EFFECTS OF STRESS AND PROMINENCE ON SPANISH STOPS AND LENITION IN L2 SPEECH OF PROFICIENT MANDARIN LEARNERS OF SPANISH

Xiaotong Xi[1], Peng Li[2]

[1]Department of Translation and Language Sciences, Univeristat Pompeu Fabra, Spain
[2]Center for Multilingualism in Society across the Lifespan (MultiLing), University of Oslo, Norway
[1]xiaotong.xi@upf.edu, [2]peng.li@iln.uio.no

ABSTRACT

This study assessed the effects of lexical stress and sentence prominence on Chinese students' phonetic realization of Spanish stops by measuring the Voice Onset Time (VOT) and the allophonic variants of voiced stops. Sixteen Chinese students and twelve Spanish natives participated in reading Spanish sentences containing /b, d, g, p, t, k/ in different utterance positions (initial vs. non-initial), lexical stress (stressed vs. unstressed) and sentence prominence (prominent vs. non-prominent). Results showed that (a) lexical stress showed no effect on learners' VOT; (b) learners could produce longer prevoicing of voiced stops in prominent positions; (c) stress and prominence did not show clear effects on learners' allophonic realization of voiced stops due to the limited cases of lenition. These results suggest that prosodic structure could potentially aid in the production of challenging L2 sounds, and that teachers may benefit from incorporating speech prominence in L2 teaching practice for training voiced stops.

Keywords: stop, lenition, prominence, L2 Spanish, Chinese students

1. INTRODUCTION

When learning voicing languages like Spanish, aspiration language speakers, such as Mandarin speakers, tend to assimilate both voiced (with negative voice onset time [VOT]) and voiceless (short positive VOT) stops to unaspirated stops (short positive VOT), as predicted by SLM-r [1] and validated by empirical studies [2], [3]. Another challenge arises from consonant lenition [3], [4]. Spanish natives often lenite voiced plosives to approximants in intervocalic positions, within a word or across word boundaries [5], but L2 learners tend to retain the plosive manner of articulation [3].

Despite the challenges faced by L2 learners at the segmental level, prosodic structures can influence the phonetic details of segments [6]. For instance, the Hyperarticulation hypothesis posits that sounds in stressed positions are hyperarticulated [7]. Empirical studies have shown that prosodically strong positions can affect the articulatory features of stops, such as shortening the VOT of voiceless stops [8] and lengthening the prevoicing period of voiced stops [9], [10]. In the context of L2 speech acquisition, lexical stress has been found to favor the production of voiced stops by English learners of Spanish [11]. However, few studies have assessed the effects of stress at both the word (lexical stress) and sentence levels (sentence prominence) on L2 Spanish stops, especially in learners whose native language lacks lexical stress and voicing contrast, such as Mandarin. Thus, it is worth exploring whether Mandarin-speaking learners would produce more nativelike Spanish stops under lexical stress or sentence prominence. Furthermore, if prosodically strong positions can strengthen the articulation of plosives, it is plausible that prosodically weak positions may facilitate lenition. Hence, would Mandarin-speaking learners lenite Spanish stops at unstressed or non-prominent positions?

Based on the literature reviewed above, this study formulates two hypotheses:

H1: Mandarin speakers would produce more native-like Spanish voiced stops, reflected by longer prevoicing, in stressed and prominent positions compared to unstressed and non-prominent positions.

H2: Mandarin speakers would be more likely to exhibit lenition of Spanish voiced stops in unstressed and non-prominent positions compared to stressed and prominent positions.

Since short-lag /p, t, k/ are shared phonemes in both Spanish and Mandarin, no specific hypothesis was formulated, but we included them in the design.

2. METHOD

2.1. Participants

We recruited 16 Mandarin learners of Spanish (12 females, M_{age} = 26.9, *SD* = 3.0) and 12 Spanish natives (8 females, M_{age} = 23.5, *SD* = 3.7). All participants signed the consent form to allow us to process their data. The Chinese students were late adult learners who began learning Spanish at the age of 18.7 years (*SD* = 1.6) and had studied Spanish for

4.3 years (*SD* = 1.1). They had formal Spanish language instruction in Spanish-speaking countries for 1.9 years (*SD* = 1.7) and had resided in Spanish-speaking countries for 5.3 years (*SD* = 2.9). Thus, they are proficient Spanish learners with substantial exposure to Spanish through study and living abroad.

2.2. Materials and procedure

We created 17 four-sentence dialogues. The first and third sentences were prompt questions to induce an answer that would have a narrow focus on the subject or object. The answers (second and fourth sentences) were two identical SVO sentences, with the target words embedded in the subject and object positions. For each of the six phonemes /b, d, g, p, t, k/, we selected four target words. The target phonemes always occurred at the beginning of the target words, with half of them being stressed and the other half unstressed. This resulted in a 2 × 2 × 2 design for each phoneme, varying in prominence (prominent vs. non-prominent), stress (stressed vs. unstressed), and sentence position (initial vs. non-initial). See (1) for an example, with stressed syllables underlined, target phonemes in boldface, and focus in uppercase.

(1) *¿Quién come la piña?* 'Who eats the pineapple?'
***B**E<u>LÉN</u> come la **<u>p</u>**<u>i</u>ña.* 'Belen eats the pineapple.'
¿Belén come la pera? 'Belen eats the pear?'
*No, **B**e<u>lén</u> come la **<u>P</u>**<u>I</u>ÑA.* 'No, Belen eats the pineapple.'

The speech data were recorded in a soundproof booth using a Zoom H4n Pro recorder and a SHURE SM35 microphone, with a sampling rate/resolution of 44.1kHz/16bits. Participants read all the sentences, but only the target sentences were analyzed.

2.3. Data coding and analysis

We obtained a total of 1,344 tokens. The first author manually annotated the VOT of all the plosives at initial and voiceless stops at non-initial positions as well as the length of the target words using Praat [12]. To avoid the effects of speech rate on the VOT production [13], we calculated the ratio of VOT to word length. However, we will plot and report the raw VOT values in the "Results" as descriptive statistics.

To test hypothesis 1, we built two Linear Mixed-Effects models (LMM) using the *lme4* package [14] in R to examine whether the two groups of speakers exhibited differences in the realizations of VOT for voiced and voiceless stops, depending on lexical stress and sentence prominence. The dependent variables for the two LMMs are the VOT ratio of sentence-initial stops and the VOT ratio of non-initial voiceless stops. Random structures were by-subject and by-item intercepts. The fixed effects for the first LMM involved voicing (voiced and voiceless), prominence (prominent and not prominent), stress (stressed and unstressed), group (Chinese students and Spanish natives), and their interactions. The second LMM included prominence, stress, group, and their interactions as fixed effects.

(2) *lmer(vot_ratio ~ voicing * prominence * stress * group + (1 | subject) + (1 | item))*
(3) *lmer(vot_ratio ~ prominence * stress * group + (1 | subject) + (1 | item))*

For the sentence-medial voiced /b, d, g/, the first author labeled the realizations by assessing the waveforms and spectrogram. Specifically, a sound was annotated as (a) voiceless plosive if the release burst was present with the aspiration phase, (b) voiced plosive if a voicing bar preceded the release burst, (c) fricative if there was a chaotic mix of random frequencies, and (d) approximant if vowel-like formant patterns could be identified. To test hypothesis 2, we ran a multinomial logistic regression (MLR) using *nnet* package [15] to examine whether the two groups of speakers differed in the allophonic realizations depending on the lexical stress and sentence prominence. The dependent variable was the allophonic realization (voiceless plosive, voiced plosive, fricative, and approximant) and the fixed effects were prominence, stress, group, and their interactions.

Significance were tested with Type II Wald chi-squared tests using *Anova()* from *car* package [16], and the group comparisons were carried out using the *emmeans* package [17] adjusted with the Bonferroni.

3. RESULTS

3.1. Effects of stress and prominence on VOT

The LMM for the VOT ratio of sentence-initial stops showed significant 3-way interactions of Voicing × Stress × Group (χ^2 = 5.7, p = .017) and Voicing × Prominence × Group (χ^2 = 11.3, $p < .001$), suggesting that the way in which the two groups distinguished voiced and voiceless stops by VOT differed across prominence or stress conditions. The LMM for the non-initial voiceless stops only revealed a significant 2-way interaction of Prominence × Group (χ^2 = 4.9, p = .027), suggesting that the VOT ratio of voiceless stops differed across group and prominence. In what follows, we will report the post-hoc results of the significant interactions related to the effects of stress and prominence.

3.1.1. Effects of stress on VOT

Regarding sentence-initial stops (Figure 1), stress had a significant effect only on the VOT ratio of voiced stops produced by Spanish natives, with the stressed position showing longer prevoicing than the unstressed position (t = 2.5, p = .027). As for the comparisons between groups, Chinese students showed significantly shorter prevoicing of voiced stops than Spanish natives in both stressed (t = 6.0, p < .001) and unstressed (t = 4.1, p < .001) positions. Moreover, Chinese students produced significantly shorter VOT ratio for voiceless stops than Spanish natives in stressed positions (t = 2.7, p = .009), but not in unstressed positions (t = 1.5, p = .138). In addition, both groups exhibited clear VOT ratio contrasts between voiced and voiceless stops regardless of stress condition (all p < .05).

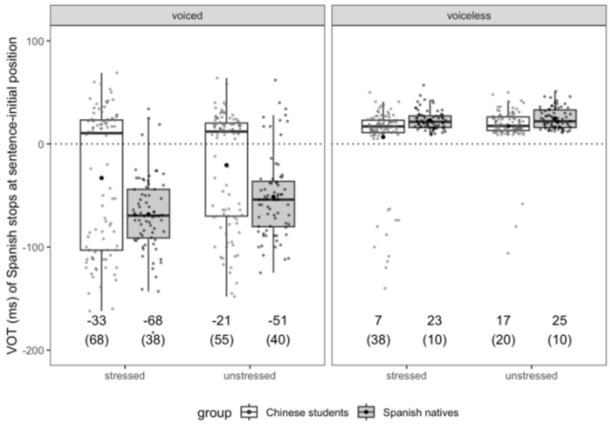

Fig. 1: Boxplots of VOT for the sentence-initial stops across group and stress conditions. Jittered points illustrate individual data points and means (SD) are labeled at the bottom.

In terms of non-initial voiceless stops (left panel in Figure 2), we did not find significant stress (χ^2 = 1.3, p = .262) or Stress × Group interaction (χ^2 = 0.1, p = .817). Both groups produced similar VOT ratio irrespective of the stress condition.

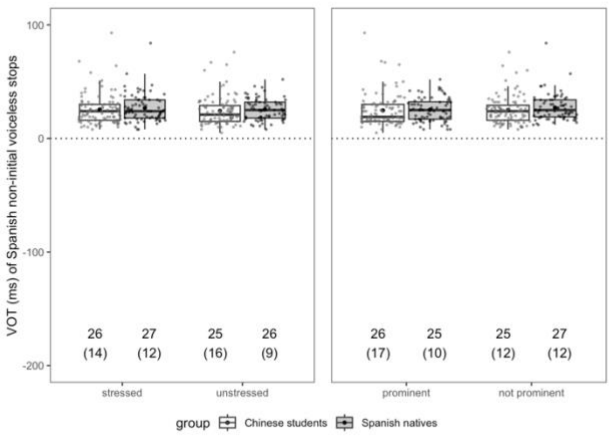

Fig. 2: Boxplots of VOT for non-initial voiceless stops across group and stress conditions (left panel) and across group and prominence conditions (right panel). Jittered points illustrate individual data points and means (SD) are labeled at the bottom.

Taken together, Chinese students' VOT was less affected by lexical stress and proficient learners still exhibited significant differences in VOT compared to Spanish natives.

3.1.2. Effects of prominence on VOT

At sentence-initial positions (Figure 3), Chinese students produced significantly longer prevoicing of voiced stops in prominent positions than in non-prominent positions (t = 5.4, p < .001). By contrast, Spanish natives did not show such a difference (t = 0.2, p = .874). As for voiceless stops, prominence did not show significant effects on VOT ratio in either group (all p > .05). In comparison with Spanish natives, Chinese students showed significantly shorter prevoicing than Spanish natives in both prominent (t = 3.4, p = .001) and non-prominent positions (t = 6.8, p < .001). Interestingly, for voiceless stops, Chinese students showed significantly shorter VOT ratio than Spanish natives in non-prominent positions (t = 2.6, p = .012) but not in prominent positions (t = 1.6, p = .114). Again, the voiced-voiceless contrast in VOT ratio was clear across group and prominence conditions (all p < .05).

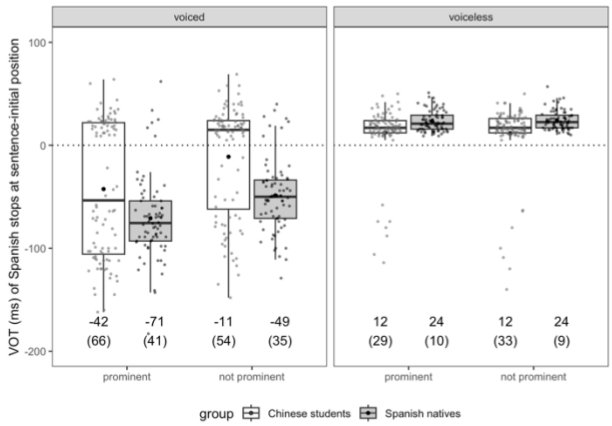

Fig. 3: Boxplots of VOT for sentence-initial stops across group and prominence. Jittered points illustrate individual data points and means (SD) are labeled at the bottom.

As for the non-initial voiceless stops (right panel in Figure 2), Spanish natives produced longer VOT ratio in prominent position (t = 4.2, p < .001), but this was not the case for Chinese students (t = 1.5, p = .131). In terms of group-level comparisons, similar to the initial voiceless stops, Chinese students showed significantly shorter VOT ratio than Spanish natives in non-prominent positions (t = 3.6, p = .001) but not in prominent positions (t = 1.5, p = .129).

3.2. Effects of stress and prominence on lenition

The MLR revealed two significant main effects. First, there was a significant main effect of prominence (χ^2 = 13.3, p = .004). Post-hoc pairwise comparisons showed that sentence non-initial voiced stops were realized more as approximants at non-prominent positions compared to prominent positions ($t = 3.7$, $p = .001$), regardless of group. Second, the main effect of group was significant ($\chi^2 = 178.0$, $p < .001$). Compared to Spanish natives, Chinese students produced a smaller portion of approximants ($t = 16.3$, $p < .001$) and a larger portion of voiced stops ($t = 3.5$, $p = .002$) and voiceless stops ($t = 12.9$, $p < .001$). No other significant main effects or interactions were found. Descriptive data can be found in Figure 4.

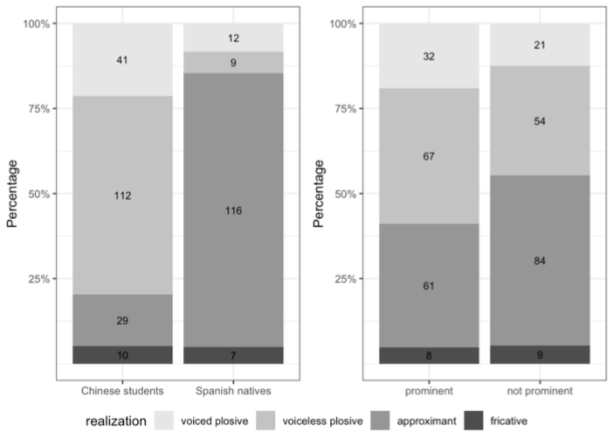

Fig. 4: Portion of allophonic realizations of the non-initial voiced stops divided by group (left panel) and prominence (right panel), with counts labeled on the bars.

4. DISCUSSION AND CONCLUSION

This study examined the impact of lexical stress and sentence prominence on Chinese students' VOT and allophonic realizations of Spanish stops using a dialogue-reading task. We found that (a) lexical stress did not show clear effects on Chinese students' prevoicing or lenition; (b) Chinese students produced longer prevoicing of sentence-initial voiced stops in prominent positions but prominence did not affect lenition as such; and (c) no clear effects of prominence or stress were observed for voiceless stops. Chinese students distinguished the voiced and voiceless stops by different VOT values, but they still showed differences from Spanish natives.

Our first hypothesis was partly confirmed that Mandarin speakers could produce more native-like L2 Spanish voiced stops in prominent positions. Firstly, unlike Spanish natives who produced longer prevoicing in stressed positions, Chinese students' production of voiced stops was not influenced by stress. This finding contrasts with [11] which showed that English speakers could increase the amount of Spanish prevoicing in stressed positions, similar to native speakers. However, it is important to note that our Chinese students had a much shorter length of residence compared to the participants in [11] (5.3 years vs. 36 years). Furthermore, English has lexical stress, and some English speakers indeed use prevoiced stops in their L1 [18], [19], but Mandarin shows little evidence for Spanish-like lexical stress or prevoicing. Therefore, the effects of stress on L2 speech may vary depending on learners' length of residence and L1 background. Secondly, Chinese students showed better pronunciation of the nonnative voiced stops under speech prominence. This finding adds to the growing body of evidence supporting the relationship between prosody and segment [8]–[10], and extends this research to L2 speech, where similar effects on Spanish vowels have been observed [20].

Notably, for the supposedly less challenging voiceless stops, we did not find significant effects of sentence prominence or lexical stress on Chinese students' VOT. In contrast, Spanish natives produced shortened VOT of voiceless stops in prominent positions, consistent with previous research [8]. Furthermore, Chinese students showed significantly shorter VOT of voiceless stops than Spanish natives in non-prominent positions. This finding suggests that although Chinese students were able to contrast voiced stops and voiceless stops by VOT values, they still showed some nonnative production patterns in the "less challenging" voiceless stops.

Our second hypothesis was not confirmed, as we did not observe more occurrence of lenition in L2 speech at unstressed and non-prominent positions compared to stressed or prominent positions. In general, more approximants were produced in non-prominent positions than in prominent positions. However, Chinese students showed a significantly smaller portion of lenition (15%) compared to Spanish natives (66%). Therefore, the effects of prominence on lenition were not evident in L2 speech. Moreover, previous research has reported a larger proportion of lenition within words by Chinese students [3], as opposed to our study where lenition occurred across word boundaries. This suggests that the lenition in L2 Spanish speech may be frequency-driven as approximants are more frequent within words than across word boundaries [5].

In conclusion, this study provides evidence that L2 learners can produce longer provoicing of Spanish voiced /b, d, g/ in sentence prominent positions, which highlights the prosody-segment interaction in L2. In practice, L2 teachers may manipulate the prosodic conditions to train the production of Spanish stops. Finally, further attention should be given to the lenition of voiced stops in L2 Spanish learning.

5. ACKNOWLEDGEMENTS

This study was supported by the Research Council of Norway through its Centres of Excellence funding scheme, project number 223265.

6. REFERENCES

[1] J. E. Flege and O. S. Bohn, "The Revised Speech Learning Model (SLM-r)," in *Second language speech learning: Theoretical and empirical progress*, R. Wayland, Ed. Cambridge University Press, 2021, pp. 84–118.

[2] Z. Liu and M. J. M. Ayuso, "Effects of attention on the production of Spanish consonants by SFL learners," *Estud. Fon. Exp.*, vol. 30, pp. 209–224, 2021.

[3] X. Xi and P. Li, "Spanish Stops And Their Allophones Produced By Proficient Mandarin Learners Of Spanish," in *Proceedings of the 25th Conference of the Oriental COCOSDA International Committee for the Co-ordination and Standardisation of Speech Databases and Assessment Techniques (O-COCOSDA)*, Nov. 2022, pp. 152–156, doi: 10.1109/O-COCOSDA202257103.2022.9997932.

[4] C. L. Nagle, "Individual developmental trajectories in the L2 acquisition of Spanish spirantization," *J. Second Lang. Pronunciation*, vol. 3, no. 2, pp. 218–241, 2017, doi: 10.1075/jslp.3.2.03nag.

[5] T. L. Face and M. R. Menke, "Acquisition of the Spanish voiced spirants by second language learners," in *Selected proceedings of the 11th Hispanic Linguistics Symposium*, 2009, pp. 39–52.

[6] D. Mücke and M. Grice, "The effect of focus marking on supralaryngeal articulation – Is it mediated by accentuation?," *J. Phon.*, vol. 44, pp. 47–61, May 2014, doi: 10.1016/j.wocn.2014.02.003.

[7] K. J. de Jong, "The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation," *J. Acoust. Soc. Am.*, vol. 97, no. 1, pp. 491–504, Jan. 1995, doi: 10.1121/1.412275.

[8] T. Cho and J. M. McQueen, "Prosodic influences on consonant production in Dutch: Effects of prosodic boundaries, phrasal accent and lexical stress," *J. Phon.*, vol. 33, no. 2, pp. 121–157, 2005, doi: 10.1016/j.wocn.2005.01.001.

[9] E. Wojtkowiak and G. Schwartz, "Prosody-segment interaction in the acoustics of Polish plosives," in *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia 2019*, 2019, pp. 3285–3289.

[10] M. Simonet, J. Casillas, and Y. Díaz, "The effects of stress/accent on VOT depend on language (English, Spanish), consonant (/d/, /t/) and linguistic experience (monolinguals, bilinguals)," in *Speech Prosody 2014*, May 2014, pp. 202–206, doi: 10.21437/SpeechProsody.2014-28.

[11] T. L. Face and M. R. Menke, "L2 Acquisition of Spanish VOT by English-Speaking Immigrants in Spain," *Stud. Hisp. Lusoph. Linguist.*, vol. 13, no. 2, pp. 361–389, Oct. 2020, doi: 10.1515/shll-2020-2034.

[12] P. Boersma and D. Weenink, "Praat: doing phonetics by computer." 2022.

[13] A. M. Schmidt and J. E. Flege, "Speaking rate effects on stops produced by spanish and english monolinguals and spanish/english bilinguals," *Phonetica*, vol. 53, no. 3, pp. 162–179, 1996, doi: 10.1159/000262196.

[14] D. Bates, M. Mächler, B. Bolker, and S. Walker, "Fitting Linear Mixed-Effects Models Using lme4," *J. Stat. Softw.*, vol. 67, no. 1, pp. 1–48, 2015, doi: 10.18637/jss.v067.i01.

[15] W. N. Venables and B. D. Ripley, *Modern applied statistics with S*. Springer, 2002.

[16] J. Fox and S. Weisberg, *An {R} Companion to Applied Regression*. SAGE, 2019.

[17] R. V Lenth, "emmeans: Estimated Marginal Means, aka Least-Squares Means." 2022.

[18] J. E. Flege, "Laryngeal timing and phonation onset in utterance-initial English stops," *J. Phon.*, vol. 10, no. 2, pp. 177–192, 1982, doi: 10.1016/s0095-4470(19)30956-8.

[19] L. Hunnicutt and P. A. Morris, "Prevoicing and Aspiration in Southern American English," *Univ. Pennsylvania Work. Pap. Linguist.*, vol. 22, no. 1, pp. 215–224, 2016.

[20] P. Li and X. Xi, "Spanish lexical stress produced by proficient Mandarin learners of Spanish," in *ISAPh 2022, 4th International Symposium on Applied Phonetics*, 2022, pp. 40–45, doi: 10.21437/ISAPh.2022-8.