts approached VOT and aspiration in a similar way, collecting descriptive statistics across various languages, and treating VOT as a key acoustic cue in distinguishing stops (Cho & Ladefoged, 1999; Flege & Eefting, 1986; Abramson & Lisker, 1972). It was only within the last couple decades that perspectives on VOT have broadened, looking at cues such as stop closure duration, lexical stress, formant levels (both F1 and F0), and unknown "others" that have an effect on the perception of both native speakers and language learners.[2]

These studies suggest that speakers are influenced by more than just the onset of voicing, and raise the question of the exact role of VOT in the perception of language learners. Learners must recognize a variety of new perceptual cues, only some of which are salient in their L1. In the case of English and Spanish, Simonet (2012) and Martínez-Celdrán (1993) found that while there are differences in the closure-duration of Spanish voiced and voiceless stops, the same distinction does not exist for English. Formant levels serve as an additional acoustic cue used to identify voicing distinctions across languages, and, like closure-duration, there is an effect of L1. Benkí (2005) reported that monolingual Spanish and English speakers, when

[2] Although these cues come out of both production and perception studies, evidence from SLA research indicates that both can provide insight into perceptual boundaries, given their interrelated nature (George, 2014; Knouse, 2012; Schoonmaker-Gates, 2017; Solon & Long, 2019).

presented with stimuli from the two languages, used F1 in similar ways to distinguish voicing; however, Llanos et al. (2013) found that English bilinguals relied on F0, while Spanish bilinguals relied on prevoicing to make that distinction. Finally, some studies have questioned the role of VOT as a distinguishing feature altogether. When considering short-lag [t], Bohn and Flege (1993) showed that bilinguals and monolingual English and Spanish speakers identified voicing consistently in stimuli, but not corresponding to VOT, or any single other acoustic dimension.

A complex combination of linguistic and extralinguistic factors seems to influence a speaker's perceptual space. Olmstead et al. (2013, p. 5) determined that, when naïve monolingual Spanish and English speakers imitated both native and non-native words, their native stops were able to reflect L1 phonemic differences in VOT, but their attempt to imitate non-native stops did not accurately reflect L2 phonemic differences. Even in an explicit imitation task, speakers' L1 served as the basis for distinction. In addition to L1, language exposure also influences L2 awareness. Llanos and Francis (2016) instructed native Spanish speakers with both high and low levels of bilingual knowledge of English to categorize Spanish stops produced by native English speakers as either voiced or voiceless. Participants answered based on *English* VOT boundaries, albeit with their success varying according to their exposure to English and the amount of context provided in the stimuli. Hearers with greater knowledge of an L2 can therefore use contextual and social information like accent to adjust VOT boundary identification.

Social factors have also been found to play a role in perception. Using stimuli with manipulated stops and varying durations of VOT, Abramson and Lisker (1972) showed that native speakers of Spanish used prevoicing as a distinctive feature in English when mediated by the duration of contact with English. One speaker, who had studied English for thirteen years and lived in the US for five, identified a boundary between /k/ and /g/ that was much closer to monolingual English than Spanish norms. Flege and Eefting (1986) further found that speaker age also affected perception – both native Spanish and English adults required longer-lag VOT than children to identify /t/, and English adults used greater prevoicing than children when producing /d/. This could have resulted from shifting acoustic boundaries or changes in processing abilities.

2.3 Acoustic and acquisitional tendencies of bilinguals

Early descriptions of bilingualism in acquisitional phonology described bilingual speakers as having a single perceptual space, with an intermediary boundary that compromised L1 and L2 norms. Caramazza et al. (1973), working with perceptual and production data from English- and French-Canadian speakers, found that bilinguals more closely approximated the norms of the language they were stronger in. They relied on VOT more than French monolinguals and less than English ones, and were not found to adapt these tendencies when moving between different language contexts. Williams (1977) argued that bilinguals distinguished voiced from voiceless stops at a single point in perception, despite maintaining VOT differences in production.

More recent studies have contributed to a view of bilingual phonetic systems as complex, taking into account individual differences such as input, cognitive load and proficiency. Birdsong (2018) argues that bilinguals are perpetually influenced by both of their language systems, and as such, it cannot be expected that two monolingual speakers operate within them. Instead, bilinguals differ categorically from monolinguals, with individual differences placing their perception in a constant state of flux. This is important, because we would therefore expect to find key differences between bilingual and monolingual systems that perceptual studies can key in on. Despite some of these individual and situational differences, such as age of acquisition, frequency, and acoustic specifications, Flege (2005) argues that learners with sufficient input and time should come to perceive phonetic properties of L2 speech accurately. Greater differences between L1 and L2 sounds may lead to new category creation rather than assimilation, but in cases with minimal differences, as in the case of Spanish and English stops, L1 and L2 categories are predicted to merge.

Comparing Spanish VOT boundaries to Quichua and Media Lengua, Stewart (2018) averaged voicing onset tendencies across the Spanish-speaking world, finding that speakers had an average 110ms difference between voiced and voiceless stops (Figure 2). Media Lengua was determined to rely more heavily on the Spanish superstrate VOT system as a result of the high cognitive load placed on bilingual speakers, who relied on the superstrate as a means of maintaining phonetic distinction in the new system they were constructing.

Stop:	p/	/t/	/k/	/b/	/d/	/g/
Average:	11	14	29	−109	−100	−89

Figure 2. Average VOT for speakers of Spanish in milliseconds (Stewart, 2018)

When analyzing the VOT of young Spanish-English bilinguals, Flege and Eefting (1987) argued that individuals had separate phonological categories for each language. L1 Spanish-speaking bilinguals were found to produce voiceless English stops differently from native speakers, which they attributed to limited input: their L1 Spanish caregivers did not produce English with native-like VOT patterns, meaning children did not acquire English phonological boundaries. An additional difficulty is that even once L2 perceptual boundaries are acquired, bilinguals do not necessarily stabilize their internal system (Amengual, 2012; Bullock & Toribio, 2009; Olson, 2013). They can instead demonstrate asymmetrical transfer, going so far as to employ L2 VOT boundaries in their L1. These studies show that social and cognitive factors play a role in the "native-ness" of bilingual speech production and perception.

2.4 Theoretical models of L2 perception

Following an organizational schema used by Schmidt (2018), I describe two important models of second language perception below in light of this study's theoretical interests: The Perceptual Assimilation Model for L2 Speakers (PAM L2: Best & Tyler, 2007) and the Speech Learning Model (SLM: Flege, 1995).

The PAM L2 serves as an extension of the original PAM model, which made predictions about naïve listeners' acoustic spaces, expanding to make predictions about language learners and the ease with which they would assimilate certain L2 categories into their native system. Best and Tyler (2007) predict that although learners start out comparable to monolingual listeners, assimilating non-L1 sounds based on the similarities to their existing system, they have the potential to develop their L2 system based on increased experience and exposure to the language. Over time, they integrate L2 cues into their system and an even approach the successful discrimination shown by native speakers. The resulting interlanguage system, melding phonetic and phonological levels, allows learners to determine the functional equivalence of phonology across both languages, even though the language-specific phonetic perception might differ. This model would predict two possible interpretations of the Spanish and English voiceless systems.

One possible description of the voicing distinction under PAM L2 would result in a Two-Category Assimilation, where "the two non-native phones are perceived as acceptable exemplars of two different native phonemes" (p. 23). If this were the case, L1 English listeners would be predicted to acknowledge both Spanish voiced and voiceless stops as akin to the English category, although with a consistent shift in VOT. The other possibility would be a Category Goodness Assimilation Contrast, where "both L2 phonological categories are perceived as equivalent to the same L1 phonological category, but one is perceived as being more deviant" (p. 29). In this case, learners would be expected to treat the prevoiced stimuli as more deviant when compared to the Spanish voiceless stop, with minimal short-lead VOT. It is unclear prima facie whether classroom learners of an L2 perceive two equally acceptable categories, or whether both Spanish stops are considered to be more and less deviant productions of the English voiced stop.

According to the second model, SLM, perceptual systems undergo equivalence classification between the L1 and L2 in bilinguals' phonological space (Flege, 1995). Phonetic information is shared across languages, and, in the case of perceived similarities, L2 sounds are equated with the closest pre-existing L1 categories. Like PAM, two L2 sounds that are assimilated into a single L1 category will be difficult to distinguish under SLM, while two sounds perceived as different lead to the development of a new phonetic category. In this way, the elements of the L1 and L2 phonetic systems of bilingual speakers exist together, such that each can impact the other and cause bidirectional interference over the course of a speaker's life. In the current case, this results in an effect resembling PAM L2. If the similarities are great, speakers -- even at a relatively low proficiency and with minimal language exposure -- should adopt the L2 Spanish voicing categories. If the differences are larger, proficiency should have less of an effect.

Both of these perceptual models permit predictions regarding the difficulty learners will have in assimilating L2 sounds into previously-established L1 categories. In both cases, the degree of perceived difference of the Spanish voicing cues from English ones should affect the difficulty speakers have in adopting a new category, or adjusting the boundaries of a pre-existing one.

2.5 Research questions

In this study, L1 English learners of Spanish at varying proficiency levels identified manipulated Spanish stimuli as voiced or voiceless. Their responses demonstrated how perceptual boundaries shift as language learners' L2 familiarity increases. A control group of native Spanish bilinguals established a prototypical boundary for L1 Spanish voicing. The expectation was that, if this division fell into the Two Category Assimilation of PAM L2 (Best & Tyler, 2007), or allowed easy Equivalence Classification under SLM (Flege, 1995), learners would group Spanish norms into pre-existing English voiced and voiceless categories, with increasingly native-like classifications tied to individual proficiency. On the other hand, if this division were better described by a Category Goodness Assimilation, equivalence classification would not be possible initially because the sounds were sorted into a single category. If this is the case, then learners would have greater confusion in separating voiced and voiceless stops, which would be reflected in difficulty distinguishing stops by voicing at varying levels of proficiency. With the difference of these two models in mind, one question that motivated this investigation was: to what extent do L1 English speakers follow native-like perception norms when identifying Spanish stops, and how does this reflect the two types of perceptual classifications described by PAM L2 or SLM?

Cues other than VOT have been shown to affect participants' ability to distinguish between voiced and voiceless stops (Benkí, 2005; Bohn & Flege, 1993; etc.). Social factors like speaker proficiency, individual differences and experience abroad also play a role (Flege, 1995; Williams, 1977), as well as linguistic cues such as place of articulation (Lisker & Abramson, 1964). Given the array of possible factors involved, the second question guiding this research was: what social and linguistic constraints govern listeners' perceptions of stop constraints, and how important overall was VOT specifically to their classification?

3. Methodology

3.1 Participants

A digital survey[3] hosted on Qualtrics was carried out at a large public university in the American Midwest during the spring and fall semesters of 2018. The survey, which participants completed in an average of 30 minutes, was (i) disseminated digitally to some students to complete outside of class as extra credit, and (ii) presented to some in-class as a "linguistic experiment." Native and graduate speakers were recruited, taking the study on their own. A total of 197 participants answered the survey; however, 36 submissions with less than 50% finished, 12 retakes by participants with technical problems, seven submissions lasting between two and 100 hours, as well as two participants who listed their gender as non-binary (resulting in a statistically unbalanced group) were excluded from the final analysis. The majority of the 140 remaining participants were students of Spanish (*n* = 118) in the fourth (*n* = 44), sixth (*n* = 36), and eighth (*n* = 38) semester of their undergraduate career. The other participants included non-native (*n* = 12) and native (*n* = 10) graduate instructors of Spanish.

3.2 Production and manipulation of stimuli

Stimuli from three speakers were recorded: one female and two male speakers of Iberian Spanish. Initially, a female Colombian Spanish speaker was also recorded, but her productions included unexpected short-lead and long-lag voicing that differed from previous studies so, in order to control for speaker differences, only individuals from a single region were included. Each speaker read a list of minimal pairs printed on a page of paper, which included stimuli beginning with stops, as well as distractors beginning with /r/ (Figure 3). Distractors were included to ensure that participants paid attention throughout the survey. Each word was saved as separate .mp3 sound files, with 20ms of preceding pause and 30ms following.

[3] Available at <https://go.iu.edu/1T1h>

(1)	*peso*	weight	(7)	*teja*	weave	(13)	*kia*	kia
(2)	*beso*	kiss	(8)	*deja*	leave	(14)	*guía*	guide
(3)	*reto*	challenge	(9)	*reja*	railing	(15)	*ría*	estuary
(4)	*pata*	paw	(10)	*trama*	plot	(16)	*cano*	white-hiared
(5)	*bata*	robe	(11)	*drama*	drama	(17)	*gano*	I win
(6)	*rata*	rat	(12)	*rama*	branch	(18)	*rana*	frog

Figure 3. List of minimal pairs used in the survey

Across the six minimal pairs, three stops were followed by a front, high/mid vowel (i.e., /i/ and /e/), and three others were followed by /a/. Inclusion of minimal pairs with stops followed by back vowels (e.g., *cofre* 'trunk,' *gofre* 'waffle') was considered, but given the low frequency of many such pairs, and the number of stimuli already included in the survey, only the above eighteen were used. The average duration of the original VOT for each phoneme, as produced by the three native speakers, is presented in Table 1.

Table 1. Average duration and variance of VOT for stimuli (ms)

Phone	VOT	St. Deviation
/p/	10.2	2
/t/	17	3.1
/k/	27.3	11.5
/b/	−103.7	14.3
/d/	−91.7	17.2
/g/	−65.9	21.3

For each stimulus beginning with a stop, the original file was altered in Praat (Boersma and Weenink, 2017) to create three "manipulated" production types that formed a spectrum between the word (e.g., *bata*), and its minimal pair (e.g., *pata*). Durations were calculated from the onset of voicing to the edge of prevoicing or VOT, and this sound was cut or pasted into files to create four voicing "types". For example, for *bata*, the audio with "Full prevoicing" had no manipulation (prevoicing = 109.4 ms, Figure 4). "Half prevoicing" had exactly half the prevoicing bar from the original sound removed (prevoicing = 54.7 ms, Figure 5). "Zero prevoicing" had all prevoicing removed (prevoicing = 0, Figure 6). Finally, "Full short lag VOT" had the VOT from the voiceless minimal pair (i.e., *pata*) copied and pasted into the audio file with zero prevoicing, creating a word with voicing resembling *pata* (VOT = 10.3 ms, Figure 7).

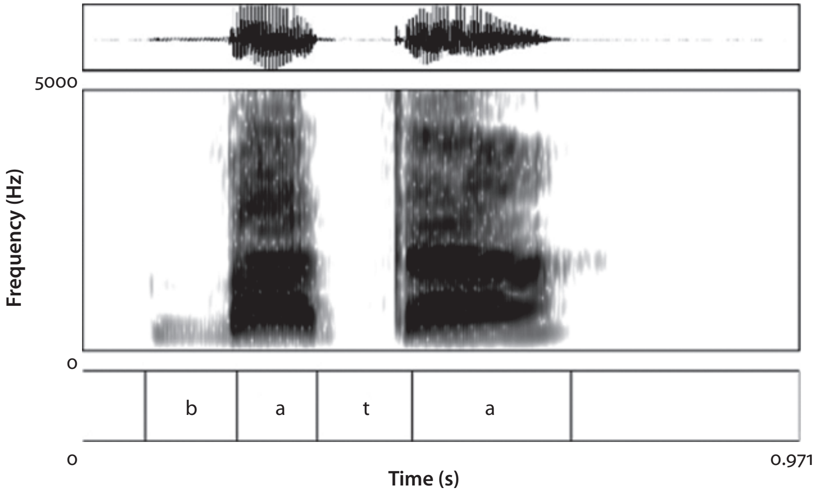


Figure 4. Full prevoicing

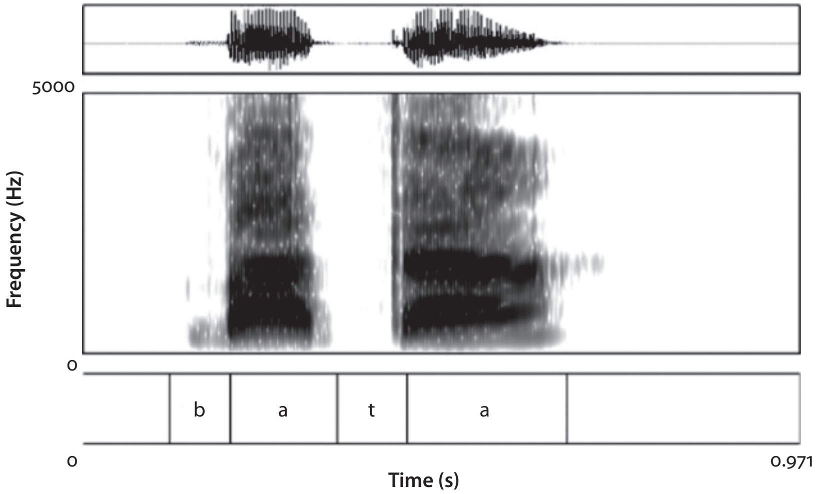


Figure 5. Half prevoicing

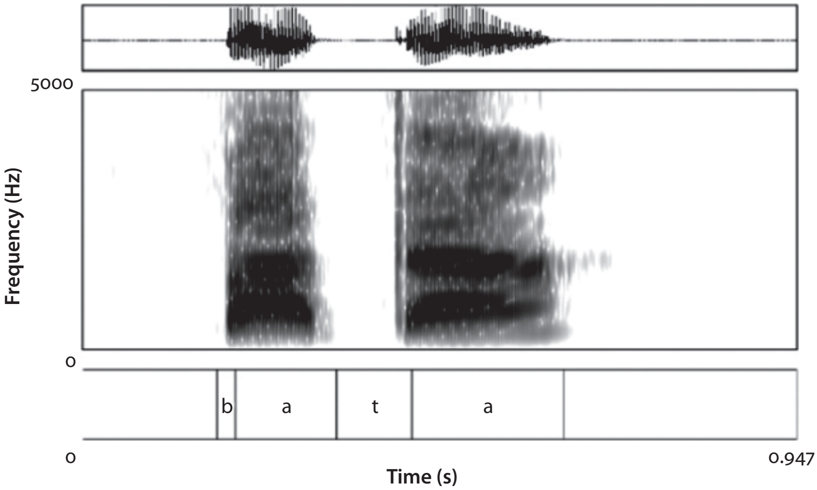


Figure 6. Zero prevoicing

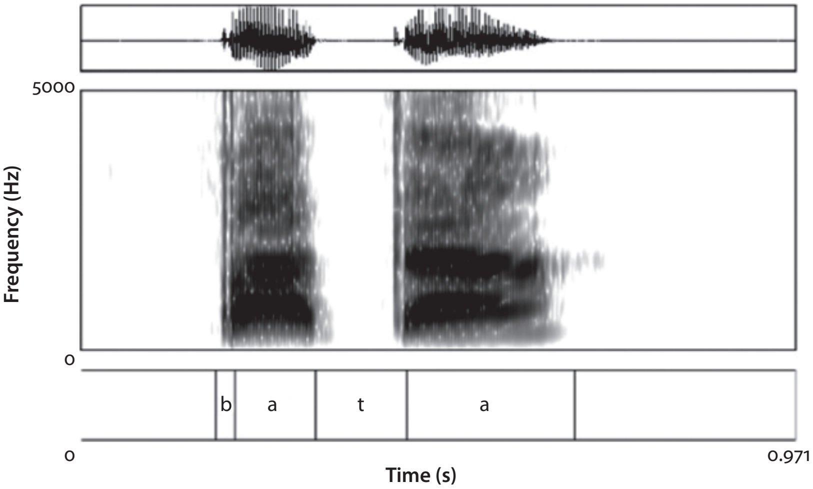


Figure 7. Full Short Lag VOT

The same process was carried out in the opposite direction as well. Thus, *pata* stimulus had (1) an unmanipulated "Full short lag VOT" stimulus, (2) a "Zero prevoicing" stimulus with VOT removed, (3) a "Half Prevoicing" stimulus with half the prevoicing of *bata* added before the burst, and (4) a "Full Prevoicing" stimulus with the prevoicing from *bata* added. Average values for the duration of each minimal pair are shown in Table 2.

Table 2. Average duration and variance of manipulated stimuli by place of articulation (ms)

	Full prevoicing		**Half prevoicing**		**Zero prevoicing**		**Full short lag VOT**	
	Duration	**Var.**	**Duration**	**Var.**	**Duration**	**Var.**	**Duration**	**Var.**
Bilabial (/p/ /b/)	−103.66	14.30	−51.83	7.15	0.00	0.00	10.24	1.99
Dental (/t/ /d/)	−91.69	17.23	−45.84	8.61	0.00	0.00	17.03	3.15
Velar (/k/ /g/)	−65.95	21.32	−32.97	10.66	0.00	0.00	27.33	11.51

In order to reduce variability between stimuli, the duration of each vowel following the stop was measured and averaged across minimal pairs. For example, if the female actor had produced an /a/ in *pata* of 100ms duration, and an /a/ in *bata* with 50ms duration, 25ms was cut from the vowel in *bata* and added to the one in *pata*, meaning that all of the stimuli for *pata* and *bata* had an /a/ with a duration of 75ms. A group of 20 focus-testers, including native and near-native Spanish speakers, were presented with the files and did not remark on infelicities concerning these vowels.

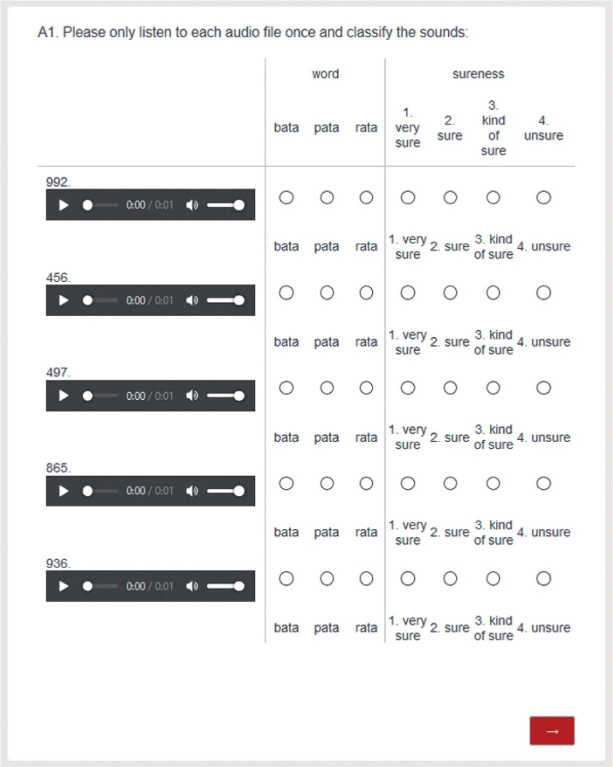
In the end, the study contained 120 total stimuli: 2 (minimal pair) × 4 (manipulation) × 2 (speaker)[4] × 6 (total pairs) = 96 experimental stimuli + 24 distractors beginning with /r/.

3.3 Instrument design

According to Thomas (2011), perception studies should employ normalized stimuli with built-in pauses and distractors to obfuscate the goal of the researcher. As a result, after a basic background questionnaire (Appendix 1) following Melero-García and Cisneros (2018), the instrument administered a vocabulary test. A secondary goal of this test, which used images for the minimal pairs in Figure 3, was to prime students with the lexical items in the survey.[5]

[4] To balance the number of tokens from male and female speakers, three of the six minimal pairs produced by Male Actor 1 and three of the six by Male Actor 2 were used, averaging out to 2 speakers for any given minimal pair.

[5] Minimal pairs were not controlled for lexical frequency or cognate status, which, as a reviewer noted, may have influenced perception (Amengual, 2012; Connine et al., 1993).



A1. Please only listen to each audio file once and classify the sounds:

	word			sureness			
	bata	pata	rata	1. very sure	2. sure	3. kind of sure	4. unsure
992.	○	○	○	○	○	○	○
	bata	pata	rata	1. very sure	2. sure	3. kind of sure	4. unsure
456.	○	○	○	○	○	○	○
	bata	pata	rata	1. very sure	2. sure	3. kind of sure	4. unsure
497.	○	○	○	○	○	○	○
	bata	pata	rata	1. very sure	2. sure	3. kind of sure	4. unsure
865.	○	○	○	○	○	○	○
	bata	pata	rata	1. very sure	2. sure	3. kind of sure	4. unsure
936.	○	○	○	○	○	○	○
	bata	pata	rata	1. very sure	2. sure	3. kind of sure	4. unsure

Figure 8. Sample survey page

Following a test page allowing students to adjust their headphone levels, the classification task began (Figure 8), interrupted by regularly-spaced intonation-based distractor sections. The task asked participants to listen to five stimuli per page, categorizing them into one of three word groups, and describing their certainty. While students were requested to listen to the audio file only once, this is a possible confound that could not be controlled for with the survey software, as some listeners may have chosen to listen to stimuli more than once. The stimuli were divided into 24 pages (i.e., five audios per page), with stimulus order randomized per page by Qualtrics to avoid ordering-based effects. Pages contained two stimuli from a male actor, two from the female actor, and one randomly-chosen distractor. At the end of the survey, a brief qualitative classification task asked participants to list cues they listened for when identifying stimuli.

3.4 Data analysis

Once the distractors were removed, the 140 participants made 11,766 identifications. The dependent variable reflected their responses: "voiced" or "voiceless." Five independent variables were analyzed.

First, "proficiency level" was determined based on participants' institutional classification (i.e., course level), divided into semesters four, six and eight of undergraduate studies, and non-native and native graduate speakers of Spanish. This variable quantified participant proficiency and experience with Spanish, given that lower-level participants were expected to perceive VOT with more L1-English-like boundaries, as various previous researchers have described their perception as closer to naïve non-native listeners (Flege, 2005; García-Sierra et al., 2009). Second, the "initial phone" (i.e., phone prior to manipulation) was analyzed to ascertain whether additional correlates continued to play a role in classification despite manipulations, and whether effects arose based on the initial place of articulation and voicing, given that some studies have identified non-VOT correlates that affect identification (Benkí, 2005; Bohn & Flege, 1993; Zampini, 1998). Previously voiced sounds, regardless of manipulation, were expected to be more likely classified as voiced. Third, "voicing type" (i.e., Figures 4–7) tracked the acoustic boundary between speakers' voiced and voiceless stops. English learners' perceptual boundary was expected to be at a higher VOT than Spanish natives (Abramson & Lisker, 1964). Spanish perceptual norms dictated that a stop with zero VOT should be closer to a "good" representative of a voiceless rather than voiced sound, unlike English norms, which should classify it as voiceless.

Two additional factors were drawn from the background questionnaire in Qualtrics. "Time abroad" reflected arguments in favor of perceptual development due to time spent in the country of the target language, showing that learners are influenced by dialectal exposure when producing phonetic differences (e.g., Schmidt, 2018; Solon & Long, 2019). This variable was divided into four sub-sections: never abroad, abroad less than three months (capturing summer programs), abroad more than three months, and native graduate students. Those with more international experience were expected to follow a similar trend to higher proficiency students, being influenced by their exposure to perceive more Spanish-like boundaries. Finally, "survey duration" in minutes served as a metric Qualtrics provided that passingly resembled reaction time: slower response times across the entire questionnaire were expected to indicate greater cognitive load, thereby showing higher L1 effects (Dupoux et al., 2008).

The response "voiced" was set as the reference value in the mixed-effects logistic regression performed in the R-based Rbrul program (Johnson, 2008). "Participant" and "Word" were both set as random effects. Finally, based on Table 3, interactions were set up between each of the first independent variables.

Table 3. List of variable levels

Variable	**Coding**		
Dependent variable			
Response	Voiced	Voiceless	
Independent variables			
1. Level	Semester 4	Semester 6	Semester 8
	Non-native Grad	Native Grad	
2. Initial phone	/p/ /b/	/t/ /d/	/k/ /g/
3. Voicing type	Full Prevoicing	Half Prevoicing	Zero Prevoicing
	Full VOT		
4. Time abroad	No time abroad	0–3 months	3+ Months
5. Survey duration	Continuous (minutes)		

4. Results

The inclusion of distractor words served both to prevent participants from guessing the goal of the study as well as to ensure they were focused on the task. Before removing the distractor, the data were checked to ensure that overall misclassifications were low (Table 4).

With the distractors removed, a mixed-effect logistic regression yielded three significant main effects and three significant interactions – Voicing Type, Initial Phone, Proficiency Level, Voicing Type:Proficiency Level, Initial Phone:Voicing Type, and Initial Phone:Proficiency Level (Table 5a, see also Appendix 2 for a more detailed table). Due to the high correlation between participant level and time abroad (high proficiency students had spent more studying in countries where the target language was spoken), the main effect and interactions for Time Abroad were not included in the final model. Duration of the study was also not found to be significant.

Table 4. Overall classification of stimuli as voiced or voiceless stop or /r/

Initial phone	**/r/**		***Response* voiced stop**		**Voiceless stop**		**Total**
	n	**%**	**n**	**%**	**n**	**%**	**n**
/b/	60	2.90%	1507	73.60%	481	23.50%	2048
/d/	66	3.20%	1466	71.60%	516	25.20%	2048
/g/	69	3.40%	1519	74.20%	460	22.50%	2048
/p/	65	3.20%	1198	58.50%	785	38.30%	2048
/t/	63	3.10%	1148	56.10%	837	40.90%	2048
/k/	70	3.40%	1291	63.00%	687	33.50%	2048
/r/	2854	92.90%	122	4.00%	96	3.10%	3072
Total	**3247**	**21.10%**	**8251**	**53.70%**	**3862**	**25.10%**	**15360**

Table 5. Mixed-Effects logistic regression with "voiced" as the reference value

Factor	Factor weight
Voicing Type ($p < 0.001$)	
Range	*74.9*
Voicing Type*Proficiency Level ($p < 0.001$)	
Range	*48.2*
Initial Phone ($p < 0.001$)	
Range	*37.4*
Initial Phone*Voicing Type ($p < 0.001$)	
Range	*35.9*
Initial Phone*Proficiency Level ($p < 0.017$)	
Range	*35.9*
Proficiency Level ($p < 0.0091$)	
Range	*13.1*
Random Intercept: Respondent	
Range	*70.4*
Random Intercept: Word	
Range	*12.7*

n = 11766 df = 62 Log Likelihood=−5367.5 AIC = 10859 R2 Fixed = 0.363 R2 Random = 0.08 R2 Total = 0.443

The variable rule analysis conducted in Rbrul provides four types of data to situate results statistically: p-values, log-odds factor weights, and variable hierarchy. When the p-value is less than 0.05, a factor is determined to be significant. Positive log-odds show that a factor favored identification as voiced. Factor weights go from zero to one, with values above 0.5 indicating that voicing was favored, and values below 0.5 indicating that it was disfavored. Finally, the fourth value, the variable hierarchy, ranks predictors in terms of most to least descriptive of the variation in the model. All but log-odds are shown in the condensed Table 5a.

First, with a p-value < 0.001 and a range (difference across factor weights) of 74.9, the most predictive independent variable selected by Rbrul was Voicing Type (Figure 9). The longer the prevoicing, the higher the chance that participants would identify the sound as voiced. Rather than a sharp slope, classifications undergo a gradual curve, as both Full (89.8%) and Half Prevoicing (85.7%) have factor weights above 0.5, while No Prevoicing was categorized slightly above chance (63.5%) with a factor weight indicating that it favored "voiceless" identification, and Full Short Lag VOT was least frequently identified as "voiceless" (34.6%).

The second most predictive independent variable in the logistic hierarchy was the interaction between Voicing Type and Proficiency Level, with a p-value < 0.001 and a range of 48.2 (Figure 10). In the model, Native Speakers and Non-Native Grads were more likely to identify Full (97.6% and 95% respectively) and Half

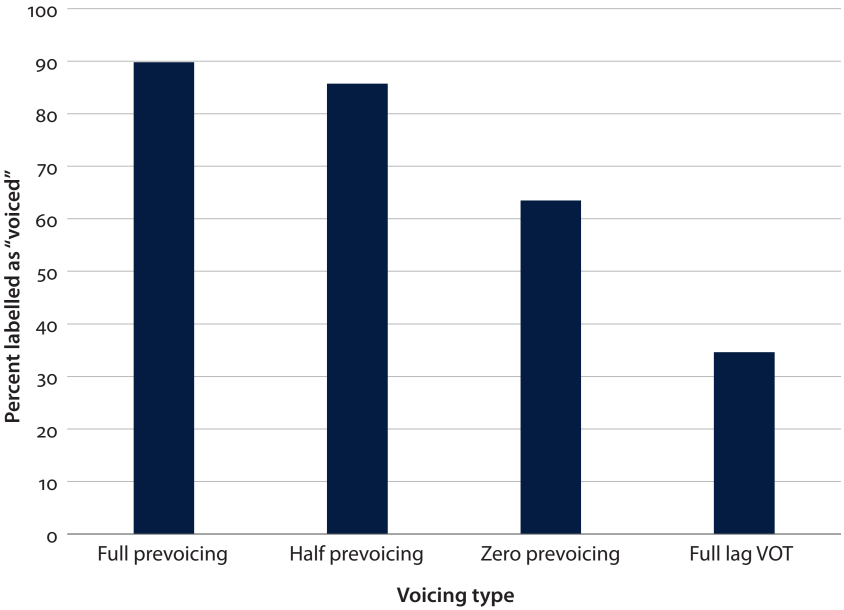


Figure 9. Stimuli identified as "voiced" by Voicing type

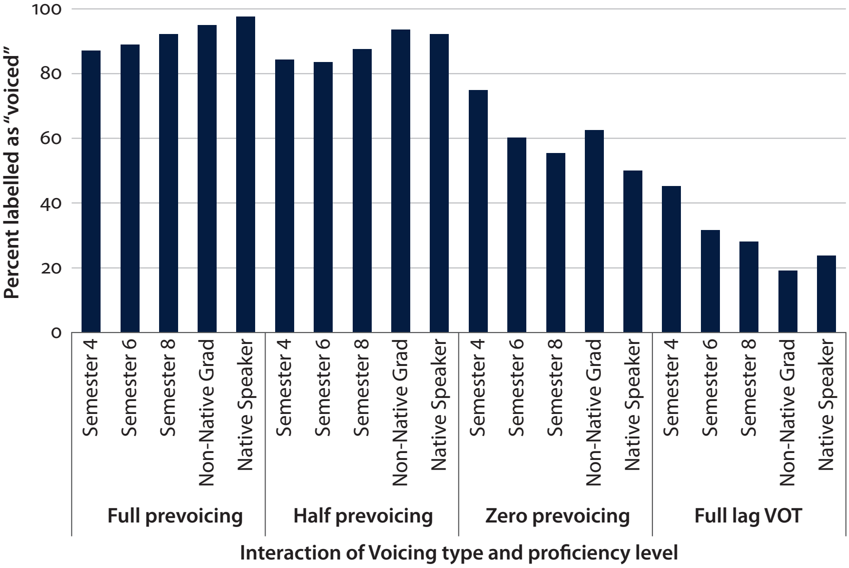


Figure 10. Stimuli identified as "voiced" by Voicing type and proficiency level

prevoicing (92.2% and 93.5%) as "voiced," while students in Semester 2 identified Zero Prevoicing (74.8%) and Full Short Lag VOT (45.3%) as "voiced" most, with the undergraduate groups performing similarly to the native speakers, and graduate groups fluctuating.

Third in the prediction hierarchy was "initial phone," with p < 0.001 and a range of 37.4 (Figure 11). The main division here was between originally voiced and voiceless stops, prior to manipulation. Despite the alteration of VOT and the normalization of following vowel duration, the initially voiced stops were labelled as "voiced" more (/g/ = 76.8% /b/ = 75.7% /d/ = 74.5%) than the voiceless ones (/k/ = 65.3% /p/ = 60.3% /t/ = 58.3%). This is reflected in the factor weight, as the first three favor a "voiced" identification while the voiceless ones disfavor it. There is also a smaller trend based on place of articulation: for both originally voiced and voiceless stops, the velar sound has the highest "voiced" identification, and the alveolar one has the lowest.

The fourth most predictive independent variable was the interaction between Initial Phone and Voicing Type, with p < 0.001 and a range of 35.9 (Figure 13). Originally voiceless stops were least often identified as "voiced" for Full, Half and Zero Prevoicing, while /g/ was most often perceived as "voiced" (Full = 92.6% Half = 92.4% Zero = 83.1%). However, for Full Short Lag VOT, this trend changes, as the alveolar stops are more often classified as "voiced" in both the originally

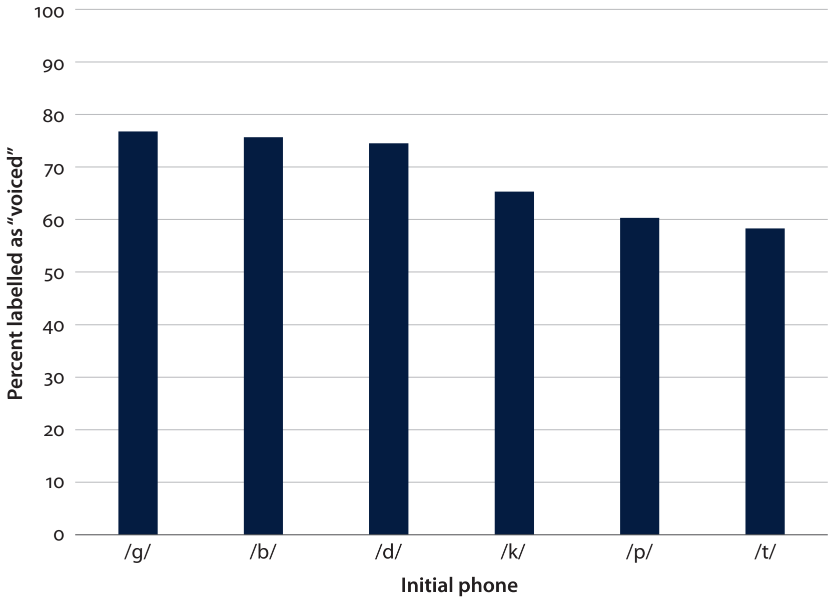


Figure 11. Stimuli identified as "voiced" by initial phone

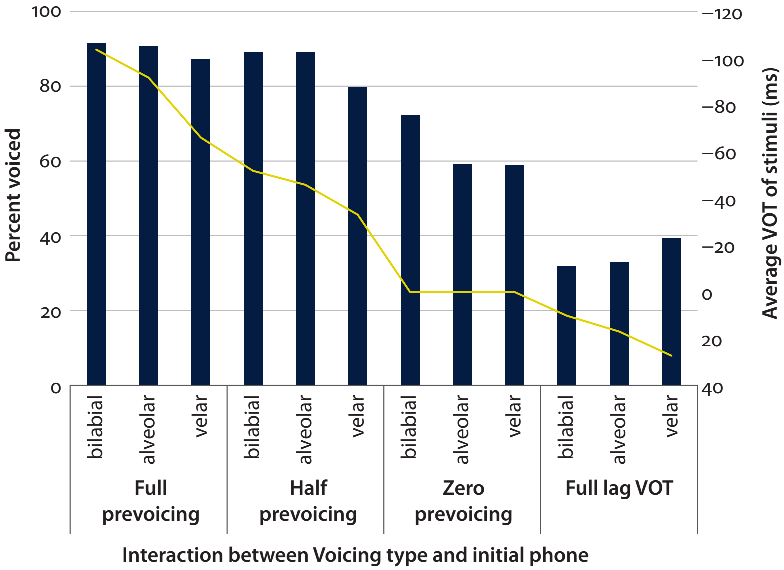


Figure 12. The bars show participant identification of sounds as voiced, while the line gives the average VOT for the stimuli

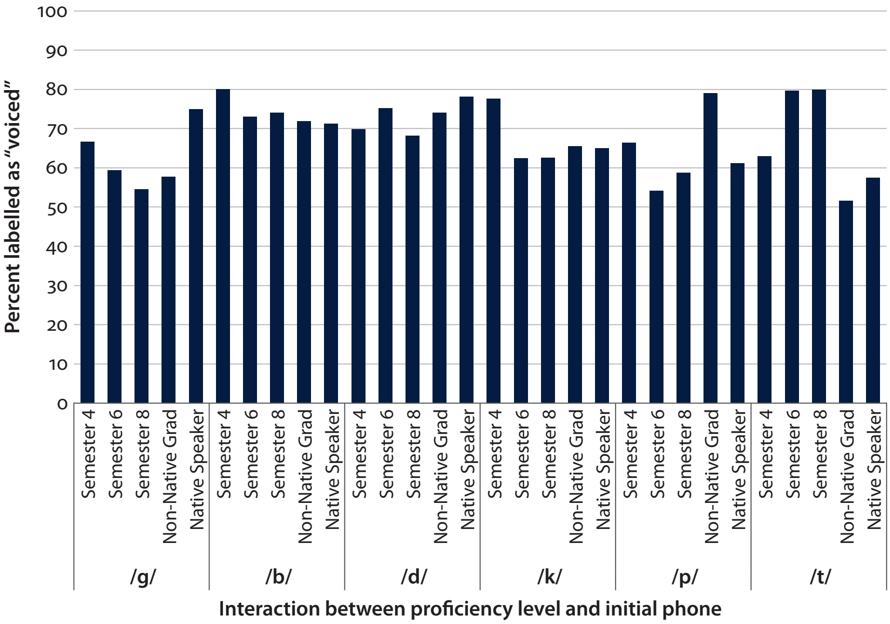


Figure 13. Stimuli identified as "voiced" by proficiency level and initial phone

voiced (/d/ = 52.3%) and voiceless (/t/ = 26.4%) groups. Participant categorization of the initial phones have been charted with bar graphs alongside a line graph averaging the VOT of the stimuli to depict the relation between actual voicing and identification as "voiced." The level of prevoicing does not appear to have a strong effect on identification as "voiced" until the VOT approaches -40ms (40ms prevoicing), with the actual shift being closer to 10ms.

The fifth factor in the hierarchy was the interaction between initial phone and proficiency level, with p < 0.017 and a range of 35.9 (Figure 13). In addition to an interaction between Voicing type, the proficiency level of participants affected the degree to which they identified sounds as voiced based on their original voicing. For some of the sounds like /g/, there is a u-shaped curve, with natives identifying the sound most-frequently as voiced, whereas in other cases, as with /k/, the lowest-proficiency group was highest, while the other groups were similar. Still another pattern, shown in /p/ and /t/, is that students with higher proficiency behave unexpectedly (Non-Native Grad for /p/ = 79%, Semester 8 for /t/ = 79.9%), while natives are more consistent (/p/ = 61.2%, /t/ = 57.5%).

The final factor in the hierarchy was Proficiency Level, with p < 0.0091 and a range of 13.1 (Figure 14). The division between the groups, holistically, is relatively minimal, with the most noticeable division between the lowest proficiency participants (Semester 4 = 72.9%) and both native speaker (66.1%) and higher-level non-native participants (Semester 6 = 66.1%, Semester 8 = 65.9%, Non-Native Grads = 67.8%) participants.

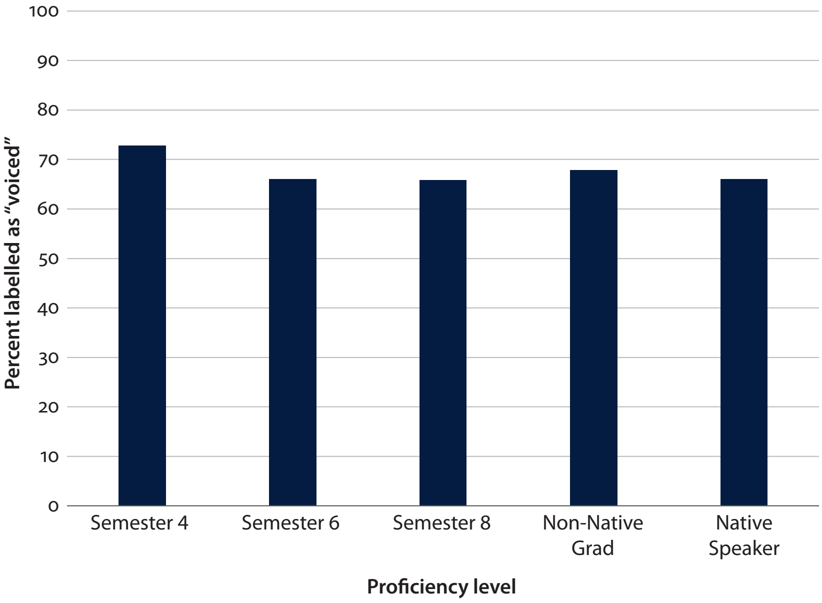


Figure 14. Stimuli identified as "voiced" by proficiency level

4.1 Qualitative description of Spanish stops

At the end of the survey, participants were asked to describe how they differentiated ten manipulated audio files from the classification task meta-linguistically. In Figures 15 through 19, word-clouds demonstrate the most-frequent terms used by each proficiency group, where size correlates with frequency (e.g., "sound" was used quite frequently).

Fourth semester participants talked about "rolled" /r/, and adjectives like "hard" and "soft" to distinguish between voiced and voiceless sounds – salient elements and general descriptors. Sixth semester participants talked about the flow of "air" and mentioned articulators (e.g., "mouth" and "tongue"). In the eighth semester, participants used linguistic explanations, sometimes abandoning "soft" and "hard" in favor of "vibration" and "tongue." The non-native graduate students (mainly students of linguistics) wrote more specifically about "voiceless" and "voiced" sounds, "stops," "trills" and voicing "duration." Finally, natives from various backgrounds were less technical in their descriptions, mentioning "flaps," "contexts," and adjectives that implied the ease of categorization (e.g., "easy," "clearly," "sure").

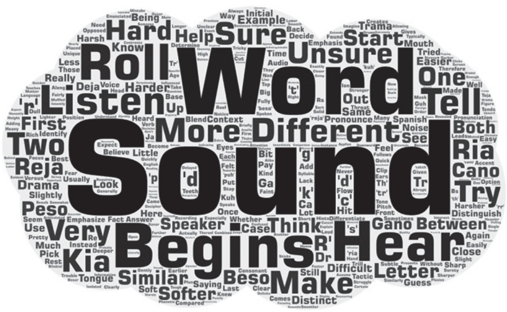


Figure 15. Semester 4 students' responses

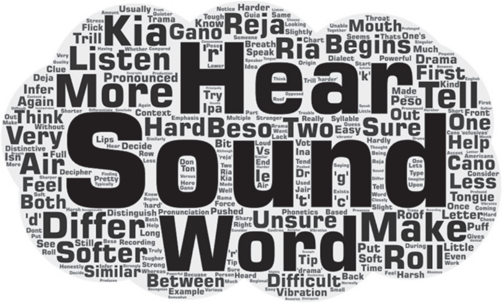


Figure 16. Semester 6 students' responses

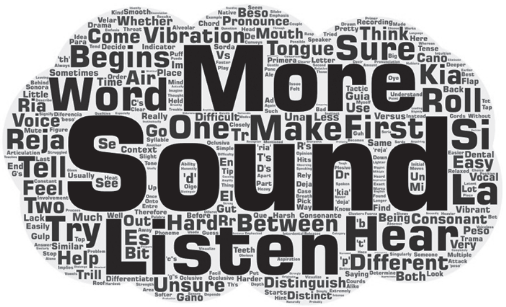


Figure 17. Semester 8 students' responses

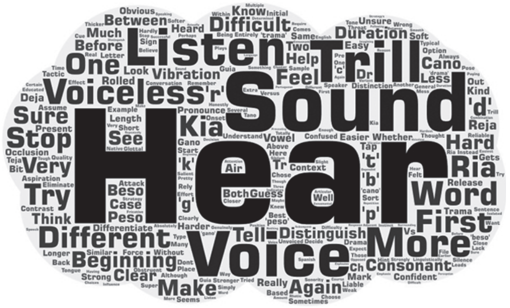


Figure 18. Non-native graduates responses

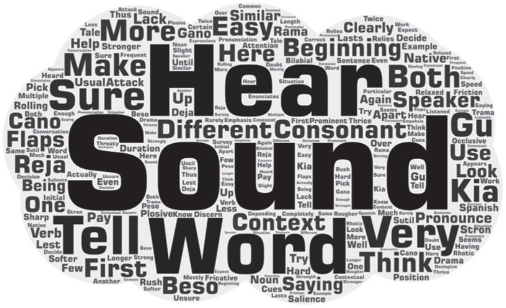


Figure 19. Native graduate responses

5. Discussion

5.1 RQ1: To what extent do L1 English speakers follow native-like perception norms when identifying Spanish stops, and how does this reflect the two types of perceptual classifications described by PAM L2 or SLM?

The factors selected as significant in the mixed-effect logistic regression indicate that proficiency level interacted with Voicing type and the initial phone, as well as being a significant main effect in the data, meaning that participants' classificatory patterns did differ based on their language proficiency. Fourth semester students behaved most differently from the other participants, while the rest of the students, both graduate and undergraduate, moved in the direction of, but did not quite reach, native-like identification patterns. Fourth Semester participants were most likely to under-classify Full and Half Prevoicing tokens and over-classify Zero Prevoicing and Full Short Lag VOT tokens as "voiced," compared to natives.

This pattern reflects Casillas and Simonet's (2018) and Mack's (1989) visualization of bilingual perceptual spaces. Both studies show that bilinguals perceive stops similarly to monolinguals when they enter a given language "mode" or hear a certain perceptual cue. Higher-proficiency participants followed this behavior, while those in Semester Four acted more like English monolinguals. Flege (1995) and Birdsong (2018) have indeed argued that language learners at low proficiencies behave more like monolinguals than bilinguals, which explains this trend. As participants were placed into only a single language mode (i.e., Spanish), the data are not sufficient to reject earlier claims (e.g., Caramazza et al., 1973) that the bilingual perceptual space is a fixed realm existing between monolingual norms for speakers' languages. However, as participants at higher levels showed an increased ability to distinguish voiced from voiceless stops following boundaries that differed from those employed in their L1, they likely can switch between English and Spanish norms, although further research would need to confirm this assertion.

It was predicted that, if Spanish and English stops were close enough to allow Equivalence Classification (Flege, 1995) and fit the Two Category Assimilation of PAM L2 (Best & Tyler, 2007), learners would group Spanish stops into pre-existing English voiced and voiceless categories with increasing success based on their proficiency level. This seemed to be the case. The alternative, that there would be increased confusion due to an inability to successfully categorize the Spanish voiced stop with lead VOT into the English "voiced" category with short lag VOT, was not observed. Instead, the abrupt shift in classification patterns between the fourth and sixth semester, and the slow trend of sixth, eighth and non-native graduate participants in the direction of (and in certain cases over-shooting) native-like perceptual norms indicates that L1 speakers can develop acoustic boundaries that approximate L2 norms, albeit with aspects of their L1 perceptual system continuing to play a role. As Birdsong (2018) argues, it is overly simplistic to expect learners to become native-like, given the perpetual influence of their L1 systems.

Learners' phonological boundaries can and do shift when distinguishing between voiced and voiceless stops in their L2, an ability that develops with increased exposure to the L2 system. Because they have to recognize voicing differences so as to distinguish between minimal pairs, nuanced perception must be developed early in acquisition. Following predictions from the phonological acquisition models, learners gain awareness of differences in L2 categories at an early proficiency level, with increased exposure to Spanish allowing the acceptable VOT boundaries to become more flexible depending on their language mode.

5.2 RQ2: What social and linguistic constraints govern listeners' perceptions of stop constraints, and how important was VOT specifically to their classification?

The initial phone was selected as significant both in main effect and in two interactions. Across these factors, it was evident that identification of stimuli differed based on the voicing of the original stop. The results showed variable tendencies for stops to be identified as "voiced" based on both the type of VOT and the proficiency of the participant. Stops with Zero Prevoicing and Full Short Lag VOT behaved variably, whereas those with Full and Half Prevoicing received more consistent identifications. When participants were uncertain of how to interpret voicing cues, other factors seemed to come into play. Similarly, identifications based on language level showed that native speakers were the only group to consistently classify the originally voiced stops /b/, /d/ and /g/ as "voiced" more frequently than the voiceless ones /p/, /t/ and /k/. This suggests the presence of Spanish voicing cues, which L1 English speakers were less likely to perceive. Fourth Semester participants in particular seemed to be relying on a different (perhaps L1) set of acoustic cues than the other participants, or to be identifying voicing at a rate approaching random chance. This reinforces arguments made by Flege (2005) and García-Sierra et al. (2009), who found that increased experience and competence in an L2 correlates to the success with which L2 categories are assimilated (or not).

Although individual differences, measured by time spent studying abroad and cognitive load (measured by duration of the survey), did not significantly affect perception, factors other than VOT did indeed play a role. The fact that originally voiced stops were categorized as "voiced" nearly ten percent more than their voiceless counterparts follows Bohn and Flege (1993), showing that there are difficult-to-identify cues that must lay behind the systematic decisions of listeners.

This consistent variability indicates that acoustic differences other than VOT were used by participants in their classifications, which merits future investigation to determine what factors specifically affect identification (e.g., Benkí, 2005; Llanos et al., 2013; Simonet et al., 2014).

6. Conclusion

This study focused on the ability of college-level L1 English speakers, learning Spanish as an L2, to perceive prevoicing and VOT in voiced and voiceless stops. Learners at intermediate language proficiencies showed perceptual boundaries that resembled native speakers, allowing them to distinguish voicing in Spanish stops. Speakers at higher proficiencies showed more-refined categories that resulted from further language learning, and a steady reduction of L1 interference, even though they never completely replicated native tendencies. The voicing distinction in Spanish for L1 English learners, under the PAM L2 or SLM models, seems to be most similar to Two Category Assimilation, allowing Equivalence Classification and encouraging acquisition.

While VOT is indeed one acoustic cue that speakers of both Spanish and English use to distinguish between voiced and voiceless stops, additional acoustic correlates affect identification. This study could not make as a strong a rejection of VOT as Bohn and Flege (1993), who argued that VOT is not the "dominant" correlate in stop identification. However, these results strongly indicate that there are (unidentified) correlates that most significantly affect the classification of stops, which could include variation in F1, F0, or even word frequency (Benkí, 2005; Connine et al., 1993; Llanos et al., 2013).[6]

References

Abramson, A. S., & Lisker, L. (1972). Voice-timing perception in Spanish word-initial stops. *Journal of Phonetics*, 1, 1–8. https://doi.org/10.1016/S0095-4470(19)31372-5

Amengual, M. (2012). Interlingual influence in bilingual speech: Cognate status effect in a continuum of bilingualism. *Bilingualism: Language and Cognition*, 15(3), 517–530. https://doi.org/10.1017/S1366728911000460

Antoniou, M., Tyler, M. D., & Best, C. T. (2012). Two ways to listen: Do L2-dominant bilinguals perceive stop voicing according to language mode? *Journal of Phonetics*, 40(4), 582–594. https://doi.org/10.1016/j.wocn.2012.05.005

[6] I would like to thank the editors of this volume, my two anonymous reviewers, and Erik Willis and Kimberly Geeslin for their comments and advice. All mistakes that remain are, of course, my own.

Benkí, J. R. (2005). Perception of VOT and first formant onset by Spanish and English speakers. In J. Cohen, K. T. McAlister, K. Rolstad, & J. MacSwan (Eds.), *Proceedings of the 4th International Symposium on Bilingualism* (pp. 240–248). Somerville, MA: Cascadilla Press.

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O. Bohn & M. J. Munro (Eds.), *Language experience in second language speech learning: In honor of James Emil Flege* (pp. 13–34). Amsterdam: John Benjamins. https://doi.org/10.1075/lllt.17.07bes

Birdsong, D. (2018). Plasticity, variability and age in second language acquisition and bilingualism. *Frontiers in Psychology*, 9(81), 1–17.

Boersma, P., & Weenink, D. (2017). *Praat: Doing phonetics by computer* [Computer program] (Version 6.0.35). Retrieved from <http://www.praat.org/> (16 March, 2020).

Bohn, O., & Flege, J. E. (1993). Perceptual switching in Spanish/English bilinguals. *Journal of Phonetics*, 21, 267–290. https://doi.org/10.1016/S0095-4470(19)31339-7

Bullock, B. E., & Toribio, A. J. (2009). Trying to hit a moving target: On the sociophonetics of code-switching. In L. Isurin, D. Winford, & K. de Bot (Eds.), *Multidisciplinary approaches to code switching* (pp. 189–206). Amsterdam: John Benjamins. https://doi.org/10.1075/sibil.41.12bul

Caramazza, A., Yeni-Komshian, G. H., Zurif, E. B., & Carbone, E. (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, 421–428. https://doi.org/10.1121/1.1913594

Casillas, J. V., & Simonet, M. (2018). Perceptual categorization and bilingual language modes: Assessing the double phonemic boundary in early and late bilinguals. *Journal of Phonetics*, 71, 51–64. https://doi.org/10.1016/j.wocn.2018.07.002

Cho, T., & Ladefoged, P. (1999). Variation and universals in VOT: Evidence from 18 languages. *Journal of Phonetics*, 27, 207–229. https://doi.org/10.1006/jpho.1999.0094

Connine, C. M., Titone, D., & Wang, J. (1993). Auditory word recognition: Extrinsic and intrinsic effects of word frequency. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 19(1), 81–94.

Dupoux, E., Sebastian-Galles, N., Navarrete, E., & Peperkamp, S. (2008). Persistent stress 'deafness': The case of French learners of Spanish. *Cognition*, 106(2), 682–706. https://doi.org/10.1016/j.cognition.2007.04.001

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* (pp. 233–277). Baltimore, MD: York Press.

Flege, J. E. (2005). Origins and development of the Speech Learning Model. *1st ASA Workshop on L2 Speech Learning*. Keynote Lecture. Vancouver, BC: Simon Fraser University.

Flege, J., & Eefting, W. (1986). Linguistic and developmental effects on the production and perception of stop consonants. *Phonetica*, 43, 155–171. https://doi.org/10.1159/000261768

Flege, J. E., & Eefting, W. (1987). Production and perception of English stops by native Spanish speakers. *Journal of Phonetics*, 15, 67–83. https://doi.org/10.1016/S0095-4470(19)30538-8

García-Sierra, A., Diehl, R. L., & Champlin, C. (2009). Testing the double phonemic boundary in bilinguals. *Speech Communication*, 51(4), 369–378. https://doi.org/10.1016/j.specom.2008.11.005

George, A. (2014). Study abroad in central Spain: The development of regional phonological features. *Foreign Language Annals*, 47, 97–114. https://doi.org/10.1111/flan.12065

Hunnicutt, L., & Morris, P. A. (2016). Prevoicing and aspiration in Southern American English. *University of Pennsylvania Working Papers in Linguistics*, 22(1). Retrieved from <http://repository.upenn.edu/pwpl/vol22/iss1/24> (16 March, 2020).

Johnson, D. E. (2008). Getting off the GoldVarb standard: Introducing Rbrul for mixed-effects variable rule analysis. *Language and Linguistics Compass*, 3(1), 359–383. https://doi.org/10.1111/j.1749-818X.2008.00108.x

Knouse, S. M. (2012). The acquisition of dialectal phonemes in a study abroad context: The case of the Castilian theta. *Foreign Language Annals*, 45(4), 512–542. https://doi.org/10.1111/j.1944-9720.2013.12003.x

Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, 20(3), 384–422. https://doi.org/10.1080/00437956.1964.11659830

Llanos, F., Dmitrieva, O., Shultz, A., & Francis, A. L. (2013). Auditory enhancement and second language experience in Spanish and English weighting of secondary voicing cues. *Journal of the Acoustical Society of America*, 134(3), 2213–2224. https://doi.org/10.1121/1.4817845

Llanos, F., & Francis, A. L. (2016). The effects of language experience and speech context on the phonetic accommodation of English-accented Spanish voicing. *Language and Speech*, 59(1), 1–24.

Mack, M. (1989). Consonant and vowel perception and production: Early English-French bilinguals and English monolinguals. *Perception & Psychophysics*, 46(2), 187–200. https://doi.org/10.3758/BF03204982

Martínez Celdrán, E. (1993). La percepción categorial de /b–p/ en español basada en las diferencias de duración. *Estudios de Fonética Experimental*, 5, 224–239.

Melero-García, F., & Cisneros, A. R. (2018). *No es tan simple como parece: The duration of one-closure rhotics on the perception of Spanish /ɾ/ and /r/* (Unpublished manuscript).

Newlin-Lukowicz, L. (2014). From interference to transfer in language contact: Variation in voice onset time. *Language Variation and Change*, 26, 359–386. https://doi.org/10.1017/S0954394514000167

Olmstead, A. J., Viswanathan, N., Aivar, M. P., & Manuel, S. (2013). Comparison of native and non-native phone imitation by English and Spanish speakers. *Frontiers in Psychology*, 4, 1–7. https://doi.org/10.3389/fpsyg.2013.00475

Olson, D. (2013). Bilingual language switching and selection at the phonetic level: Asymmetrical transfer in VOT production. *Journal of Phonetics*, 41, 407–420. https://doi.org/10.1016/j.wocn.2013.07.005

Pallier, C., Christophe, A., & Mehler, J. (1997). Language-specific listening. *Trends in Cognitive Sciences*, 1(4), 129–132. https://doi.org/10.1016/S1364-6613(97)01044-9

Ringer-Hilfinger, K. (2012). Learner acquisition of dialect variation in a study abroad context: The case of the Spanish [θ]. *Foreign Language Annals*, 45, 430–446. https://doi.org/10.1111/j.1944-9720.2012.01201.x

Schmidt, L. B. (2018). L2 Development of perceptual categorization of dialectal sounds: A study in Spanish. *Studies in Second Language Acquisition*, 40(4), 857–882. https://doi.org/10.1017/S0272263118000116

Schoonmaker-Gates, E. (2012). Foreign accent perception in L2 Spanish: The role of proficiency and L2 experience. In J. Levis & K. LeVelle (Eds.), *Proceedings of the 3rd Pronunciation in Second Language Learning and Teaching Conference, Sept. 2011* (pp. 84–92). Ames, IA: Iowa State University.

Schoonmaker-Gates, E. (2013). The interplay between native Spanish dialect exposure and foreign accent perception. In A. M. Carvalho & S. Beaudrie (Eds.), *Selected proceedings of the 6th Workshop on Spanish Sociolinguistics* (pp. 169–176). Somerville, MA: Cascadilla Proceedings Project.

Schoonmaker-Gates, E. (2017). Regional variation in the language classroom and beyond: Mapping learners' developing dialectal competence. *Foreign Language Annals*, 50(1), 177–194. https://doi.org/10.1111/flan.12243

Simonet, M. (2012). The L2 acquisition of Spanish phonetics and phonology. In J. I. Hualde, A. Olarrea, & E. O'Rourke (Eds.), *The handbook of Hispanic linguistics* (pp. 729–746). Hoboken, NJ: Blackwell. https://doi.org/10.1002/9781118228098.ch34

Simonet, M., Casillas, J. V., & Díaz, Y. (2014). The effects of stress/accent on VOT depend on language (English, Spanish), consonant (/d/, /t/) and linguistic experience (monolinguals, bilinguals). In N. Campbell, D. Gibbon, & D. Hirst (Eds.), *7th International Conference on Speech Prosody* (pp. 202–206). Retrieved from <http://fastnet.netsoc.ie/sp7/> (16 March, 2020).

Solon, M., & Long, A. Y. (2019). Acquisition of phonetics and phonology abroad: What we know and how. In C. Sanz & A. Morales-Front (Eds.), *The Routledge handbook of study abroad research and practice* (pp. 126–145). Abingdon: Routledge.

Stewart, J. (2018). Voice onset time production in Ecuadorian Spanish, Quichua, and Media Lengua. *Journal of the International Phonetic Association*, 48(2), 173–197. https://doi.org/10.1017/S002510031700024X

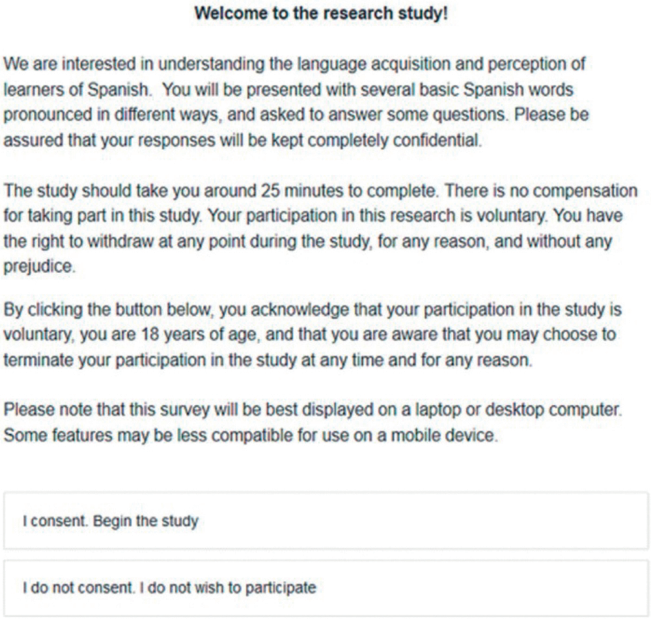
Strange, W., & Shafer, V. L. (2008). Speech perception in second language learners: The re-education of selective perception. In J. G. Hansen Edwards & M. L. Zampini (Eds.), *Phonology and second language acquisition* (pp. 153–191). Amsterdam: John Benjamins. https://doi.org/10.1075/sibil.36.09str

Thomas, E. (2011). *Sociophonetics: An introduction*. London: Palgrave Macmillan. https://doi.org/10.1007/978-1-137-28561-4

Williams, L. (1977). The perception of stop consonant voicing by Spanish-English bilinguals. *Perception & Psychophysics*, 21(4), 289–297. https://doi.org/10.3758/BF03199477

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 1. Consent and demographic survey questions



Welcome to the research study!

We are interested in understanding the language acquisition and perception of learners of Spanish. You will be presented with several basic Spanish words pronounced in different ways, and asked to answer some questions. Please be assured that your responses will be kept completely confidential.

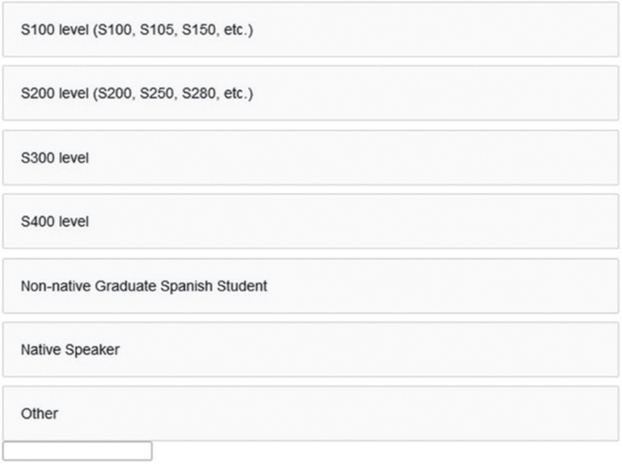
The study should take you around 25 minutes to complete. There is no compensation for taking part in this study. Your participation in this research is voluntary. You have the right to withdraw at any point during the study, for any reason, and without any prejudice.

By clicking the button below, you acknowledge that your participation in the study is voluntary, you are 18 years of age, and that you are aware that you may choose to terminate your participation in the study at any time and for any reason.

Please note that this survey will be best displayed on a laptop or desktop computer. Some features may be less compatible for use on a mobile device.

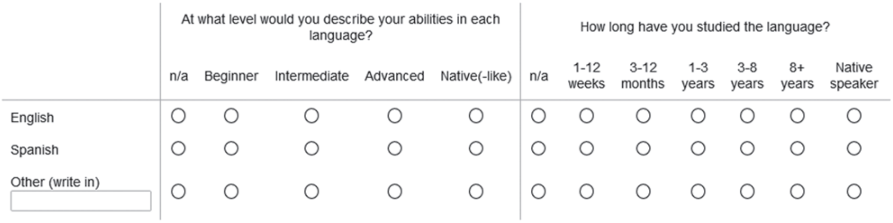
I consent. Begin the study
I do not consent. I do not wish to participate

Spanish level:



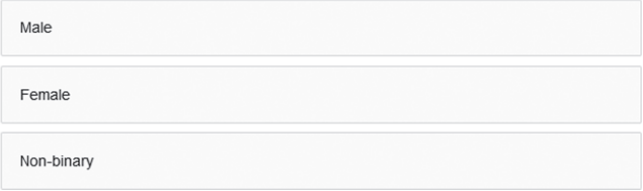
S100 level (S100, S105, S150, etc.)
S200 level (S200, S250, S280, etc.)
S300 level
S400 level
Non-native Graduate Spanish Student
Native Speaker
Other

What are your strongest languages? How long have you studied them?



	At what level would you describe your abilities in each language?					How long have you studied the language?						
	n/a	Beginner	Intermediate	Advanced	Native(-like)	n/a	1-12 weeks	3-12 months	1-3 years	3-8 years	8+ years	Native speaker
English	○	○	○	○	○	○	○	○	○	○	○	○
Spanish	○	○	○	○	○	○	○	○	○	○	○	○
Other (write in)	○	○	○	○	○	○	○	○	○	○	○	○

The gender I identify as:

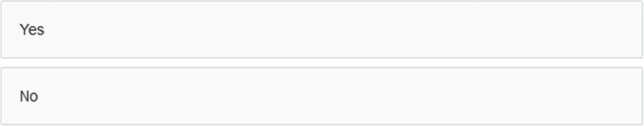


Male

Female

Non-binary

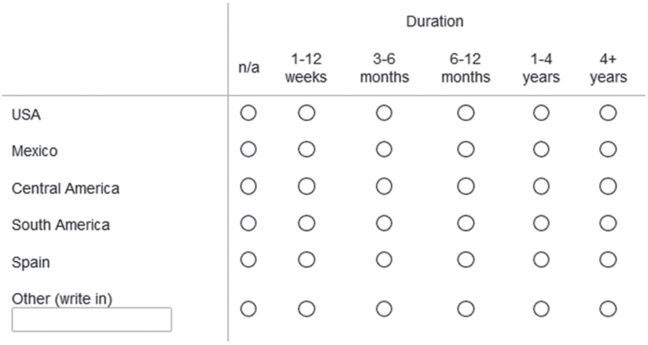
I've studied or lived outside of the U.S.



Yes

No

Which regions and countries have you studied or lived in? For how long?



	Duration					
	n/a	1-12 weeks	3-6 months	6-12 months	1-4 years	4+ years
USA	○	○	○	○	○	○
Mexico	○	○	○	○	○	○
Central America	○	○	○	○	○	○
South America	○	○	○	○	○	○
Spain	○	○	○	○	○	○
Other (write in)	○	○	○	○	○	○

Appendix 2. Mixed-effects logistic regression results of factors

Table 1. Mixed-Effects logistic regression with "voiced" as application value, "Participant" and "Word" as random effects, and Voicing Type, Level and Initial Phone as main effects, with three 2-way interactions between them

Factor	Log-odds	Tokens	Percent voiced	Factor weight
Voicing Type ($p < 0.001$)				
Full Prevoicing	1.683	3068	89.80%	0.843
Half Prevoicing	1.251	2828	85.70%	0.777
Zero Prevoicing	−0.665	2937	63.50%	0.34
Full Short Lag VOT	−2.269	2933	34.60%	0.094
Full Short Lag VOT				74.9
Interaction: Voicing Type*Proficiency Level ($p < 0.001$)				
Half Prevoicing:Native Speaker	1.098	115	92.20%	0.75
Full Short Lag VOT:Semester 8	0.929	883	28.10%	0.717
Full Prevoicing:Semester 6	0.71	855	89.00%	0.67
Half Prevoicing:Semester 6	0.644	783	83.50%	0.656
Full Short Lag VOT:Semester 4	0.373	1019	45.30%	0.592
Full Prevoicing:Non-Native Grad	0.327	100	95.00%	0.581
Zero Prevoicing:Native Speaker	0.25	120	50.00%	0.562
Zero Prevoicing:Semester 4	0.217	1017	74.80%	0.554
Full Short Lag VOT:Non-Native Grad	0.107	94	19.10%	0.527
Zero Prevoicing:Semester 8	−0.013	891	55.40%	0.497
Full Prevoicing:Semester 4	−0.026	1061	87.00%	0.493
Full Prevoicing:Semester 8	−0.067	927	92.20%	0.483
Zero Prevoicing:Non-Native Grad	−0.105	96	62.50%	0.474
Half Prevoicing:Non-Native Grad	−0.329	92	93.50%	0.418
Zero Prevoicing:Semester 6	−0.349	813	60.10%	0.414
Full Short Lag VOT:Native Speaker	−0.403	118	23.70%	0.4
Half Prevoicing:Semester 4	−0.564	984	84.30%	0.363
Half Prevoicing:Semester 8	−0.849	854	87.50%	0.3
Full Prevoicing:Native Speaker	−0.944	125	97.60%	0.28
Full Short Lag VOT:Semester 6	−1.005	819	31.60%	0.268
Range				48.2
Initial Phone ($p < 0.001$)				
/g/	0.731	1963	76.80%	0.675
/b/	0.473	1972	75.70%	0.616
/d/	0.314	1950	74.50%	0.578
/k/	−0.186	1961	65.30%	0.454
/p/	−0.489	1966	60.30%	0.38
/t/	−0.842	1954	58.30%	0.301
Range				37.4

Factor	Log-odds	Tokens	Percent voiced	Factor weight
Interaction: Initial Phone*Voicing Type ($p < 0.001$)				
/k/:Full Short Lag VOT	0.762	491	24.40%	0.682
/g/:Zero Prevoicing	0.489	492	83.10%	0.62
/t/:Half Prevoicing	0.448	490	70.40%	0.61
/d/:Zero Prevoicing	0.324	491	65.20%	0.58
/p/:Full Prevoicing	0.305	494	88.50%	0.576
/t/:Full Short Lag VOT	0.277	489	26.40%	0.569
/b/:Half Prevoicing	0.225	494	93.10%	0.556
/p/:Zero Prevoicing	0.204	486	46.10%	0.551
/d/:Half Prevoicing	0.059	490	89.00%	0.515
/g/:Full Prevoicing	0.051	610	92.60%	0.513
/t/:Full Prevoicing	0.016	489	83.60%	0.504
/b/:Full Prevoicing	−0.03	492	92.70%	0.493
/b/:Full Short Lag VOT	−0.074	493	44.80%	0.481
/d/:Full Prevoicing	−0.099	495	90.70%	0.475
/b/:Zero Prevoicing	−0.121	493	72.20%	0.47
/p/:Half Prevoicing	−0.141	494	85.20%	0.465
/k/:Zero Prevoicing	−0.155	489	61.10%	0.461
/g/:Half Prevoicing	−0.228	367	92.40%	0.443
/k/:Full Prevoicing	−0.243	488	90.20%	0.44
/d/:Full Short Lag VOT	−0.284	474	52.30%	0.429
/g/:Full Short Lag VOT	−0.312	494	39.30%	0.423
/k/:Half Prevoicing	−0.363	493	85.60%	0.41
/p/:Full Short Lag VOT	−0.368	492	20.90%	0.409
/t/:Zero Prevoicing	−0.741	486	52.70%	0.323
Range				35.9
Interaction: Initial Phone*Proficiency Level ($p < 0.017$)				
/t/:Non-Native Grad	0.776	64	51.60%	0.685
/p/:Semester 6	0.517	548	54.20%	0.626
/p/:Semester 4	0.474	682	66.40%	0.616
/g/:Native Speaker	0.402	80	75.00%	0.599
/b/:Native Speaker	0.295	80	71.20%	0.573
/d/:Semester 8	0.158	595	68.20%	0.539
/t/:Semester 8	0.115	676	79.90%	0.529
/d/:Semester 4	0.11	682	69.90%	0.527
/k/:Semester 4	0.099	587	77.70%	0.525
/g/:Semester 8	0.085	603	54.60%	0.521
/t/:Native Speaker	0.083	80	57.50%	0.521
/k/:Semester 8	0.054	588	62.60%	0.514
/b/:Semester 6	0.03	549	73.00%	0.507
/b/:Semester 4	0.021	689	80.10%	0.505

Factor	**Log-odds**	**Tokens**	**Percent voiced**	**Factor weight**
/b/:Non-Native Grad	0.001	64	71.90%	0.5
/d/:Semester 6	−0.012	539	75.30%	0.497
/k/:Native Speaker	−0.014	80	65.00%	0.496
/d/:Native Speaker	−0.028	684	78.10%	0.493
/g/:Semester 4	−0.039	78	66.70%	0.49
/p/:Semester 8	−0.051	592	58.80%	0.487
/k/:Semester 6	−0.067	547	62.50%	0.483
/k/:Non-Native Grad	−0.083	64	65.60%	0.479
/g/:Semester 6	−0.213	64	59.40%	0.447
/d/:Non-Native Grad	−0.214	548	74.10%	0.447
/p/:Non-Native Grad	−0.222	62	79.00%	0.445
/g/:Non-Native Grad	−0.253	539	57.70%	0.437
/t/:Semester 6	−0.26	64	79.70%	0.435
/b/:Semester 8	−0.347	590	74.10%	0.414
/t/:Semester 4	−0.691	668	62.90%	0.334
/p/:Native Speaker	−0.727	80	61.20%	0.326
Range				35.9
Participant Level ($p < 0.0091$)				
Non-Native Grad	0.227	382	67.80%	0.557
Native Speaker	0.175	478	66.10%	0.544
Semester 4	0.067	4081	72.90%	0.517
Semester 8	−0.172	3555	65.90%	0.457
Semester 6	−0.297	3270	66.10%	0.426
Range				13.1
Random Intercept: Respondent				
96-Semester 4	1.898	96	97.90%	0.87
109-Semester 4	1.655	96	95.80%	0.84
46-Semester 8	1.439	95	87.40%	0.809
81-Semester 4	1.149	96	90.60%	0.76
100-Semester 4	1.149	96	90.60%	0.76
… (130 participants omitted)	…	…	…	…
47-Semester 4	−0.922	77	55.80%	0.285
116-Non-Native Grad	−0.956	95	55.80%	0.278
32-Semester 8	−1.03	60	43.30%	0.264
42-Semester 8	−1.284	60	38.30%	0.217
39-Semester 4	−1.617	58	39.70%	0.166
Range				70.4

Factor	Log-odds	Tokens	Percent voiced	Factor weight
Random Intercept: Word				
guia	0.249	981	80.30%	0.563
beso	0.22	988	79.10%	0.556
Kia	0.218	978	69.10%	0.555
pata	0.187	982	63.40%	0.548
teja	0.15	977	61.40%	0.538
drama	0.087	963	76.40%	0.523
deja	−0.102	987	72.60%	0.475
trama	−0.152	977	55.20%	0.463
peso	−0.187	984	57.10%	0.454
cano	−0.223	983	61.50%	0.445
bata	−0.232	984	72.30%	0.443
gano	−0.26	982	73.20%	0.436
Range				12.7

n = 11766 *df* = 62 Log Likelihood = −5367.5 AIC = 10859 R2 Fixed = 0.363 R2 Random = 0.08 R2 Total = 0.443