GESTURAL DEVELOPMENT IN ADULT SECOND LANGUAGE ACQUISITION: TARGETS AND TIMING

by

Alexander C. Aldrich

A Dissertation Submitted to the Faculty of the

DEPARTMENT OF SPANISH AND PORTUGUESE

In Partial Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY
WITH A MAJOR IN SPANISH

In the Graduate College

THE UNIVERSITY OF ARIZONA

2019

THE UNIVERSITY OF ARIZONA
GRADUATE COLLEGE

As members of the Dissertation Committee, we certify that we have read the dissertation prepared by Alexander C. Aldrich, titled Gestural Development in Adult Second Language Acquisition: Targets and Timing and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy.

Miquel Simonet — Date: May 7, 2019

Natasha Warner — Date: May 7, 2019

Sonia Colina — Date: May 7, 2019

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copies of the dissertation to the Graduate College.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

Miquel Simonet — Date: May 7, 2019
Dissertation Committee Chair
Department of Spanish & Portuguese

ACKNOWLEDGEMENTS

Firstly, I thank God not only for providing me with the capacity and the opportunity to have accomplished the daunting task of completing a dissertation, but also for accompanying me through the hardest and happiest times of my life, especially graduate school. In the words of Paul, I know that "I can do all things through Christ which strengtheneth me."

I'm thankful to my loving and supportive parents and stepparents. Mom, your example of enduring challenges without giving up has been a blessing to me through this process. The day that you made me re-enroll in geometry after I dropped out because it seemed impossibly hard was a turning point for the rest of my life. Dad, you always told me that I could accomplish anything if I put my mind to it, and, after hearing it enough times, I believed you.

There is only one woman in the universe that would have endured all of these years of college at my side, and, somehow, I was lucky enough to find her. Thank you for your support, love, and companionship. You always listened to me when I needed someone to talk to, even when it was about linguistics. These years of graduate school have been enriched by the gift of fatherhood that you have given me. ♫ *You are the love of my life*. ♫

My scholarship has benefited from the gifted and dedicated mentorship of several individuals. Thank you, Miquel Simonet, for your enthusiasm for experimental design, bilingualism, and writing. Your guidance on this project and on my previous academic work has been fundamental to their success. Thank you, Natasha Warner, for your help with learning statistics and phonetics. Thank you, Antxon Olarrea, for helping me improve my writing. All of that *antxonear* was well worth it. Thank you, Kristin Doran, for entrusting me with Spanish 251 and D2L for a year. Lastly, throughout this work, I have stood on the shoulders of giants. Thank you to all those who have helped pave the way.

And, of course, thank you to all of my peers. The friendship, laughter, and commiseration that we have shared has made it all worth it.

DEDICATION

For my children.

Dream greatly!

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

This dissertation deals with the development of place of articulation and of rhythm in adult second language learners. English and Spanish, while sharing similar phonological stop inventories, have coronal stops that differ as a matter of gestural target: In English, the coronal stops have an alveolar place of articulation, whereas they have a dental target in Spanish. Previous research on the acquisition of second language stops, particularly as a function of the acoustic correlates of voicing, have found that second language learners find acquiring stop contrasts particularly challenging; however, to date, no published data exist on the acquisition of L2 stops that have unique gestural targets. English and Spanish also differ typologically as a matter of their rhythm, with English being categorized as stress-timed and Spanish as syllable-timed. Previous research has shown that adults are capable of acquiring aspects of their L2 rhythm; however, the findings have been mixed. Furthermore, detailed developmental data on L2 rhythm is lacking. The present dissertation addresses both the acquisition of place of articulation and of rhythm in a cross-sectional design.

The production data of 41 female, adult native speakers of English learning Spanish were collected and compared to those of 9 female adult native bilinguals of English and Spanish. The learners were assigned to four different study conditions based on their linguistic experience. The results suggest that the native bilinguals produce Spanish and English with language-specific gestural targets for their coronal stops and language-specific rhythm patterns in the direction expected for each language. The data from the language learners imply that, on the one hand, adults learning a second language are capable of developing increasingly target-like gestural targets in their L2 Spanish coronal stop production as they gain experience with the second language. On the other hand, the rhythm data reveal that speakers of a second language quickly learn to make adjustments to their rhythm toward target-like norms, but that evidence of transfer remains. The implications of the findings are discussed with respect to second language phonological theory.

CHAPTER 1
INTRODUCTION

This dissertation deals with the development of place of articulation and of rhythm in native speakers of English learning Spanish as a second language. Spanish and English differ phonologically in many ways: For instance, although they share phonologically similar stop inventories—each has a set of bilabial, coronal, and velar stops that contrast phonologically in voicing: /b, d, g/ and /p, t, k/—phonetically, these stops are greatly dissimilar between the two languages. Although much literature exists on the nature of the development of stop voicing in an L2 (Reeder, 1998; Stevens, 2001; Zampini, 1998), this study is concerned principally with the acquisition of stops that are phonologically similar to those of the L1 but which differ in regard to gestural target. Specifically, it investigates the development of the production of L2 Spanish coronal stops, which have a dental place of articulation, vis-à-vis L1 English coronal stops, which are alveolar. Beyond stop inventories, Spanish and English are typologically different in regard to their rhythm: Spanish is a so-called syllable-timed language, whereas English is a so-called stress-timed language, and this dissertation presents rhythm data on the L2 speech of learners whose native and second language differ typologically as a matter of rhythm (L1 English vs. L2 Spanish).

1.1 THE DEVELOPMENT OF PLACE OF ARTICULATION

In the realm of first language segmental development, stops are one of the first classes of consonants acquired in speech production (Gerken, 2009; Stoel-Gammon, 1998). Gerken (2009, p. 74), summarizing prior research on this topic, states that this is due to a combination of factors, but that "articulatory constraints" are a leading factor as to why stops are usually learned first, suggesting that they are gesturally easier to acquire than other manners of articulation. For instance, it has been documented among the world's languages that the words for *father*, which often begin with a bilabial or coronal stop consonant (e.g., *papá*), are learned before the word for *mother*, which often begin with a bilabial nasal consonant (*mamá*), even though children tend to spend more time interacting with the latter than the former (as reviewed in Stoel-Gammon, 1998). Despite the seeming ease with which stops are acquired in a first language, they are notoriously challenging to acquire in a second language (Flege, 1987).

Indeed, the acquisition of the stop phonemes of a second language has been a common thread among the body of literature in second language speech (Zampini, 2013), likely due to their perceived phonological similarity between languages (Flege, 1995). There are many phonological features that distinguish stop consonants from one another, including voicing (i.e., [+ voice]) and place of articulation (e.g., [coronal]) (Chomsky & Halle, 1968). Although the acquisition of stop phonemes in a second language has often concentrated on the production and perception of cues to stop voicing categories, such as voice onset time (VOT) (Zampini, 2013), research has also shown that stops that differ as a matter of place of articulation can be distinguished by acoustic correlates such as formant transitions (Delattre, Liberman, & Cooper, 1955; Ladefoged, 2006), closure duration (Port, 1979), and spectral information in the stop release (Jongman & Blumstein, 1985), at least in monolinguals. In fact, spectral cues have proven to be effective in distinguishing within-voicing-category stops (i.e., [b, d, g] or [p, t, k]) that differ slightly in regard to exact articulatory target but which share the same phonological feature [coronal], as in the language Malayalam, which has contrasting dental and alveolar stops (Jongman & Blumstein, 1985).

Unlike in the case of phonological vowel inventories, which have been theorized to contain contrasts that are *maximally* different in perception (Liljencrants & Lindblom, 1972), segments in consonant inventories differ *minimally* in terms of phonological features (Ohala, 1980), and even the seemingly minute differences of place of articulation between alveolar and dental stop consonants can be captured robustly by measuring spectral information in the acoustic signal (Jongman & Blumstein, 1985). Native speakers of English, whose coronal stop has an alveolar place of articulation, that are learners of Spanish, whose coronal stop has a dental place of articulation, face a challenge described in Escudero's (2005, 2009) Second Language Linguistic Perception model (L2LP) as the *Similar Scenario*: acquire L2 sounds for which a phonological equivalent exists in the L1, that differ only slightly as a matter of gestural target. Escudero's model predicts that this task would be difficult for second language learners, as they would have to adjust the phonetic categories of their L2 perception grammar, whose initial state was a fully copy of the L1, to match the phonetic categories of the target language. To date, it remains unknown whether speakers of a second language learn to update the phonetic categories relating to unique gestural targets for similar stop phonemes in a Similar Scenario.

There also exists abundant evidence that linguistic experience is a factor in the outcomes of L2 segmental acquisition. For instance, Casillas (2016) and Zampini (1998), in longitudinal studies, and Reeder (1998) and Stevens (2001), in cross-sectional studies, reported that the samples of native English speakers learning Spanish in their studies made gains toward target-like stop production, in terms of the acoustic correlates to stop voicing, with increasing exposure to the L2. The present dissertation presents data in a cross-sectional design of the development of gestural targets in L2 stop phonemes, which will be a valuable contribution to what is known currently about the effect of linguistic experience on L2 segmental acquisition.

Specifically, one of the sets of research questions driving the presentation of this dissertation are the following: Are learners of a second language capable of acquiring the phonetic differences for similar phonemes (i.e., Spanish /t/ and English /t/) whose articulatory target differs as a matter of place of articulation (i.e., dental vs. alveolar)? If so, at what stage in the learning trajectory does this acquisition take place, and what extralinguistic factors, if any, are relevant to it? One of the objectives of the present chapter is to address these questions via a cross-sectional study of the development of articulatory targets—specifically, the Spanish voiceless, dental /t/—in adult learners of a second language.

1.2 THE DEVELOPMENT OF L2 RHYTHM

Beyond the development of segmental aspects of second language speech, this dissertation also deals with the development of prosody, those linguistic phenomena not restrained to a single unit of sound, such as pitch, accent, stress, fluency, tone, loudness, intonation, and rhythm (Colantoni, Steele, & Escudero, 2015, p. 314). There is evidence that prosodic phenomena are learned very early in the trajectory of human language development. For instance, research has found that fetuses are capable of learning language-specific acoustic properties of pitch in utero (Partanen et al., 2013). Likewise, newborn babies possess the capability to perceive (Mehler et al., 1988) and produce (Mampe, Friederici, Christophe, & Wermke, 2009) language-specific prosodic cues. These facts together suggest that prosodic phenomena form an integral part of L1 development and, therefore, L1 phonology.

Indeed, the languages of the world have often been categorized typologically as a matter of at least one prosodic phenomenon—rhythm, or the strong and weak patterns perceived in spoken language. Cross-linguistic differences in rhythm were first documented by Lloyd James (1940), and, although it is said that rhythm is a manifestation of different abstract aspects of the phonology of the language, including syllable structure and lexical stress (Dauer, 1983), investigators have proposed means by which rhythm can be studied empirically (Grabe & Low, 2002). With the ability to quantify rhythm, several studies have investigated the acquisition of rhythm in learners of a second language whose rhythm class is typologically different from that of their L1 (Carter, 2005; Grenon & White, 2008; Gut, 2003; Kinoshita & Sheppard, 2008, 2011; Li & Post, 2014; Lin & Wang, 2005; Mok & Dellwo, 2008; Nava & Zubizarreta, 2009; Rasier & Hiligsmann, 2007; Stockmal, Markus, & Bond, 2005; Thomas & Carter, 2006; Tortel & Hirst, 2010; Trofimovich & Baker, 2006; Valls-Ferrer, 2011; White &

Mattys, 2007), although little has been documented regarding the acquisition of L2 Spanish rhythm in native speakers of English (White & Mattys, 2007). Linguistic experience has been shown to be relevant to the acquisition of L2 rhythm (Benet, Gabriel, Kireva, & Pešková, 2012), and this justifies, in part, the study of L2 rhythm in English speakers learning Spanish.

Furthermore, while the studies cited previously generally agree that some rhythm patterns of the L2 can be acquired, the findings have been mixed in regard to which aspects of rhythm are learned. Even fewer data are available on the chronology of the acquisition of rhythm, as only a handful of studies have examined the acquisition of rhythm over time (Stockmal et al., 2005; Tortel & Hirst, 2010; Trofimovich & Baker, 2006), none of which analyzed Spanish as a second language. The present study fills this gap by presenting cross-sectional data on the acquisition of L2 Spanish rhythm in native speakers of English, languages that differ typologically in terms of rhythm. Although there is no formal theory for the second language acquisition of prosodic phenomena, one can extrapolate from the studies cited on segmental acquisition that increasing linguistic experience will lead to gains toward more target-like rhythm. In particular, it would be expected that learners of Spanish would exhibit rhythm patterns in their L2 speech that resemble those of their L1, but that, with increasing linguistic experience, the rhythm patterns would begin to resemble the norms of native speakers of the target language.

Specifically, the final set of research questions driving this dissertation are the following: Do native speakers of English acquire the rhythm patterns of their L2 Spanish, a language that differs typologically from English in regard to rhythm? If so, are gains made toward target-like rhythm patterns with increasing linguistic experience?

These questions, and those presented in the previous subsection, will be discussed at length in the upcoming chapters. First, however, a description of the sampling procedures carried out in this research is needed to justify the cross-sectional, developmental contribution of this research.

CHAPTER 2

SAMPLING PROCEDURES

2.1 PARTICIPANTS

In order to measure the development of gestural targets and of rhythm in the L2, a cross-sectional research design was implemented, as increased experience with the L2 has consistently been shown to lead to gains in L2 phonological acquisition (Flege & Fletcher, 1992; Flege, Munro, & MacKay, 1995). The objective of this chapter is to describe in detail the sampling procedures carried out and the method used to assign participants into study conditions. Throughout the process, the apparent Spanish proficiency of the participant, including the individual's experience taking Spanish in college, knowledge of Spanish vocabulary, self-reported speaking proficiency in Spanish, and bilingual balance, were submitted to statistical models to organize the speakers in terms of increasing linguistic experience.

To this end, a series of questionnaires were used during recruitment to obtain measures relating to the linguistic experience of the participants. A sample of native English/Spanish bilinguals (BL) was recruited as the control condition, and a sample of native speakers of English learning Spanish as a second language, termed here as second language learners (LL), was recruited as part of study condition. All those who took part in the experiments were recruited from the University of Arizona campus and were enrolled as undergraduate students. Recruitment took place in and around campus, largely in visits to classes in the Department of Spanish and Portuguese. All participants were compensated with a small sum of money for their participation in the research.

2.1.1 Recruitment

LLs were recruited among four different levels of Spanish courses at the University of Arizona. The first two levels from which they were recruited were Second Semester Spanish and Fourth Semester Spanish. Traditionally, students enrolled in these classes are fulfilling a university-level, second-language requirement. The third level from which LLs were recruited was, firstly, Intermediate Grammar and Writing and, secondly, Translation and Interpretation: Social Justice and Practice. Both of these courses have the same pre-requisite course. Students in these classes are usually those seeking a minor in Spanish in the Department of Spanish and Portuguese. The fourth and final level from which LLs were recruited were from a multitude of upper-division, 300- and 400-level Spanish classes, all of which required that the individual had taken the Intermediate Grammar and Writing course prior to being enrolled. Students enrolled in these classes are traditionally minoring or majoring in Spanish.

The BL participants were recruited from the Spanish as a Heritage Language program and the upper-division courses in the Department of Spanish and Portuguese. A sample from this population is included in the study for two reasons: (a) as a control condition, to which the LL conditions will be compared, and (b) as a study condition, as there is no published data to date on the nature of place of articulation of English and Spanish coronal stops in native English/Spanish bilinguals, nor is there a complete understanding of this same population's rhythm patterns in both of their languages. A sample of BLs has been chosen as the control condition instead of monolingual Spanish speakers, as is usually selected in SLA studies, under the assumption that LLs have the desire to be, are, or will become bilinguals themselves, albeit in adulthood. Furthermore, as pointed out by Casillas (2016), both BLs and LLs tend to exhibit a level of interlanguage in regard to their phonetic performance, producing acoustic values for multiple cues that lie in between those of monolinguals of either language. In other words, BLs more closely resemble the end-state of language learning of the LLs vis-à-vis monolingual Spanish speakers. Lastly, by recruiting BLs for the control condition, as opposed to having two monolingual control conditions, the experiments are entirely within-subjects, as they analyze data from each speaker's English and Spanish production. This design will increase the power of the statistical analyses and thus allow for greater generalizability of the results.

2.1.2 Inclusion/Exclusion Criteria

In order to make an appointment to take part in the research, the potential participants first took an online survey via Qualtrics, which covered basic inclusion and exclusion criteria for the study. Potential participants were asked details about their linguistic experience, including whether (and, if applicable, which) languages in addition to English were spoken in the home, what Spanish class the individual was enrolled in, and whether the individual had any vision, hearing, or speech difficulties. The same survey was administered to all potential participants, regardless of whether it was anticipated that they were native bilingual speakers. Those that met the following criteria were invited to participate in the research and were considered LLs:

1. 18 years of age or older
2. Self-identified as learners of Spanish as a second language enrolled in a Spanish class as previously described
3. Reported that no language besides English was spoken in the home
4. Not fluent in any language besides English and Spanish
5. Did not have any vision, hearing, or speech difficulties

Those that met the following criteria were invited to participate in the research and were considered native bilinguals of English and Spanish:

1. 18 years of age or older
2. Reported that English and Spanish (and only English and Spanish) were spoken in the home since s/he was a child
3. Not fluent in any language besides English and Spanish
4. Did not have any vision, hearing, or speech difficulties

(See Appendix A for more detail.) After completing the initial survey, the potential participants were redirected immediately to a website to sign up for an available time to visit the lab to take part in the research.

2.1.3 Bilingual Language Profile

Once in the lab, all participants filled out the Bilingual Language Profile (BLP) (Birdsong, Gertken, & Amengual, 2012) on a computer. As stated on the website, "The BLP is an **open and free** assessment tool for researchers, educators, and anyone with an interest in assessing language dominance" (Birdsong et al., 2012, bold in original). The BLP asks the participant questions relating to his/her use of both English and Spanish in environments such as the home, school, work, and with friends, and also asks questions relating to linguistic identity. The BLP also asks the participant questions relating to age of onset of learning of both languages, years of grammar classes taken in each language, number of years living in a country whose dominant language is either Spanish or English, and asks them to self-assess their speaking, listening, reading, and writing proficiency in both languages. (See Appendix B for a fully copy of the BLP.) The survey calculates a dominance score centered around 0: The closer to 0 the score is, the more balanced the bilingual is between English and Spanish. Positive scores suggest English dominance, whereas negative scores suggest Spanish dominance. The BLP is a standardized test that has been used in multiple studies, particularly those whose study conditions were samples of early bilinguals (Banov, 2014; Casillas, 2015; Casillas & Simonet, 2015; Nadeu & Renwick, 2016; Simonet, 2014), but also those in which they served as a control condition (Casillas, 2016). It has also been used to measure the language dominance of LLs (Casillas & Simonet, 2015), as done in the present study, which allows a comparison of language dominance between the BL and LL conditions.

2.1.4 Background questionnaire

In addition to the BLP, a background questionnaire, which gathered information not collected by the BLP regarding the participant's linguistic experience and relevant personal information, was filled out by each participant on a computer. For instance, participants provided details regarding their current enrollment in Spanish classes, responded whether a parent, guardian, or close family member spoke to him/her habitually in Spanish (or any other language) in the home while growing up, and provided information on place of birth, place of raising, what other foreign languages s/he had studied, experience living abroad, and experience with taking Spanish phonetics or linguistics courses in the university. (See Appendix C for the full background questionnaire.)

2.1.5 LexTALE and LexTALE-Esp

In light of the fact that each learner of a foreign language is affected in unique ways by his/her individual differences—including affective factors such as motivation, attitude, and so forth (Wong-Fillmore, 1979)—and the finding that measured language proficiency of students enrolled in separate but similar courses of the same level differ from one course or style of instruction to the next (Young, 2008), it was considered expedient to include additional measures of proficiency for the participants of the present study, beyond those data already obtained regarding Spanish class enrollment, language use, and language experience (as measured by the BLP and the background questionnaire). Thus, all participants took part in standardized tests of "vocabulary knowledge and proficiency" (Lemhöfer & Broersma, 2012, p. 325) in each language (Spanish and English). Specifically, the participants took the English LexTALE (Lemhöfer & Broersma, 2012) (henceforth LexTALE) and Spanish-language LexTALE-Esp (Izura, Cuetos, & Brysbaert, 2013) (henceforth LexTALE-Esp), which were administered on a computer as part of the experiment phase of this research. Lemhöfer and Broersma (2012) validated the LexTALE as an effective measure of second-language proficiency by comparing the results to participants' own L2 self-assessment ratings, their accuracy in a translation test, and their placement in an L2 student placement exam and found that the LexTALE scores correlated with measures of L2 grammatical knowledge as measured by a grammar translation test. The LexTALE was later adapted to other languages, including to Spanish, which is known as the LexTALE-Esp (Izura et al., 2013). LexTALE is a lexical decision task: it asks the participant to answer whether the particular word in question is a real word in the target language, and to only answer 'yes' if she is sure that it is a word, even if she cannot define it. A percentage score is calculated from the results by taking the average correct for both words and non-words. The English version contains 60 experimental items (20 non-words), and the Spanish version contains 90 experimental items (30 non-words).

In the present experiment, all participants were administered the LexTALE and LexTALE-Esp in PsychoPy 2 (Peirce et al., 2019). The participants read the instructions first and then were presented each lexical item one at a time on the computer screen, with each item appearing in the same order for every participant as recommended by the creators of LexTALE. The participants were given as much time to respond 'Yes' or 'No' to each item, and a fixation cross lasting 500 ms was shown in between each. The instructions and items for the LexTALE English and LexTALE-Esp versions are found in Appendix D.

2.1.6 Sampling Procedures

In light of the individual differences between speakers in regard to learning outcomes in a second language, an agglomerative nesting (i.e., hierarchical cluster) model was used to assign participants into their respective study conditions. The following subsections report the regression analyses used to determine which extralinguistic and proficiency data to include in the cluster model, in addition to the outcomes of the model and the condition assignments.

2.1.6.1 Simple and multiple linear regression models for hierarchical clustering

A multiple regression model was fit in R (R Core Team, 2015) via RStudio (RStudio Team, 2016) using the function *lm* to determine which extralinguistic and proficiency data to include in the cluster model prior to computation. For all models, only data relevant to Spanish, the language of study of the experiment, were considered, including BLP scores, Spanish class level (SCL), self-reported speaking proficiency, and LexTALE-Esp. Previous cross-sectional research on L2 phonological development has often assigned participants into conditions as a matter of their experience with formal language instruction (Reeder, 1998). However, as discussed previously, the foreign-language class in which one is enrolled is not always a reliable indication of his/her proficiency. Taking these facts into account, several linear regression models were fit with SCL as the response variable coded numerically[1] as follows:

- 0 = BLs, which are the presumed end state of the LLs
- 1 = Participants enrolled in upper-division, 300- and 400-level Spanish classes
- 2 = Participants enrolled in Intermediate Grammar and Writing or Translation and Interpretation
- 3 = Participants enrolled in 4th-semester Spanish
- 4 = Participants enrolled in 2nd-semester Spanish

Initially, a simple linear model was fit to determine whether the LexTALE-Esp predicts SCL with SCL as response variable and LexTALE-Esp as the explanatory variable. The model showed that LexTALE-Esp explains a significant amount of variance in SCL ($R^2 = 0.24$, $F(1, 48) = 15.42$, $p < .001$) and the LexTALE-Esp coefficient was significant ($b = –.08$, $t(48) = –3.93$, $p < .001$), suggesting that, as SCL decreased (or class level moved toward less experienced), LexTALE-Esp scores also decreased. Furthermore, a linear model fit with SCL as response and

[1] In these models, in order to fit a linear model, it is assumed that the difference between conditions is equal. This same assumption is made for linear regression models in Chapters 3 and 4.

BLP scores as the explanatory variable suggested that BLP scores explain a significant amount of variance in SCL ($R^2 = 0.44$, $F(1, 48) = 37.94$, $p < .001$), and the BLP scores coefficient was significant ($b = .02$, $t(48) = 6.16$, $p < .001$), implying that, as class level moved towards less experience, the English dominance (increased BLP scores) increased. Additionally, a linear model fit with SCL as response and Self-reported Spanish Speaking Proficiency (SRSP) as the explanatory variable suggests that the latter explains a significant amount of variance in SCL ($R^2 = 0.25$, $F(1, 48) = 16.17$, $p < .001$), and the SRSP coefficient was also significant ($b = -.58$, $t(48) = -4.02$, $p < .001$), denoting that self-ratings of proficiency decreased as experience with formal Spanish instruction decreased. Next, a multiple regression model was fit with SCL as response and BLP scores and Self-reported Spanish Speaking Proficiency as explanatory variables. The model suggested that the two explanatory variables combined explain a significant amount of variance in SCL ($R^2 = 0.45$, $F(2, 47) = 19.01$, $p < .001$). However, a partial F test comparing the nested model (SCL ~ BLP scores) to that of the full model (SCL ~ BLP scores + SRSP) suggested that adding SRSP to the multiple regression model did not significantly improve it ($F < 1$). Therefore, as follow up, a multiple regression model was then fit with SCL as the response variable and BLP scores and LexTALE-Esp as explanatory variables. The results suggested that BLP scores and LexTALE-Esp predict a significant amount of variance in SCL ($R^2 = 0.50$, $F(2, 47) = 23.40$, $p < .001$). A partial F test comparing the nested model (SCL ~ BLP scores) with that of the full model (SCL ~ BLP scores + LexTALE-Esp) suggested that BLP scores combined with LexTALE-Esp predict a significantly greater amount of variance in SCL than BLP scores alone ($F(1, 47) = 5.40$, $p < .03$). Lastly, adding SRSP to this full model did not improve it significantly ($F(2, 46) = 2.86$, $p > .05$). Thus, it appears that BLP scores and LexTALE-Esp are good predictors of SCL for the participants in this study. In light of the results of these regression models combined, the variables LexTALE-Esp scores, BLP scores, and SCL were all included in the cluster model.

2.1.6.2 Agglomerative nesting model

A distance matrix was computed to calculate the agglomerative nesting model. Using the *cluster* package (Maechler et al., 2018) in R, the function *daisy* was used to compute a dissimilarity distance matrix for the continuous variables LexTALE-Esp scores and BLP scores and the factor SCL. The *daisy* function allows for a mixed design of continuous and categorical variables. It first converts the categorical variable into a dummy numeric variable, then standardizes the variables, and, finally, computes Gower's (1971) general dissimilarity coefficient for pairwise differences across each row of the data frame. Next, the distance matrix was submitted to the *agnes* (i.e., agglomerative nesting) algorithmic function of the *cluster* package to calculate an agglomerative nesting model using the complete linkage method. The complete linkage method measures distances between clusters as the greatest distance between individual observations within each cluster. The *agnes* algorithm, as opposed to traditional hierarchical cluster algorithms, such as that found in the *hclust* algorithmic function in R, takes a more holistic and globally accurate method as each new cluster is created in the model (Spector, 2011). A dendrogram of the nesting model can be found in Figure 2.1.

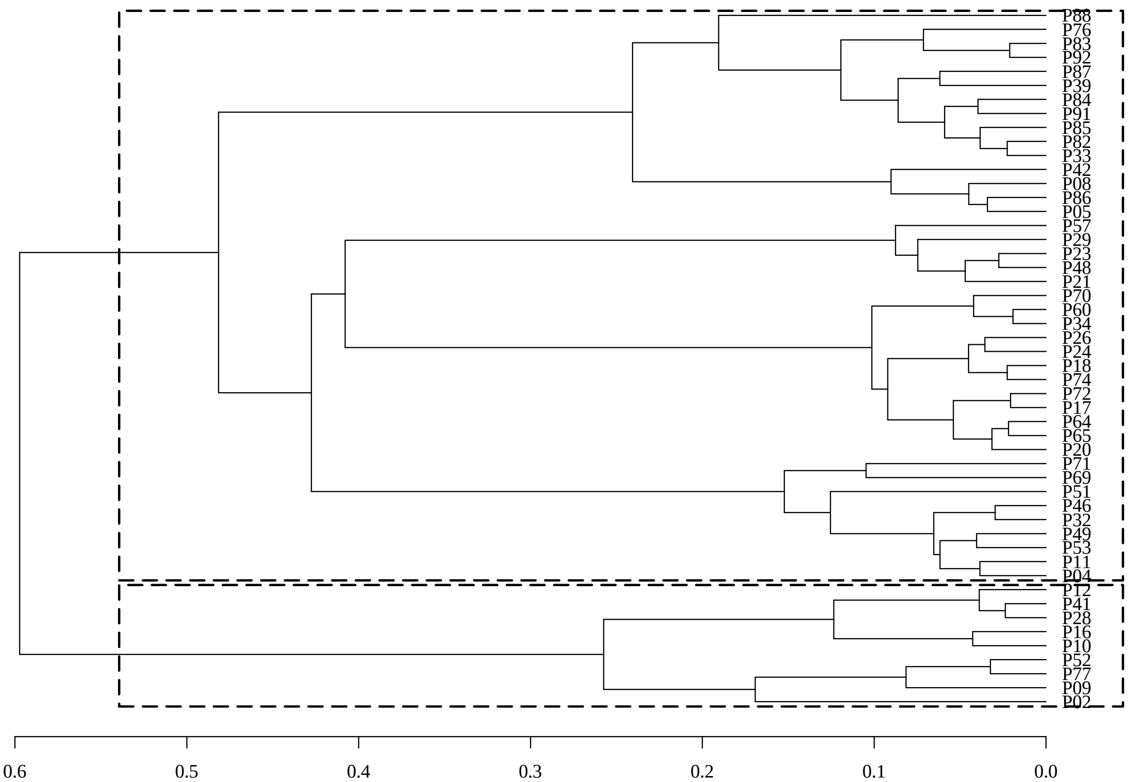


Figure 2.1 Dendrogram of agglomerative nesting model displaying participants grouped into clusters based on BLP scores, LexTALE-Esp scores, and Spanish Class Level. Participant numbers (P##) are listed to the right of their corresponding line in the model. The model clustered all BLs together (small dotted rectangle) and all LLs together (big dotted rectangle) separately.

The dendrogram visualizes how the nesting model allows for any number of clusters to be extracted based on the number of lines available along *y* at any arbitrary value of *x*. From Figure 2.1, it is evident that two natural clusters are initially formed, which are comprised of, firstly, the BLs (small dotted rectangle), and, secondly, the LLs (large dotted rectangle). A five-cluster call from the model separates the participants into conditions corresponding to their initial Spanish class enrollment. That is, all those enrolled in 2nd-semester Spanish are clustered together, all those enrolled in 4th-semester Spanish are clustered together, and so on.

2.1.7 Study Conditions

In order to achieve a cross-sectional design to observe linguistic development, a total of four study conditions (all LLs) and one control condition (BLs) were included in the experiment. The different conditions of linguistic experience (LE) of the study, as determined by an agglomerative nesting model, in order of increasing linguistic experience, are as follows: LL1 (n = 5), LL2 (n = 12), LL3 (n = 9), LL4 (n = 15), and BL (n = 9). A summary of the study conditions, sample sizes, relevant extralinguistic data, and proficiency measures are found in Table 2.1.

Table 2.1 (Above) Descriptive statistics (mean followed by standard deviation in parentheses) of extralinguistic data, including years of Spanish-as-a-foreign-language class taken, years living abroad in a Spanish-speaking country, age of acquisition of English, and age of acquisition of Spanish for each Linguistic Experience (LE) condition. (Below) Descriptive statistics of proficiency and dominance measures, including self-reported speaking proficiency in English (scale of 0–6), self-reported speaking proficiency in Spanish (scale of 0–6), LexTALE (% correct), LexTALE-Esp (% correct) for each condition. Values in bold represent significant differences vis-à-vis the BL condition.

LE Condition	Spanish Class Enrollment	Years of Spanish Class	Years Living Abroad	Age of Acquisition of English (Years)	Age of Acquisition of Spanish (Years)
LL1	2nd-semester	2.8 (1.3)	0.0 (0.0)	0.0 (0.0)	14 (2)
LL2	4th-semester	5.3 (1.9)	0.0 (0.0)	0.0 (0.0)	12 (4)
LL3	3rd-year	8.1 (2.2)	0.1 (0.2)	0.0 (0.0)	11 (2)
LL4	4th-year	7.5 (2.7)	0.5 (0.7)	0.0 (0.0)	13 (3)
BL	N/A	2.7 (2.9)	0.9 (1.8)	3.8 (1.1)	0 (0)

LE Condition	Self-reported *English* Speaking Proficiency	Self-reported *Spanish* Speaking Proficiency	LexTALE (Eng)	LexTALE-Esp	Bilingual Language Profile
LL1	5.8 (0.5)	**2.2** (0.8)	86 (10)	**49** (3)	**132** (15)
LL2	5.9 (0.3)	**3.0** (0.7)	86 (8)	**50** (3)	**145** (22)
LL3	5.9 (0.3)	3.1 (1.5)	85 (9)	**49** (5)	**124** (24)
LL4	6 (0)	3.6 (0.9)	87 (10)	57 (9)	**118** (27)
BL	5.4 (0.7)	4.2 (0.8)	81 (14)	60 (11)	50 (28)

2.1.7.1 Extralinguistic data

As observed in Table 2.1, extralinguistic data provided by the participants show a general trend of increasing mean number of years of classes of Spanish as a foreign language as linguistic experience increases, with the exception of the BL condition, which have a mean similar to that of the LL1 condition. Regarding the number of years spent living abroad in a Spanish-speaking country[2], the LL1 and LL2 conditions reported never having lived abroad. However, beginning with the LL3 and on to the BL speakers, a trend of increasing mean number of years spent living in a Spanish-speaking country is observed, as would be the case for upper-division students who are majors or minors that have begun to attend study-abroad trips, and, in the case of the BLs, likely spending time living in the country with which they identify culturally and linguistically, which, in this case, would be Mexico. In terms of age of acquisition (AoA), the pattern for the BLs and LLs is inverse between English and Spanish. For English, all LLs report their AoA as age 0, whereas the BLs report an average age of 3.8. Considering that the BLs were born and raised in Spanish-speaking households, this difference in AoA is not surprising and is likely due to the BLs' not learning English until entering preschool or kindergarten. For all LLs, the age of onset of learning Spanish is about that of adolescence, whereas all BLs reported having learned Spanish from birth.

2.1.7.2 Proficiency and dominance measures

Measures of proficiency and language dominance—including self-reported speaking proficiency (English and Spanish), LexTALE, LexTALE-Esp, and BLP—generally showed trends patterning in the expected direction in relation to the presumed linguistic experience of the given speakers (see Table 2.1). In regard to self-reported speaking proficiency, which is scaled from 0–6, with 6 meaning that the participant reports speaking the given language 'very well,' participants from all conditions reported similar self-ratings in English speaking proficiency; however, for Spanish, there is a trend of increasing mean of self-rating from LL1 to the BL speakers. The self-reported speaking proficiency in English was submitted to a one-

[2] These data were extracted from the Bilingual Language Profile. In this profile, the participants are asked how many years they have spent living in Spanish-speaking country. One of the possible answers is "Less than one year," which is greater than 0 but less than 1. In all cases, when this option was selected by the participant, the data were coded as 0.5 years.

way, between-subjects ANOVA with Self-reported Speaking Proficiency (English) as the dependent variable and LE (LL1, LL2, LL3, LL4, BL) as the independent variable. The ANOVA reported a significant main effect of LE ($F(4, 45) = 3.12$, $p < .03$).[3] However, none of the follow-up *t* tests[4] to explore pairwise comparisons between the LL conditions and the BL condition reported significant differences. In terms of Spanish, the self-reported proficiency was submitted to a one-way, between-subjects Analysis of Variance (ANOVA) with Self-reported Speaking Proficiency (Spanish) as the dependent variable and LE (LL1, LL2, LL3, LL4, BL) as independent variable ($F(4, 45) = 4.11$, $p < .01$). Follow-up *t* tests for pairwise comparisons showed that the self-reported speaking proficiency between the LL1 and BL conditions was significant ($t(8.36) = -4.34$, $p < .005$), as was the difference between the LL2 and BL conditions ($t(16.13) = -3.50$, $p < .005$). The differences between the LL3 and the BL and those between the LL4 and the BL conditions were not significant (LL3 × Bilingual: $t(12.33) = -1.91$, p > . 05; LL4 × Bilingual: $t(18.22) = -1.71$, p > .10). These results denote that the SRSP of the participants in the LL3 and LL4 conditions was not different from that of the BL participants. However, the SRSP reported by the LL1 and LL2 participants was. The mean SRSP rating for the BL speakers ($M = 4.2$ $SD = 0.8$) may initially be perceived as low. However, it has been reported that native English/Spanish bilinguals living in the United States overwhelmingly consider their proficiency in one of either language to be subpar (Martínez, 2010; Toribio & Durán, 2018) relative to monolinguals of the given language. In

[3] The *ezANOVA* function from the *ez* package (Lawrence, 2016) in R is used in all omnibus ANOVAs. This function tests for homogeneity of variance in fully between-subjects designs, in addition to violations of the assumption of sphericity in within-subjects and/or mixed designs and provides corresponding corrections and adjusted *p* values where necessary. Importantly, *ezANOVA* also allows the user to specify the type of sums of squares approach to take in the case of unbalanced conditions (as is the case here). In all tests, type is set to 3, which models after that used in commercial statistical software such as SPSS (Lawrence, 2016).

[4] In all cases, corrections for familywise error is applied only when the number of pairwise comparisons is greater than $\alpha - 1$, where α is the number of conditions (levels) in a variable (factor). Here, $\alpha = 5$ and four pairwise comparisons are made, and thus # of pairwise comparisons is not greater than $\alpha - 1$ ($4 \ngtr \alpha - 1$ (4)).

this light, the low mean of self-reported speaking proficiency in the BLs and lack of difference between this value and that of the LL3 and LL4 speakers may not be surprising.

The mean value (% correct) for each condition on the two LexTALE (LexTALE and LexTALE-Esp) tests showed differing patterns (see Table 2.1). All conditions scored similarly on the LexTALE. To confirm this pattern, LexTALE scores were submitted to a one-way, between-subjects ANOVA with LexTALE as dependent variable and LE (LL1, LL2, LL3, LL4, BL) as the independent variable. The main effect of LE was not significant ($F(4, 45) = 0.63$, $p > .60$), suggesting that the LL and BL speakers do not have different vocabulary proficiencies in English overall. The pattern for LexTALE-Esp is noticeably different, however. The mean value (% correct) on the LexTALE-Esp is roughly at chance for the LL1, LL2, and LL3 speakers, but deliberate gains are made by the LL4 condition, and slightly more so by the BL speakers. This relationship was analyzed with a one-way, between-subjects ANOVA with LexTALE-Esp as the dependent variable and LE (LL1, LL2, LL3, LL4, BL) as the independent variable. The results suggest a significant main effect of LE ($F(4, 45) = 4.33$, $p < .01$) after applying Levene's Test for Homogeneity of Variance. Follow-up t tests for pairwise comparisons suggest that the differences between the LL1, LL2, and LL3 speakers' LexTALE-Esp scores and those of the BL speakers were all significant (LL1 × BL: $t(10.16) = -2.80$, $p < .02$; LL2 × BL: $t(9.00) = -2.86$, $p < .02$; LL3 × BL: $t(11.05) = -2.73$, $p < .02$). However, the difference between the LL4 speakers' LexTALE-Esp and that of the BL speakers was not significant ($t(14.67) = -0.76$, $p > .40$). These results suggest that, in terms of vocabulary development in Spanish as a second language, the participants in this study showed gains toward vocabulary proficiency similar to that of native early bilinguals during their fourth year of studying Spanish in the university.

Similar to the BL speakers' SRSP, the low LexTALE-Esp score (% correct) for the BLs ($M = 60$, $SD = 11$) may also be initially surprising. The LexTALE and LexTALE-Esp are visual lexical decision tasks that therefore rely on the participant's knowledge of formal orthographic rules in the language. However, this sample's experience with formal education in Spanish is likely limited due to their having been born and raised in the United States, a region in which English is the dominant language used for instruction in schooling. In fact, as reported previously, the BL speakers reported taking a similar number of Spanish classes overall (*M* = 2.7, *SD* = 2.9) as the LL1 speakers (*M* = 2.8, *SD* = 1.3). Given that the speakers between the two conditions have similar formal education in Spanish, and that their LexTALE-Esp scores are significantly different, it is clear that the BL speakers possess vocabulary intuitions in Spanish that the LL1 (and LL2 and LL3) speakers do not. Prior research on language development in bilinguals has shown that bilinguals are different from monolingual speakers in terms of vocabulary size and ease of lexical access, both in childhood (Hoff et al., 2012) and adulthood (Bialystok, Craik, & Luk, 2008), with monolinguals displaying a significant advantage.[5] Importantly, however, Hoversten, Brothers, Swaab, and Traxler (2017), in a study on visual word recognition in bilinguals, found LexTALE-Esp trends in their study similar to those of the present. Specifically, the early English/Spanish bilinguals reported in Hoversten et al. (2017), which were similar in linguistic experience to those of this study (early English/Spanish bilinguals leaning slightly toward English dominance) scored a mean of 63.5 (*SD* = 10.6) on the LexTALE-Esp, which is comparable to that of those BLs in this study (*M* = 60, *SD* = 11). In sum, prior research helps explain the seemingly low LexTALE-Esp score obtained by the BLs in this study.

In regard to language dominance, mean BLP scores generally display a trend toward decreasing English dominance as presumed experience with Spanish increases among the LLs (see Table 2.1). However, there is a noticeable drop of English dominance, moving toward more balanced bilingualism, in the BL speakers (cf., LL4's BLP [*M* = 118, *SD* = 27] and BL's

[5] The trends in the data show that this is true in the case of English for the participants in this study. In other words, the mean LexTALE for the BL speakers is noticeably lower than that of any LL condition; however, the difference as analyzed by ANOVA suggests that this difference is not robust.

BLP [$M = 50$, $SD = 28$]). A one-way, between-subjects ANOVA with BLP score as dependent variable and LE (LL1, LL2, LL3, LL4, BL) as independent variable was run to explore this pattern. The ANOVA reported a significant main effect of LE ($F(4, 45) = 23.32$, $p < .001$). Follow-up t tests for pairwise comparisons suggested that the differences between the BLP scores of the LL1, LL2, LL3, and LL4 speakers and that of the BL speakers were all significant (LL1 × BL: $t(12) = 7.55$, $p < .001$; LL2 × BL: $t(14.65) = 8.80$, $p < .001$; LL3 × BL: $t(15.64) = 6.35$, $p < .001$; LL4 × BL: $t(16.36) = 6.18$, $p < .001$). These findings indicate that the LLs of this study are different from the BLs in terms of English dominance and bilingualism, with the LLs showing a significantly greater English dominance and reduced measure of bilingualism than the BLs. A boxplot of the BL speakers' BLP scores, along with individual scores overlaid in black points and the condition mean in white, is found in Figure 2.2.

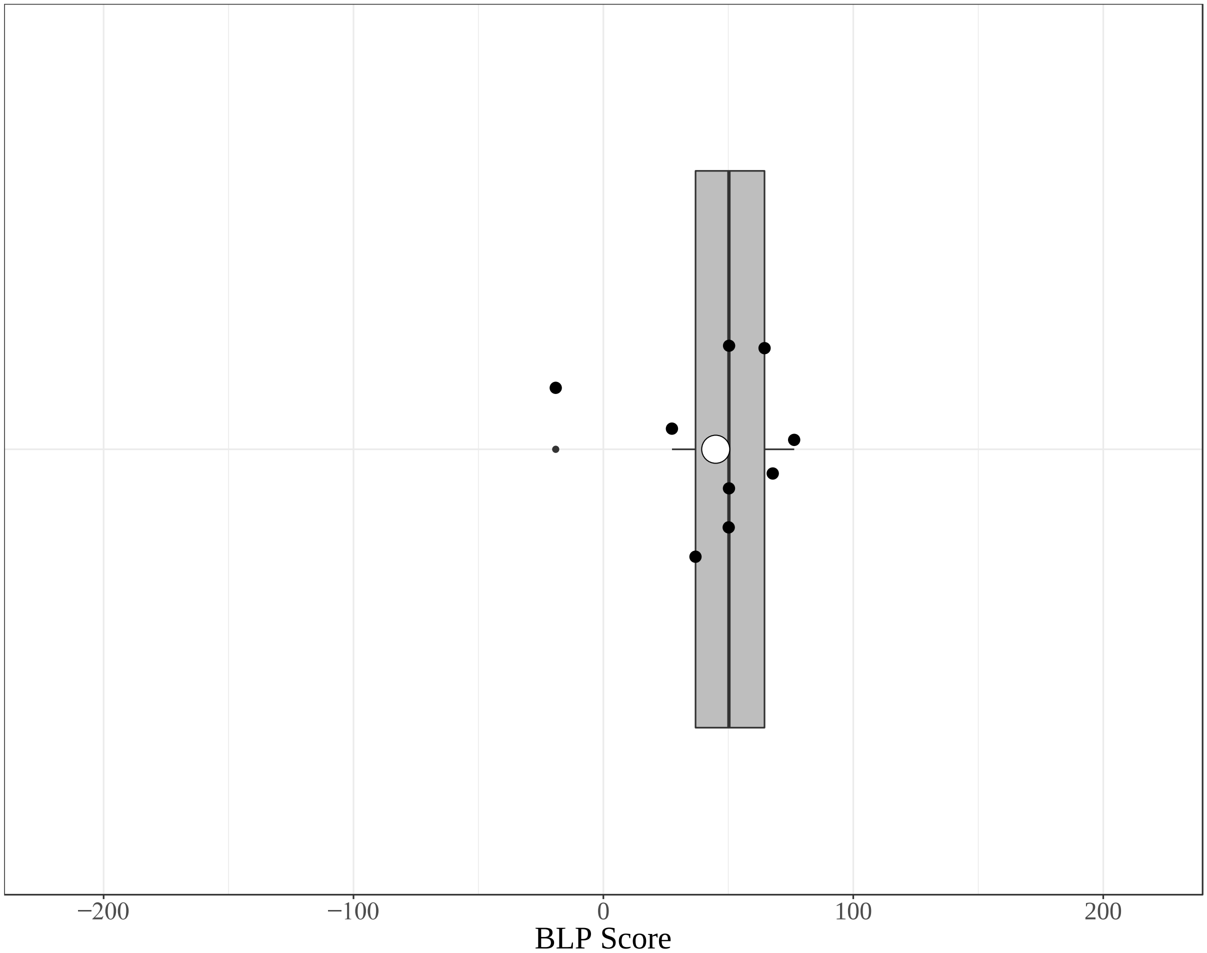


Figure 2.2 Boxplot of BL speakers' BLP scores overlaid with individual scores in black points and the mean in large white point. Along the *x* axis, positive values suggest English dominance, whereas negative values suggest Spanish dominance. A value at 0 suggests balanced bilingualism. The *y* axis is an arbitrary representation used to distinguish one point from another.

From Figure 2.2, one can see that most BLs are leaning English dominant, with the exception of one outlier, which is slightly Spanish dominant. Previous research has found a similar pattern of BLP scores in participants of similar linguistic background. The sample of early English/Spanish bilinguals reported in Casillas (2016), for example, included a majority of English-dominant and a few Spanish-dominant bilinguals. Thus, a sample consisting of a combination of native English/Spanish bilingual speakers with English-leaning and Spanish-leaning dominance is taken to well represent this population for the purposes of this study and the inclusion of both is considered justified.

2.1.8 Summary of Sampling Procedure

In summary, the present study includes five conditions—four learner conditions and one native bilingual condition—the participants assigned to which are all adults recruited from the student population at the University of Arizona: (a) The LL1 condition (*n* = 5), enrolled in a 2nd-semester Spanish class; (b) the LL2 condition (*n* = 12), enrolled in a 4th-semester Spanish class; (c) the LL3 condition (*n* = 9), enrolled in a 3rd-year Spanish class, (d) the LL4 condition (*n* = 15), enrolled in a 4th-year Spanish class, and (e) the BL condition (*n* = 9), comprised of native English/Spanish bilinguals. In regard to English learning and proficiency, all learner conditions began learning English at birth, whereas the BL speakers began around age four on average. None of the conditions differed in regard to their self-reported speaking proficiency in English, neither did they differ in their LexTALE scores (% correct).

In regard to Spanish learning and proficiency, there were robust differences between the LL and the BL speakers. For instance, all LL speakers reported learning Spanish in their adolescence, whereas the BL speakers reported learning Spanish at birth. The self-reported speaking proficiency in Spanish was low for all conditions; however, a trend toward increasing self-reported speaking proficiency as linguistic experience increased was observable and also confirmed statistically. Specifically, LL1 and LL2 speakers perceive their speaking abilities as significantly different (lower) than those of the LL3, LL4, and BL speakers, but the BL speakers' self-reported proficiency is greatest of all. In regard to LexTALE-Esp, the LL1, LL2, and LL3 speakers all scored at around chance levels, and significant gains in Spanish vocabulary proficiency were not made until the LL4 condition, whose score did not differ from that of the BL. Language dominance, as measured by the BLP, followed an expected pattern: generally, a decreased English dominance as linguistic experience in Spanish increased, and a sudden drop toward balanced bilingualism with the BL speakers, whose score was significantly lower from that of all the LL speakers.

In short, the descriptive and inferential statistics presented in this chapter describe the relationship between the participants and their linguistic experience. They also justify the sampling procedure and the method used to assign participants into study conditions. Importantly, the chapter details the way in which this dissertation will operationalize linguistic experience and/or development via a cross-sectional study.

CHAPTER 3

THE DEVELOPMENT OF L2 PLACE OF ARTICULATION

3.1 INTRODUCTION

The experiment presented in this chapter deals with the development of place of articulation in second language learners. Specifically, it analyzes the production of L2 Spanish coronal stops in native speakers of English to investigate the degree to which the participants developed a unique gestural target for their L2 voiceless coronal stop /t/. Importantly, the full voiceless stop category (/p, t, k/) from the English and Spanish production of learners is investigated. Additionally, the production of English and Spanish voiceless stops is analyzed in a sample of native bilingual speakers of English and Spanish as a baseline comparison. Gestural targets are operationalized via a series of spectral measures, including the relative intensity of the burst, and the center of gravity, standard deviation, skewness, and kurtosis of the distribution of energy of the burst of the stop consonant.

3.1.1 THE PHONOLOGY AND PHONETICS OF ENGLISH AND SPANISH STOPS

In this subsection, a brief characterization of the phonology of English and Spanish stop consonants is provided, followed by a detailed description of the phonetic characteristics relevant to the place of articulation of English and Spanish stops. Although the main focus of this chapter are the differences between English and Spanish voiceless coronal stop consonants, stops from three places of articulation (bilabial, coronal, velar) are analyzed in the experiment, and, consequently, coronal stops are discussed within the context of the full stop consonant inventory of both languages.

At the phonological level, it can be said that consonants differ as a matter of their laryngeal (voicing), manner, and place features (Chomsky & Halle, 1968). English and Spanish each have two categories of stop consonants ([+consonant]) that contrast as a matter of their voicing feature: /b, d, g/, which are [+ voice], and /p, t, k/, which are [– voice]. In both languages, /b/ and /p/ are [√ LABIAL], or, in other words, have a bilabial place of articulation, whereas /g/ and /k/ are [√ DORSAL], or, in other words, have a velar place of articulation. A difference is found, however, with the phonemes /d/ and /t/. Although in English and Spanish these stops share a [√ CORONAL] specification, the gestures required for their production are different in each language. On the one hand, in Spanish, the active articulator is the tip of the tongue, which touches the passive articulator—the back of the upper teeth (Hualde, 2005, p. 47; Quilis, 1993, p. 197). On the other hand, in English, the active articulator is either the tip or the blade of the tongue, which is then pressed against the passive articular—the alveolar ridge (Ladefoged, 2006, p. 11). Thus, Spanish has an apico-dental coronal stop, whereas that of English can be either apico- or lamino-alveolar.

These differences in voicing and place of articulation between stops, both within languages and cross-linguistically, can be observed phonetically via temporal and spectral cues, including: (a) voice onset time, (b) stop closure duration, (c) spectral peak, (d) spectral moments, (e), relative intensity, and, lastly, (f) formant transitions, all of which will be considered in the following paragraphs. Subsequently, evidence from empirical studies that measured these cues—cross-linguistically and in native bilinguals—to determine differences in place of articulation in phonologically similar stops will be reviewed.

Voice onset time (VOT), which is the measure of time in milliseconds between the release of the stop and the onset of periodic voicing (Lisker & Abramson, 1964), is a temporal correlate of stop voicing and an acoustic property that is the result of coordination between the supralaryngeal gestures produced by the active and passive articulators and the laryngeal gestures of periodic voicing produced by the vocal cords during the production of stop consonants. The measure can be positive or negative, and it is usually expressed in milliseconds (ms): In the case that the onset of voicing occurs after the release of the burst, a positive value is assigned to the duration of VOT, whereas a negative value is assigned if voicing begins before the burst. Stop consonants that have a short VOT value are also called short-lag stops, whereas those with a long VOT duration are termed long-lag stops. A negative VOT is also known as prevoicing, or that voicing was present during the stop closure, and stops that exhibit this behavior are also called long-lead or prevoiced stops (Lisker & Abramson, 1964, 1967).

Many languages of the world, including Spanish and English, exhibit a two-way contrast in which phonologically voiced (/b, d, g/) and voiceless (/p, t, k/) stops differ in terms of VOT. Spanish, for example, maintains a long-lead (/b, d, g/), short-lag (/p, t, k/) contrast, while English has a short-lag (/b, d, g/), long-lag (/p, t, k/) contrast (see Table 3.1). The voiceless English long-lag stops are produced with an audible aspiration, whereas the voiceless short-lag stops of Spanish are not. Thus, it can be said that, phonetically, voicing in each of these languages creates a distinction in which pairs of stops between voicing categories (e.g., /d/ and /t/) also differ as a matter of VOT: In Spanish, phonologically voiceless stops have short-lag VOT, whereas in English they have long-lag VOT (Table 3.1).

Table 3.1 VOT values (ms) for Spanish and English voiced and voiceless stops, as reported in Lisker and Abramson (1964).

	Voiced			Voiceless		
	/b/	/d/	/g/	/p/	/t/	/k/
English	1	5	21	58	70	80
Spanish	−138	−110	−108	4	9	29

Studies on the production and perception of stops in monolingual speakers have shown VOT to be an acoustic correlate of phonological voicing and place of articulation,[6] and that

[6] Throughout this dissertation, VOT will be considered without explicit mention of context; however, the default context, unless otherwise stated, is word initial, one of the few contexts in which VOT contrasts between phonologically voiced stops in English are found (Ladefoged, 2006). Ladefoged (2006) clarifies that, in word

it is affected by prosodic elements of speech. In a seminal paper, Lisker and Abramson (1964) formally presented VOT as a single metric for distinguishing phonological stop voicing categories among the world's languages, including English and Spanish. They found VOT to be a reliable metric for this purpose and, additionally, for distinguishing within-category stops, as well, as "the velars seem to have consistently higher values than the other stops" (Lisker & Abramson, 1967, p. 399). These facts have also been supported for Spanish in separate studies (Castañeda Vicente, 1986). A discussion of the reasons why VOT values increase as place of articulation advances farther toward the velum goes beyond the scope of this dissertation; however, this phenomenon has been well documented (for a review, see Abramson & Whalen, 2017) and is considered to be the norm among most speakers of most languages (Chodroff, Golden, & Wilson, 2019). In addition to production, VOT is also a relevant acoustic cue to the perception of stop voicing categories in English (Aslin, Pisoni, Hennessy, & Perey, 1981; Eilers, Wilson, & Moore, 1976) and Spanish (Abramson & Lisker, 1973).

Furthermore, multiple studies have shown that prosodic phenomena, including lexical stress and speech rate, affect VOT. Specifically, in English (Cole, Choi, Kim, & Hasegawa-Johnson, 2003; Simonet, Casillas, & Díaz, 2014) and in Spanish (Castañeda Vicente, 1986; Simonet et al., 2014), lexical stress has a lengthening effect on the VOT of long-lag (English) or long-lead stops (Spanish); however, it seems as though the VOT of short-lag stops is immune to prosody-induced lengthening (Beckman, Helgason, McMurray, & Ringen, 2011; Simonet et al., 2014). Specifically, Simonet, Casillas, and Díaz (2014) found, on the one hand, that the VOT of the long-lag, English voiceless coronal stop /t/ to be significantly longer in stressed (M = 78 ms, SD = 7 ms) than in unstressed (M = 71 ms, SD = 7 ms) syllables, whereas this same effect was not found for the short-lag, voiced coronal /d/. On the other hand, they

---

initial position in English, aspiration is contrastive (cf. *pie* [pʰɑɪ] and *buy* [pɑɪ]). Intervocalically, voiced stops surface as prevoiced (e.g., *a buy*), and voiceless stops become unaspirated after /s/ (e.g., *spy* [spɑɪ]) and in word-final position (cf., *nap* [næp] and *nab* [næːp]) (2006, pp. 56–57).

found that the VOT of the short-lag, Spanish voiceless coronal stop /t/ was not significantly longer in stressed than in unstressed syllables, but that the VOT of the long-lead, Spanish voiced coronal stop /d/ was. In terms of speech rate, an inverse relationship with VOT has been found: The slower the speech rate, the greater the VOT values vis-à-vis a faster speech rate, and this has been found for English (Kessinger & Blumstein, 1997; Magloire & Green, 1999; Schmidt & Flege, 1996) and Spanish (Magloire & Green, 1999), among other languages. Interestingly, the effect of speech rate on VOT appears to be similar to that of stress on VOT: generally the short-lag category stops go unaffected, whereas there is a robust effect for long-lag or long-lead stops (Kessinger & Blumstein, 1997; Magloire & Green, 1999). These findings relating to the effect of speech rate on VOT suggest that measures of VOT should be normalized to account for changes in speech rate between speakers and across conditions.

In sum, VOT is an acoustic correlate of phonological voicing among the world's languages.[7] It has been found to be a reliable predictor of between-voicing-category differences in production and perception, and likewise serves as a cue to place of articulation within voicing categories: As the place of articulation advances to a more posterior location, VOT values increase, such that those for velar stop consonants are predicted to be longer than coronal and/or bilabial stops for most speakers. Lastly, VOT is affected by prosodic phenomena such as stress and speech rate, although the effect appears to be less robust for short-lag stops than it is for long-lag or long-lead stops.

Similar to VOT, the timing of the closure interval of stop consonants has been proven to be a reliable acoustic correlate of stop voicing and place of articulation. The findings cross-linguistically generally agree that phonologically voiced stops have shorter closure durations than voiceless stops, in perception and in production, and this has been found to be true for English (Flege, Munro, & Skelton, 1992; Lisker, 1957; Port, 1979) and for Spanish (Álvarez

---

[7] I use the notion of phonological voicing in this dissertation to refer to the contrast in VOT values between homorganic stops that is found in many of the world's languages.

González, 1978; Martínez Celdrán, 1991, 1993; Navarro Tomás, 1918; Torreblanca, 1979).[8] Likewise, closure duration is an acoustic correlate of place of articulation within voicing categories. For English and Spanish, the literature has found a general trend in which the closure for coronal stops is shorter than that for velar stops, which is in turn shorter than that for labials (Dorman & Raphael, 1980; Luce & Charles-Luce, 1985; Martínez Celdrán, 1991; Navarro Tomás, 1918; Repp, 1984). In other words, labial stops consistently have longer closure durations than velar stops. In sum, closure duration and VOT have been found to be acoustic correlates of both stop voicing and place of articulation, findings that are relevant to English and Spanish; however, beyond these two temporal-acoustic correlates, there are also spectral correlates that are relevant to the distinction of place of articulation in stops.[9]

Spectral information, including that which is derived from the burst of the stop consonant—specifically, spectral peak (Jongman, Wayland, & Wong, 2000) and spectral moments (Watt, 2013)—distinguish stops of different places of articulation within and across voicing categories in native-language speech cross-linguistically (Casillas, Díaz, & Simonet, 2015; Forrest, Weismer, Milenkovic, & Dougall, 1988; Jongman & Blumstein, 1985; Sundara, 2005). Likewise, relative intensity (Casillas et al., 2015) and formant transitions (see Delattre, Liberman, & Cooper, 1955; Ladefoged, 2006; Port, 1979 for English; Quilis, 1993 for Spanish) can distinguish stops as a matter of voicing and place of articulation. These will be discussed at length in the following paragraphs.

A voiceless stop, such as the English /t/ and Spanish /t/ that are the focus of this study, have two sources of sound: a) the burst, which originates at the place of articulation, and b)

---

[8] In Spanish, voiced stops /b, d, g/ weaken phonetically in intervocalic position to the allophones [β, ð, ɣ], respectively (Hualde, 2005, p. 64). Most of the studies cited here that investigated closure duration differences between voiced and voiceless stop phonemes in Spanish did so in a context in which the voiced stop allophones do not weaken ([b, d, g]). However, Martínez Celdrán (1993) found that, even intervocalically, shorter closure durations led to a stop being identified as /b/ (i.e., [β]) instead of /p/.

[9] Other acoustic cues to phonological voicing also exist, such as onset f0, as documented in Dmitrieva, Llanos, Shultz and Francis (2015), among other papers.

the aspiration following the burst, derived from the glottis (Johnson, 2011, p. 174). Under the assumption that the vocal tract acts as an acoustic tube, the constriction that takes place during stop production divides the tube into two—a front and a back (Johnson, 2011, p. 177). The closer the constriction is to the edge of the mouth, the smaller the front tube becomes, and the reverse is true for the back tube. Thus, a lingual constriction in the dental region of the mouth (e.g., in the case of a dental stop) would create a smaller front tube than a lingual constriction at the alveolar ridge (in the case of an alveolar stop). These differences in tube length are reflected in the spectra produced by each one during the burst, respectively: The smaller the tube in front of the place of constriction, the higher the frequency peak of spectral energy, or *spectral peak* (henceforth SP), will be (Johnson, 2011, p. 176). Therefore, it would be expected that a stop with a dental place of articulation would have a higher SP than one with an alveolar place of articulation, which is slightly more posterior vis-à-vis a dental constriction.

The burst is short relative to the duration of the following aspiration of any given long-lag or short-lag stop, often lasting between two and three ms, followed by the onset of aspiration (Johnson, 2011, p. 174). Shortly after the release of the stop consonant, the space in front and behind the place of articulation become acoustically coupled and, therefore, the signal after the burst no longer includes noise originating from the place of articulation only but also noise proceeding from the glottis (i.e., aspiration) (Johnson, 2011, p. 175), suggesting that the first few milliseconds following the release of the stop consonant are the most relevant in terms of detecting place of articulation via spectral energy.

Interest in the spectral correlates of place of articulation, particularly of stops, is not new; indeed, it has been studied in depth, both qualitatively and quantitatively, across a variety of languages over the course of the past many decades (e.g., Delattre et al., 1955). Specifically, researchers set out to determine which, if any, acoustic features of the speech signal were invariant correlates to place of articulation. In a series of experiments in which judges classified stop consonants using the Fast Fourier Transform (FFT) spectra of the stop burst, based on a set of predetermined parameters derived from the distribution of spectral energy, it was found that, even visually, spectral data encoded in the burst of the stop consonant were sufficiently invariant—even when differing as a function of vocalic contexts, voicing categories, and speech rate—to accurately determine its place of articulation above chance (Blumstein & Stevens, 1979; Kewley-Port, 1983; Kewley-Port & Luce, 1984). Later, using *spectral moments* (discussed in the following paragraphs), researchers applied the findings from these qualitative studies to quantitative-based experiments whose end goal was to determine a systematic procedure for identifying the invariant acoustic properties of place of articulation from the burst spectra of stop consonants (Forrest et al., 1988; Kobatake & Ohtani, 1987).

Under the assumption that the distribution of spectral energy in the burst of a stop consonant could be taken to represent a probability density function (Keppel & Wickens, 2004, p. 133), Forrest, Wesimer, Milenkovic, and Dougall (1988) applied a mathematical measure called *moments* to the spectral data obtained from the burst of stop consonants. A moment is a metric derived from the density function of a distribution of continuous data. The first moment is the center of gravity (or mean), the second is the variance, the third is the skewness, and, lastly, the fourth moment is the kurtosis (Boersma, 2002a). Each moment is a function of a similar base formula, and each provides information about the shape of the distribution. The first moment, the center of gravity (henceforth COG) of the distribution, is a measure of "how high the frequencies in a spectrum are on average" (Boersma, 2007). The second moment, variance, has commonly been replaced (Forrest et al., 1988, p. 118) by a measure of standard deviation (henceforth SD), or the square root of the variance (Boersma, 2002a). The third moment, skewness (henceforth SK), is a measure of the asymmetry (Keppel & Wickens, 2004, p. 144) in the shape of the spectrum above and below the COG (Boersma, 2002b). Lastly, the fourth moment, kurtosis (henceforth KT), is a measure of the outliers of the distribution (Westfall, 2014). Spectral moments, in addition to SP, are considered *invariant* acoustic correlates of stop place of articulation: The information they provide is independent of "speakers, phonetic contexts, and languages" (Lahiri, Gewirth, & Blumstein, 2005, p. 391).

Beyond invariant correlates, context-dependent spectral cues, including relative intensity and formant transitions, also carry information relating to place of articulation. Relative intensity (henceforth RI) is the intensity (dB) of the burst relative to that of the following vowel (Casillas et al., 2015; Stoel-Gammon, Williams, & Buder, 1994; Sundara, 2005). With RI, the larger the value is, the smaller the intensity of the burst is, and the inverse is also true—the smaller the RI, the greater the intensity of the burst. Vowel formant peaks prior to and after the stop closure (in the case of intervocalic stops) are also relevant to the distinction of place of articulation. As the mouth closes, the shape of the tube changes, thus affecting the frequency of the formants. This movement of formant peaks across time before and after the stop closure are often referred to as formant transitions, or formant loci, and, because the spectral information obtained from them relies on the vowel that is being produced, which, in turn, has its own place of articulation, the type of vowel also affects the direction of the formant transitions (Johnson, 2011, p. 177).

The spectral measures summarized here—spectral peak (SP), center of gravity (COG), standard deviation (SD), skewness (SK), kurtosis (KT), and relative intensity (RI)—provide information about the shape of the distribution of energy of the burst of a given stop consonant. Given the facts based on the tube model regarding how energy is affected relative to tube size, it would be expected that the shape of the burst would vary along one or more dimensions as a function of place of articulation,[10] and this assumption has been born out in

[10] It's important to note that there is a difference between the measure of peak frequency of the burst and the spectral moments of the burst. As reviewed from Johnson (2011, p. 176), peak frequency should be higher in dental stops than in alveolar stops, but, on the other hand, this is not necessarily true for the moments of the burst. For example, as will be discussed in more detail, several studies have found that the mean frequency of the dental stop was lower than the mean frequency of the alveolar stop (as an example, see the Stoel-Gammon et al., 1994), opposite of what would be expected from the values of spectral peak.

several studies. Forrest et al. (1988) found that COG, SK, and KT together were able to accurately classify stop consonants in English with 92% accuracy in their study. Jongman and Blumstein's (1985) research on Malayalam, a language that contrasts dental and alveolar voiceless stop phonemes, found—using a study-specific metric that they call an *energy ratio*, which is a normalized value of the root mean square of spectral energy at the onset of the vowel and that of the burst of the stop consonant—that they were able to accurately identify over 90% of the tokens in their study as either dental or alveolar stops. In a series of work, Ali and colleagues (Ali, 1999; Ali, Van der Spiegel, & Mueller, 2001) developed an algorithm for the detection of place of articulation in stops in continuous speech in English. The results of their discriminant analyses found that SP was the most valuable correlate to detecting place; importantly, however, accurate detection of place of articulation increased drastically when spectral information of the following vowel was included, suggesting that there is no one singular invariant acoustic-phonetic correlate to place, but rather that humans likely use a combination of spectral data for discriminating place perception (Ali et al., 2001).

In a series of experiments in which synthetic stimuli were created based on the formant loci of stop consonants, Liberman, Delattre, and colleagues (Delattre et al., 1955; Liberman, Delattre, & Cooper, 1952; Liberman, Delattre, Cooper, & Gerstman, 1954; Liberman, Harris, Hoffman, & Griffith, 1957) found that the formant loci were indeed acoustic correlates of place of articulation in production and perception. Delattre, Liberman, and Cooper (1955), for instance, showed how the first (F1) and second formants (F2) are particularly useful in distinguishing place of articulation in English. When a bilabial consonant is followed by a low vowel, the F2 transition has a positive, or upward, slope. When a low vowel follows a coronal or velar stop, the F2 transition is negative, but the slope is more pronounced following a velar than a coronal. The ability of formant transitions to distinguish the place of articulation of stop consonants is, apparently, a cross-linguistic universal as understood by the two-tube model summarized previously, and, likewise, is also relevant to Spanish (Quilis, 1993, p. 209).

Several studies have done cross-linguistic research with monolinguals of languages that share phonologically similar phonemes (e.g. /t/) that differ phonetically as a matter of the gestural target (i.e. alveolar vs. dental place of articulation) to determine whether spectral measures are capable of differentiating the stops. Specifically, these studies have focused on the difference in place of articulation between the English alveolar /t/ and (a) the Swedish dental /t/ (Stoel-Gammon et al., 1994), (b) the French-Canadian dental /t/ (Sundara, 2005), and (c) the Spanish dental /t/ (Casillas et al., 2015). For instance, Stoel-Gammon, Williams, and Buder (1994), in their data for a coronal stop followed by a high back vowel [u], report that the SD of the energy of the burst was the most reliable predictor of differences in place of articulation between the alveolar /t/ of the American English speakers and the dental /t/ of the Swedish speakers. Likewise, Sundara (2005), in her data for a coronal stop followed by a front or back mid vowel, found, similar to Stoel-Gammon et al. (1994), that SD was the most reliable spectral predictor of place of articulation between these languages. Of special interest to this study is that of Casillas, Díaz, and Simonet (2015), who studied the acoustic correlates of the English alveolar /t/ and Spanish dental /t/. In this study, like the previous ones already reviewed, COG was greater for the alveolar stop than for the dental stop. Furthermore, using RI, COG, and SD combined, 87% of the tokens in a testing subset of the data were accurately identified as either a Spanish or English stop consonant by a logistic regression model.

Along similar lines, spectral measures have been used to investigate whether early bilinguals that speak two languages with phonologically similar stop consonants make a distinction in their gestural targets for each language, the findings from which are relevant to the study of place of articulation in second language learners, particularly native speakers of English learning Spanish as a second language. Sundara, Polka, and Baum (2006) studied simultaneous bilinguals of Canadian French and Canadian English, which, similar to Spanish and English, differ in stop consonant place of articulation and voicing categories: the coronal stops of Canadian French have a dental place of articulation, whereas they have an alveolar place of articulation in Canadian English. The specific consonants studied were the Canadian English /d/ and the Canadian French /t/, both of which are short-lag stops. The authors reported that the SD and KT measures were significantly different between the bilinguals' Canadian English and Canadian French dental stops. Furthermore, these same bilinguals were rated by native speakers of each language as sounding native like in the respective language. Taken together, these results suggest that the bilinguals reported in Sundara et al. (2006) produce language-specific gestural targets between the coronal stops in their two languages and that this difference is robust enough to be detected through spectral measures.

From the studies reviewed here, despite the relative proximity in regard to constriction location between alveolar and dental /t/, spectral measures are robust enough to detect differences in place of articulation both within languages and cross-linguistically. The spectral measures discussed (RI, COG, SD, SK, and KT) will be used, among others, as a means to measuring the development of the place of articulation of the Spanish dental /t/ in second language learners whose native language, English, has an alveolar /t/.

3.1.2 The Acquisition of Stop Consonants in a Second Language

3.1.2.1 Theories of Second Language Speech Acquisition

Over the past several decades, many theories have emerged that attempt to capture the nature of second language speech acquisition, at least of segmental phenomena. Of these theories, which touch on the interface between perception and production, some of the more prominent include Best's (1995) Perceptual Assimilation Model (PAM), Flege's (1995) Speech Learning Model (SLM), and Escudero's (2005, 2009; see also van Leussen & Escudero, 2015) Second Language Linguistic Perception (L2LP) model. Although overlapping points exist between these three theories mentioned, each model makes unique contributions to the understanding of the nature of second language speech in general.

3.1.2.1.1 The Perceptual Assimilation Model (PAM)

Best's (1995) PAM, originally proposed as a framework to be related to cross-linguistic speech perception in naïve listeners, was later extended to be applied to second language learners, and was given the title PAM-L2[11] (Best & Tyler, 2007). The PAM posits that learners are continually updating perceptual categories over their lifetime, both in the L1 and in the L2. It takes a direct-realist approach to speech perception and thus claims that the primitives, or basic units, of speech perception are the actual distal articulatory gestures themselves (Best, 1995)—including articulator coordination, and the place and degree of constriction of the articulation—and this information is extracted from the speech directly by the perceiver's knowledge of the human vocal tract (Best, 1995). Hence, the PAM claims that perceivers make use of lower-order gestural coordination and higher-order phonological categories for speech perception. The PAM posits that the L1 and L2 phonemes are stored in a shared phonological space; however, in regard to production, given that an L2 phoneme has been equated with that of an L1, the PAM claims that learners are capable of maintaining separate phonetic realizations for the merged L1/L2 phoneme.

The PAM makes hypotheses regarding the discrimination of nonnative sound contrasts. The first pattern is called Two-Category Assimilation, which occurs when each member of a nonnative contrast is assimilated to a native contrast, such as, for example, if an L1 English learner were to assimilate the L2 Spanish /p, b/ contrast directly to the corresponding English pair. This pattern is expected to portray excellent discriminability. Importantly, the PAM supposes that discriminating similar sounds in this context is relatively easy. The second pattern is Category-Goodness Difference, a two-to-one assimilation, in which both members of a nonnative contrast are assimilated to the same native phoneme. One of the members of the contrast is considered to be a better fit than the other, and one's ability to discriminate the

[11] For the sake of simplicity, the PAM-L2 is referred to here simply as the PAM.

two nonnative sounds is "moderate to very good" (Best, 1995, p. 195). The third pattern is Single-Category Assimilation, which, much like the Category-Goodness Difference, is a two-to-one assimilation; however, in this pattern, individual members of the contrast are discerned equally well or equally poorly and the contrast between the two will be poorly discriminated. The fourth pattern is Both Uncategorizable, in which neither member of the nonnative contrast is assimilated to a native phoneme. The perceiver's ability to discriminate this nonnative contrast will vary depending on how close the nonnative sounds are in relation to native sounds phonologically. The fifth pattern is Uncategorized versus Categorized, in which only one of the members of the nonnative contrast are assimilated to a native phoneme, whereas the other remains uncategorized. In this case, discriminability "is expected to be very good" (Best, 1995, p. 195). Lastly, the sixth pattern is called Nonassimilable, in which, as the name suggests, neither nonnative sound is assimilated, but rather both are considered to be non-speech sounds.

In short, in the PAM, which approaches second language speech learning from a contrastive approach, the primitives to perception are the articulatory gestures of speech, and the abstract representations of sound contrasts are stored in a shared phonological space.

3.1.2.1.2 The Speech Learning Model (SLM)

Flege's (1995) SLM, designed with experienced learners in mind, explains the "ultimate attainment of L2 pronunciation" (p. 238). Unlike the PAM, the SLM discusses L2 perception in terms of individual segments. At the model's base is the idea that "phonetic categories" (Flege, 1995, p. 239), composed of the phonetic properties of speech sounds for a given language, are stored in long-term memory. These dynamic phonological representations are updated over time to account for the variability observed in their production among speakers. Ultimately, the phonetic composition of an L2 sound is what drives perceptual acquisition.

The SLM lays out a series of postulates and derived hypotheses regarding the nature of sound categories, their formation, and their evolution. Flege postulates that the L1 and L2 sound categories are stored perceptually in the same phonological space, and, as a result, learners will "strive to maintain [a] contrast between [them]" (1995, p. 239). Likewise, cross-linguistic interference (CLI), or the effect that one language may have on the perception or production of another within the same speaker, may be bidirectional, with the L1 categories impacting those of the L2 and vice versa. Furthermore, Flege hypothesizes that the acoustic-phonetic makeup of a sound determines its perceptual location in relation to other sounds. Thus, sounds will be located in the phonological space more closely to those they resemble phonetically than to those from which they are dissimilar.

These postulates lead Flege (1995) to hypothesize at least two principal outcomes for speech perception. The first is that an L2 sound may be successfully perceived as different from the L1 sound most similar to it. In this event, the greater the perceived phonetic differences between the two sounds, the greater the likelihood for the successful creation of a new L2 category. However, the new L2 category, Flege proposes, may be dissimilar from that of monolinguals of the L2 if perceptual distance is created between the L2 sound and a similar L1 sound or if the acoustic cues employed by the speaker for identification are different from those used by monolinguals (Flege, 1995, p. 239). Furthermore, the older the individual is at the outset of L2 speech learning, the more likely it is that phonetic differences between L1 and L2 phonemes will not be perceived and, consequently, new category formation will not be achieved. The second potential outcome is that an L2 sound may be difficult to perceive and is considered equal to the corresponding L1 sound that it most closely resembles (Escudero, 2009, p. 168), and, consequently, the creation of a new category is "blocked by the mechanism of equivalence classification" (Flege, 1995, p. 239). In this event, what would have been two separate abstract representations for the sounds would fuse together to form a single, merged perceptual category. Lastly, perception and production are directly linked, such that, over time, how a sound is produced reflects how that sound has been perceived and stored categorically.

3.1.2.1.3 The Second-Language Linguistic Perception Model (L2LP)

A third prominent theory of second language speech acquisition is Escudero's (2005, 2009) L2LP model, which, while containing elements of the SLM and the PAM, presents a theory aimed at explaining the entire phylogeny of second language speech, from hypothesizing about the beginning state of the L2 phonology in naïve listeners to the very end state of acquisition. The L2LP proposes that, at the beginning stages of acquisition, L2 learners work from an L2 perception grammar that is a full copy of that of the L1, an idea known as the *full copying* hypothesis. Therefore, at least in the initial stages, the L2 is perceived entirely through the learner's linguistic experience and the lens of the L1 sound categories. Because the L2 perception grammar is a separate copy from that of the L1, it can be said that the L2LP affords two independent phonological spaces where canonical forms of the L1 and the L2 are maintained and developed separately and, therefore, will not influence one another. When learners do behave in ways intermediate between their L1 and L2 sound categories, it is said that both the L1 and L2 perception grammars are being activated simultaneously, not that one is influencing the other. The L2LP also makes claims regarding the interface between production and perception: The acoustic-phonetic information in the speech signal, which are the primitives of speech perception, drive L2 perception and, in turn, L2 production represents the state of the learner's L2 perception grammar. As the learner is exposed to increasingly more exemplars of the L2 sound categories, her own L2 perception grammar is updated such that, over time, "L2 learners can achieve native-like L2 competence" (Escudero, 2009, p. 183)

Similar to the PAM, the L2LP concentrates on nonnative contrasts and proposes patterns that are similar to those in the PAM, with some important deviations. Firstly, the

L2LP proposes the New Scenario, similar to the PAM's Single-Category Assimilation (Best, 1995), in which a nonnative contrast is phonetically most similar to only one native sound (2:1 assimilation). In this pattern of assimilation, which is considered difficult, the learner will either successfully create a new L2 category in the perception grammar or, alternatively, split the L1 category existing from the L1 copy. Secondly, learners may face a Similar Scenario, known in the PAM as Two-Category Assimilation, in which "L2 sounds … have the same number of counterparts in the L1 but have different production distributions" (Escudero, 2009, p. 171). For example, native speakers of English learning Spanish as a second language face the Similar Scenario with regard to stops: English and Spanish each have three voiceless and three voiced stops, but they differ from one another phonetically. In this medium-difficulty pattern, the learners may adjust the parameters of the L1 contrast (existent in the full copy L2 perception grammar) to match those of the L2 contrast. Under this premise, L1 English speakers, for example, using the copy of the L1 /b, p/ short-lag, long-lag contrast, make adjustments over time so that the sounds more closely resemble the L2 Spanish /b, p/ long-lead, short-lag phonetic properties. Lastly, learners may face a Subset Scenario, in which a single L2 sound undergoes multiple category assimilation to two or more L1 sound categories. This pattern is hypothesized to be relatively easy for the listener, as the L2 perception grammar remains unaffected.

In summary, the L2LP traces the perceptual speech learning of individuals ranging from the very beginning stages of development to the final state of learning. Similar to the PAM, it concentrates on the assimilation of nonnative contrasts, proposing three patterns of potential perceptual learning: i) the New Scenario (2:1 assimilation), ii) the Similar Scenario (2:2 assimilation), and iii) the Subset Scenario (1:2+ assimilation). Unlike the PAM, the primitives of speech perception are the acoustic signal. This model maintains a *full copying* hypothesis in which the initial state of the L2 perception grammar is a copy of that of the L1;

thus, learners initially perceive their L2 through the lens of their L1, but they develop the L1 and L2 sound systems independent from one another.

3.1.2.2 Summary of theories

Whether the primitives to speech perception are the articulatory gestures (Best, 1995), or whether the abstract representations are stored in two separate (Escudero, 2005) or one single phonological space (Flege, 1995), goes beyond the scope of the present study. However, a key point from these theories are the hypotheses that they make and how they relate to outcomes in the development of second language speech. When taken together, generally speaking, these theories agree that the abstract categories that have formed in the phonology of the learner can be observed via second language speech production. Likewise, in one way or another, some level of phonological interference from the L1 to the L2 is at play, particularly in sounds that have the potential to be perceived as equal between the two languages, and this is expected to be observed in perception as well production.

Another common thread among these theories (with the exception of the PAM) is that the primitives to speech perception are derived from the acoustic signal. Due to their orthographic, acoustic, and gestural similarities (and, particularly important, especially among the latter two, are the ways in which these differ) among many of the languages of the world, stop consonants have often been used as a medium for studying the development of abstract segmental categories in a second language. There are multiple acoustic correlates that are relevant to place of articulation in these sounds, and learners may exhibit an acquisition of stop phonemes via a target-like, or near-target-like, production and/or perception of these correlates more quickly than others (Bohn, 1995; Flege, Bohn, & Jang, 1997). In most cases, the acoustic correlates of stop consonants are the product of the coordination of gestures. For example, VOT is an acoustic correlate of stop voicing and the product of the coordination of laryngeal and supralaryngeal articulatory gestures. And, although the acquisition of *stops* per se is not the focus of this experiment, but rather the acquisition of differing articulatory targets in similar stop phonemes between languages, a review of the literature that has used the acoustic correlates of stop phonemes heretofore discussed is needed.

3.1.2.3 Empirical evidence

A great body of the literature on second language speech has concentrated on the acquisition of voicing contrasts. Although voicing does not form a principal part of the questions relating specifically to this dissertation, it is relevant to them in several ways. Firstly, voicing traditionally has been used to measure the acquisition of L2 segmental phenomena, particularly stops, as does this dissertation, and, consequently, the effect that language experience and linguistic phenomena have on the acquisition of voicing categories in second-language learners can inform on expected outcomes in the experiments in this study. For instance, some of the literature has researched the effect of prosodic phenomena, such as stress and speech rate, on the production of voicing categories. Secondly, other research has been framed around extralinguistic factors that may lead to the successful acquisition of voicing contrasts, such as the level of instruction, in the case of classroom learners, or the age at which L2 learning took place, in the case of non-classroom learners. These findings are relevant to the contributions of this study, which presents data on the phonological development of an L2 stop phoneme (/t/) that is similar to that of the L1, but which differs slightly as a function of its gestural target.

Cross-sectional data have provided some insight into the nature of the development of the acquisition of L2 Spanish stop voicing differences, operationalized via the measurement of VOT, both in and outside of a traditional classroom setting. Reeder (1998) studied the acquisition of the production of L2 Spanish stops in students of Spanish as a second language in a traditional foreign-language classroom across four different levels of university instruction: 1) beginning learners in their first semester of instruction, 2) intermediate learners in their third semester, 3) advanced learners enrolled in upper-division or graduate courses, and 4) highly advanced learners who were lecturers at the university. The study found that mean VOT values in production significantly improved only between the intermediate and advanced learners. Thus, this study reveals that changes in L2 stop production may occur in between the intermediate and advanced stages of classroom instruction, at least for English-speaking learners of Spanish, and that gains are made in conjunction with linguistic experience.

Zampini (1998) studied the production of L2 Spanish stops in participants enrolled in a Spanish phonetics course. The participants' production was recorded in a reading task that was administered at three separate times over the course of the semester. The results indicate that the participants began the semester already at an advanced level, as their production of the L2 Spanish /p/ was close to native like. Notably, the results revealed that the difference between the L1 English /b/ and L2 Spanish /p/ production, while significant at week 3 and 6, did not differ at week 15. Zampini, interpreting these results in the SLM (Flege, 1995) framework, suggests that, although the participants may have started out with two different categories for English /b/ and Spanish /p/, these two categories over time became equated with one another and merged into one. In short, Zampini found, at least in voiceless stops, consistent improvement of L2 Spanish stops in production over the course of a semester.

Casillas (2016) collected production data over the course of eight experimental sessions in a longitudinal study of the acquisition of L2 Spanish stops in English-speaking beginner learners enrolled in a domestic immersion program. The results of the production study indicate that learners significantly improved their production of most L2 Spanish stops each week over the course of the program. When comparing the production data to that recorded at the initial stage of the program, it was found that a significant improvement took place beginning at week three. However, despite the gains made in production accuracy, the VOT of most of the L2 Spanish stops in the participants' final productions were still significantly different from that of an English/Spanish bilingual control condition. These findings—which suggest that the speech of L2 learners, both initially and over time, exhibited interference from their L1 in their production of L2 stops—support the SLM's (Flege, 1995) and PAM's (Best, 1995) accounts of a shared phonological space, and, in the language of the L2LP, the idea that both the L1 and L2 perception grammars can be activated simultaneously (Escudero, 2009). In sum, there is evidence to suggest that linguistic experience is a factor in L2 segmental acquisition.

Thus, the state of the literature at present is informative in regard to what is known about the acquisition of L2 stop phonemes in terms of their voicing distinction, even though only some of the information most relevant in this respect has been summarized here. From longitudinal and cross-sectional data, it has been shown that, generally speaking, linguistic experience is a factor in the outcomes of L2 speech acquisition (Flege & Eefting, 1987; Flege, Frieda, Walley, & Randazza, 1998), but that, over time, L2 learners may eventually begin to differentiate between their L1 and L2 stop inventories in production (Casillas, 2016; Reeder, 1998; Zampini, 1998). Studies investigating the acquisition of L2 stop phonemes as measured by the acoustic correlates of stop place of articulation, such as closure duration, formant transitions, and spectral measures, are rare, however, and, to this author's knowledge, no published research exists currently regarding the development of place of articulation of L2 stops that differ slightly in regard to gestural target from that of the L1. The questions driving this experiment are the following:

1. Are learners able to acquire unique gestural targets in their L2 stops? Specifically, do native speakers of English, whose voiceless coronal stop /t/ has an alveolar place of articulation, learn a new articulatory target for the L2 Spanish voiceless dental stop /t/?
2. If so, what is the pattern of development of the gestural distinction and how does it differ, if at all, from the pattern of development of other characteristics of stops, including voicing?
3. Lastly, are gains made toward target-like production as linguistic experience increases?

One can extrapolate from the literature reviewed here and the hypotheses set forth in the PAM, SLM, and L2LP that it is possible that learners of a second language can develop patterns of a unique gestural target in the second language for a stop phoneme that is similar in the L1 and L2. It is hypothesized that, at the initial stages of learning, the L2 perception grammar of native speakers of English learning Spanish will still be very similar to that of the L1. Thus, their production of L2 Spanish dental /t/ will likely resemble that of their L1, without much distinction in gestural target. However, it would be expected that the more linguistic experience and exposure to the second-language an individual has, and therefore the more opportunity one has had for category adjustment in the L2 perception grammar from exposure to the L2, the more likely it is that she will produce the Spanish /t/ with a gestural target different from that of her L1 in the direction of native speakers of the target language.

In addition to the production of English /t/ and Spanish /t/, the learners' production of English and Spanish bilabial and velar voiceless stops (/p, k/) is also analyzed. As a hypothesis, under the assumption that the bilabial and velar stops between English and Spanish have similar gestural targets, the place of articulation of /p/ and /k/ between English and Spanish will not differ among the learners, regardless of linguistic experience and gains made toward target-like production of other acoustic correlates of stops, such as VOT and closure duration.

Similarly, the production of English and Spanish voiceless stops is investigated in native English/Spanish bilinguals. The literature reviewed previously (i.e., Sundara et al., 2006) provides a hypothesis for the outcomes of these speakers: Native bilinguals will exhibit language-specific patterns of place of articulation between their English and Spanish coronal stops, which differ as a matter of gestural target. Conversely, under the same assumption as for the learners, the place of articulation for the bilabial and velar stops will not differ between English and Spanish.

The field of second language speech has benefited greatly from research on the acquisition of voicing distinctions cross-linguistically in stops that are similar phonologically; however; the findings from this study will contribute important data on the nature of the development of place of articulation in learners whose second language has a stop phoneme that is conceptually similar to that of its L1, but which differs fundamentally in its gestural target.

3.2 METHOD

3.2.1 PARTICIPANTS

The participants, method of recruitment, and sampling procedure for this experiment are reported in Chapter 2.

3.2.2 MATERIALS

The controls employed in the materials for this experiment closely modelled other previous research on the invariant spectral properties of place of articulation (Casillas et al., 2015; Sundara, 2005). A list was created of real English and Spanish words that began with one of the three phonologically voiceless stops of English and Spanish, differing as a function of place of articulation: /p, t, k/. The words were controlled for stress, such that about half of the materials had lexical stress on the first syllable and the other half had lexical stress on the second syllable. Similarly, none of the words chosen contained secondary stress. Because acoustic correlates of place of articulation of stops are often affected by the preceding or following vowel, the following controls were implemented: The target vowel following the initial, target stop, was, in both English and Spanish, a low vowel. Spanish has only one low vowel, /a/, which is central, whereas English has two: /æ/, a front vowel, and /ɑ/, a back vowel. The vowel preceding the target stop was controlled by inserting the target word into the carrier phrase *I say ______ again.* Primary stress, secondary stress, and the vowel employed in American English were confirmed by consulting Upton and Kretzschmar Jr. (2017).

Furthermore, an effort was made to exclude items that were (a) cognates between languages and/or (b) words in which spirantization (Harris, 1984), or the lenition of phonologically voiced stops to fricatives, may occur. These controls made it rather difficult to find an equal number of words between language (English, Spanish), consonant (/p, t, k/), and stress (stressed, unstressed) conditions. Nevertheless, fifty unique words in Spanish and 59 unique words in English were included in the experiment, for a total of 109 different items. The full list of materials is found in Table 3.2.

Table 3.2 Materials for the experiment on the development of place of articulation, as a function of Language (English, Spanish), Consonant (/p, t, k/), and Stress (Stressed, Unstressed).

	/p/		/t/		/k/	
	Spanish					
	Stressed	Unstressed	Stressed	Unstressed	Stressed	Unstressed
1	pacto	papel	tanque	tamaño	casa	camión
2	paso	pasado	tanto	también	cambio	canción
3	pata	patente	taza	tampoco	campo	candela
4	patria	pantera	tacha	tapioca	canto	capricho
5	paño	pandilla	tambo	tacada	capa	cascada
6	paja	pastoso	tapa	tapamos	casco	casete
7	panza	patriota	tacto	tapete	caso	castaño
8	Paco	pasillo	tajo	tacaño	cancha	casero
9					cana	camisa
	English					
	Stressed	Unstressed	Stressed	Unstressed	Stressed	Unstressed
1	pacify	pontific	tabernacle	tomfoolery	cabbage	cascading
2	package	pontoon	tabloid	tattoo	canopy	campaign
3	paddle	posterity	tacit	taxation	capsize	cashier
4	padlock	pandemic	tackle	tactician	capsule	cockade
5	pageant	paprika	tandem	tangential	captain	confound
6	pancake	Pangaea	tamper	Tasmanian	caption	cosmetic
7	pander	pomposity	tantrum	toxicity	captive	canteen
8	passable	passivity	tangent	taxonomy	casket	cavalier
9	patty	Pompeii	tacky	tactility	casting	caffeine
10	passion		taffy		castrate	captivity
11						Kathleen

3.2.3 RECORDINGS

The recording of materials took place after the completion of the background questionnaire and BLP (see Chapter 2). The English-language materials were recorded first, after completing the LexTALE and following the recording of English-language materials for the experiment reported in Chapter 4. The Spanish-language materials were recorded second, after completing the LexTALE-Esp and following the recording of the Spanish-language materials for the experiment reported in Chapter 4. In other words, for all participants, the direction of recording was the following: firstly, English-language materials; secondly, Spanish-language materials.

The materials were presented in a different, random order to each participant using PsychoPy 2 (Peirce et al., 2019). In all cases, for the English and Spanish portions of the experiment, the instructions were written in English. This was done to ensure that all LLs of all proficiency levels that were taking part in the experiment were able to clearly understand the task. In the instructions provided for each language condition, the participants were told that the materials that they were about to read were in English or Spanish, respectively. However, the participants were not given access to the materials prior to the experiment, nor were they coached on how to say them. The participants were asked to read each item aloud at a natural pace, each of which was presented on the screen for four seconds, followed by a red fixation cross that appeared in the center of the screen for 500 milliseconds. Each item was repeated three times and, therefore, each participant provided 327 tokens to the data set ((50 Spanish items + 59 English items) × 3 repetitions) for a study total of 16,350 tokens (327 tokens × 50 participants). However, the second repetition was the only token of the three repetitions that was submitted to analysis (unless there was a mispronunciation, in which case the third repetition was submitted in its stead). Thus, the reduced data set consisted of 5,450 tokens (50 participants × 109 items × 1 repetition).

All recordings were carried out in a sound-attenuated booth using a head-mounted, dynamic microphone and a Fostex DC-R302 digital recorder with built-in microphone pre-amplification and analog-to-digital interface. The signals were digitized at 44.1 kHz, 16-bit quantization.

3.2.4 Segmentation of Acoustic Data

The acoustic data were segmented by hand in the speech-signal processing software Praat (Boersma & Weenink, 2018) using a combination of spectral information and waveform shape to guide the segmentation. Using the synchronized spectrogram and waveform display, four landmarks separating three intervals were placed on a tier, as follows:

- *Onset of stop closure:* Marked at the offset of the preceding vowel at the point at which a noticeable decrease in F2 intensity, corresponding with a decrease in waveform height, was observed.
- *Onset of stop burst:* Marked at the point at which a noticeable and abrupt upward departure was observed, associated with the onset of aperiodicity in the waveform.
- *Onset of vowel:* Marked at the zero crossing of first upward departure in which modal voicing and formant structure were observed.
- *Offset of vowel:*
  - *Before nasals:* The point at which a decrease in formant intensity and waveform height was observable.
  - *Before fricatives:* The point at which waveform structure resembled aperiodic noise, corresponding with an end of formant structure.
  - *Before stops:* The point at which a noticeable drop in formant structure and waveform height was observable, just prior to closure.

Some tokens, or the data associated with a specific acoustic correlate, were removed due to mispronunciations or disfluencies. For example, tokens during which the speaker yawned, coughed, and so forth, were excluded from the data. Likewise, because prevoicing can affect the quality of the burst (Sundara, 2005), the relatively small number of tokens (n = 48) in which the stop consonant was prevoiced was removed from the data set. Prevoicing was determined firstly through visual inspection during the original segmentation process. To ensure that all prevoiced tokens were eliminated, prevoicing was double checked via a Praat script in which the F0 (Hz) at the onset of the fourth quartile of the closure leading up to the stop was calculated. Any token whose F0 at this point in the closure was greater than 75 Hz (Ladefoged, 2003, p. 81) was re-inspected for prevoicing and removed, if necessary.

Some speakers paused between saying *I say/Digo* and the target word. In the relatively few cases in which these pauses were due clearly to a disfluency, the stop closure duration was not calculated for that token. In the cases in which the speaker exhibited a seemingly long stop closure duration and it was not possible to objectively identify a pause as being due to a disfluency, the stop closure duration was calculated. However, as is described in the following subsection, outliers were later removed prior to data analysis. Similarly, some speakers hesitated during the enunciation of the target word, paused briefly, and then, without starting at the beginning of the carrier phrase again, repeated the target word. In these cases, stop closure duration could not be calculated.

Lastly, to conclude, a few remaining notes on the data segmentation are necessary. Firstly, some speakers, particularly LL speakers, appeared to stress the word on the incorrect syllable for a small number of tokens in English and Spanish. Nevertheless, in the few cases in which this did occur, the token was maintained in the data set, even though the materials were controlled for stress, in order to capture the actual speech production of the learners without biased judgments from the investigator. Lastly, as noted by Fischer-Jørgensen (1954), velar stops have a tendency to exhibit two distinct releases coupled close together. In the case of a double release in a velar stop, the first of the two was considered the onset of the burst in all cases. An example of the segmentation procedure is available in Figure 3.1.

3.2.5 CODING AND STATISTICAL ANALYSIS

In total, after eliminating tokens for the reasons described previously, a total of 5,354 tokens were submitted to the analysis ((50 participants × 109 items × 1 repetition) – 96 misses). Prior to coding, the digitized sound files were low pass filtered at 11.025 kHz (Casillas et al., 2015). From the data, several acoustic correlates were extracted for each participant, as follows: Firstly, three rhythm measures were calculated: closure duration (CD), VOT, and vowel duration (VD). CD was calculated as the difference between the onset of the burst and the offset of the preceding vowel, VOT was the difference between the onset of modal voicing and the onset of the burst, and VD was the difference between the offset of the vowel and the onset of modal voicing. Prior to analysis, the data were converted to a relative value to account for tempo differences from token to token and across speakers. The relative value of each measure is equal to its duration (ms) divided by the sum of the duration of the CD, VOT, and VD, as follows:

$$Rx = \frac{x}{closure\ duration + vot + vowel\ duration}$$

In other words, for a given acoustic measure x (i.e., VOT, CD, VD), relative x (Rx) is equal to the value of x (ms) for a given observation divided by the sum of the CD (ms), VOT (ms), and VD (ms) of the CV structure from which x was measured (e.g., CV{…} in *caption*) from the same item.

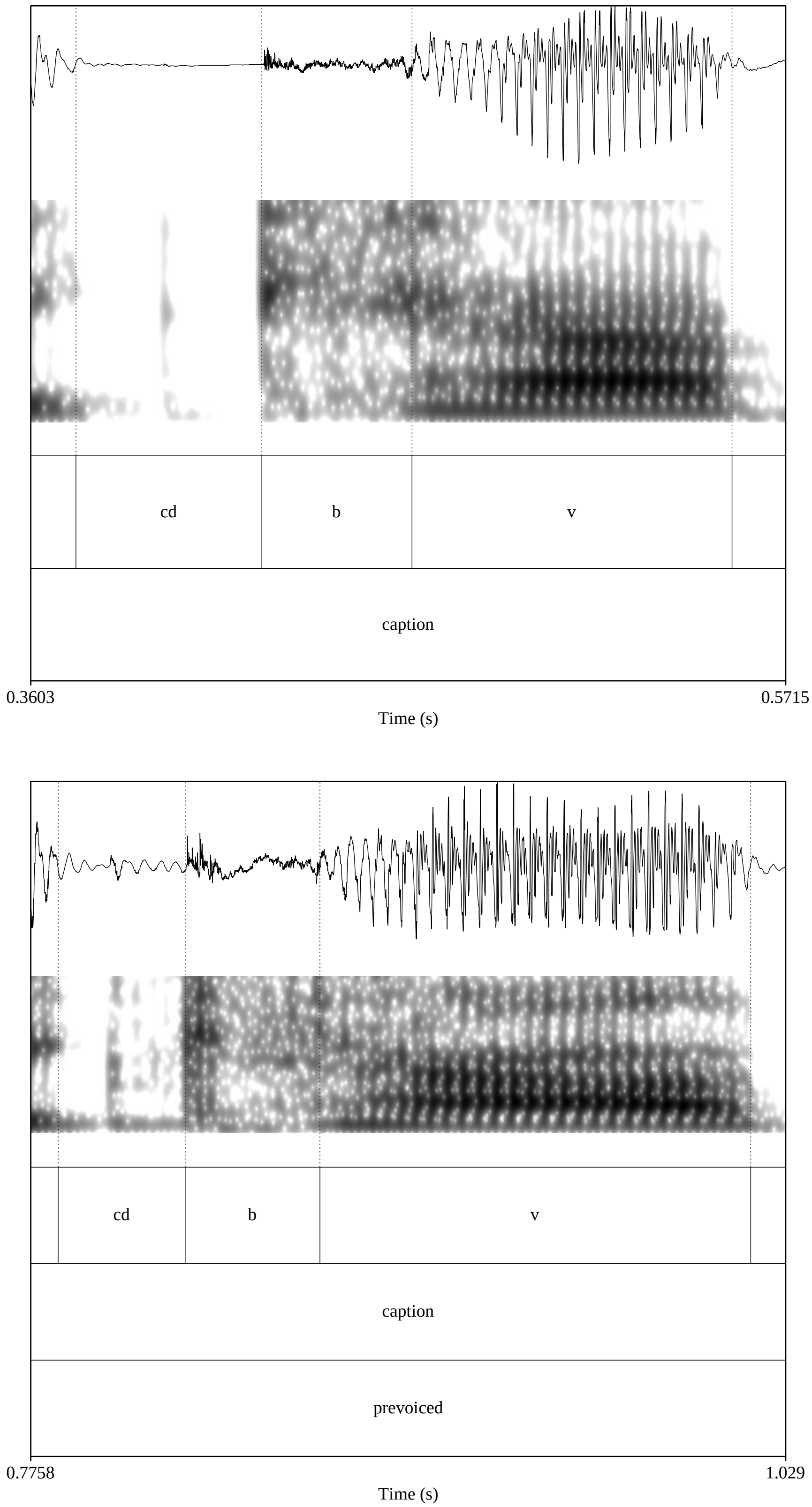


Figure 3.1 Illustrative example of data segmentation of the word *caption*. (Above) Four landmarks creating three distinct intervals—measuring closure duration ("cd"), voice onset time ("b"), and vowel duration ("v"), were coded in a TextGrid in Praat with the aid of a synchronized spectrogram and waveform display. (Below) Tokens with prevoicing were eliminated from the data set prior to analysis.

Secondly, spectral measures derived from the burst, as reviewed in the Introduction, were calculated as follows: The burst was calculated automatically for short-lag stops of 10 ms or less as being equal to the duration of VOT, whereas, for those stops whose VOT was greater than 10 ms, the duration of the burst was considered to be equal to 10 ms. The spectral moments (COG, SD, SK, and KT) were calculated from the envelope of the burst using an FFT of the spectra. The RI of the burst was calculated by subtracting the intensity (dB) of the burst from that of the vowel. The second formant transitions (F2T) were calculated by obtaining the value in Bark at the onset of the vowel subtracted from that of the midpoint of the vowel. Lastly, vowel quality (F2 Bark) was determined by obtaining the second formant value in Bark at the midpoint of the vowel.

The data were then submitted to linear regression models that were fit using the *lm* function in R (R Core Team, 2015) via the RStudio (RStudio Team, 2016) graphical user interface. Unless specified otherwise, the data were averaged over items as a function of Language (English, Spanish), Consonant (/p/, /t/, /k/), and Stress (Stressed, Unstressed) prior to analysis, such that the data for each of the 50 participants were aggregated down to 12 observations (2 × 3 × 2) for each participant. Similarly, the default baseline for the regression models was set to Language = English, Consonant = /t/, and, if included, Stress = Stressed. To determine differences between languages for /p/ and /k/, separate models were run with the reference level of Consonant changed to the corresponding phoneme. In all the models that were fit with the LL data, the LE explanatory variable was coded numerically in the following manner: LL1 = –1.5, LL2 = –0.5, LL3 = 0.5, LL4 = 1.5. The adjusted R^2 is reported for all models.

Following the analysis of the individual acoustic correlates mentioned, a separate analysis was conducted to determine which of the spectral measures were the best predictors of differences in place of articulation between the alveolar English /t/ and dental Spanish /t/ in the BL speakers. A data set for training, comprised of 75% of the original BL English and

Spanish /t/ data, chosen randomly, and a testing data set, comprised of 25% of the original data, were selected randomly. Then, a hybrid forward–backward selection logistic regression (James, Witten, Hastie, & Tibshirani, 2017, p. 210) was run with the English and Spanish /t/ data coded dichotomously (English = 0, Spanish = 1). The model selected from the training data set was then run on the testing data set to determine predictability. This same procedure was done for the /p/ and /k/ data.

3.3 RESULTS

The results for the BL data are presented first in order to have a baseline understanding of the acoustic correlates of place of articulation in adults that grew up speaking two languages whose phonological voiceless /t/ differ phonetically as a matter of gestural target, namely English, which has an alveolar /t/, and Spanish, which has a dental /t/. Following the BL results, the results for the LL speakers are presented to gauge the nature of the development of place of articulation in adults who grew up speaking English, a language with a phonetically alveolar stop, and who, later in life, learned to speak Spanish, a language with a phonetically dental stop.

The descriptive and inferential statistics of the data are presented in tabular form and are summarized in the text. An initial fitting of the linear regression models indicated that, for many response variables, the coefficient of the factor Stress (Stressed, Unstressed) was not significant; therefore, in these cases, this factor was removed, and the models were fit a second time, and only the second model is reported. Lastly, although the Consonant × Language (× LE) interaction was included in each model that was fit to allow for an interpretable Language (Spanish) (× LE) coefficient vis-à-vis the baseline, its indication of the presence or absence of a significant change from the English baseline stop to a different Spanish stop is irrelevant to the questions of this study and therefore is not discussed.

3.3.1 Results: BL Speakers

3.3.1.1 Voice onset time, closure duration, vowel duration, and F2 Bark

3.3.1.1.1 Descriptive statistics

The data presented in Table 3.3 suggest that, generally, the BL speakers' production of stops and vowels patterns in the direction that would be expected for native bilinguals of English and Spanish. On the one hand, the VOT for both English and Spanish follow a consistent pattern intralingually: as the place of articulation progresses to a more posterior location, the VOT values increase. Interlingually, the English and Spanish VOT values display the expected difference: All VOT values in the English voiceless stops are apparently greater than those of the corresponding Spanish voiceless stops. On the other hand, cross-linguistically, the CD of the Spanish voiceless stops appears to be longer overall than that of the English voiceless stops. Intralingually, the bilabial stop had a longer CD than the velar stop. Lastly, in terms of vowel quantity and quality, the duration of the vowel following the target consonant appears to be greater in stressed syllables in English and in Spanish, and the duration of the vowel in Spanish is evidently longer overall than in English. Conversely, the vowel quality, as measured by F2 Bark, is apparently greater in English, implying a more fronted vowel than in Spanish. In English, it is greater in stressed syllables, whereas it is greater in unstressed syllables in Spanish.

Table 3.3 Descriptive statistics (mean with standard deviation in parentheses) of (left) VOT and Closure Duration (CD) as a function of Language (English, Spanish) and Consonant (/p, t, k/) and of (right) Vowel Duration (VD) and F2 Bark as a function of Language (English, Spanish) and Stress (Stressed, Unstressed) for the BL speakers. Boxplots are provided in Appendix G.

	VOT		CD			VD		F2 Bark	
	Eng	Span	Eng	Span		Eng	Span	Eng	Span
/p/	0.23	0.05	0.35	0.47	*Stressed*	0.43	0.49	12.06	11.14
	(0.07)	(0.02)	(0.08)	(0.05)		(0.06)	(0.05)	(0.92)	(0.72)
/t/	0.27	0.05	0.31	0.49	*Unstressed*	0.40	0.45	11.67	11.39
	(0.07)	(0.03)	(0.09)	(0.06)		(0.08)	(0.06)	(1.37)	(0.72)
/k/	0.28	0.10	0.30	0.42					
	(0.06)	(0.04)	(0.08)	(0.06)					

3.3.1.1.2 *Inferential statistics*

Separate linear regression models were fit using the VOT and CD data for the BL speakers as the response variable and the explanatory factors Stress (Stressed, Unstressed), Consonant (/p, t, k/), and Language (English, Spanish), and the Language × Consonant interaction. As observed in Table 3.4, among both models, these factors combined predicted a significant amount of variance in the response. The Stress (Unstressed) coefficient was not significant, meaning that VOT was significantly greater in stressed syllables than in unstressed in English. Compared to the baseline of the corresponding phoneme in English, the Language (Spanish) coefficient was significant for all consonants for VOT, each with a negative *b*, suggesting that the VOT of each of the Spanish voiceless stops was significantly less than the corresponding English phoneme. Similarly, for CD, the Language (Spanish) coefficient was significant for all consonants, but in the opposite direction: The CD of all voiceless Spanish stops was significantly longer than the CD of the corresponding consonant in English. Importantly, a separate model was run in which Consonant was coded numerically (/p/ = –1, /t/ = 0, /k/ = +1) in order to analyze whether CD changes within each language. A regression model was fit with CD as the response and Consonant (coded numerically), Language (English, Spanish), and the Consonant × Language interaction as the explanatory variables. These variables combined predict a significant amount of variance in the response (R^2 = .66, $F(3, 104) = 69.03$, $p < .001$). With the baseline set to either English or Spanish, the Consonant coefficient was significant (English: $b = -0.03$, $t(104) = -3.46$, $p < .001$; Spanish: $b = -0.03$, $t(104) = -3.05$, $p < .03$), suggesting that CD decreases across each stop consonant as place progresses more posteriorly in both languages.

Another regression model was fit with VD as response and the explanatory factors Stress (Stressed, Unstressed), Language (English, Spanish), and the Stress × Language interaction. These factors combined predicted a significant amount of variance in VD. The Stress (Unstressed) coefficient was significant, with the English unstressed vowels being significantly shorter than their stressed counterparts. The Language (Spanish) coefficient was also significant, suggesting that the stressed vowels in Spanish were longer than the stressed vowels in English. With the Language reference level set to Spanish, unstressed vowels in Spanish were also significantly shorter in duration than stressed vowels in Spanish.

Additionally, a model was fit with F2 Bark as the response and the explanatory factors Stress (Stressed, Unstressed), Language (English, Spanish), and the Stress × Language interaction. These variables combined predict a significant amount of variance in F2 Bark. The Stress (Unstressed) coefficient was significant, suggesting that F2 Bark decreased significantly in English unstressed vowels, meaning that unstressed vowels were more back than stressed vowels. The Language (Spanish) coefficient was significant, suggesting that in stressed syllables the vowel spoken in English was significantly different from that of Spanish for the BL speakers. With the reference level set to Spanish, the Stress (Unstressed) coefficient was also significant, implying that the vowel in Spanish unstressed syllables was more back compared to that of Spanish stressed syllables; however, the effect was seemingly smaller than that for English.

Table 3.4 Inferential statistics of VOT, Closure Duration (CD), Vowel Duration (VD), and F2 Bark for the BL speakers. For VOT and CD, separate models were fit with each consonant /p, t, k/ (in English) as baseline. For VD and F2 Bark, separate models were fit with each language (Spanish, English) as baseline. Significant coefficients are in bold.

	Regression Model				Regression Coefficients				
	R^2	*df*	*F*	*p*	Coefficient	*b*	*SE*	*t*	*p* value
VOT	.88	(6, 101)	133.6	< .001	Intercept	0.23	0.009		
					Span /p/	**–0.18**	**0.012**	**–14.88**	**< .001**
					Intercept	0.27	0.009		
					Span /t/	**–0.22**	**0.012**	**–17.57**	**< .001**
					Intercept	0.29	0.009		
					Span /k/	**–0.19**	**0.012**	**–15.27**	**< .001**
CD	.72	(6, 101)	47.09	< .001	Intercept	0.23	0.009		
					Span /p/	**0.12**	**0.017**	**6.83**	**< .001**
					Intercept	0.30	0.012		
					Span /t/	**0.18**	**0.016**	**11.23**	**< .001**
					Intercept	0.29	0.009		
					Span /k/	**0.12**	**0.017**	**7.25**	**< .001**
VD	.38	(3, 104)	22.58	< .001	Intercept	0.43	0.008		
					Unstressed	**–0.03**	**0.011**	**–2.54**	**< .02**
					Lang (Span)	**0.06**	**0.011**	**5.45**	**< .001**
					Intercept	0.49	0.008		
					Unstressed	**–0.04**	**0.011**	**–3.62**	**< .001**
F2 Bark	.36	(3, 104)	21.34	< .001	Intercept	12.07	0.09		
					Unstressed	**–0.43**	**0.12**	**–3.49**	**< .001**
					Lang (Span)	**–0.94**	**0.12**	**–7.63**	**< .001**
					Intercept	11.13	0.09		
					Unstressed	**0.24**	**0.12**	**1.99**	**< .05**

3.3.1.2 Spectral correlates of place of articulation

In this section, the descriptive and inferential statistics for the spectral correlates of place of articulation, including RI, F2 transition (F2T), COG, SD, SK, and KT are discussed.

3.3.1.2.1 Descriptive statistics

From the data presented in Table 3.5, it appears as though the BL speakers use a variety of different spectral correlates to distinguish place of articulation in their speech. On the one hand, the RI measures in English seem to be greater for each place of articulation than in Spanish, and in both languages the values increase as the place of articulation becomes increasingly posterior. This effect is outlined in more detail in the discussion. On the other hand, F2T generally follows the expected pattern: There is a negative slope for /t/ and /k/ in English and Spanish, although the effect appears to be greater in English than in Spanish. The slope for English /p/, however, is negative, whereas the slope for Spanish /p/ is positive. This unexpected pattern is discussed in detail in the Discussion.

In terms of the first four spectral moments, a pattern suggesting unique places of articulation for English and Spanish /t/ is evident for COG, SK, and SD. The COG and SK values for /p/ in English and Spanish appear to be similar amongst themselves, whereas, although similar, those for /k/ are slightly higher for COG and slightly lower for SK in English. A seemingly large difference exists between the values for English and Spanish /t/ for both COG and SK, however. Conversely, although English and Spanish /p/ pattern similarly, the differences in the SD values between English and Spanish /t/ and English and Spanish /k/ are both seemingly large. No apparent pattern emerges for the values obtained from KT.

Table 3.5 Descriptive statistics (mean with standard deviation in parenthese) of spectral correlates of place of articulation—Relative Intensity (RI), F2 Transition (F2T), Center of Gravity (COG), Standard Deviation (SD), Skewness (SK), and Kurtosis (KT)—analyzed in this study as a function of Language (English, Spanish) and Consonant (/p, t, k/) for the BL speakers. Boxplots are provided in Appendix G.

	RI		F2T		COG		SD		SK		KT	
	Eng	Span	Eng	Span	Eng	Span	Eng	Span	Eng	Span	Eng	Span
/p/	8.58	3.19	–0.62	1.53	217	203	442	400	15.7	14.0	680	467
	(3.10)	(1.92)	(1.07)	(0.90)	(268)	(139)	(428)	(266)	(9.9)	(6.6)	(935)	(514)
/t/	9.17	3.61	–0.79	–0.27	3633	587	2584	1121	1.4	7.6	26	97
	(2.84)	(1.82)	(0.87)	(0.84)	(2630)	(689)	(1014)	(611)	(4.6)	(4.9)	(218)	(148)
/k/	10.28	6.90	–1.08	–0.01	1785	1279	2028	1274	2.4	3.9	9	25
	(2.59)	(2.10)	(1.03)	(0.68)	(1005)	(481)	(651)	(559)	(1.4)	(1.2)	(16)	(18)

3.3.1.2.2 Inferential statistics

From the inferential statistics for the spectral correlates of place of articulation presented in Table 3.6, including RI, F2T, COG, SK, and KT, a pattern of using multiple correlates for distinguishing the English and Spanish coronal stops emerges among the BL speakers.

A model was fit with F2T as the response and the explanatory factors Stress (Stressed, Unstressed), Consonant (/p, t, k/), and Language (English, Spanish), and the Consonant × Language interaction. The Stress (Unstressed) coefficient was significant, suggesting that F2T increased significantly in unstressed syllables vis-à-vis stressed syllables in English. The Language (Spanish) coefficients were significant with a positive *b* for all baseline stop phonemes, indicating that F2T increased significantly for each stop in Spanish compared to English.

A similar trend as F2T, but in the opposite direction, was observed from the model fit with RI as response and the factors Consonant (/p, t, k/), Language (English, Spanish), and the Consonant × Language interaction: The Language (Spanish) coefficients were significant, but with negative *b* values, for all consonants in reference to the corresponding English baseline, denoting a significant decrease in RI in Spanish versus English. Because this pattern and direction appeared to model that of VOT for these speakers, several linear regression models were also fit, each with a different spectral measure (RI, COG, SD, SK, and KT) as the response variable and the explanatory variables VOT, Language (English, Spanish), and the VOT × Language interaction. These variables together predicted a significant amount of variance in RI ($R^2 = .65$, $F(3, 104) = 68.49$, $p < .001$) and the Language (Spanish) coefficient and VOT × Language interaction were significant (Language: $b = -6.46$, $t(104) = -4.37$, $p < .001$; Interaction: $b = 38.88$, $t(104) = 4.25$, $p < .001$) only for RI. These results suggest that there is a linear relationship between RI and VOT across languages and place of articulation: as RI increases, so does VOT. This effect is not observed for any of the other spectral measures.

Separate models using the first four spectral moments—COG, SD, SK, and KT—as the response variable and the factors Consonant (/p, t, k/), Language (English, Spanish), and the Consonant × Language interaction were also fit. The model fit for KT did not produce a significant Language (Spanish) coefficient for baseline /t/, and, consequently, follow-up models with the other stop consonants as baseline were not fit. The Language (Spanish) coefficient with baseline set to /p/ was not significant for COG, SD, or SK. With the baseline set to /k/, the Language (Spanish) coefficient was significant for SD, although it was not so for COG or SK, suggesting that SD decreased significantly in Spanish /k/ with reference to English /k/ for the BL speakers. However, for COG, SD, and SK, the Language (Spanish) coefficient was significant with baseline /t/, and, importantly, it was significant for COG and SK only with baseline /t/, indicating significant changes (decrease in COG and increase in SK in Spanish) in COG and SK between English and Spanish uniquely for the /t/ phoneme.

Table 3.6 Inferential statistics of spectral correlates of place of articulation—Relative Intensity (RI), F2 Transition (F2T), Center of Gravity (COG), Standard Deviation (SD), Skewness (SK), and Kurtosis (KT)—for the BL speakers. Separate linear regression models were fit with each consonant /p, t, k/ (in English) as the baseline. Significant coefficients are in bold.

	Regression Model				Regression Coefficients				
	R^2	*df*	*F*	*p*	Coefficient	*b*	*SE*	*t*	*p* value
RI	.70	(5, 102)	49.7	< .001	Intercept	8.62	0.422		
					Span /p/	**–5.44**	**0.597**	**–9.11**	**< .001**
					Intercept	9.19	0.422		
					Span /t/	**–5.56**	**0.597**	**–9.32**	**< .001**
					Intercept	10.30	0.422		
					Span /k/	**–3.40**	**0.597**	**–5.70**	**< .001**
F2T	.83	(6, 101)	88.35	< .001	Intercept	–0.72	0.099		
					Span /p/	**2.12**	**0.135**	**15.14**	**< .001**
					Intercept	–0.88	0.099		
					Span /t/	**0.50**	**0.130**	**3.16**	**< .003**
					Intercept	–1.19	0.099		
					Span /k/	**1.07**	**0.135**	**8.47**	**< .001**
COG	.63	(5, 102)	37.78	< .001	Intercept	218	207.9		
					Span /p/	–12	294.1	–0.04	> .90
					Intercept	3554	207.9		
					Span /t/	**–2962**	**294.1**	**–10.07**	**< .001**
					Intercept	1773	207.9		
					Span /k/	–494	294.1	–1.68	> .09
SD	.79	(5, 102)	80.58	< .001	Intercept	444	95.94		
					Span /p/	–42	135.68	–0.31	> 70
					Intercept	2568	95.94		
					Span /t/	**–1438**	**135.68**	**–10.60**	**< .001**
					Intercept	2028	95.94		
					Span /k/	**–755**	**135.68**	**–5.56**	**< .001**
SK	.73	(5, 102)	57.37	< .001	Intercept	15.67	0.800		
					Span /p/	–1.77	1.132	–1.56	> .10
					Intercept	1.43	0.800		
					Span /t/	**6.04**	**1.132**	**5.34**	**< .001**
					Intercept	2.38	0.800		
					Span /k/	1.49	1.132	1.32	> .10
KT	.45	(5, 102)	18.27	< .001	Intercept	25.99	66.26		
					Span (/t/)	69.40	93.70	0.74	> 40

3.3.1.3 Relative contribution of burst measurements in BL speakers

As the focus of this experiment is the difference in place of articulation between the English alveolar /t/ and the Spanish dental /t/, a logistic regression model was fit to determine which spectral moments—including COG, SD, and SK[12]—are the best predictors of the identity of these two stops. The hybrid forward-backward selection logistic regression model fit using the training data (a 75% subset of original data containing 212 observations) eliminated SK and kept COG and SD as the predictors for the best-fit model. This model (COG and SD as predictors) was then fit with the testing data (a 25% subset of original data containing 75 observations) to determine how well they can predict the identity of either English /t/ or Spanish /t/. With the prediction criterion set to above chance (> 50%), and using the data only from COG and SD, the model accurately predicted the identity of 79% of the English/Spanish coronal stop testing data (21% error rate).[13] In other words, using the spectral measures of COG and SD alone, 79% of English and Spanish coronal stops, which differ in place of articulation, can be predicted accurately in BL speakers.

However, since there was some evidence of a unique gestural target for the velar stops between English and Spanish in the BL speakers, a separate logistic regression model was fit using only the /k/ data using the same methods as for the /t/ data. The training subset of the /k/ data (75% of the the original data set) consisted of 244 observations. The hybrid forward–backward selection logistic regression model selected COG, SD, and SK as the best predictors of the identity of /k/ (i.e., whether it was English /k/ or Spanish /k/). These were then used to predict the language identity of the velar stops from the testing subset of the data (25% of

[12] RI was not included as a possible predictor to avoid any confounds with VOT, and KT was not included because it was found to not be different between English and Spanish.

[13] When including RI, this number is much higher—91%. However, RI was different for all consonants between English and Spanish and thus is likely not a reliable predictor of small differences in place of articulation between coronal stops.

original data set), consisting of 86 observations. The model accurately predicted the identity of 85% of the velar stops (15% error rate). This finding is detailed further in the discussion.

To ensure that that these models were predicting differences in place of articulation, the same cross-validation method was used on the /p/ data. However, the hybrid forward–backward selection logistic regression model eliminated all of the variables (COG, SD, and SK), indicating that none of the spectral moments were reliable predictors of place of articulation for /p/ between English and Spanish. Thus, any model fit with the /p/ data would be identifying the stop at chance.

3.3.2 RESULTS: LL SPEAKERS

3.3.2.1 Voice onset time, closure duration, vowel duration, and F2 Bark

3.3.2.1.1 Descriptive statistics

The descriptive statistics of the data VOT, CD, VD, and F2 Bark data presented in Table 3.7 generally reflect trends that would be expected from native speakers of English learning Spanish later in life: Some approximation toward target-like norms with patterns typical of interference from the L1 phonology. On the one hand, in English, the VOT values for all LL conditions are similar and pattern after one another for within-voicing-category stops: VOT increases as the place of articulation becomes increasingly more posterior. In Spanish, while the VOT values increase in a similar fashion as they do in English, the overall values appear to be less than in English. From Figure 3.2, a pattern of behavior in the Spanish VOT of all three Spanish voiceless stops of the LL speakers is apparent in which VOT increases from LL1 to LL2 and then begins to decrease as linguistic experience increases. This pattern was explored visually by plotting the VOT data for the LL2 condition separately in a violin plot, which is presented in Appendix F. A visual inspection of the violin plot suggests that the LL2 condition is bimodal in their production of Spanish /t/ in regard to VOT, as two different peaks are apparent. This may explain the peculiar trend of development observed in the VOT data. On the other hand, much like the BL speakers, CD seems to be greater in Spanish than in English overall, and a pattern of decreasing closure duration as place of articulation becomes more posterior is observed across the LL conditions.

In terms of vowel quantity and quality, a pattern similar to that exhibited by the BL speakers is apparent in the LL speaker data. Vowel duration seems to be greater in stressed than in unstressed syllables in English and in Spanish for the LL speakers. Evidently, as with the BL speakers, vowel duration is longer in Spanish than in English overall. Lastly, F2 Bark is also apparently greater overall in English than in Spanish, implying a more fronted vowel in the former.

Table 3.7 Descriptive statistics (mean with standard deviation in parenthese) of (left) VOT and Closure Duration (CD) as a function of Language (English, Spanish), Consonant (/p, t, k/), and Linguistic Experience (LE) (LL1, LL2, LL3, LL4) and of (right) Vowel Duration (VD) and F2 Bark as a function of Language, LE, and Stress (Stressed, Unstressed) for the LL speakers. Boxplots of individual variables are provided in Appendix G.

		VOT		CD			VD		F2 Bark	
		Eng	Span	Eng	Span		Eng	Span	Eng	Span
LL1	/p/	0.24	0.09	0.37	0.47	*Stressed*	0.42	0.46	12.12	11.05
		(0.07)	(0.07)	(0.07)	(0.07)		(0.06)	(0.08)	(1.09)	(0.74)
	/t/	0.27	0.09	0.31	0.47	*Unstressed*	0.39	0.41	11.58	11.57
		(0.08)	(0.05)	(0.05)	(0.09)		(0.09)	(0.08)	(1.46)	(0.77)
	/k/	0.29	0.15	0.31	0.43					
		(0.07)	(0.07)	(0.06)	(0.09)					
LL2	/p/	0.23	0.14	0.31	0.37	*Stressed*	0.47	0.50	12.22	10.88
		(0.08)	(0.08)	(0.07)	(0.09)		(0.08)	(0.08)	(1.34)	(0.87)
	/t/	0.26	0.16	0.28	0.38	*Unstressed*	0.46	0.45	11.76	11.41
		(0.06)	(0.10)	(0.08)	(0.11)		(0.10)	(0.09)	(1.54)	(0.90)
	/k/	0.29	0.21	0.24	0.31					
		(0.07)	(0.08)	(0.08)	(0.09)					
LL3	/p/	0.26	0.13	0.31	0.40	*Stressed*	0.45	0.49	11.88	10.69
		(0.07)	(0.09)	(0.07)	(0.07)		(0.08)	(0.09)	(1.49)	(0.88)
	/t/	0.32	0.15	0.26	0.40	*Unstressed*	0.41	0.43	11.76	11.20
		(0.06)	(0.10)	(0.09)	(0.10)		(0.08)	(0.08)	(1.50)	(1.04)
	/k/	0.32	0.21	0.25	0.33					
		(0.07)	(0.10)	(0.06)	(0.09)					
LL4	/p/	0.26	0.09	0.31	0.41	*Stressed*	0.45	0.51	12.02	10.85
		(0.07)	(0.07)	(0.06)	(0.08)		(0.06)	(0.08)	(1.28)	(0.70)
	/t/	0.29	0.09	0.26	0.44	*Unstressed*	0.44	0.45	11.70	11.36
		(0.07)	(0.06)	(0.06)	(0.09)		(0.08)	(0.09)	(1.44)	(0.84)
	/k/	0.31	0.17	0.24	0.35					
		(0.07)	(0.09)	(0.07)	(0.10)					

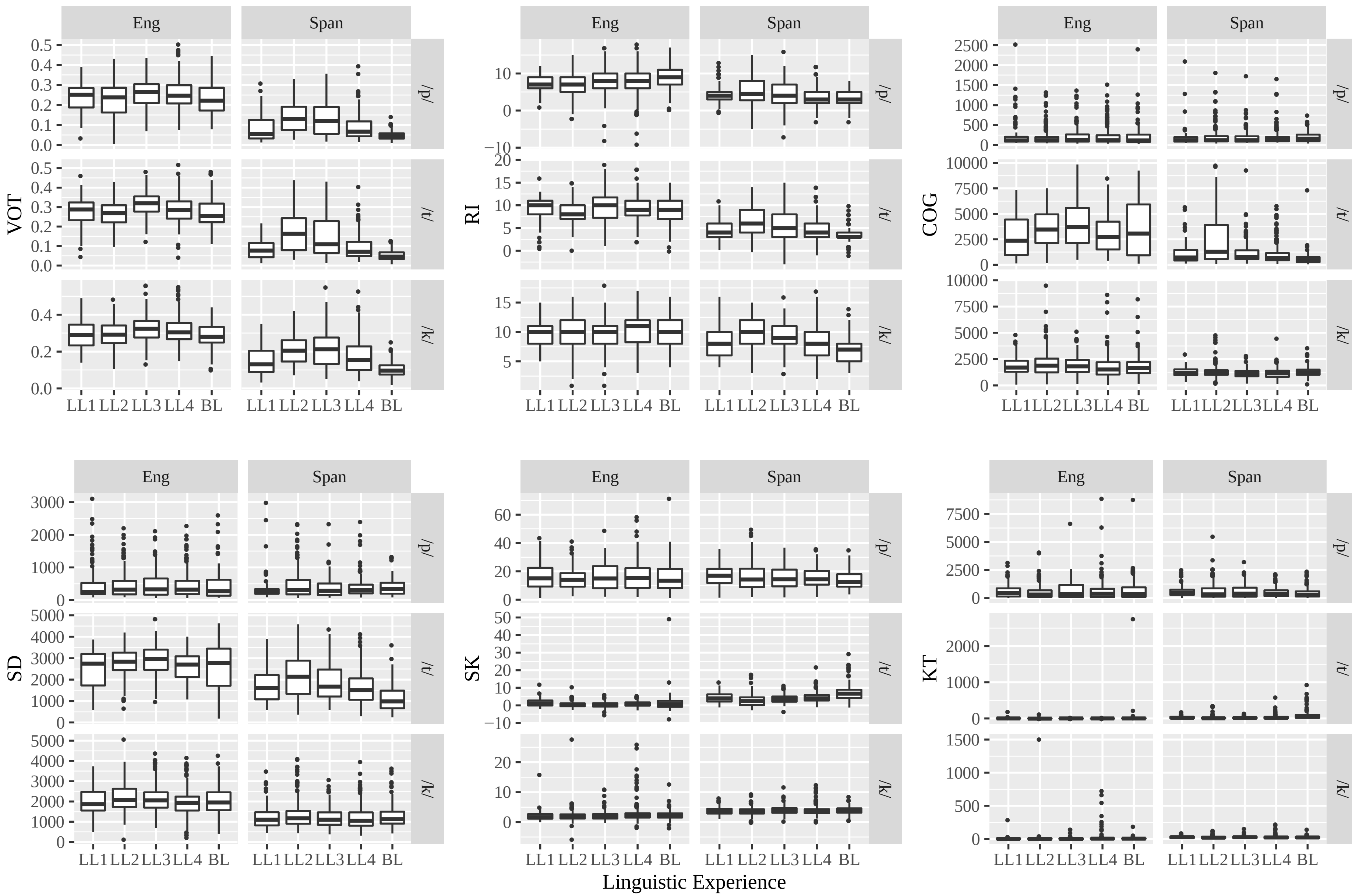


Figure 3.2 Boxplots of VOT, Relative Intensity (RI), Center of Gravity (COG), Standard Deviation (SD), Skewness (SK), and Kurtosis (KT) from LL and BL data. Scales of y-axes are adjusted for each consonant to enhance visual representation. Boxplots of each variable separately with equal y-axis scales across all consonants are provided in Appendix G.

3.3.2.1.2 Inferential statistics

Separate linear regression models were fit using the VOT and CD data from the LL speakers as the response variable and explanatory factors Stress (Stressed, Unstressed), Consonant (/p/, /t/, /k/), Language (English, Spanish), and numeric variable LE, and the Language × Consonant × LE interaction. Separate models were run for each level of consonant as the baseline. Thus, the baseline for each model was English {Consonant} in a stressed syllable. As can be seen in Table 3.8, these variables combined predict a significant amount of variance in VOT and CD.

For VOT, the Language (Spanish) coefficient was significant with all stops as the baseline, suggesting that, controlling for LE, VOT decreased in Spanish for /p/, /t/, and /k/. The LE coefficient was not significant for any of the consonants, meaning that VOT was not different across LL conditions in English stressed syllables for any of the stops. The Language (Spanish) × LE coefficient was significant with baseline English /p/ and English /t/, indicating that an increase in LE was associated with a decrease in VOT for Spanish /p/ and /t/, in stressed syllables. However, this coefficient was not significant with English /k/ as the baseline, suggesting that the decrease observed across LL conditions was not significant. These findings were similar when removing Stress from the model.

Table 3.8 Inferential statistics of VOT, Closure Duration (CD), Vowel Duration (VD), and F2 Bark for the LL speakers. Separate linear regression models were fit with each consonant /p, t, k/ (in English) as the baseline. Significant coefficients are in bold.

Regression Model					Regression Coefficients				
	R^2	*df*	*F*	*p*	Coefficient	*b*	*SE*	*t*	*p*
VOT	.60	(12, 479)	63.29	< .001	Intercept (/p/)	0.24	0.008		
					Lang (Span)	**–0.13**	**0.010**	**–12.64**	**< .001**
					LE	0.01	0.006	1.71	> .05
					Span /p/ × LE	**–0.02**	**0.009**	**–2.14**	**< .04**
					Intercept (/t/)	0.28	0.008		
					Lang (Span)	**–0.15**	**0.010**	**–15.42**	**< .001**
					LE	0.01	0.006	1.88	> .05
					Span /t/ × LE	**–0.02**	**0.009**	**–2.72**	**< .01**
					Intercept (/k/)	0.30	0.008		
					Lang (Span)	**–0.11**	**0.010**	**–11.29**	**< .001**
					LE	0.01	0.006	1.55	> .10
					Span /k/ × LE	–0.01	0.009	–1.44	> .10
CD	.51	(12, 479)	43.5	< .001	Intercept (/p/)	0.30	0.008		
					Lang (Span)	**0.09**	**0.010**	**8.27**	**< .001**
					LE	**–0.01**	**0.006**	**–1.98**	**< .05**
					Span /p/ × LE	0.01	0.009	1.09	> .20
					Intercept (/t/)	0.26	0.008		
					Lang (Span)	**0.14**	**0.010**	**13.57**	**< .001**
					LE	**–0.01**	**0.006**	**–2.13**	**< .05**
					Span /t/ × LE	0.02	0.009	1.84	> .06
					Intercept (/k/)	0.24	0.008		
					Lang (Span)	**0.09**	**0.010**	**8.66**	**< .001**
					LE	–0.01	0.006	–1.83	> .05
					Span /k/ × LE	0.00	0.009	0.26	> .70
VD	.14	(7, 484)	11.64	< .001	Intercept (Eng)	0.45	0.006		
					Unstressed	**–0.02**	**0.009**	**–2.75**	**< .01**
					Lang (Span)	**0.04**	**0.009**	**4.39**	**< .001**
					LE	0.00	0.005	0.50	> .60
					Lang (Span) × LE	0.01	0.008	1.24	> .20
					Intercept (Span)	0.49	0.006		
					Unstressed	**–0.05**	**0.009**	**–6.38**	**< .001**
F2 Bark	.42	(7, 484)	81.61	< .001	Intercept (Eng)	12.08	0.049		
					Unstressed	**–0.40**	**0.070**	**–5.66**	**< .001**
					Lang (Span)	**–1.23**	**0.070**	**–17.56**	**< .001**
					LE	–0.07	0.045	–1.53	> .10
					Lang (Span) × LE	0.02	0.063	0.29	> .70
					Intercept (Span)	10.86			
					Unstressed	**0.52**	**0.070**	**17.56**	**< .001**

The model fit with CD as response shows that, similar to VOT, the Language (Spanish) coefficient was significant for all consonants, although in the opposite direction. In other words, controlling for LE, CD increased in Spanish for each stop relative to its English counterpart, in stressed syllables. Furthermore, the LE coefficient was significant with /p/ and /t/ set as the baseline, but not for /k/. This denotes that, for the bilabial and coronal stops in English, CD decreased as LE increased, in stressed syllables. These effects were similar when Stress was removed from the model. Additionally, unlike for VOT, the Language (Spanish) × LE coefficient was not significant for any of the consonants, implying that, in Spanish, CD did not differ for any stop as LE increased. To determine whether CD changed as a matter of place of articulation across LL conditions within each language, a separate model was fit with CD as response and Consonant (coded numerically, as done with the BL speakers), Language (English, Spanish), LE, and the Consonant × Language × LE interaction as the explanatory variables. These variables together predicted a significant amount of variance in the response ($R^2 = .46$, $F(7, 484) = 59.03$, $p < .001$). The Consonant coefficient was significant (English: $b = –0.03$, $t(484) = –6.04$, $p < .001$; Spanish: ($b = –0.03$, $t(484) = –5.67$, $p < .001$) and the Consonant × LE interaction was not (English: $b = 0.00$, $t(484) = 0.11$, $p < .90$; Spanish: $b = –0.00$, $t(484) = –0.68$, $p < .40$) in English and Spanish, which, taken together, suggest that CD decreased as place of articulation progressed from the lips to the velum in English and in Spanish, and that this pattern was not different across the LL conditions.

Additionally, a linear regression model was fit with VD as response and the explanatory factors Stress (Stressed, Unstressed) and Language (English, Spanish), numeric variable LE, and the Stress × Language × LE interaction. These variables combined predict a significant amount of variance in VD. The Stress (Unstressed) coefficient was significant, suggesting that, controlling for LE, the unstressed vowels in English were shorter in duration than stressed vowels in English. The Language (Spanish) coefficient was also significant, denoting that, controlling for LE, stressed vowels in Spanish were longer than stressed vowels in English.

The LE coefficient was not significant, meaning that stressed English vowels did not change in duration across LE. The Language (Spanish) × LE interaction coefficient was not significant, indicating that stressed VD did not significantly increase in Spanish with increasing LE compared to English. With the Language reference level set to Spanish, the Stress (Unstressed) coefficient was also significant, meaning that, controlling for LE, Spanish unstressed vowels were shorter in duration than Spanish stressed vowels. The effect size for vowel reduction (quantity) in English and Spanish was apparently greater in Spanish than in English.

Furthermore, a model was fit with F2 Bark as the response and the explanatory variables Stress (Stressed, Unstressed), Language (English, Spanish), and LE and the Stress × Language × LE interaction. These variables combined predicted a significant amount of variance in F2 Bark. The Stress (Unstressed) coefficient was significant, denoting that, controlling for LE, F2 Bark was significantly smaller in unstressed syllables than in stressed in English, denoting a more fronted vowel in stressed syllables than in unstressed. The Language (Spanish) coefficient was significant, suggesting that, controlling for LE, F2 Bark was lower in Spanish than in English in stressed syllables, meaning that the vowel in Spanish was more back than that of English. The LE coefficient was not significant, indicating that the F2 Bark of stressed vowels in English did not change significantly across LL conditions. Lastly, the Language (Spanish) × LE coefficient was not significant, meaning that F2 Bark did not significantly increase (i.e., become more fronted) in stressed vowels in Spanish compared to English as linguistic experience increased. With the Language reference set to Spanish, the Stress (Unstressed) coefficient was also significant, denoting that, controlling for LE, the F2 Bark of the vowel (i.e., vowel quality) in unstressed syllables in Spanish was significantly less from that of stressed syllables in Spanish, implying a more back vowel in unstressed syllables, and this effect was apparently greater than it was for English.

3.3.2.2 Spectral correlates of place of articulation

3.3.2.2.1 Descriptive statistics

From the data presented in Table 3.9, it is apparent that the pattern of development of place of articulation in LL speakers is much more nuanced than it is for the BL speakers. Firstly, RI appears to follow a similar pattern as the data from the BL speakers: Across LL conditions, the values from one condition to the next are about the same for each place of articulation in English. Furthermore, RI seems to be greater in English than in Spanish overall, and the values appear to consistently increase as place of articulation progresses more toward the back of the vocal tract. From a visual inspection of the data in Figure 3.2, a pattern similar to that of the VOT production emerges: For RI, COG, and SD, the LL data for the Spanish /t/ begin at the LL1 by approaching target-like norms; however, they appear to regress toward those of the L1 in the LL2 condition, making their way back down to approaching seemingly target-like perform by the LL4 condition. The LL2 data for RI, COG, and SD are also visualized in separate violin plots in Appendix F. Similar to the VOT data, there is some evidence to suggest that participants in the LL2 condition are also bimodal, as for RI, COG, and (although slightly less so) SD, two different peaks appear in the data. This bimodality of the LL2 condition may explain the pattern of development observed for these metrics in regard to the development of place of articulation among the LL speakers.

Likewise, the trends in the F2T data seem to pattern closely after those found in that of the BL speakers: In English, all the stop consonants have a negative slope. In Spanish, for all speakers, the /p/ has a positive slope and /t/ and /k/ have negative ones, and the effect appears to be greater in English than in Spanish. This possibly surprising finding will be detailed further in the discussion. Furthermore, there is a trend of linear development toward target-like values for Spanish /t/ in the SK data among the LL speakers; however, much like with the BL data, it is difficult to decipher any trends in the KT data for the LL conditions.

Table 3.9 Descriptive statistics (mean with standard deviation in parentheses) of spectral correlates of place of articulation—Relative Intensity (RI), F2 Transition (F2T), Center of Gravity (COG), Standard Deviation (SD), Skewness (SK), and Kurtosis (KT)—analyzed in this study as a function of Language (English, Spanish), Consonant (/p, t, k/), and LE (LL1, LL2, LL3, LL4) for the LL speakers. Boxplots of individual variables are provided in Appendix G.

		RI		F2T		COG		SD		SK		KT	
		Eng	Span	Eng	Span	Eng	Span	Eng	Span	Eng	Span	Eng	Span
LL1	/p/	7.24	4.25	–0.54	0.71	261	192	512	365	16.23	17.10	621	636
		(2.47)	(2.68)	(1.22)	(1.30)	(375)	(276)	(596)	(447)	(9.71)	(7.36)	(658)	(551)
	/t/	8.85	4.78	–0.89	–0.58	2854	1142	2508	1767	1.48	4.35	7	30
		(2.96)	(2.45)	(1.0)	(0.73)	(2130)	(1125)	(847)	(842)	(2.29)	(2.86)	(23)	(34)
	/k/	9.74	8.03	–1.09	–0.19	1897	1246	2019	1219	2.10	3.78	9	30
		(1.94)	(2.39)	(1.24)	(0.86)	(900)	(456)	(681)	(554)	(1.77)	(1.35)	(30)	(20)
LL2	/p/	6.80	4.93	–0.74	0.64	199	218	443	473	14.73	16.13	506	615
		(3.26)	(3.91)	(1.50)	(1.20)	(207)	(249)	(387)	(449)	(7.57)	(9.71)	(592)	(747)
	/t/	8.63	6.51	–0.94	–0.77	3555	2351	2797	2167	0.49	2.86	2	22
		(2.65)	(2.95)	(1.32)	(1.0)	(1797)	(2357)	(655)	(974)	(1.55)	(3.50)	(10)	(51)
	/k/	9.92	9.84	–1.39	–0.40	2039	1304	2192	1354	2.08	3.66	13	24
		(2.76)	(2.69)	(1.11)	(0.94)	(1162)	(667)	(691)	(688)	(2.11)	(1.46)	(95)	(20)
LL3	/p/	7.87	4.23	–0.75	0.36	228	200	462	396	16.40	15.82	679	613
		(3.76)	(3.81)	(1.22)	(1.27)	(241)	(207)	(404)	(339)	(9.54)	(7.93)	(808)	(602)
	/t/	9.65	5.86	–1.03	–0.95	3934	1193	2875	1896	0.35	3.85	2	23
		(3.03)	(3.28)	(1.36)	(0.94)	(2132)	(1168)	(729)	(883)	(1.54)	(2.48)	(6)	(29)
	/k/	9.53	9.22	–1.20	–0.34	1896	1121	2140	1214	2.11	3.95	8	26
		(2.48)	(2.50)	(1.17)	(0.80)	(915)	(431)	(676)	(513)	(1.59)	(1.52)	(20)	(20)
LL4	/p/	7.70	3.89	–0.72	0.88	208	190	462	381	16.13	15.65	601	502
		(3.57)	(2.75)	(1.31)	(1.10)	(211)	(180)	(392)	(294)	(9.43)	(6.77)	(850)	(462)
	/t/	9.31	4.81	–1.13	–0.65	3139	1006	2643	1658	0.84	4.50	2	33
		(2.80)	(2.60)	(1.29)	(0.96)	(1948)	(956)	(647)	(780)	(1.48)	(2.85)	(6)	(55)
	/k/	10.55	8.35	–1.16	–0.35	1679	1124	1957	1204	2.80	3.88	19	29
		(2.64)	(2.96)	(1.14)	(0.93)	(1060)	(496)	(657)	(597)	(2.91)	(1.68)	(72)	(30)

3.3.2.2.2 Inferential statistics

The results of the inferential statistics for the spectral measures are presented among two tables: RI, F2T, and COG in Table 3.10, and SD, SK, and KT in Table 3.11. From these results, a pattern of the development of the place of articulation in Spanish dental /t/ is apparent. For all spectral measures (RI, F2T, COG, SD, SK, and KT), a similar linear regression model was fit with the explanatory variables Consonant (/p, t, k/), Language (English, Spanish), LE, and the Consonant × Language × LE interaction. These variables combined predicted a significant amount of variance in the response in each model. Different models were run with each level of Consonant as the reference separately (with English as the reference level of the factor Language) in all cases. Thus, the baseline for all models was English {Consonant}.

The results of the linear regression models for RI and F2T were similar. In both models, the LE coefficient was not significant for any consonant, meaning that RI and F2T did not change significantly with increasing LE in English for any stop. The Language (Spanish) coefficient was significant for RI and F2T for all consonants, meaning that, controlling for LE, in the case of F2T, the F2 transition was greater in Spanish than in English and, conversely, RI was less in Spanish than in English, for each consonant. Likewise, the Language (Spanish) × LE coefficients were not significant for any consonant for RI or F2T, suggesting that changes in F2T or RI between English and Spanish across LL conditions were not different.

Separate models were also fit for the spectral moments COG, SD, SK, and KT as response variables. Firstly, the LE coefficient was not significant for any consonant across all spectral moments, implying that the LL conditions did not behave differently from one another in their English. Secondly, none of the Language (Spanish) and Language (Spanish) × LE coefficients were significant with /p/ set as the baseline, suggesting that the behavior for the production of /p/ was not different in English and Spanish for the LL speakers. Thirdly, the Language (Spanish) × LE coefficient was not significant for any spectral moment for /k/;

however, the Language (Spanish) coefficient was significant for COG and SK for /k/, and the change went in the same direction as the change did for the BL speakers: lower COG and higher SK values in Spanish vis-à-vis English. These results denote that the behavior for the production of /k/, while seemingly stable across LL conditions, was different between English and Spanish for the LL speakers. Lastly, when /t/ was set as the baseline, the Language (Spanish) coefficient was significant for COG, SD, and SK, meaning that, controlling for LE, the values for these spectral moments were different between English and Spanish. In all cases, the direction was in the same direction as was observed for the BL speakers. Importantly, at least for COG and SD, the Language (Spanish) × LE coefficient was also significant, and in the same direction as occurred for the BL speakers. These findings for the LL speakers' Spanish /t/ production suggest that, as linguistic experience increases, the COG and SD data for Spanish /t/ decrease relative to that of English /t/. In other words, there is robust evidence of development in the place of articulation of L2 Spanish dental /t/ among the late learners in this study.

Table 3.10 Inferential statistics of Relative Intensity (RI), F2 transition (F2T), and Center of Gravity (COG) for the LL speakers. Separate linear regression models were fit with each consonant /p, t, k/ (in English) as baseline. Significant coefficients are in bold.

Regression Model					Regression Coefficients				
	R^2	*df*	*F*	*p*	Coefficient	*b*	*SE*	*t*	*p*
RI	.47	(11, 480)	39.79	< .001	Intercept (/p/)	7.31	0.25		
					Lang (Span)	**–2.85**	**0.358**	**–7.91**	**< .001**
					LE	0.30	0.228	1.32	> .10
					Span /p/ × LE	–0.60	0.323	–1.87	> .06
					Intercept (/t/)	9.04	0.253		
					Lang (Span)	**–3.40**	**0.358**	**–9.50**	**< .001**
					LE	0.24	0.228	1.03	> .30
					Span /t/ × LE	–0.56	0.323	–1.75	> .08
					Intercept (/k/)	9.96	0.253		
					Lang (Span)	**–0.92**	**0.358**	**–2.57**	**< .02**
					LE	0.27	0.228	1.16	> .20
					Span /k/ × LE	–0.50	0.323	–1.56	> .10
F2T	.57	(12, 479)	55.32	< .001	Intercept (/p/)	-0.70	0.065		
					Lang (Span)	**1.35**	**0.092**	**14.68**	**< .001**
					LE	–0.03	0.059	–0.53	> .50
					Span /p/ × LE	0.10	0.083	1.24	> .20
					Intercept (/t/)	–1.12	0.065		
					Lang (Span)	**0.25**	**0.092**	**2.67**	**< .01**
					LE	–0.08	0.059	–1.38	> .10
					Span /t/ × LE	0.08	0.083	0.97	> .30
					Intercept (/k/)	–1.25	0.065		
					Lang (Span)	**0.92**	**0.092**	**9.97**	**< .001**
					LE	0.04	0.059	0.61	> .50
					Span /k/ × LE	–0.06	0.083	–0.66	> .50
COG	.64	(11, 480)	81.61	< .001	Intercept (/p/)	220	94.45		
					Lang (Span)	–16	133.58	–0.12	> .90
					LE	–6.56	85.32	–0.08	> .90
					Span /p/ × LE	–0.74	120.66	–0.01	> .90
					Intercept (/t/)	3425	94.45		
					Lang (Span)	**–1875**	**133.58**	**–14.04**	**< .001**
					LE	–14.55	85.32	–0.17	> .80
					Span /t/ × LE	**–292**	**120.66**	**–2.42**	**< .02**
					Intercept (/k/)	1903	94.45		
					Lang (Span)	**–692**	**133.58**	**–5.18**	**< .001**
					LE	–112	85.32	–1.31	> .10
					Span /k/ × LE	49.06	120.66	0.41	> .68

Table 3.11 Inferential statistics of Standard Deviation (SD), Skewness (SK), and Kurtosis (KT) for the LL speakers. Separate linear regression models were fit with each consonant /p, t, k/ (in English) as baseline. Significant coefficients are in bold.

	Regression Model				Regression Coefficients				
	R^2	*df*	*F*	*p*	Coefficient	*b*	*SE*	*t*	*p*
SD	.83	(11, 480)	210.8	< .001	Intercept (/p/)	467	45.1		
					Lang (Span)	–49	63.7	–0.78	> .40
					LE	–3	40.7	–0.08	> .90
					Span /p/ × LE	–16	57.6	–0.26	> .70
					Intercept (/t/)	2721	45.1		
					Lang (Span)	**–811**	**63.7**	**–12.72**	**< .001**
					LE	1	40.7	0.03	> .90
					Span /t/ × LE	**–12**	**57.6**	**–2.14**	**< .04**
					Intercept (/k/)	467	45.1		
					Lang (Span)	–49	63.7	–0.77	> .40
					LE	–3	40.7	–0.08	> .90
					Span /k/ × LE	23.06	57.6	0.40	> .60
SK	.81	(11, 480)	189.6	< .001	Intercept (/p/)	15.69	0.35		
					Lang (Span)	0.40	0.50	0.80	> .40
					LE	0.22	0.32	0.68	> .40
					Span /p/ × LE	–0.56	0.45	–1.23	> .20
					Intercept (/t/)	0.71	0.35		
					Lang (Span)	**3.04**	**0.50**	**6.05**	**< .001**
					LE	–0.06	0.32	–0.19	> .80
					Span /t/ × LE	0.41	0.45	0.91	> .30
					Intercept (/k/)	2.24	0.35		
					Lang (Span)	**1.55**	**0.50**	**3.10**	**< .003**
					LE	0.26	0.32	0.82	> .40
					Span /k/ × LE	–0.19	0.45	–0.42	> .60
KT	.60	(11, 480)	67.66	< .001	Intercept (/p/)	587.49	25.19		
					Lang (Span)	2.38	35.62	0.07	> .90
					LE	14.76	22.74	0.65	> .50
					Span /p/ × LE	–59.50	32.18	–1.85	> .06
					Intercept (/t/)	2.77	25.19		
					Lang (Span)	23.61	35.62	0.66	> .50
					LE	–0.98	22.75	–0.04	> .90
					Span /t/ × LE	3.38	32.18	0.12	> .90
					Intercept (/k/)	12.48	25.19		
					Lang (Span)	14.36	35.62	0.40	> .60
					LE	2.94	22.75	0.13	> .80
					Span /k/ × LE	–1.66	32.18	–0.05	> .90

3.3.2.3 Development of place of articulation in LL speakers

The final step in the analysis was to determine the predictability of the English alveolar and Spanish dental stops in the LL speakers using the same acoustic correlates as those used to predict the identity of English/Spanish coronal stops in BL speakers, including COG and SD. Thus, a logistic regression model was fit with English /t/ (= 0) and Spanish /t/ (= 1) as the response variable and the explanatory variables COG and SD for the data from each LL condition separately. With the prediction criterion set to above chance (> 50%), the model accurately predicted the identity of the English alveolar/Spanish dental stop among the LL conditions as follows: (a) LL1: 69% accuracy, (b) LL2: 71% accuracy, (c) LL3: 82% accuracy, and, lastly, (d) LL4: 80% accuracy. Once again, these results suggest that the LL speakers do make gains in their development of the place of articulation in Spanish dental voiceless stops and that, generally, greater target-like production is observed as linguistic experience increases.

The LL English and Spanish /k/ data were analyzed using the same method as was used on the /t/ data for each condition of LE separately. The same variables as selected in the model with the BL data for /k/ (COG, SD, and KT) were used to predict the identity of the velar stop production of the LL speakers. The model produced the following results: (a) LL1: 83% accuracy, (b) LL2: 81% accuracy, (c) LL3: 83% accuracy, and, lastly, (d) LL4: 78% accuracy. These results suggest that the gestural target is different between the English and Spanish velar stops, but that the production stays relatively stable as linguistic experiences increases.

In light of the fact that the hybrid forward–backward selection logistic model on the /b/ data for the BL speakers eliminated all predictors, it was determined unnecessary to run any logistic regression model on the LL speakers' /b/ data.

3.4 DISCUSSION

The objective of this experiment was to investigate the development of place of articulation in learners of a second language whose gestural target for the voiceless coronal stop phoneme

(Spanish dental /t/) differs slightly from that of their native language (English alveolar /t/). As a baseline comparison, native English/Spanish bilinguals were also investigated. Place of articulation was measured from production data obtained for English and Spanish voiceless stops /p, t, k/ and was operationalized as a function of, among other acoustic correlates of place, the first four spectral moments derived from the envelope of the burst of the coronal stop /t/ (COG, SD, SK, KT), in addition to the intensity of the burst relative to that of the following vowel (RI). The descriptive and inferential statistics presented in the Results section lead to the conclusions described in the following subsections.

3.4.1 BL Speakers

3.4.1.1 Primary findings for BL speakers

3.4.1.1.1 Summary of primary findings

The primary findings from the BL speakers' production data in the present experiment can be summarized as follows:

1. Robust evidence for a unique gestural target between the English alveolar /t/ and the Spanish dental /t/ was found: The measures of center of gravity, standard deviation, and skewness of the spectral energy derived from the burst were all different between English and Spanish for /t/, and, importantly, differences between English and Spanish were found for center of gravity and skewness *only* for the coronal stop phonemes.
2. Center of gravity and standard deviation combined predicted the identity of either the English alveolar /t/ or Spanish dental /t/ with 79% accuracy on a testing subset (25%) of the BL data.
3. Some evidence of a unique gestural target between the English /k/ and Spanish /k/ was found: The standard deviation of the distribution of the energy of the burst of this velar consonant was greater in English than in Spanish.

4. The relative intensity of the burst was greater in English than in Spanish for all voiceless stop consonants, and, in both languages, relative intensity increased as place of articulation moved toward the velum.

3.4.1.1.2 Interpretation and implications of primary findings

Most relevant to the present study are the differences in gestural target exhibited by the BL speakers in the phonologically similar phonemes between their native English and Spanish speech. The findings discussed here are based on the results of the inferential statistics for the spectral moments center of gravity, standard deviation, and skewness. The hypothesis for speakers from this population was that there would be effects of language for the English and Spanish coronal voiceless stops, but not for the bilabial or velar stops. The results partly support this hypothesis. Firstly, no effects of language were found for center of gravity, standard deviation, or skewness for the bilabial stops. Likewise, none of these were good predictors of the identity of /p/ as determined by the hybrid forward-backward selection logistic regression model. This is a simple but important finding: The place of articulation for the bilabial stops was not different between English and Spanish for the BL speakers. Secondly, effects of language were found for standard deviation for the velar stops. This finding was, initially, unexpected; however, as will be addressed in the following paragraphs, it is likely gesturally motivated. Thirdly, and lastly, effects of language were found for center of gravity, standard deviation, and skewness for the coronal stops, and center of gravity and skewness were different between English and Spanish only for the coronal stop phonemes. Furthermore, center of gravity and standard deviation together predict the identity of the coronal stop, whether it be English alveolar /t/ or Spanish dental /t/, with 79% accuracy. This is a principal finding of the present experiment: Robust evidence has been found that native bilinguals of English and Spanish produce the English alveolar /t/ and Spanish dental /t/ with unique gestural targets, suggesting that, phonologically, these two sounds have unique abstract representations in the BL speakers of this study.

Sundara (2006), who measured the spectral moments of the burst of the Canadian-English alveolar /t/ and the French-Canadian dental /t/ production to determine differences in place of articulation among coronal stops in early English/French bilinguals, found that standard deviation and kurtosis displayed significant differences, suggesting that, similar to the BL speakers in the present study, native bilingual speakers of English and French also exhibit language-specific gestural targets for coronal stops. However, although kurtosis was found to be a relevant correlate to place in the speakers in Sundara (2006), it was not so for the BL speakers in this study. Similar to Sundara, however, there were language effects for standard deviation, and, furthermore, the step-wise logistic regression model included standard deviation as a predictor in the best-fit model for the prediction of the identity of alveolar and dental stops in the present study. Taking the results from the present study and those of Sundara (2006) into account, apparently, and intuitively, the spectral correlates to language-specific gestural targets differ as a matter of linguistic experience in native bilinguals, even among samples of populations that share phonologically similar sounds.

Contrary to the initial hypothesis, some evidence of a unique gestural target between English and Spanish /k/ was found. However, this finding has a simple explanation based on coarticulation. Previous research has shown that the gestural target for velar stops can vary cross-linguistically (Butcher & Tabain, 2004). Within a language, articulatory data (Recasens & Espinosa, 2009) has shown that /k/ has different gestural targets due to effects of coarticulation. Specifically, Recasens and Espinosa (2009) found that the target for /k/ was more fronted when preceded by front vowels than when preceded by back vowels. In the present study, in English, the velar /k/ was always preceded by the same front vowel /eɪ/ (i.e., *say* [seɪ]), whereas, in Spanish, it was always preceded by the same back vowel /o/ (i.e., *digo* [ˈdi.ɣo]). According to the findings in Recasens and Espinosa, it would be expected that the gestural target for the /k/ in the English production would be more fronted, due to the coarticulation with the preceding fronted vowel, than it would be in Spanish, due to the preceding back vowel. In short, the finding that the place of articulation for the English /k/ was different, at least as measured acoustically, in some respects compared to that of the Spanish /k/ is likely a result of the materials of the study and an expected outcome of coarticulation.

The remainder of this subsection deals with the findings on relative intensity, which is an auditory metric that has been shown to be relevant to the distinction of place of articulation of coronal stops in monolingual speakers intralingually (Jongman & Blumstein, 1985), cross-linguistically (Casillas et al., 2015; Sundara et al., 2006), and in native bilinguals (Sundara et al., 2006). However, the direction of the findings has been mixed, and, consequently, a discussion regarding the findings for the relative intensity in the BL speakers in the present study is merited. Jongman and Blumstein (1985) found that the dental /t/ of Malayalam had a greater relative intensity than the alveolar /t/ in speakers of the same language. Sundara (2006) reports a similar finding for the alveolar /t/ of monolingual and bilingual speakers of Canadian English and the dental /t/ of monolingual and bilingual speakers of Canadian French. Conversely, in Casillas et al. (2015), the relative intensity of the dental /t/ of the monolingual speakers of Spanish, however, was less than that of that of the alveolar /t/ of monolingual English speech, and the results of the present study pattern with those of Casillas et al. Furthermore, although the relative intensity difference between the English and Spanish coronal stops was different, it was different for the bilabial and velar stops, as well. The direction appears to follow the same direction and pattern as VOT for both languages: Relative intensity and VOT are greater for English than for Spanish for each stop, and the value of each measure increases as place of articulation progresses toward a more posterior location in the cavity of the mouth. Casillas et al. and Sundara both controlled for a potential confound of VOT on relative intensity by measuring stops of similar voicing categories (short-lag, or voiced stop /d/, in English, and short-lag, or voiceless stop /t/, in Spanish and French, respectively). Although it was clearly outside of the objective of the research of those papers to do so, the relative intensity of the burst for the full inventory of one or both stop voicing categories for English, Spanish, or French were not reported by Casillas et al. or Sundara. Nevertheless, the findings from the present study denote that VOT linearly predicts relative intensity across place of articulation in voiceless stops in English and Spanish in the BL speakers. In other words, this study found that relative intensity cannot be used as a reliable predictor of place of articulation for coronal stops.

It is important to note that an increase in relative intensity signifies a decrease in burst intensity. Thus, although the relative intensity increases from bilabial to alveolar to velar, the burst intensity itself decreases in the same direction in this study. It can be said, then, that the intensity of the bursts in the production of voiceless stops by BL speakers in this study are greatest for /b/, less so for /t/, and less so for /k/. One plausible explanation for this finding is, perhaps, that intraoral pressure for bilabial stops is greater than for other places of articulation, and the pressure release during the burst causes a greater intensity for labials than other stops. However, Eshghi, Alemi, and Zajac (2017) found, in a recent study on voiceless stops, that intraoral pressure is greatest for velars and smallest for bilabial stops, whereas Müller and Brown (1980) found that intraoral pressure remained constant across place of articulation. Thus, it seems unlikely that the greater intensity observed in bilabials is due to a greater intraoral pressure vis-à-vis stops of other places of articulation. Perhaps a more probable explanation can be found simply in what is being measured and its relationship to place of articulation. Considering that burst intensity, measured in dB, is a logarithmic measure of perceived loudness, and the noise of the burst originates at the place of articulation (Johnson, 2011, p. 174), then the finding that bilabials have a greater burst intensity than coronals, which, in turn, have a greater burst intensity than velars, is intuitive—the perceptual intensity of the burst may be decreasing as the sound wave travels through the vocal tract. A visual inspection of the data implies that this is plausible, as the relative intensity difference between the bilabial and coronal stops for both languages, a relatively short distance, is smaller than that between the coronal stops and velars, a relatively long distance.

The fact that the relative intensity effect is greater in English—and, therefore, that the intensity of the burst itself overall is smaller—than in Spanish may have to do with differences in the phonetic realization of phonological voicing categories between the two languages. On the one hand, it may be that, in English, VOT serves as a more salient psychoacoustic feature (Strange & Shafer, 2008) than burst intensity to place of articulation among voiceless stops in English for the BL speakers. This is reasonable to assume considering that English employs long-lag stops, the aspiration of which is rather salient in and of itself. On the other hand, it may be that burst intensity has a greater psychoacoustic salience in the phonetic distinction of voiceless stops in Spanish for the BL speakers, especially considering that the language employs short-lag VOT, which, intuitively, is not as salient as that which is found in long-lag categories. In any event, without relative intensity data for the phonologically voiced stops for these BL speakers, and without detailed data reported previously in the literature, this discussion is only speculation. Further research is needed to explore these findings for relative intensity.

3.4.1.2 Secondary findings for BL speakers

3.4.1.2.1 Summary of secondary findings

Beyond the findings relating directly to the questions of the research, there were several secondary findings that relate to the acoustic correlates of stop consonants and the vowels surrounding them. Generally, the results indicate that the BL speakers have a unique phonological categorization for vowels, as measured by F2 transitions and F2 Bark, and consonants, based on the measures of VOT and closure duration, between their English and Spanish. These findings are summarized as follows:

1. In English and Spanish, stressed vowels were longer than unstressed vowels, and stressed vowels in Spanish were longer than those in English. The effect of unstressed vowel reduction (quantity) was apparently greater in Spanish than in English.
2. F2 Bark at the midpoint of the vowel in stressed syllables in Spanish was significantly less than the same measure in English, implying that the BL speakers used a more back vowel in Spanish than in English. F2 Bark of the vowel was significantly different in unstressed syllables compared to stressed syllables for English and Spanish, suggesting vowel reduction (quality) in both languages, although the effect was evidently greater in English than in Spanish.
3. VOT was greater for each voiceless stop consonant place of articulation in English than the corresponding voiceless stop consonant in Spanish. In both languages, VOT increased as place of articulation became more posterior.
4. In English and Spanish, closure duration linearly decreased as place of articulation progressed to a more posterior location. Closure duration for each place of articulation (/p, t, k/) was greater for Spanish voiceless stops than for English voiceless stops.
5. F2 transitions in English were negative for all voiceless stop consonants, which were followed by either /æ/ or /ɑ/. In Spanish, the F2 transition was positive for /p/ and negative for /t/ and /k/, which were all followed by /a/. The size of the transition was greater for each stop consonant in English than in Spanish.

3.4.1.2.2 Interpretation and implications of secondary findings

Previous research has found that native English/Spanish bilinguals have language-specific phonological systems for stop consonants (García-Sierra, Ramírez-Esparza, Silva-Pereyra, Siard, & Champlin, 2012; Gonzales & Lotto, 2013), and that they produce their vowels (Kehoe, 2002) and stop consonants using language-specific acoustic correlates (Brown & Copple, 2018), although their production can be phonetically different from monolinguals of the same language (Simonet et al., 2014). Unstressed vowel reduction in this population has also been investigated and discussed elsewhere (Ronquest, 2013). These secondary findings (1–3)—which dealt with vowel quantity, vowel quality, and the VOT of Spanish and English voiceless stops in the BL speakers—go beyond the scope of the present dissertation and have been well documented previously; therefore, they will not be discussed at length here.

The finding that closure duration is longer in Spanish than in English for voiceless stops does merit brief attention, however, as it deals more closely with the question of place of articulation in native bilinguals. It has been found previously that, between languages that exhibit similar phonetic patterns between stop voicing categories, such as Spanish and French, which both have long-lead voiced and short-lag voiceless stops, the closure duration of corresponding voiceless stops is different (Torreira & Ernestus, 2011). Very little detailed data, if any, however, exist in published literature comparing the closure duration of voiceless stops in English and Spanish (but see Green, Zampini, & Magloire, 1997; Zampini, 2002). Although the direction of the pattern of closure duration in voiceless stops (especially vis-à-vis voiced stops) between English and Spanish is expected to be similar (see 3.1.1 for a more detailed discussion), Zampini (2002) reports that the monolingual Spanish speakers in her study appeared to rely on stop closure duration for cues to place of articulation in voiceless stops in Spanish to a greater degree than did the monolingual English speakers in English. Similarly, Green, Zampini, and Magloire (1997) report that both monolingual English speakers and English/Spanish bilinguals in their study produced similar closure duration patterns in English between voiced and voiceless stops. The finding in this study that closure duration was greater for Spanish voiceless stops than for English may support these previous findings that it is a cue to voicing in Spanish more so than it is in English. It may be that the BL speakers here use closure duration as a cue to stop voicing in Spanish, but not in English, and thus require greater closure duration for voiceless stops in their Spanish to allow for enough phonetic malleability to maintain the voicing distinction vis-à-vis English.

However, considering that closure duration as analyzed here is a measure that is relative to the duration of VOT, another explanation based on the facts known from the data collected in the present study is that the greater closure duration observed in Spanish may be a result of the smaller VOT values vis-à-vis the shorter closure duration and longer VOT in English. That is, it may be that, within either language, a movement in the direction of smaller VOT values leads to greater closure duration, whereas movement in the direction of greater VOT leads to smaller closure duration. In any event, without CD data from voiced stops for these speakers, or detailed data on the CD of voiced and voiceless stops in either monolingual or bilingual speakers of English and Spanish, this remains conjecture.

Lastly, in terms of formant transitions, one peculiar finding resulted from the F2 transition measurements: Contrary to expectations (Delattre et al., 1955), the F2 transition for the English /p/ was negative (although so were those for /t/ and /k/ in both English and Spanish, which was expected). Conversely, and as expected (Quilis, 1993, p. 209), the F2 transition for the Spanish /p/ was positive. However, this divergence has a possible, simple explanation. Ladefoged (2006, p. 193) states that formant transitions in voiceless, aspirated stops take place prior to the onset of modal voicing during aspiration (see also Quilis, 1993, p. 209). However, since Spanish short-lag stops, which have little to no aspiration, and English long-lag stops were both included in the present study, the F2 transition calculation was derived from the value of F2 Bark at the onset of the vowel as the beginning of the transition. Furthermore, the target vowel in English was necessarily different among some of the materials, changing between the low, front /æ/ and the low, back /ɑ/. These differences in target vowel will affect the slope of the F2 transition. In contrast, the short-lag Spanish /p/ was always followed by the low, central vowel /a/. In sum, the negative F2 transition observed for the English /p/ may well have been an artifact of the method or materials of the experiment, or both. And, considering that different vowels were used between English and Spanish, the effect size, which was greater in English than in Spanish, may well be due to the same reasoning.

3.4.2 LL SPEAKERS

3.4.2.1 Primary findings for LL speakers

3.4.2.1.1 Summary of primary findings

Two of the principal research questions driving this experiment where (a) whether adult learners of Spanish as a second language can acquire unique gestural targets for the L2 Spanish dental stop /t/ vis-à-vis their L1 English alveolar stop /t/, and, if so, (b) whether gains are made toward target-like production as linguistic experience increases. Based on previous research regarding the development of L2 stop acquisition as measured by the acoustic correlates of stop voicing, and also taking into account second language speech theory, the initial hypothesis was that learners will be able to make some progress toward a unique gestural target for this sound in their L2 with increasing linguistic experience, although it may initially resemble the L1 alveolar stop. The principal findings from this experiment support these hypotheses, and they are summarized as follows:

1. Center of gravity and standard deviation were significantly different only for coronal stops, in the same direction as for the BL speakers, as linguistic experience increased, suggesting that the LL speakers developed a unique gestural target for their Spanish /t/. However, relative intensity, skewness, kurtosis, and F2 transitions did not change significantly between English and Spanish for any stop consonant across LL conditions.
2. Using the same predictors used to model the testing data set for BL speakers (center of gravity and standard deviation), the identity of either the English alveolar /t/ or Spanish dental /t/ was predicted with 69% accuracy for the LL1, 71% accuracy for the LL2, 82% for the LL3, and 80% for the LL4 speakers.

3.4.2.1.2 Interpretation and implications of primary findings

In the language of the L2LP, the native speakers of English learning Spanish in this study were facing a Similar Scenario in their learning of L2 Spanish stops. English and Spanish each have bilabial, coronal, and velar stops that contrast phonologically as a matter of voicing, but which, particularly between languages, differ as a matter of their phonetic realization. The principal finding of this experiment was that there exists evidence of the development of a unique gestural target for the L2 Spanish dental /t/ vis-à-vis the L1 English alveolar /t/. This suggests that the learners were capable of updating the phonetic categories of their L2 perception grammar and that this happened with increasing linguistic experience. As reviewed in the Introduction, previous research has found that L2 learners are capable of making gains toward target-like production in their acquisition of phonologically similar sounds, particularly in regard to stop voicing, and that gains are made as experience with the language increases (Casillas, 2016; Reeder, 1998; Stevens, 2001; Zampini, 1998). However, these data are the first to account directly for the acquisition of the gestural targets in stops whose phonological representation is similar but whose phonetic realization differs slightly in terms of place of articulation.

Prior research has shown that the acoustics of coronal stops between English and Spanish are different (Casillas et al., 2015), and, therefore, the initial state of the L2 perception grammar in Spanish, which *a priori* was copied from the L1, would be inherently different for these sounds. The results presented here suggest that LL speakers use different—or, at least, fewer—cues to distinguish the place of articulation of the Spanish dental /t/ from the English alveolar /t/ at the beginning stages of learning. Specifically, the BL speakers exhibited significant differences in relative intensity, center of gravity, standard deviation, and skewness between their English and Spanish coronal stops; however, the LL speakers displayed development of only center of gravity and standard deviation in terms of acoustic correlates of stop place of articulation. The finding that different acoustic cues to the distinction of place are used between the samples of BL and LL speakers was corroborated by the logistic regression model that was fit with the data from each LL condition separately. The degree to which the LL speakers used the same cues that were selected as the best predictors of place of articulation between English and Spanish coronal stops in the BL speakers increased as a function of linguistic experience, remaining low for the LL1 and LL2 conditions, and then increasing abruptly for the LL3 speakers to the point of overshooting the norm established by the model for the BL speakers, and, lastly, decreasing for the LL4 speakers back to a degree similar to that of the native bilinguals. These findings are in line with other research that suggests that L2 learners weigh acoustic cues differently than native speakers (for a review, see Strange & Shafer, 2008). In other words, the findings from this study suggest that, when learners are faced with a Similar Scenario, the L2 perceptual categories are different from those of native speakers of the language, but that, over time and with increasing linguistic experience, these categories are updated and adjusted to become more in line with those of target-like production. Importantly, this research has shown that the phonetic categories for place of articulation in a Similar Scenario are updated to increasingly reflect target-like production with greater exposure to the language.

A final comment is merited regarding the pattern of development observed in the visual inspection of the correlates to voicing and place of articulation in which, initially, LL1 production appeared to resemble target-like production, but then a regression in the direction of the L1 was observed in the LL2 data, followed by a trend toward increasing target-like production in the LL3 and LL4 conditions. This pattern in the data may be due to individual differences in the speakers of the LL1 or the LL2 conditions, or both, that are not generalizable outside of the sample of speakers in this study. Firstly, the LL2 speakers displayed high variability for the second and third spectral moments (standard deviation and skewness, respectively) relative to that of the other conditions. Because of this, the LL2 data were explored further by means of violin plots for each metric separately. A clear pattern of bimodality emerges in the LL2 condition's center of gravity data, suggesting that the LL2 condition may be composed of two separate groups of speakers: One in which the speakers largely produce their L2 Spanish /t/ with more target-like place of articulation, and one in which they do not. Secondly, and lastly, another plausible explanation—which complements, rather than supplements, the previous—lies in the LL1 data. In the case of VOT and the spectral moments of the Spanish /t/, the LL1 speakers' production appears to resemble that of the LL4 condition (which, in turn, resembles that of the BL condition) the closest relative to the LL2 and LL3 conditions. In other words, these speakers appear to be more advanced in their production of the Spanish /t/ than the LL2 and the LL3 conditions, which is also supported by the low variability present in their data. It may be that the LL1 speakers are simply more advanced than what would be expected for their level. In any event, let it suffice to say that evidence of a pattern suggesting that the LL speakers developed a unique gestural target, and, likewise, a unique abstract representation, for their L2 Spanish dental /t/ as been found.

3.4.2.2 Secondary findings for LL speakers

3.4.2.2.1 Summary of secondary findings

Beyond the principal findings of this study, the LL speakers displayed tendencies of L1 interference in their L2 production, particularly in terms of vowel quantity, vowel quality, and of VOT, but robust evidence of the effect of linguistic experience on gains toward target-like production was also found. These secondary findings are summarized as follows:

1. In English and Spanish, controlling for LE, stressed vowels were longer than unstressed vowels. The effect size of unstressed vowel reduction (quantity) was seemingly similar in both languages.
2. F2 Bark was significantly smaller in Spanish than in English when controlling for LE, implying that the LL speakers used a more back vowel in Spanish overall; however, when taking LE into account, there was not a significant trend of development. Also,

controlling for LE, in unstressed syllables, F2 Bark was significantly different in both languages, although the effect was apparently greater in Spanish than in English, suggesting a pattern of transfer of unstressed vowel reduction (quality) in the L2.

3. VOT in Spanish decreased significantly as linguistic experience increased for /p/ and /t/; however, there was no significant development of VOT for /k/ across levels of LE.
4. Closure duration did not differ significantly between English and Spanish for any stop consonant across levels of LE; however, it did decrease in each language as place of articulation progressed toward the back of the mouth in both languages.
5. Controlling for LE, the F2 transitions were different in English and Spanish.
6. Robust evidence of a unique gestural target for /k/ was also found; however, the pattern suggests that it is the result of coarticulation rather than phonological learning.

3.4.2.2.2 Interpretation and implications of secondary findings

The evidence of unstressed vowel reduction, both in quality and quantity, in adult learners of Spanish as a second language is not new and has been well documented in previous research (Cobb & Simonet, 2015; Menke & Face, 2010; Stevens, 2011). Similarly, the acquisition of VOT in adults learning Spanish as a second language has also received much attention (Casillas, 2016; González-Bueno, 1997; Reeder, 1998; Stevens, 2001). Generally speaking, the findings in the present study are congruent with those of previous studies in this regard. Nevertheless, these findings, vowel reduction and stop voicing were not related to the principal questions of the present study, and, therefore, they will not be discussed further.

The development of stops as measured by the production of closure duration, however, is a topic that is both relevant to the questions of this research and has received little attention, and, therefore, it warrants brief discussion. The finding, which was that the closure duration of the bilabial stops was greater than that of the coronal stops, which was in turn greater than that of the velar stops, in both languages, was the expected direction based on previous research, as reviewed in the Introduction. However, unlike the BL speakers, the LL speakers in this study did not produce their English and Spanish stops with differing closure durations, and this did not change across LL conditions. In other words, in terms of closure duration, the stops produced in English were similar to those in Spanish for all LL speakers, including the coronal stops. There is very little published data on the acquisition of closure duration in voiceless stops between languages to compare these findings to, as much of the research deals with the acquisition of voicing distinctions (e.g., Flege & Port, 1981). In any event, the findings here suggest that the LL speakers in this study did not produce language-specific properties of stop closure duration in their L2 Spanish. These findings compared to those of the development of unique gestural targets of stops imply that the phonetic categories relating to place of articulation are updated sooner than those relating to closure duration in the L2 perception grammar.

The F2 transitions were different in English and Spanish controlling for LE in the LL speakers; however, there was no trend of development for any of the stop consonants. In general, F2 transitions change as a function of place of articulation, but they are also sensitive to changes in the following vowel, as the F2 values are derived from the following vowel itself. Like the F2 transition results, in this study, while controlling for LE, the F2 Bark at the midpoint of the vowel was different between English and Spanish, with a more back vowel in Spanish; however, it did not differ across LL conditions. The F2 transition results here may be a finding relating more closely to the phonological acquisition of vowels than the gestural targets of stop consonants. Apparently, the quality of the target vowel was different in Spanish than in English across LL speakers' production, but it did not differ with increasing linguistic experience, which is why the findings for the F2 Bark and F2 transition are similar. This may also be why there was robust evidence found for a unique gestural target between the English and Spanish /k/, as discussed in the discussion of the BL speaker data. The logistic regression model also supports the idea that the unique velar target was due to coarticulation and not learning: The model predicted the identity of the velar stop with seeming similar accuracy across all LL conditions.

3.5 CONCLUSION

The present study investigated the development of place of articulation in adult learners of a second language (Spanish) whose coronal stop differs in place of articulation from that of their native language. As a baseline comparison, native English/Spanish bilinguals were also included in the study. The main findings indicate that the BL speakers have unique phonological categories for their English and Spanish coronal stops and display differing gestural targets for each. Importantly, the LL speakers also exhibited evidence of having developed a unique gestural target for their L2 Spanish dental /t/ from their L1 English alveolar /t/. These findings suggest that L2 speakers, when confronted with a Similar Scenario in stop acquisition, are capable of updating their L2 perception grammar to approach the phonetic categories—specifically, the gestural targets—of native speakers of the target language, and that this process is aided by an increase in linguistic experience.

CHAPTER 4

THE DEVELOPMENT OF L2 RHYTHM

4.1 INTRODUCTION

The objective of the experiment presented in this chapter is to explore the nature of the development of rhythm in the speech of adults learning a second language (Spanish) whose rhythm is typologically different from that of their first language (English). The speech of native bilinguals of English and Spanish is also analyzed. Rhythm is operationalized here via a series of metrics that capture the variability in the timing of vocalic and consonantal sequences of segments in the participants' utterances in English and Spanish.

4.1.1 RHYTHM

Rhythm, a prosodic characteristic of speech and the focus of this chapter, deals principally with the "perceived regularity of prominent units in speech" (Crystal, 2008, p. 417), or, in other words, the apparent up and down beat of the strong and weak units of a given language. As far as this author is aware, while the facts of the rhythm of English have been known for centuries (Steele, 1775), the idea of cross-linguistic rhythmic variation was first documented by Arthur Lloyd James in his work as a linguistic advisor to the Royal Air Force of Great Britain during World War II to aid in the development of communication systems. In his brief manual on speech, two rhythm categories were proposed: "Morse code rhythm" and "machine gun rhythm" (Lloyd James, 1940, p. 25), in reference to the short (dot) and long (dash) pattern of Morse and the even beat of a machine gun. Lloyd James theorized that rhythm was a combination of linguistic patterns of duration, lexical stress, and melodic contour (p. 26), and that prominent changes to these phenomena from one period to the next constitute the Morse-code or machine-gun beat of a given language. Thus, according to this proposal, languages similar to English had Morse code rhythm, while those similar to French (i.e., Spanish) had machine gun rhythm.

The notion of cross-linguistic rhythmic variation was later supported by a series of rhythm perception experiments in children by Mehler and colleagues. Mehler et al. (1988) studied the ability of four-day-old French and two-month-old American infants to discriminate between-category rhythm. The study found that, even after low-pass filtering the stimuli, the French and American infants (whose languages, in the phrasing of Lloyd James, were machine gun rhythm and Morse code rhythm, respectively) were able to distinguish the speech of their native language from that of a language whose rhythm was typologically different—Russian (Morse code rhythm) and Italian (machine gun rhythm), respectively. Robust evidence for the existence of differing rhythm patterns cross-linguistically was presented in Nazzi, Bertoncini, and Mehler (1998), who studied the perception of rhythm in five-day-old infants born to French-speaking families. The infants in this study were able to discriminate low-pass-filtered utterances of their native language, French, from those of Japanese, a language whose rhythm is typologically different.[14] Importantly, these infants were not able to discriminate between within-category rhythm languages (English and Dutch). Taken together, the findings from these two studies show that infants can perceive differences in language rhythm between languages with seemingly different rhythm patterns, supporting the notion of a cross-linguistic rhythm typology.

Later, Pike (1947), while addressing the idea of rhythmic units, elaborated further on the idea of categorical rhythm classes. On the one hand, some languages, such as English, he proposed, have a tendency in their rhythm units—defined as groupings of syllables in a given utterance separated by pauses—toward “somewhat uniform” (p. 34) durations of time between stressed syllables. In other words, three contiguous rhythmic units may be composed of a different number of syllables; however, the duration between the beginning of the stressed syllable of the first rhythmic unit to that of the second, and from the second to that of the third, would be of roughly equal duration. Thus, a phrase containing more syllables than

---

[14] See footnote 15 for more information.

another may be the same duration as another phrase with fewer syllables. Languages whose rhythm units are largely stress based after this manner were termed *stress-timed*. On the other hand, "many non-English languages" (p. 35),[15] such as Spanish, have rhythm units at the syllable level, such that one syllable to the next is of roughly equal duration. Consequently, phrases containing more syllables than another are of longer duration than another phrase with fewer syllables. To these types of languages, he gave the name *syllable-timed*. Pike's major contribution in regard to the phenomenon of rhythm—the idea of roughly equal durations of time between the prominent syllables of consecutive rhythm units—has come to be known as, appropriately, *isochrony*.[16]

However, a strict interpretation of isochrony has not been upheld by empirical studies on rhythm unit duration. Dauer (1983), for instance, found no evidence that the duration of interstress intervals in English, a stress-timed language, were significantly different from those of Spanish, a syllable-timed language. Furthermore, for Spanish, a strict interpretation of isochrony has not been found in prior research. For example, it has been found that closed syllables are longer than open syllables (Aldrich & Simonet, n.d.; Hoequist Jr., 1983) and, similarly, that stressed syllables in Spanish are longer in duration than unstressed syllables (Cuenca, 1996; Ortega-Llebaria & Prieto, 2007). Taken together, these facts for Spanish suggest that neither syllables nor entire interstress intervals could be of equal duration. Nevertheless, the degree to which interstress intervals are considered equal may also depend on the perspective of whether a language truly exhibits patterns of isochrony (Dauer, 1983),

---

[15] Apparently, at the time of Pike's writing, the world's languages had been categorized into only two rhythm classes: those that patterned after English (i.e., stress-timed) and those that patterned after Spanish (i.e., syllable-timed). For instance, Abercrombie states, "As far as is known, every language in the world is spoken with one kind of rhythm or with the other" (1967, p. 97). However, while not discussed at length here, a third rhythm class—mora-timed languages—has also since been proposed in the literature (Han, 1962).

[16] Stemming from the adjective *isochronous*, derived from *isochron*, composed of the Greek words for equal, ίσος ('isos') and time, χρόνος ('k^{h}ronos') (See ("Equal," n.d.).

and some studies have found at least partial support for isochrony in production in English and even stronger tendencies in perception (Lehiste, 1973).

In fact, the very notion of rhythm categories themselves has been put to test. On the one hand, Rathcke and Smith (2015) investigated trained phoneticians' ability to determine the rhythm class—whether it be stress-timed or syllable-timed—of different dialects of their native language, British English. Although previous research had classified the dialects of the study according to their rhythm, the phoneticians were able to determine whether a given dialect was syllable-timed or stress-timed only slightly above chance. On the other hand, Dauer (1983) proposes that all languages exhibit some evidence of what Pike (1947) described as stress timing and proposes that the standard, categorical nomenclature be replaced with a continuum of "stress-based rhythm" (pp. 59-60). In this model, languages traditionally classified as syllable timed would fall to the left on the scale for being less stress based and those traditionally classified as stress timed would fall to the right. The data from Grabe and Low (2002) support the notion of gradient language rhythm: These authors found that some languages that had previously not been categorized into rhythm categories seemed to overlap between the two. Thus, language rhythm may be better described as a continuum rather than purely categorical.

In any event, setting aside the notions of isochrony and categoricity, what is certain is that stress-rhythm languages and syllable-rhythm languages[17] do tend to differ in regard to syllable structure, phonological vowel reduction (particularly in the case of English and Spanish), and the acoustic correlates of stress (Dauer, 1983). On the one hand, syllable structure in English allows for up to three consonants in the onset (e.g., *straight*) and up to four in the coda (e.g., *twelfths*) (Whitley, 2002, p. 34), whereas, in Spanish, up to two consonants are permitted in the onset (e.g., *criar*) and up to two in the coda (e.g., *vals*) (Hualde, Olarrea,

[17] To avoid confusing with the notion of isochrony, *syllable-timed* and *stress-timed* will be referred to here as *syllable-rhythm* and *stress-rhythm* languages henceforth.

Escobar, & Travis, 2010, p. 101). On the other hand, English has phonological vowel reduction in unstressed syllables (Ladefoged, 2006, p. 94), whereas Spanish does not (Hualde, 2005, pp. 126–127). Lastly, in English, stress affects vowel duration: vowels in stressed syllables are longer than in unstressed syllables (de Jong, 2004; Van Summers, 1987). Although a similar pattern of stress-induced vowel lengthening is found in Spanish, as well (Marín Gálvez, 1995; Navarro Tomás, 1917), the effect is greater in English than in Spanish due to the tendency of unstressed syllables in English to reduce in quality and quantity (Whitley, 2002, p. 68). Working under the view taken by Dauer (1983) in which rhythm is composed, in part, by a language's syllable structure, propensity toward vowel reduction, and the phonetic realization of stress, both of which are observable via duration measures, these distinctions between English and Spanish are relevant to the empirical analysis of rhythm.

4.1.2 MEASURING RHYTHM

The findings that seemingly contradicted the notion of isochrony called into question any possibility of capturing what appeared to be phonological distinctions of rhythm cross-linguistically using phonetic properties of the language alone (Dauer, 1983). Nevertheless, Dauer (1987) offered an alternative phonetic–phonological approach to classifying rhythm cross-linguistically. "It seems that an adequate description of rhythm in a language or across languages requires both phonetic and phonological information" (p. 447), she said, proposing, including those already discussed from Dauer (1983), eight components of language rhythm, each consisting of a tripartite classification: +, 0, and –. Although a description of the various components goes beyond the scope of this chapter, suffice it to say that they dealt, broadly speaking, with phonetic patterns of duration, pitch, quality, and accent (stress) (Dauer, 1987, pp. 448–449). The phonological component of Dauer's proposal enters in with the tripartite classification: Based on the features described, English would receive a rhythm score of +6, whereas French would receive a score of –5 (Arvaniti, 2009, p. 54). This phonetic–phonological approach led to the development of other metrics, more closely aligned to the phonetic properties of language, by Ramus and colleagues, described in the following paragraphs.

In a seminal paper, Ramus, Nespor, and Mehler (1999) proposed phonetic metrics for operationalizing language rhythm in languages that had been previously described as stress timed (e.g., English, Dutch, Polish) and syllable timed (e.g., Spanish, Italian, Catalan, French). The foundational argument in Ramus et al. stemmed from previous findings suggesting that infants largely attend to sequences of vowels in their development of the first language (Mehler, Dupoux, Nazzi, & Dehaene-Lambertz, 1996). They concluded, "We thus assume that the infant primarily perceives speech as a succession of vowels of variable durations and intensities, alternating with periods of unanalyzed noise (i.e. consonants)" (p. 270). From this assumption, each metric, defined further on, applies a statistic to the duration of vocalic and consonantal sequences, also called vocalic/consonantal intervals (White & Mattys, 2007). Specifically, these measures were the following: (a) %V, the proportion of a given utterance that is vocalic, (b) ΔV, the standard deviation of vocalic intervals, and (c) ΔC, the standard deviation of consonantal intervals (Ramus et al., 1999, p. 272). As a result of measuring vocalic and consonantal intervals independent from their relationship to one another (see the Pairwise Variability Indexes below), it can be said that these three metrics—%V, ΔV, and ΔC—measure rhythm patterns *globally*.

Considering that stress-rhythm languages tend to exhibit greater consonantal complexity—allowing for, and leading to, greater variability among consonantal intervals—and, similarly, have a tendency toward vowel reduction—leading to greater variability among vocalic intervals—it is expected that languages such as English would have relatively high ΔV and ΔC scores. Conversely, syllable-rhythm languages, which have simpler syllable structure and often lack vowel reduction, are expected to have relatively low ΔV and ΔC scores. In terms of %V, it is hypothesized that syllable-rhythm languages would have higher overall vowel proportions than stress-rhythm languages due to their general lack of vowel reduction

(Ramus et al., 1999). Ramus et al. (1999) plotted the three metrics (%V × ΔC, %V × ΔV, and ΔC × ΔV) and found, in all cases, that the data showed evidence supporting the notion that there exist rhythm categories, although the languages within each category appeared to be related on a continuum to one another. In sum, although Pike's (1947) original proposal was largely anecdotally based, and other accounts of rhythm class proposed that the phenomenon was not as categorical as originally hypothesized, acoustic data have shown that languages such as Spanish and English do behave differently in regard to their rhythm patterns.

Subsequent studies on rhythm using Ramus et al.'s (1999) metrics revealed that ΔV and ΔC are affected by speech rate. For instance, Barry, Andreeva, Russo, Dimitrova, and Kostadinova (2003) found an inverse relationship between the standard deviation of vocalic and consonantal intervals and tempo: As speech rate increased, ΔV and ΔC decreased for the same speakers within the same language condition. These findings led Dellwo (2006) to introduce a coefficient of variation, or a metric of "relative variation to the norm" (p. 232) for consonantal intervals. This *variation coefficient,* or simply *varco*, has come to be known as VarcoV (White & Mattys, 2007) for vocalic intervals and VarcoC (Henriksen & Fafulas, 2017; White & Mattys, 2007) for consonantal ones. It is calculated by taking the standard deviation of the vocalic or consonantal intervals, divided by the mean of the intervals, multiplied by 100 (to provide a whole, interpretable number relative to the standard deviation). In sum, ΔV and ΔC calculate the standard deviation of vocalic and consonantal intervals, respectively, and VarcoV and VarcoC are these same metrics relative to their mean. All four metrics are global rhythm measures of interval variability.

While global rhythm metrics measure the variability of vocalic and consonantal intervals independent from one another, pairwise variability metrics calculate *local* rhythm variability from one vocalic/consonantal interval to the next. Specifically, the Pairwise Variability Index (PVI)—proposed in Low and Grabe (1995) and further developed in Low, Grabe and Nolan (2000) and Grabe and Low (2002), among other studies—takes the absolute difference in duration between consecutive intervals, whether they be vocalic (PVI-V) or consonantal (PVI-C), summed across all samples in an utterance, divided by the number of samples. This metric can also be normalized (nPVI-V or nPVI-C) by dividing the absolute difference between consecutive intervals by the average of the two prior to summing the value across samples (Grabe & Low, 2002). Stress-rhythm languages, which exhibit greater variability in consonantal and vocalic intervals, are expected to have a greater PVI score than syllable-rhythm languages, which display less variability (Low et al., 2000). Based on previous studies in which it was found that changes in speech rate have a greater effect on the duration of vocalic segments than on consonantal ones (Gay, 1978), Grabe and Low (2002) propose that the nPVI, normalized for speech rate, should be used to measure rhythm between vocalic intervals (nPVI-V) and that the raw PVI is best used for consonantal intervals (rPVI-C). In light of this, nPVI-V and rPVI-C have been used by default in many subsequent studies on rhythm (Henriksen & Fafulas, 2017; White & Mattys, 2007).

Rhythm has also been investigated in contexts such as dialectal variation and early bilingualism. For instance, it was found that Singapore English exhibits patterns more closely resembling syllable-rhythm languages, whereas British English displays variability typical of a stress-rhythm language (Low et al., 2000). There has also been some research on the nature of rhythm in bilingual speakers who from birth have spoken languages with differing rhythm patterns. Specifically, this research has asked if bilinguals use a unique rhythm pattern for each language or if only one rhythm pattern is used for both, whether the two languages be within the same rhythmic category (Whitworth, 2002) or between (Carter & Wolford, 2016). Whitworth (2002) found that early German/English bilingual children, in each language separately, produced rhythm patterns approaching those specific to that language, but still differed from adult native speakers of each. Carter and Wolford (2016), in a cross-sectional study of different generations of Mexican-American speakers of Spanish and English living in the United States, found that the younger the generation of the speaker, the more the participants exhibited a stress-rhythm pattern in their Spanish, and the reverse for older generations, who spoke Spanish with a more syllable-rhythm pattern. More recently, Henriksen and Fafulas (2017) studied the rhythm patterns of Yagua–Spanish bilinguals living in Amazonian Peru. Yagua, prior to the study, had not been classified in regard to its rhythm category, but it was found that, compared to monolingual Spanish, Yagua showed evidence of a more stress-rhythm pattern. The researchers found that the Yagua-dominant bilinguals, who were the older generation, displayed more Yagua-like rhythm in their Spanish, whereas the Spanish-dominant bilinguals, who were the younger generation, displayed more Spanish-like rhythm in their Spanish. Taken together, the results from these studies suggest that early bilinguals speak with tendencies similar to those specific to the given language, but not without an influence from their dominant language.

Despite the proliferation of these metrics in the literature on rhythm, Arvaniti (2009) does not agree that they actually capture the nuances that exist between rhythm classes cross-linguistically, or that the conception of rhythm as it has been defined is accurate. She presents data that contradict the findings from previous studies that measured rhythm in typologically different languages. For instance, using the same metrics described previously (ΔV, ΔC, VarcoV, VarcoC, %V, nPVI-V, and rPVI-C), in addition to the calculation of Euclidean distances between them with English as the reference point, she found that German (stress-rhythm), Korean (syllable-rhythm), and Greek (likely syllable-rhythm) were not different from one another, and, in one case, that Japanese (mora-rhythm) and English (stress-rhythm) were rhythmically similar. The mixed findings between this and other studies, Arvaniti argues, is due to rhythm being much more than a measure of duration alone, including elements of speech that are seemingly impossible to quantify with only one metric, including stress, tone, pitch, and intonation. Citing evidence from studies on the perception of musical rhythm, she calls into question the psychological reality of something such as a syllable-rhythm language at all. Arvaniti concludes her argument with a call to rely less on durational measures to capture linguistic rhythm and that researchers take a more holistic, psychological approach to understanding, beyond its manifestation in speech only, how it is perceived by native speakers of the language in question. However, in the absence of an alternative approach to capture phonetically the rhythm of languages, the present study employs those metrics reviewed previously, including ΔV, ΔC, VarcoV, VarcoC, %V, nPVI-V, and rPVI-C, to investigate the development of rhythm in second language learners.

4.1.3 Second Language Rhythm

Unlike with the acquisition of segmental phenomena, as reviewed in Chapter 3, no formal theory for the acquisition of prosodic phenomena has been established for second language speech. Therefore, as other authors have done (Trofimovich & Baker, 2006), one is left to conjecture that the outcomes of prosodic development observed in the literature would parallel those modeled theoretically by the likes of Flege (1995), Best (1995), and/or Escudero (Escudero, 2005; van Leussen & Escudero, 2015). Evidence from research on the acquisition of L2 rhythm in samples of individual populations of learners, in addition to studies on the development of L2 rhythm over time suggest that, although often exhibiting behavior resembling interference from the L1, second language learners have been found to make some gains toward target-like rhythm production. However, as will be reviewed in the following subsections, the findings from these studies have been largely incongruous with one another.

4.1.3.1 The acquisition of L2 rhythm

Knowing that the rhythm of even within-rhythm-category languages is likely to vary along a continuum (Grabe & Low, 2002), Benet, Gabriel, Kireva, and Pešková (2012) found that native speakers of Italian learning Spanish outside of an academic context produced their L2 Spanish with rhythm patterns similar to those of Italian. This finding patterns closely with that of White and Mattys (2007), who investigated L2 English rhythm patterns in L1 Dutch speakers and found that, relative to speakers of an L2 that was rhythmically dissimilar from the

L1, few gains toward L2-like rhythm were made. Taken together, these findings imply that durational properties of rhythm are transferred to the L2 and, in the case of typologically similar languages, appear to support Flege's (1995) notion of equivalence, but at the suprasegmental level.

Conversely, several studies investigated the L2 acquisition of rhythm in between-rhythm-category languages and, generally, found that learners were able to make gains away from L1 patterns toward those of the L2. Lin and Wang (2005) found that native speakers of Mandarin Chinese (syllable-rhythm), who spoke L2 English at an advanced level, produced rhythm patterns different from their own L1, approaching trends of the L2 (see also Mok & Dellwo, 2008). More relevant to the present study are investigations on the acquisition of syllable-rhythm languages in native speakers of a stress-rhythm language. White and Mattys (2007) researched the L2 Spanish rhythm patterns in native speakers of English that spoke Spanish at an advanced level and found that they displayed evidence of consonantal duration patterns similar to those of native speakers of Spanish and that, at some level, although less so than for consonantal variability, the vocalic duration of L2 Spanish. Specifically, their study found their advanced learners produced global vocalic duration (ΔV and VarcoV) patterns intermediate between native speakers of Spanish and native speakers of English, suggesting a trend of learning. On the other hand, global consonantal duration (ΔC) in L2 Spanish was not different from that of native speakers of Spanish, but both native Spanish and L2 Spanish were different from native speakers of English; however, once normalized, rhythm differences between all conditions, including between native speakers of English and Spanish, were no longer observed. In other words, global consonantal duration in native speakers of English learning Spanish was affected by speech rate and normalizing the rhythm metrics had a neutralizing effect on the patterns observed for second language learners. Patterns similar to those observed in global rhythm were also seen for local, pairwise variability, as well. In sum, it appears that the native speakers of English learning Spanish in White and Mattys had learned consonantal duration patterns similar to those of native speakers of Spanish, but that their vocalic duration patterns resembled an interlanguage status between native speakers of Spanish and native speakers of English.

4.1.3.2 The development of L2 rhythm

Beyond investigating the acquisition of L2 rhythm patterns taken from individual snippets in time in a given population's learning, the field has also benefited from findings stemming from research on the development of L2 rhythm. Prior research on the development of segmental phenomena, as reviewed in Chapter 3, suggests that linguistic experience has an effect on outcomes in second language speech development. Although the data on the relationship between linguistic experience and the development of L2 rhythm is not abundant like it is for the development of segmental phenomena, there is a body of literature, which will be reviewed in the following paragraphs, that suggests that linguistic experience leads to gains in the development of L2 rhythm. For instance, Stockmal, Markus, and Bond (2005) investigated rhythm in native speakers of Latvian (at the time, Latvian's rhythm class had not been identified) and ethnic Russians (Russian being a stress-rhythm language) that spoke Latvian, the latter of which were divided into two conditions as a basis of self-reported measures of language proficiency and use: low-proficiency and high-proficiency. The researchers found that the low-proficiency speakers of Latvian differed from the high-proficiency speakers in terms of raw global variability measures (ΔV and ΔC) and in normalized, local vocalic variability (nPVI-V) and raw, local consonantal variability (rPVI-C), whereas the highly proficient speakers of Latvian were similar to the native speakers of Latvian in most rhythm metrics. Furthermore, low-proficiency speakers spoke more slowly than the high-proficiency speakers of Latvian, suggesting that linguistic experience also has an effect on speech rate.

A large majority of what is known about the development of rhythm has been explored in learners of English as a second language who spoke a syllable-rhythm language natively. Trofimovich and Baker (2006), for example, carried out a cross-sectional study with native speakers of Korean, a syllable-rhythm language, dealing with the effect of linguistic experience on the development of L2 English prosodic phenomena, including speech rate and rhythm. The participants were divided into three groups, each of which had lived for an increasing amount of time in the U.S.: (a) an inexperienced group, (b) a moderately experienced group, and (c) an experienced group. Each participant provided data from six spoken sentences in L2 English. To measure rhythm, the authors developed a metric that took the ratio of the average duration of stressed and unstressed syllables from all tokens. As reviewed previously, the effect of stress on vowel duration and, consequently, syllable duration, would be expectedly greater in a stress-rhythm language than in a syllable-rhythm language. Thus, the lower the ratio, the more the speech of that learner was considered to be patterning after a stress-rhythm language, whereas the closer the ratio was to 1, the more syllable-rhythm the speech was considered to be. The results indicate that only the experienced group was not different from the native English speakers' rhythm. The findings from this study suggest that native speakers of a syllable-rhythm language are capable of acquiring at least one acoustic cue of a stress-rhythm, second language—duration of stressed and unstressed syllables—after a significant amount of experience living in the target culture.

Furthermore, Tortel and Hirst (2010) investigated native speakers of French speaking L2 English. The L2 English rhythm values obtained from the L1 French speakers in their study were different enough to accurately predict to which proficiency condition (low or high) a given individual belonged. Specifically, both consonantal and vocalic normalized interval variability scores (VarcoV and VarcoC) suggested a pattern of development toward target-like rhythm. On the other hand, Li and Post (2014) found that their L1 Mandarin speakers of L2 English exhibited a pattern of development only in vocalic duration patterns, both globally and locally (VarcoV and nPVI-V), whereas there were no significant developmental gains made toward target-like consonantal variability patterns.

4.1.3.3 L2 speech rate

Lastly, in light of the findings discussed previously that rhythm measures are sensitive to changes in speech rate (Barry et al., 2003), a brief discussion on the patterns of speech rate in the L2 is needed. Guion, Flege, Liu, and Yeni-Komshian (2000) elicited speech from 240 native speakers of Italian living in Canada learning English as an L2. They found that length of residence was correlated with speech rate: The longer the individuals had lived in Canada, the faster their speech rate in L2 English was. Similar results have been obtained from other studies measuring speech rate in the L2, as well (Paul Lennon, 1990; Towell, Hawkings, & Bazergui, 1996; Trofimovich & Baker, 2006). This finding implies that speech rate increases in line with linguistic experience, a fact that is relevant to the study of the development of rhythm in second language learners.

4.1.4 Summary of Second Language Rhythm

Taken together, the findings from the studies summarized here regarding the acquisition and development of rhythm patterns in the L2 are mixed. Firstly, the large number of rhythm metrics used in language-rhythm studies—%V, ΔC, ΔV, VarcoV, VarcoC, PVI, rPVI, nPVI, nPVI-V, and rPVI-C, among other study-specific metrics—brings to light the fact that there is still no agreement in the literature as to which are the best at capturing rhythm differences for monolinguals, early bilinguals, or late bilinguals, which can be concluded from the numerous studies cited above (see Henriksen and Fafulas, 2017, for the same conclusion).

Secondly, much of what is known about the development of L2 rhythm comes from studies on native speakers of a syllable-rhythm language learning a stress-rhythm language, the latter often being English. The results from these studies largely suggest that these learners can make gains away from patterns of their L1 toward those of the L2; however, the specific aspects of rhythm that are learned seem to vary from one language to the other, even among typologically similar languages.[18] On the one hand, advanced L1 Spanish speakers of L2 English learned to produce consonantal duration patterns similar to native speakers of English, but the case was not the same for their L2 English vocalic duration patterns (White & Mattys, 2007). On the other hand, second-year university students (presumably of a lower proficiency than the advanced speakers reported in White and Mattys), whose native language was French, made steady gains toward both vocalic and consonantal duration patterns of their L2 English (Tortel & Hirst, 2010). Similarly, Mandarin (syllable-rhythm) speakers learning L2 English made gains only in vocalic duration patterns, but not in consonantal duration (Li & Post, 2014), whereas Korean (also syllable-rhythm) speakers appeared to struggle with the vocalic duration patterns of their L2 English (Trofimovich & Baker, 2006). These disparate findings justify, in part, the present study, whose objective is to better understand the nature of the development of rhythm in second language learners.

Thirdly, while there is data on the relationship between linguistic experience and the development of L2 rhythm, the picture remains incomplete. Is the pattern of development linear? The studies on the development of rhythm largely investigated rhythm in up to two conditions—usually, a low proficiency sample and a high proficiency sample—whose levels of proficiency and how proficiency was calculated was not always clear. Thus, the need for a greater understanding of the pattern of development over time of L2 rhythm justifies this study, as well.

Lastly, unlike the case of segmental phenomena, “research on the acquisition of prosodic features is still in need of a fully-fledged, well-grounded model,” (Simonet, 2013, p. 741) and, while the present study does not set out to do so, the basis of the research is justified

[18] Arvaniti (2009) came to a similar conclusion in her review of rhythm studies in monolingual speakers.

in that it yields further data in this regard to allow for a greater understanding of L2 prosody generally.

To summarize, the goal of the study in the present chapter is to better understand the nature of the development of second-language rhythm through a cross-sectional design of English-speaking, second-language learners of Spanish, specifically in terms of which aspects of rhythm are acquired (vocalic and/or consonantal, and, for that matter, global or local), and at what stages in the trajectory of the learning process the gains are made. As a baseline for comparison, the rhythm patterns of native English/Spanish bilinguals are also investigated. Previous research reviewed here leads to the following hypotheses:

1. Native English/Spanish bilinguals will exhibit evidence of language-specific rhythm patterns between English and Spanish. Specifically, rhythm metrics will capture greater vocalic and consonantal duration variability both locally and globally in English vis-à-vis Spanish.
2. Learners of Spanish will show gains of increasing speech rate in their L2 Spanish as linguistic experience increases.
3. Learners of Spanish as a second language will initially display behavior suggesting interference from their L1: high vocalic and high consonantal duration variability in their L2 Spanish, similar to their L1 English. As linguistic experience increases, however, the learners are expected to produce L2 Spanish with increasingly more target-like (i.e., in the direction of the native bilinguals) rhythm patterns.

4.2 METHOD

4.2.1 PARTICIPANTS

The participants, method of recruitment, and sampling procedure for this experiment are reported in Chapter 2.

4.2.2 MATERIALS

Prior studies examining rhythm patterns in spoken language have varied in their use of laboratory and spontaneous speech for analysis. For example, some studies used only laboratory speech (Dellwo, 2006; Grabe & Low, 2002; Low et al., 2000; Ramus et al., 1999), while others collected or elicited spontaneous speech (Carter & Wolford, 2016; Gut, 2003; Henriksen & Fafulas, 2017; Warner & Arai, 2002; Whitworth, 2002). The data analyzed in this study come from laboratory materials, which are comprised of complete sentences spoken in English and Spanish—eight in each language, all of which were designed by the author of this study. Native speakers of each language confirmed that they were legal utterances in the given language. An effort was made to control the materials for mean number of syllables between languages (cf., English [M = 17.8, SD = 1.3] and Spanish [M = 18.8, SD = 1.5]). Likewise, none of the materials in either language included rhotic (/r, ɾ, ɝ/) or lateral phonemes (/l/), neither did they include glides (i.e., [j, w]) at syllable boundaries. Considering the sonorant quality of these sounds, the latter control was implemented in order to allow for reliable and reproducible segmentation criteria between vocalic and consonantal sequences in the acoustic data. Despite the controls implemented, the materials were created with the intention of being a representative sample of speech in these two languages. To this end, the distribution of stressed and unstressed syllables between languages was not controlled for, neither was pitch accent. The materials do not follow any particular theme or context, but rather were designed as contextually independent one from another; however, in some cases, despite best endeavors to avoid otherwise, some contain words also present in others. The complete set of materials included in this study are found in Table 4.1.

Table 4.1 Materials for experiment on development of rhythm as a function of language, with mean and standard deviation of number of syllables of each item per language.

	English (M = 17.8, SD = 1.3)
1	Dingoes can be viewed as pests at times because they feed on sheep and goats.
2	It's common to think that the most educated succeed in commanding the ship.
3	The sound of my pet cat dying in the basement keeps haunting me.
4	I'm excited to visit my cousin that has a cabin by the sea.
5	Baby chimpanzees enjoy picking and eating bananas as a snack.
6	My mom says that anchovies and sausage taste good as pizza toppings.
7	Dad posted on Facebook that he got a new job as a mechanic.
8	I vote that he should give back the money that they donated to the campaign.
	Spanish (M = 18.8, SD = 1.5)
1	Te digo que Pepe nunca pasa tiempo en esta discoteca. *I'm telling you that Pepe never spends time in this club.*
2	Vivimos con una muchacha que ni sabe dónde deja sus cosas. *We live with a girl that has no clue where she leaves her things.*
3	A mí no me gusta que tú no sepas dónde queda mi casa. *(Emphasis) I don't like that you don't know where my house is.*
4	Hace mucho tiempo que Paco no come tacos de gambas. *It's been a long time since Paco has eaten shrimp tacos.*
5	Mis cuñados viajan de vez en cuando desde San Agustín a Nevada. *My brothers (and sisters)-in-law sometimes travel from San Agustin to Nevada.*
6	Me cuenta que sus hijos necesitan zapatos estas Navidades. *(S/he) tells me that their children need shoes this Christmas.*
7	Viven muchos tipos de pequeños gusanos en humanos. *Lots of types of small worms live in humans.*
8	Es un éxito cuando tenemos muchas ovejas y vacas. *It's a success when we have lots of sheep and cows.*

4.2.3 RECORDINGS

The recording of materials took place after the completion of the background questionnaire and BLP (see Chapter 2). The English-language materials were recorded first, after completing the LexTALE and prior to recording English-language materials for the experiment reported in Chapter 3. The Spanish-language materials were recorded second, after completing the LexTALE-Esp and before recording the Spanish-language materials for the experiment reported in Chapter 3. In other words, for all participants, the direction of recording was the following: firstly, English-language materials; secondly, Spanish-language materials.

The materials were presented to each participant using PsychoPy 2 (Peirce et al., 2019). In all cases, for the English and Spanish portions of the experiment, the instructions were written in English. This was done to ensure that all LLs of all proficiency levels that were taking part in the experiment were able to clearly understand the task. In the instructions provided for each language condition, the participants were told that the materials that they were about to read were in English or Spanish, respectively. Following White and Mattys (2007), participants were instructed to repeat each sentence aloud in their normal, conversational voice at a rate that felt natural and comfortable. They were asked to say the entire sentence without pausing while speaking and to repeat the sentence if they made a mistake. This experiment was self-paced: Each participant had as much time as needed to read the sentence and say it aloud before moving on to the next, which was done by pressing a button on the keyboard. Participants were told to ensure that they had finished saying the sentence before pressing the button to eliminate additional noise in the sound booth. Following each sentence, a red fixation cross appeared in the center of the screen for 500 milliseconds.

Each sentence is considered an individual token, and each was repeated three times in a different, random order. Thus, each participant provided 48 tokens (8 sentences × 3 repetitions × 2 languages), for a study total of 2,400 tokens (48 tokens × 50 participants). For each participant, the most fluent of the three repetitions of each token, defined as that which contained the fewest pauses and the most natural-sounding speech, was selected as the representative sample of that token for that individual and was that which was submitted to analysis. Thus, for each speaker, 16 tokens were submitted to the final data set, for a total of 800 tokens (50 participants × 8 sentences × 2 languages).

All recordings were carried out in a sound-attenuated booth using a head-mounted, dynamic microphone and a Fostex DC-R302 digital recorder with built-in microphone pre-amplification and analog-to-digital interface. All signals were digitized at 44.1 kHz, 16-bit quantization.

4.2.4 Segmentation of Acoustic Data

The acoustic data were segmented by hand in Praat (Boersma & Weenink, 2018) using a combination of spectral information and waveform shape to guide the segmentation. Contiguous vowels were considered a single vocalic interval, and, likewise, contiguous consonants were considered a single consonantal interval. Vowel boundaries were marked using the following criteria:

Onset

- Phrase initial: Zero crossing of first upward departure in which modal voicing and formant structure were observable.
- After nasals: Point at which an increase in formant intensity and waveform height, indicating vowel onset, were observable.
- After fricatives: Zero crossing of first upward departure in which modal voicing and decreased aperiodic noise were observable.
- After stops: Zero crossing of first upward departure in which modal voicing was observable following the aspiration.

Offset

- Phrase final: Point at which a noticeable decrease in F2 intensity, corresponding with a decrease in waveform height, was observed.
- Before nasals: The point at which a decrease in formant intensity and waveform height is observable.

- Before fricatives: The point at which waveform structure resembled aperiodic noise, corresponding with an end of formant structure.
- Before stops: The point at which a noticeable drop in formant structure and waveform height was observable, just prior to closure.

For most English speakers, there is a phonological tendency to insert a glottal stop prior to a word-initial vowel (Bissiri, Lecumberri, Cooke, & Volín, 2011), particularly in continuous speech, and this pattern was true for portions of the data collected from LLs in this study. An opposite pattern is observed in monolingual Spanish, a language in which, due to resyllabification, word boundaries rarely begin with a vowel in connected speech (Hualde, 2005, p. 87). In light of these facts, when a glottal stop appeared prior to a word-initial vowel, the corresponding silent period was considered as part of the vowel phonologically and, therefore, as part of the vocalic sequence. One exception to this is when a glottal stop allophone of a stop phoneme in word-final position surfaced before a word-initial vowel (e.g., mom says tha[ʔ] anchovies), and objectively distinguishing between the word-final glottal stop and potential word-initial, vowel-induced glottal stop was not possible: In this scenario, any creaky voice leading up to the glottal stop was considered vocalic, whereas the entirety of the silent period was consonantal. For similar phonological motives, devoiced vowels were also considered as part of the vocalic sequence. However, in those cases in which vowels were deleted in favor of syllabic consonants, such as the case of the deleted schwa in *cousin* [ˈkʌzn̩], for instance, the syllabic consonant was considered as part of the consonantal sequence.

Objective criteria were used to determine the boundaries of utterances. Some of the materials begin or end with a consonant, while others begin or end with a vowel. The initial boundary of all utterances was considered the onset of modal voicing of the first vowel of the utterance and was marked at the upward zero crossing. Due to the difficulty of reliably segmenting phrase-final vowels, which display a decrease of formant intensity, possibly due to their propensity to lengthen in phrase-final position (Klatt, 1976), the final boundary of all utterances was determined to be the offset of the F2 of the second-to-last vowel of the utterance. In other words, all utterance durations began and ended with a vowel. On the other hand, if the beginning of the utterance were always the first segment of the item, consonantal rhythm measures may be affected, as detecting the beginning of the stop closure in utterance-initial stops would be guesswork, at best. These criteria allowed for more reliable and reproducible segmentation across tokens, speakers, and language conditions.

Any intra-sentential pauses produced by the speakers were excluded from the analysis, as were sequences affected by pre-pausal lengthening. In the event that a pause took place between two consonants, the durations of the preceding and following consonant sequences were summed together and counted as one consonantal sequence. In the event that a pause followed a vowel and preceded a stop consonant, and, therefore, an objective way of distinguishing the pause from the closure duration of the stop was not possible, the segmentation method followed the criteria reported in White and Mattys (2007): The short duration of the interval between the end of formant structure of the preceding vowel (where the vowel offset would have been marked regardless of the presence of a pause) and on to the onset of silence was summed to the duration of the following stop beginning with the release of the burst, and, together, both were considered as part of the consonantal sequence. In other words, the silent period of the pause and any closure duration of the stop were both removed, and the short interval between the end of the preceding vowel and the onset of silence was added to that of the following consonant segment. In regard to pre-pausal lengthening, in the event that the removal of a sequence due to this phenomenon would have created one, compounded sequence of either vowels or consonants (i.e., a single, lengthened sequence combined of two similar sequences prior to and following the pause), an additional sequence (respective of the type of sequence lengthening in question) was excluded from the syllable in question in order to avoid artificially inflating any of the vocalic or consonantal duration ratios, and, subsequently, the rhythm metrics. An example of the segmentation procedure is available in Figure 4.1.

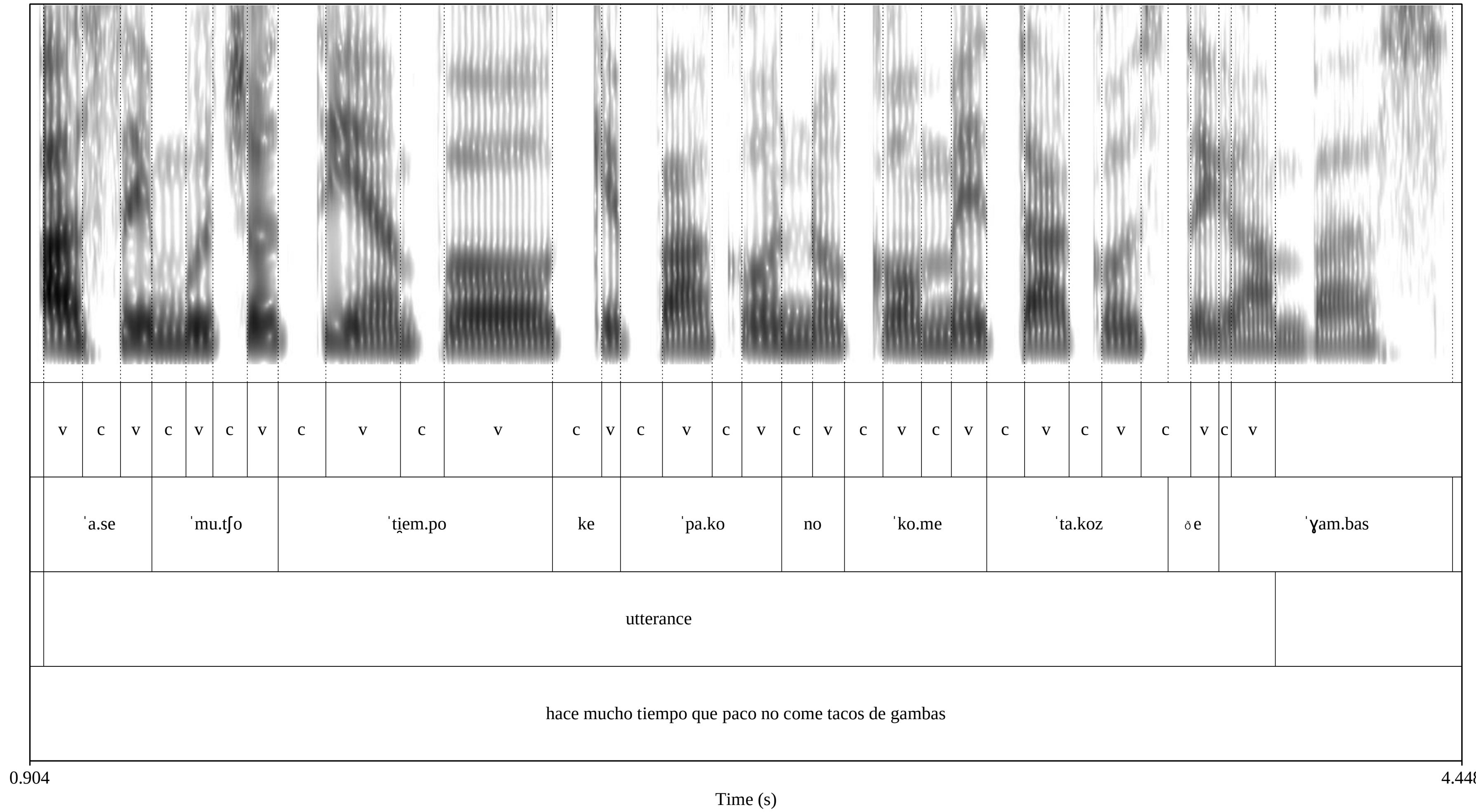


Figure 4.1 Illustrative example of data segmentation from the item *Hace mucho tiempo que Paco no come tacos de gambas* “It’s been a long time since Paco has eaten shrimp tacos.” Vocalic sequences are marked as V and consonantal sequences are marked as C. A broad phonetic transcription of the presumed intended speech is provided. Utterance duration is marked as beginning at the point of upward departure from zero crossing of the onset of model voicing of the first vowel in the sentence and ending at the end of the formant structure of the second-to-last vowel in the sentence.

4.2.5 Coding and Statistical Analysis

In total, after eliminating some tokens due to mispronunciations (n = 11), a total of 789 tokens were included in the final analysis ((50 participants × 8 sentences × 2 languages) – 11 misses). Several different rhythm metrics—ΔV, ΔC, %V, VarcoV, VarcoC, nPVI-V, and rPVI-C, as described earlier—were derived from these data using a Praat script. The names of these measures and how they are calculated are summarized here in Table 4.2.

Table 4.2 Rhythm measures calculated in this study: ΔV, ΔC, %V, VarcoV, VarcoC, nPVI-V, rPVI-C, and speech rate (Dellwo, 2006; Grabe & Low, 2002; Ramus et al., 1999; White & Mattys, 2007).

Metric	Calculation
ΔV	SD of vocalic intervals, multiplied by 1000.
ΔC	SD of consonantal intervals, multiplied by 1000.
%V	Vocalic intervals summed and divided by utterance duration, multiplied by 100.
VarcoV	SD of vocalic intervals, divided by the mean of vocalic intervals, multiplied by 100.
VarcoC	SD of consonantal intervals, divided by the mean of consonantal intervals, multiplied by 100.
nPVI-V	Absolute difference in duration between sequential vowel intervals, divided by the mean of the two intervals, summed across all samples, divided by the number of samples, multiplied by 100.
rPVI-C	Absolute difference in duration between sequential consonantal intervals, summed across all samples, divided by the number of samples, multiplied by 1000.
Speech Rate	Sum of number of vocalic and consonantal intervals divided by utterance duration.

As reviewed in the Introduction to this chapter, it is expected that native speakers of English would exhibit greater values for ΔV, ΔC, VarcoV, VarcoC, nPVI-V, and rPVI-C relative to native speakers of Spanish. Conversely, for %V, English is hypothesized to have a lower value than Spanish. Lastly, speech rate is theorized to be greater (i.e., faster) for native speakers of English and Spanish than for learners of Spanish.

The data were then submitted to regression models that were fit using the *lm* function in R (R Core Team, 2015) via the RStudio (RStudio Team, 2016) graphical user interface. Prior to analysis, the data were averaged over items as a function of Language (English, Spanish) and LE (LL1, LL2, LL3, LL4, BL), such that there were in total two observations of each rhythm metric for each speaker: one for English and one for Spanish. Next, several linear regression models were fit to explore the relationship between each of the rhythm metrics and each condition of Language (English, Spanish), and, where applicable, LE. The BL and the LL data were analyzed separately. In the case of the LL conditions, LE was coded as a numeric variable from −1.5 to +1.5—such that the conditions centered around 0—in the following manner: LL1 = −1.5, LL2 = −0.5, LL3 = +0.5, LL4 = +1.5. Language was submitted to the analyses as a categorical variable with English as the baseline. For the LL data, the full model included the LE × Language interaction, whereas the nested model excluded it. A partial *F* test between the full and the nested models was used to determine the best-fit model. The response variables for which the full model was the better fit are indicated in bold in Table 4.6. The adjusted R^2 is reported for all models.

4.3 RESULTS

The results from BL data are presented first in order to have a baseline understanding of the rhythm patterns of English and Spanish using a within-subjects design in which the participants were early learners of both languages. Following the BL data, that of the LL speakers were analyzed. For all conditions (BL and LL), the data are discussed in the following order, from first to last: (a) speech rate, (b) vocalic and consonantal interval variability (ΔV, ΔC, %V, Varco C, and VarcoV), and (c) vocalic and consonantal pairwise variability (nPVI-V and rPVI-C). Grouping the results in this way allows for an analysis of changes in speech rate across speakers between languages, followed by a comparison of the vocalic and consonantal rhythm patterns globally (i.e., at the interval level), and, subsequently, a comparison of the vocalic and consonantal rhythm patterns locally (i.e., comparisons of one vowel or consonant sequence to the next). Together, these results inform about these patterns in speakers that grew up speaking both languages from an early age and those who began learning Spanish as an adult. Importantly, by including speech rate in the analysis, one can determine whether the learning of L2 rhythm is taking place independent from gains in speech rate in the target language.

4.3.1 RESULTS: BL SPEAKERS

4.3.1.1 Descriptive statistics

The descriptive statistics presented in Table 4.3 for speech rate, the vocalic and consonantal interval variability rhythm metrics (ΔV, ΔC, VarcoV, and VarcoC), and the vocalic and pairwise variability metrics (nPVI-V and rPVI-C) suggest, overall, similar speech rates between English and Spanish and greater vocalic and consonantal variability in English than in Spanish among the BL speakers. However, %V, the percentage of the utterance that is vocalic, was similar in English and Spanish. This seemingly peculiar finding is outlined further in the following section and in the discussion.

Table 4.3 Descriptive statistics for speech rate, vocalic and consonantal **interval** variability (ΔV, ΔC, %V, VarcoV, and VarcoC) and vocalic and consonantal **pairwise** variability (nPVI-V and rPVI-C) for BL speakers (mean followed by standard deviation in parentheses).

	English	Spanish
Speech rate		
Speech Rate	10.08 (2.18)	10.73 (1.55)
Interval metrics		
ΔV	43 (15)	30 (11)
ΔC	52 (15)	35 (10)
%V	49 (5)	50 (4)
VarcoV	43 (10)	32 (8)
VarcoC	47 (7)	36 (8)
Pairwise metrics		
nPVI-V	50 (11)	35 (9)
rPVI-C	59 (21)	39 (13)

4.3.1.2 Inferential statistics

The results from the inferential statistics reported in Table 4.4 are summarized as follows. Firstly, a model was fit with Speech Rate as the response variable and Language (English, Spanish) as a categorical explanatory variable. Language does not predict a significant amount of variance in Speech Rate and Speech Rate does not significantly increase in Spanish. These results suggest that the BL speakers maintained a similar speech rate in both English and Spanish.

Table 4.4 Inferential statistics for BL speakers' speech rate, vocalic and consonantal **interval** rhythm metrics, and vocalic and consonantal **pairwise** rhythm metrics. Linear regression models were fit with the rhythm statistic (speech rate, ΔV, ΔC, VarcoV, VarcoC, %V, nPVI-V, rPVI-C) as response variable separately and Language (English, Spanish) as categorical explanatory variable with English as the baseline. Adjusted R^2 is reported for all models. Significant coefficients are bolded.

	Regression Model				Regression Coefficient (Spanish)				
	R^2	*df*	*F* ratio	*p* value	*b*	*SE*	*t*	*df*	*p* value
Speech rate									
Speech Rate	.00	(1, 16)	1.10	> .30	.69	0.67	1.10	16	> .30
Interval metrics									
ΔV	.46	(1, 16)	15.36	< .002	**–13**	**3.31**	**–3.92**	**16**	**< .002**
ΔC	.56	(1, 16)	22.3	< .001	**–17**	**3.52**	**–4.72**	**16**	**< .001**
VarcoV	.64	(1, 16)	31.77	< .001	**–11**	**1.92**	**–5.64**	**16**	**< .001**
VarcoC	.84	(1, 16)	86.8	< .001	**–11**	**1.21**	**–9.32**	**16**	**< .001**
%V	< .01	(1, 16)	< 1	> .80	.20	1.01	0.21	16	> .80
Pairwise metrics									
nPVI-V	.69	(1, 16)	38.92	< .001	**–15**	**2.44**	**–6.24**	**16**	**< .001**
rPVI-C	.59	(1, 16)	25.55	< .001	**–20**	**3.98**	**–5.05**	**16**	**< .001**

Secondly, separate models were fit with ΔV, ΔC, VarcoV, VarcoC, nPVI-V, and rPVI-C as the response variable and Language (English, Spanish) as a categorical explanatory variable. Language predicts a significant amount of variance in all six of these rhythm metrics and the Language (Spanish) coefficient was also significant in all models. These results suggest that ΔV, ΔC, VarcoV, VarcoC, nPVI-V, and rPVI-C decrease significantly in Spanish compared to English. The direction of these results can be summarized as follows: (a) in the BL speakers, the standard deviation and the standard deviation relative to the mean, of both vocalic and consonantal intervals, is greater in English than in Spanish; and, (b) differences of absolute duration from one vowel sequence to another and from one consonant sequence to another were greater in English than in Spanish. Together, these findings imply that the native BL speakers in this study exhibit unique rhythm patterns in their English and Spanish and that these patterns follow the direction that would be expected from these languages based on previous literature.

Lastly, a model was fit for %V as the response variable and Language (English, Spanish) as a categorical explanatory variable. Unlike the others reported thus far, %V does not predict a significant amount of variance in %V (R^2 < .01, F < 1) and %V in Spanish does not significantly change compared to English (b = .20, t(16) = 0.21, p > .80). In other words, the vowel percentage of an utterance in Spanish is not significantly less than that of an utterance in English for the BL speakers.

4.3.2 RESULTS: LL SPEAKERS

As there were no significant differences in %V between English and Spanish in the BL speakers, the data for %V is considered irrelevant for the LLs and, thus, were excluded from the analysis for the LL speakers.

4.3.2.1 Vocalic and consonantal interval variability

4.3.2.1.1 Descriptive statistics and data visualization

The descriptive statistics reported in Table 4.5 for speech rate and the vocalic and consonantal interval variability rhythm patterns suggest the following: (a) relatively static speech rate across all LL conditions in English, but an increase in speech rate in Spanish as linguistic experience increased; (b) greater vocalic variability in English than in Spanish, overall, and a trend of decreasing vocalic variability in Spanish as linguistic experience increases among the LL conditions; and (c) slightly greater consonantal variability in English than in Spanish, overall, and a trend of decreasing variability in Spanish as linguistic experience increases among the LL conditions.

Table 4.5 Descriptive statistics for speech rate, vocalic and consonantal **interval** rhythm metrics—ΔV, ΔC, VarcoV, and VarcoC—and vocalic and consonantal **pairwise** rhythm metrics—nPVI-V, rPVI-C, and nPVI-C—for LL conditions (mean followed by standard deviation in parentheses).

Metric	Conditions	English	Spanish
Speech Rate			
Speech Rate	LL1	10.46 (1.47)	8.53 (0.85)
	LL2	10.56 (1.54)	9.22 (1.20)
	LL3	10.63 (1.89)	9.66 (1.39)
	LL4	10.71 (1.61)	10.25 (1.23)
Interval metrics			
ΔV	LL1	40 (10)	40 (11)
	LL2	41 (11)	37 (14)
	LL3	41 (13)	36 (13)
	LL4	41 (11)	33 (10)
ΔC	LL1	49 (11)	49 (15)
	LL2	47 (13)	43 (15)
	LL3	50 (13)	42 (14)
	LL4	47 (14)	38 (12)
VarcoV	LL1	44 (9)	34 (8)
	LL2	44 (9)	32 (9)
	LL3	46 (9)	34 (10)
	LL4	45 (9)	34 (8)
VarcoC	LL1	47 (6)	40 (10)
	LL2	46 (7)	39 (11)
	LL3	48 (6)	40 (11)
	LL4	46 (7)	38 (10)
Pairwise metrics			
nPVI-V	LL1	53 (13)	38 (7)
	LL2	54 (11)	36 (10)
	LL3	55 (10)	37 (9)
	LL4	54 (10)	36 (8)
rPVI-C	LL1	58 (20)	58 (21)
	LL2	54 (20)	51 (18)
	LL3	58 (19)	48 (17)
	LL4	54 (19)	43 (13)
nPVI-C	LL1	58 (14)	47 (13)
	LL2	55 (12)	46 (12)
	LL3	58 (12)	47 (12)
	LL4	55 (12)	43 (12)

In terms of ΔV, the standard deviation of the duration (ms) of vocalic intervals, and ΔC, the standard deviation of the duration (ms) of consonantal intervals, a similar trend was observed: firstly, in English, the values were all similar; and, secondly, in Spanish, the LL1 condition's English and Spanish were similar, but Spanish variability decreased steadily as linguistic experience increased. These data, along with those of the BL speakers for comparison, are presented in a scatter plot in Figure 4.2.

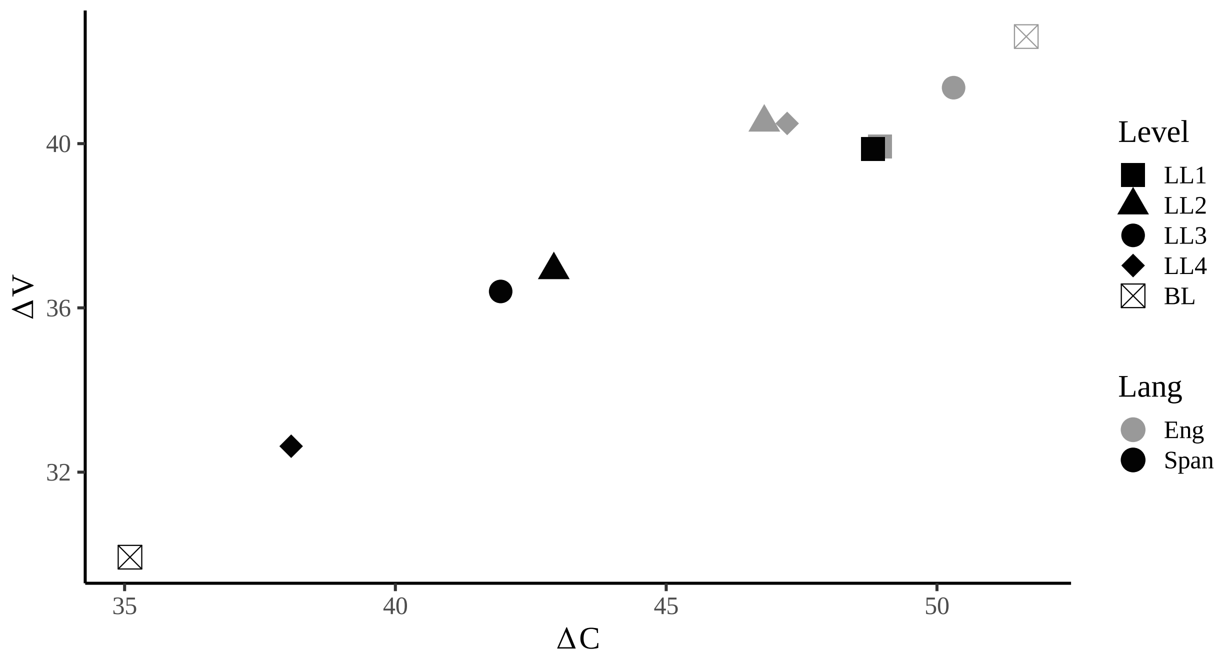


Figure 4.2 Scatter plot of the mean ΔC and mean ΔV of the LL and BL conditions. English data are plotted in grey points and Spanish data in black. The Linguistic Experience conditions are represented as follows: LL1 = square, LL2 = triangle, LL3 = circle, LL4 = diamond, BL = box.

VarcoV and Varco C, the variation coefficients of ΔV and ΔC (or, in other words, ΔV and ΔC relative to their mean value, respectively) displayed the following pattern: firstly, Varco V and VarcoC were greater in English than in Spanish for all conditions, and, secondly, VarcoV and VarcoC slightly decreased as linguistic experienced increased. These data, along with those of the BL speakers for comparison, are presented in a scatter plot in Figure 4.3.

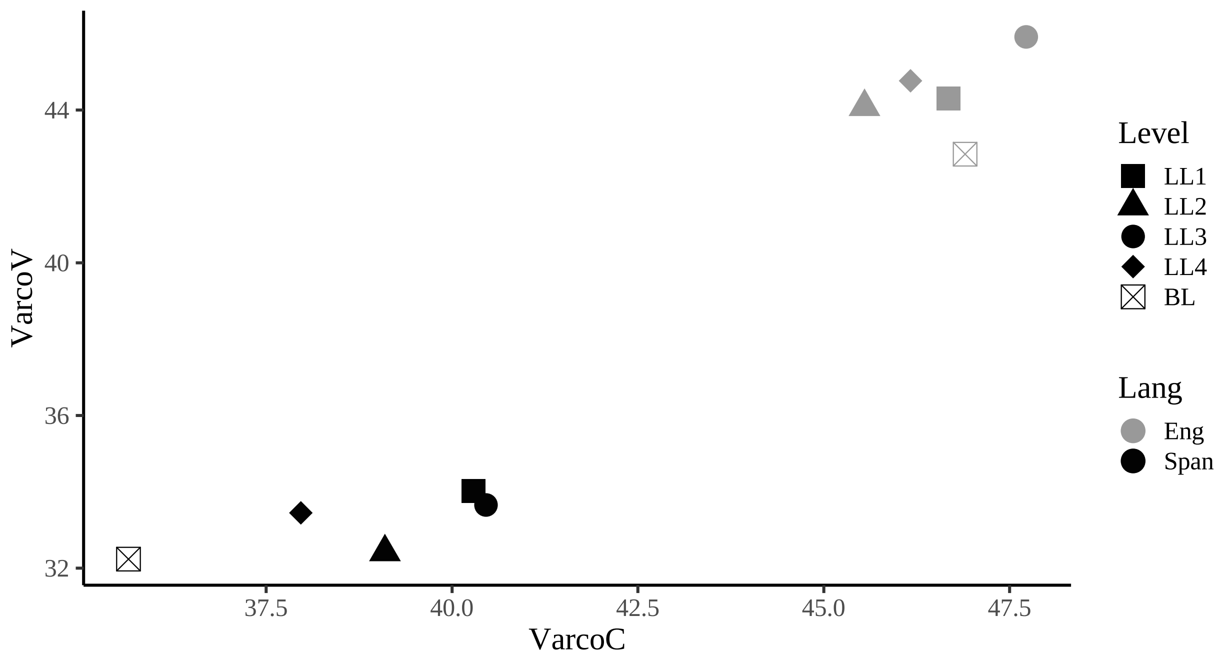


Figure 4.3 Scatter plot of the mean VarcoC and mean VarcoV of the LL and BL conditions. English data are plotted in grey points and Spanish data in black. The Linguistic Experience conditions are represented as follows: LL1 = square, LL2 = triangle, LL3 = circle, LL4 = diamond, BL = box.

4.3.2.1.2 Inferential statistics

The results of the linear regression models for the interval rhythm metrics, as presented in Table 4.6, can be summarized as follows: Firstly, a model was fit with Speech Rate as the response variable and LE (LL1, LL2, LL3, LL4), Language (English, Spanish), and the LE × Language interaction as explanatory variables. The model suggests that these variables together predict a significant amount of variance in Speech Rate. The LE coefficient was not significant, suggesting that speech rate did not increase significantly in English as linguistic experience increased. The Language (Spanish) coefficient was significant, implying that, controlling for LE, speech rate decreased significantly in Spanish. Lastly, the LE × Language interaction was also significant, which suggests that speech rate increased significantly in Spanish as linguistic experienced increased, and the direction was toward that of the Spanish of the BL speakers.

Table 4.6 Inferential statistics for LL speakers' speech rate, vocalic and consonantal **interval** rhythm metrics (ΔV, ΔC, VarcoV, VarcoC), and vocalic and consonantal **pairwise** rhythm metrics (nPVI-V, rPVI-C, and nPVI-C). Multiple linear regression models were fit with the rhythm statistic as response variable and LE (coded numerically), Language (English, Spanish; ref = English), and LE × Language as explanatory variables. The regression models determined to be the better fit as per a partial *F* test are bolded. Coefficient estimates whose *df* of *t* = 79 are reported from the nested model (LE + Lang); those whose *df* of *t* = 78 are reported from the full model (LE + Lang + LE × Lang). Significant coefficients are in bold.

	Regression Model					Regression Coefficients					
		R^2	*df*	*F* ratio	*p* value	coefficient	*b*	*SE*	*t*	*df*	*p* value
Speech Rate	~ LE + Lang	.29	(2, 79)	15.83	< .001	LE	0.06	0.14	0.58	78	> .50
	(+ LE × Lang)	**.31**	**(3, 78)**	**12.94**	**< .001**	**Lang (Span)**	**–1.84**	**0.42**	**–4.44**	**78**	**< .001**
						LE × Lang (Span)	**0.46**	**0.20**	**2.33**	**78**	**< .03**
ΔV	**~ LE + Lang**	**.17**	**(2, 79)**	**9.46**	**< .001**	LE	–1.02	0.60	–1.71	79	> .05
	(+ LE × Lang)	.20	(3, 78)	7.86	< .001	**Lang (Span)**	**–5.07**	**1.27**	**–4.00**	**79**	**< .001**
						LE × Lang (Span)	–2.34	1.18	–1.99	78	= .051
ΔC	~ LE + Lang	.27	(2, 79)	16.15	< .001	LE	–0.14	0.82	–0.17	78	> .80
	(+ LE × Lang)	**.32**	**(3, 78)**	**13.44**	**< .001**	Lang (Span)	–1.20	2.46	0.49	78	> .60
						LE × Lang (Span)	**–2.84**	**1.16**	**–2.45**	**78**	**< .02**
VarcoV	**~ LE +Lang**	**.78**	**(2, 79)**	**144.8**	**< .001**	LE	0.20	0.32	0.62	79	> .50
	(+ LE × Lang)	.78	(3, 78)	95.41	< .001	**Lang (Span)**	**–11.54**	**0.68**	**–17.01**	**79**	**< .001**
						LE × Lang (Span)	–0.16	0.65	–0.25	78	> .80
VarcoC	**~ LE + Lang**	**.69**	**(2, 79)**	**91.23**	**< .001**	LE	–0.24	0.26	–0.95	79	> .30
	(+ LE × Lang)	.69	(3, 78)	61.95	< .001	**Lang (Span)**	**–7.29**	**0.54**	**–13.47**	**79**	**< .001**
						LE × Lang (Span)	–0.67	0.51	–1.31	78	> .10
nPVI-V	**~ LE + Lang**	**.82**	**(2, 79)**	**190.4**	**< .001**	LE	–0.06	0.43	–0.14	79	> .80
	(+ LE × Lang)	.82	(3, 78)	125.7	< .001	**Lang (Span)**	**–17.74**	**0.91**	**–19.51**	**79**	**< .001**
						LE × Lang (Span)	–0.38	0.87	–0.44	78	> .60
rPVI-C	~ LE + Lang	.26	(2, 79)	15.04	< .001	LE	–0.73	1.08	–0.67	78	> .50
	(+ LE × Lang)	**.30**	**(3, 78)**	**12.44**	**< .001**	Lang (Span)	–0.80	3.23	–0.25	78	< .001
						LE × Lang (Span)	**–3.60**	**1.53**	**–2.35**	**78**	**< .03**
nPVI-C	**~ LE + Lang**	**0.72**	**(2, 79)**	**103.8**	**< .001**	**LE**	**–0.75**	**0.36**	**–2.08**	**79**	**< .05**
	(+ LE × Lang)	.72	(3, 78)	69.93	< .001	**Lang (Span)**	**–10.83**	**0.76**	**–14.26**	**79**	**< .001**
						LE × Lang (Span)	–0.82	0.72	–1.14	78	> .20

Secondly, separate models were fit with ΔV and ΔC as the response variable and LE (LL1, LL2, LL3, LL4), Language (English, Spanish), and the LE × Language interaction as explanatory variables. These variables combined predict a significant amount of variance in ΔV and ΔC. On the one hand, the LE coefficient was not significant for either metric, suggesting that ΔV and ΔC in English did not increase significantly across LL conditions. The Language (Spanish) coefficient was significant for ΔV, implying that, controlling for LE, ΔV decreased in Spanish vis-à-vis English. However, the Language (Spanish) coefficient was not significant for ΔC, meaning that, controlling for LE, the standard deviation of consonantal intervals did not decrease in Spanish compared to English. Although the LE × Language interaction was not significant for ΔV with $\alpha = 0.05$, there was a trend of development closely approaching statistical significance ($p = 0.051$) across LL conditions for ΔV. However, for ΔC, the LE × Language interaction was significant, meaning that the slope for English and that for Spanish were significantly different across the LL conditions. This finding suggests that a significant trend of development was observed for ΔC: As linguistic experience increased, the values for ΔC in Spanish decreased, and they decreased toward the values obtained from the BL speakers in Spanish.

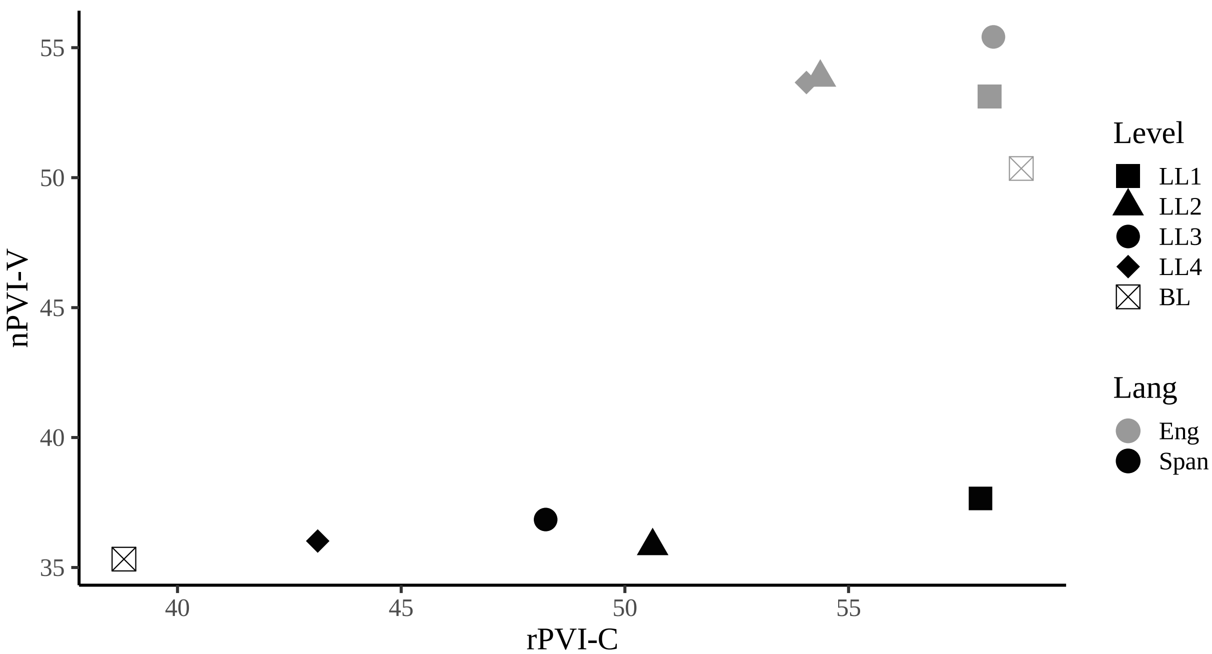
Subsequently, separate models were fit with VarcoV and VarcoC as the response variable and LE (LL1, LL2, LL3, LL4), Language (English, Spanish), and the LE × Language interaction as explanatory variables. The results were the same for both metrics: The variables together predicted a significant amount of variance in the response. The LE coefficient was not significant, suggesting that neither VarcoV nor VarcoC increased significantly in English as a function of LL condition. The Language (Spanish) coefficient was significant, however, suggesting that VarcoV and VarcoC decreased in Spanish vis-à-vis English, controlling for LE. However, the LE × Language interaction was not significant for either metric, meaning that the slope for English and that for Spanish were not significantly different across LL conditions for VarcoV and VarcoC. Together, these results indicate that, controlling for LE, VarcoV and VarcoC decrease in Spanish vis-à-vis English toward target-like values, but that there is not a significant trend of development for either metric as linguistic experience increases.

4.3.2.2 Vocalic and consonantal pairwise variability

4.3.2.2.1 Descriptive statistics

The descriptive statistics presented in Table 4.5 for the vocalic and consonantal pairwise variability rhythm metrics largely show a trend of development: While the values for all metrics are relatively the same in English, the values in Spanish show a trend toward a decrease in vocalic and consonantal pairwise variability as linguistic experience increases.

A visual inspection of the data in Figure 4.4 suggests that, on the one hand, nPVI-V, the normalized pairwise-variability difference for vowel intervals, showed a trend in which nPVI-V was less in Spanish than in English for all conditions and decreased slightly as linguistic experience increased. On the other hand, rPVI-C displayed a trend of development: as linguistic experience increased, rPVI-C steadily decreased. However, because speech rate changed significantly across conditions in Spanish, it was determined necessary to calculate a normalized consonantal pairwise variability index (nPVI-C) to control for changes in speech rate (descriptive statistics are provided in Table 4.5). A visual inspection of the data suggests that the pattern between rPVI-C and nPVI-C differs slightly: a greater developmental trend is observed in the raw data than in the normalized data, although some evidence of a trend remains in the latter.



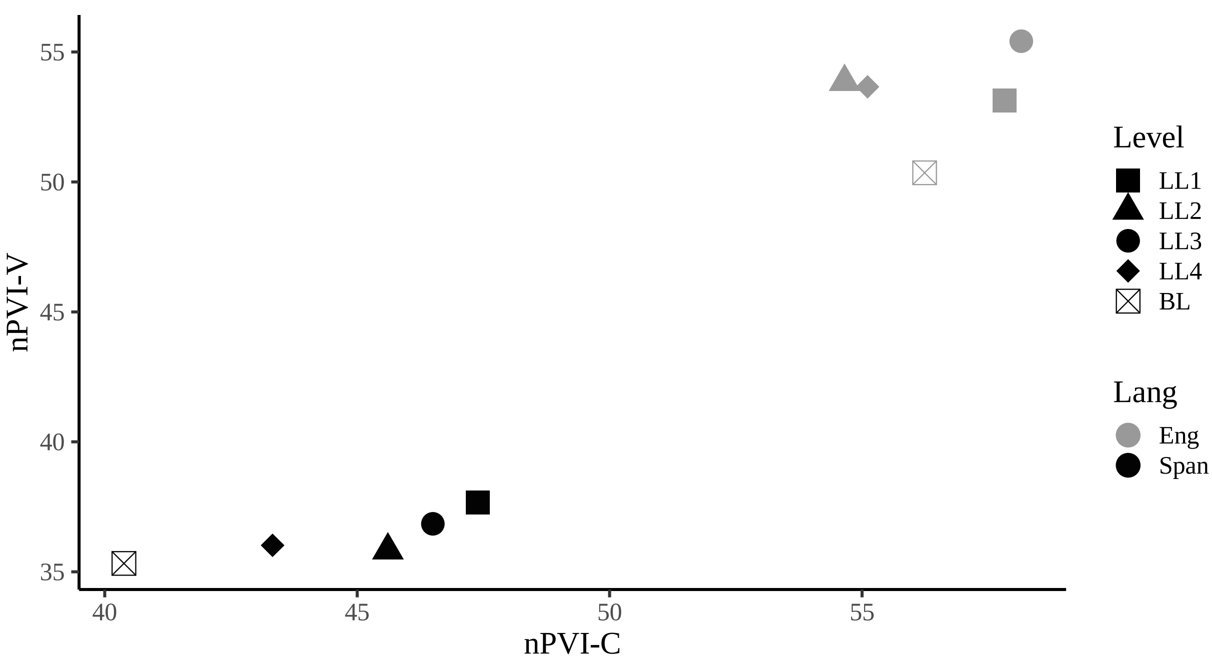


Figure 4.4 Scatter plot of the mean rPVI-C and mean nPVI-V (above) and mean nPVI-C and mean nPVI-V (below) of the LL and BL conditions. English data are plotted in grey points and Spanish data in black. The Linguistic Experience conditions are represented as follows: LL1 = square, LL2 = triangle, LL3 = circle, LL4 = diamond, BL = box.

4.3.2.2.2 Inferential statistics

The inferential statistics presented in Table 4.6 for the pairwise variability metrics can be summarized as follows: Firstly, separate models were fit with nPVI-V and rPVI-C as the response variable and LE (LL1, LL2, LL3, LL4), Language (English, Spanish), and the LE × Language interaction as explanatory variables. The results from the models suggest that these variables together predict a significant amount of variance in nPVI-V and rPVI-C. The LE coefficient was not significant for either metric, suggesting that nPVI-V and rPVI-C in English did not increase significantly across LL conditions. The Language (Spanish) coefficient was significant for nPVI-V only, suggesting that, controlling for LE, nPVI-V decreased in Spanish vis-à-vis English, but that rPVI-C did not. The LE × Language interaction was not significant for nPVI-V, suggesting that the slope for English and that for Spanish were not significantly different across LL conditions for nPVI-V, implying the absence of development. However, the LE × Language interaction was significant for rPVI-C, meaning that the slope for English and that for Spanish were significantly different across LL conditions for rPVI-C. These results suggest that, controlling for LE, nPVI-V and rPVI-C were significantly less in Spanish than in English, and that this effect was in the direction of the target-like norms exhibited by the BL speakers. However, a significant trend of development was observed only for rPVI-C: As linguistic experience increased, rPVI-C values decreased toward target-like production.

Lastly, as mentioned previously, it was determined necessary to also calculate a normalized consonantal pairwise variability metric—nPVI-C—for the LL conditions. Thus, a model was fit with nPVI-C as the response variable and LE (LL1, LL2, LL3, LL4), Language (English, Spanish), and the LE × Language interaction as explanatory variables. The variables combined predict a significant amount of variance in nPVI-C. Unlike with the other rhythm metrics, the LE coefficient was significant, suggesting that nPVI-C did decrease significantly in English across LL conditions in the direction of L2 Spanish. The Language (Spanish) coefficient was also significant, suggesting that, controlling for LE, nPVI-C decreased in Spanish vis-à-vis English. However, the LE × Language interaction was not significant, meaning that the slope for English and that for Spanish are not significantly different across LL conditions for nPVI-C. The finding that the LE coefficient was significant, and that the effect is in the direction of target-like rhythm patterns, implies the possibility of L1–L2 bidirectionality in the LL speakers.

4.4 DISCUSSION

In this study, the development of the rhythm patterns of a second language (Spanish) whose rhythm category (syllabled-timed) differs from that of the first language (English, stress-rhythm) has been investigated in adult speakers in a within-subjects design. Likewise, rhythm was measured in both English and Spanish in early English/Spanish bilinguals. Rhythm patterns were analyzed globally (i.e., considering the variability patterns of either vocalic or consonantal intervals at large) via interval variability metrics, including ΔV, ΔC, VarcoV, VarcoC, and %V. They were also analyzed locally (i.e., considering the absolute difference between sequential vocalic and, separately, consonantal, intervals) via pairwise variability metrics, including nPVI-V, rPVI-C, and nPVI-C. Linear regression models were fit using the rhythm metric as the response variable and, in the case of the BL speakers, Language (English, Spanish) as the explanatory variable, or, in the case of the LL conditions, LE (LL1, LL2, LL3, LL4), Language (English, Spanish), and the LE × Language interaction as the explanatory variables. The descriptive statistics (presented in Table 4.3 and Table 4.5) together with the inferential statistics (presented in Table 4.4 and Table 4.6) lead to the conclusions outlined in the following subsections.

4.4.1 BL Speakers

4.4.1.1 Summary of findings

The findings for the BL speakers were rather straightforward, and they can be summarized as follows:

1. Speech rate was not different between English and Spanish for the BL speakers.
2. The proportion of the utterance that was vocalic (%V) in English was not different from that which was vocalic in Spanish.
3. As measured by ΔV, ΔC, VarcoV, VarcoC, nPVI-V, and rPVI-C, the BL speakers displayed evidence of language-specific rhythm patterns between their English and Spanish patterning in the direction expected for native speakers of each: Greater vocalic and consonantal variability in English than in Spanish.

4.4.1.2 Interpretation and implications

The finding that the speech rate was not different in English and Spanish for the BL speakers implies that, phonetically, the BL speakers have rather balanced proficiency and articulatory control in both of their languages. This finding also allows for a reliable interpretation of the results for the BL speakers as a baseline comparison to the LL speakers: The former displays evidence of balanced phonetic facility between their two languages.

In terms of %V, or the proportion of vowels in the utterance, prior research suggests that Spanish would have a higher value than English (Grabe & Low, 2002). Due to the fact that English has a more complex syllable structure, leading to greater consonant clusters, and phonological vowel reduction, which leads to a reduction in vowel quantity in unstressed syllables, whereas Spanish does not, the proportion of time occupied by vowels in a given Spanish utterance is expected to be higher than it would be in a given English utterance. However, the materials used for this study were controlled for several factors, such as the exclusion of various speech sounds, including rhotics, laterals, and glides. Similarly, consonantal (and vocalic) segments were removed at the beginning and end of utterances (see 4.2.4 Segmentation of Acoustic Data). Although these same controls were implemented in Spanish, the materials in Spanish were fundamentally different than those of English, and, as was reviewed in 4.4.2, English and Spanish differ in the way that consonants and vowels are affected in terms of duration. The removal of these consonants likely decreased the proportion of consonants and resulted in a higher %V ratio in English, approaching that of what was observed for Spanish in the same BL speakers.

The finding that all vocalic and consonantal rhythm metrics displayed systematically higher values in English than in Spanish for the BL speakers is not surprising. Prior research has shown that early bilinguals have a clear advantage over late learners of a language in terms of phonetic production (Baker & Trofimovich, 2005) and that they display patterns that are phonetically similar to monolinguals (Mack, 1989; Piske, Flege, MacKay, & Meador, 2002). In other words, phonetically, early bilinguals tend to display patterns that would be expected from monolinguals, at least in how they are perceived by others. On the one hand, English has a more complex syllable structure than Spanish, has phonological vowel reduction, and has a greater effect of stress-induced lengthening. On the other hand, Spanish has simpler syllable structure, no phonological vowel reduction, and smaller effects of stress-induced lengthening, and, consequently, vocalic and consonantal rhythm measures are expectedly higher in English than in Spanish. Furthermore, although some data did exist prior to this study on the rhythm patterns of early bilinguals who grew up speaking languages that differ as a matter of rhythm (see Carter, 2005; Carter & Wolford, 2016; Henriksen & Fafulas, 2017; Whitworth, 2002), such as English and Spanish, the present study adds a greater understanding of how prosodic phenomena are manifested in these populations. In short, the initial hypothesis that native English/Spanish bilinguals would display evidence of language-specific rhythm patterns in their two languages was confirmed.

4.4.2 LL Speakers

4.4.2.1 Summary of findings

Compared to the BL speakers, the findings for the LL speakers were more nuanced. They can be summarized as follows:

1. For the LL speakers, speech rate did not differ in English across conditions; however, in Spanish, speech rate increased significantly as linguistic experience increased.
2. Using raw rhythm metrics, evidence of L2 rhythm development was found. Furthermore, even after normalizing for speech rate, the LL speakers displayed robust evidence of having acquired target-like rhythm patterns in their L2 Spanish. Learners seemingly learned to reduce vocalic variability quickly relative to consonantal variability in Spanish vis-à-vis their L1 English.
3. Some evidence of bidirectional interference was observed in the consonantal rhythm patterns of the LL speakers' L1 English.

4.4.2.2 Interpretation and implications

Finding 1—that the LL speakers made gains in their speech rate in Spanish with increasing linguistic experience—supports previous research on this topic, which suggests that linguistic experience has an effect on L2 speech rate: Speakers of a lower proficiency display slower speech rates than those with higher proficiency in a given language. For instance, Lennon (1990) found that the speech rate of native speakers of German in their L2 English sped up after living in England for six months. Likewise, Towell, Hawkins, Bazergui (1996) found that the speech rate of native speakers of English in their L2 French increased after living in France for six months. Along the same lines, Guion et al. (2000) found that the length of residency in a foreign country, in addition to the age at which their participants began to live in the target country, were positively correlated with gains in speech rate in the target language. Trofimovich and Baker (2006) reported similar findings as those of Guion et al. The findings from the present study parallel the findings summarized here from prior research that an increase in linguistic experience yields an increase in speech rate in learners of a second language. While this finding is interesting, speech rate was measured in this study largely in order to control for the learning of L2 rhythm specifically.

The principal finding of this study—the evidence of the development of L2 rhythm—can be interpreted in at least two ways. On the one hand, the findings in this study that the LL speakers significantly increased their speech rate in Spanish with increasing linguistic experience can suggest that the pattern of development observed in the raw rhythm metrics may need to be interpreted cautiously. For both the vocalic and consonantal measures of standard deviation (ΔV and ΔC), taking speech rate into account (VarcoV and VarcoC) drastically increased the amount of variance that the regression model was able to account for. Along this line of interpretation, the apparently clear pattern of development observed in the ΔV and ΔC rhythm values was likely confounded with a development of increasing speech rate.[19] One the other hand, although the findings from Barry et al. (2003) suggest that it is relevant to rhythm in the L1, speech rate may not be random noise in the development of L2 rhythm. For instance, the development of L2 speech rate was statistically and visually linear: L2 speech rate increased along with an increase in linguistic experience. Furthermore, rhythmic variability, as measured by ΔV and ΔC, was also statistically and visually linear: It was greatest for the LL1 condition, less so for the LL2, and so on as linguistic experience

---

[19] A similar result was found with the pairwise vocalic and consonantal variability measures: Accounting for speech rate between rPVI-C and nPVI-C greatly increased the amount of variance that the regression model was able to predict and, although to a lesser degree than with the interval measures, the developmental effect was reduced, as well. Furthermore, these findings suggest that normalizing all rhythm metrics as a function of speech rate should be considered in research on L2 rhythm, including those measuring consonantal variability, despite the current trend in the literature to measure, at least for pairwise variability, only nPVI-V and rPVI-C (Henriksen & Fafulas, 2017; White & Mattys, 2007). The methodology of normalizing only vocalic pairwise variability in studies on rhythm appears to have begun with Grabe and Low (2002) who—citing Gay (1978), a study that had found that vowel gestures were more affected by changes in speech rate than were consonantal gestures—justified the calculation of only rPVI-C. The findings here indicate that consonantal rhythm is also affected by speech rate.

increased. On the other hand, as will be discussed in more detail further on, the analysis of the Varco rhythm metrics was only statistically linear, whereas, visually, the variability did not follow a logical progression: Variability was greatest for the LL1 condition, followed by a decrease in the LL3 condition, which was then followed by a further reduction in variability in the LL2 condition, followed lastly by the LL4 speakers. In other words, accounting for L2 speech rate, the development of which was robustly linear in the LL speakers, did not result in a visually linear development of rhythm in the normalized rhythm metrics. In short, although it can be said that the development of rhythm observed in the raw rhythm metrics was confounded with the development of L2 speech rate, the data are not entirely clear in this regard; therefore, these findings should not be dismissed as trivial.

Nevertheless, what is certain is that the learning of L2 vocalic and consonantal rhythm patterns globally and locally, even after normalizing for speech rate, did take place among the LL speakers. Apparently, the LL speakers quickly learned to reduce their vocalic variability toward target-like norms, beginning with the LL1 condition. That is, rapid gains in the acquisition of L2 rhythm were made, beginning at low-levels of linguistic experience, and this pattern stayed relatively static as linguistic experience increased. At the segmental level, acquiring the phonetic properties of L2 vowels has been shown to be challenging. Multiple studies have shown that L1 English speakers struggle to adapt the vowel *quality* of individual vowels and/or vowel contrasts of their L2 Spanish (Aldrich, 2014; Cobb & Simonet, 2015; Menke & Face, 2010). Fewer studies, however, have documented their acquisition of vowel quantity (Simonet, 2013). Nevertheless, the findings here appear to be in line with those of prior research on cue weighting in second language vowel learning. For instance, Flege, Bohn, and Jang (1997) found that L1 Korean speakers displayed a reliance on duration over spectral information in both their production and their perception of L2 English vowels. Although the participants in that study were native speakers of a syllable-rhythm language learning a stress-rhythm language, the reverse of those in the present, Bohn (1995, pp. 294–295), citing evidence from prior research in which L2 vowel acquisition could not be explained by cross-linguistic interference from the L1 alone, goes one step further and theorizes that a reliance on vowel duration in second language speech is universal, regardless of the linguistic background of the learner. While spectral information from the production data in this experiment was not analyzed, it is evident that the participants in this study, despite that their native language employs phonological vowel reduction, made gains toward target-like vocalic rhythm variability, and that this learning took place early in the development of the L2, a finding that could be due to, in part, the participants' reliance on durational cues in their learning of the vocalic properties of their L2 Spanish.

In terms of the acquisition of consonantal variability, although no interaction between Linguistic Experience (LL1, LL2, LL3, LL4, and BL) and Language (English, Spanish) was found for normalized consonantal rhythm metrics, patterns of learning and development are apparent in the data. Similar to the vowel rhythm data, gains toward target-like consonantal variability were made globally and locally at the early stages of linguistic experience. Unlike the vowel rhythm data, these gains were not made as abruptly across LL conditions. Rather, a trend of development was observed in both the VarcoC and the nPVI-C results. For both measures, the L1 condition displays greater consonantal variability than LL2, LL3, and LL4. The development pattern between LL1 and LL4 does not appear to be linear: for both VarcoC and nPVI-C, the LL3 condition displays slightly greater variability (more L1 like) than the L2 condition. In all cases, however, the first three LL conditions appear to pattern together, whereas the LL4 condition tends to display rhythm values intermediate between those of the BL speakers and those of the other LL conditions. The fact that the LL speakers in this study appeared to make fewer gains in consonantal rhythm vis-à-vis vocalic rhythm may seem, initially, to be counterintuitive. However, considering that the consonantal rhythm measures in this study were exactly that—measures of duration of sequences of consonants—prior research regarding the L2 acquisition of consonants may shed light on these results. It has been found in many studies that learners tend to transfer the acoustic correlates of stop voicing from their L1 to their L2 (Casillas, 2016; Flege & Port, 1981; Zampini, 1998), in both word-initial (Flege & Port, 1981; Gass, 1984) and word-final (Flege & Port, 1981) positions. The materials for the present study contained many syllable-initial, intervocalic voiceless stops, which, in English, are manifested as long-lag stops and, in Spanish, as short-lag stops, as reviewed in Chapter 3. Although VOT was not analyzed in the materials from the present experiment, the propensity that these learners have shown in the previous chapter toward transfer of the acoustic correlates of stop voicing, including VOT, from their L1 to their L2 may explain, in part, some of the L2 consonantal rhythm patterns observed here.

However, as Dauer (1983) and Lloyd James (1940) agree, rhythm is a matter of more than a simple combination of segmental phenomena alone. Rather, suprasegmental elements of speech, including stress-induced lengthening effects, play a role in what is perceived as language rhythm. Likewise, the consonantal variability patterns observed in this study may also be due to the transfer of prosody induced lengthening effects from the L1 to the L2. For instance, Turk and colleagues (Turk & Sawusch, 1997; Turk & White, 1999) found robust consonantal lengthening effects on syllables, and surrounding syllables, containing pitch accent in English: "consonants are longer when the unit they belong to is accented" (Turk & White, 1999, p. 178). On the other hand, in a similar study on the effect of accent-induced lengthening in Spanish, Ortega-Llebaria and Prieto (2007) found no pattern of lengthening on the syllable or unit containing the pitch accent. Thus, it may be that the LL speakers in this study are transferring non-salient acoustic correlates of pitch accent—specifically, accent-induced consonantal lengthening—from their L1 English to their L2 Spanish. Although the analysis of pitch accent in the data from this experiment goes beyond the scope of the present study, the laboratory nature of the materials, and, consequently, the fact that the data were extracted from the same materials for all speakers, this possibility of a consistent but steadily decreasing transfer of prosodic patterns from the L1 to the L2 is conceivable.

The findings from the present study differ from those of White and Mattys (2007), who, in their L1 English speakers, found that the rhythm patterns of the L2 Spanish data suggested that, of vocalic and consonantal rhythm, vocalic rhythm patterns were the more challenging of the two. As the present study measured rhythm in multiple samples of language learners of differing linguistic experience, a direct comparison to the results of the single sample of L1 English/L2 Spanish speakers in White and Mattys isn't feasible. What the present study has accomplished, however, is a panoramic view of the development of rhythm in a sample of this population, and it is evident that adult native speakers of English acquire language-specific phonetic detail of their L2 Spanish rhythm.

Lastly, finding 3—evidence of bidirectional interference in rhythm—has implications for theory revolving L2 speech. It has been documented previously that the phonetic properties of an L1 can be affected, or pulled, by those of the L2, at least in regard to the production of segments (Sancier & Fowler, 1997). In other words, cross-linguistic interference may affect the L1 and the L2 production of segments in both languages by attracting, in a sense, the acoustic characteristics of the segment. There has also been evidence of bidirectional interference in the production of prosodic phenomena. Mennen (2004) elicited speech from native speakers of Dutch learning Greek. She measured the peak alignment of tonal phones in the participants' L1 Dutch and L2 Greek and compared both to native speakers of each language. The results suggested that, not only did the Dutch speakers of Greek produce Greek tonal alignment with patterns seemingly being pulled by the L1 Dutch, but their L1 Dutch also displayed evidence of being pulled in the direction of their L2 Greek. The findings from the present study support those from Mennen: Apparently, even at the suprasegmental level, bidirectional interference is possible. Although no formal theory for L2 prosody acquisition exists, the fact that the L1 and L2 can influence each other at the prosodic level implies that the abstract representations for rhythm in the L1 and the L2 may possibly lie in the same phonological space, as Flege (1995) has proposed for L2 segmental acquisition. On the other hand, in the language of the L2LP (Escudero, 2009), it may be that the L1 and L2 perception grammars are being simultaneously activated for these particular correlates of rhythm.

To summarize, the LLs in this study showed patterns suggesting a developmental pattern of L2 rhythm. Furthermore, robust evidence was found that that they learned target-like rhythm patterns, both globally and locally. The results suggest that L2 Spanish vocalic duration variability is learned at the early stages of linguistic experience. Along the same lines, consonantal duration variability also displays learning early on, although less so than for vocalic rhythm, and that these patterns appear to display a pattern of development: As linguistic experience increases, consonantal duration variability decreases, although this effect is attenuated when controlling for speech rate. Importantly, the present study provides cross-sectional data across four learner conditions on the development of L2 rhythm and shows that the learning of L2 rhythm does happen. Considering the lack of a formal theory on the acquisition of prosodic phenomena in L2 speech, as mentioned in the Introduction to this chapter, one can only extrapolate so far as to the source of the patterns observed in the data.

Nevertheless, these findings have implications for the theory and future research of the field of second language speech as a whole. Firstly, L2 speech rate may not be random noise in the development of L2 rhythm. Secondly, bidirectional interference (a.k.a., simultaneous activation of L1 and L2 perception grammar) is possible in the acquisition of L2 rhythm, which implies that the phonological space between the L1 and L2 prosody may be shared (or each activated simultaneously). Thirdly, while linguistic experience plays a subtle role in the development of rhythm, the learning that takes place largely occurs at early stages of L2 exposure. Lastly, L2 consonantal rhythm patterns appear to be more susceptible to L1 interference than L2 vocalic rhythm patterns, denoting that the learning of L2 rhythm is sectional and that the underlying patterns—whether they be speech rate, transfer of vocalic rhythm, or transfer of consonantal rhythm—should be taken into account.

4.4.3 SUMMARY OF DISCUSSION

This study explored the rhythm patterns of early English/Spanish bilinguals and adult learners of Spanish as a second language. The main findings suggest that early bilinguals make clear distinctions between their two languages in terms of rhythm, in vocalic and consonantal duration variability, both globally and locally. The findings for the LL speakers are more nuanced. Firstly, as found in other research, speech rate in the L2 rose as linguistic experience increased. Speech rate also had an effect on the standard deviation (global variability) and pairwise (local variability) differences between L1 English and L2 Spanish. Nevertheless, controlling for speech rate, the learning of both vocalic and consonantal rhythm patterns does happen, and it happens early in the linguistic experience of the learner. However, these effects of learning are greater for vocalic sequences than for consonantal ones. The developmental pattern observed for consonantal rhythm may be due to L1 transfer of acoustic correlates of stop voicing and, or additionally, may also be due to L1 transfer of accent-induced consonantal lengthening. These findings have implications for second language speech theory.

4.5 CONCLUSION

The world's languages are said to differ typologically as a matter of rhythm, which, fundamentally, is a manifestation of a language's propensity toward vowel reduction, its syllable structure, and its phonetic realization of stress. Rhythm measures of vowel and consonant sequences, including the standard deviation, normalized standard deviation, proportion of vocalic segments, and absolute difference between consecutive intervals, has been shown to be a reliable way of operationalizing rhythm among languages. Furthermore, research has shown that native speakers that are learning a language that is typologically different from their native language in terms of rhythm are able to acquire aspects of the duration variability patterns of the L2. These gains have been shown to be made as a function of linguistic experience, although the direction of the development apparently varies from language to language. The present study researches the rhythm patterns of adult, native speakers of English learning Spanish as a second language and sheds light on the patterns of rhythm development across four learner conditions, each with increasing linguistic experience in the L2. To achieve a base measure of understanding of these patterns in adults that speak both languages from an early age, the rhythm of early English/Spanish bilinguals was also investigated.

The main findings from this experiment suggest that early bilinguals display duration variability patterning after what would be expected in either language and that adults learning Spanish as a second language do differentiate their L1 rhythm from that of their L2. On the one hand, early bilinguals displayed greater vocalic and consonantal variability in English than in Spanish while maintaining a similar speech rate between the two languages. On the other hand, late learners of Spanish—LL1, LL2, LL3, and LL4—while maintaining similar speech rates in English, showed gains in speech rate as linguistic experience increased. These changes in speech rate in the L2 across conditions appeared to have an effect on rhythm. Measures not controlled for speech rate, such as the ΔV and ΔC metrics, which calculate the standard deviation of vocalic and consonantal intervals, respectively, displayed patterns of development toward target-like rhythm. Once these measures were controlled for speech rate (VarcoV and VarcoC), however, while initial gains were made, the effect of development was attenuated. A similar pattern emerged among the LL conditions for the measures of pairwise rhythm variability—nPVI-V, rPVI-C, and nPVI-C. These findings have implications for second language speech acquisition: Late learners of a second language that is rhythmically different from the first are capable of making phonetic distinctions in speech production between L1 and L2 rhythm, even when controlling for speech rate.

CHAPTER 5

CONCLUSION

This dissertation has presented two experiments analyzing the effect of linguistic experience on the acquisition of segmental and suprasegmental speech phenomena. The speech of native bilinguals of English and Spanish was also analyzed as a baseline comparison.

The first experiment dealt with the development of gestural targets in speakers of a native language, English, whose voiceless coronal stop has an alveolar place of articulation, that were learning Spanish, whose voiceless coronal stop has a dental place of articulation. The native bilinguals displayed robust evidence of having language-specific gestural targets between their English and Spanish coronal stops. The principal finding of the experiment, however, were the results from the language learner data: The native speakers of English learning Spanish also exhibited robust evidence of a unique gestural target for the coronal voiceless stops of English and Spanish. Importantly, these learners made significant gains toward target-like gestural targets for the Spanish /t/ as their linguistic experience with L2 Spanish increased. This dissertation presents the first published data on the acquisition of stop consonants that are phonologically similar but differ in regard to gestural target. In the language of the L2LP, these learners, who faced a Similar Scenario of language acquisition in which the phonological stop inventory of their L2 was similar but phonetically distinct from that of their L1, appeared to learn to update the phonetic categories of place of articulation of their L2 perception grammar toward target-like norms.

The second experiment dealt with the development of rhythm in a second language, Spanish, that is typologically different in regard to rhythm from the L1, English. The native bilinguals displayed robust evidence of language-specific rhythm patterns between their English and their Spanish, in regard to both vowels and consonants. The findings for the language learners were more nuanced, although they also displayed evidence of having made gains toward target-like rhythm. Firstly, language learners showed gains toward faster speech rate, approaching the norm of their L1 English and that of the Spanish of the native bilingual speakers, with increasing linguistic experience. More importantly, on the one hand, the second language speakers learned to significantly reduce their vocalic variability at early stages in their linguistic development, such that patterns of development with increasing linguistic experience were not statistically present. On the other hand, although significant patterns of development were also not revealed by the analyses, the second language speakers learned to reduce their consonantal variability early on in their linguistic experience; however, the effect was seemingly less robust than it was for the learning of L2 vocalic rhythm. Although there is no formal theory of L2 prosodic learning, the results from this can be related to theories of segmental acquisition in second language speech. For instance, in the words of the L2LP, it can be said that the second language learners in this study updated the phonetic categories of the perception grammar of their L2 rhythm, and this process occurred rather quickly in the trajectory of their language development.

Although increasingly target-like L2 gestural targets and L2 rhythm were observed in conjunction with an increase in linguistic experience, it should be pointed out that the exact nature of the L2 experience obtained among the various levels of speakers is not entirely clear. Although it is certain that an increase in exposure and formal learning to the L2 Spanish led to gains toward target-like production, other individual differences not measured in this study could have also contributed. For instance, its plausible that some of these speakers could have received some form of explicit feedback in regard to their production of their Spanish /t/ or their Spanish rhythm. Likewise, individual differences, such as motivation and attitudes toward the target language and culture, could have had an effect on these outcomes. These factors fall outside of the principal questions of this study, and, therefore, they were not analyzed here; however, future research should consider including them to further understand the nature of gestural development in the second language.

Taken together, the results from these two experiments suggest that adult learners of a second language make gains toward target-like gestural targets and rhythm as their experience with the language increases.

APPENDIX A

INCLUSION/EXCLUSION CRITERIA

1. Are you at least 18 years old?
2. Since you were a child, were any other languages spoken in your home besides English?
3. Select any languages (besides English) that were spoken in your home since you were a child. If the language isn't listed, please select "Other" and type the name of the language.
   a. Spanish
   b. Other
4. Are you currently enrolled in a Spanish class at the University of Arizona?
5. Select the Spanish class that you are currently enrolled in (Spring 2018). If you are enrolled in any 300-level Spanish class, select "A Spanish 300-level class." If you are enrolled in any 400-level Spanish class, select "A 400-level Spanish class." If the class you are currently enrolled in is not on the list and you are not enrolled in a 300- or 400-level Spanish class, select, "My class isn't listed."
   a. Spanish 102
   b. Spanish 202
   c. A Spanish 300-level class
   d. A Spanish 400-level lass
   e. Spanish 103 – Oral skills for Heritage Learners of Spanish
   f. I am not currently enrolled in a Spanish class.
6. Do you consider yourself a fluent speaker of any language besides English or Spanish?
7. Do you have a speech or hearing impairment or a vision impairment not easily corrected with corrective lenses?

APPENDIX B

BILINGUAL LANGUAGE PROFILE

Bilingual Language Profile: English-Spanish

We would like to ask you to help us by answering the following questions concerning your language history, use, attitudes, and proficiency. This survey was created with support from the Center for Open Educational Resources and Language Learning at the University of Texas at Austin to better understand the profiles of bilingual speakers in diverse settings with diverse backgrounds. The survey consists of 19 questions and will take less than 10 minutes to complete. This is not a test, so there are no right or wrong answers. Please answer every question and give your answers sincerely. Thank you very much for your help.

I. Biographical Information

Name ______________________________	Today's Date ____/____/______

Age____	☐Male / ☐Female	Current place of residence: city/state_______________ country____________

Highest level of formal education:	☐ Less than high school	☐ High school	☐ Some college
	☐ College (B.A., B.S.)	☐ Some graduate school	☐ Masters
	☐ PhD/MD/JD	☐ Other: ____________	

Please cite as :

Birdsong, D., Gertken, L.M., & Amengual, M. *Bilingual Language Profile: An Easy-to-Use Instrument to Assess Bilingualism*. COERLL, University of Texas at Austin. Web. 20 Jan. 2012. <https://sites.la.utexas.edu/bilingual/>.

II. Language history

In this section, we would like you to answer some factual questions about your language history by placing a check in the appropriate box.

1. At what age did you **start learning** the following languages?

English

☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Since birth	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20+

Spanish

☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Since birth	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20+

2. At what age did you **start to feel comfortable** using the following languages?

English

☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
As early as I can remember	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20+	not yet

Spanish

☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
As early as I can remember	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20+	not yet

3. How many years of **classes (grammar, history, math, etc.)** have you had in the following languages (primary school through university)?

English

☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20+

Spanish

☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20+

4. How many years have you spent in a **country/region** where the following languages are spoken?

English

☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20+

Spanish

☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20+

5. How many years have you spent in a **family** where the following languages are spoken?

English

☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20+

Spanish

☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20+

6. How many years have you spent in a **work environment** where the following languages are spoken?

English

☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20+

Spanish

☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20+

III. Language use

In this section, we would like you to answer some questions about your language use by placing a check in the appropriate box. Total use for all languages in a given question should equal 100%.

7. In an average week, what percentage of the time do you use the following languages **with friends**?

English	☐ 0%	☐ 10%	☐ 20%	☐ 30%	☐ 40%	☐ 50%	☐ 60%	☐ 70%	☐ 80%	☐ 90%	☐ 100%
Spanish	☐ 0%	☐ 10%	☐ 20%	☐ 30%	☐ 40%	☐ 50%	☐ 60%	☐ 70%	☐ 80%	☐ 90%	☐ 100%
Other languages	☐ 0%	☐ 10%	☐ 20%	☐ 30%	☐ 40%	☐ 50%	☐ 60%	☐ 70%	☐ 80%	☐ 90%	☐ 100%

8. In an average week, what percentage of the time do you use the following languages **with family**?

English	☐ 0%	☐ 10%	☐ 20%	☐ 30%	☐ 40%	☐ 50%	☐ 60%	☐ 70%	☐ 80%	☐ 90%	☐ 100%
Spanish	☐ 0%	☐ 10%	☐ 20%	☐ 30%	☐ 40%	☐ 50%	☐ 60%	☐ 70%	☐ 80%	☐ 90%	☐ 100%
Other languages	☐ 0%	☐ 10%	☐ 20%	☐ 30%	☐ 40%	☐ 50%	☐ 60%	☐ 70%	☐ 80%	☐ 90%	☐ 100%

9. In an average week, what percentage of the time do you use the following languages **at school/work**?

English	☐ 0%	☐ 10%	☐ 20%	☐ 30%	☐ 40%	☐ 50%	☐ 60%	☐ 70%	☐ 80%	☐ 90%	☐ 100%
Spanish	☐ 0%	☐ 10%	☐ 20%	☐ 30%	☐ 40%	☐ 50%	☐ 60%	☐ 70%	☐ 80%	☐ 90%	☐ 100%
Other languages	☐ 0%	☐ 10%	☐ 20%	☐ 30%	☐ 40%	☐ 50%	☐ 60%	☐ 70%	☐ 80%	☐ 90%	☐ 100%

10. When you talk to yourself, how often do you **talk to yourself** in the following languages?

English	☐ 0%	☐ 10%	☐ 20%	☐ 30%	☐ 40%	☐ 50%	☐ 60%	☐ 70%	☐ 80%	☐ 90%	☐ 100%
Spanish	☐ 0%	☐ 10%	☐ 20%	☐ 30%	☐ 40%	☐ 50%	☐ 60%	☐ 70%	☐ 80%	☐ 90%	☐ 100%
Other languages	☐ 0%	☐ 10%	☐ 20%	☐ 30%	☐ 40%	☐ 50%	☐ 60%	☐ 70%	☐ 80%	☐ 90%	☐ 100%

11. When you count, how often do you **count** in the following languages?

English	☐ 0%	☐ 10%	☐ 20%	☐ 30%	☐ 40%	☐ 50%	☐ 60%	☐ 70%	☐ 80%	☐ 90%	☐ 100%
Spanish	☐ 0%	☐ 10%	☐ 20%	☐ 30%	☐ 40%	☐ 50%	☐ 60%	☐ 70%	☐ 80%	☐ 90%	☐ 100%
Other languages	☐ 0%	☐ 10%	☐ 20%	☐ 30%	☐ 40%	☐ 50%	☐ 60%	☐ 70%	☐ 80%	☐ 90%	☐ 100%

IV. Language proficiency
In this section, we would like you to rate your language proficiency by giving marks from 0 to 6.

	0=not well at all						*6=very well*
12. a. How well do you speak **English**?	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
b. How well do you speak **Spanish**?	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
13. a. How well do you understand **English**?	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
b. How well do you understand **Spanish**?	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
14. a. How well do you read **English**?	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
b. How well do you read **Spanish**?	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
15. a. How well do you write **English**?	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
b. How well do you write **Spanish**?	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6

V. Language attitudes
In this section, we would like you to respond to statements about language attitudes by giving marks from 0-6.

	0=disagree						*6=agree*
16. a. I feel like myself when I speak **English**.	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
b. I feel like myself when I speak **Spanish**.	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
17. a. I identify with an **English-speaking** culture.	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
b. I identify with a **Spanish-speaking** culture.	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
18. a. It is important to me to use (or eventually use) **English** like a native speaker.	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
b. It is important to me to use (or eventually use) **Spanish** like a native speaker.	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
19. a. I want others to think I am a native speaker of **English**.	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
b. I want others to think I am a native speaker of **Spanish**.	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6

APPENDIX C

BACKGROUND QUESTIONNAIRE

1. What is your full name?
2. Where were you born? (City, State, Country)
3. Where were you raised? (If the same as above, please respond “Same as above.”)
4. Are you currently enrolled in a Spanish class at the University of Arizona?
5. If you answered 'Yes' to the previous question, in which Spanish class are you currently enrolled? (If you answered 'No' to the previous question, please enter 'n/a'.)
6. How many years of classes in which Spanish was taught as a foreign language (i.e. classes whose objective it was to teach how to speak, read, write, and/or comprehend Spanish as a second or foreign language) have you taken, elementary school through university (in years)?
7. Have you ever taken foreign language classes for any language other than Spanish? If yes, please respond with the name(s) of the language(s). If no, please respond “No.”
8. If you currently take or have taken a foreign language class(es) for any language other than Spanish, how many years of these classes (i.e. classes whose objective it was to learn how to speak, read, write, and/or comprehend the other foreign language(s)) have you taken from elementary school through university (in years)? (If you responded ‘No’ to the previous question, enter 0.)
9. Have you ever lived in a Spanish-speaking country?
10. If you answered ‘Yes’ to the question above, how long did you live in a Spanish-speaking country? (If you answered 'No' on the previous question, select 0.)
11. Growing up, did either of your parents, a guardian, or a close family member speak to you habitually in a language other than English?

12. If you answered ‘Yes’ to the previous question, which language were you spoken in habitually growing up? (If you answered ‘No’ to the previous question, please enter ‘n/a’.)
13. Have you ever taken a Spanish phonetics class (e.g. Span 340, Span 343) (i.e. classes whose objective it was to improve your pronunciation of Spanish sounds)?
14. If you answered ‘Yes’ to the previous question, when did you take the Spanish phonetics course (Semester, Year)? (If you responded ‘No’ to the previous question, please enter ‘n/a’.)
15. Have you ever taken an Introduction to Hispanic Linguistics class (e.g. Span 452) or a Hispanic Linguistics class (e.g. Span 459)?
16. If you answered ‘Yes’ to the previous question, when did you take the Intro to Hispanic Linguistics or Hispanic Linguistics course (Semester, Year)? (If you responded ‘No’ to the previous question, please enter ‘n/a’.)

APPENDIX D

LEXTALE

Task Instructions (1/2)

This first task is a test that consists of about 60 trials, in each of which you will see a string of letters. Your task is to decide whether this is an existing English word or not. If you think it is an existing English word, press the left arrow on the keyboard for "yes", and if you think it is not an existing English word, press the right arrow on the keyboard for "no".

If you are sure that the word exists, even though you don't know its exact meaning, you may still respond "yes". But if you are not sure if it is an existing word, you should respond "no".

Press any key to continue.

Task Instructions (2/2)

In this experiment, we use British English rather than American English spelling. For example: "realise" instead of "realize"; "colour" instead of "color", and so on. Please don't let this confuse you. This experiment is not about detecting such subtle spelling differences anyway.

You have as much time as you like for each decision. This part of the experiment will take about 5 minutes.

If everything is clear, you can now start the experiment by pressing any key.

LexTALE Materials:

platery (not scored)	30. skave
denial (not scored)	31. plaintively
generic (not scored)	32. kilp
1. mensible	33. interfate
2. scornful	34. hasty
3. stoutly	35. lengthy
4. ablaze	36. fray
5. kermshaw	37. crumper
6. moonlit	38. upkeep
7. lofty	39. majestic
8. hurricane	40. magrity
9. flaw	41. nourishment
10. alberation	42. abergy
11. unkempt	43. proom
12. breeding	44. turmoil
13. festivity	45. carbohydrate
14. screech	46. scholar
15. savoury	47. turtle
16. plaudate	48. fellick
17. shin	49. destription
18. fluid	50. cylinder
19. spaunch	51. censorship
20. allied	52. celestial
21. slain	53. rascal
22. recipient	54. purrage

23. exprate	55. pulsh
24. eloquence	56. muddy
25. cleanliness	57. quirty
26. dispatch	58. pudour
27. rebondicate	59. listless
28. ingenious	60. wrought
29. bewitch	

LexTALE-Esp

Task Instructions (1/2)

This next task is a test that consists of about 90 trials, in each of which you will see a string of letters. Your task is to decide whether this is an existing Spanish word or not.

If you think it is an existing Spanish word, press the left arrow on the keyboard for "yes", and if you think it is not an existing Spanish word, press the right arrow on the keyboard for "no".

Press any key to continue.

Task instructions (2/2)

If you are sure that the word exists, even though you don't know its exact meaning, you may still respond "yes". But if you are not sure if it is an existing word, you should respond "no".

You have as much time as you like for each decision. This part of the experiment will take about 5 minutes. If everything is clear, you can now start the experiment. Press any key to continue.

LexTALE-Esp Materials:

1. terzo
2. pellizcar
3. pulmones
4. batillón
5. zapato
6. tergiversar
7. pésimo
8. cadeña
9. hacha
10. antar
11. cenefa
12. asesinato
13. helar
14. yunque
15. regar
16. abracer
17. floroso
18. arsa
19. brecedad
20. ávido
21. capillo
22. lacayo
23. lampera
24. látigo
25. bisagra

46. martillo
47. cartinar
48. ladrón
49. ganar
50. flamida
51. candado
52. camisa
53. vegada
54. fomentar
55. nevar
56. musgo
57. tacaño
58. plaudir
59. besar
60. matar
61. seda
62. flaco
63. esposante
64. orgulloso
65. bizcocho
66. hacido
67. cabello
68. alegre
69. engatusar
70. temblo

26. secuestro	71. polvoriento
27. acutación	72. pemición
28. merodear	73. hervidor
29. decar	74. cintro
30. alardio	75. yacer
31. pandilla	76. atar
32. fatacidad	77. tiburón
33. pauca	78. frondoso
34. aviso	79. tropaje
35. rompido	80. hormiga
36. loro	81. pozo
37. granuja	82. empirador
38. estornudar	83. guante
39. torpe	84. escuto
40. alfombra	85. laúd
41. rebuscar	86. barato
42. cadallo	87. grodo
43. canela	88. acantilado
44. cuchara	89. prisa
45. jilguero	90. clavel

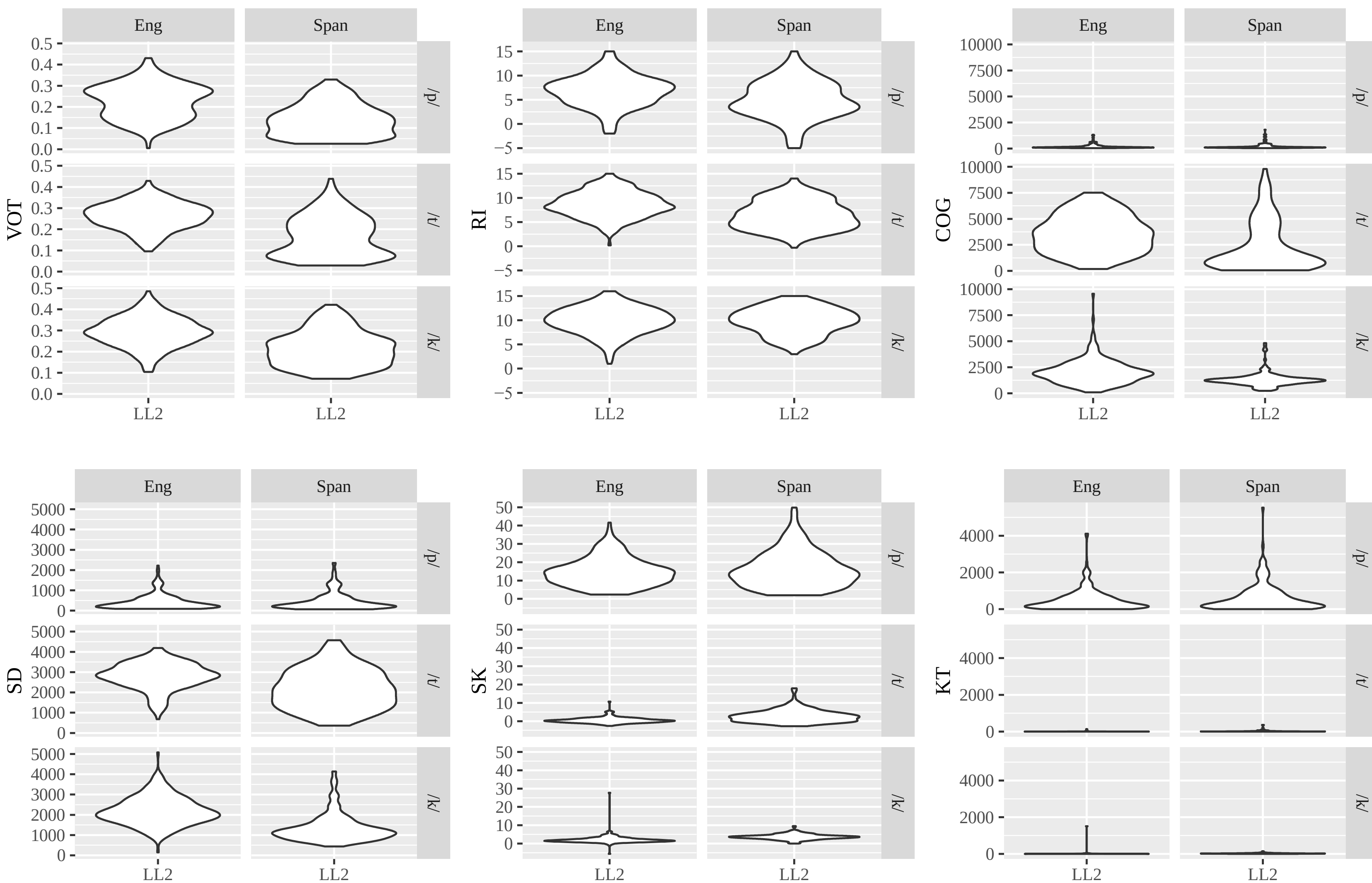
APPENDIX E

ABBREVIATIONS AND ACRONYMS

1. BL: Native English/Spanish bilingual
2. LL: second language learner
3. BLP: Bilingual Language Profile
4. SCL: Spanish class level
5. SRSP: Self-reported Spanish speaking proficiency
6. LE: Linguistic experience
7. LL1: Lowest linguistic experience
8. LL2: Second lowest LE
9. LL3: Second highest LE
10. LL4: Highest LE
11. AoA: Age of acquisition
12. ANOVA: Analysis of Variance
13. M: mean
14. SD: standard deviation
15. PAM: Perceptual Assimilation Model
16. SLM: Speech Learning Model
17. L2LP: Second Language Linguistic Perception Model
18. SP: Spectral peak
19. RI: Relative intensity
20. COG: Center of gravity
21. SK: Skewness
22. KT: Kurtosis
23. F1: First formant
24. F2: Second formant
25. F2T: F2 transition

APPENDIX F

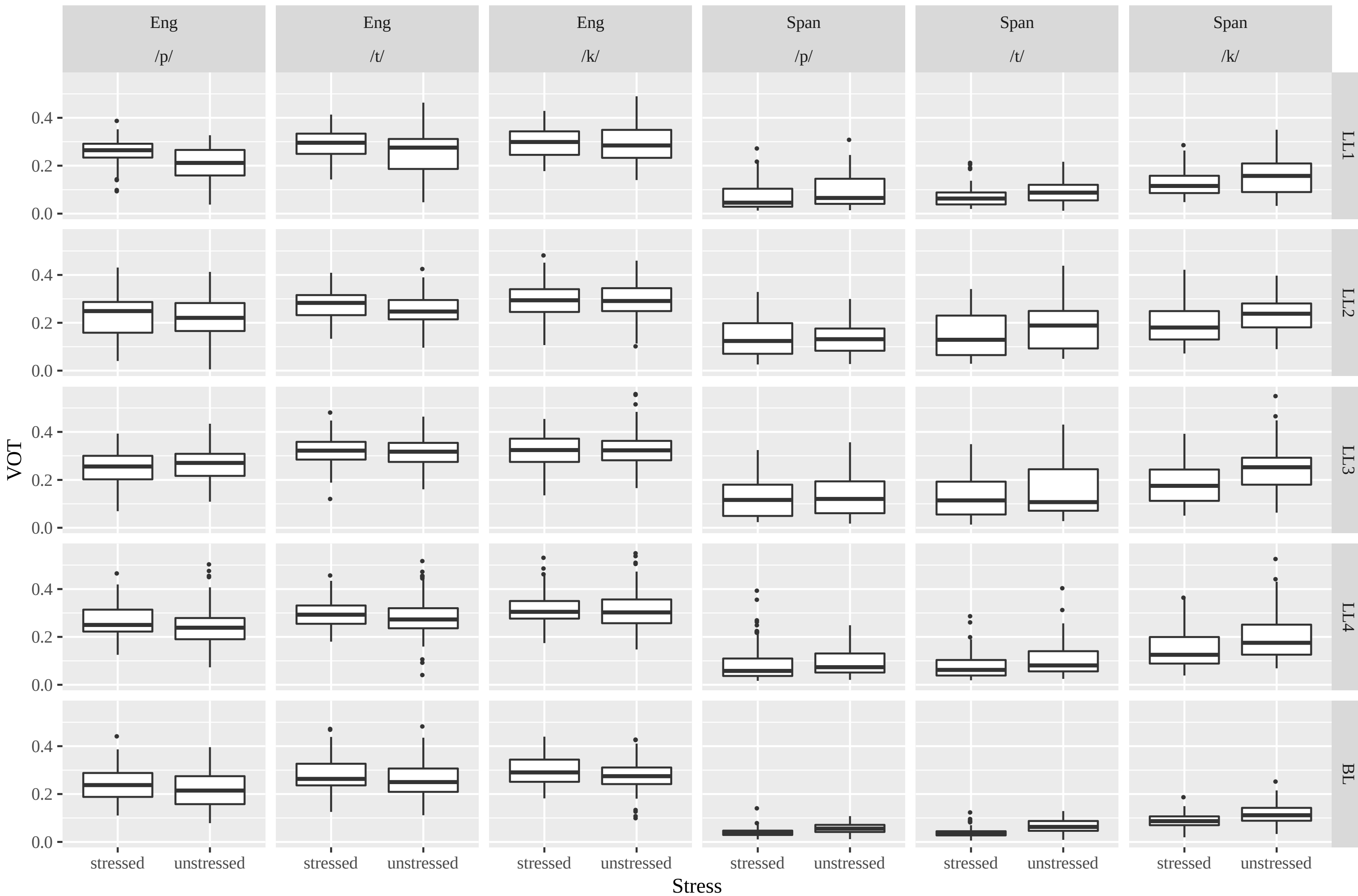
VIOLIN PLOT OF LL2 DATA



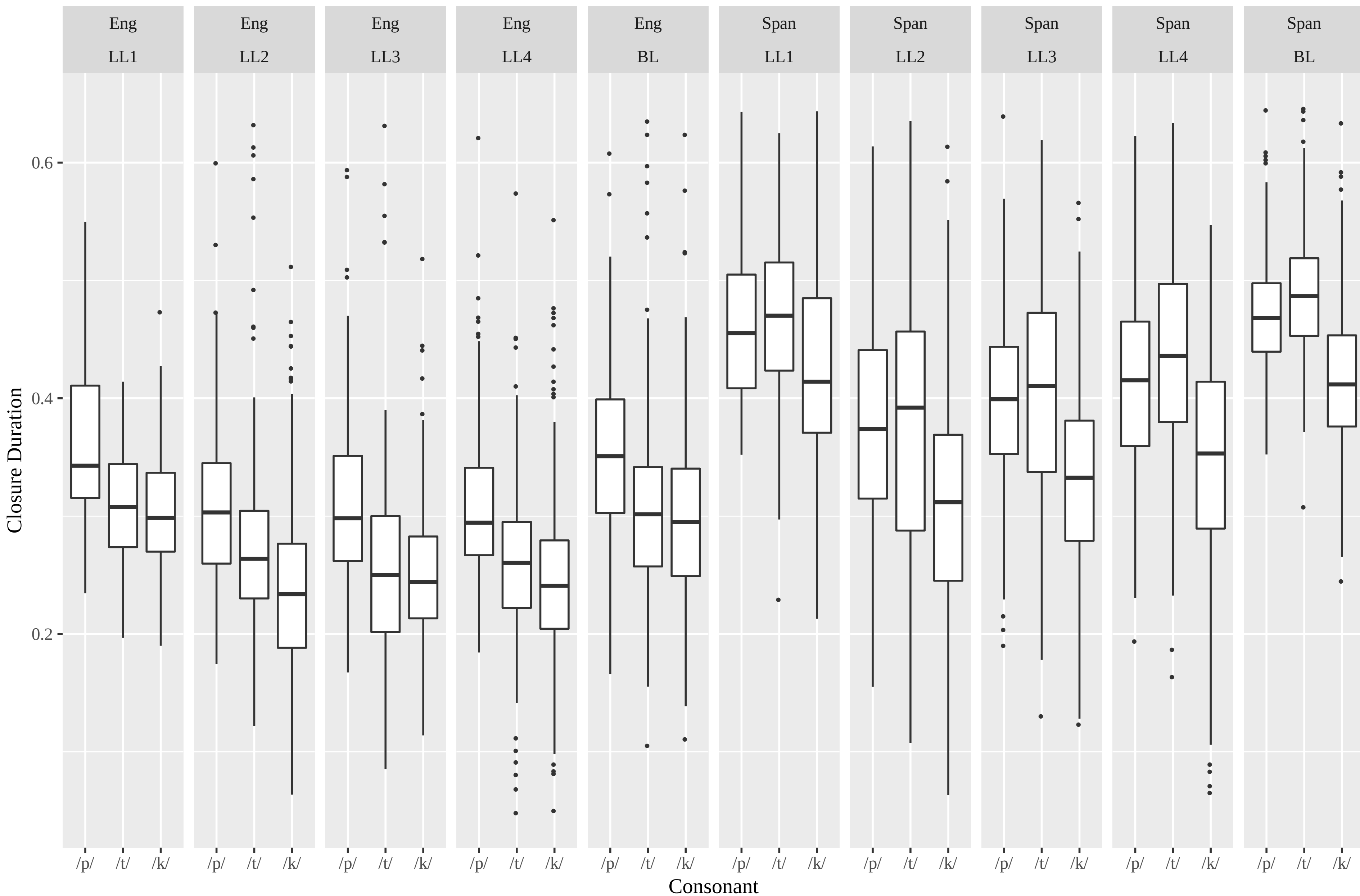
Above: Violin plot of VOT, Relative Intensity (RI), Center of Gravity (COG), Standard Deviation (SD), Skewness, (SK), and Kurtosis (KT) for the LL2 condition.

APPENDIX G

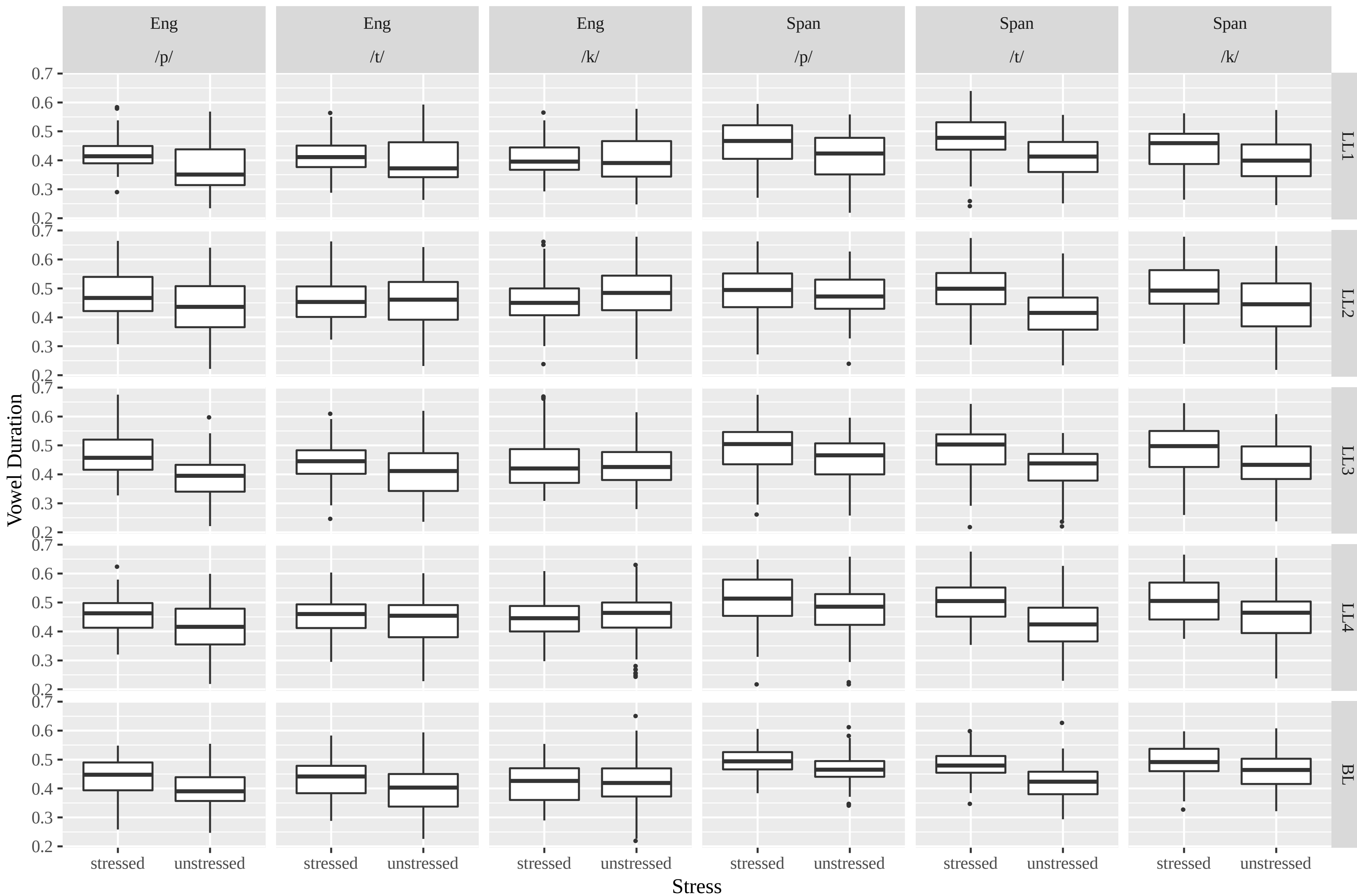
ADDITIONAL FIGURES



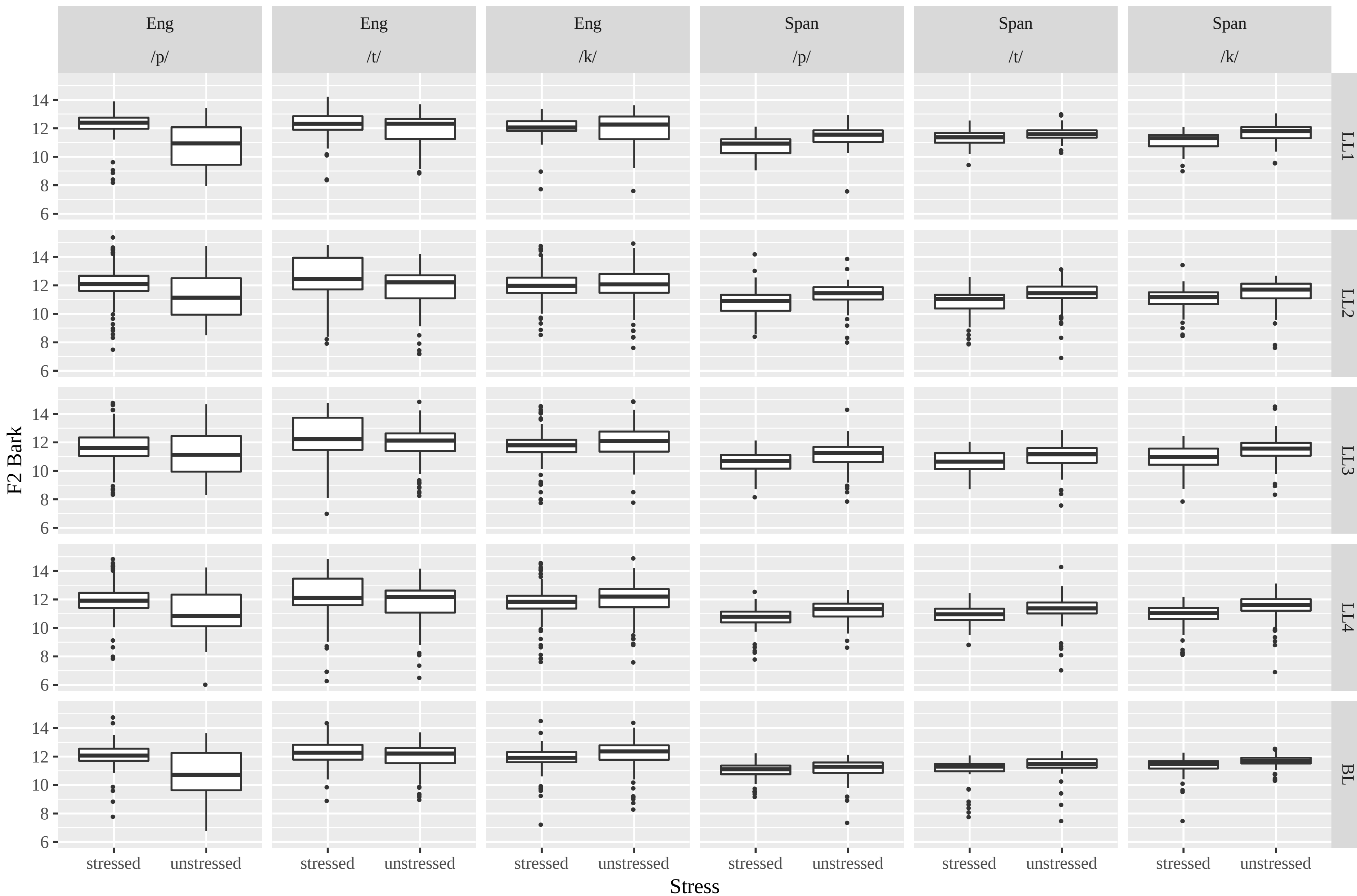
Above: Boxplot of VOT as a function of Linguistic Experience (LL1, LL2, LL3, LL4, BL), Language (English, Spanish), Consonant (/p, t, k/), and Stress (Stressed, Unstressed).



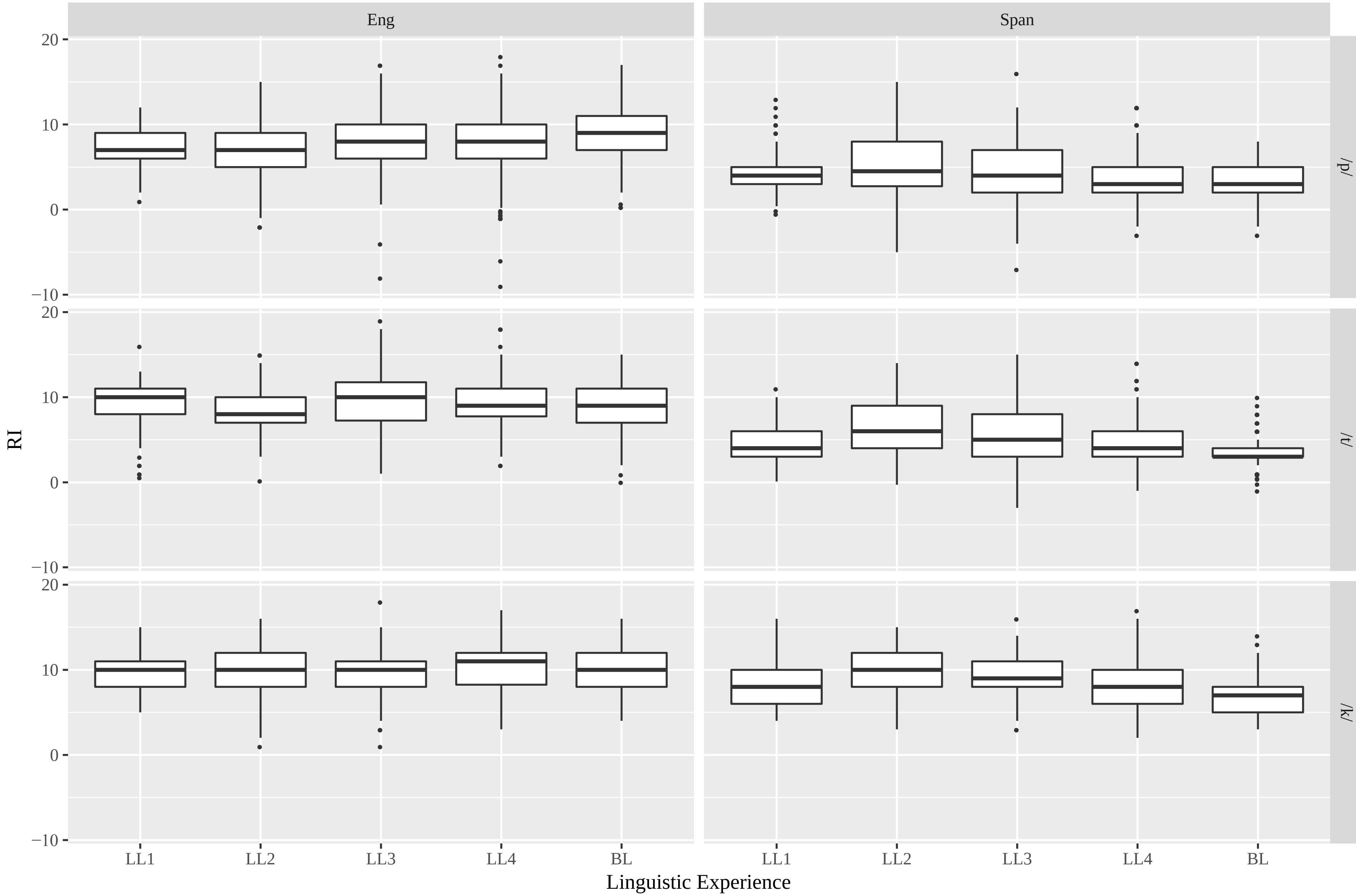
Above: Boxplot of Closure Duration as a function of Linguistic Experience (LL1, LL2, LL3, LL4, BL), Language (English, Spanish), and Consonant (/p, t, k/).



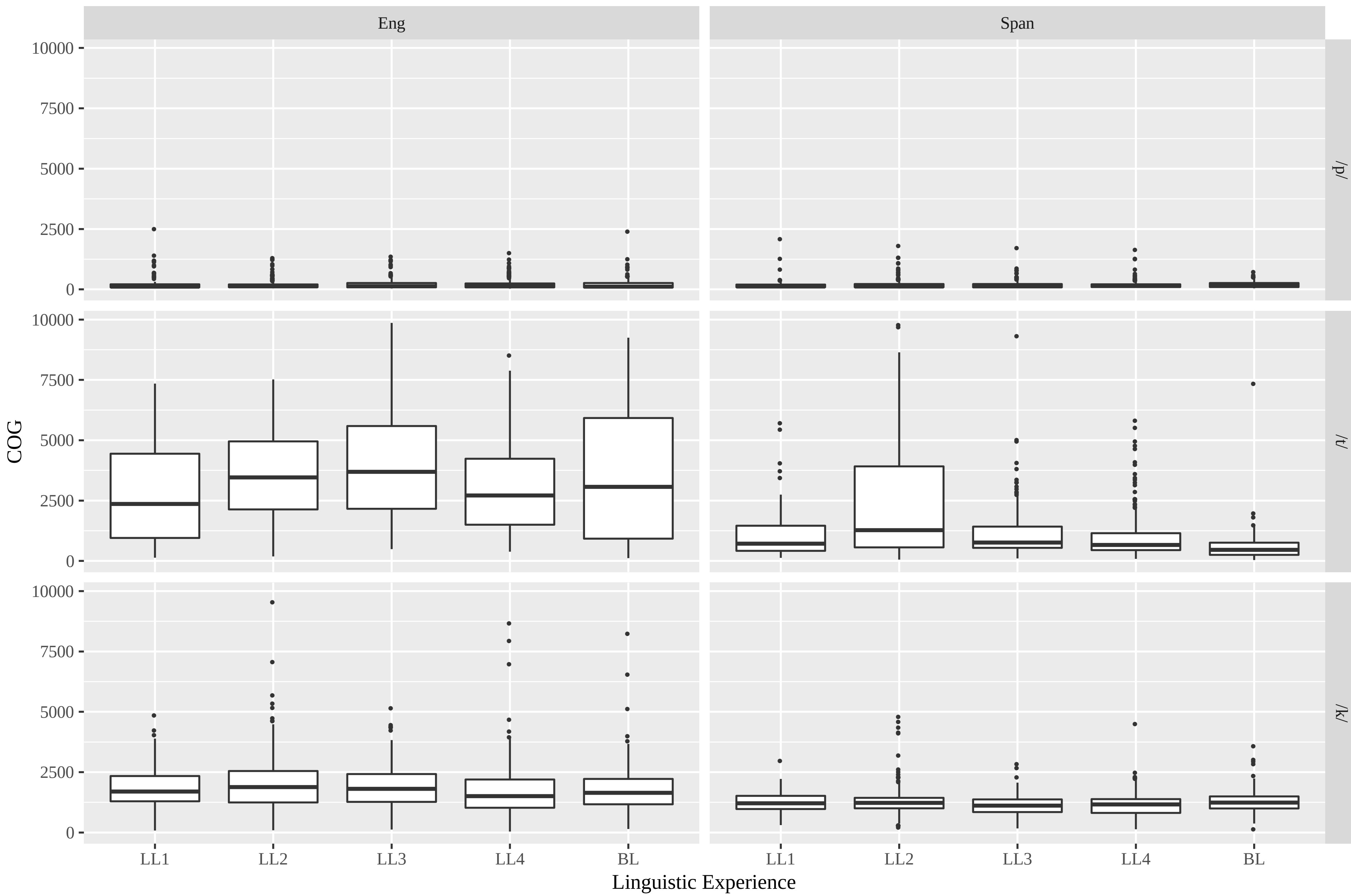
Above: Boxplot of Vowel Duration as a function of Linguistic Experience (LL1, LL2, LL3, LL4, BL), Language (English, Spanish), Consonant (/p, t, k/), and Stress (Stressed, Unstressed).



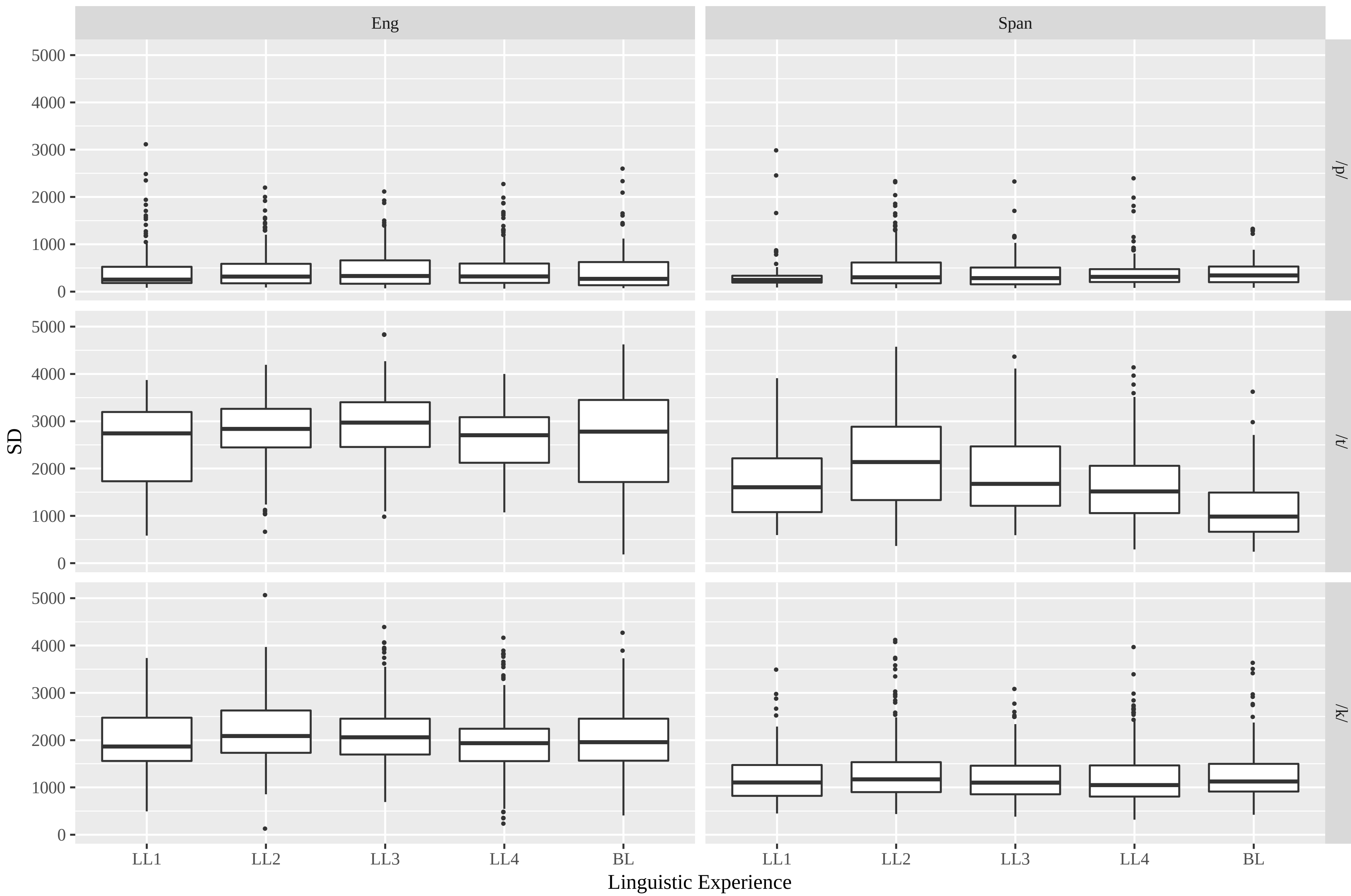
Above: Boxplot of F2 Bark as a function of Linguistic Experience (LL1, LL2, LL3, LL4, BL), Language (English, Spanish), Consonant (/p, t, k/), and Stress (Stressed, Unstressed).



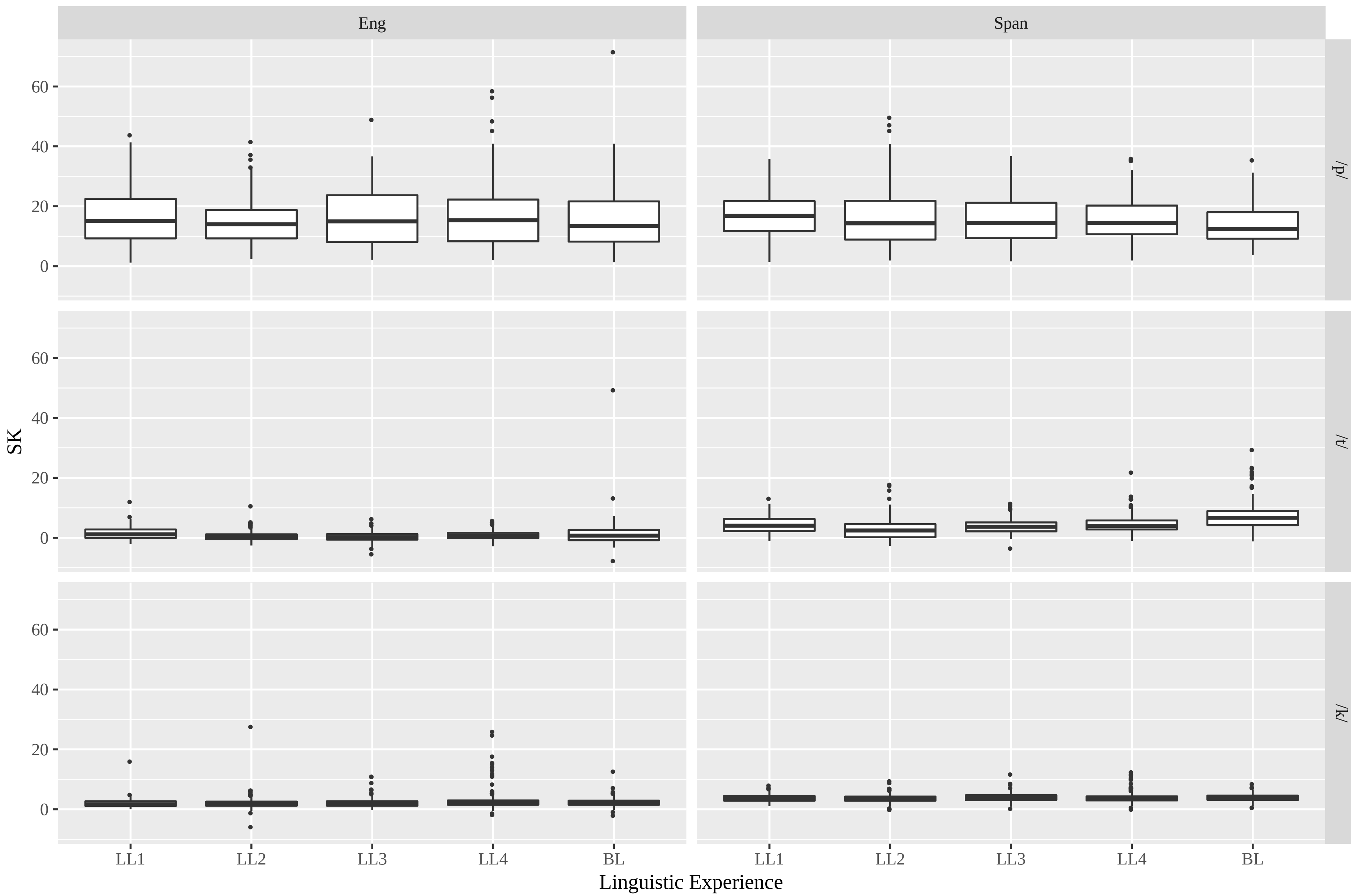
Above: Boxplot of Relative Intensity (RI) as a function of Linguistic Experience (LL1, LL2, LL3, LL4, BL), Language (English, Spanish), and Consonant (/p, t, k/).



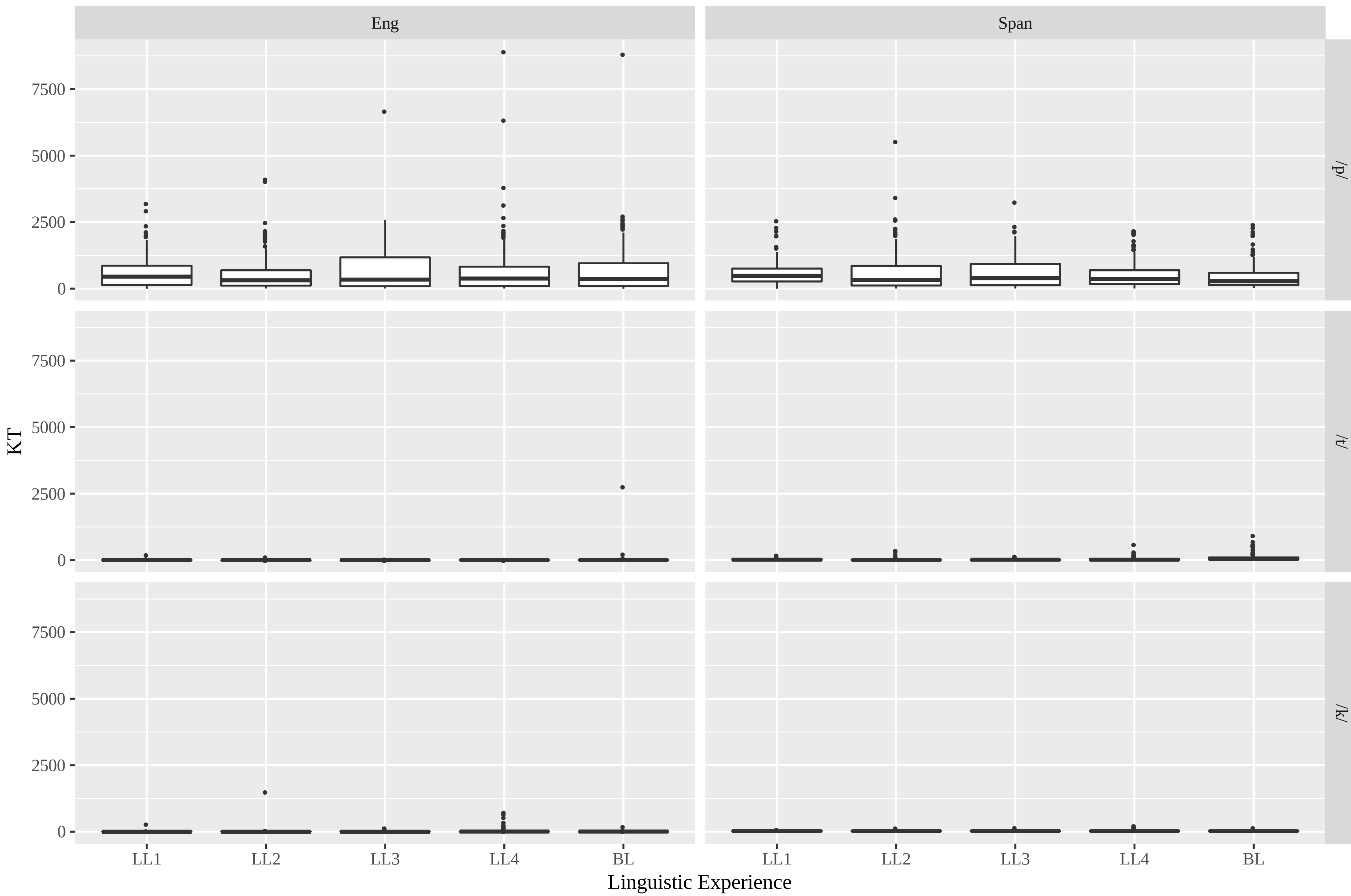
Above: Boxplot of Center of Gravity (COG) as a function of Linguistic Experience (LL1, LL2, LL3, LL4, BL), Language (English, Spanish), and Consonant (/p, t, k/).



Above: Boxplot of Standard Deviation (SD) as a function of Linguistic Experience (LL1, LL2, LL3, LL4, BL), Language (English, Spanish), and Consonant (/p, t, k/).



Above: Boxplot of Skewness (SK) as a function of Linguistic Experience (LL1, LL2, LL3, LL4, BL), Language (English, Spanish), and Consonant (/p, t, k/).



Above: Boxplot of Kurtosis (KT) as a function of Linguistic Experience (LL1, LL2, LL3, LL4, BL), Language (English, Spanish), and Consonant (/p, t, k/).

REFERENCES

Abercrombie, D. (1967). *Elements of General Phonetics*. Chicago: Aldine Publishing Company.

Abramson, A. S., & Lisker, L. (1973). Voice-timing perception in Spanish word-initial stops. *Journal of Phonetics*, *1*, 1–8.

Abramson, A. S., & Whalen, D. H. (2017). Voice Onset Time (VOT) at 50: Theoretical and practical issues in measuring voicing distinctions. *Journal of Phonetics*, *63*, 75–86. https://doi.org/10.1016/j.wocn.2017.05.002

Aldrich, A. C. (2014). *Acquisition of L2 Phonology: An Acoustic Analysis of the Centralization of L2 Spanish /a/ in Adult L1 English-Speaking Learners*. (Master's thesis). Brigham Young University, Provo, UT.

Aldrich, A. C., & Simonet, M. (n.d.). Duration of syllable nuclei in Spanish. *Studies in Hispanic and Lusophone Linguistics*.

Ali, A. M. A. (1999). *Auditory-based acoustic-phonetic signal processing for robust continuous speech recognition*. (Doctoral dissertation). University of Pennsylvania, Philadelphia, PA.

Ali, A. M. A., Van der Spiegel, J., & Mueller, P. (2001). Acoustic-phonetic features for the automatic classification of stop consonants. *IEEE Transactions on Speech and Audio Processing*, *9*(8), 833–841. https://doi.org/10.1109/89.966086

Álvarez González, J. A. (1978). El análisis por síntesis. *Revista Española de Lingüística*, *8*(1), 117–124.

Arvaniti, A. (2009). Rhythm, timing and the timing of rhythm. *Phonetica*, *66*(1–2), 46–63. https://doi.org/10.1159/000208930

Aslin, R. N., Pisoni, D. B., Hennessy, B. L., & Perey, A. J. (1981). Discrimination of Voice

Onset Time by Human Infants: New Findings and Implications for the Effects of Early Experience Richard. *Child Development*, *52*(4), 1135–1145.

Baker, W., & Trofimovich, P. (2005). Interaction of Native- and Second-Language Vowel System (s) in Early and Late Bilinguals. *Language and Speech*, *48*(1), 1–27. https://doi.org/10.1177/00238309050480010101

Banov, I. K. (2014). *The Production of Voice Onset Time in Voiceless Stops by Spanish-English Natural Bilinguals*. (Master's thesis). Brigham Young University, Provo, UT.

Barry, W., Andreeva, B., Russo, M., Dimitrova, S., & Kostadinova, T. (2003). Do Rhythm Measures Tell us Anything about Language Type? *Proceedings of XVth ICPhS, Barcelona, Spain*, *1*(November 2015), 2693–2696.

Beckman, J., Helgason, P., McMurray, B., & Ringen, C. (2011). Rate effects on Swedish VOT: Evidence for phonological overspecification. *Journal of Phonetics*, *39*(1), 39–49. https://doi.org/10.1016/j.wocn.2010.11.001

Benet, A., Gabriel, C., Kireva, E., & Pešková, A. (2012). Prosodic transfer from Italian to Spanish: Rhythmic Properties of L2 Speech and Argentinean Porteño. *Proceedings of Speech Prosody 2012*, (1), 438–441.

Best, C. (1995). A direct-realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: issues in cross-language research* (pp. 171–206). Timonium, MD: York Press.

Best, C., & Tyler, M. (2007). Nonnative and Second-Language Speech Perception: Commonalities and Complementarities. *Language Experience in Second Language Speech Learning*. https://doi.org/10.1075/lllt.17.07bes

Bialystok, E., Craik, F. I. M., & Luk, G. (2008). Lexical access in bilinguals: Effects of vocabulary size and executive control. *Journal of Neurolinguistics*, *21*(6), 522–538.

https://doi.org/10.1016/j.jneuroling.2007.07.001

Birdsong, D., Gertken, L. M., & Amengual, M. (2012). Bilingual Language Profile: An Easy-to-Use Instrument to Assess Bilingualism. Retrieved May 1, 2018, from https://sites.la.utexas.edu/bilingual/

Bissiri, M. P., Lecumberri, M. L., Cooke, M., & Volín, J. (2011). The Role of Word-Initial Glottal Stops in Recognizing English Words. *Proceedings of the 12th Annual Conference of the International Speech Communication Associationg (Interspeech)*, 165–168.

Blumstein, S. E., & Stevens, K. N. (1979). Acoustic invariance in speech production: Evidence from measurements of the spectral characteristics of stop consonants. *The Journal of the Acoustical Society of America*, *66*(4), 1001–1017. https://doi.org/10.1121/1.383319

Boersma, P. (2002a). Spectrum: Get central moment. Retrieved April 16, 2019, from http://www.fon.hum.uva.nl/praat/manual/Spectrum__Get_central_moment___.html

Boersma, P. (2002b). Spectrum: Get skewness. Retrieved April 16, 2019, from http://www.fon.hum.uva.nl/praat/manual/Spectrum__Get_skewness___.html

Boersma, P. (2007). Spectrum: Get centre of gravity. Retrieved April 16, 2019, from http://www.fon.hum.uva.nl/praat/manual/Spectrum__Get_centre_of_gravity___.html

Boersma, P., & Weenink, D. (2018). Praat: Doing phonetics by computer [Computer Program]. Retrieved from http://praat.org

Bohn, O. (1995). Cross Language Speech Perception in Adults First Language Transfer Doesn't Tell It All. In W. Strange (Ed.), *Speech perception and linguistic experience: issues in cross-language research* (pp. 278–300). Timonium, MD: York Press.

Brown, E. K., & Copple, M. T. (2018). Constructing two phonological systems: A phonetic analysis of /p/, /t/, /k/ among early Spanish–English bilingual speakers. *International Journal of Bilingualism*, *22*(1), 51–68. https://doi.org/10.1177/1367006916651983

Butcher, A., & Tabain, M. (2004). On the back of the tongue: Dorsal sounds in Australian languages. *Phonetica*, *61*(1), 22–52. https://doi.org/10.1159/000078661

Carter, P. M. (2005). Quantifying rhythmic differences between Spanish, English, and Hispanic English. In R. S. Gess & E. J. Rubin (Eds.), *Theoretical and experimental approaches to Romance Linguistics: Selected papers from the 34th Linguistic Symposium on Romance Languages (LSRL)* (pp. 63–75). Philadelphia, PA: John Benjamins.

Carter, P. M., & Wolford, T. (2016). Cross-generational prosodic convergence in South Texas Spanish. *Spanish in Context*, *13*(1), 29–52. https://doi.org/10.1075/sic.13.1.02car

Casillas, J. V. (2015). Production and perception of the /i/-/I/ vowel contrast: The case of L2-dominant early learners of English. *Phonetica*, *72*(2–3), 185–205. https://doi.org/10.1159/000431101

Casillas, J. V. (2016). *The Longitudinal Development of Fine Phonetic Detail in Late Learners of Spanish*. University of Arizona.

Casillas, J. V., Díaz, Y., & Simonet, M. (2015). Acoustics of Spanish and English Coronal Stops. *Proceedings of the 18th International Congress of Phonetic Sciences*, 1–5.

Casillas, J. V., & Simonet, M. (2015). Production and perception of the English /ae/-/a/ contrast in switched-dominance speakers. *Second Language Research*, *32*(2), 171–195. https://doi.org/10.1177/0267658315608912

Castañeda Vicente, M. L. (1986). El V.O.T. de las oclusivas sordas y sonoras españolas. *Estudios de Fonética Experimental*, *2*, 92–110.

Chodroff, E., Golden, A., & Wilson, C. (2019). Covariation of stop voice onset time across languages: Evidence for a universal constraint on phonetic realization. *The Journal of the Acoustical Society of America*, *145*(1), EL109–EL115. https://doi.org/10.1121/1.5088035

Chomsky, N., & Halle, M. (1968). *The Sound Pattern of English*. Cambridge, Massachusetts: MIT Press. https://doi.org/10.1037/031431

Cobb, K., & Simonet, M. (2015). Adult Second Language Learning of Spanish Vowels. *Hispania*, *98*(1), 47–60. https://doi.org/10.1353/hpn.2015.0026

Colantoni, L., Steele, J., & Escudero, P. (2015). Second language speech: theory and practice. Cambridge, United Kingdom: Cambridge University Press.

Cole, J. S., Choi, H., Kim, H., & Hasegawa-Johnson, M. (2003). The effect of accent on the acoustic cues to stop voicing in Radio News speech. *5th ICPhS Barcelona*, 15–18.

Crystal, D. (2008). *A Dictionary of Linguistics and Phonetics*. John Wiley & Sons.

Cuenca, M. H. (1996). Análisis instrumental de la duración de las vocales en español. *Philologia Hispalensis*, *11*, 295–307.

Dauer, R. M. (1983). Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics*, *11*, 51–62.

Dauer, R. M. (1987). Phonetic and phonological components of language rhythm. *Proceedings of the XIth International Congress of Phonetic Sciences*, *5*, 447–450.

de Jong, K. (2004). Stress, lexical focus, and segmental focus in English: Patterns of variation in vowel duration. *Journal of Phonetics*, *32*(4), 493–516. https://doi.org/10.1016/j.wocn.2004.05.002

Delattre, P. C., Liberman, A. M., & Cooper, F. S. (1955). Acoustic Loci and Transitional Cues for Consonants. *The Journal of the Acoustical Society of America*, *27*(4), 769–773. https://doi.org/10.1121/1.1908024

Dellwo, V. (2006). Rhythm and Speech Rate: A Variation Coefficient for deltaC. In P. Karnowski & I. Szigeti (Eds.), *Language and language-processing: Proceedings of the 38th Linguistics Colloquium* (pp. 231–241). Frankfurt, Germany: Peter Lang. https://doi.org/10.5167/uzh-111789

Dmitrieva, O., Llanos, F., Shultz, A. A., & Francis, A. L. (2015). Phonological status, not voice onset time, determines the acoustic realization of onset f0 as a secondary voicing cue in Spanish and English. *Journal of Phonetics*, *49*, 77–95. https://doi.org/10.1016/j.wocn.2014.12.005

Dorman, M. F., & Raphael, L. (1980). Distribution of acoustic cues for stop consonant place of articulation in VCV syllables. *The Journal of the Acoustical Society of America*, *64*(4), 1333–1335. https://doi.org/10.1121/1.384186

Eilers, R. E., Wilson, W. R., & Moore, J. M. (1976). Discrimination of synthetic prevoiced labial stops by infants and adults. *The Journal of the Acoustical Society of America*, *60*, S91. https://doi.org/10.1121/1.2003601

Equal. (n.d.). Retrieved March 27, 2019, from http://www.wordreference.com

Escudero, P. (2005). *Linguistic Perception and Second Language Acquisition: Explaining the attainment of optimal phonological categorization*. (Doctoral dissertation). Universiteit Utrecht, Utrecht, Netherlands.

Escudero, P. (2009). The linguistic perception of similar L2 sounds. In P. Boersma & S. Hamann (Eds.), *Phonology in Perception* (pp. 152–190). New York: Mouton de Gruyter.

Eshghi, M., Alemi, M., & Zajac, D. J. (2017). Aerodynamic and laryngeal characteristics of place-dependent voice onset time differences. *Folia Phoniatrica et Logopaedica*, *68*(5), 239–246. https://doi.org/10.1159/000478523

Fischer-Jørgensen, E. (1954). Acoustic Analysis of Stop Consonants. *Le Maître Phonétique*, *32*(69), 42–59.

Flege, J. E. (1987). The production of “new” and “similar” phones in a foreign language: evidence for the effect of equivalence classification. *Journal of Phonetics*, *15*(1), 47–65. https://doi.org/0095a4470/87/010047!

Flege, J. E. (1995). Second language speech learning: theory, findings and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: issues in cross-language research* (pp. 233–277). Timonium, MD: York Press.

Flege, J. E., Bohn, O.-S., & Jang, S. (1997). Effects of experience on non-native speakers’ production and perception of English vowels. *Journal of Phonetics*, *25*(4), 437–470. https://doi.org/10.1006/jpho.1997.0052

Flege, J. E., & Eefting, W. (1987). Production and perception of English stops by native Spanish speakers. *Journal of Phonetics*, *15*(1), 67–83.

Flege, J. E., & Fletcher, K. L. (1992). Talker and listener effects on degree of perceived foreign accent. *The Journal of the Acoustical Society of America*, *91*(1), 370–389. https://doi.org/10.1121/1.402780

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*(2), 155–187.

Flege, J. E., Munro, M. J., & MacKay, I. R. (1995). Factors affecting strength of perceived foreign accent in a second language. *The Journal of the Acoustical Society of America*, *97*(5 Pt 1), 3125–3134. https://doi.org/10.1121/1.413041

Flege, J. E., Munro, M. J., & Skelton, L. (1992). Production of the word-final English /t/-/d/ contrast by native speakers of English, Mandarin, and Spanish. *The Journal of the Acoustical Society of America*, *1*, 128–143.

Flege, J. E., & Port, R. (1981). Cross-language phonetic interference: Arabic to English. *Language and Speech*, *24*(2), 125–146. https://doi.org/10.1177/002383098102400202

Forrest, K., Weismer, G., Milenkovic, P., & Dougall, R. N. (1988). Statistical analysis of word-initial voiceless obstruents: Preliminary data. *The Journal of the Acoustical Society of America*, *84*(1), 115–123. https://doi.org/10.1121/1.396977

García-Sierra, A., Ramírez-Esparza, N., Silva-Pereyra, J., Siard, J., & Champlin, C. A. (2012). Assessing the double phonemic representation in bilingual speakers of Spanish and English: An electrophysiological study. *Brain and Language*, *121*(3), 194–205. https://doi.org/10.1016/j.bandl.2012.03.008

Gass, S. M. (1984). Development of speech perception and speech production abilities in adult second language learners. *Applied Psycholinguistics*, *5*(01), 51–74. Retrieved from http://www.journals.cambridge.org/abstract_S0142716400004835%5Cnhttp://journals.cambridge.org/production/action/cjoGetFulltext?fulltextid=2733760

Gay, T. (1978). Effect of speaking rate on vowel formant movements. *The Journal of the Acoustical Society of America*, *63*(1), 223. https://doi.org/10.1121/1.381717

Gerken, L. (2009). *Language Development*. San Diego, CA: Plural Publishing.

Gonzales, K., & Lotto, A. J. (2013). A Bafri, un Pafri: Bilinguals' pseudoword identifications support language-specific phonetic systems. *Psychological Science*, *24*(11), 2135–2142. https://doi.org/10.1177/0956797613486485

González-Bueno, M. (1997). The Effects of Formal Instruction on the Acquisition of Spanish Stop Consonants. In W. R. Glass & A. T. Pérez-Leroux (Eds.), *Contemporary Perspectives on the Acquisition of Spanish, Volume 2: Production, Processing, and Comprehension* (pp. 57–75). Somerville, MA: Cascadilla Press.

Gower, J. C. (1971). A General Coefficient of Similarity and Some of Its Properties.

Biometrics, *27*(4), 857–871. https://doi.org/10.1109/ultsym.1987.199076

Grabe, E., & Low, E. L. (2002). Durational Variability in Speech and the Rhythm Class Hypothesis. *Papers in Laboratory Phonology* 7, 515–546. https://doi.org/10.1515/9783110197105.2.515

Green, K. P., Zampini, M. L., & Magloire, J. (1997). An examination of word–initial-stop closure interval in English, Spanish, and Spanish–English bilinguals. *The Journal of the Acoustical Society of America*, *102*, 3136.

Grenon, I., & White, L. (2008). Acquiring rhythm: A comparison of L1 and L2 speakers of Canadian English and Japanese. In H. Chan, H. Jacob, & E. Kapia (Eds.), *BUCLD 32: Proceedings of the 32nd annual Boston University Conference on Language Development* (pp. 155–166). Somerville, MA: Cascadilla Press.

Guion, S. G., Flege, J. E., Liu, S. H., & Yeni-Komshian, G. H. (2000). Age of learning effects on the duration of sentences produced in a second language. *Applied Psycholinguistics*, *21*(2), 205–228. https://doi.org/10.1017/S0142716400002034

Gut, U. (2003). Prosody in second language speech production: The role of native language. *Fremdsprachen Lehren Und Lernen*, *32*, 133–152.

Han, M. S. (1962). The feature of duration in Japanese. *Onsei No Kenkyu (The Study of Sounds)*, *10*, 65–80.

Harris, J. W. (1984). La espirantización en castellano y la representación fonológica autosegmental. *Estudis Gramaticals*, *1*, 149–167.

Henriksen, N., & Fafulas, S. (2017). Prosodic timing and language contact: Spanish and Yagua in Amazonian Peru. *Studies in Hispanic and Lusophone Linguistics*, *10*(2), 225–257.

Hoequist Jr., C. (1983). Durational correlates of linguistic rhythm categories. *Phonetica*, *40*, 19–31.

Hoff, E., Core, C., Place, S., Rumiche, R., Señor, M., & Parra, M. (2012). Dual language exposure and early bilingual development. *Journal of Child Language*, *39*(1), 1–27. https://doi.org/10.1017/S0305000910000759

Hoversten, L. J., Brothers, T., Swaab, T. Y., & Traxler, M. J. (2017). Early processing of orthographic language membership information in bilingual visual word recognition: Evidence from ERPs. *Neuropsychologia*, *103*(June), 183–190. https://doi.org/10.1016/j.neuropsychologia.2017.07.026

Hualde, J. I. (2005). *The sounds of Spanish*. New York, NY: Cambridge University Press.

Hualde, J. I., Olarrea, A., Escobar, A. M., & Travis, C. (2010). *Introducción a la lingüística hispánica* (2nd Ed.). Cambridge, UK;New York; Cambridge University Press.

Izura, C., Cuetos, F., & Brysbaert, M. (2013). Lextale-Esp: A Test to Rapidly and Efficiently Assess the Spanish Vocabulary Size. *Psicológica: International Journal of Methodology and Experimental Psychology*, *35*(1), 49–66. https://doi.org/10.7498/aps.63.104215

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2017). An Introduction to Statistical Learning. https://doi.org/10.1007/978-1-4614-7138-7

Johnson, K. (2011). *Acoustic and auditory phonetics* (3rd Ed.). John Wiley & Sons.

Jongman, A., & Blumstein, S. E. (1985). Acoustic properties for dental and alveolar stop consonants: a cross-language study. *Journal of Phonetics*, *13*, 235–251.

Jongman, A., Wayland, R., & Wong, S. (2000). Acoustic Characteristics of English Fricatives. *Working Papers of the Cornell Phonetics Laboratory*, *12*(3), 195–205. https://doi.org/10.1121/1.1288413

Kehoe, M. (2002). Developing vowel systems as a window to bilingual phonology. *International Journal of Bilingualism*, *6*(3), 315–334.

Keppel, G., & Wickens, T. D. (2004). *Design and Analysis: A Researcher's Handbook* (4th Ed.). Upper Saddle River, NJ: Pearson Prentice Hall.

Kessinger, R. H., & Blumstein, S. E. (1997). Effects of speaking rate on voice-onset time in Thai, French, and English. *Journal of Phonetics*, *25*(2), 143–168. https://doi.org/10.1006/jpho.1996.0039

Kewley-Port, D. (1983). Time-varying features as correlates of place of articulation in stop consonants. *The Journal of the Acoustical Society of America*, *73*(1), 322–335. https://doi.org/10.1121/1.388813

Kewley-Port, D., & Luce, P. A. (1984). Time-varying features of initial stop consonants in auditory running spectra: A first report. *Perception & Psychophysics*, *35*(4), 353–360. https://doi.org/10.3758/BF03206339

Kinoshita, N., & Sheppard, C. (2008). The acquisition of lexical rhythm and duration by Japanese second language learners. *Laboratory Phonology*, (11), 79–80.

Kinoshita, N., & Sheppard, C. (2011). Validating acoustic measures of speech rhythm for second language acquisition. *Proceedings of the 17th International Congress of Phonetic Sciences (ICPhS XVII)*, (August), 1086–1089.

Klatt, D. H. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *The Journal of the Acoustical Society of America*, *59*(5), 1208. https://doi.org/10.1121/1.380986

Kobatake, H., & Ohtani, S. (1987). Spectral transition dynamics of voiceless stop consonants. *The Journal of the Acoustical Society of America*, *81*, 1146–1151. https://doi.org/10.1121/1.394635

Ladefoged, P. (2003). *Phonetic Data Analysis: An Introduction to Fieldwork and Instrumental Techniques*. Malden, MA: Blackwell Publishing Ltd.

Ladefoged, P. (2006). *A Course in Phonetics* (5th Ed.). Boston, MA: Michael Rosenberg.

Lahiri, A., Gewirth, L., & Blumstein, S. E. (2005). A reconsideration of acoustic invariance for place of articulation in diffuse stop consonants: Evidence from a cross-language study. *The Journal of the Acoustical Society of America*, *76*(2), 391–404. https://doi.org/10.1121/1.391580

Lawrence, M. A. (2016). ez: Easy Analysis and Visualization of Factorial Experiments. R package version 4.4-0. Retrieved from https://cran.r-project.org/package=ez

Lehiste, I. (1973). Rhythmic units and syntactic units in production and perception. *The Journal of the Acoustical Society of America*, *54*(5), 1228–1234. https://doi.org/10.1121/1.1914379

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English. *Behavior Research Methods*, *44*(2), 325–343. https://doi.org/10.3758/s13428-011-0146-0

Li, A., & Post, B. (2014). L2 Acquisition of Prosodic Properties of Speech Rhythm. *Studies in Second Language Acquisition*, *36*(02), 223–255. https://doi.org/10.1017/S0272263113000752

Liberman, A. M., Delattre, P. C., & Cooper, F. S. (1952). The Role of Selected Stimulus-Variables in the Perception of the Unvoiced Stop Consonants. *The American Journal of Psychology*, *65*(4), 497–516.

Liberman, A. M., Delattre, P. C., Cooper, F. S., & Gerstman, L. J. (1954). The Role of Consonant-Vowel Transitions in the Perception of the Stop and Nasal Consonants. *Psychological Monographs: General and Applied*, *68*(8), 1–13.

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, *54*(5), 358–368. https://doi.org/10.1037/h0044417

Liljencrants, J., & Lindblom, B. (1972). Numerical Simulation of Vowel Quality Systems: The Role of Perceptual Contrast. *Language*, *48*(4), 839–862.

Lin, H., & Wang, Q. (2005). Vowel quantity and consonant variance: A comparison between Chinese and English. *Proceedings of Between Stress and Tone*, 1–24.

Lisker, L. (1957). Closure duration and the intervocalic voiced-voiceless distinction in English. *Language*, *33*(1), 42–49. Retrieved from http://www.jstor.org/stable/10.2307/410949

Lisker, L., & Abramson, A. S. (1964). A Cross-Language Study of Voicing in Initial Stops: Acoustical Measurements. *WORD*, *20*(3), 384–422. https://doi.org/10.1080/00437956.1964.11659830

Lisker, L., & Abramson, A. S. (1967). Some Effects of Context on Voice Onset Tme in English Stops. *Language and Speech*, *10*, 1–28. https://doi.org/10.1177/002383096701000101

Lloyd James, A. (1940). *Speech signals in telephony*. London, UK: Sir Isaac Pitman & Sons, LTD.

Low, E. L., & Grabe, E. (1995). Prosodic patterns in Singapore English. *Proceedings of the 13th International Congress of Phonetics Sciences (ICPhS)*, *3*, 636–639.

Low, E. L., Grabe, E., & Nolan, F. (2000). Quantitative Characterizations of Speech Rhythm: Syllable-Timing in Singapore English. *Language and Speech*, *43*(4), 377–401.

Luce, P. A., & Charles-Luce, J. (1985). Contextual effects on vowel duration, closure duration, and the consonant/vowel ratio in speech production. *The Journal of the Acoustical Society of America*, *78*(6), 1949–1957. https://doi.org/10.1121/1.392651

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

Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., Hornik, K., Studer, M., … Kozlowski, K. (2018). Cluster. Retrieved from https://cran.r-project.org/web/packages/cluster/cluster.pdf

Magloire, J., & Green, K. P. (1999). A cross-language comparison of speaking rate effects on the production of Voice Onset Time in English and Spanish. *Phonetica*, *56*(3–4), 158–185. https://doi.org/10.1159/000028449

Mampe, B., Friederici, A. D., Christophe, A., & Wermke, K. (2009). Newborns' Cry Melody Is Shaped by Their Native Language. *Current Biology*, *19*(23), 1994–1997. https://doi.org/10.1016/j.cub.2009.09.064

Marín Gálvez, R. (1995). La duración vocálica en español. *Estudios de Lingüística. Universidad de Alicante (ELUA)*, *10*, 213–226.

Martínez Celdrán, E. (1991). Duración y tensión de las oclusivas no iniciales del español: un estudio perceptivo. *Revista Argentina de Lingüística*, 7, 51–71.

Martínez Celdrán, E. (1993). La percepción categorial de /b–p/ en español basada en las diferencias de duración. *Estudios de Fonética Experimental*, *5*, 223–239.

Martínez, R. A. (2010). Spanglish as a Literacy Tool: Toward an Understanding of the Potential Role of Spanish–English Code-Switching in the Development of Academic Literacy. *Research in the Teaching of English*, *45*(2), 124–149.

Mehler, J., Dupoux, E., Nazzi, T., & Dehaene-Lambertz, G. (1996). Coping with linguistic diversity: the infant's viewpoint. In J. L. Morgan & K. Demuth (Eds.), *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition* (pp. 101–116). Mahwah, NJ: Lawrence Erlbaum Associates.

Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoncini, J., & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition*, *29*, 143–178. https://doi.org/10.1016/0010-0277(88)90035-2

Menke, M., & Face, T. (2010). Second language Spanish vowel production: an acoustic analysis. *Studies in Hispanic and Lusophone Linguistics*, *3*(1), 181–214.

Mennen, I. (2004). Bi-directional interference in the intonation of Dutch speakers of Greek. *Journal of Phonetics*, *32*(4), 543–563. https://doi.org/10.1016/j.wocn.2004.02.002

Mok, P. P. K., & Dellwo, V. (2008). Comparing native and non-native speech rhythm using acoustic rhythmic measures: Cantonese, Beijing Mandarin and English. *Proceedings of Speech Prosody 2008*, (January 2016), 423–426.

Müller, E. M., & Brown, W. S. (1980). *Variations in the Supraglottal Air Pressure Waveform and Their Articulatory Interpretation*. *Speech and Language: Advances in Basic Research and Practice* (Vol. 4). ACADEMIC PRESS, INC. https://doi.org/10.1016/b978-0-12-608604-1.50013-6

Nadeu, M., & Renwick, M. E. L. (2016). Variation in the lexical distribution and implementation of phonetically similar phonemes in Catalan. *Journal of Phonetics*, *58*, 22–47. https://doi.org/10.1016/j.wocn.2016.05.003

Nava, E., & Zubizarreta, M. L. (2009). Order of L2 Acquisition of Prosodic Prominence Patterns: Evidence from L1 Spanish/L2 English Speech. In *Proceedings of the 3rd Conference on Generative Approaches to Language Acquisition North America (GALANA*

2008) (pp. 175–187). Somerville, MA: Cascadilla Proceedings Project. Retrieved from http://www.lingref.com/cpp/galana/3/paper2318.pdf

Navarro Tomás, T. (1917). Cantidad de las vocales inacentuadas. *Revista de Filología Española*, *4*, 371–388.

Navarro Tomás, T. (1918). Diferencias de duración entre las consonantes españolas. *Revista de Filología Española*, *5*, 367–393.

Nazzi, T., Bertoncini, J., & Mehler, J. (1998). Language discrimination by newborns: Toward an understanding of the role of rhythm. *Journal of Experimental Psychology: Human Perception and Performance*, *24*(3), 756–766. https://doi.org/10.1037/0096-1523.24.3.756

Ohala, J. J. (1980). Moderator's summary of symposium "Phonetic universals in phonological systems and their explanation." *Proceedings of the 9th International Congress of Phonetic Sciences*, *II*(1), 181–184.

Ortega-Llebaria, M., & Prieto, P. (2007). Disentangling stress from accent in Spanish: Production patterns of the stress contrast in deaccented syllables. In P. Prieto, J. Mascaró, & M. Solé (Eds.), *Segmental and prosodic issues in Romance phonology* (pp. 155–176). https://doi.org/10.1075/cilt.282.11ort

Partanen, E., Kujala, T., Näätänen, R., Liitola, A., Sambeth, A., & Huotilainen, M. (2013). Learning-induced neural plasticity of speech processing before birth. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(37), 15145–15150. https://doi.org/10.1073/pnas.1302159110

Paul Lennon. (1990). Investigating Fluency in EFL: A Quantitative Approach. *Language Learning*, *40*(3), 387–417. https://doi.org/10.1111/j.1467-1770.1990.tb00669.x

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., … Kristoffer

Lindeløv, J. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*, 195–203.

Pike, K. L. (1947). *The Intonation of American English*. Ann Arbor: University of Michigan Press.

Piske, T., Flege, J. E., MacKay, I. R. A., & Meador, D. (2002). The production of English vowels by fluent early and late Italian-English bilinguals. *Phonetica*, *59*(1), 49–71. https://doi.org/10.1159/000056205

Port, R. F. (1979). The influence of tempo on stop closure duration as a cue for voicing and place. *Journal of Phonetics*, 7(1), 45–56.

Quilis, A. (1993). Tratado de fonología y fonética españolas. Madrid: Gredos.

R Core Team. (2015). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.r-project.org/

Ramus, F., Nespor, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, *73*(3), 265–292. https://doi.org/10.1016/S0010-0277(99)00058-X

Rasier, L., & Hiligsmann, P. (2007). Prosodic transfer from L1 to L2. Theoretical and methodological issues. *Nouveaux Cahier s de l Inguistique Française*, *28*, 41–66.

Rathcke, T. V, & Smith, R. H. (2015). Rhythm class perception by expert phoneticians. *Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS 2015)*, 1–5.

Recasens, D., & Espinosa, A. (2009). An articulatory investigation of lingual coarticulatory resistance and aggressiveness for consonants and vowels in Catalan, *125*(4), 2288–2298. https://doi.org/10.1121/1.3089222

Reeder, J. T. (1998). English speakers' acquisition of voiceless stops and trills in L2 Spanish.

Texas Papers in Foreign Language Education, *3*(3), 101–118.

Repp, B. H. (1984). Stop Consonant Manner and Place of Articulation. *Language and Speech*, *27*(3), 245–254.

Ronquest, R. E. (2013). An Acoustic Examination of Unstressed Vowel Reduction in Heritage Spanish. In C. Howe, S. E. Blackwell, & M. Lubbers Quesada (Eds.), *Selected Proceedings of the 15th Hispanic Linguistics Symposium* (pp. 157–171). Somerville, MA: Cascadilla Proceedings Project.

RStudio Team. (2016). RStudio: Integrated Development for R. Boston, MA: RStudio, Inc. Retrieved from http://www.rstudio.com/

Sancier, M. L., & Fowler, C. A. (1997). Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *Journal of Phonetics*, *25*(4), 421–436. https://doi.org/10.1006/jpho.1997.0051

Schmidt, A. M., & Flege, J. E. (1996). Speaking rate effects on stops produced by spanish and english monolinguals and spanish/english bilinguals. *Phonetica*, *53*(3), 162–179. https://doi.org/10.1159/000262196

Simonet, M. (2013). The L2 Acquisition of Spanish Phonetics and Phonology. In J. I. Hualde, A. Olarrea, & E. O'Rourke (Eds.), *Blackwell Handbooks in Linguistics: Handbook of Hispanic Linguistics*. John Wiley & Sons.

Simonet, M. (2014). Phonetic consequences of dynamic cross-linguistic interference in proficient bilinguals. *Journal of Phonetics*, *43*(1), 26–37. https://doi.org/10.1016/j.wocn.2014.01.004

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). *Social and Linguistic Speech Prosody*, 202–206.

Spector, P. (2011). Cluster Analysis. Retrieved April 26, 2019, from https://www.stat.berkeley.edu/~s133/Cluster2a.html

Steele, J. (1775). *An essay towards establishing the melody and measure of speech to be expressed and perpetuated by peculiar symbols*. London: W. Bowyer and J. Nichols, for J. Almon.

Stevens, J. J. (2001). Study abroad learners' acquisition of the Spanish voiceless stops. *MIFLC Review*, *10*, 137–151.

Stevens, J. J. (2011). Vowel duration in second language Spanish vowels: study abroad versus at-home learners. *Arizona Working Papers in SLA & Teaching*, *18*, 77–104.

Stockmal, V., Markus, D., & Bond, D. (2005). Measures of native and non-native rhythm in a quantity language. *Language and Speech*, *48*(1), 55–63. Retrieved from http://las.sagepub.com/content/48/1/55.short

Stoel-Gammon, C. (1998). Sounds and words in early language acquisition: The relationship between lexical and phonological development. In R. Paul (Ed.), *Exploring the speech/language connection* (pp. 25–52).

Stoel-Gammon, C., Williams, K., & Buder, E. (1994). Cross-language differences in phonological acquisition: Swedish and American /t/. *Phonetica*, *51*(1–3), 146–158. https://doi.org/10.1159/000261966

Strange, W., & Shafer, V. L. (2008). Speech perception in second language learners: The re-education of selective perception. In J. G. Hansen Edwards & M. L. Zampini (Eds.), *Phonology and Second Language Acquisition* (pp. 153–191). John Benjamins Publishing.

Sundara, M. (2005). Acoustic-phonetics of coronal stops: A cross-language study of Canadian English & Canadian French. *The Journal of the Acoustical Society of America*, *118*(2), 1026–1037. https://doi.org/10.1121/1.1953270

Sundara, M., Polka, L., & Baum, S. R. (2006). Production of coronal stops by simultaneous bilingual adults. *Bilingualism: Language and Cognition*, *9*(1), 97–114. https://doi.org/10.1017/S1366728905002403

Thomas, E. R., & Carter, P. M. (2006). Prosodic rhythm and African American English. *English World-Wide*, *27*(3), 331–355. https://doi.org/10.1075/eww.27.3.06tho

Toribio, A. J., & Durán, L. (2018). Understanding and Leveraging Spanish Heritage Speakers' Bilingual Practices. In K. Potowski (Ed.), *The Routledge Handbook of Spanish as a Heritage Language* (pp. 284–298). https://doi.org/10.4324/9781315735139-19

Torreblanca, M. (1979). Un rasgo fonológico de la lengua española. *Hispanic Review*, *47*(4), 455–468.

Torreira, F., & Ernestus, M. (2011). Realization of voiceless stops and vowels in conversational French and Spanish. *Laboratory Phonology*, *2*, 331–353.

Tortel, A., & Hirst, D. (2010). Rhythm metrics and the production of English L1/L2. *Proceedings of Speech Prosody 2010*, 1–4.

Towell, R., Hawkings, R., & Bazergui, N. (1996). The Development of Fluency in Advanced Learners of French. *Applied Linguistics*, *17*(1), 84–119. https://doi.org/10.1093/applin/17.1.84

Trofimovich, P., & Baker, W. (2006). Learning Second Language Suprasegmentals: Effect of L2 Experience on Prosody and Fluency Characteristics of L2 Speech. *Studies in Second Language Acquisition*, *28*(1), 1–30.

Turk, A. E., & Sawusch, J. R. (1997). The domain of accentual lengthening in American English. *Journal of Phonetics*, *25*, 25–41.

Turk, A. E., & White, L. (1999). Structural influences on accentual lengthening in English.

Journal of Phonetics, *27*, 171–206.

Upton, C., & Kretzschmar Jr., W. A. (2017). *The Routledge Dictionary of Pronunciation for Current English* (2nd Ed.). Taylor and Francis Group.

Valls-Ferrer, M. (2011). *The development of oral fluency and rhythm during a study abroad period*. (Doctoral dissertation). Universitat Pompeu Fabra, Barcelona, Spain.

van Leussen, J.-W., & Escudero, P. (2015). Learning to perceive and recognize a second language: the L2LP model revised. *Frontiers in Psychology*, *6*(August), 1000. https://doi.org/10.3389/fpsyg.2015.01000

Van Summers, W. (1987). Effects of stress and final-consonant voicing on vowel production: Articulatory and acoustic analyses. *The Journal of the Acoustical Society of America*, *82*(3), 847–863. https://doi.org/10.1121/1.395284

Warner, N., & Arai, T. (2002). The role of the mora in the timing of spontaneous Japanese speech. *The Journal of the Acoustical Society of America*, *109*(3), 1144–1156. https://doi.org/10.1121/1.1344156

Watt, D. (2013). Research Methods in Speech Acoustics. In M. J. Jones & R.-A. Knight (Eds.), *The Bloomsbury Companion to Phonetics* (pp. 79–97). Bloomsbury Academic, NY.

Westfall, P. H. (2014). Kurtosis as Peakedness, 1905–2014. R.I.P. *The American Statistician*, *68*(3), 191–195. https://doi.org/10.1080/00031305.2014.917055

White, L., & Mattys, S. L. (2007). Calibrating rhythm: First language and second language studies. *Journal of Phonetics*, *35*(4), 501–522. https://doi.org/10.1016/j.wocn.2007.02.003

Whitley, S. M. (2002). *Spanish/English Contrasts: A Course in Spanish Linguistics*. Washington, D.C.: Georgetown University Press.

Whitworth, N. (2002). Speech rhythm production in three German–English bilingual families. In D. Nelson (Ed.), *Leeds Working Papers in Linguistics and Phonetics* (Vol. 9, pp. 175–205).

Wong-Fillmore, L. (1979). Individual Differences in Second Language Acquisition. In C. Fillmore, D. Kempler, & W. Wang (Eds.), *Individual differences in language ability and language behavior* (pp. 203–228). San Diego: Academic Press Inc.

Young, D. J. (2008). Learning on Student Outcomes in a Redesigned. *CALICO Journal*, *26*(1), 160–181.

Zampini, M. L. (1998). The Relationship between the Production and Perception of L2 Spanish Stops. *Texas Papers in Foreign Language Education*, *3*(3), 85–100.

Zampini, M. L. (2002). Sensitivity to voiceless closure in the perception of Spanish and English stop consonants. *The Journal of the Acoustical Society of America*, *112*, 2383.

Zampini, M. L. (2013). Voice Onset Time in Second Language Spanish. In K. L. Geeslin (Ed.), *Handbook of Spanish Second Language Acquisition* (pp. 113–129). Wiley-Blackwell.