Studies in Second Language Acquisition (2026), 1–28
doi:10.1017/S0272263126101879

RESEARCH ARTICLE

Language aptitudes and VOT development of L3 Spanish stops in Chinese-English bilinguals

Linxi Zhang, Alfonso Morales-Front and Cristina Sanz

Department of Spanish and Portuguese, Georgetown University, Washington, DC, USA
Corresponding author: Linxi Zhang; Email: lz391@georgetown.edu

(Received 12 October 2025 UTC; Revised 12 March 2026 UTC; Accepted 29 May 2026 UTC)

Abstract

Voicing contrast is a difficult feature for Chinese-English emergent bilingual learners of L3 Spanish. Adopting a time-series approach, this study investigates developmental trajectories of L3 Spanish voiced and voiceless VOT integrating the dynamic roles of language aptitude components.To trace their developmental paths, thirty classroom learners of L3Spanish in China produced VOT samples in a wordlist reading task five times during a five-month period. Language aptitude indicators were integrated in mixed effect models. Results showed improvement on voiced stop VOT accuracy as proficiency increases, with significant individual differences in developmental paths. Statistical analysis revealed a facilitative effect of phonetic coding ability on pronunciation accuracy and an effect of sequence recognition ability in more advanced stages of learning.This study contributes to a dynamic view of voicing contrast development in multilinguals, highlighting the differential role of explicit and implicit language aptitude in distinct stages of segment pronunciation development.

Keywords: VOT; L3 Acquisition; Phonological Development; Spanish

Voice Onset Time (VOT), defined as the time period between the release of the stop and the onset of voice in the following vowel, is an acoustic measure that varies across languages and is known to be vulnerable to cross-linguistic influence among multilinguals (e.g., Flege, 1991, 1995; Flege & Eefting, 1988; Llama et al., 2010; Wrembel, 2011, 2014, 2015; Wunder, 2010). For instance, English has longer VOT in /p, t, k/ sounds than Spanish. It is not unusual for first language (L1) English-speaking learners to transfer long VOT to their second language (L2) Spanish (e.g., Zampini, 1998; Nagle, 2017, 2019), pronouncing words like "taco" with distinctive long aspiration.

For Chinese-speaking learners of Spanish, learning to perceive and produce the negative VOT in Spanish voiced stops /b, d, g/ is known to be more challenging, as learners tend to assimilate Spanish /b, d, g/ to Chinese /p, t, k/ and produce short positive VOT due to their similarity (Bravo-Díaz, 2020; Chen, 2007; Liu, 2016, 2019; Liu & Machuca Ayuso, 2021; Zhang, 2022; Zhang et al., 2023). Lack of voicing in both initial

© The Author(s), 2026. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 17 Nov 2026 at 17:08:57, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263126101879

and sentence-medial positions (i.e., pronouncing "*bollo*" in a way that is perceived as "*pollo*" by native Spanish speakers) is seen as a stereotype of Chinese-accented Spanish.

However, here as everywhere else in foreign language pronunciation, a tremendous degree of individual variation can be observed. While native-like pronunciation is generally thought to be out of reach for many second language learners (L2ers) who are exposed to the additional language after a fairly young age (6–7 years, e.g., Granena & Long, 2013), it is also not so uncommon for some adult learners of a new language to achieve excellent and even near-native pronunciation. Previous studies investigating Spanish VOT produced by Chinese university students identified individual learners who developed target-like negative VOT in Spanish voiced stops (Chen, 2007; Zhang, 2022; Zhang et al., 2023). Considering that Spanish learners in China almost uniformly start learning the language in their first year of college, factors of individual differences (IDs) other than age of acquisition must be at play to give some individuals an advantage.

Previous research on L2 speech has pointed to language aptitude as an important factor that impacts the rate and outcome of speech development (e.g., Granena, 2016; Saito, 2017, 2019; Skehan, 2002). However, there is a dearth of longitudinal and multi-wave studies that tap into how language aptitude interacts with time to shape pronunciation development. The present study aims to revisit language aptitude in an L3 classroom setting investigating the roles of language aptitude in interaction with development by applying a time-series method to trace VOT trajectories over 5 months. In particular, previous research pointed to delayed acquisition of negative VOT (prevoicing) in the target language when the feature does not prevail in the learner's previous language(s) (e.g., Face & Menke, 2009; Bravo-Díaz, 2020). This study follows the VOT production of L3 Spanish learners across a range of proficiency levels—from beginning to advanced—to trace how VOT patterns develop over time, both at the level of the full participant sample and in the individualized trajectories of learners. In doing so, the study examines which language aptitude components predict the accuracy and rate of prevoicing development in the production of Spanish voiced stops.

VOT and voicing contrast in Spanish

VOT refers to the time-lapse between the release of the consonant closure and the onset of voicing of the following segment. Lisker and Abramson (1964) were the first study to propose using VOT as a unified phonetic measure of the phonemic voicing contrast. Their cross-language study demonstrated three ways that voicing contrast is mediated through VOT values: 1) negative VOT (i.e., voicing begins before the release of the stop, also known as "prevoicing") is typical in voiced stops; 2) short-lag VOT (i.e., voicing begins very shortly after the release of the stop, typically between 0 and ~30 ms) indicate voiceless, unaspirated stops; 3) long-lag positive VOT (i.e., voicing starts much later, typically 40 to 100 ms or more) is characteristic of voiceless, aspirated stops.

The combination of L1 Mandarin-L2 English-L3 Spanish involves all three VOT scenarios. Spanish is known as a voicing language, in which voiced stops in word-initial position typically show negative VOT (prevoicing) ranging from −60 to −100 ms depending on the place of articulation, e.g., "***bollo***" (bread roll). Voiceless stops have short, positive VOT typically shorter than 30 ms, e.g., "***pollo***" (chicken) (see Rosner et al., 2000). Although English has a phonemically voiced stop category /b, d, g/, these stops are often realized as voiceless with short-lag VOT by the majority of English speakers in word-initial position (e.g., bike), although some individuals prefer

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 17 Nov 2026 at 17:08:57, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263126101879

Table 1. Mean VOTs of Stops in English, Mandarin, and Spanish.

	/p^h/	/t^h/	/k^h/	/p/	/t/	/k/	/b/	/d/	/g/
Mandarin Liao (2005) Rochet & Fei (1991)	99.6	98.7	110.3	17.9	18.6	21			
English Lisker & Abramson (1964)				58	70	80	1 −101	5 −102	21 −88
Spanish Rosner et al. (2000)				13.1	14	26.5	−91.5	−91.6	−73.7

prevoicing (Lisker & Abramson, 1964; Herd, 2020). English voiceless stops /p, t, k/ are aspirated in word-initial positions with long-lag VOT typically between 50 and 90 ms. Although some southern Chinese languages, such as Wu and Min, preserve voiced-voiceless stop contrasts from ancient Middle Chinese (Han, 2025; Zhai, 2022), modern Mandarin is generally assumed not to have a phonemically voiced category, contrasting two voiceless series of stops differentiated by aspiration—the unaspirated stops /p, t, k/ have short-lag VOT, e.g., "***baigu***" (skull) and the aspirated /p^h, t^h, k^h/ have long-lag VOT, e.g., "***paigu***" (rib). Previous studies showed that Mandarin aspirated stops generally have longer VOTs than their English counterparts (Liao, 2005; Rochet & Fei, 1991). Table 1 summarizes mean VOTs of stops in English, Mandarin, and Spanish according to different authors.

The Perceptual Assimilation Model for L2 (PAM-L2; Best & Tyler, 2007) proposes that late-learned additional languages are perceived through the filter of previously acquired phonological categories, typically those of the L1. According to the model, learners develop a unified interlanguage phonological system in which contrasts in the new language are interpreted relative to prior language knowledge. As in the original PAM model for monolingual perception (Best, 1994), non-native contrasts may be assimilated in several ways: two-category assimilation (the two sounds are mapped onto distinct native categories), single-category assimilation (both sounds map onto the same native category), category-goodness assimilation (both sounds map to the same native category but differ in perceived fit), and uncategorized assimilation (a sound does not match any native category).

In the case of L3 Spanish acquisition by L1 Mandarin speakers with prior experience in L2 English, additional cross-linguistic interactions may influence assimilation patterns (cf. PAM-L3; Wrembel et al., 2019). Since fully voiced stops with negative VOT do not exist in Mandarin, and given that English provides partially overlapping stop categories (e.g., short-lag VOT voiceless stops), learners may initially map Spanish voiced stops /b, d, g/ onto their L1 voiceless unaspirated categories, which exhibit short-lag VOT values somewhat closer to the negative VOT of Spanish stops. This prediction aligns with empirical findings showing that Chinese learners often produce both Spanish /b, d, g/ and /p, t, k/ with short-lag VOT values, without differentiating voicing contrasts (Bravo-Díaz, 2020; Chen, 2007; Zhang, 2022; Zhang et al., 2023). Nevertheless, these studies have also documented individuals who successfully produce prevoicing in Spanish, particularly among more advanced learners. The core argument drawn from PAM-L2 for the present study is that acquiring Spanish prevoicing is not merely a matter of adjusting phonetic details, but rather a structurally difficult process of breaking a "single-category" assimilation. To successfully produce negative VOT, learners must establish an entirely new phonetic category—a cognitive hurdle that

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 17 Nov 2026 at 17:08:57, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263126101879

explains why prevoicing development is often delayed, and why individual differences in language aptitude may be critical in helping learners overcome this initial L1 perceptual filter.

While perception typically guides and overlaps with production (Nagle, 2018; Nagle & Baese-Berk, 2022), the ability to produce target-like prevoicing implies that some learners have established new L2 phonetic categories and acquired the articulatory control necessary for accurate production. The Speech Learning Model-revised (SLM-r; Flege et al., 2021) offers a complementary framework for understanding these production outcomes, emphasizing that phonetic category formation is shaped by factors such as amount and quality of input, age of acquisition, language learning experience, and individual differences. According to SLM-r, successful acquisition of new phonetic categories in production depends not only on the perceptual similarity between L1 and L2 sounds, but also on learners' ability to detect phonetic distinctions, receive sufficient input, and develop the requisite auditory-motor skills (see discussions in Saito et al., 2024). Therefore, investigating individual difference variables, especially in an interactive and dynamic manner, is crucial for understanding variability in L2 and L3 pronunciation and phonological category development (Nagle, 2022).

Language aptitude and phonological development

Foreign language aptitude (LA) generally refers to a set of cognitive abilities that affects the rate, speed, and ease with which one individual, usually an adult, learns a foreign language, and has long been studied as a predictor of foreign language learning success (e.g., Carroll, 1962, 1981; Granena & Long, 2013; Li, 2015, 2016; Skehan, 2002, 2016). Although the field has not yet reached a complete agreement on the cognitive components of LA, three major abilities, as measured by the five subsets of the MLAT (Modern Language Aptitude Test, Carroll and Sapon, 1959, 2002) are considered essential to SLA, namely phonetic coding ability (i.e., sound-symbol mapping), language analytic ability (i.e., grammatical sensitivity and inductive learning ability), and memory. Research has demonstrated that LA is independent of other individual variables such as motivation and is specific to the domain of language learning, although it overlaps with general intelligence (Li, 2015, 2016). Evidence has shown an association between LA and general second language proficiency (e.g., Carroll, 1962, 1981; Dörnyei & Skehan, 2003; Granena & Long, 2013; Sasaki, 1996). With the aim to theorize LA within the frame of SLA, Skehan (2016) proposed the acquisition-aptitude model, hypothesizing that different types of aptitude are uniquely tied to different stages of L2 learning, i.e., phonemic coding is linked to analyzing incoming input, associative memory is linked to automatizing partially acquired knowledge, and sequence recognition is linked to attaining advanced-level use of the language.

Over the years, the majority of empirical studies on LA have focused on how different components of aptitude relate to the acquisition of grammar and morpho-syntactic structures (e.g., Granena, 2013; Yilmaz & Granena, 2016; Yalçın & Spada, 2016; see Li, 2015, for meta-analytic review), while relatively fewer studies have examined the relationship between aptitude and L2 phonological acquisition. And among these, the majority focus on learners' overall speaking abilities, most commonly taking foreign accent as the dependent variable. Granena and Long (2013) studied a group of 65 Chinese learners of Spanish living in Barcelona and investigated the relationship between age of onset, length of residence, language aptitude, and ultimate L2 attainment in three linguistic domains. In the domain of phonology, they found

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 17 Nov 2026 at 17:08:57, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263126101879

aptitude sub-tests measuring phonetic coding ability and grammatical inference to have the strongest correlation with learners' degree of foreign accent as rated by native speakers. Another study, Hu et al. (2013), also found phonetic coding ability, as measured by MLAT-spelling cues, to be a predictor of advanced learners' rated foreign accent.

Research examining a wider range of variables related to oral production has uncovered the link between LA components and different aspects of speech development. Saito (2017) examined the relationship between aptitude components measured using the LLAMA subtests (Meara, 2005; Meara & Rogers, 2019) and a range of pronunciation, fluency, vocabulary, and grammar measures of elicited picture descriptions produced by Japanese ESL learners. Results indicated that explicit aptitudes, specifically phonetic coding ability (LLAMA E), associative memory (LLAMA B), and language analytic ability (LLAMA F), were moderately associated with phonological accuracy, fluency, and lexicogrammar complexity in oral production. In particular, phonetic coding ability was significantly associated with the segmental measure of pronunciation rating. These results highlighted the link between phonetic coding ability and phonological accuracy in L2 speech, especially in limited-input classroom conditions, as it "might help L2 learners notice and integrate new linguistic knowledge more efficiently via optimizing the processing of incoming linguistic input" (p. 683).

There has been a dearth of studies on the role of LA in the acquisition of segmental features. Saito (2019) investigated the relationship between different components of aptitude and the attainment of different dimensions of English /ɹ/ by Japanese learners in reading tasks. Learners with higher phonetic coding ability (LLAMA E) performed better in a relatively easy dimension of English /ɹ/ (lower F2 for tongue retraction), while those with greater associative memory (LLAMA B) demonstrated an advantage in a more difficult dimension of English /ɹ/ (longer transition duration for phonetic length; lower F3 for labial/alveolar/pharyngeal constrictions). In another study, Saito et al. (2019a) investigated how aptitude and experience factors are associated with English oral production proficiency of Chinese international students in the UK. Phonetic coding ability (LLAMA E) emerged as a predictor for segmental attainment, along with neurophysiological measures of implicit aptitude. These results highlight the important roles of explicit language aptitudes in L2 learners' segmental attainment, and pointed to an interaction between aptitude and segment difficulty and/or learnability.

Up to now, only one longitudinal study has been conducted to uncover the dynamic link between LA components and pronunciation development. In a year-long investigation of a group of English learners in Japan, Saito et al. (2019b) identified both phonetic coding ability (LLAMA E) and associative memory (LLAMA B) as contributors to general comprehensibility in the first semester. However, during the second semester, as learners' exposure to English continued to accumulate, sequence recognition ability (SRA), an implicit aptitude component measured by LLAMA D subset, played a more decisive role in segmental accuracy. Their findings on oral proficiency development at a macro level corroborate Skehan's hypothesis that explicit aptitudes are stronger predictors for attainment in L2 oral production during earlier stages of learning, while implicit aptitude characterized by the capacity of inducing patterns from natural input might benefit learners in more advanced stages. Similar conclusions were reached by Serafini and Sanz (2016) for morphosyntactic development.

By conducting a longitudinal examination of a specific segmental feature (VOT) at the micro-level, this study aims to contribute to the existing literature in three distinct ways. First, while previous studies have largely confirmed the role of explicit language

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 17 Nov 2026 at 17:08:57, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263126101879

aptitude in predicting macro-level pronunciation outcomes like overall foreign accent, this study tests whether those findings hold true for the micro-level articulatory control required for L3 prevoicing. Second, by tracking learners over 5 months, this research addresses a methodological gap; it moves beyond cross-sectional snapshots to uncover whether different aptitude components (specifically explicit versus implicit aptitudes) exert their influence at different developmental stages, as hypothesized by Skehan's (2016) acquisition-aptitude model. Finally, this study generates new knowledge regarding L3 phonological development, revealing how individual cognitive differences mitigate or exacerbate the cross-linguistic challenges faced by Mandarin-English bilinguals acquiring Spanish voiced stops.

In light of these objectives, the present study is mainly guided by the following questions:

RQ1. ***What developmental paths do Mandarin-English bilingual learners follow in their Spanish voiced and voiceless VOT over 5 months?***

> We expect asymmetric patterns in the VOT growth paths of voiced and voiceless stops. Because Spanish voiceless stops align perceptually with Mandarin unaspirated stops (both exhibiting short-lag VOT), we predict learners will demonstrate target-like accuracy for voiceless stops from the beginning. Conversely, because negative VOT (prevoicing) does not prevail in Mandarin, we hypothesize its development in Spanish voiced stops will be delayed and characterized by high levels of inter-subject variation over time. Moreover, the fact that prevoicing is more marked (due to its articulatory effort and lower perceptibility) than aspiration also makes us expect asymmetric patterns.

RQ2. ***To what extent do individual differences in language aptitude predict Mandarin-English bilingual learners' production of prevoicing?***

a. Which language aptitude indicator(s) predicts accuracy in prevoicing production?

Based on previous findings that highlighted the role of explicit aptitude in segmental pronunciation, we expect stronger phonetic coding ability (PCA) and associative memory (AM) to facilitate production of prevoicing. Specifically, stronger PCA should help learners notice and process fine-grained acoustic cues like prevoicing in the input, while AM should assist in storing these new sound-to-symbol mappings in short-term memory during early acquisition stages.

b. Which language aptitude indicator(s) predicts the rate of development in prevoicing production?

Following Skehan's (2016) acquisition-aptitude model, initial input processing is tied to explicit aptitudes, whereas sequence recognition ability (SRA)—an implicit aptitude—is linked to extracting regularities from natural input and proceduralizing knowledge over time. Because our 5-month longitudinal design captures the continuous accumulation of L3 exposure, we expect SRA to interact with time and proficiency, serving as the primary predictor for the *rate* of accelerated prevoicing development.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 17 Nov 2026 at 17:08:57, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263126101879

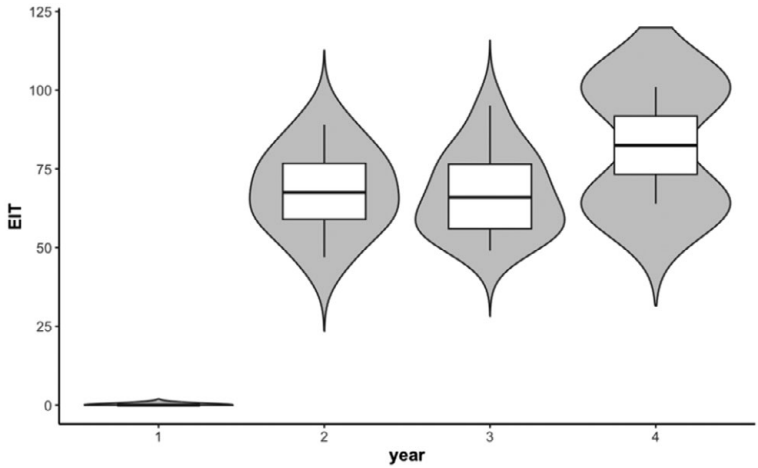


Figure 1. Violin boxplot of initial EIT scores of participants by year.

Methodology

Participants

An initial group of 33 sequential trilingual speakers of L1 Mandarin-L2 English-L3 Spanish participated in the study; 30 of these participants completed all data collection ($N = 30$). Their ages range from 17 to 21 ($M = 19.6$, $SD = 1.21$). They were all university students majoring in Spanish in four institutions of higher education located in northern China. All participants are native speakers of Mandarin dialects who typically started learning English in primary school (age, $M = 6.9$, $SD = 2.0$), and started learning Spanish when entering university (age, $M = 17.9$, $SD = 0.6$). Participants were enrolled in different years (1–4) in their undergraduate program and had varied proficiency levels in Spanish; their proficiency levels were measured using the Spanish Elicited Imitation Test (EIT, Ortega et al., 2002). At the time of first data collection, year-1 students ($n = 5$) were beginners with barely any previous contact with the language; year-2 students ($n = 8$) had studied the language for two semesters; year-3 ($n = 15$) students for four semesters; and year-4 students ($n = 2$), six semesters. However, participants' EIT oral proficiency test scores indicate that their Spanish proficiency was not always aligned with their academic placement groups. For this reason, participants were not divided into groups by their placement. Instead, we examined their proficiency (EIT scores) as a continuous variable. Figure 1 illustrates the distribution of initial Spanish proficiency (EIT score at beginning of the study with a range of 0 to 120).

Only three participants had studied abroad. Participants 17 and 25 (both year-4) had spent 10 and 7 months, respectively, studying in Spain, but had returned to China by the time data collection started. Participant 10 was the only student who was abroad during the length of the study. The first data collection coincided with her first month studying in Santiago de Compostela, Spain, and the last data collection was shortly after she completed her semester abroad and returned to China. No other participants reported living in or traveling to English or Spanish-speaking countries for more than one month.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 17 Nov 2026 at 17:08:57, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263126101879

Three participants reported minimal knowledge of additional foreign languages, including Korean (Participant 3, AoA = 14), Portuguese (Participant 3, AoA = 20), Japanese (Participant 23, AoA = 18), and German (Participant 27, AoA = 15). Their self-rated oral proficiency in these additional languages was between 1 and 3 on a scale of 0 to 7. Participant 18 was the only speaker of Wu—a southern Chinese dialect that preserves voiced stops, but indicated minimal exposure to the language after moving to Beijing during elementary school.

Experimental procedure

The protocol for the study was approved by the university's IRB (MOD00005370). Participants provided informed consent online when they first reported to the experiment and completed an online Language History Questionnaire. They agreed to meet with the primary researcher five times, once per month. At each session, participants were recorded reading word lists in English, Mandarin, and Spanish, as well as a short text in Spanish, as part of a larger project. Their Spanish oral proficiency was assessed twice during the study, at Time 1 and Time 5. At the midpoint of the study (Time 3), participants completed the LLAMA language aptitude tests online. The present study analyzes only the recordings of the Spanish word list readings.

Word list reading task

Word lists

The word lists contained voiced and voiceless stop consonants in word-initial position in Mandarin, English, and Spanish, but only Spanish data were analyzed for the current study. Word lists have been used as instruments in previous studies on stop consonants as a measure to gather VOT in a controlled way both in L2 (Flege, 1991; Yavas & Wildermuth, 2006) and in L3 studies (Llama & Cardoso, 2018; Tremblay, 2007; Wrembel, 2014). Most of these studies looked at stops in initial position because the difference in VOT is the most prominent both word-initially and in the onset of stressed syllables.

All the words are frequent real words selected from Spanish textbooks used by Chinese college Spanish majors (*Español Moderno* I and II). A pre-study survey completed by a separate group of Chinese college students ($n = 7$) with similar profiles assessed how familiar they were with the words, based on the frequency with which they heard or pronounced them. The included words received an average familiarity ranking of 4.88 out of 5 ($SD = 0.14$). The list includes 5 items for each of the 6 consonants—/b, d, g, p, t, k/—resulting in 30 target words with stop consonants in stressed, word-initial position (e.g., /pa.to/, "duck"). Twenty distractors that do not contain stop consonants were added to the word lists. The words were presented embedded in a short carrier sentence "*Voy a decir*, #_____" (*I will say*). The carrier sentences were designed to ensure that the stop consonants were located in word-initial position after a pause. The word list is available in the IRIS database (Zhang, 2026).

Recording procedure

Due to COVID-19 restrictions, data collection was conducted remotely via Zoom. Participants joined individual sessions in which the experimenter shared a screen

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 17 Nov 2026 at 17:08:57, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263126101879

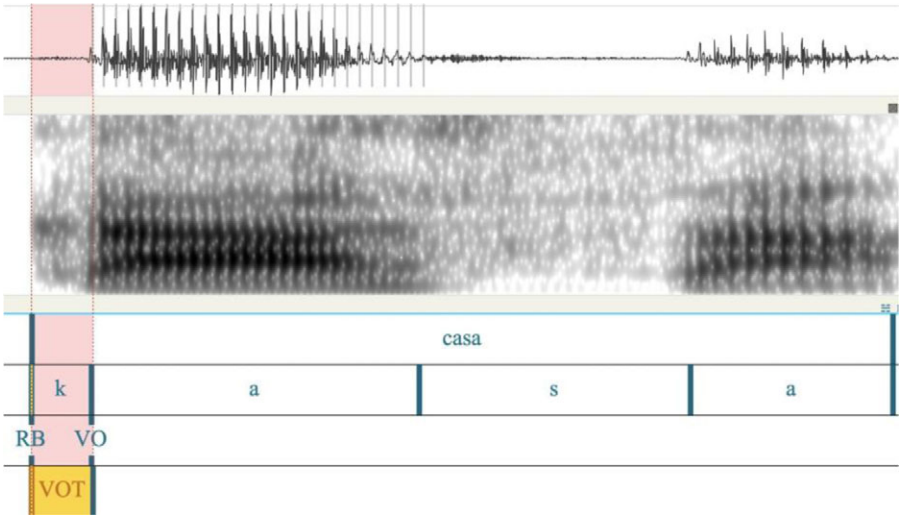


Figure 2. Positive VOT of voiceless stop.

displaying randomized words presented with PsychoPy software (Peirce et al., 2019). As each target word appeared within the carrier sentence, participants read the sentences aloud. Audio was recorded locally on the experimenter's computer using Praat (Boersma & Weenink, 2018), in mono mode at 44,100 Hz, thereby avoiding compression artifacts or bandwidth-related distortions from online transmission.

Since the primary measure was voice onset time (VOT), a durational parameter relatively robust to minor variations in recording quality (Ge et al., 2021), the procedure was well suited to the study's aims. When connectivity issues led to noticeable lags or dropouts, immediate repetitions were elicited to ensure data quality. The recording procedure was repeated five times at one-month intervals.

Coding of VOT

To answer RQ1, data from the Spanish word list collected at five points during the semester were coded for VOT values. A total of 4,500 stop tokens were annotated manually using Praat and extracted using the Praat script VOT logger (DiCanio, 2011). Of those, 131 tokens (2.9%) were excluded from the analysis due to mispronunciation or poor recording quality, resulting in 4,369 tokens. Following Lisker and Abramson (1964), VOT was annotated as the time lapse between the release burst of the stop (RB) and the onset of the following vowel (VO), marked by the onset of periodic waveform and vowel formant. Stops pronounced with positive VOT show a period of aspiration before the periodicity of the vowel with no voicing during the stop closure (Figure 2), while stops produced with prevoicing are characterized by periodicity before the onset of the vowel, marked as the negative VOT (Figure 3).

Aptitude tests

Participants' language aptitude was measured by Version 3 of the LLAMA tests released in 2019 (Meara & Rogers, 2019; available at https://www.lognostics.co.uk/tools/

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 17 Nov 2026 at 17:08:57, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263126101879

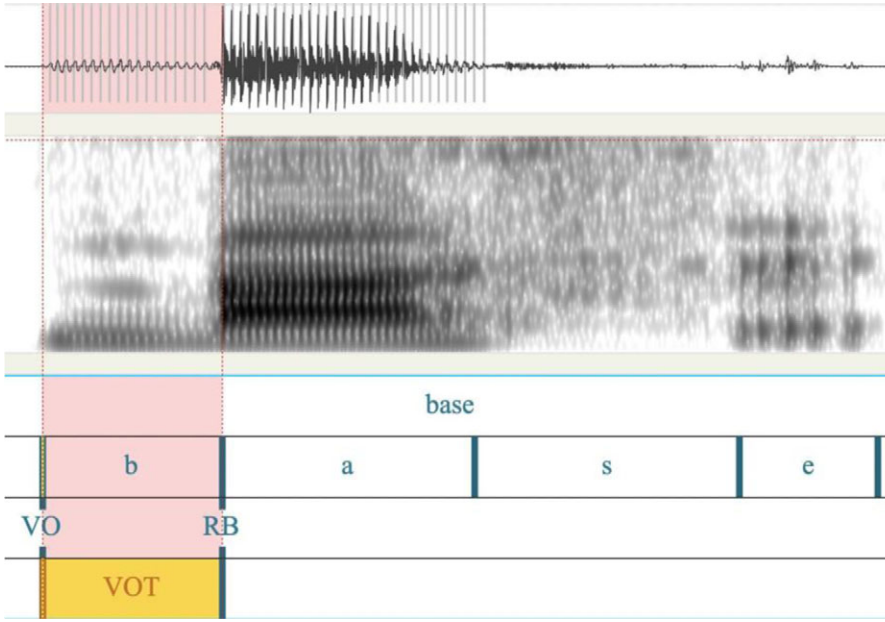


Figure 3. Negative VOT of voiced stop.

LLAMA_3/). This version of the tests is completely web based and has made important changes to the original version (Meara, 2005; Rogers et al., 2023), addressing criticisms concerning its internal validity (Bokander & Bylund, 2020; Bokander, 2023). Recent analyses conducted by the LLAMA lab based on large-scale tester data have shown that LLAMA v.3 has improved internal consistency and reliability over LLAMA v.1 (Bokander et al., 2023). According to Skehan (2016), the LLAMA is considered a domain-specific language aptitude test, in which linguistic materials are used to target four language-specific learning abilities: phonetic coding ability (LLAMA E), associative memory (LLAMA B), grammatical inference ability (LLAMA F), and sequence recognition ability (LLAMA D). The tasks use an artificial language in a way that simulates the experience of learning a new, unfamiliar language. The LLAMA tests provide researchers with an L1-neutral and open-source tool that has been increasingly used in SLA studies, generating a pattern of generally consistent and intuitive results (Artieda & Muñoz, 2016; Granena, 2013; Granena & Long, 2013; Saito, 2017, 2019; Saito et al., 2019a, 2019b).

Following Saito (2019), subsets LLAMA_E, LLAMA_B, and LLAMA_D were selected. These subsets are believed to highlight different modes of L2 learning: LLAMA_E and LLAMA_B measure language aptitude in an explicit and intentional learning context, giving participants a practice phase to study the new language intentionally before the testing phase, while LLAMA_D taps into more implicit aptitude component, characterized by the absence of practice phase and explicit learning of materials, also believed to require a larger amount of input to take effect (see Saito, 2019 for detailed introduction and meta-analysis of Li, 2015 for a review of cross-test validity). Following Saito's suggestion, participants completed the subtests online in the following order: LLAMA_D, followed by LLAMA_E, and then

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 17 Nov 2026 at 17:08:57, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263126101879

LLAMA_B, to ensure that implicit aptitude was measured first to avoid participants' awareness and attention to the task. The LLAMA aptitude tests are automatically scored once participants complete each subset of the tests online. The maximum score is 20 for each subtest.

In addition to the word list reading task and online LLAMA aptitude tests, participants also completed a language history questionnaire (adapted from Li et al., 2020) at the beginning of the study and a language use survey before each data collection (both can be found in the IRIS database; Zhang, 2026).

Statistical analysis

Linear mixed-effects models were run in R (R Core Team, 2019) and the lme4 package (Bates et al., 2015) to fit the growth paths of voiced (vd) and voiceless (vl) VOT, respectively. Prior to model fitting, the data were examined to ensure statistical assumptions were met; visual inspection of Q-Q plots and residual plots confirmed the normality of residuals and the linearity of predictors, with no significant outliers distorting the data. Following the procedure suggested by Singer and Willet (2003), we first fit unconditional linear growth models without any predictor variables to examine the effect of time, controlling words and Participant ID as random effects. Then, by-participant random slopes for time were added to inspect variabilities in trajectories over time. Lastly, we fit conditional LMM models with Spanish proficiency (EIT score) and language aptitude scores as predictors and words and Participant ID as random effects to examine what variables may affect the initial point and changing slope of VOT.

Results

The goal of the study is to understand L3 phonological development by investigating the developmental paths that learners show in Spanish VOT and the ways in which language aptitude predicts them.

RQ1. *What developmental paths do Mandarin-English bilingual learners follow in their Spanish voiced and voiceless VOT over 5 months?*

To answer RQ1, and to single out the effect of time, we conducted an unconditional LMM with Time as a fixed effect and Participant ID and Word as random effects on voiced (vd) and voiceless (vl) VOT, respectively. We first present group trends followed by individual analysis that highlights inter-subject variability in VOT trajectories.

Group paths

The LMM model for voiced stops yielded a significant effect of Time ($\beta = -0.05$, $t = -3.5$, $p < 0.01$, 95% CI [−0.07, −0.02]), indicating that participants' voiced VOT significantly decreased by 2.24 ms on average between each time point over the period. To further investigate whether participants' developmental trajectories varied significantly, by-participant random slopes for time were added to the model. ANOVA for model comparison revealed significantly better goodness of fit ($p < 0.01$), indicating significant variability in individual trajectories over time.

The same procedure was applied to voiceless VOT. Unconditional LMM revealed lack of significant effect for Time ($\beta = 0.00$, $p = 0.68$, 95% CI [−0.01, 0.02]). However, the model was significantly improved by adding a by-participant random slope for

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 17 Nov 2026 at 17:08:57, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263126101879

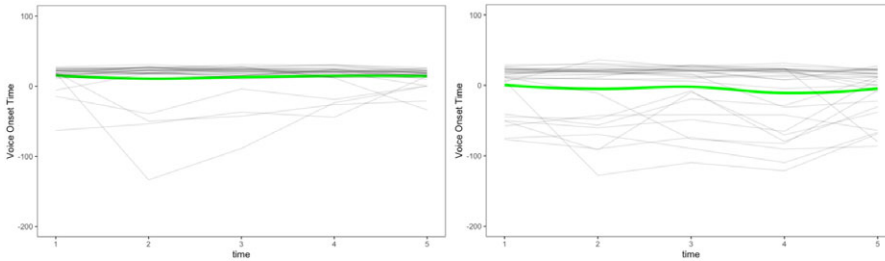


Figure 4. Group path of voiceless (left) and voiced (right) VOT.

Time ($p < 0.01$), meaning that although Time did not have an overall effect, there was significant variability in individual voiceless VOT trajectories.

Figure 4 graphs interpolation lines that demonstrate group voiceless and voiced VOT paths.

Individual paths

Individual paths demonstrated a great deal of variation in relation to both intercept—their initial VOT values—and slope—the direction and rate of VOT change over time (see Figure 5).

It is notable to observe that while the group curve for voiced VOT demonstrated improvement over time, the majority of participants showed flat lines of both voiceless (vl) and voiced (vd) VOT, with the two largely overlapping (Participants 1, 4, 5, 9, 13, 14, 27, 28, 33). These participants demonstrated "neutralization" between voiced and voiceless stop VOTs in Spanish, pronouncing both groups of stops with short-lag VOT, as observed in Chen (2007), Zhang (2022), and Zhang et al. (2023). However, other participants showed clear signs of prevoicing development from the beginning or during the course of study. The majority of these "prevoicing" participants (Participants 2, 3, 7, 8, 12, 16, 19, 23) consistently produced negative VOT, but the values of their VOTs showed "W-shaped" patterns marked by ups and downs over time. This demonstrates a destabilization of the emerging voice-lead category.

Further inspection of participants' profiles demonstrated that an increase in proficiency did not guarantee prevoicing development. For instance, Participant 14 was a third-year student and was in his fifth semester as a Spanish major. He had an EIT proficiency score of 85 out of 120 (70.8%) at Time 1 of the study. His VOT trajectory did not show any development of prevoicing over the course of the study. In contrast, Participant 2 was a first-year student in her first semester of college learning Spanish *ab initio*. However, since Time 2, her VOT trajectory showed clear prevoicing development. Subsequent analysis of the dynamic roles of language aptitude will shed light on inter- and intra-subject variation exemplified here.

RQ2. ***To what extent do individual differences in language aptitude predict Mandarin-English bilingual learners' production of prevoicing?***

RQ2 is concerned with the effect of language aptitude components on the development of prevoicing in Spanish voiced stops, and how aptitude components interact with Spanish Proficiency and Time to shape the rate of development. To answer these questions, conditional LMM models were applied to investigate the

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 17 Nov 2026 at 17:08:57, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263126101879

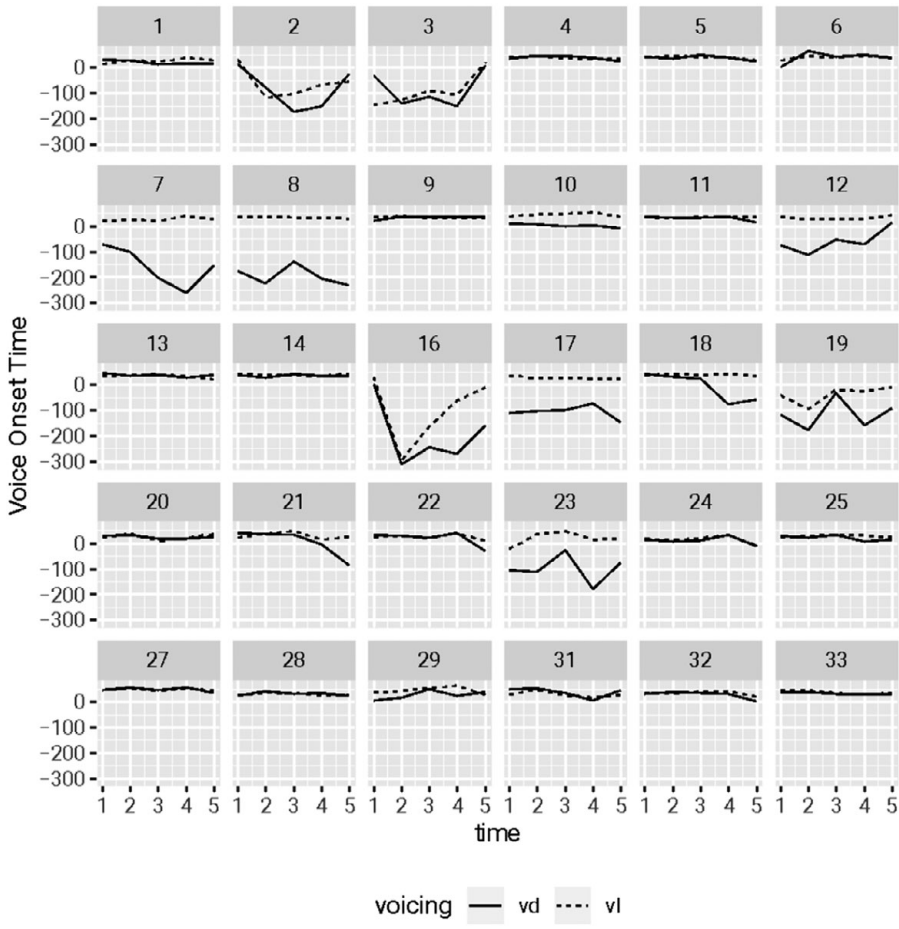


Figure 5. Individual paths of voiceless (vl) and voiced (vd) VOT.

effect of language aptitude measures—LLAMA D, LLAMA B, and LLAMA E—on participants' production of voiced VOT, as well as the interaction between each indicator and time.

Statistical profile of the variables

Table 2 summarizes scores from the LLAMA sub-tests (out of 20 for each sub-test) and pre/post test scores from the EIT tests (out of 120); it shows a striking level of interpersonal variability. Data were scaled by subtracting mean and dividing by standard deviation (*SD*) when entered into statistical analysis.

A Spanish proficiency gain score was calculated for each participant (SP EIT T5 – SP EIT T1). As displayed in Table 2, participants' Spanish oral proficiency score increased by an average of 17.6 points from the first to the fifth months, although not all participants improved their Spanish every semester (four 3rd-year learners showed a decrease in their proficiency score).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 17 Nov 2026 at 17:08:57, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263126101879

Table 2. Descriptive Statistics of Variables.

Variable	Mean	SD	Min.	Max.
Language Aptitude				
LLAMA D	10.60	4.13	0	15
LLAMA B	10.43	5.18	2	20
LLAMA E	7.63	4.83	0	17
Spanish EIT				
Time 1	57.2	29.1	0	101
Time 5	74.8	21.2	31	118
Gain	17.6	22.0	−11	82

Table 3. Correlation Coefficients (Pearson's r) Among Variables.

	LLA. D	LLA. E	LLA. B	EIT.T1	EIT.T5	EIT gain
LLA. D	1.00					
LLA. E	0.16	1.00				
LLA.B	0.26	0.10	1.00			
EIT.T1	0.09	−0.27	0.12	1.00		
EIT.T5	0.10	0.07	**0.40***	**0.66****	1.00	
EIT gain	−0.01	**0.42***	0.22	**−0.69****	0.10	1.00

Correlation Matrix Analysis (see Table 3) was conducted primarily to check for multicollinearity. According to Plonsky and Oswald's (2014) field-specific benchmarks for correlation strength in L2 research, correlation coefficients (r) close to .25 should be considered small, .40 medium, and .60 large. Not surprisingly, participants' EIT score at Time 5 was highly correlated with their score at Time 1 ($r = 0.66$, $p < .01$). At the same time, EIT gains showed a large-strength negative correlation with EIT T1 ($r = -0.69$, $p < .01$), indicating that participants who had lower proficiency at the beginning of the study demonstrated more improvement during one semester. To avoid multicollinearity issues, we included SP EIT T1 as the sole proficiency measure in the models. The matrix also revealed significant moderate correlations between specific aptitude components and proficiency measures—namely, LLAMA E with EIT gain and LLAMA B with final EIT T5 scores; all of which aligns well with the premise that explicit aptitude facilitates overall language development. It is worth noting that the LLAMA sub-test scores were not significantly correlated with each other, confirming the previous statement that the sub-tests LLAMA D, B, and E each target a different aspect of language aptitude (e.g., Saito, 2019).

Model selection was conducted using the Anova() function in R in an iterative process until the minimally adequate model (i.e., with the least number of predictors and with the best goodness of fit) was selected. The final selected model was as follows:

```
Model <- lmer (VOT ~ time +
LLAMA_e + LLAMA_b + LLAMA_d + SPN_EIT_t1 +      # predictors
time*LLAMA_d + time*SPN_EIT_t1 +                # interactions
(1 | id) +                          # by-participant random intercepts
(1 | word),                         # by-word random intercepts
data = df.vd)
```

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 17 Nov 2026 at 17:08:57, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263126101879

RQ2a. ***Which language aptitude indicator(s) predicts accuracy in prevoicing production?***

LMM with vd VOT as response (see Table 4) yielded a significant effect for Time ($\beta = -0.05$, $p < .001$, 95% CI [−0.07, −0.02]), indicating that participants' overall voiced VOT progressively and significantly decreased toward negative values over time. LLAMA E emerged as a significant main effect predictor for voiced VOT ($\beta = -0.35$, $p < 0.01$, 95% CI [−0.61, −0.09]). Because the iterative model selection process eliminated the non-significant interaction between LLAMA E and Time, this main effect indicates an overall, consistent influence across the entire 5-month period. Controlling for proficiency, participants with stronger phonetic coding ability (PCA), as operationalized by higher LLAMA E scores, produced significantly lower overall voiced VOT across all data collection points (Figure 6). LLAMA B scores were associated with higher VOT ($\beta = 0.35$, $p < 0.01$, 95% CI [0.10, 0.60]) (Figure 7). No main effect was isolated for sequence recognition ability (LLAMA D).

RQ2b. ***Which language aptitude indicator(s) predicts the rate of development in prevoicing production?***

Against our prediction, none of the aptitude components interacted significantly with proficiency. However, sequence recognition ability (SRA), measured by LLAMA D test, showed a significant interaction with Time ($\beta = -0.04$, $p < 0.01$, 95% CI [−0.06, −0.01]) (Figure 8). Participants with stronger SRA showed accelerated decrease of voiced VOT over time, evidencing faster development of prevoicing over the 5 months of study. These results mirror previous research that highlighted the link between PCA and segment production accuracy at the macro-level (Granena, 2013; Saito, 2017, 2019; Saito et al., 2019a) and corroborate findings suggesting that SRA exerts its effect with continuous language exposure over time (Granena, 2020; Saito et al., 2019b).

Additionally, Spanish oral proficiency operationalized as EIT score at Time 1 showed a significant effect on Spanish voiced VOT ($\beta = -0.42$, $p < 0.01$, 95% CI [−0.69, −0.16]). As expected, participants at higher proficiency level produced significantly lower voiced VOT, suggesting more target-like prevoiced tokens with negative VOT values. Proficiency also interacted with time to predict the slope of change. The significant interaction between Time and EIT score ($\beta = 0.09$, $p < .001$, 95% CI [0.06, 0.11]) showed that participants with higher initial Spanish proficiency had a slower rate

Table 4. Summary of Significant Parameters in the LMM of the Overall vd VOT Data.

Predictors		Coefficient (β)	*p*-value	t-value
Fixed Effects				
	Intercept	−0.06	0.67	−0.42
	time	**−0.05*****	<.001	−3.54
	LLAMA E	**−0.35****	0.008	−2.67
	LLAMA B	**0.35****	0.007	2.70
	SP EIT T1	**−0.42****	0.002	−3.11
	time* SP EIT T1	**0.09*****	<.001	6.46
	time*LLAMA D	**−0.04****	0.006	−2.75
Random Effects		**Variance**	**SD**	
	words	0.02	0.15	
	subject	0.44	0.67	
	Residual	0.77	0.88	

Note: $p < .1$; $p < .05$; $p < .01$**; $p < .001$***

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 17 Nov 2026 at 17:08:57, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263126101879

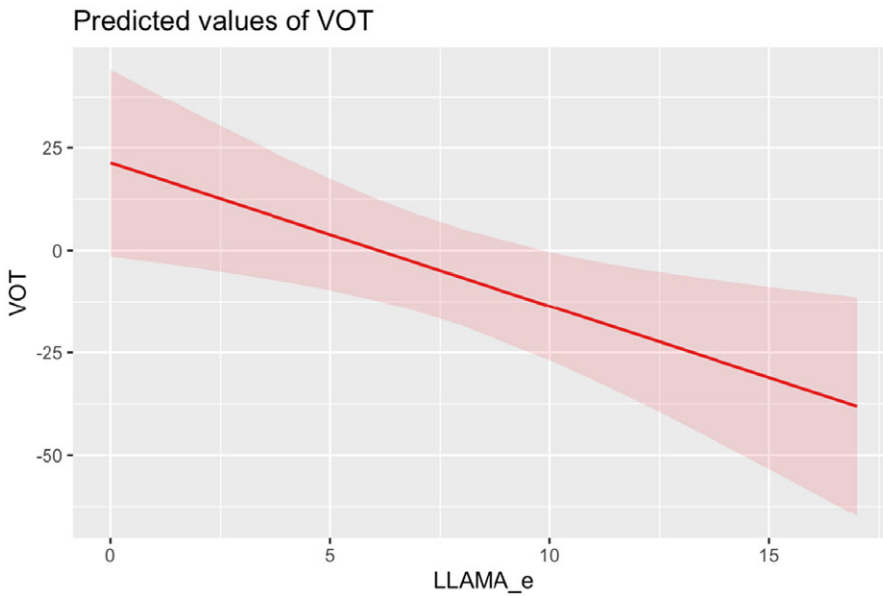


Figure 6. Predicted voice onset time (VOT) as a function of LLAMA E scores based on the linear mixed-effects model. Shaded areas represent 95% confidence intervals.

of change over time, while those with lower proficiency demonstrated a more prominent change.

Discussion

This study aims to investigate, firstly, the VOT developmental paths of Spanish voiced (vd) and voiceless (vl) stops produced by Mandarin-English sequential bilingual learners; secondly, how language aptitude as an individual difference variable may influence the development of prevoicing in voiced stops, a difficult feature for this learner population.

RQ1. *What developmental paths do Mandarin-English bilingual learners follow in their Spanish voiced and voiceless VOT over 5 months?*

At the group level, a mixed-effect model was applied to vd and vl VOTs separately to examine the effect of Time and showed distinct developmental paths for vd and vl stops as predicted. Specifically, as time progressed for 5 months, participants' mean vd VOT decreased significantly from 0.37 ms to −8.69 ms, showing improvement toward target-like prevoicing in Spanish. In contrast, vl VOT showed a flat pattern with target-like short-lag VOT sustained from Time 1 to Time 5. In fact, none of the participants produced /p, t, k/ with long-lag VOT, including the four *ab initio* learners (Participants 2, 16, 21, and 29).

The target-like short-lag VOT in vl stops confirms that Chinese learners map Spanish /p, t, k/ to Mandarin unaspirated /p, t, k/ from initial stages of learning. This differs from the pattern observed in English-speaking learners of Spanish, who typically map Spanish /p,t, k/ onto English aspirated stops [p^h, t^h, k^h], resulting in the production of relatively long VOT values. Such distinction might be a consequence of an enhanced

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 17 Nov 2026 at 17:08:57, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263126101879

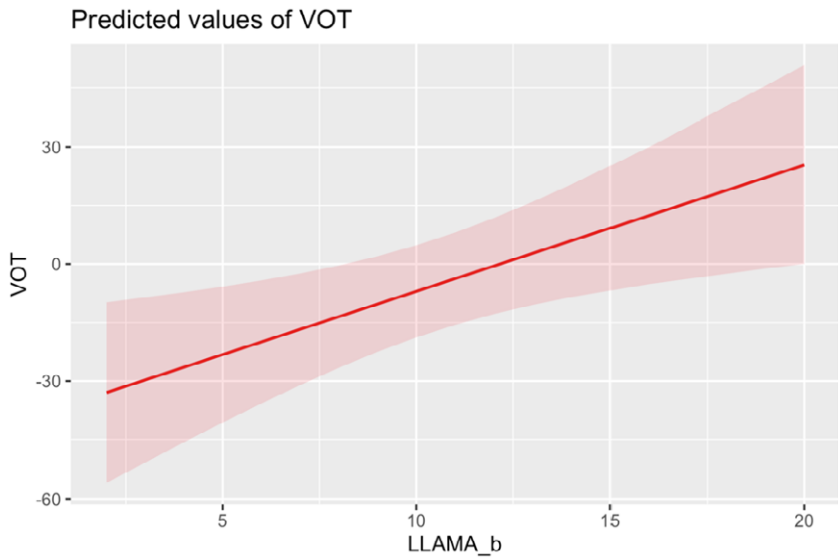


Figure 7. Predicted voice onset time (VOT) as a function of LLAMA B scores based on the linear mixed-effects model. Shaded areas represent 95% confidence intervals.

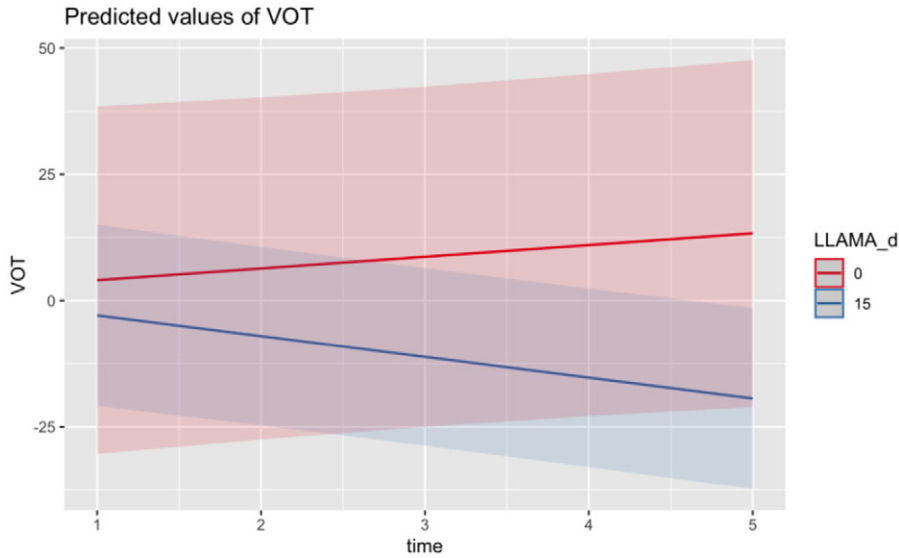


Figure 8. Predicted values from the linear mixed-effects model illustrating the interaction between Time and LLAMA D. Lines represent predicted trajectories for learners with low (−1 *SD*), mean, and high (+1 *SD*) LLAMA D scores. Shaded areas indicate 95% confidence intervals.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 17 Nov 2026 at 17:08:57, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263126101879

perceptual distance because VOTs in aspirated stops are longer in Mandarin than in English (Chen, 2007). Another factor is the orthographic system. Although in word-initial positions, Spanish /p, t, k/ are perceptually more similar to English /b, d, g/, orthographic representations may lead English speakers to link the unaspirated Spanish "p" to aspirated English "p", while this scenario is less likely to occur with L1 Mandarin learners given that Chinese's writing system is logographic. Additionally, explicit instructions about phonetic systems widely present in language curricula in China may contribute to higher awareness of the long-short VOT distinction (Zhang et al., 2023).

The vd group path shows a decreasing trend but the mean values are far from the target values (−60 to −100 ms). At the same time, progress seems to be mostly driven by a few participants who developed prevoicing when the majority did not. Compared to the unaspirated-aspirated distinction that exists in Mandarin, the prevoiced-unaspirated distinction in Spanish is less salient, making the Spanish voiced and voiceless VOT contrast more difficult to discern for L1 Mandarin learners. Previous studies on English-speaking learners of Spanish also found the development of VOT in voiced stops to be delayed. Learners are more successful in adjusting phonetic details (i.e., reducing long VOTs in the voiceless stops) than creating a new negative VOT category (Nagle et al., 2016; Zampini, 1998).

At the individual level, there was significant variability in trajectories over time. Individual plots in Figure 5 revealed three distinct patterns according to trends of change and categorical assimilation. We hypothesize three scenarios to account for these patterns:

1. Flat line, i.e., short positive VOT values in both vl and vd stops with no development over time. This pattern is characteristic of 50% of the participants across proficiency levels. We hypothesize that they map both voiced and voiceless stops onto the unaspirated stop category in Mandarin, consistent with single category or category goodness assimilation from the PAM-L2 model (Best & Tyler, 2007). This assimilation in perception blocks the formation of a negative VOT category in their mental representation and causes equivalent assimilation in production (SLM; Flege, 1995). As noted earlier, this pattern spans across experience levels; for example, Participant 14, despite being an advanced 3rd-year student with high overall Spanish proficiency, demonstrated complete neutralization. This suggests that without the necessary perceptual shifts or aptitude-driven interventions, early-stage category assimilation can fossilize regardless of accumulated classroom experience. Figure 9 illustrates two examples of this pattern from Participants 14 and 32.

2. "W" shape development, i.e., vd VOTs demonstrate great destabilization in the negative VOT domain and vl VOTs tend to partially overlap with vd VOT showing similar prevoicing (Participants 2, 3, 16, 18, 19). We hypothesize that two categories, a vd and a vl stop category, were formed in their mental representation, but their level of motor articulatory control was not advanced enough to consistently differentiate vd and vl stops with prevoicing and short-lag VOT, respectively, causing them to produce comparable prevoicing in both categories. Articulatory control can improve with time, as observed in Participant 16, whose vl and vd VOT gradually drove apart from each other from Time 2 to Time 5; the vl VOT increased to reach positive values, while the vd VOT remained negative. This pattern illustrates the dynamic process of category formation. Examining individual profiles reveals that experience plays varied roles here. For instance, Participant 2, an *ab initio* learner,

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 17 Nov 2026 at 17:08:57, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263126101879

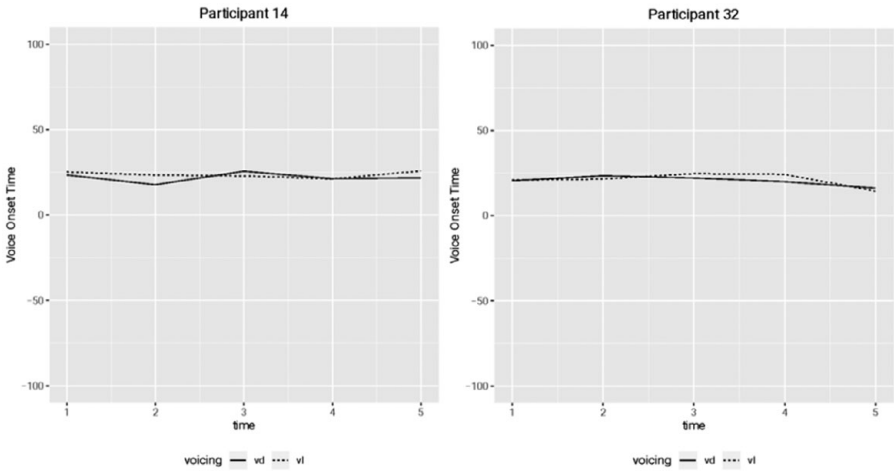


Figure 9. Examples of flat-line pattern from Participants 14 (left) and 32 (right).

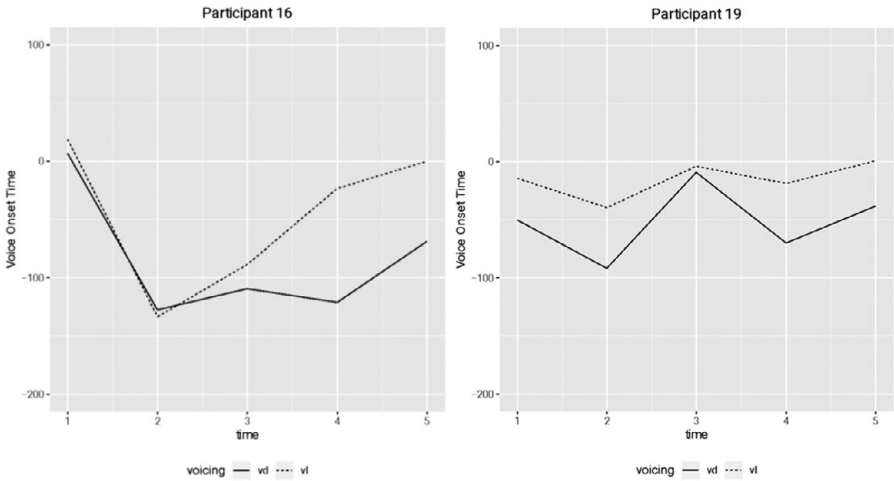


Figure 10. Examples of "W"-shape pattern from Participants 16 (left) and 19 (right).

began exhibiting this destabilized prevoicing as early as her second month of learning, capturing the genesis of the phonetic category. Conversely, Participant 18's presence in this group may be influenced by prior linguistic experience, as she reported minimal but existent childhood exposure to the Wu dialect, which preserves voicing contrasts. Figure 10 illustrates two examples of this pattern from Participants 16 and 19.

3. Two-category development, i.e., vd and vl VOT lines do not overlap and follow distinct trajectories (Participants 7, 8, 10, 12, 17). These participants appear to be able to clearly differentiate vd and vl VOTs producing target-like prevoicing for vd

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 17 Nov 2026 at 17:08:57, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263126101879

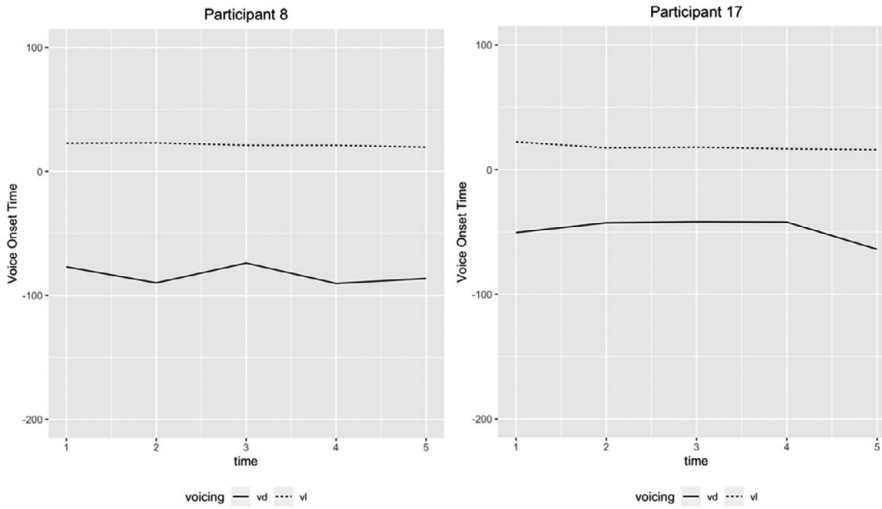


Figure 11. Examples of two category development from Participants 8 (left) and 17 (right).

stops and short-lag VOTs for vl stops. It is likely that in perception, they assimilate the vl VOTs to the unaspirated stops in Mandarin, but could recognize a difference in the vd stops and created a new negative VOT category that is specific to Spanish. This new category coupled with strong articulatory control helps them transduce acoustic information in the input into articulatory gestures in production. Auditory processing ability—a set of perceptual, cognitive, and motoric skills including the perception of fine-grained acoustic detail, the selective attention to relevant dimensions, and the integration of auditory input into motor action (e.g., Saito & Tierney, 2024; Saito et al., 2024)—may offer a domain-general explanation for the aptitude-related differences underlying two-category versus W-shaped developmental trajectories. When examining the role of experience, it is interesting to note that this successful pattern includes learners with immersive study abroad experience. Participant 17, who consistently produced target-like prevoicing from Time 1, had previously spent ten months studying in Spain. Similarly, Participant 10 was studying abroad in Spain for the duration of the data collection. This strongly indicates that for some learners, the negative VOT category was indeed developed through prior, intensive immersive experience. Figure 11 illustrates two examples of this pattern from Participants 8 and 17.

RQ2. *To what extent do individual differences in language aptitude predict Mandarin-English bilingual learners' production of prevoicing?*

RQ2a. *What language aptitude indicator(s) predicts the accuracy in prevoicing production?*

To examine the roles of language aptitude components on the accuracy of prevoicing, mixed effect models were applied with voiced VOT as continuous dependent variables. Results suggest that PCA emerged as a predictor for lower VOT values and more target-like prevoicing. In contrast, AM was linked to higher vd VOT values and

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 17 Nov 2026 at 17:08:57, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263126101879

less target-like prevoicing. Additionally, SRA—an implicit component of language aptitude— mediated prevoicing development: higher SRA predicted more rapid growth in the production of prevoicing.

PCA is a component of language aptitude that is linked to analyzing incoming input and that maps sounds to corresponding symbols (Skehan, 2016). Previous studies have found PCA to correlate with foreign-accentedness (Granena & Long, 2013; Hu et al., 2013), and is significantly associated with segmental measure of pronunciation rating (Saito, 2017, 2019). In a classroom learning setting, stronger phonetic coding ability can help learners to better notice and discern phonetic features and differences. Thus, those with stronger PCA tend to be more likely to notice prevoicing as characterized by a period of voicing during the consonant closure, and map this feature to the vd stop category in Spanish. While the relationship between accurate perception and accurate production remains a subject of ongoing discussion (see Flege & Bohn, 2021; Nagle, 2018; Nagle & Baese-Berk, 2022), our results suggest that stronger ability in coding and analyzing phonetic features in the input facilitates more accurate production of VOT.

Our results also indicate AM to associate with higher VOT values and less prevoicing. AM is believed to be linked to the short-time store component of working memory (Li, 2016) and phonological short-term memory (Saito, 2019). Previous studies found that AM also contributes to the learning of easy syntactic structure (past progressive) in oral production (Yalçın & Spada, 2016) as well as to fluency, i.e., articulation rate in spontaneous speech, and complexity, i.e., clause-to-AS-unit ratio (Saito, 2017). That is to say, in spontaneous speech (i.e., picture naming task), a speaker with stronger rote and associative memory may be able to store and retrieve linguistic information, i.e., vocabulary and syntactic structure, at a higher rate, which permits more efficient processing of input and more fluent and complex oral production. However, in the context of L3 phonological acquisition, this robust storage capacity may actually entrench non-target phonetic categories from previously learned languages. According to PAM-L2, L1 Mandarin learners typically assimilate novel Spanish voiced stops to their native unaspirated stops, which are characterized by short-lag positive VOT. We hypothesize that learners with higher AM may more rapidly stabilize and routinize this non-target sound-to-category mapping. Thus, the positive main effect of LLAMA B—where stronger memory predicts higher, positive VOTs—may reflect the successful, albeit inaccurate, establishment of an L1-transferred phonetic category. Nevertheless, the current dataset limits our ability to definitively test this mechanism. Future research incorporating perception tasks is necessary to confirm whether high associative memory directly correlates with the stabilization of non-target interlanguage representations during both initial and later learning stages. However, it appears that the development of production of a specific segmental feature such as prevoicing in the voiced Spanish stops seems to rely not only on storage functions, but also specific phonological analytical skills that may allow learners to capture the phonetic feature in the input, and further match it to the corresponding phonemes. Such ability seems to be more in line with the PCA measured by LLAMA E.

In interpreting the negative association of AM to prevoicing, it is also imperative to consider the nature of the task and target items in this study. During the word reading task, participants focused on producing the sounds corresponding to the graphic letters they saw on the screen. While this production process would involve PCA, i.e., processing graphic input and matching the forms to their corresponding sounds, AM may be less relevant since little information is being stored or retrieved in and out of the short-term memory system. With these considerations, we cautiously interpret the observed association as potentially reflecting task-related limitations on the role of AM in this specific production context.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 17 Nov 2026 at 17:08:57, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263126101879

RQ2b. ***What language aptitude indicator(s) predicts the rate of development in prevoicing production?***

Skehan (2016) proposes that in the first half of the acquisition process, input processing, noticing, and pattern identification are likely linked to working memory capacity and PCA. A longer working memory span may allow learners to retain more input for phonological analysis, while PCA supports learners in identifying phonetic features—such as prevoicing—and mapping them to appropriate phonological categories. In contrast, the later stages of acquisition, including pattern extension, automatization, and lexicalization, are thought to engage implicit learning abilities, such as SRA, though AM may also contribute to automatization in speech development (Saito, 2017).

Although this theoretical framework implies a growing influence of SRA and AM with increasing proficiency, the present findings did not reveal a statistically significant interaction between these aptitude components and proficiency level. One possible explanation is that the acquisition of prevoicing among L1 Mandarin speakers may be constrained by fossilization processes in segmental development. Due to perceptual assimilation and/or articulatory limitations, some learners may be less likely to revise their phonetic categories or articulatory routines even as their general L2 proficiency increases. This interpretation is supported by the individual learner trajectories (see Figure 5), which suggest variability in prevoicing development despite overall group-level improvement. That is, while mean VOT values decreased with proficiency, individual-level data show that increased oral proficiency does not uniformly predict segmental refinement such as VOT.

In this context, the global effect of PCA appears to play an important role in predicting overall continued development, which may help explain why certain learners successfully avoid plateauing in prevoicing production. Extrapolating from these overarching statistical trends, we hypothesize that learners with stronger phonetic coding skills may be better equipped to analyze fine-grained acoustic input and establish or maintain the distinct phonetic categories observed in the two-category developmental pattern.

At the same time, across the entire sample, participants with higher SRA scores demonstrated more rapid progress towards target-like VOT values over time, as indicated by a steeper decline in their overall voiced VOT trajectories. While our current statistical model evaluates the group as a whole, this overarching trend suggests a compelling hypothesis for the individual trajectories identified in RQ1: for learners who were not blocked by perceptual assimilation and who already showed some evidence of prevoicing (e.g., the "W-shape" or "two-category" patterns), stronger SRA may have supported more efficient proceduralization and refinement of articulatory patterns. This finding aligns with results from a pioneer longitudinal study by Saito et al. (2019b), which found that while in the earlier instructional phases, explicit aptitude (PCA and AM) helped students enhance their fluency and prosody, it was incidental aptitude (SRA measured by LLAMA D) that significantly predicted the extent to which learners continued to improve their segmental accuracy and ultimately attain advanced-level comprehensibility.

Importantly, "stage of learning" in the present study can be understood along two dimensions: macro-level proficiency differences across learners and micro-level developmental change over time. On the one hand, the sample included learners at different proficiency and placement levels, representing both earlier and later stages of acquisition. On the other hand, all participants continued to accumulate language

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 17 Nov 2026 at 17:08:57, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263126101879

exposure over the 5-month study period. The significant SRA × Time interaction, together with the absence of an SRA × proficiency interaction found in this study, suggests that the role of SRA became increasingly important with continued language exposure regardless of learners' initial proficiency levels. This finding appears to support the micro-level interpretation of Skehan's (2016) acquisition-aptitude model.

It is worth noting that previous studies (e.g., Saito, 2017; Saito, 2019) did not find any link between implicit aptitude and phonological development, probably due to their cross-sectional design. Our findings tentatively support the hypothesis that the influence of SRA is likely to be developmentally mediated—that is, its impact may be more observable with increased language experience. This is in line with findings from previous studies on morphosyntactic acquisition, where implicit language aptitudes have been found to exert greater influence at later stages of acquisition (e.g., Abrahamsson & Hyltenstam, 2008; Granena, 2013, 2020; Serafini & Sanz, 2016).

Nonetheless, these findings should be treated with caution, considering that the reliability of LLAMA v.3, although improved from previous versions, is still subject to further examination (Bokander et al., 2023), and additional longitudinal studies are needed to clarify the developmental trajectory of aptitude effects on L2 segmental learning. It is also important to note that although the LMM model controls other predictors when analyzing effects of the aptitude indicators, an array of other confounding factors might be at play to shape multilingual learners' phonetic acquisition, including previous study abroad experience in Spanish-speaking countries and exposure to Chinese dialects that preserve voicing contrasts. For instance, Participant 18's experience with the voicing Wu dialect from childhood and family, though reported to be minimal, might exert an influence on her development of prevoicing during the study.

Conclusion and future directions

The study explored the shape of 5-month developmental trajectories in L3 Spanish VOT among Mandarin-English bilinguals, with results that suggest learners generally improve on VOT in parallel with proficiency, with high levels of individual variation driven by individual difference factors. Specifically, it was clear that individual plots could be grouped according to common patterns showing how learners at different stages can share neutralization, destabilization, or development in their VOT production. Part of this variation can be explained by the facilitative role of phonetic coding ability in the development of prevoicing production, as well as the role of sequence recognition ability that facilitates a more rapid development of prevoicing as the learner's experience and exposure increase, particularly once a certain perceptual or articulatory threshold has been crossed.

Of note is the sizable proportion of participants who showed little or no change in VOT over the 5-month period. Rather than indicating a methodological limitation, this plateau may reflect meaningful learner-internal constraints—such as perceptual assimilation patterns or articulatory routinization—that merit further investigation. The findings highlight the need to better understand the conditions under which prevoicing development does or does not occur, and how this interacts with both aptitude and developmental stage.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 17 Nov 2026 at 17:08:57, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263126101879

Future research can benefit from following individual learners over a longer period to scrutinize their developmental trajectories more comprehensively, including the investigation of perception and the link between perception and production. It remains unclear whether perceptual assimilation results in fossilization in production, and the locus of influence of the language aptitude component discussed—whether it is perception, production, or both. Replication studies in immersive contexts, such as study abroad programs, could also provide valuable insights into how language aptitude interacts with contextual factors, thereby offering a more holistic understanding of the effects of explicit and implicit language aptitude components on segmental development in distinct learning contexts. Future research should also consider assessing auditory processing abilities—such as sensitivity to pitch, rhythm, and temporal structure in non-linguistic sounds—as these domain-general skills may help explain individual differences in the perception–production link and phonetic category development (Saito & Tierney, 2024).

In spite of its limitations—sample size and uneven sample distribution across proficiency levels, which may have constrained the scope of the proficiency-related analyses—as one of the first time-series studies in L3 phonological development, the study unveiled individual development paths and contributed to our understanding of how these paths are influenced by language aptitude components. Crucially, while three distinct developmental trajectories were identified qualitatively, the sample size constraints of the current study precluded the ability to run separate inferential statistical models for each subgroup. Future research with larger cohorts should investigate whether specific aptitude components interact differently depending on the learner's distinct developmental trajectory.

Author Note. **Linxi Zhang** Georgetown University, Washington, DC, USA.
Alfonso Morales-Front Georgetown University, Washington, DC, USA.
Cristina Sanz Georgetown University, Washington, DC, USA.

Declaration of Competing interest. The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Abrahamsson, N., & Hyltenstam, K. (2008). The robustness of aptitude effects in near-native second language acquisition. *Studies in Second Language Acquisition*, *30*(4), 481–509. https://doi.org/10.1017/S027226310808073X

Artieda, G., & Muñoz, C. (2016). The LLAMA tests and the underlying structure of language aptitude at two levels of foreign language proficiency. *Learning and Individual Differences*, *50*, 42–48. https://doi.org/10.1016/j.lindif.2016.06.023

Bates, D., Mächler, M., Bolker. B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Best, C. T. (1994). The emergence of native-language phonological influences in infants: A perceptual assimilation model. *The Development of Speech Perception: The Transition from Speech Sounds to Spoken Words*, *167*(224), 233–277.

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

Boersma, P. & Weenink, D. (2018): Praat: doing phonetics by computer [Computer program]. Version 6.0.37, retrieved 14 March 2018 from http://www.praat.org/

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 17 Nov 2026 at 17:08:57, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263126101879

Bokander, L. (2023). Exploring the predictive validity of the LLAMA (v1) language aptitude tests: A research synthesis. In Z. (E.) Wen, P. Skehan, & R. L. Sparks (Eds.), *Language aptitude theory and practice* (pp. 94–116). Cambridge University Press. https://doi.org/10.1017/9781009076463.007

Bokander, L., & Bylund, E. (2020). Probing the internal validity of the LLAMA language aptitude tests. *Language Learning*, *70*(1), 11–47. https://doi.org/10.1111/lang.12368

Bokander, L., Rogers, V., Meara, P., & Rogers, B. (2023). Internal validity of the new LLAMA (v. 3) aptitude tests. In *EuroSLA 32*. 30 August–2 September 2023, University of Birmingham.

Bravo-Díaz, C. (2020, December). The Acquisition of Stop Consonants by Chinese Learners of Spanish: A Cross-sectional Study. Paper presented at the *Hispanic and Lusophone Symposium*.

Carroll, J. B. (1962). The prediction of success in intensive foreign language training. In R. Glaser (Ed.), *Training and research in education* (pp. 87–136). University of Pittsburgh Press.

Carroll, J. B. (1981). Twenty-five years of research on foreign language aptitude. *Individual Differences and Universals in Language Learning Aptitude*, *83*, 113.

Carroll, J. B., & Sapon, S. M. (1959). *Modern language aptitude test*. Psychological Corporation.

Carroll, J. B. & Sapon, S. (2002). *Modern Language Aptitude Test: Manual 2002* Edition. N. Bethesda, MD: Language Learning and Testing Foundation.

Chen, Y. (2007). A Comparison of Spanish Produced by Chinese L2 Learners and Native Speakers---an Acoustic Phonetics Approach. Doctoral dissertation, University of Illinois at Urbana-Champaign. http://hdl.handle.net/2142/86153

DiCanio, C. (2011). VOT logger 1. [Praat script]. Retrieved from https://www.acsu.buffalo.edu/~cdicanio/scripts/Get_duration_2.0.praat

Dörnyei, Z., & Skehan, P. (2003). Individual differences in second language learning. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 589–630). Blackwell Publishing. https://doi.org/10.1002/9780470756492.ch18

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

Flege, J. E. (1991). Age of learning affects the authenticity of voice-onset time (VOT) in stop consonants produced in a second language. *Journal of the Acoustical Society of America*, *89*, 395–411. https://doi.org/10.1121/1.400473

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. *Speech Perception and Linguistic Experience: Issues in Cross-language Research*, *92*, 233–277.

Flege, J. E., & Eefting, W. (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*, 729–40. https://doi.org/10.1121/1.396115

Flege, J. E., Aoyama, K., & Bohn, O.-S. (2021). The revised speech learning model (SLM-r) applied. In R. Wayland (Ed.), *Second language speech learning: Theoretical and empirical progress* (pp. 84–118). Cambridge University Press. https://doi.org/10.1017/9781108886901.004

Flege, J. E., & Bohn, O.-S. (2021). The Revised Speech Learning Model (SLM-r). In R. Wayland (Ed.), *Second Language Speech Learning: Theoretical and Empirical Progress* (pp. 3–83). Cambridge: Cambridge University Press. https://doi.org/10.1017/9781108886901.002

Ge, C., Xiong, Y., & Mok, P. (2021). How Reliable Are Phonetic Data Collected Remotely? Comparison of Recording Devices and Environments on Acoustic Measurements. In *Interspeech* (pp. 3984–3988).

Granena, G. (2013). Cognitive aptitudes for second language learning and the LLAMA language aptitude test. *Sensitive Periods, Language Aptitude, and Ultimate L2 Attainment*, *35*, 105–129.

Granena, G. (2016). Cognitive aptitudes for implicit and explicit learning and information-processing styles: An individual differences study. *Applied psycholinguistics*, *37*(3), 577–600.

Granena, G. (2020). Implicit language aptitude. In *Elements in second language acquisition*. Cambridge University Press. https://doi.org/10.1017/9781108625616

Granena, G., & Long, M. H. (2013). Age of onset, length of residence, language aptitude, and ultimate L2 attainment in three linguistic domains. *Second Language Research*, *29*(3), 311–343. https://doi.org/10.1177/0267658312461497

Han, S. (2025). *Transferencia fónica y percepción auditiva de oclusivas en español: Análisis del efecto del dialecto wu en la adquisición del español por sinohablantes* [Undergraduate thesis, Universitat Autònoma

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 17 Nov 2026 at 17:08:57, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263126101879

de Barcelona]. Dipòsit Digital de Documents (DDD). https://ddd.uab.cat/pub/tfg/2024/tfg_8434891/TFG_Shuai_Han.pdf

Herd, W. (2020). Sociophonetic voice onset time variation in Mississippi English. *The Journal of the Acoustical Society of America*, *147*(1), 596–605. https://doi.org/10.1121/10.0000545

Hu, X., Ackermann, H., Martin, J. A., Erb, M., Winkler, S., & Reiterer, S. M. (2013). Language aptitude for pronunciation in advanced second language (L2) learners: Behavioural predictors and neural substrates. *Brain and Language*, *127*(3), 366–376. https://doi.org/10.1016/j.bandl.2012.11.006

Li, P., Zhang, F., Yu, A., & Zhao, X. (2020). Language History Questionnaire (LHQ3): An enhanced tool for assessing multilingual experience. *Bilingualism: Language and Cognition*, *23*(5), 938–944. https://doi.org/10.1017/S1366728918001153

Li, S. (2015). The associations between language aptitude and second language grammar acquisition: A meta-analytic review of five decades of research. *Applied Linguistics*, *36*(3), 385–408. https://doi.org/10.1093/applin/amu054

Li, S. (2016). The construct validity of language aptitude: A meta-analysis. *Studies in second language acquisition*, *38*(4), 801–842. https://doi.org/10.1017/S027226311500042X

Liao, S. J. (2005). *Interlanguage production of English stop consonants: A VOT analysis*. Unpublished master's thesis, National Kaohsiung Normal University, Kaohsiung, Taiwan.

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

Liu, Z. (2016). Exploring cross-linguistic influence: Perception and production of L1, L2 and L3 bilabial stops by Mandarin Chinese speakers. [Master's thesis, Universitat Autónoma de Barcelona]. https://ddd.uab.cat/record/167902

Liu, Z. (2019). *Análisis de las obstruyentes en chino y en español como L3: Estudio acústico y perceptivo para la categorización de errores*. [Doctoral thesis, Universitat Autònoma de Barcelona]. https://hdl.handle.net/10803/667728

Liu, Z., & M. J., Machuca-Ayuso. (2021). Effects of attention on the production of Spanish consonants by SFL learners. *Estudios de fonética experimental*, *30*, 209–224. https://revistes.ub.edu/index.php/experimentalphonetics/article/view/44007

Llama, R., & Cardoso, W. (2018). Revisiting (Non-) Native Influence in VOT Production: Insights from Advanced L3 Spanish. *Languages*, *3*(3), 30. https://doi.org/10.3390/languages3030030

Llama, R., Cardoso, W., & Collins, L. (2010). The influence of language distance and language status on the acquisition of L3 phonology. *International Journal of Multilingualism*, *7*(1), 39–57. https://doi.org/10.1080/14790710902972255

Meara, P. (2005). *LLAMA language aptitude tests: The manual*. Swansea: Lognostics.

Meara, P. M., & Rogers, V. E. (2019). *The LLAMA Tests v3. LLAMA_d & LLAMA_e*, v3. Cardiff: Lognostics.

Nagle, C. (2017). Individual developmental trajectories in the L2 acquisition of Spanish spirantization. *Journal of Second Language Pronunciation*, *3*(2), 218–241. https://doi.org/10.1075/jslp.3.2.03nag

Nagle, C. L. (2018). Examining the temporal structure of the perception–production link in second language acquisition: A longitudinal study. *Language Learning*, *68*(1), 234–270.

Nagle, C. (2019). A longitudinal study of voice onset time development in L2 Spanish stops. *Applied Linguistics*, *40*(1), 86–107. https://doi.org/10.1093/applin/amx011

Nagle, C. L. (2022). Individual difference factors for second language pronunciation. In S. Li, P. Hiver, & M. Papi (Eds.), *The Routledge handbook of second language acquisition and individual differences* (pp. 269–281). Routledge. https://doi.org/10.4324/9781003270546-23

Nagle, C. L., & Baese-Berk, M. M. (2022). Advancing the state of the art in L2 speech perception-production research: revisiting theoretical assumptions and methodological practices. *Studies in Second Language Acquisition*, *44*(2), 580–605. https://doi.org/10.1017/S0272263121000371

Nagle, C. L., Morales-Front, A., Moorman, C., & Sanz, C. (2016). Disentangling Research on Study Abroad and Pronunciation: Methodological and Programmatic Considerations. In D. M. Velliaris & D. Coleman-George (Eds.), *Handbook of research on study abroad programs and outbound mobility*. (pp. 673–695). IGI Global. https://doi.org/10.4018/978-1-5225-0169-5.ch027

Ortega, L., Iwashita, N., Norris, J. M., & Rabie, S. (2002, October). An investigation of elicited imitation tasks in crosslinguistic SLA research. In *Second Language Research Forum*, Toronto.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 17 Nov 2026 at 17:08:57, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263126101879

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., & Lindeløv, J. K. (2019). PsychoPy2: experiments in behavior made easy. *Behavior Research Methods.*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Plonsky, L., & Oswald, F. L. (2014). How big is "big"? Interpreting effect sizes in L2 research. *Language learning*, *64*(4), 878–912.

R Core Team. (2019). *R: a language and environment for statistical computing*, version 3.0. 2. Vienna, Austria: R Foundation for Statistical Computing, 2013.

Rochet, B. L., & Fei, Y. (1991). Effect of consonant and vowel context on Mandarin Chinese VOT: production and perception. *Canadian Acoustics*, *19*(4), 105–106. https://jcaa.caa-aca.ca/index.php/jcaa/article/view/691

Rogers, V., Meara, P., & Rogers, B. (2023). Testing language aptitude: LLAMA evolution and refinement. In Z. (E.) Wen, P. Skehan, & R. L. Sparks (Eds.), *Language aptitude theory and practice* (pp. 47–72). Cambridge University Press. https://doi.org/10.1017/9781009076463.005.

Rosner, B. S., López-Bascuas, L. E., García-Albea, J. E., & Fahey, R. P. (2000). Voice-onset times for Castilian Spanish initial stops. *Journal of Phonetics*, *28*(2), 217–224. https://doi.org/10.1006/jpho.2000.0113

Saito, K. (2017). Effects of sound, vocabulary, and grammar learning aptitude on adult second language speech attainment in foreign language classrooms. *Language Learning*, *67*(3), 665–693. https://doi.org/10.1111/lang.12244

Saito, K. (2019). The role of aptitude in second language segmental learning: The case of Japanese learners' English/ɹ/pronunciation attainment in classroom settings. *Applied Psycholinguistics*, *40*(1), 183–204. https://doi.org/10.1017/S0142716418000528

Saito, K., Sun, H., & Tierney, A. (2019a). Explicit and implicit aptitude effects on second language speech learning: Scrutinizing segmental and suprasegmental sensitivity and performance via behavioural and neurophysiological measures. *Bilingualism: Language and Cognition*, *22*(5), 1123–1140. https://doi.org/10.1017/S1366728918000895

Saito, K., Suzukida, Y., & Sun, H. (2019b). Aptitude, experience, and second language pronunciation proficiency development in classroom settings: A longitudinal study. *Studies in Second Language Acquisition*, *41*(1), 201–225. https://doi.org/10.1017/S0272263117000432

Saito, K., & Tierney, A. (2024). Domain-general auditory processing as a conceptual and measurement framework for second language speech learning aptitude: A test-retest reliability study. *Studies in Second Language Acquisition*, *46*(4), 1206–1230.

Saito, K., Kachlicka, M., Suzukida, Y., Mora-Plaza, I., Ruan, Y., & Tierney, A. (2024). Auditory processing as perceptual, cognitive, and motoric abilities underlying successful second language acquisition: Interaction model. *Journal of Experimental Psychology: Human Perception and Performance*, *50*(1), 119–138. https://doi.org/10.1037/xhp0001166

Sasaki, M. (1996). *Second language proficiency, foreign language aptitude, and intelligence: Quantitative and qualitative analyses* (Vol. *6*). Peter Lang Pub Incorporated.

Serafini, E. J., & Sanz, C. (2016). Evidence for the decreasing impact of cognitive ability on second language development as proficiency increases. *Studies in Second Language Acquisition*, *38*(4), 607–646. https://doi.org/10.1017/S0272263115000327

Singer, J. D., & Willet, J. B. (2003). A framework for investigating change over time. *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*, *315*, 115–139.

Skehan, P. (2002). Theorising and updating aptitude. *Individual Differences and Instructed Language Learning*, *2*, 69–94.

Skehan, P. (2016). Foreign language aptitude, acquisitional sequences, and psycholinguistic processes. In G. Granena, D. Jackson, & Y. Yilmaz (Eds.), *Cognitive individual differences in L2 processing and acquisition* (pp. 15–38). Benjamins.

Tremblay, M. (2007, August). L2 influence on L3 pronunciation: Native-like VOT in the L3 Japanese of English-French bilinguals. Paper presented at the *L3 Phonology Satellite Workshop of ICPhS XVI*, Freiburg, Germany.

Wrembel, M. (2011). Cross-linguistic influence in third language acquisition of voice onset time. In L. Wai-Sum & E. Zee (Eds.), *Proceedings of the 17th International Congress of Phonetic Sciences ICPhS* (pp. 2157–2160). Hong Kong: City University of Hong Kong.

Wrembel, M. (2014). VOT patterns in the acquisition of third language phonology. *Concordia Papers in Applied Linguistics (COPAL)*, *5*(2014), 751–771.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 17 Nov 2026 at 17:08:57, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263126101879

Wrembel, M. (2015). Cross-linguistic influence in second vs. third language acquisition of phonology. In U. Gut, R. Fuchs, & E.-M. Wunder (Eds.), *Universal or diverse paths to English phonology* (pp. 41–70). Mouton De Gruyter. https://doi.org/10.1515/9783110346084-004

Wrembel, M., Marecka, M., & Kopečková, R. (2019). Extending perceptual assimilation model to L3 phonological acquisition. *International Journal of Multilingualism*, *16*(4), 513–533.

Wunder, E. M. (2010). Phonological cross-linguistic influence in third or additional language acquisition. In K. Dziubalska-Kołaczyk, M. Wrembel, & M. Kul (Eds.), *New sounds 2010: Proceedings of the 6th international symposium on the acquisition of second language speech* (pp. 566–571). Poznań: Adam Mickiewicz University.

Yalçın, Ş., & Spada, N. (2016). Language aptitude and grammatical difficulty: An EFL classroom-based study. *Studies in Second Language Acquisition*, *38*(2), 239–263. https://doi.org/10.1017/S0272263115000509

Yavaş, M., & Wildermuth, R. (2006). The effects of place of articulation and vowel height in the acquisition of English aspirated stops by Spanish speakers. *International Review of Applied Linguistics in Language Teaching*, *44*(3), 251–263. https://doi.org/10.1515/IRAL.2006.011

Yilmaz, Y., & Granena, G. (2016). The role of cognitive aptitudes for explicit language learning in the relative effects of explicit and implicit feedback. *Bilingualism: Language and Cognition*, *19*(1), 147–161. https://doi.org/10.1017/S136672891400090X

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

Zhai, P. (2022). *El aprendizaje de las oclusivas del español por alumnos sinohablantes de distintos perfiles dialectales: Estudio acústico y perceptivo* (Doctoral dissertation, Universitat Autònoma de Barcelona). Dipòsit Digital de Documents (DDD). https://ddd.uab.cat/record/275024

Zhang, L. (2022). ¿Pollo o Bollo? Acquisition of voicing contrast by Chinese learners of L3 Spanish. *Hispania* *105*(4), 595–613. https://doi.org/10.1353/hpn.2022.0118

Zhang, L. (2026). *VOT measure materials*. Appendices from "Language aptitudes and VOT development of L3 Spanish stops in Chinese-English bilinguals." IRIS Database, University of York, UK. https://doi.org/10.48316/F6ENK-1CSCA

Zhang, L., Morales-Front, A., & Sanz, C. (2023). Dynamics in cross-linguistic influences - evidence from VOT development of Mandarin-English bilingual learners of L3 Spanish. In M. Brown, E. Fernández-Berke, & S. Flynn (Eds.), *L3 development after the initial state* (pp.121–149). John Benjamins. https://doi.org/10.1075/sibil.65.06zha

Cite this article: Zhang, L., Morales-Front, A., & Sanz, C. (2026). Language aptitudes and VOT development of L3 Spanish stops in Chinese-English bilinguals. *Studies in Second Language Acquisition,* 1–28. https://doi.org/10.1017/S0272263126101879

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 17 Nov 2026 at 17:08:57, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263126101879