Nagle, C. & Bruun, S. (2020). Examining predictors of phonetic variation in semi-spontaneous L2 Spanish Speech. In O. Kang, S. Staples, K. Yaw, & K. Hirschi (Eds.), *Proceedings of the 11th Pronunciation in Second Language Learning and Teaching conference*, ISSN 2380-9566, Northern Arizona University, September 2019 (pp. 163–172). Ames, IA: Iowa State University.

EXAMINING PREDICTORS OF PHONETIC VARIATION IN SEMI-SPONTANEOUS L2 SPANISH SPEECH

Charles Nagle, Iowa State University
Shelby Bruun, Iowa State University

Longitudinal research on second language (L2) sound learning demonstrates that speakers' production of challenging L2 sounds can improve in the absence of targeted instruction. Although the growing body of longitudinal work on this topic provides insight into the rate and shape of development, the factors that shape L2 phonetic production in spontaneous speech are not yet well understood. The data described here are part of a longitudinal data set collected from native English speakers enrolled in introductory-level Spanish language courses. For the current analysis, 16 participants were selected based on the availability of data coinciding with the beginning and end of the first semester and end of the second semester. Participants completed a simplified picture description task. Voice Onset Time (VOT) was annotated and measured for all instances of Spanish /p, t, k/ in the speech samples. Participants' instructors were recorded during two class periods. Recordings were transcribed and a frequency measure was calculated based on the resulting teacher speech data. After controlling for a range of linguistic factors known to affect VOT, modeling demonstrated that time and frequency were not significantly related to VOT.

INTRODUCTION

Current models of second language (L2) pronunciation development posit that phonetic learning remains possible for late-acquired languages. For instance, the Speech Learning Model (SLM; Flege, 1995) relates production accuracy to phonetic category formation. According to the SLM, if L2 speakers distinguish cross-linguistically similar native (L1) and L2 sounds, then they will create a new category for the L2, which should enable them to produce the L2 sound accurately. Conversely, if they equate L2 sounds with L1 categories, then category formation will be blocked, resulting in compromised phonetic productions that reflect both L1 and L2 phonetic norms. Although the SLM is arguably the most widely employed framework in L2 pronunciation learning, it is important to bear in mind that it was formulated to account for cross-linguistic interference in bilingual speakers. As a result, the focus of this model is on end states, or the extent to which L1 and L2 categories interact in highly proficient L2 users. This focus is certainly warranted, given the bilingual contexts in which many individuals live and work. At the same time, examining phonetic learning at its onset can shed light on precisely when and where changes first appear in the developing L2 system. Addressing this gap, the current study took a data-driven approach to understanding the factors that influence phonetic variation in L2 stop consonant production during the first two semesters of university-level foreign language learning.

A growing body of longitudinal research suggests that learners' pronunciation of L2 sounds can improve in the absence of targeted pronunciation training, a finding that applies to both second (e.g., Munro et al., 2015) and foreign language contexts (Casillas, 2016; Hanzawa, 2018; Nagle, 2019). Irrespective of differences in rate and shape of learning, the common finding that learners' pronunciation can improve given the right input conditions begs the question: What mechanisms underlie L2 phonetic learning? Whereas some scholars have suggested that L2 speakers learn abstract features (e.g., Olson, 2019), we were specifically interested in an exemplar- or frequency-based approach. According to this approach, words are stored in robust phonetic detail, resulting in exemplars that emerge as statistical regularities over sets of accumulated phonetic traces (see, e.g., Pierrehumbert, 2001, 2003). Full presentation of an exemplar-based approach is beyond the scope of this paper, but within such an approach, frequency is crucial, insofar as more frequent words should be produced more accurately since their exemplar cloud would be more densely populated. Though the majority of work in this area has concentrated on L1 speakers, at least one study suggests that L2 learners are sensitive to the frequency characteristics of the input. Trofimovich et al. (2012) profiled the frequency of English /ð/ in teacher talk, examining learners' production in relation to teachers' patterns. They found that the proportion of /ð/ tokens that students produced across phonetic contexts was aligned with the frequency distribution of /ð/ in teacher talk. These findings run counter to Flege et al. (1998), who did not find a significant relationship between a variety of lexical factors (e.g., frequency, familiarity, estimated age of acquisition, etc.) and Spanish speakers' production of English /t/. These contradictory results could be due to methodological factors, such as the way frequency was operationalized and differences in the populations sampled—classroom ESL learners in Trofimovich et al. (2012) versus native Spanish speakers who had lived in the United States between 5–13 years on average in Flege et al. (1998).

The Current Study

In this study we examined whether English-speaking learners of Spanish produced more accurate L2 voiceless stops over time. We were also interested in whether learners would produce more accurate stops in words that were more frequent in their teachers' speech.

We defined production accuracy in terms of Voice Onset Time (VOT), an acoustic measure that refers to the interval between the release burst of the stop and the onset of voicing in the following segment (Lisker & Abramson, 1964). In English, voiceless stops are produced with a long delay in voicing, whereas in Spanish, voicing begins shortly after stop release. VOT is highly variable depending on a variety of factors such as speech style, stop consonant point of articulation, and phonetic context. Canonical VOT values of 60–90 milliseconds (ms) are common for word-initial voiceless stops in English citation forms, whereas shorter values in the 0–30 ms range are typical of Spanish (Casteñada, 1986; Lisker & Abramson, 1964). However, in running speech, VOT may be considerably shorter, bringing English VOT values closer to the values observed for Spanish. For example, Lisker and Abramson (1964) reported VOT values of 28–43 ms and 4–25 ms for English and Spanish voiceless stops in sentences (see also, Lisker & Abramson, 1967). With respect to point of articulation, it has been well documented that VOT increases the further back the closure, with longer VOT for /k/ than /p/ and /t/ (Cho & Ladefoged, 1999). VOT of English stops has also been shown to vary according to prosodic factors (Kim et al., 2018). For instance, English but not Spanish voiceless stops are produced with longer VOT in stressed syllables (Simonet et al., 2014). A comprehensive account of contextual factors on VOT is beyond the scope of this study. For our purposes, given that we examined stops in spontaneous speech, we coded tokens for a range of phonetic features (lexical stress, following vowel height, etc.) so that we could account for them in our statistical models.

Research Questions and Predictions

1. Do English speakers produce more accurate L2 voiceless stops over their first two semesters of Spanish language coursework?

   Given accumulated research suggesting that L2 learners' pronunciation improves in the absence of targeted training, we predicted that learners' production of Spanish /p, t, k/ would improve, as evidenced by shorter VOT.

2. Does input frequency shape the accuracy with which L2 learners produce L2 voiceless stops? What other factors shape L2 phonetic production?

   Following Trofimovich et al. (2012), we hypothesized that learners would produce shorter, more Spanish-like VOT in words that were more frequent in teacher talk. We also predicted that longer VOT would be observed for /t, k/, for word-initial stops, and for stops occurring in stressed syllables.

METHOD

This study was part of a larger longitudinal project examining individual differences in the perception and production of L2 Spanish stops. As part of the larger project, we recruited 30 native English speakers from multiple sections of a first-semester university Spanish course taught by both native and non-native instructors, all of whom followed the same communicative curriculum. Participants took part in the study during their first two semesters of communicative Spanish language coursework. In total, there were seven data points, three in the fall and four in the spring, spaced at approximately one-month intervals.

At each session, participants completed a variety of perception, production, and individual difference measures. In this study, we report and analyze data elicited via a simplified picture description task. We included data from sessions 1, 3, and 7, which coincided with the beginning and end of the first semester and the end of the second semester course. These three data points allowed us to capture longitudinal changes in stop consonant production during students' first (session 1 to 3) and second (session 3 to 7) semesters.

Participants

Of the 30 participants included in the larger study, we selected a sample of 16 (3 males) who participated in all three of the target sessions. On average, participants had begun learning Spanish in high school (*M* age of onset in years = 13.31, *SD* = 3.70) and had taken just over one and a half years of Spanish (*M* = 1.63, *SD* = 1.99) prior to enrolling in the study.

We recorded two instructors, one native and one non-native, teaching two 50-minute classes, resulting in 200 minutes of spontaneous teacher talk. These recordings were made near the end of the first-semester course and were thus most representative of the content that participants were studying at session 3.

Materials and Procedure

At each session, participants completed a simplified picture description task (see Saito et al., 2017). Participants received a set of six images, including one practice image, and were given 20 seconds to examine each image before describing it in as much detail as possible. Participants were not allowed to take notes during the planning time.

To ensure that participants could complete the task, we selected visually similar images that novice L2 learners would be able to describe (e.g., a man fishing at the beach, two people preparing dinner, etc.), and we included up to three words or short phrases with each image as vocabulary aides. Following this procedure, we created seven image sets to allow for counterbalancing across participants and sessions.

A member of the research team met with each participant individually in a sound-treated room. Participants wore a Shure SM10A dynamic, head-mounted microphone that was connected to a desktop computer through an XRL to USB converter. All recordings were made in Audacity and exported as WAV files.

Data Processing and Coding

A member of the research team who was a native Spanish speaker created an orthographic transcription of the picture description files. The Montreal Forced Aligner (McAuliffe et al., 2017) was then used to generate rough Praat text grids (Boersma & Weenik, 2015) to be checked and corrected by hand by another member of the research team. Once the text grids had been verified, the second author annotated VOT for all voiceless stops as well as the duration of the target word in which the stop occurred. A Praat script was used to extract all measurements.

Due to the semi-spontaneous nature of the picture description task, it was not possible to control for the phonetic environment in which the target stops occurred. Consequently, following Stuart-Smith et al. (2015), we coded target words for the following features: place of articulation, lexical stress, following vowel height, and word duration.

Due to the variable number of words that participants produced over time, we included time-aligned data only if the participant produced at least three viable voiceless stop tokens at a given session. We employed this screening criterion to enhance the reliability of person-by-time estimates in our models. These screening procedures resulted in a final data set consisting of 468 observations of word-initial stops in stop-vowel sequences.

Instructor audio recordings were sent to TranscribeMe to generate preliminary orthographic transcriptions, which were verified by the second researcher. Transcriptions were transformed into word lists, resulting in a data set consisting of 12,765 word tokens. We opted for a lemma-based frequency measure, which would better reflect the contexts in which a stop occurred in a set of morphologically related forms. For example, the entry for the verb *tener* (‘to have’), the most frequent lemma in the teacher data set (lemma frequency = 227), included forms such as *tienes* (‘you have’), *tengo* (‘I have’), etc. The most frequent lemmas for word-initial /p/ and /k/ were *por* (‘for,’ ‘because’) and *con* (‘with’), with frequencies of 123 and 92, respectively. We derived the teacher talk frequency measure by looking up the lemma frequency of each of the student-generated words in the teacher talk data. If a word did not occur in the data set, the teacher-talk lemma frequency of that word was set at 0.1, to allow for log transformation of the frequency measure.

Approach to Statistical Analysis

We fit mixed-effects models in R version 3.6.1 (R Core Team, 2019) using the lme4 package (Bates et al., 2014). Models included fixed effects for session and the teacher-talk frequency measure, by-participant and by-word random intercepts, and the following covariates: (1) age of onset and years of previous Spanish study (both continuous) to control for relationships between participant background characteristics and VOT production; and (2) place of articulation (categorical, three levels: /p/, /t/, /k/), lexical stress (categorical, two levels: unstressed, stressed), following vowel height (categorical, two levels: nonhigh, high), and word duration (continuous) to control for the phonetic properties of the target word in which the stop occurred. Categorical variables containing two levels were contrast-coded (Linck & Cunnings, 2015), and continuous predictors were transformed into z-scores, such that positive numbers indicate values above the mean and negative numbers values below the mean. Bilabial /p/ was set as the baseline in the three-level place of articulation predictor; thus, model estimates for /t/ and /k/ refer to differences with /p/. We used QQ plots to check the assumption that model residuals were normally distributed. We tested competing random-effects structures by progressively integrating random effects and performing likelihood ratio tests on nested models.

RESULTS

Descriptive statistics are reported in Table 1. The median number of observations per cell was 25, and the range was 7–53. There were no observations of unstressed /t/ at session 1. Descriptive data indicate that participants produced shorter VOT in /p/ than in /t/ and /k/ and that VOT was longer and more variable in stressed syllables (see Figure 1).

Table 1

Means and Standard Deviations for VOT According to Phone, Lexical Stress, and Session

	/p/		/t/		/k/	
	U	S	U	S	U	S
Session 1	44 (24)	68 (36)	–	89 (42)	71 (32)	81 (47)
Session 3	54 (22)	82 (39)	80 (28)	84 (40)	63 (22)	93 (57)
Session 7	59 (30)	63 (43)	69 (22)	95 (31)	65 (21)	87 (47)

Note. U = Unstressed, S = Stressed.

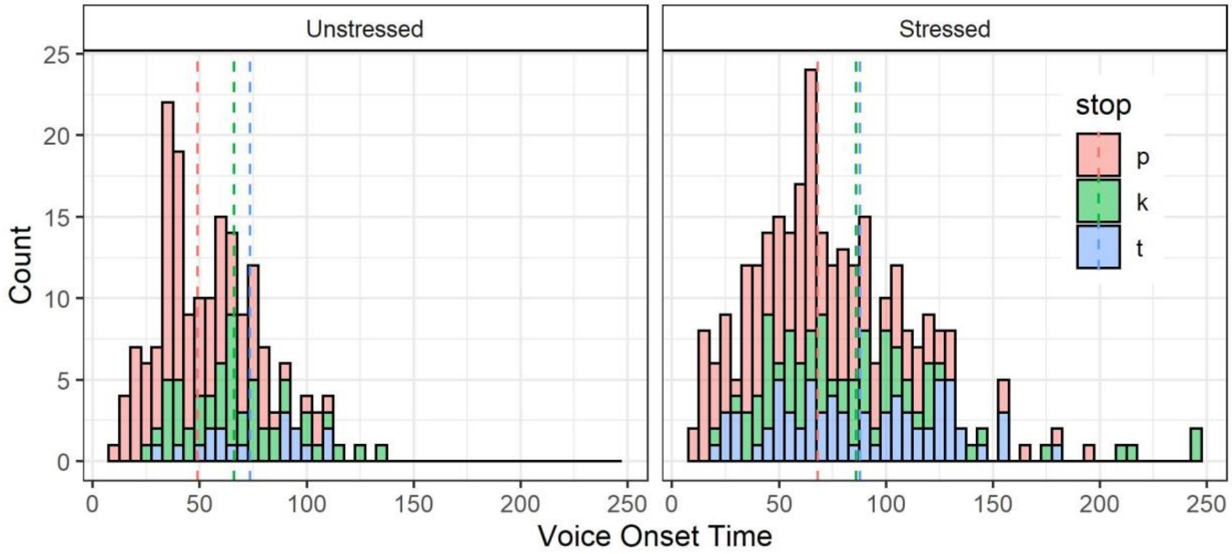

Figure 1. Distribution of VOT values by phone in unstressed and stressed syllable. Dashed lines indicate means.

Model Fit to All Word-Initial Stops

As summarized in Table 2, predictors for the session and teacher frequency measures missed significance, demonstrating that participants did not produce shorter VOT over time or in words that were more frequent in teacher talk. In line with previous research, many of the phonetic variables were highly significant, and effects were in the expected direction. VOT was longer in /t/ and /k/ than /p/, and longer VOT was observed in stops occurring in stressed syllables, in longer words, and before a high vowel.

Table 2

Model Summary: VOT in All Word-Initial Voiceless Stops

						Random effects	
	Fixed effects					Subject	Word
Parameter	Estimate	*SE*	95% CI	*t*	*p*	*SD*	*SD*
Intercept (/p/)	66.80	5.90	[55.24, 78.35]	11.33	< .001	20.55	10.31
Session	–1.81	2.50	[–6.72, 3.10]	–.72	.48	7.89	
Log(teacher)	.20	1.74	[4.78, 3.60]	.12	.91		
PoA: /t/	13.49	4.44	[4.78, 22.20]	3.04	.003		
PoA: /k/	16.71	3.61	[9.64, 23.79]	4.63	< .001		
Lexical Stress	17.31	6.20	[5.16, 29.47]	2.79	.01	20.12	
Vowel Height	12.71	3.89	[5.09, 20.33]	3.27	.001		
Word Duration	7.26	1.43	[4.46, 10.05]	5.09	< .001		
Age Onset	4.03	5.90	[–7.53, 15.59]	.68	.51		
Previous Exp.	–3.38	6.62	[–16.35, 9.60]	–.51	.62		

Note. Log(teacher) = The log-transformed word frequency in teacher talk; PoA = Place of Articulation; Duration = Duration of the host word; Age Onset = Age at which participants began learning Spanish; Previous Exp. = Number of years of Spanish courses participants had taken prior to enrolling in the study. Stress and Vowel Height were contrast-coded. Log(teacher), Duration, Age Onset, and Previous Exp. Were scaled and centered.

Model Fit to Word-Initial Stops in Stressed Syllables

We hypothesized that development might be most evident in word-initial stops in stressed syllables since in that context we would expect English and Spanish voiceless stops to show the greatest differences in VOT. We therefore fit a follow-up model, summarized in Table 3, to a subset of the data containing the 291 observations of word-initial stressed stops. The follow-up model confirmed initial findings. However, it is worth noting that the confidence interval for the session predictor shrank and shifted toward negative values, suggesting a downward trend in VOT despite the null finding.

Table 3
Model Summary: VOT in Word-Initial Voiceless Stops in Stressed Syllables

						Random effects	
	Fixed effects					Subject	Word
Parameter	Estimate	*SE*	95% CI	*t*	*p*	*SD*	*SD*
Intercept (/p/)	78.44	8.51	[61.77, 95.12]	9.22	< .001	30.06	10.52
Session	–4.61	3.24	[–10.96, 1.75]	–1.42	.18	9.60	
Log(teacher)	1.63	2.62	[–3.51, 6.76]	.62	.54		
PoA: /t/	6.77	6.23	[–5.43, 18.97]	1.09	.28		
PoA: /k/	16.92	5.60	[5.95, 27.88]	3.02	.003		
Vowel Height	15.39	5.27	[5.06, 25.72]	2.92	.004		
Word Duration	9.12	2.00	[5.20, 13.03]	4.57	< .001		
Age Onset	5.09	8.59	[–11.74, 21.92]	.59	.56		
Previous Exp.	–4.32	9.67	[–23.27, 14.62]	–.45	.66		

Note. Log(teacher) = The log-transformed word frequency in teacher talk; PoA = Place of Articulation; Duration = Duration of the host word; Age Onset = Age at which participants began learning Spanish; Previous Exp. = Number of years of Spanish courses participants had taken prior to enrolling in the study. Vowel Height was contrast-coded. Log(teacher), Duration, Age Onset, and Previous Exp. Were scaled and centered.

DISCUSSION

This study explored predictors of phonetic accuracy in semi-spontaneous speech collected from a sample of English-speaking university students who were enrolled in introductory Spanish language courses. We were interested in whether participants' production of L2 stops would improve over time and if the frequency of forms in teacher talk would have an impact on learners' production accuracy.

Modeling demonstrated a downward trend in VOT over the yearlong study, particularly in word-initial stops in stressed syllables, but this trend was not statistically significant. This finding runs counter to previous research reporting significant gains in phonetic accuracy in the absence of targeted instruction. Two methodological differences can account for this difference. First, in many previous studies, speech was elicited used controlled tasks such as word reading, sentence reading, and delayed repetition, which allow speakers to focus exclusively on their pronunciation. In contrast, in this study speakers had to coordinate semantic, grammatical, and phonetic processing to complete the picture description task. On this cognitively demanding task, we might expect less accurate pronunciation. In fact, research suggests that even bilingual speakers' pronunciation shifts toward native language norms on tasks involving semantic processing (Gustafson et al., 2013). Second, previous research has typically examined learners in input- and interaction-rich environments, such as learners participating in a short-term immersion program (Casillas, 2016) or learners enrolled in content-based language instruction (Hanzawa, 2018). In these environments, the extensive input that learners receive may catalyze rapid shifts in production accuracy over a relatively short developmental window, especially since learners who opt to enroll in immersion and content-based language programs may exhibit particularly high levels of motivation. In contrast, in the present study participants' introductory-level course was their sole point of contact with Spanish.

Our second research question focused on the role of frequency in shaping phonetic accuracy. Following Trofimovich et al. (2012), we hypothesized that participants might produce more accurate VOT in voiceless stops in words that were more frequent in their teachers' speech. Contrary to our prediction we found no evidence of a significant relationship between the teacher talk frequency measure and participants' VOT production. One potential explanation for differences between our study and Trofimovich et al. (2012) is the salience and perceived importance of the target structure: whereas /ð/ is a highly salient feature for many L2 English learners, akin to /r/ in Spanish, voiceless stops are not. Thus, it could be that both the frequency and salience of target structures influence L2 pronunciation learning. However, our findings should be interpreted with caution. A robust test of frequency effects would require sampling teacher talk longitudinally and aligning learner and teacher samples, as Trofimovich et al. (2012) did. Time by frequency interaction terms would then need to be modeled to examine if more frequent words show greater gains over time. It was not possible to include these interactions in the present study given due to the limitations of the data set. We should also point out that we did not conduct an acoustic analysis of voiceless stops in teacher talk. While we can attest to the fact that the non-native instructors in our study were near-native speakers of Spanish with excellent pronunciation, future research should quantify phonetic variation in the input.

The methodology of the current study is one of its strengths. In many pronunciation studies, particularly research involving acoustic analysis, production is elicited using controlled tasks, and target forms are closely matched for a range of phonetic features. This approach is advantageous because it ensures that all participants produce a consistent number of tokens while avoiding confounding factors. At the same time, while this research design can speak to pronunciation accuracy under optimal conditions, it does not shed much light on pronunciation in spontaneous speech. Indeed, one central issue that pronunciation research must address is the conditions under which L2 speakers are actually able to use pronunciation features spontaneously. Put another way, if we hope to understand L2 phonetic learning in all of its complexity, we need to understand how pronunciation varies over time and across phonetic contexts in spontaneous speech. In line with this view, we opted to embrace the chaos of spontaneous data by controlling for relevant variables via the introduction of covariates, which we identified after we had selected target forms for analysis. This approach is one of many that could prove fruitful for exploring and explaining phonetic variation in complex data sets.

ACKNOWLEDGMENTS

This study was funded by a Language Learning Early Career Research Grant to the first author and an Iowa State University Dean's High Impact Award for Undergraduate Research to the first and second authors. We would like to thank Alexandra Urbanski, Laura Valderrama, Sonca Vo, and Ziwei Zhou for their help with data collection and processing.

ABOUT THE AUTHORS

Charles Nagle is an Assistant Professor of Spanish and the Director of the Spanish Language Program at Iowa State University. His main area of research is second language pronunciation and individual differences. He is particularly interested in how learners' pronunciation develops over time. He is also passionate about statistics, particularly mixed-effects modeling. At Iowa State, he teaches all levels of Spanish language and upper-level linguistics courses.

Shelby Bruun is an undergraduate student at Iowa State University studying Spanish Language, International Culture and Society, and Linguistics. She is primarily interested in second language pronunciation in relation to social and psychological factors.

REFERENCES

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

Boersma, P., & Weenik, D. (2015). Praat: doing phonetics by computer (Version 5.4.08) [Computer program]. Retrieved from http://www.praat.org/

Casillas, J. V. (2016). *The longitudinal development of fine phonetic detail in late learners of Spanish.* [Unpublished doctoral dissertation]. The University of Arizona, Tuscon, AZ.

Casteñada, V. M. L. (1986). El V.O.T. de las oclusivas sordas y sonoras españolas. *Estudios de fonética experimental, 2*, 92-110.

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

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

Flege, J. E., Frieda, E. M., Walley, A. C., & Randazza, L. A. (1998). Lexical factors and segmental accuracy in second language speech production. *Studies in Second Language Acquisition, 20*, 155–187.

Gustafson, E., Engstler, C., & Goldrick, M. (2013). Phonetic processing of non-native speech in semantic vs non-semantic tasks. *Journal of the Acoustical Society of America, 34*(6), EL506–512.

Hanzawa, K. (2018). The development of voice onset time (VOT) in a content-based instruction university program by Japanese learners of English: A longitudinal study. *Canadian Modern Language Review, 74*(4), 502–522.

Kim, S., Kim, J., & Cho, T. (2018). Prosodic-structural modulation of stop voicing contrast along the VOT continuum in trochaic and iambic words in American English. *Journal of Phonetics, 71*, 65–80.

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

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

Lisker, L., & Abramson, A. S. (1967). Some effects of context on Voice Onset Time in English stops. *Language and Speech, 10*(1), 1–28.

McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., & Sonderegger, M. (2017). Montreal Forced Aligner [Computer program]. Version 0.9.0. http://montrealcorpustools.github.io/Montreal-Forced-Aligner/.

Munro, M. J., Derwing, T. M., & Thomson, R. I. (2015). Setting segmental priorities for English learners: Evidence from a longitudinal study. *International Review of Applied Linguistics in Language Teaching, 53*(1), 39–60.

Nagle, C. (2019). A longitudinal study of voice onset time development in L2 Spanish stops. *Applied Linguistics, 40*(1), 86–107.

Olson, D. J. (2019). Feature acquisition in second language phonetic development: Evidence from phonetic training. *Language Learning, 69*(2), 366–404.

Pierrehumbert, J. B. (2001). Exemplar dynamics: Word frequency, lenition, and contrast. In J. L. Bybee & P. J. Hopper (Eds.), *Frequency and the emergence of linguistic structure* (pp. 137–157). Philadelphia, PA: John Benjamins.

Pierrehumbert, J. B. (2003). Phonetic diversity, statistical learning, and acquisition of phonology. *Language and Speech, 46*(2-3), 115–154.

Saito, K., Dewaele, J. M., & Hanzawa, K. (2017). A longitudinal investigation of the relationship between motivation and late second language speech learning in classroom settings. *Language and Speech, 60*(4), 614–632.

Simonet, M., Casillas, J. V., & Díaz, Y. (2014). The effects of stress/accent on VOT depend on language (English, Spanish), consonant (/d/, /t/) and linguistic experience (monolinguals, bilinguals). In *Proceedings of the International Conference on Speech Prosody* (pp. 202–206). International Speech Communications Association.

Stuart-Smith, J., Sonderegger, M., Rathcke, T., & Macdonald, R. (2015). The private life of stops: VOT in a real-time corpus of spontaneous Glaswegian. *Laboratory Phonology, 6*(3-4), 505–549.

Team, R. C. (2019). R: A language and environment for statistical computing (Version 3.6.1). Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/

Trofimovich, P., Collins, L., Cardoso, W., White, J., & Horst, M. (2012). A frequency-based approach to L2 phonological learning: Teacher input and student output in an intensive ESL context. *TESOL Quarterly, 46*(1), 176–186.