Title
Methodological, Linguistic, and Social Effects in Language Alternation: Evidence from Voice Onset Time in Spanish-English Bilinguals

Author
Gutierrez Topete, Ernesto Rafael

Publication Date
2023

Methodological, Linguistic, and Social Effects in Language Alternation: Evidence from Voice Onset Time in Spanish-English Bilinguals

By

Ernesto R. Gutiérrez Topete

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Hispanic Languages and Literatures

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Associate Professor Justin Davidson, Chair
Professor Terry Regier
Professor Keith Johnson
Associate Professor Daniel J. Olson

Summer 2023

Methodological, Linguistic, and Social Effects in Language Alternation: Evidence from Voice Onset Time in Spanish-English Bilinguals

Abstract

Methodological, Linguistic, and Social Effects in Language Alternation: Evidence from Voice Onset Time in Spanish-English Bilinguals

by

Ernesto R. Gutiérrez Topete

Doctor of Philosophy in Hispanic Languages and Literatures

University of California, Berkeley

Associate Professor Justin Davidson, Chair

In phonetics research, language alternation–including code switching (speaker-initiated) and cued switching (researcher-prompted)–can be used as a tool to investigate various aspects of speech production and perception in bilingual or multilingual speakers (Bullock & Toribio, 2009a). Studies on the production of voice onset time (VOT) during language alternation have demonstrated that bilingual speakers—for example, Spanish-English bilinguals—have a convergence effect, in which VOT of a given language near switch sites becomes more similar to the VOT norms of the other language (e.g., Toribio et al., 2005; Bullock et al., 2006; Balukas & Koops, 2015; Piccinini & Arvaniti, 2015; Olson, 2016). As Olson (in press) summarizes, some general patterns have been identified in the literature: (1) "unidirectional interference has been the most common finding in the literature. When unidirectional interference is reported, the language with long-lag VOTs shifts in the direction of the short-lag language", and (2) "when bidirectional interference has been found, the magnitude of the switch differs between the two languages. The degree of shift is generally larger for the long-lag language than the short-lag language" (Olson, in press, p. 7). However, regardless of these general patterns, it is evident that a great number of distinct outcomes has been reported, namely, convergence presence and directionality (no convergence, unidirectional convergence, or bidirectional convergence), the language(s) more prone to convergence (only long-lag language or both), and the magnitude of the effect (small or large shifts). The wide array of different task types and methodologies used to study this phenomenon makes it all the more difficult to pinpoint the cause of the aforementioned inconsistent directionality, propensity, and magnitude of the convergence effect. For example, some studies relied on word list reading tasks (e.g., Olson, 2013) or passage reading tasks (e.g., Toribio et al., 2005; Bullock et al., 2006), and others used speech produced during sociolinguistic interviews (e.g., Balukas & Koops, 2015), inter-subject conversations (e.g., Balukas & Koops, 2015; Piccinini & Arvaniti, 2015), or puzzle tasks (e.g., Piccinini & Arvaniti, 2015).

In order to shine light on the possible effect of task type on the directionality and magnitude of convergence effects in VOT bilingualism research, the present study analyzes VOT productions across four of the most popular research tasks (i.e., word list reading, passage reading, puzzle–spot the difference–and casual interview tasks) from a single group of Spanish-English bilingual speakers to obtain VOT measurements. In addition to the aspect of research task, the present study incorporates linguistic (i.e., language, place of articulation, and speech rate) and social (i.e., language history, language proficiency, language usage, and language attitudes) factors that could help predict VOT production patterns in language alternation. A total of 60 Spanish-English bilingual subjects participated in the four aforementioned tasks for this study, which yielded nearly 65 hours of recorded code-switched speech, in addition to a standardized demographic questionnaire: the Bilingual Language Profile (Birdsong et al., 2012). Data collection took place in a sound booth at the Berkeley PhonLab. The audio for the word list and passage reading tasks were annotated by hand. The audio for the spot-the-difference puzzle and the interview were segmented into Spanish and English speech. A Python script generated automatic transcriptions for each language using OpenAI's Whisper language model for automatic speech recognition (Radford et al., 2022). The data for all four tasks were then forced aligned using the Montreal Forced Aligner (McAuliffe et al., 2017). Through these aligned annotations, VOT measurements were obtained for word-initial voiceless stops /p t k/ in non-cognate words in both languages using AutoVOT software (Keshet et al., 2014).

Two statistical models were performed for this study–one for the methodological and linguistic factors and another for social factors. The results from the first model indicate that there is an effect of research task, with passage reading and interview tasks displaying the highest degree of convergence–in English, specifically–compared with the word list and puzzle tasks. In addition, there were significant results of language (English yielded longer VOT), place of articulation (for English, /p/ was shorter than /t k/; for Spanish, /p t/ were shorter than /k/), and speech rate (slower speech rate led to longer VOT). Finally, the second statistical model reports relevant factors of language proficiency (participants who report higher ability scores for reading and writing–regardless of language–display less convergence than participants who report higher scores for speaking and comprehension abilities), language history (participants who learned Spanish later in life display more convergence), language usage (participants who use more English with friends and in the workplace display less convergence), and language attitudes (participants who more closely identify with a Spanish-speaking culture display more convergence than participants who identify with an English-speaking culture). The social model also indicated that participants who had higher exposure to Spanish spoken with an American English accent display less convergence than participants with exposure to any other type of accented speech.

The results from this study are discussed in relation to previous empirical studies and predictions made by relevant theoretical frameworks for (Spanish-English) bilingualism and language alternation. In particular, this study elaborates on (a) the potential language processing mechanism at play across the different types of research tasks during language alternation, (b) the significance of divergent VOT patterns of /t/ between the two languages, and (c) the effects of exposure to accented speech on a speaker's own speech patterns. All in all, this research study provides (1) a thorough analysis and comparison of the research methodologies typically used in code switching studies in order to uncover task effects in production studies and (2) a better understanding of the language processing mechanisms that are engaged during language alternation behaviors.

To my family

Este logro es por y para ustedes.
Gracias por su apoyo y cariño incondicional.

Contents

List of Figures

List of Tables

Acknowledgments

As I reflect on my academic career at UC Berkeley, I am filled with gratitude to everyone who has had an impact on my graduate student experience, which has culminated in the completion and submission of this dissertation. From my family to friends and colleagues to mentors and advisors, I thank you for your positive influence.

I want to start with my dissertation committee: Dr. Davidson, since the beginning I knew that you would be my biggest advocate and cheerleader. Through your advising, you have shaped me into a scholar who is methodologically informed and socially conscious. From encouraging me to follow my curious mind to pushing me to become involved in academic conferences and publications, you have helped me develop a research program that I've truly enjoyed (and yes, I'm including this dissertation–even after the countless hours of work it required). Ultimately, your flexibility with my academic trajectory, gave me the opportunity to explore other areas and acquire skills that I can incorporate into my own scholarship, allowing me to grow as a linguist with a varied toolkit at my disposal. This year you saw your first two advisees graduate, but I'm confident that you will lead the many students to come through the finish line, sending forth an army of linguists ready to tackle the big social and linguistic questions of our modern world.

Dr. Regier, from the first computational linguistics course I took with you, where I often felt overwhelmed with the wave of new terminology and techniques, to our discussion about my dissertation statistical models, in which I felt better-versed, my improvement has been palpable. The new set of skills I acquired through coursework and office hours consultations with you has invaded my scholarship ever since, which in my view has made my projects more interesting. Thank you for your patience towards a novice coder and for your never-ending willingness to advise me and participate in my milestone requirements (that is, QE's and dissertation).

Dr. Johnson, whenever I got stuck on a phonetics question or if I had an unsuccessful search through the previous literature, I knew that I could always reach out to you for guidance. Your extensive and unparalleled knowledge of the discipline and prior studies makes you an invaluable resource for a new phonetician. Thank you for your willingness to meet and guide me so that I could grow as a scholar. Even though you're now retired, I'm glad you're still involved in academia, for that gives others the opportunity to take advantage of your wealth of knowledge.

Dr. Olson, even though we only started communicating two years ago as you joined my dissertation committee, I must say that I've learned a lot from you. From reading your publications to exchanging personal correspondence, your knowledge and advice–especially as it relates to code switching, a topic I'm deeply interested in–has taught me a lot and has improved my dissertation project. I thank you for sharing this knowledge with me (and the world), and I hope to continue our discussions on this topic in the future, as I further my research on code switching.

Dr. Lin, as I've mentioned to you before, the first course I took with you (Phonetic Theory) led me into the study of phonetics. The combination of the interesting contents of the course and your teaching had a positive impact on my trajectory, ushering me in a direction I was not expecting. Your great teaching skills are only surpassed by your amazing mentoring. Thank you for your invaluable guidance and support on my QP1, QE's, and initial stages of my dissertation. I will never forget the many ways you helped me advance and complete my graduate studies.

I am also grateful to my many colleagues and friends from the HLL, RLL and Linguistics programs. Interactions with you in academic and social settings helped me maintain my sanity and stay motivated throughout my graduate student career. I particularly want to thank you, Annie Helms, for providing advice on and assistance with manuscripts, practice talks, research ideas in their infancy, and many, many other things; I'm forever indebted to you. Also thank you for being a great friend I could always count on for taking an afternoon to decompress away from campus or for getting help with plant-sitting. Ben Papadopoulos, thank you, friend, for always being there when I needed to talk or for celebrating all my accomplishments. You motivated me to keep going full steam ahead. Gabriella Licata, from the day we started our respective programs in the Department of Spanish and Portuguese, you have been someone I can lean on. You helped me prepare my lesson plans as a new instructor, review homework assignments for classes we took together, and de-stress during our outings outside of campus. Thank you for making my graduate experience better.

I also want to thank the undergraduate research assistants who helped me make this project possible. Jesus Duarte, thank you for your contributions towards the development of the research stimuli for this word list and passage reading tasks. Fátima Torres, Anna Knall, Martha Tellez Reza, and Joseph Olivas, thank you for your assistance segmenting the audio data for the word list, puzzle, and interview tasks. Your assistance facilitated the completion of this project. I also want to thank you, Jacky Duong, for helping me create the image used for the puzzle activity. Thank you all for your contributions towards this project.

Por último, quiero agradecer a mis papás, Rafael Gutiérrez y Gloria Topete, igual que a mis hermanos, Oscar, Ricky e Isabel. Su apoyo y cariño incondicional me han dado la motivación para seguir adelante. Me han echado porras en cada paso que doy y han tomado iniciativa cada vez que he necesitado ayuda. Las palabras no pueden expresar mi agradecimiento por tenerlos en mi vida.

Chapter 1

Introduction

In phonetics research, language alternation–including code switching (speaker-initiated) and cued switching (researcher-prompted)–can be used as a tool to investigate various aspects of speech production and perception in bilingual or multilingual speakers (Bullock & Toribio, 2009a). In particular, the study of language alternation during speech production allows us to explore the perceptual, acoustic, and articulatory characteristics and adjustments during the alternation of two or more languages within the same conversation. Such research endeavors can shed light on how the brain processes and organizes language(s) and how different language systems interact with each other during this linguistic behavior (Fricke et al., 2016). The inclusion of natural switches, as is the case with code switching, and prompted switches, as is the case with cued switching, can further help us capture more naturalistic speech with all of its complexities as well as identify and isolate specific, individual factors for manipulation and analysis (Bullock & Toribio, 2009b). Additionally, language alternation can be used to explore the manner in which phonetic and acoustic properties are affected by social, cultural, and contextual factors, such as language attitudes, age, gender, and cultural background (Auer, 1984, 2005). Overall, the study of language alternation in phonetic research–through methodologies that combine naturalistic and experimental approaches–can provide valuable insights into the complex nature of language use and the cognitive mechanisms that underlie it.

Voice onset time (VOT) is a measure of the timing relationship between the release of a stop consonant and the onset of vocal fold vibrations during speech production. This articulatory/acoustic speech feature is the most commonly studied voice property in language alternation research as it can reveal important information about how bilingual and multilingual speakers switch between languages.[1] Linguistic differences in the VOT systems that exist across language pairs provide us with the opportunity to compare potential phonetic adjustments in speech production as a result of cross-linguistic influence or other cognitive and social factors. In fact, research has shown that bilingual and multilingual speakers may

[1] All studies referenced in this dissertation–from (Grosjean & Miller, 1994) to (Toribio et al., 2005) to (Piccinini & Arvaniti, 2015), and many others mentioned below–study this phonetic feature as the variable of interest.

adjust their VOT productions when switching between languages, depending on factors such as language acquisition order, language dominance, and linguistic context (e.g., Olson, 2013; Balukas & Koops, 2015; Piccinini & Arvaniti, 2015). For instance, when switching from a language with a shorter VOT range to a language with a longer VOT range, a speaker may lengthen their VOT to accommodate the phonetic expectations of the second language, in anticipation, and vice versa; likewise, the second language with longer VOT may also display shorter VOT productions immediately following a switch due to perseverative effects from the first language with a typically shorter VOT, and vice versa (Toribio et al., 2005). Accordingly, VOT can provide valuable insights into the complex mechanisms–cognitive, linguistic, and social–that underlie language alternation and can help us better understand how bilingual and multilingual speakers use and process language.

However, there are some discrepancies among studies on language alternation and VOT due to a number of factors. For instance, a wide array of research methodologies has been used for code switching (Balukas & Koops, 2015; Piccinini & Arvaniti, 2015) and for cued switching (Toribio et al., 2005; Antoniou et al., 2011; Olson, 2013) speech production. Moreover, group comparisons have focused on language acquisition order: L1 versus L2 (Balukas & Koops, 2015) or language dominance (Piccinini & Arvaniti, 2015; Olson, 2016). Furthermore, data collection has taken place in controlled laboratory settings (Olson, 2013, 2016; Piccinini & Arvaniti, 2015) and in the field (Balukas & Koops, 2015). Thus, it comes as no surprise that these studies have reported inconsistent results: no difference between monolingual and bilingual speech (Grosjean & Miller, 1994), tandem shift (Piccinini & Arvaniti, 2015), unidirectional convergence (Olson, 2013; Antoniou et al., 2011; Balukas & Koops, 2015), and bidirectional convergence (Toribio et al., 2005; Olson, 2016), just to name some of the main differences reported in the literature on the phonetics of language alternation.

In light of the inconsistent results of VOT measurements during language alternation, some questions that come to mind are: How are research methodologies influencing speech behaviors? How is the comparison of different bilingual populations affecting production results? And what are some of social factors that mediate such behavior? The discrepancies among studies on language alternation and VOT express the critical need for more systematic and standardized approaches to the study of this phenomenon. By taking into account the methodological and contextual factors that may influence VOT production measurements, we can develop a more comprehensive understanding of the complex mechanisms that underlie language alternation. Accordingly, in the present study, I investigate VOT production during four of the most common language alternation tasks: word list reading, passage reading, puzzle completion, and casual interview (the first two are part of the cued switching experimental paradigm, and the last two are part of the code switching experimental paradigm), in order to identify potential task artifacts. I further analyze demographic and linguistic profile data to discern the social factors mediating phonetic patterns. In short, the purpose of this study is to identify methodological effects of research task in speech production during language alternation in a single group of Spanish-English bilinguals in a way that controls (methodological) variables that may confound cross-study comparisons in the literature.

Nonetheless, it is needless to say that language alternation is not simply governed by task or speech mode–monolingual or bilingual mode. Rather, there are additional cognitive, linguistic, and social factors that may influence speakers' speech behaviors–in monolingual speech and language alternation–across a wide array of tasks, alike. Therefore, the remainder of this dissertation provides: (a) the background and theoretical framework that provide the foundation for the present study, including the cognitive, linguistic, and social factors identified as relevant to the present research study in chapter 2, (b) an in-depth description of the research questions that motivated this study and the methodology that guided the implementation of this project in chapter 3, (c) the results and discussion obtained for each of the different sections examined in this study in chapter 4, and (d) the conclusion and implications of the present study for the general bilingualism research community in chapter 5.

Chapter 2

The phonetics of language alternation

The term *code switching* is typically used in the literature to refer to the phenomenon in which a speaker engages two languages in natural conversation–often within the same utterance. Such behavior is initiated by the speaker themselves. Conversely, the term *cued switching* refers to the actions by a bilingual speaker in which they switch from one language to another when prompted or cued by another individual, typically a researcher; this most often occurs in experimental research studies on bilingualism. In general terms, the former phenomenon refers to spontaneously produced speech, whereas the latter refers to speech produced from scripted text read aloud, following a distinction made by others (e.g., Piccinini & Arvaniti, 2015). In this dissertation, I will use *language alternation* (also bilingual mode) as an umbrella term to refer to the former two behaviors described above. Throughout this work, I will additionally use the terms cued switching to refer to specific instances where research tasks were used to prompt speakers to a switch. Alternatively, the term code switching will only refer to cases where a language switch is initiated by the speaker in academic research or otherwise.[1]

2.1 Voice onset time

Many languages categorize their stop consonants based on their voice onset time (VOT) realizations, with a subset of these consonants having a substantially (or at the very least quantitatively) longer VOT than the rest (Lisker & Abramson, 1964; Keating, 1984). Such is the case with English and Spanish, although specific VOT values vary across the two languages. For instance, even though both Spanish and English have voiced and voiceless coronal stops (i.e., /d t/), the realization of these sounds differs between the two languages. While Spanish has prevoiced [d] and short-lag [t] stops, English has short-lag [d] and long-lag [t] plosives. In other words, glottal vibrations in the productions of /d/ begin prior to the

[1] Note that *language alternation* refers to the practice of alternating languages on cue or at will, whereas a *language switch* refers to a single instance of language alternation, regardless of context (i.e., cued switching or code switching).

stop burst in Spanish speech and immediately after the burst in English speech. Meanwhile, voicing during the production of /t/ begins immediately after the production of the stop consonant in Spanish speech and after some delay in English speech–a delay that is filled with aspiration or turbulent airflow prior to the onset of the following phonetic segment. By and large, the voiced stops (i.e., /b d g/) are produced with prevoicing in Spanish and with a short-lag in English, and the voiceless stops (i.e., /p t k/) are produced with a short-lag in Spanish and a long-lag in English, albeit with some variation as is natural in speech production. Note that there is a place of articulation difference between coronal stops in Spanish and English. Spanish has dental /t̪ d̪/, whereas English has alveolar /t d/ (Hualde, 2005; Hammond, 1999). I will use the term *coronal* stops, generally, to refer to these stops in both languages, and I will refer to the specific place of articulation only when a distinction is important to make.[2]

However, the codification of VOT categories for stop consonants becomes muddled when we compare linguistic profiles. Reportedly, VOT productions can differ between monolingual and bilingual speakers. For example, Flege and Eefting (1987) report that, on the one hand, adult Spanish monolinguals can produce prevoiced stops with an average VOT of -82 to -68 milliseconds (ms) and short-lag stops between 18 to 39 ms. On the other hand, the recorded VOT averages for prevoiced consonant stops by adult Spanish-English early bilinguals is -75 to -64 ms; average short-lag VOT values are 17 to 31 ms for this same group. While the differences are not very wide for Spanish data, English data display a starker contrast. Adult English monolinguals produce short-lag stops with an average VOT of 18 to 30 ms and long-lag stops with an average of 78 to 94 ms. For the adult Spanish-English early bilinguals, the average short-lag productions vary between 18 to 35 ms and 57 to 75 ms for long-lag stops (Flege & Eefting, 1987, pp. 74–76).

The definition of *phonetic convergence* that this study uses derives directly from the literature in linguistics. In general terms, convergence is "an increase in similarity among linguistic components... *accumulating common ground, or alignment*" (emphasis in original, Pardo, 2006, p. 2383). In bilingual phonetics, this phenomenon is represented by an increase in similarity between the acoustic properties of comparable or parallel sound segments across two languages. In other words, a production instance of a sound in a given language displays acoustic properties that are more similar to the properties found in the other language. For instance, with regard to VOT, Spanish /p/ is said to be converged with English /p/ if it is produced with longer VOT, approximating typical measurements in English; thus, Spanish /p/ can be said to be more English-like. Convergence between two languages may occur when a segment in only one language approaches the acoustic properties typical of the other language (i.e., unidirectional convergence) or when the segments in both languages shift towards each other (i.e., bidirectional convergence).

Thus, revisiting the results in Flege and Eefting (1987), we find that this group of speakers produced converged values (i.e., values that are more similar to the categories in the opposite

[2]The diacritic specifying a dental production / ̪ / will not be used in this dissertation beyond this point. A dental constriction of a coronal stop will only be indicated in prose hereafter.

language) in both languages, albeit more marked in English speech, in particular in the voiceless category. This phenomenon–converged VOT productions in the speech of bilingual or multilingual speakers–has been referred to as "compromise categories" (Flege, 1995). In addition to the group of adult Spanish-English early bilinguals reported above, this linguistic pattern (i.e., convergence of VOT productions among bilingual speakers) is also attested in other populations with different linguistic profiles. For instance, convergence has also been found among speakers of different age groups (e.g., Flege & Eefting, 1987; Khattab, 2000), speakers with a different L2 age of acquisition (e.g., Thornburgh & Ryalls, 1998), and speakers of different languages or language pairs (e.g., Sundara et al., 2006; S. A. S. Lee & Iverson, 2012).

Moreover, this phenomenon is not static and is not found only in monolingual speech (by bilingual and multilingual speakers). For instance, in a similar fashion to Flege and Eefting (1987), Amengual (2021) reports that English-Japanese and Japanese-English bilinguals as well as Spanish-English-Japanese trilinguals are able to produce distinguishable VOT values in each of their languages, maintaining distinct phonological categories for each language. However, these categories are subject to dynamic cross–linguistic influence. When the researcher elicited a bilingual or trilingual mode (i.e., engaging more than one language in communication), both the bilingual and trilingual speakers produced VOT values with some convergence, compared with data collection sessions where only monolingual speech was elicited. Thus, speech mode (monolingual versus bilingual) for multilingual speakers was determined to influence the magnitude of phonetic convergence, highlighting the dynamic nature of phonetic productions by multilingual speakers in monolingual or bi-/multilingual speech.

In the last few decades of the 20th century, however, the study of the phonetics of bilingualism focused on the monolingual productions of multilingual speakers.[3] Grosjean and Miller (1994) first reported an experimental production study on language alternation in which they analyzed the VOT distributions in the production of /k/ surrounding prompted language switches between French and English and compared them to monolingual productions of the same speakers. The researchers reported that there were no differences in voicing of monolingual and code-switched productions and posited that there were immediate and complete switches from one language system to the other. However, some aspects of the study design raise some questions about the validity of this conclusion: the study design allowed participants to rehearse the sentences prior to the production task, and while switch stimuli included all three stop consonants, surrounding stimuli included only velar stops. In furthermore criticizing its methodology, Bullock et al. (2006) argue that in Grosjean and Miller (1994), the authors (a) only provided the insertion of a single word within an entire utterance in another language, which may be interpreted as a borrowing by some speakers, resulting in the integration of the words into the lexicon of the other language and the uti-

[3]To my knowledge, the earliest reported experimental studies on phonetics/phonology revolving around language alternation were interested in phonetic categorization and categorical perception (Bürki-Cohen et al., 1989), as well as acoustic cues that signaled to a listener which language was produced in the stimuli, through gating methods (Grosjean, 1980).

lization of the phonetic characteristics typical of that language; and (b) have "a directionality assumption that the language that you start in affects the language that you switch into, i.e., that the 'base' affects the 'guest' language" (Bullock et al., 2006, p. 10). Finally, Bullock and Toribio (2009c) note that the stimuli included cross-linguistic homophones, which may have led the speakers to use extremely careful speech, accentuating an acoustic contrast that may not be typical of natural speech.

It was not until a decade later that researchers explored VOT productions more methodically. Initially, research studies took inspiration from the methodology of Grosjean and Miller (1994), while strengthening the systematic generation of research stimuli and investigation of the dependent variable. Considering the comparable methodologies, at first, researchers reported relatively consistent results. For example, Toribio et al. (2005), in a read-aloud task of code-switched Spanish-English sentences, reported that a convergence effect was present near switch sites in the productions of both languages, where the stops that immediately followed a switch approximated VOT in the previous language more so than consonants further away from the switch point. Another study by Bullock et al. (2006) with a similar design and target population found asymmetrical results, with unidirectional VOT convergence of English towards Spanish. The effect was found to be more pronounced in anticipation of a switch from English towards Spanish, relative to perseverative effects in the reverse direction.

Subsequent bilingualism studies have corroborated the unidirectional patterns reported by Bullock and colleagues–at least when analyzing Spanish and English. For instance, Balukas and Koops (2015) conducted a study of VOT productions in monolingual and spontaneous code-switched speech. The authors found that the VOT values for stop consonants in Spanish and English were consistent with previous research on monolingual speakers of these languages. However, when the languages were mixed within the same utterance, the VOT values showed significant variability, with some instances of code switching resulting in convergence of VOT. More specifically, the study concluded that early Spanish-L1-English-L2 bilinguals had a convergence effect only in their L2, not their L1. Finally, Piccinini and Arvaniti (2015) conducted another study on spontaneous code switching in peer conversations by Spanish-English early bilinguals who are English-dominant (L2). Two tasks were conducted, one without distractions and another in which the participants had to complete a puzzle while conversing. The researchers found a shortening of VOT in voiceless stops in both languages in a tandem shift effect, although the effect was greater in the speakers' L2 (or dominant language).

Admittedly, language dominance and acquisition order are somewhat complicated predictors of VOT productions in language alternation. Similar to the findings reported for Spanish-English bilinguals, Antoniou et al. (2011) indicate that Greek-English bilinguals also have an unidirectional convergence effect in which the speakers' L2 (i.e., English) shows more L1-like (i.e., Greek-like) VOT productions during a read-aloud task of code-switched utterances that included syllables with bilabial or alveolar stops. Along the same lines, in a picture-naming language switching task, Olson (2013) found that there was convergence in VOT productions after a language switch for Spanish-English bilinguals. Yet, the unidirectional effect was seen in either Spanish or English, whichever was the speaker's dominant language. Contrastively, in a code-switched passage reading task, Olson (2016) found that English-dominant speakers had a convergence effect in their dominant language (i.e., their L1), whereas Spanish-dominant participants displayed a bidirectional effect.

In short, we find that when there is unidirectional convergence, English is the language that moves towards the other language, and when there is bidirectional convergence, English displays a greater shift towards the center. As such, English, the long-lag VOT language, appears to be more vulnerable to undergo phonetic convergence, compared to other short-lag languages (e.g., Spanish and Greek) (for a discussion on parallel results from other long-lag languages, see Olson, in press). In explaining this asymmetrical pattern, Bullock and Toribio (2009c) discuss the flexibility provided by "phonetic latitude" of long-lag languages. That is to say, the range of acceptable VOT productions between short- and long-lag languages allows for more movements. For example, "the English voicing contrast permits a great deal of latitude in the expression of voiceless stops" (Bullock & Toribio, 2009c, p. 204). Olson (in press) further develops this point by saying that:

> Asymmetrical interference relates to the range of acceptable productions within a given phonetic category (e.g., phonetic latitude). Notably, short-lag voiceless stops have a much smaller range of acceptable VOTs (approximately 30 ms) than long-lag voiceless stops (up to 100 ms). Bilinguals may 'permit' a degree of interference, so long as the production stays within the expected category variability. Long-lag voiceless stops, with greater acceptable ranges, are most likely to show evidence of cross-linguistic interference. In contrast, short-lag voiceless stops, with a relatively reduced range, are less likely to evidence cross-linguistic interference and do so only when monolingual productions leave 'room' (i.e., are sufficiently short) to allow for convergence. (Olson, in press, pp. 21–22)

Nonetheless, these results are not undisputed. At the core of this debate is the question of whether or not experimental tasks, typically cued switching task, can be compared with code switching. For example, Fricke et al. (2016) state that (within the cued switching paradigm) "[l]anguage switching studies typically examine single word production, where the target language can vary at random and is determined by the experimenter [...] During code switching, by contrast, grammatical planning mechanisms are fully engaged, the language of all lexical and morphosyntactic elements is fully under the control of the speaker, and in a normal conversational setting, production must additionally be coordinated with comprehension of the interlocutor's speech" (2016, p. 2). Put differently, Piccinini and Arvaniti summarize these concerns as follows: "The use of different tasks in combination with different populations of bilinguals is likely to have further compounded discrepancies among studies" (2015, p. 122). While it is true that the mechanisms engaged during prompted switches and natural code switching may differ, Piccinini and Arvaniti conducted a comparative analysis of VOT values between stops near or far away from switch points, and found similar results to the ones reported from experimental tasks. The authors, thus, concluded that "code-switching effects reported in earlier studies are not an experimental artifact but apply in spontaneous speech as well" (Piccinini & Arvaniti, 2015, p. 131). However, no study has been conducted to compare bilingual speech behaviors within the same population across speech modes and tasks (e.g., cued switching versus code switching tasks), leaving open the question of whether research task can influence the magnitude of production results in this area.

2.2 Cognitive and communicative factors

When examining the methodology in Piccinini and Arvaniti (2015), we are reminded that the authors used a casual peer conversation (naturalistic code switching) and a peer conversation with a distraction (a puzzle; experimental code switching), with both tasks yielding comparable results regarding convergence: tandem shift towards lower VOT productions in both languages (Piccinini & Arvaniti, 2015).[4] This notion is supported by the results of the study by Balukas and Koops (2015), which analyzes VOT productions in naturalistic code switching; the authors conclude that "the fact that we find code-switching effects on the VOT duration of English-language tokens suggests that such effects do indeed occur in spontaneous code-switching, and not only in laboratory-induced code-switching" (2015, p. 438).

However, phonetic adjustments during speech production (including patterns observed in language alternation activities) do not occur in a vacuum. There is a range of factors that mediate the linguistic output of speakers beyond settings (i.e., experimental or naturalistic). Some of these factors may relate to the cognitive processing mechanisms that are involved in speech processing of naturalistic versus experimental tasks, while others may relate to the cognitive mechanisms involved in the social interactions that occur between interlocutors in more naturalistic settings versus the experimental tasks that are often devoid of social interaction (e.g., in read aloud tasks). Cognitive factors are a huge influence on linguistic production, especially as it relates to the several studies described above–in this section and the previous one. For instance, the code-switched peer conversations without and with distractors can have cognitive effects that affect the acoustic productions of the speakers by adding to the speakers' cognitive overload. It is difficult to determine which research task has a greater cognitive overload on speakers during experimental linguistic research. However, in general, puzzle completion may require more cognitive processing than a casual interview, as it involves problem-solving and spatial reasoning.

There is some evidence to suggest that completing puzzles can engage cognitive processes such as attention, working memory, and executive function (e.g., Diamond, 2013; Ganley et al., 2014). While these processes are also involved in linguistic processing, the level of engagement may depend on the specific linguistic task being performed. On the other hand, a casual interview may be less cognitively demanding in terms of problem-solving and spatial reasoning, but it still requires attention and working memory to process and respond to

[4] Note that in both tasks, speakers were able to switch languages at will, making both activities part of the code switching paradigm.

questions in real-time. The level of cognitive demand may also depend on the complexity of the questions and the individual's familiarity with the topic (e.g., Tyler & Feldman, 2004). All in all, both puzzle completion and casual interviews can engage cognitive processes during linguistic research, and the degree of cognitive overload may vary depending on the specific tasks and individuals involved. Nonetheless, the added distractions provided by the puzzle task very often increase the cognitive overload, affecting the amount of attention the speakers can place on other tasks such as monitoring their own speech.

Expanding on the previous points on the cognitive demands associated with some speech activities compared to others, among cognitive influences, we find what Labov (2011) calls the "attention to speech" factor. This theory explains the ways in which people adapt their speech according to the social context and the degree of attention paid to them by their interlocutors. According to Labov, attention to speech is a crucial factor in determining the linguistic variation that exists at the micro (within a given conversation) and macro (a language community) level. In short, Labov summarizes the influence of attention to speech in the following way:

> The factor "Attention to speech" is realized by stylistic ratings on the following well-known scale, used to classify the degree of formality within a sociolinguistic interview:
>
> 1. casual speech
> 2. careful speech
> 3. group
> 4. elicited
> 5. reading text
> 6. word lists
> 7. minimal pairs

In other words, as the scale number increases (i.e., as we move down the list), speakers are expected to be more attentive to their speech patterns, especially with shorter utterances such as single words, which allows for more close monitoring of individual sound segments. For example, when describing the production of the approximant /ɹ/ among English speakers in New York City across word list and minimal pair tasks, Labov found that "(r) is pronounced in the formal contexts of a word list, but it does not receive the full attention of the reader. But in minimal pairs such as *dock* and *dark*, *guard* and *god*, *source* and *sauce*, *bared* and *bad*, (r) is the sole differentiating element, and it therefore receives maximum attention" (Labov, 2006, p. 63).

In general, the difference in allophonic production (i.e., the production or omission of /ɹ/) was related to the perceived degree of attention paid to the speaker by the listener, following the scale displayed above. That is, Labov argued that, through internal perceptual processes (i.e., psycholinguistically), when speakers perceive that their interlocutors are paying close attention to their speech, they are more likely to increase the rate of usage of the standard, prestige variant of a linguistic variable, compare to contexts with less perceived external monitoring. This is because speakers perceive that their status and credibility are being evaluated by the listener, and they want to present themselves in the best possible light. In contrast, when speakers perceive that their interlocutors are not paying close attention to their speech, they are more likely to increase the rate of usage of non-standard variants of a linguistic variable. This is because speakers do not feel the need to conform to the norms of the standard language in this given context.

Moreover, it must be emphasized that attention to speech cannot be stripped from its sociocultural influences. Labov's theory of attention to speech has been influential in the field of sociolinguistics and has been used to explain a wide range of linguistic phenomena. For example, researchers have found that speakers are more likely to produce a higher rate of standard forms of a linguistic variable in formal settings where they perceive that their speech is being closely monitored (Eckert, 2012). Similarly, speakers are more likely to increase the rate of non-standard forms of a linguistic variable in informal settings, such as casual conversations with friends, where they perceive that their speech is not being evaluated in the same way (Rickford & McNair-Knox, 1994). One of the strengths of the attention to speech theory is its emphasis on the social context of linguistic variation. Labov argued that linguistic variation is not random, but rather is shaped by social factors such as class, race, and gender. He demonstrated that linguistic variation is not simply a matter of individual choice or preference, but is instead a reflection of broader social structures and power relations.

However, Labov's theory has also been criticized for its emphasis on the role of attention in shaping linguistic variation. Some researchers have argued that attention may not be the only factor that influences linguistic variation, and that other social and psychological factors may also be important (e.g., Coupland, 2007). Coupland (2007), for instance, argues that Labov's attention to speech principle provides a good account of intra-speaker variation but fails to encompass the effect of social interactions in speech variation. Instead, the author proposes a shift towards the *audience design* paradigm (Bell, 1984) and *accommodation theory* (Giles, 1973; Giles & Ogay, 2007), which better account for intra- and -inter-speaker variation.

Communication Accommodation Theory (CAT) suggests that people adjust their communication style in order to facilitate understanding and build rapport with others. This theory has been applied to the study of language accommodation, which refers to the way in which speakers adjust their language use in response to their audience. According to Giles and colleagues, CAT "provides a wide-ranging framework aimed at predicting and explaining many of the adjustments individuals make to create, maintain, or decrease social distance in interaction" (Giles & Ogay, 2007, p. 297). In short, this theory is a tool for understanding how people communicate with each other in many contexts, as it posits explanations for a wide range of communication phenomena, including:

1. Language variation: The way that people speak can vary depending on who they are talking to. For example, people may use more formal language when talking to someone they do not know well, and more informal language when talking to a friend.

2. Inter-group communication: The way that people from different groups communicate with each other can be influenced by their perceptions of each other's group memberships. For example, people from different ethnic groups may accommodate their speech to each other in order to reduce social distance or to assert their own identities.

3. Inter-personal relationships: The way that people communicate with each other in close relationships can be influenced by their goals for the relationship. For example, people who are in a romantic relationship may accommodate their speech to each other in order to show their love and affection.

Similarly, the audience design paradigm is a theoretical framework that focuses on how speakers adapt their language use to suit the communicative needs of their audience. As Bell postulates:

> [Audience design] assumes that persons respond mainly to other persons, that speakers take most account of hearers in designing their talk. The speaker is first person, primary participant at the moment of speech, qualitatively apart from other interlocutors. The first person's characteristics account for speech differences between speakers. However, speakers design their style for their audience. Differences within the speech of a single speaker are accountable as the influence of the second person and some third persons, who together compose the audience to a speaker's utterances. (1984, p. 159)

In other words, the audience design framework proposes that speakers make a series of strategic choices when adapting their language use to suit their audience. These choices include lexical choices, grammatical structures, and discourse organization. For example, a speaker may use simpler vocabulary and shorter sentences when communicating with a child, or they may use more complex vocabulary and longer sentences when communicating with a highly educated adult. The audience roles can be distinguished and ranked based on whether the individuals are known, ratified, or directly addressed by the speaker. The primary audience member is the second person, while there may be additional third persons present but not directly addressed. Known and ratified individuals in the group are referred to as auditors, while third parties whose participation is acknowledged by the speaker but not ratified are called overhearers. Finally, individuals whose presence is unknown are termed eavesdroppers, regardless of whether their listening is intentional or accidental.

All in all, CAT suggests that individuals adjust their language use in response to their conversational partner, using strategies of convergence, divergence, or maintenance. This theory has important implications for intergroup communication and highlights the importance of understanding the factors that influence language accommodation in order to promote effective communication. On the other hand, the audience design paradigm suggests that speakers adapt their language use to suit the communicative needs of their audience, making strategic choices in order to facilitate understanding and build rapport.

Among the literature on sociolinguistic aspects of code switching, language choice has been a key consideration, which has in part been studied under CAT:

> [L]anguage choices cannot be explained sufficiently by entirely referring to the situational factors and categories of a society. Explanations for language choice have to take aspects of interpersonal relationship and the psychological forces which influence an individual's behavior into consideration (Appel & Muysken, 1987). Thus some researchers consider code switching within communication accommodation theory (CAT). (Nguyen, 2014)

However, as Ng and He (2004) indicate, code switching studies "under the rubric of CAT or its predecessor, the Speech Accommodation Theory" are rare; "[n]onetheless, some of the general theoretical notions [of CAT] are applicable to [code switching] between languages (Ng & He, 2004, p. 30). Just like in accommodation of accent or style, accommodation in code switching will result in either divergence or convergence of the interlocutors' speech, telling us more about the social association between both parties: "Language divergence is a marker of social disapproval, distancing, or disidentification, and is often perceived as such for the lack of solidarity and respect that it conveys. Language convergence, by contrast, conveys social approval of and identification with the addressee" (Ng & He, 2004, pp. 30–31). While sociolinguistic studies are typically interested in language choice itself or switches at the lexical level, we can extend this principle to the phonetic level in language alternation. That is, in trying to display a greater or lesser social distance, interlocutors may rely on the phonetic segments of their speech and converge or diverge their acoustic properties in order to display social distance while alternating languages, just like they would in monolingual speech.

Bell (1984) also addresses the matter of language choice within the audience design framework by accommodating to the audience's communicative needs to initiate and maintain a social interaction, especially as it pertains to bilingual contexts in which a monolingual individual is introduced. For instance, Bell says that "[w]ith monolingualism increasing among household members, accommodating to a single monolingual auditor reduces geometrically the occasions for using the alternative language," given that "[a] speaker in a bilingual situation is ... bound to take account of the audience's linguistic repertoire" (Bell, 1984, p. 176). As such, to decrease the possibility of exclusion of one or more (monolingual) audience members, bilingual speakers are expected, under the audience design framework, to resort to the language shared by the entire audience because:

> The sharper the linguistic differences between codes, the larger the issue of intelligibility looms, the stronger are the pressures to accommodate to the audience, and hence the greater the influence of peripheral members on the speaker. Use of a language which is unintelligible to any interlocutor defines that person out of the audience. (Bell, 1984, p. 176)

Nonetheless, sociolinguistic studies on code switching under the audience design framework focus on aspects outside of the speaker's own desires or motivations, as Wei expresses: "To some extent, ... the audience design model ... focus[es] on 'external' factors affecting language choices" (Wei, 2005, p. 376). One such example is the study by Mugari (2014) who investigates singers' code switching choice in song lyrics. The author argues that the decision to engage in such linguistic behavior is to appeal to a young audience that frequently engages in this practice. However, while this framework might help us better understand language choice in code switching in bilingual situations or style choice in other situations, typically contexts with larger audiences are required in order to fully understand the social negotiations taking place; this framework is thus very limited in experimental settings where only the research subject and researcher are in communication.

By contrast, according to the attention to speech principle, we expect speakers to rely on self-monitoring to adjust their productions at the phonetic level, specifically, in order to shape the image they want to present to the interlocutor(s). Moreover, this principle applies to natural contexts (e.g., two or more friends interacting in school) or experimental settings (e.g., a research subject completing a data collection setting in a laboratory). As such, the attention to speech factor appears to be better suited to investigate and explain the phenomena observed in language alternation behavior in naturalistic and experimental settings, especially when investigating productions at the phonetic level. In addition, this principle combines the elements of social interactions with the usage and expenditure of cognitive resources associated with or required by each distinct task during the interaction. Therefore, the attention to speech factor seems better capable of analyzing subjects' speech patterns and cognitive costs in academic research–naturalistic or experimental.

2.3 Linguistic factors

Several linguistic factors have been identified as influencing VOT measurements in speech.[5] Cho and Ladefoged (1999) investigated cross-linguistic patterns in a set of 18 languages from distinct language families. The authors found that there is a great deal of variation in VOT across languages. However, they also found some general patterns. The researchers argue that the variation in VOT is due to a combination of phonetic and phonological factors. Phonetic factors, specifically, include the place of articulation (POA) of the stop consonant, the voicing contrast in the language, and the speaking rate.

Expanding on these linguistic factors, Cho and Ladefoged (1999) report that, first, voiceless stops tend to have longer VOT than voiced stops, and this difference is larger for stops with more retracted places of articulation (e.g., velar stops) than for stops with more advanced POA (e.g., bilabial stops). Second, the difference in VOT between voiceless and

[5]In this dissertation, I define linguistic factors as the various elements and aspects of language itself that can influence or affect communication, understanding, and interpretation. In particular, this study is interested in speech features at the segmental level (i.e., place of articulation) and speech characteristics at the suprasegmental level (i.e., speaking rate or speech rate).

voiced stops is larger for stops in syllable-initial position than for stops in syllable-final position. Third, the difference in VOT between voiceless and voiced stops is larger for stops in languages that have a phonemic contrast between voiceless and voiced stops than for stops in languages that do not have this contrast.

The POA differences have been long-established by previous research (e.g., Lisker & Abramson, 1964; Volaitis & Miller, 1992). Regarding the place-dependent VOT differences, Cho and Ladefoged (1999) allude to previously reported explanations for these differences that range from aerodynamics, speed of articulatory movement, and differences in the mass of articulators. For example, as suggested by other phoneticians, the supraglottal cavity–the space between the glottis and the constriction site during stop production–may dictate VOT production patterns. There are two important details to remember: first, the supraglottal cavity in more posterior constrictions has a smaller volume, and second, the cavity in front of the constriction is bigger for more posterior constrictions. For posterior stops (e.g., velar stops), the smaller supraglottal spaces may lead to the generation of higher air pressure. Thus, as the air rushes through the vocal tract to stabilize air pressure after the stop burst, (1) "it will take a longer time for the pressure behind the closure to fall and allow an adequate transglottal pressure for the initiation of the vocal fold vibration" (p. 210). Or (2) the greater cavity in front of the constriction, irrespective of air pressure amount, may results in a slower transglottal pressure stabilization due to the higher volume of air that needs to be pushed out of the vocal track, allowing for longer voicing.

In addition, articulation constraints have been suggested as a potential cause of the cross-linguistic place-dependence of VOT productions patterns. For instance, Hardcastle (1973) suggests that the movement of the articulators themselves, and in particular articulators' mass, may influence the place-dependent VOT differences; he states that articulators such as the tip of the tongue and the lips are faster than the back of the tongue, potentially as a result of their general mass making them more nimble. Moreover, with regard to the extend of articulatory contact, Cho and Ladefoged discuss Stevens's (1999) aerodynamic explanation:

> Stevens (1999) provides an aerodynamic explanation for these differences. His main point is that the rate of change in intraoral pressure following the release depends on the rate of increase in cross-sectional area at the constriction. This is significantly different for different places of articulation, primarily due to the differences in the extent of articulatory contact. When there is a long narrow constriction the Bernoulli effect causes the articulators forming the constriction to be sucked together. Because the velar stop has extensive contact between the tongue body and the palate, there is a larger Bernoulli force so that the change in cross-sectional area is relatively slow compared with that for the bilabial or alveolar stops. Consequently, the decrease in intraoral pressure after the closure is gradual for the velar and rapid for the bilabial. (Cho & Ladefoged, 1999, p. 211)

Furthermore, for aspirated voiceless stops (e.g., English voiceless stops), Stevens (1999) proposes that the glottis opening during aspiration decreases more slowly in velar than alveolar and bilabial stops due to a slower intraoral pressure drop for velar stops, leading to faster adduction movements of the vocal folds for anterior consonants and resulting in delayed VOT for more posterior stops.

Lastly, according Maddieson (1997), anterior stop consonants tend to have a longer closure duration, which is likely due to the difference in tolerance of intraoral air pressure provided by the physiology; for alveolar, and especially for bilabial sounds, the flexibility of the cheek tissue allow for an increase in cavity volume. As the author states, "if the consonant gesture is timed in some way that directly relates to the time of the pressure peak, then broadly speaking, the further back in the oral cavity a stop closure is formed, the shorter its acoustic closure duration will be" (Maddieson, 1997, p. 630). This would lead to a negative correlation between closure duration and VOT measurement. Alternatively, given reports by Weismer (1980) that bilabial and velar stops have comparable intervals between the onset of stop closure and VOT, Maddieson (1997) proposes that "[t]here is an abduction-adduction cycle of the vocal cords for voiceless stops which is longer in duration than the closure and has a constant time course, anchored to the onset of closure" (p. 621).

Ultimately, Cho and Ladefoged (1999)

Speech rate (also speaking rate) is another crucial factor that affects the production of speech sounds. Generally speech rate refers to the speed at which language is produced. Different quantitative metrics have been used to identify speech rate in studies analyzing the influence of speech rate on various speech features, most commonly phonemes per second and/or syllables per second (Roach, 1998; Hewlett & Rendall, 1998; Verhoeven et al., 2004; Kahng, 2014; Tendera et al., 2019), although preceding vowel duration has been used before (as a proxy of speech rate) (e.g., Abdelli-Beruh, 2004). There is evidence suggesting that different speakers may rely on distinct mechanisms to manipulate speech rate (e.g., segment duration of specific sound categories: Crystal and House, 1982; articulatory transition speed: Kuehn and Moll, 1976; degree of coarticulation: Matthies et al., 2001; etc.). Nonetheless, it is clear that changes in speech rate are not merely the incorporation or extension of pauses in otherwise normal speech (J. L. Miller et al., 1984).

Studies on speech rate and VOT duration have demonstrated that this general pattern holds: as speech rate increases, articulators move faster and segmental duration decreases; as a results, VOT measurements decrease with faster speech rates (Kessinger & Blumstein, 1997; J. L. Miller et al., 1986; Nagao & de Jong, 2007). For instance, Theodore et al. (2009), in trying to identify whether place of articulation and speech rate are talker-specific predictors of VOT productions, analyze these measurements in word-initial voiceless stops in English in a range of speech rates. The authors found that as speech rate decreased, VOT values increased for all subjects, but the magnitude of the increase varied across speakers, leading the researchers to conclude that this contextual influence is talker-specific. Nevertheless, while the authors also reported place-dependent VOT values (i.e., bilabial $<$ alveolar $<$ velar), there was no statistical significant variability among the three POA's across all speakers, indicating that the effect of speech rate on VOT remains constant across the different places. Or in the authors' words, "unlike speaking rate, the contextual influence of place of articulation on VOT appears not to be talker-specific" (Theodore et al., 2009, p. 3980).

In addressing the implications of their findings, Theodore et al. remind us that "listeners retain talker-specific acoustic-phonetic information in memory (e.g., Goldinger, 1998) and that familiarity with a particular talker's speech can facilitate word recognition (e.g., Nygaard et al., 1994)" (p. 3980). Thus, the authors continue their discussion with a theoretical explanations for the process in which listeners become familiar with a speaker's linguistic patterns, as it pertains to VOT productions:

> This finding raises the possibility that listeners may customize stop voicing categories based on individual talkers' characteristic VOTs. However, given contextual influences on VOT, listeners would need to consider a talker's characteristic VOTs not in an absolute manner, but with respect to context. Indeed, it is well established that at a general level, listeners do process VOT in relation to numerous contextual factors, including both speaking rate and place of articulation. These contextual influences systematically affect both the boundaries between phonetic categories and the best exemplars of a given phonetic category (e.g., Lisker and Abramson, 1970; J. L. Miller and Volaitis, 1989; Summerfield, 1981; Volaitis and Miller, 1992). We do not yet know whether such context-dependent processing is tuned to the speech of individual talkers, but the results of the current experiments place constraints on the type of exposure listeners might require for such perceptual tuning. (Theodore et al., 2009, p. 3980)

More specifically, the authors argue that hearing a speaker's VOT productions for a given stop at a given rate cannot provide optimal information to deduce the speaker's VOT productions for said stop at a novel speaking rate. Given that modifications of VOT as a function of speech rate remain relatively constant across places of articulations, a listener must, then, only decipher the VOT adjustment slope for a single stop consonant and make generalizations to a speaker's speech rate contextual influences overall.

2.4 Social factors

Over the last several decades, sociolinguistics has identified countless ways in which social factors influence speakers' speech production and perception. Starting with the foundational department store study in the mid 1960's on the production of /ɹ/ in New York City by Labov (2006), many sociolinguists have now followed with additional studies to cover other language communities, linguistic features, and social contexts. The next two sections provide background from the literature that covers socio-cultural factors that influence linguistic behaviors. While the first section will cover speaker-internal factors, the second section will cover environmental variables.[6]

[6]While speaker-internal variables deal with factors that are inherent to the speaker themselves (e.g., age of language acquisition or linguistic attitudes), environmental factors are those that may not be (completely)

Speaker-internal social factors

In addition to linguistic profile, as discussed above, several other speaker-specific social factors have been reported to influence speakers' VOT. Namely, social factors such as gender, age of acquisition of an L2 among bilingual speakers, and linguistic attitudes have been shown to impact VOT productions. With respect to gender, studies have found that males tend to produce shorter VOT values than females for certain stop consonants, such as coronal stops (i.e., /t d/) (Swartz, 1992). For this study, the mean difference among the voiced tokens were minimal, with females producing a VOT that was approximately 4.3 ms longer (F: 17.21 ms; M: 12.87). However, among the voiceless consonants, the distinction has more drastic. Females produced VOT measurements about 17.6 ms longer (F: 82.62 ms; M: 64.96 ms). While males were reported as having a faster speech rate, which has been associated with shorter VOT, the researcher discarded speaking rate as a contributor to the observed gender-based linguistic difference, given that no statistically significant correlation was identified. While the author attributes the gender difference for /d/ on males' higher usage of prevoicing than females, he acknowledged that such explanation fails to account for the even greater difference observed on the voiceless sound /t/.

Moreover, Robb et al. (2005) conducted another study in which they analyzed data on short-lag voiced and long-lag voiceless stops /p b t d k g/ from CV monosyllabic tokens in a read-aloud task that collected productions from 60 participants, 30 males and 30 females, who were recorded speaking in two different settings: a laboratory setting (inside a sound booth) and a non-laboratory setting (outside of a sound booth). The results showed that there were significant gender differences in VOT only in voiceless tokens, with females producing longer VOTs than males, unlike Swartz (1992) (and others such as Ryalls et al., 1997), who found a difference in both voiced and voiceless stops. Robb et al. also found significant environmental setting differences in VOT, with participants producing longer VOTs in the laboratory setting than in the non-laboratory setting. The authors determined there were no CV duration differences (i.e., speaking rate) in the laboratory setting between the two groups, excluding speech rate as a factor influencing VOT differences between the two groups. However, a difference was observed in the non-laboratory setting. Nonetheless, the researchers argued that "[t]he finding of a consistent pattern of VOT production among men and women across both environmental settings, in spite of variable production of CV durations, would suggest that speaking rate had no clear effect on VOT production" (Robb et al., 2005, p. 131). This pattern is corroborated by the findings presented by Allen et al. (2003), who found a gender-based differences even when controlling for speaking rate.

Robb et al. (2005) suggested two potential reasons for the attested gender-dependent linguistic patterns on VOT productions. First, the authors attributed the production distinctions to anatomical differences between males and females, expressing that their findings "[lend] additional support for a biological influence on VOT" (p. 131). Second, they asserted that female participants' persistence in producing longer VOT values for (primarily) voiceless stops regardless of setting "[lends] support to the suggestion that females tend to use

under the control of the speaker (e.g., the language used in school or the workplace).

more carefully articulated speech in experimental settings compared with males" (p. 131), a pattern that has previously been reported in other sociophonetic studies (Byrd, 1994; Henton & Bladon, 1985).

Age of acquisition of an L2 is another social factor that mediates VOT productions.[7] While age, itself, has been studied in regard to potential differential VOT patterns, in general, comparative studies have not found average VOT differences across age groups (e.g. Sweeting & Baken, 1982; Neiman et al., 1983). Nevertheless, Sweeting and Baken (1982) do report that older speakers have a propensity to display more variability in their speech, as illustrated by their wider standard deviations. However, when it comes to bilingualism, age of acquisition of an L2 has a much greater effect on VOT categorization and production. For instance, Flege (1991) investigates the production of VOT in a group of Spanish-English bilinguals who learned English, their L2, during childhood or adulthood, to make an age-of-acquisition comparison. The author compared their Spanish and English speech to the productions of monolingual speakers in each language. The researcher found that the VOT values of the bilinguals who learned English as children were significantly more similar to those of native English speakers than the VOT values of the bilinguals who learned English as adults. Flege suggests that the differences in VOT production between the early and late L2 learners may be due to the fact that the early L2 learners had more time to acquire the L2 phonological system and may have also had more exposure to L2 speech, which could have helped them reach L2 VOT production that are comparable to the monolingual group. In short, this study suggests that the age of acquisition of an L2 can have a significant impact on the acquisition of L2 phonological features and underlies the importance of analyzing (or at the very least controlling for) the age of acquisition of an L2 when exploring the phonological systems of bilingual populations.

As discussed in an earlier section, the aforementioned results among bilingual speakers as it relates to age of acquisition have been replicated in countless studies (e.g., Thornburgh & Ryalls, 1998; Khattab, 2000; Sundara et al., 2006).[8] However, many of these studies focused on, analyzed, and compared the linguistic productions of advanced, early L2 speakers with those of intermediate, late L2 speakers. Stölten et al. (2015) reports results on the VOT productions of voiceless stops of Spanish-L1 speakers who were early and late near-native speakers of Swedish as an L2. In this study, the researchers identified two groups of L2 speakers whose Swedish speech was rated by native speakers and deemed *native speakers* of Swedish, regardless of age of acquisition. The data for that study included three POA's (i.e., bilabial, coronal, velar), and when VOT values were controlled for speech rate (i.e., turning the raw VOT value into a percentage of the word duration), the researchers still found an age effect on the production of VOT when comparing the two groups. The authors concluded that:

[7] In this study, I categorize age of acquisition as a speaker-internal social factor given that this characteristic is inherent to the speaker themselves and cannot be manipulated by other contexts or individuals.

[8] The reader is referred to Flege (1995) for a thorough summary of studies on VOT productions–among other features–as it relates to age of acquisition.

> [E]ven if some L2 learners are sometimes perceived as native speakers, nonnativelike features can still be detected when the analysis goes beyond evaluations based on everyday oral communication. However, statistically significant age effects for all three places of stop articulation became apparent only when VOT was related to word duration, thereby neutralizing speaking rate effects on VOT that have been reported in the literature (e.g., Beckman et al., 2011; Kessinger & Blumstein, 1997, 1998; Magloire & Green, 1999). (Stölten et al., 2015, p. 90)

The authors ultimately argued for the consideration of speaking rate in studies comparing early and late bilingual speakers' VOT productions, as seemingly native-like L2 speakers still produce non-native features.

Finally, linguistic attitudes have been explored as drivers for language variation and change. As Appel and Muysken (2005) explain, "Linguistic behaviour and attitudes towards languages in a bilingual society often give further insight into social norms and values" (p. 8). The social evaluations of languages, and by extension the subjective linguistic attitudes towards them (and their speakers) are associated with the identities of the social groups that speak them. These attitudes become more salient in cases where there is a strong relationship between the language and its speakers. Appel and Muysken continue with a description of theoretical approaches to language attitudes, naming the two main vines: the behaviorist view and the mentalist view. Behaviorist research tries to obtain information regarding listeners' responses to the observations of language X in use; Mentalist research, on the other hand, attempts to reach the internal, mental state of the listeners, which may lead to specific behaviors when they observe language X in use. Regarding the latter, the authors describe the two main approaches most commonly used research: the matched-guise technique, which obtains evaluations or ratings the listeners associate with the voice/speaker in the stimulus, and the questionnaire, which directly inquires about the subjects' views of language X, its use, and its speakers.

However, some languages or language varieties are often linked to particular contexts or situations, as is the case of Spanish in the US, which is often associated with a home environment. For instance, Carranza and Ryan (1975) analyze linguistic evaluations of Spanish and English in two contexts: home and school. The evaluations were provided by Mexican-American bilingual speakers (Spanish L1) and Anglo American bilingual speakers (English L1). As the authors report, in general, English was rated more favorably than Spanish in solidarity and status scales. Furthermore, Spanish received higher evaluations in the home context while English received higher evaluations in the school context. The researchers rationalize these findings as a potential internalization of higher value associated with English as a result of living and interacting with others in an English-dominant society that places higher value in the English language.

Related to the behaviorist view, Appel and Muysken (2005) indicate that there are some linguistic behaviors within bilingual communities such as language choice that may reflect linguistic attitudes but are nonetheless determined by a myriad of social factors. At the community level, there are factors that may determine language usage patterns, as is the case in diglossic contexts. Among the person-oriented motivations of language choice, however, the authors name *decision tree* and *accommodation*. In the decision tree model "the speaker is faced with a hierarchical set of binary choices, which can be represented formally as a tree... Factors such as the ethnicity of the interlocutor, the style, the topic of conversation determine which language is finally chosen" (Appel & Muysken, 2005, p. 27). While this model provides descriptive clarity, it lacks rigidity. As the authors express, "[i]n many situations more than one language is possible, often speakers are observed to make choices that are not exactly predicted by the tree model, and the model seems to exclude the use of two languages at the same time in one situation (code switching)" (p. 27). Accordingly, Sankoff (1972), inspired by Gumperz and Hernández-Chávez (1971), proposed modifications to this deterministic model to yield a more interpretive model. That is, in addition to the expected or unmarked choice, Sankoff allows for the inclusion of a marked choice where the speaker may indicate intention, irony, style change, or some other motivator for said unexpected or marked choice. Finally, the second model, accommodation, we have covered above (recall the CAT framework by Giles, 1973). As such, we will not replicate its introduction.

Taking all that in consideration, we can now shift our attention to current linguistic attitudes of Spanish-English bilinguals in the United States (US). L. Miller (2017), for example, analyzed attitudes towards Spanish and English among school-age children from data obtained through attitude questionnaires. The researcher found that the youngest participants, first graders, showed no difference in the way they see Spanish and English, providing comparable ratings for both languages. Remarkably, second, third, and fourth graders started displaying a preference for English through higher positive ratings. However, a distinction was not found in a matched-guise task, suggesting that while children may have a preference for English, they do not appear to have a negative view of Spanish and/or its speakers. Finally, L. Miller reported that a shift in language preference appears to prelude a shift in language dominance, as this dominance shift was found among older children, suggesting that the shift in language preference leads the shift in language dominance.

This pattern appears to carry over into adolescence and adulthood. Galindo (1995) analyzed the language attitudes of Spanish-English bilingual teenagers who had mostly third-generation status, meaning that their parents were born and raised in the US. In their Texas communities, the adolescents were in constant contact with English and Spanish speakers. The data for the study were collected through a language attitude questionnaire that include open- and close-ended questions, to which the researcher applied a qualitative analysis. Galindo found that a majority of the subjects self-identified as English-dominant. Furthermore, the subjects expressed a link between Spanish (or Spanish dominance) with a first-generation immigrant status, an attitude that is leading speakers to prefer English, as foreshadowed by the higher levels of English dominance among the participants.

Another study by Achugar and Pessoa (2009) explored the linguistic attitudes towards Spanish by members of a graduate program in bilingual creative writing in Southwest Texas. Through a discourse analysis, this study inquired about Spanish as spoken in academic and casual settings by bilingual and monolingual speakers. In general, the authors report that members of this academic community have a positive view of Spanish use and bilingualism when spoken in academic settings. However, their views of local varieties of Spanish, especially when spoken by monolingual speakers, were mostly negative. The researchers conclude that power and space are important social concepts that influence linguistic evaluations and attitudes towards language varieties, their usage, and their speakers.

Studies by Toribio (2002) and Anderson and Toribio (2007) also explored Spanish-English bilinguals' language attitudes and attributes towards code switching. First, Toribio (2002), in a case study of four Spanish-English bilingual speakers, reported a generally "low prestige associated with code-switching" (p. 115). Nonetheless, covert prestige drives the practice of this linguistic behavior, as it is a marker of identity by which users are able to signal their ethnicity. However, this linguistic practice "is not an essential trait of US Latino speech." The author adds, "some speakers do not practice code-switching, being far removed from communities in which code-switching is the norm" (p. 151). Other bilinguals refrain from engaging in this linguistic behavior due to the internalized stigma and stereotypes affiliated with this practice, such as the indication of a linguistic deficiency. The wide range of linguistic attitudes towards language alternation reported by Toribio (2002) were replicated in Anderson and Toribio (2007). In the latter study, the researchers analyzed linguistic attitudes more systematically through an experimental study that obtained linguistic ratings of the text of the fairy tale Little Red Riding Hood. Five versions of this story were created and presented to the subjects:

1. "Medium-register Spanish"
2. "Bilingual Spanish with insertions of specialized English-language lexical items"
3. "Bilingual Spanish with insertion of core English-language lexical items"
4. "Felicitous Spanish-English code-switching"
5. "Infelicitous Spanish-English code-switching"

Through a scalar judgments survey–matched-guise–that inquired about the personality characteristics of the "author of each text," the researchers found that the subjects "viewed the forms in question along a continuum, with single-noun insertions more positively evaluated than code-switching" (Anderson & Toribio, 2007, p. 234). Furthermore, bilingual proficiency also contributed to the participants' ratings. "Among high proficiency judges, referentially specific insertions and felicitous code-switching elicited the most favorable ratings... Interestingly, texts incorporating these contact forms were preferred over monolingual Spanish guises, perhaps reflecting participants' recognition of the inexorable mutual influence of English on Spanish in the U.S" (2007, pp. 234–235). Finally, the researchers identified that those subjects who reported higher bilingual proficiency were also able to make a distinction between felicitous and infelicitous Spanish-English code switching, with the former raking more favorable evaluations than the latter.

Speaker-external social factors

In this section, we will review the impact of language input, especially accented speech, in an L2 as it pertains to category formation and cue weighting of the acoustic properties of the input. As such, the current section will introduce the Speech Learning Model, a theoretical framework that captures the importance of language input in phonetic categorization and speech production. In Flege (1995), the author provides a comprehensive overview of the field of L2 speech learning in sequential bilingualism. It begins with a discussion of the different theoretical approaches to L2 speech learning, and then reviews the empirical research on L2 speech ranging from consonantal to vowel sound productions. The main theoretical objective of this piece, however, is to introduce the *Speech Learning Model* (SLM), a framework theorized from decades of empirical studies on L1-L2 speech patterns and the acquisition of L2 sounds across the lifespan; it short, it aims to "account for age-related limits on the ability to produce L2 vowels and consonants in a native-like fashion" (p. 237).[9] The SLM argues that L2 speech learning is a process of phonetic category formation. In other words, L2 learners must create new phonetic categories in their mental lexicon for the sounds of the L2. Nevertheless, Flege also points out that there are a number of challenges to achieving native-like L2 pronunciation. One primary challenge is that the sounds of different languages can be very similar, but they can also be very different. This can make it difficult for L2 learners to distinguish between the sounds of the L2 and the sounds of their L1. The author argues that:

> Learners of an L2 may fail to discern the phonetic differences between pairs of sounds in the L2, or between L2 and L1 sounds, either because phonetically distinct sounds in the L2 are "assimilated" to a single category (see Best this volume), because the L1 phonology filters out features (or properties) of L2 sounds that are important phonetically but not phonologically, or both. (Flege, 1995, p. 238)

Another challenge is that L2 learners who start learning L2 at a young age–typically before puberty–will be more likely to achieve native-like L2 pronunciation than L2 learners who start learning L2 at an older age. However, the author argues against the idea of maturation effects as a result of a critical period, given the wide range of acquisition outcomes in the empirical data, even when controlling for age of acquisition. He expresses that, "if L2 production accuracy were limited by maturation, such a limitation would apply across the board to the full range of L2 sounds that differ phonetically from sounds in the L1" (Flege, 1995, p. 241). As a whole, the author encapsulates his model into four postulates and seven hypothesis, which are all displayed in table 2.1.

[9] Although the SLM is not a sociolinguistic theory in and of itself, it is concerned with social and environmental factors such as input characteristics (e.g., monolingual-like versus accented speech), as we will see below. As such, this theory is introduced in this section as the primary theoretical framework for the analysis of speaker-external social factors.

	Postulates
P1	The mechanisms and processes used in learning the L1 sound system, including category formation, remain intact over the life span, and can be applied to L2 learning.
P2	Language-specific aspects of speech sounds are specified in long-term memory representations called *phonetic categories*.
P3	Phonetic categories established in childhood for L 1 sounds evolve over the life span to reflect the properties of all L1 or L2 phones identified as a realization of each category.
P4	Bilinguals strive to maintain contrast between L1 and L2 phonetic categories, which exist in a common phonological space.

	Hypotheses
H1	Sounds in the L1 and L2 are related perceptually to one another at a position-sensitive allophonic level, rather than at a more abstract phonemic level.
H2	A new phonetic category can be established for an L2 sound that differs phonetically from the closest L1 sound if bilinguals discern at least some of the phonetic differences between the L1 and L2 sounds.
H3	The greater the perceived phonetic dissimilarity between an L2 sound and the closest L1 sound, the more likely it is that phonetic differences between the sounds will be discerned.
H4	The likelihood of phonetic differences between L1 and L2 sounds, and between L2 sounds that are noncontrastive in the L1, being discerned decreases as AOL increases.
H5	Category formation for an L2 sound may be blocked by the mechanism of equivalence classification. When this happens, a single phonetic category will be used to process perceptually linked L1 and L2 sounds (diaphones). Eventually, the diaphones will resemble one another in production.
H6	The phonetic category established for L2 sounds by a bilingual may differ from a monolingual's if: 1) the bilingual's category is "deflected" away from an L1 category to maintain phonetic contrast between categories in a common L1-L2 phonological space; or 2) the bilingual's representation is based on different features, or feature weights, than a monolingual's.
H7	The production of a sound eventually corresponds to the properties represented in its phonetic category representation.

Table 2.1: Formulation of the SLM, adapted from (Flege, 1995, p. 239)

In summary, the SLM proposes that any L2 learner of any age is able to access input from the L2 and utilize the same mechanisms and processes used for L1 speech learning, which includes the capacity to create new phonetic categories. However, L2 learners do not always reach native-like productions of said sound categories, a limitation driven by two factors: degree of similarity between the L2 sound and the sound categories in the L1 inventory and age of first exposure to the L2. For the former, the general pattern is that the more dissimilar sound categories are between the L1 and L2, the easier it is for L2 learners to acquire the L2 category. For the latter, the older an L2 learner is when they begin acquiring the new language, the more difficult it is for them to discern the phonetic differences across the two languages. Nevertheless, the author does not provide a way to quantify or operationalize the quantification of degree of cross-linguistic dissimilarity perception.

Accordingly, to account for some of the limitations of the model in its initial form and to incorporate the results from research studies the following two and a half decades, Flege and Bohn (2021) reintroduced the SLM as the *revised* Speech Learning Model (SLM-r). While still focusing on L2 sequential acquisition and maintaining the mechanisms and processes for how speech sounds are learned, the SLM-r replaces the "age hypothesis" stated in the original SLM that is intended to explain the age-related effects in the acquisition of an L2. In the authors' own words,

> The core premises of the SLM-r are that (1) the phonetic categories which are used in word recognition and to define the targets of speech production are based on statistical input distributions; (2) L2 learners of any age make use of the same mechanisms and processes to learn L2 speech that children exploit when learning their L1; and (3) native versus nonnative differences in L2 production and perception are ubiquitous not because humans lose the capacity to learn speech at a certain stage of typical neuro-cognitive development but because applying the mechanisms and processes that functioned "perfectly" in L1 acquisition to the sounds of an L2 do not yield the same results. A difference in L1 and L2 learning outcomes will necessarily arise because
>
> 1. L1 sounds initially "substitute" L2 sounds because the L2 sounds are automatically linked to sounds in the L1 phonetic inventory;
> 2. preexisting L1 phonetic categories interfere with, and sometimes block, the formation of new phonetic categories for L2 sounds; and
> 3. the learning of L2 sounds is based on input that differs from the input that monolingual native speakers of the target L2 receive when learning the same sounds. (Flege & Bohn, 2021, p. 23)

Of particular interest to us is the first core premise–statistical input distributions. "The SLM proposed that L2 learners gradually 'discern' L1–L2 phonetic differences as they gain experience using the L2 in daily life, and that the accumulation of detailed phonetic information with increasing exposure to statistically defined input distributions for L2 sounds will lead to the formation of new phonetic categories for certain L2 sounds" (Flege & Bohn, 2021, pp. 31–32). However, the model did not provide an adequate methodology for measuring the accumulation of phonetic information by the L2 learner nor a quantification of the amount of phonetic input necessary to give rise to the formation of a new L2 phonetic category. As the authors of the SLM-r point out,

> The model simply pointed to years of L2 use as a metric to quantity of L2 input... [H]owever, immigrants' length of residence (LOR) in a predominantly L2-speaking environment is problematic because it does not vary linearly with the phonetic input that L2 learners receive and because it provides no insight into the quality of L2 input that has been received. (p. 32)

Flege and Bohn add that,

> It is universally accepted that infants and preliterate children attune to the phonetic categories of the ambient language through 'exposure to a massive amount of distributional information' (Aslin, 2014, p. 2; see also Kuhl et al., 2005). For the SLM-r, input is also crucial for the formation of language-specific L2 phonetic categories and composite L1–L2 phonetic categories. The SLM-r defines phonetic input as the sensory stimulation associated with L2 speech sounds that are heard and seen during the production by others of L2 utterances in meaningful conversations. (p. 32)

While the authors state that quantity and quality as input dimensions are difficult to measure (or operationalize their measurement in a systematic manner), they "observe that full-time equivalent (FTE) years of L2 input provides a somewhat better estimate of input than LOR alone does" (p. 32). The authors quantify FTE by obtaining the product of LOR and the proportion of L2 use as reported by subjects in linguistic research questionnaires. Thus, two individuals with equal LOR in an L2-speaking environment with distinct L2 usage will have distinct FTE values, which is part is attributed to contributing to the more accurate formation of an L2 category that resembles native speakers of the L2.

Although equally important, quality of input, as the authors posit, "has been largely ignored in L2 speech research even though it may well determine the extent to which L2 learners differ from native speakers" (p. 32). For example, Flege and Bohn reference research that shows that Spanish-L1-English-L2 adults who learned English as children but were exposed to a higher level of accented L2 speech displayed more converged VOT categories, putting them on par with English L2 learners who were exposed to English as adults. Thus, the authors emphasize the effect of input quality (native-like versus accented) for the creation of a new sound category in the L2.

Regarding input quality, the SLM proposed that a phonetic distinction may exist between the phonetic category formed by an L2 learner and a monolingual native speaker of the target language if said category relied on "features... not exploited" in the learner's L1 closest L1 category or if the features (in other words, perceptual cues) that define the L2 phonetic category are "weighted differently" in the L2 compared to the features that define the closest L1 phonetic category (Flege, 1995, pp. 239–243). Now, Flege finds the "feature hypothesis" incongruent with the first postulate of the model, namely that "the mechanisms and processes used in learning the L1 sound system, including category formation, remain intact over the life-span and can be applied to L2 learning" (Flege, 1995, p. 239). The SLM-r moves away from the earlier defined feature hypothesis given that new studies have demonstrated that late learners can and do gain access to some of the features utilized to define the L2 sound even if not exploited in the L1. The revised model now "adopts the "full access" hypothesis proposed by Flege (2005)" (Flege & Bohn, 2021, p. 45).

In monolingual research, it has been reported that the different cues that define a phonetic category are weighted optimally during the development of said category for its correct categorization. Moreover, identifying the optimal integration process for multiple perceptual cues in L1 sounds requires several years of input (e.g., Morrongiello et al., 1984; Nittrouer, 2004). Flege and Bohn (2021) further add that:

> Differences in cue weighting depend importantly on cue reliability (Idemaru & Holt, 2011; Strange, 2011), which, in turn, depends on the statistical properties of input distributions to which individuals have been exposed during L1 acquisition (Holt & Lotto, 2006, 3060–3062). Individual differences in cue weighting among monolinguals are likely to arise from exposure to different input distributions of tokens specifying a phonetic category (Clayards, 2018; H. Lee & Jongman, 2019). (Flege & Bohn, 2021, p. 45)

The authors further expand on the environmental factors that may influence their perceptual weighting:

> The cue weighting patterns specified in phonetic categories are not applied rigidly by monolinguals during the categorization of sounds in their native language. Human speech perception is necessarily adaptive (Aslin, 2014), enabling listeners, for example, to better understand foreign-accented renditions of their L1 after a brief exposure to foreign-accented talkers (e.g., Bradlow & Bent, 2008). Also important is the fact that cue weighting may adapt dynamically to what has been heard recently (Lehet & Holt, 2017; Schertz et al., 2016), and can be modified through training (Francis et al., 2008). (Flege & Bohn, 2021, p. 45)

Thus, for monolingual speakers, this adaptive cue weighting mechanism allows them to accommodate to their environment when exposed to new or different input.

Nonetheless, this adaptability expands beyond the perceptual realm and influences speech productions as well. For example, Nielsen (2011) found that monolingual English speakers were more likely to produce /p/ with longer VOT values after being exposed to stimuli that were artificially manipulated with a lengthened VOT in a hearing experiment (see also Clarke & Luce, 2005). In light of the adaptive cue weighting mechanism that has been identified in monolingual speakers, "the SLM-r predicts that individual differences will also be evident in the production and perception of L2 sounds that are perceptually linked to L1 sounds via the mandatory and automatic mechanism of interlingual identification" (Flege & Bohn, 2021, p. 47). In fact, other researchers have already reported on such cue weighting differences observed in beginner (Idemaru et al., 2012; Kim, 2012) and more advanced bilinguals (Chandrasekaran et al., 2010; Schertz et al., 2016). Accordingly,

> The SLM-r proposes that the influence of L1 cue weighting patterns will be stronger for L2 sounds which remain perceptually linked to an L1 category than for L2 sounds for which a new L2 phonetic category has been formed. Cue weighing patterns for newly formed L2 phonetic categories are expected to develop as in monolingual L1 acquisition, that is, to be based on the reliability of multiple cues to correct categorization that were present in input distributions. (Flege & Bohn, 2021, p. 47)

Therefore, this converged VOT categories observed in bilingual speakers may be the result of this linkage between L1 /p/, for example, and L2 /p/, if these two sounds are associated by the bilingual speakers as the closest sound categories across the two languages; and the phenomenon is replicated for each place of articulation of the stop consonants. But also, as mentioned above for monolingual speakers, the cue weighting observed in L2 learners may be the result of input that was received more recently (Lehet & Holt, 2017; Schertz et al., 2016).

To conclude the general sociolinguistic section, I connect the speaker-internal and -external factors around the importance of language input. While the influence of speaker-internal social factors, such as gender and age of acquisition of an L2, on the production of VOT during code-switched speech can be hypothesized more straightforwardly (e.g., women will likely produce more careful speech, resulting in less convergence, while more advanced age of acquisition will likely result in even larger convergence), the effects of language attitudes on the linguistic productions of monolingual and bilingual speech remain unclear, in particular, due to the fact that external factors will likely influence–or complicate–such linguistic attitudes. For instance, exposure to accented speech may evoke negative attitudes such as those mentioned by Achugar and Pessoa (2009). However, Fought (2006) also points to Chicano English, among other linguistic varieties, as a system that resulted from exposure to accented speech, as she describes in the following excerpt:

> Historically, Chicano English is the result of language contact between Spanish and English. When groups of Mexican immigrants arrived in California and other areas through the early part of the twentieth century, many of them learned English as a second language.The variety they spoke was a learner variety of English, as described above, heavily influenced by phonological and other patterns from Spanish. Children of these immigrants born in the USA, however, generally

> grew up using both Spanish and English. The non-native English of the community became the basis for a new dialect of English, Chicano English, now spoken natively by those born in the community. (Fought, 2006, pp. 79–80)

Hence, sufficient exposure of accented speech, or other environmental conditions, may overcome negative evaluations and permeate the language community, until those aspects are ultimately adopted by the larger community.

As shown and exemplified in the current section, speaker-internal and -external factors (e.g., demographic characteristics, linguistic attitudes, quality and quantity of input, etc.) can influence linguistic behaviors. The theoretical frameworks referenced in this chapter indicate that some of the environmental conditions, such as exposure to accented speech, may also have cognitive effects (e.g., the adjustment of cue weighting in speech perception), which ultimately may influence speech productions. Thus, we can conclude the current chapter by indicating that each of these factor classes influences the others in an interconnected linguistic system. As such, chapter 3 provides the description of the methodological design and execution of the current study, which attempts to implement a holistic analysis of methodological, linguistic, and social influences on phonetic productions during language alternation.

2.5 Summary

In general, the current chapter explored theoretical models and empirical research that spanned the cognitive, linguistic, and social domains. While some of the information shared here provides important foundational background, other contents of this chapter make clear predictions for speech productions, especially in language alternation contexts, which are of great interest for the present study. In this section, I will lay out a summary of the relevant background and its implications for the current study.

First, at the beginning of this chapter, we saw a section with the general empirical data that compare Spanish and English phonetic characteristics, especially as they relate to VOT productions. The studies shared in said section also included information pertaining to monolingual and bilingual mode productions by multilingual speakers. All in all, the relevant studies on VOT and the phonetics of language alternation in bilingualism research tells us that:

1. VOT: English has longer VOT and a wider range than Spanish (Lisker & Abramson, 1964; Keating, 1984).

2. VOT in Bilingual mode: Multilingual speakers display an even greater magnitude of convergence in bi/multilingual mode, compared with monolingual mode (Amengual, 2021; Olson, in press).

3. Long-lag convergence: Long-lag languages (e.g., English) tend to display greater convergence than short-lag languages in language alternation (Bullock & Toribio, 2009c; Amengual, 2021; Olson, in press).

In light of the aforementioned empirical data, which are straightforward and corroborated by countless studies, we can make the following corresponding predictions for speech produced in language alternations:

1. Speaker will produce longer VOT and more variability in English than Spanish, following general trends in this language.
2. Language alternation among Spanish-English bilingual speakers will lead to phonetic convergence, grounded in the well-documented pattern of convergence in bilingual mode.
3. Among Spanish-English bilinguals, convergence will be observed at least in English speech, but potentially in Spanish as well, albeit with a lower magnitude.

Second, the next section of this chapter provided information with respect to cognitive mechanisms utilized for different research tasks such as interview and puzzle completion activities. The section continued with a description of some theoretical frameworks that try to explain speech behaviors, especially phonetic adjustments that can be interpreting as (a) convergence in accommodation to the interlocutor's productions or divergence as a way to signal social distance, (b) accommodation to the interlocutor's communicative needs, or (c) self-monitoring mechanisms associated with the amount of attention paid by the speaker to their own speech. In general, this section can be summarized in the following three points:

1. Speech processing: While uninterrupted conversations and conversations with distractors (e.g., with a puzzle distractor) rely on similar cognitive processes (i.e., attention, working memory, and executive function), it is expected that activities such as conversations with distractors will be more cognitive demanding (Diamond, 2013; Ganley et al., 2014).
2. Attention to speech: Activities that require more careful speech or elicit shorter utterances will lead to more self-monitoring, resulting in more careful pronunciations (Labov, 2011).
3. Communication accommodation theory and audience design: People adjust their communication style in order to facilitate understanding and build rapport with others or to suit the communicative needs of their audience (Giles, 1973; Giles & Ogay, 2007; Bell, 1984).

Based on the arguments posited by these theories or empirical results, we can draw the following predictions for speech production in language alternation:

1. Activities that are more cognitive demanding are expected to result in less speech monitoring by the speaker themselves, decreasing the probability of speech that contains conscious phonetic adjustments. In language alternation, this should manifest itself as more convergence in more cognitive demanding activities.

2. Activities that lead to more self-monitoring (e.g., those without distractors or shorter scripted stimuli) will result in more careful pronunciations, or in other words, more standard forms. In language alternation, this should result in less convergence.

3. Whether speakers are trying to emphasize the social distance with or adjust to the needs of the interlocutor, speakers should display similar phonetic patterns in all interactive activities, compared to activities without an interlocutor, regardless of language mode (i.e., mono/bilingual).

Third, the following section introduced two linguistic factors that have been shown to influence the production of VOT: specific POA of stop consonants and speech rate of the utterance where the stop was produced. The literature referenced in this section can be condensed to the three general points below:

1. POA: There is a place-dependent VOT pattern observed cross-linguistically. That is to say, more posterior stops are linked to longer VOT productions. Thus, the order of VOT length from shorter to longer is as follows: bilabial $<$ coronal $<$ velar. (Cho & Ladefoged, 1999)

2. Speech rate: Studies have demonstrated consistently that speaking rate influences phonetic features such as VOT, where slower speaking rate typically results in longer VOT values (Kessinger & Blumstein, 1997; J. L. Miller et al., 1986; Nagao & de Jong, 2007).

3. POA and speech rate: Nonetheless, VOT adjustments as a result of speech rate apply to all POA's consistently across the board. That is, when VOT increases in slower speech, the feature increases at approximately the same rate for all POA's. Therefore, there does not appear to be an interaction of speaking rate and POA (Theodore et al., 2009).

This recapitulation of the linguistic factors of interest for the present study gives rise to the following corresponding predictions for language alternation productions of VOT among bilingual speakers:

1. Given that this place-dependent pattern has been found in both English and Spanish speech, it is expected that it will be repeated in bilingual speech (i.e., language alternation), with more anterior POA's producing shorter VOT for both languages, compared with more posterior POA's.

2. Given the well-documented pattern of speech rate influence on VOT cross-linguistically, it is expected that the pattern will be repeated in language alternation, with slower speech yielding longer VOT in both languages.

3. Again, it is not expected that an interaction between speaking rate and POA in language alternation will occur, following the same speech behavior observed in monolingual speech.

Fourth, and finally, below I encapsulate the important elements concerned with the explanation of phonetic patterns driven by social factors, both internal and environmental:

1. Gender: There appears to be a gender-based VOT production pattern, with females typically producing longer VOT values than males (Swartz, 1992; Ryalls et al., 1997; Allen et al., 2003; Robb et al., 2005).

2. Age of acquisition: Generally, a more advanced age of acquisition of an L2 is linked to less monolingual-like phonetic productions in the L2, although age of acquisition by itself is not a deterministic predictor of category formation in the L2. Other important aspects are quantity and quality of input (Flege, 1995).

3. Linguistic attitudes: Spanish-English bilingual teenagers associate Spanish dominance with first-generation immigrant status (Galindo, 1995). While bilinguals value formal Spanish and Spanish-English bilingualism in academic settings, Spanish spoken with monolingual characteristics receives receiver lower evaluations (Achugar & Pessoa, 2009). Finally, Spanish-English code switching has low overt prestige but high covert prestige among bilingual speakers; code switching has become a marker of ethnic identity (Toribio, 2002).

4. Cue weighing: Exposure to diverse sources of linguistic input, including linguistic idiosyncrasies, dialectal differences, and accented speech, can lead to an adaptation of the cue weighting mechanism, resulting in modifications to their phonetic categories, at the individual level, and to the creation of new linguistic varieties, at the community level (Flege & Bohn, 2021; Fought, 2006).

Taking all of this into consideration, I now provide, in respective order, the predictions for language alternation that can be drawn from these social aspects that influence language production.

1. While females tend to produce longer VOT than males, there is no evidence that indicates this pattern would differ in bilingual speech, suggesting that females will probably still produced longer VOT. However, it is not expected that this pattern will depend on the speaking activity in which people are engaged while alternating languages (e.g., read-aloud, interview, etc.). That is, females are expected to produce longer VOT consistently across activities.

2. Because late bilinguals already have what is called "compromised categories" (i.e., their L2 VOT is more L1-like), it is expected that they will display less convergence, compared to bilinguals who have farther-apart VOT categories. Since bilingual speakers have consistently demonstrated to acquire and maintain distinct phonetic categories, they may be less likely to converge their VOT productions during language alternation in order to maintain this phonetic distinction between their already closer categories.

3. Considering that code switching has low overt but high covert prestige, it is likely that, in bilingual mode, speakers will display less convergence in read-aloud activities with scripted stimuli, compared to spontaneous speech, as speakers may be more inclined to use a higher rate of standard forms (e.g., shorter VOT in Spanish and longer VOT in English) in laboratory settings, especially for activities that draw more attention to their own pronunciations.

4. Spanish-English bilinguals who are exposed to more accented speech (i.e., Spanish spoken with an American English accent and/or English spoken with a Spanish accent) are expected to display more convergence, given that their VOT categorization range for stop consonants is likely to be wider due to the greater level of variation they are exposed to on a regular basis. As such, they are expected to have more leeway within their phonetic categories.

Chapter 3

Methodology

Chapter 2 provides a broad discussion on numerous factors that have been recognized as predictors of speech production patterns, especially those related to language alternation matters. As a reminder, those factors range from cognitive to linguistic to social. As such, it is evident that there are many factors at play when it comes to identifying the impetus that leads to the phonetic convergence observed in speech production during language alternation. Considering all the background information provided by the previous section, we now continue to the next subsection which discusses the research questions and hypothesis that drove the development and completion of the present study. After that, this chapter will continue with a description of the participants whose speech was recorded and analyzed in this production study, the materials and instruments used for the data collection, the procedures followed during the data collection, and finally, the analyses performed on the data obtained for the present study.

3.1 Research questions and hypotheses

The research question driving this proposed study, in its broadest form, is: do specific experimental tasks influence phonetic convergence in language alternation studies? But more specifically, this study aims to determine which research design aspects, if any, contribute to an increase (or decrease) in convergence observed in the speech produced by bilinguals. However, considering the previously mentioned cognitive, linguistic, and social factors, a more meticulously devised list of research questions guiding the present study is:

1. To what extent are experimental tasks influencing phonetic convergence in language alternation studies? More specifically,

    a. which research design aspects contribute to an increase (or decrease) in the pattern of convergence observed in bilingual speech?

b. what do these findings indicate to us about the cognitive language processing mechanisms involved in language alternation studies–cued switching *and* code switching?

c. how does this new knowledge impact our understanding of the results reported in the literature up to date and how can these results influence our methodological setups moving forward?

2. Which linguistic factor (namely, POA or speech rate) is more salient and influential for mediating convergence effects in language alternation? If these patterns differ from monolingual speech phenomena,

   a. what are the variables that differ?

   b. what are the language-specific contexts that may lead to such outcome?

   c. what may be the universal patterns observed in language alternation behavior cross-linguistically?

3. What are the most prevalent speaker-inherent social factors that influence convergence effects? That is, language history, language proficiency, and language attitudes. In particular,

   a. from these aforementioned areas, which ones can more accurately predict linguistic behavior in language alternation?

   b. are they associated with a particular language or language profile?

4. What are the most prevalent external social (i.e., environmental) factors that can predict convergence effects? For example, exposure to 'accented' speech or language usage in specific environment. In particular,

   a. from these aforementioned areas, which ones can more accurately predict language production patterns in language alternation?

   b. are they associated with a particular language or environmental context?

The experimental conditions the present study examines are (1) methodological mode: (a) code switching versus (b) cued switching and (2) attention to speech,[1]. Of paramount importance is the fact that his study employs a within-subjects design, as opposed to a between-subjects design, wherein participants engage in multiple tasks, yet all tasks are subsequently compared collectively, as if executed by a uniform group of individuals. This design across all tasks allows us to answer a far too frequent question in bilingualism studies: is this pattern for a given activity an idiosyncratic effect in a particular individual or population group, or are the observed effects more indicative of a generalizable pattern in bilingualism.

[1]I follow the "attention to speech" scale provided by Labov (2011, p. 265) to make assessments about the amount of monitoring taking place in each task, as discussed in the literature review section.

In so doing, this study provides us with a better understanding of patterns associated with phonetic adjustment and convergence as a result of language alternation behavior.

Guided by the results from relevant studies described in chapter 2 or predictions inspired by said studies, the following predictions and/or hypotheses were formulated for the research questions listed above and tested in this study:

1. There will be task effects of methodological mode and attention to speech.
    a. Research participants will display a greater convergence effect of VOT in cued switching tasks than code switching tasks as a result of language switching costs associated with laboratory language switching tasks (e.g., Olson, 2017).
    b. Within each methodological modes (i.e., cued switching or code switching), the activities that allow for more attention to speech will display a lower convergence effect because speakers will monitor their own speech more closely and more closely approximate monolingual speech productions compared to tasks that allow for less attention to speech (Labov, 2011).
2. The linguistic factors (namely, speaking rate and POA) will influence all language alternation data in all tasks, following the patterns observed in monolingual speech.
    a. Speaking speech (i.e., faster speech rate) will be correlated with shorter VOT values (e.g., J. L. Miller et al., 1986; Nagao & de Jong, 2007).
    b. VOT measurements will increase as the place of articulation shifts towards a more posterior position (i.e., bilabial $<$ coronal $<$ velar) (Cho & Ladefoged, 1999).
    c. These linguistic effects will manifest in the research tasks of both methodological modes with the same magnitude.
    d. The patterns observed in this study will resemble the direction and patterns reported in the literature for monolingual speech.
3. There will be speaker-internal social effects in the participants' linguistic productions.
    a. Subjects who report more positive attitudes towards each language (i.e., English or Spanish) will display less convergence in said language due to their desire to meet the monolingual production expectations in the language.
    b. Subjects who report an earlier age of acquisition of the L2 will display higher levels of convergence as a result of larger phonetic latitude afforded by their farther-apart VOT categories (Bullock & Toribio, 2009c; Olson, in press).
    c. Speakers will display convergence (or a greater degree of it) in their dominant language, although English-dominant speakers may show even greater convergence due to the language's phonetic characteristics (long-lag languages have more phonetic latitude) (Olson, in press).

4. There will be environmental effects in participants' VOT patterns.
    a. Participants who have more exposure to accented speech will display more VOT convergence than those who report less exposure due to shifts in their cue weighting to accommodate to the accented speech (Flege & Bohn, 2021).

To answer these research questions and test their respective hypotheses, research subjects in this study participated in the completion of four tasks, two of which can be categorized as researcher-prompted cued switching while the other two can be categorized as speaker-initiated code switching. This study explores VOT production in voiceless occlusives /p t k/ in word-initial position. The sections below provide a more detailed description of the different components of the methodology for the proposed study.

3.2 Participants

Recruitment for the present study started in the summer of 2022 and continued until early fall of the same year. The recruitment process included invitation of the researcher's personal contacts, flyer[2] dissemination in listservs across UC Berkeley, and snowball sampling. The target population for this study is bilinguals who self-assess as proficient in Spanish and English, both in speaking and reading, for reasons I will cover below. In addition to the methodological variables, this study also analyzes some linguistic and social factors. A total of 62 bilingual individuals showed interested in the study, scheduled appointments, and participated in the data collection process. However, given that two of those participants were unable to completed all production tasks as instructed, their data were removed from the analysis. In total, data from 60 Spanish-English bilinguals were analyzed.

Three qualifications were provided during the recruitment process. To participate in this study, individuals needed to be:

1. 18 years of age or older at the time of data collection.
2. bilingual in Spanish and English.
3. able to read in both languages at a basic level.

The first qualification is self-explanatory, thus no clarifications were provided. For the second, people were advised in the recruitment literature that, for the purposes of this study, someone was considered "bilingual" if they were able to hold a conversation on various topics in both languages. Finally, regarding the third qualification, individuals who were "able to read in both languages at about 4th grade level or higher" were invited to participate. In response to follow-up questions on this last statement, they were advised that being able to text in English and Spanish with friends and family was a good measure of "basic reading level."

[2]The flyer is included at the end of this work in appendix A.

In total, there were 41 participants who identified as female and 19 who identified as male. The mean age was 22.05 years (SD=5.45) with a minimum of 18 and a maximum of 50. The general dominance score provided for the entire group by the Bilingual Language Profile (BLP)[3] is 25.5 (SD=46.54), indicating that overall the group tended to be slightly more English-dominant. There was a maximum dominance score of 118.51 and a minimum score of -100.16.[4] See figure 3.1 for a general distributions of these values; a division at the 0 dominance score value is provided as a vertical red line. The blue, curvy line is the kernel density estimate plot.

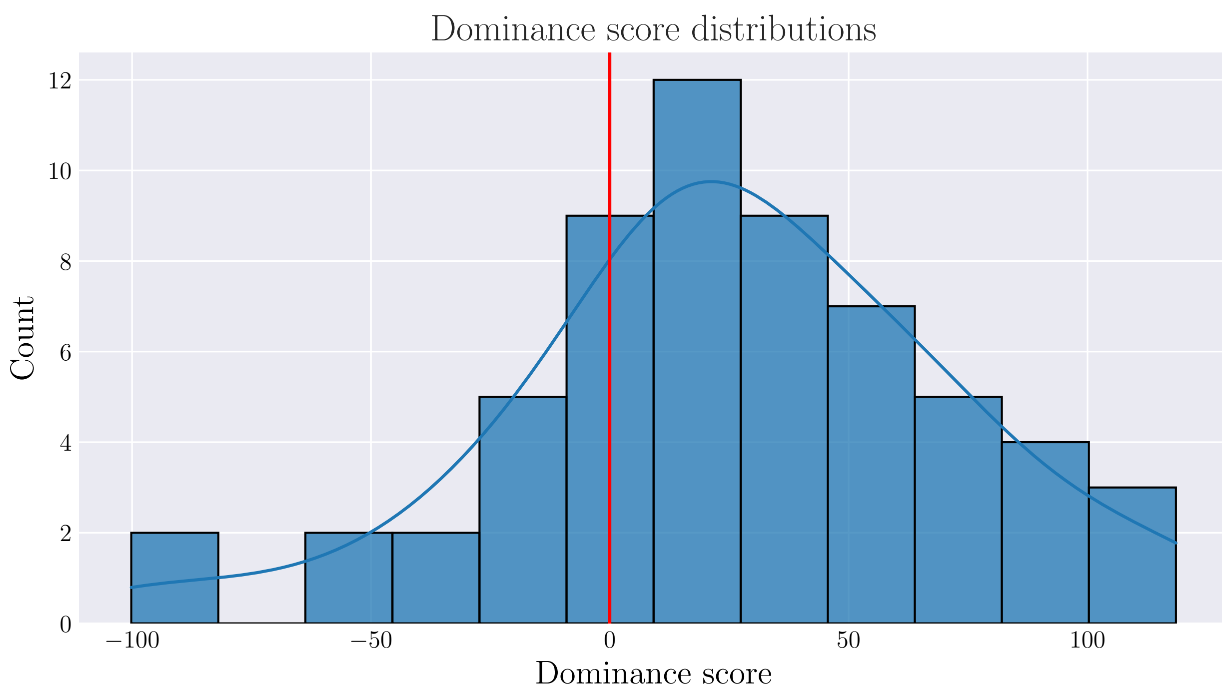


Figure 3.1: Distribution of participants' dominance scores from BLP

Furthermore, while participants reported differing levels of education, ranging from having completed "less than high school" to obtaining a "master's," the majority of participants reported having completed some college or currently being enrolled in college–an expected distribution considering the primary recruitment location for this study was a college campus, albeit snowball sampling brought in several members from the outside community with

[3]The Bilingual Language Profile is a survey that collects information about a participant's language profile, including language proficiency, and calculates their dominance score from their answers to these questions. A positive score indicates more English dominance, whereas a negative score indicates Spanish dominance. More information will be provided in the sections below.

[4]In these data, a positive value is associated with more English dominant and a negative value is associated with more Spanish dominant.

differing levels of education. Figure 3.2 displays the general distributions for the reported levels of education that each participant has completed or is currently on the process of completing.

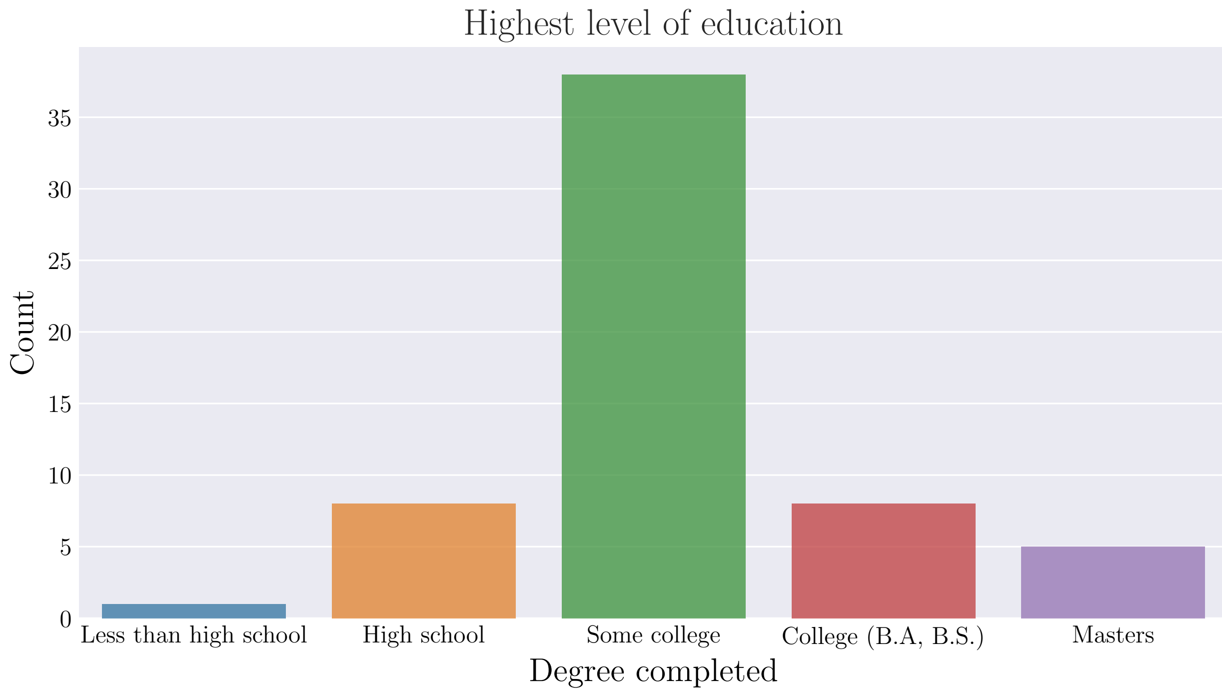


Figure 3.2: Distribution of participants' highest level of education

Finally, in the self-reported scores in the BLP for language proficiency across the four modes of communication–namely, speaking, understanding, reading, and writing–the subjects consistent listed English as their more proficient language across all modes. As figure 3.3 shows, using a scale from 0 to 6, with 6 indicating highest proficiency level, English data received higher scores compared to the Spanish data, with most of the English scores hovering around 5 and 6, especially for understanding and reading. Spanish data, on the other hand, show more variability in the scores it received, especially as it relates to speaking and writing. Table 3.1 displays the average scores reported for each mode of communication for each language, followed by the standard deviation, the minimum reported score, and the maximum recorded score for each category in parentheses.

All in all, although the English proficiency scores are generally higher than the Spanish scores, the subjects nonetheless indicated an overall high proficiency in both languages, with the Spanish averages across all modes of communication ranging from 4.5 to 5.5 and English averages ranging from 5.6 to 5.8.

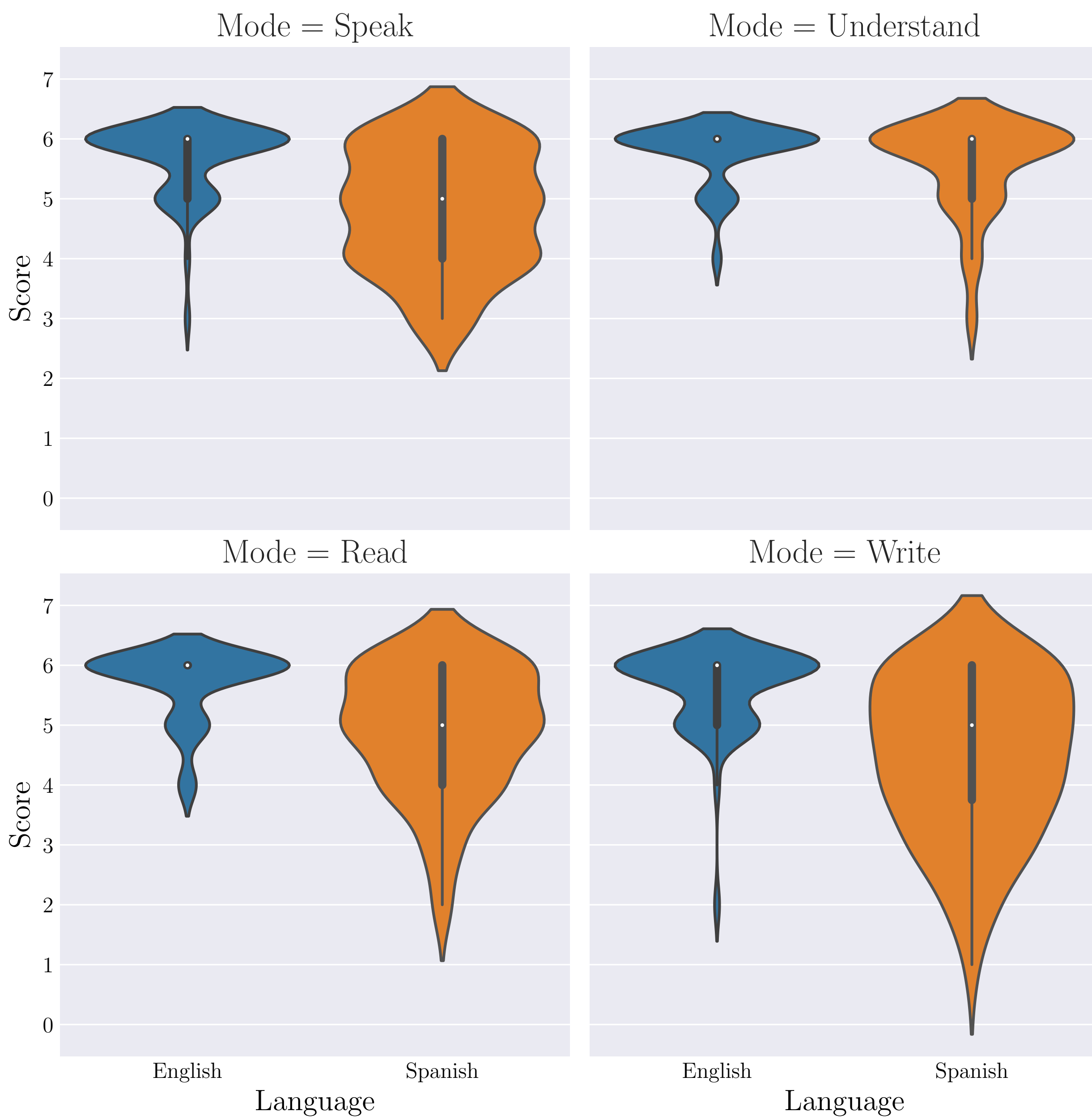


Figure 3.3: Proficiency scores per mode of communication

3.3 Research tasks, materials and instruments

As a reminder, subjects participated in a variety of production activities while alternating languages while their speech was being recorded in order to identify their speech behaviors during each of those activities. The participants also provided information regarding their

	English	**Spanish**
Speaking	5.68 (SD=0.6, Min=3, Max=6)	4.8 (SD=0.99, Min=3, Max=6)
Understanding	5.77 (SD=0.5, Min=4, Max=6)	5.55 (SD=0.77, Min=3, Max=4)
Reading	5.7 (SD=0.59, Min=4, Max=6)	4.88 (SD=1.06, Min=2, Max=6)
Writing	5.62 (SD=0.6, Min=2, Max=6)	4.43 (SD=1.32, Min=1, Max=6)

Table 3.1: Proficiency means per mode of communication

language profiles in order to determine the influence from social factors. Accordingly, the data collection process can be summarized as follows:

1. It elicits production data in two cued switching tasks: word list and passage read aloud activities.

2. It elicits production data in two code switching tasks: puzzle completion (spot the difference) with a speaking component and a casual interview.

3. It obtains speakers' demographic and linguistic background information through an exit survey.

In this study, the sounds of interest are the phonemic voiceless stop consonants /p t k/ in Spanish and English, in particular their VOT realization. This section provides a detailed description of the activity designs and materials used in their execution.

Word list reading

The word list reading task is composed of a total of 240 words, 120 of which (that is, 50%) are critical stimuli starting with a voiceless stop consonant with the remainder being fillers or distractors to ensure participants do not deduce the identity of the sounds of interest for this research study; that is, the fillers started with sounds other than voiceless stop consonants to increase the diversity of word-initial sounds. While it is possible for participants to identify one or more of the sounds of interest (/p t k/), many steps were taken during the design of the activity to minimize this risk, including the incorporation of filler words, the randomization of all items in the word list, and the inclusion of critical items, especially for velar-initial words, that started with different graphemes (e.g., <c>, <k>, <qu>) in their orthography; more details on this will be provided below. POA and language were stratified and fully crossed to ensure an equal number stimuli in each category (see table 3.2 for precise lexical item counts). For consistency, only word-initial stops that precede a vowel were included in the stimuli. In Spanish, each of the three voiceless stops was followed by one of the five phonemic vowels: /i e a o u/, with four stimuli per vowel. In English, each of the three voiceless stops was followed by one of the following five vowels: /i ɛ æ o͡ʊ u/. These five

	English	Spanish
Bilabial /p/	20	20
Coronal /t/	20	20
Velar /k/	20	20

Table 3.2: Stratification of language and place of articulation of word list stimuli

vowels[5] were selected because they most closely resembled their Spanish counterparts in their articulation and acoustics. The stimuli carried the primary stress on the first syllable with no exceptions. All words in both languages were either monosyllabic or disyllabic, with one exception–párpado (*eye lid*). Refer to appendix B for a complete list of English and Spanish stimuli for the word list reading task.

Only non-cognate words are used in this cued switching task. A cognate effect has been identified in bilingual studies, which increases the convergence of phonetic categories in bilingual speech.[6] To avoid a confound, only non-cognate words are presented in the stimuli. In this activity, the stimuli and fillers were randomized into a single list and presented using a presentation preparation software. The lexical terms were alternated between English and Spanish to ensure that each lexical item comes after another lexical item in the other language, thus ensuring that all tokens appear immediately after a switch site. While the general list was randomized and included critical stimuli and distractors, all participants received the same order. Table 3.3 includes a small sample of what the data presentation looked like for a small subset of critical stimuli. Fillers have been omitted for simplicity. Each row represent a single slide in the presentation shown on screen to the participants, meaning that participants could only see one word at a time. More information on the activity instructions and stimulus presentation are provided in subsequent sections.

Passage reading

The passage reading task is composed of a total of 10 passages, each containing a language switch after which the next language starts with a voiceless stop-initial word. The passages were devised so that they contained a total of six switches, with the following word containing one of the three stop consonants in each language. While some switches occurred in an utterance-initial position, others are found in utterance-medial position. However, this pattern (i.e., a variety of utterance placements) was present in all passages, across POA, and

[5]Note that the diphthong /o͡ʊ/ is included in this set given that the monophthong /o/ is not phonemic in English. While the monophthong /ɔ/ is an option, the phonological mapping of the verb *to mop* as *mapear* in this English loanword into Spanish suggests that, at least in some cases, /ɔ/ can be mapped to a low central vowel in Spanish. However, other English loanwords such as *bot* are incorporated into Spanish as *bot*, thus the murkiness of this vowel steered me into excluding it from the list of stimuli.

[6]To mention just a few notable studies, the reader is referred to Flege and Munro (1994), Amengual (2012), and Younes and Gathercole (2020) for a more detailed account of this phenomenon.

English	Spanish
[t]ackle	
	[p]eso
[k]eg	
	[k]odo
[p]oking	
	[t]ina
...	
	...

Table 3.3: Sample word list presentation

for both languages, in order to maintain setups that were as consistent as possible as to not insert confounds in the stimuli while still maintaining a naturalistic narrative flow. Below is a sample passage that represents the stimuli:

> *When I was in middle school,* [t]`odos mis amigos vivían lejos de mí, así que para ellos` [k]*oming over to play at my house wasn't always possible.* [p]`or eso mi mamá a veces iba por ellos a sus casas, pero` [t]*aking them back afterwards was even more challenging.* [k]`omo ella trabajaba en la noche, salía de casa a las 6 de la tarde. Por suerte,` [p]*olitely asking my older sister to take my friends home usually worked, though.*[7]

Accordingly, the language and POA were cross-balanced for equal number of stimuli in each category, as represented in table 3.4. Similar to the word list stimuli, switch sites were followed by voiceless stop-initial words where the initial consonant was followed by one of the five vowels in Spanish: /i e a o u/ or these vowels in English /i ɛ æ o͡ʊ u/. Each critical stimulus word followed a vowel, a nasal sound or a sibilant sound; that is, the last sound prior to a switch site belonged to one of these three sound categories. Similar to the word list stimuli, non-cognates were used for all critical lexical items. However, unlike the previous task, not all critical stimuli were stress-initial words, given that that complicated the passage creation in a way that all critical items met all restrictions and still allowed for sensible narrative flow. Of the 60 stimuli, 49 words were stress-initial (81.7%). Table 3.5 illustrates the number of stimuli in both languages that begin with a stressed or unstressed syllable across all 10 passages. Note that 29 of the 30 English stimuli (96.7%) and 20 of the 30 Spanish stimuli (66.7%) were stress-initial words. Nonetheless, for both languages, a majority of the passage stimuli were stress-initial, allowing for a fair comparison with the

[7]For this passage, the first language used (i.e., English) is visualized here in italics font style, with the following language in monospace font style. The font styles alternate back and forth to indicate language switches. These font styles are used in this example to facilitate the identification of language distinctions and switch sites.

	English	**Spanish**
Bilabial /p/	10	10
Coronal /t/	10	10
Velar /k/	10	10

Table 3.4: Stratification of language and place of articulation of passage stimuli

	English		**Spanish**	
	Stressed	Unstressed	Stressed	Unstressed
Passage 1	3	0	2	1
Passage 2	2	1	3	0
Passage 3	3	0	3	0
Passage 4	3	0	2	1
Passage 5	3	0	2	1
Passage 6	3	0	1	2
Passage 7	3	0	3	0
Passage 8	3	0	1	2
Passage 9	3	0	2	1
Passage 10	3	0	1	2

Table 3.5: Count of passage stimuli with a stressed or unstressed initial syllable

stimuli from the word list. Finally, while no switches were provided with distractor sounds, the fact that the velar stop can be represented in writing through several graphemes in word-initial position (i.e., <c, k, qu>) facilitated the disguise of the the critical stimuli. Refer to appendix C for a complete list of English-Spanish switched passages for the second reading task.

Puzzle completion

For this and the following activity, the methodological mode changes from cued switching to code switching, which means that rather than engaging in cued switches prompted by the researchers, research participants were encourage to engage in language alternation on their own terms, on their own time. Accordingly, these activities were designed with this detail in mind. The puzzle task was a spot the difference activity, also known as diapix, commonly used for spontaneous speech elicitation in behavioral and speech research studies (Van Engen et al., 2010; Baker & Hazan, 2011). An original image was created for this activity and designed to include primarily objects whose names are voiceless stop-initial and non-cognate words in English and Spanish. Table 3.6 displays a subset of objects included in the puzzle image. The image does not include orthographic representations of the names of

English	**Spanish**
[t]urkey	[p]avo
[k]ough	[t]os
[p]eel	[k]áscara

Table 3.6: Sample items represented in the spot the difference puzzle image

critical objects, as the purpose of the study is to have participants rely on lexical object recall from memory rather than reading, which was the mechanism from the previous two tasks. While one /k/-initial word (<keep>) is provided in writing in the image, this word was not included in the analysis of the present study. The few other orthographic representations in the image are ignored altogether given that they do not begin with a stop consonant sound.

Note that, even though spot the difference activities typically include two very similar images with a few differences across the images, in this activity only one image was created and provided to the participants. This was done because, rather than eliciting the production of a handful of words–only those objects that differ between the two images–the purpose of this activity was to have the research subjects name all objects shown in the image as a way of eliciting as much data as possible for this activity. More information is provided in the next section explaining how the spot the difference task was carried out in this research project with only one image. Refer to appendix D to access the image used for this spot the difference puzzle task.

Casual interview

A list of questions on a variety of topics was devised to formulate the basis for a semi-structured casual interview with the participants. The questions were designed with a college student profile in mind, but they also allowed for a fruitful and engaging conversation with research subjects from outside of the higher education system. Thus, not all questions were asked to all participants, for example, questions related to college enrollment for participants who have not attended college. Moreover, follow-up questions were included in the interview to clarify parts or entire answers or to elicit additional information on a topic. The interview covered mundane topics such as work duties or scholarly interests, past and future travel plans, typical hypothetical questions such as "what would you do with a million dollars?", and inquiries regarding language usage and bilingualism as well as conversations about linguistic backgrounds and linguistic opinions. Refer to appendix E for a complete list of interview questions used as the starting point of this semi-structure casual interview activity.

While the list of questions in appendix E are presented in monolingual English, for clarity purposes, the questions presented to the participants orally were code-switched. The researcher engaged in language alternation behavior when conversing with the research subjects in this task to elicit the same type of linguistic behavior from the participants. As such, the questions that the subjects received looked more like the ones presented below:

1. *Are you a student?*
2. `¿Y qué estudias?` (and what do you study?)
3. `¿Qué tipo de trabajo` *do you want to get* `después de que` *you graduate?*
   (what kind of job do you want to get after you graduate)
4. `Si tuvieras` *one million dollars,* `y yo sé que` *that's not that much money nowadays,* `pero si lo tuvieras,` *what would you do with it* `y por qué?`
   (If you had one million dollars, and I know that's not that much money nowadays, but if you had it, what would you do with it and why?)

This subset of code-switched interview questions is presented with alternating font style changes to facilitate the identification of language switches for the reader. While the representation of the code-switched questions above provides clear switch sites, note that there was no set of switch boundaries for this task. Rather, the researcher focused on naturally produced switches chosen on-line during the completion of the interview activity, thus providing slightly different input forms for the research participants.

Bilingual Language Profile

The last instrument used for the data collection process of this research study was the Bilingual Language Profile (BLP; Birdsong et al., 2012). The BLP is a self-report questionnaire that is used to assess language dominance in bilingual speakers and is a survey often used in bilingualism studies (e.g., Coetzee et al., 2015; Perpiñán, 2017; Olson, 2017; Shen et al., 2020).[8] This open-access tool determines language dominance through a statistical algorithm and elicits participants' answers for a variety of areas associated with their language profiles, namely attitudes towards each language, proficiency across communication modes (i.e., listening, speaking, reading, and writing), language usage in a multitude of social contexts (e.g., school, work, with family or friends, etc.), order and age of language acquisition.

The BLP consists of 20 items that are divided into four categories: language history, language use, language proficiency, and language attitudes. The items in each category ask participants to rate their language skills and experiences on a scale of 1 to 6 or through percentage indicators, except for questions related to age, which provide a different scale. This survey takes about 10 minutes to complete, and it produces a continuous dominance score that ranges from -218 to +218. A score of 0 indicates that the two languages are equally dominant, while a score of -218 indicates that the first language is dominant and a score of +218 indicates that the second language is dominant. In general, the BLP produces a bilingual profile that takes into account a variety of linguistic variables, such as age of acquisition, language exposure, and language use.

[8] For a comparative and systematic analysis of the validity of this research tool, refer to Olson (2023), which provides a thorough evaluation and affirmation of its reliability for obtaining language dominance scores for linguistics research.

Finally, an additional question was added to this exit survey that inquires about the number and types of accent varieties that the participants are exposed to in Spanish and English (e.g., Indian-accented English or American English-accented Spanish) on a regular basis.

Knowledge of all these (speaker-inherent and environmental) social factors allows us to shed more light on the social-cultural variables that are often cited as leading factors in bilingual performance or behaviors.

3.4 Procedures

Each subject began by reading and filling out a consent form informing them of the purpose of the study, the procedures, the risks associated with their participation, and information regarding the compensation for their participation. In order to achieve full participation in this study, subjects completed all four linguistic production tasks–namely, word list reading, passage reading, puzzle, and casual interview tasks–as well as the exit BLP survey. The order of methodological mode (i.e., cued switching or code switching) alternated for each subject,[9] but the activity associated with least attention to speech in either mode was presented first, followed by the activity expected to draw most attention to speech. Below is a representation of the two orders that were followed:

- Order 1:
    1. Passage reading
    2. Word list reading
    3. Puzzle completion
    4. Interview
    5. BLP questionnaire
- Order 2:
    1. Puzzle completion
    2. Interview
    3. Passage reading
    4. Word list reading
    5. BLP questionnaire

[9]While the activity order was not randomly assigned to participants, the alternating selection of order worked as a proxy for randomization.

Following the "attention to speech" theory by Labov (2011), this study was designed under the assumption that executing the least-attention task first minimized the possibility of the subjects identifying the sounds of interest, ensuring that they do not modify their speech behavior as a result. This disposition was particularly motivated by the (small but nonetheless existent) possibility that participants identified the sounds of interest in the cued switching tasks: passage reading and word list reading.

Data collection took place in the Berkeley PhonLab from the Linguistics Department at UC Berkeley. All four production tasks were recorded in a large double-walled WhisperRoom sound booth using a AKG C535EB Condenser Microphone placed on a desktop microphone stand sitting on a table and about one foot away from the speaker's mouth. The lab setup also included a Steinberg UR22 audio interface and AKG K240 Studio professional over-ear, semi-open studio headphones for communicating with the researcher outside the sound booth. The speech was recorded on Praat as mono audio with a 44,100 Hz sampling frequency. A link to the BLP questionnaire was provided via a QR code that participants were asked to scan with their own personal smartphones.[10] Outside of the sound booth, participants sat at the lab's conference table while completing the survey. All in all, data collection lasted approximately two hours per participant, including consent form, speech production tasks, and demographic/linguistic questionnaire.

Among the cued switching tasks, subjects read the ten switched passages containing switches followed by voiceless stop-initial stimuli in both English and Spanish,[11] all presented in a randomized order. The passages were displayed to the participants using Adobe Reader, a commercial software used for the display of documents in PDF format. A Dell UltraSharp 1900FP 190 flat-panel monitor sitting approximately one and a half feet from the participants was used to visually display the passages. Participants were instructed to read each passage aloud when displayed on the screen, and once they finished reading each passage, the next passage would be automatically displayed without additional verbal cues. That meant that they needed to begin reading the next passage once visually available. The researcher was actively listening to the subjects' productions to initiate the transition to the next stimulus and/or provide additional clarification if or when necessary using the communication interface.

Additionally, participants read the Spanish-English word list aloud, which included the 120 voiceless stop-initial critical stimuli and distractors in a randomized order, although all participants received the same order. The list alternated Spanish and English words, and it was displayed to the participants using a similar setup described for the passage reading task above, except in this case LibreOffice's Impress application was used, a free and open-source software that provides slide deck preparation and presentation functionalities. On

[10]All participants had a personal smartphone in their possession for completing the survey; however, a laptop computer was available to participants to use, should they not have a smartphone in their possession at the time of data collection.

[11]The text was color coded in dark orange for the first displayed language and dark green for the second displayed language in order to maximize the color contrast between the two languages, aiding participants with the identification of language switches as they prepare for the transitions.

the screen, English words were all presented on the left side of the screen under the word "English" written in smaller font size, printed in green.[12] Alternatively, Spanish words were all presented on the right side under the word "Español" in smaller font size, printed in blue. None of the critical stimulus or filler words were color coded; rather, all were printed in black text. Participants were instructed to read the words aloud as they appeared on the screen. While monitoring the subjects' productions, the researcher switched to the new word once the correct word was produced. If participants failed to produce the correct word, the stimulus remained on the screen for the subjects to try again.[13]

Among the code switching tasks, subjects participated in a puzzle task and a casual interview. For the former, participants engaged in the completion of a spot the difference activity. Only one image was created with at least 53 physical items or actions whose names are voiceless stop-initial and non-cognates. Two copies of the same image were printed out. While the images were concealed from the view of the participants, they were shown the blank reverse side of the images, which were labeled as "A" and "B." Participants were then instructed to choose one of the two images, and once they made a decision, that images was given to them. The instructions proceeded with an explanation of how to complete the spot the difference puzzle activity:

> You have received one image and I have the second one. They're both very similar but have a few differences. Our job is to identify all of the differences. However, because I need your speech more than my own, you'll have to do most of the talking. So you will need to name all the objects you see and describe them to me. I will then let you know if the objects on my version of the image are similar to yours or differ from your description.[14]

The rest of the activity then continued with the researcher and subjects communicating using the communication interface. During these interactions, the researcher made up differences for subjects to believe the images in fact differed.

Finally, for the interview, the participants answered a series of questions posed by the researcher using a set of predetermined questions for this semi-structured interview; however, follow-up questions for clarifying answers or eliciting more in-depth information were present in all interviews. To ensure the elicitation of code switching, the questions were asked using code switching. No predetermined switches were engineered; rather, the researcher posed each question while alternating languages on the spot, yielding different switches for a unique input for each of the subjects. The range of conversation topics during the interview and

[12]The language cue words were color coded for aiding with the identification of the language, in addition to the location distinction.

[13]Because most of the words in the word list were shorter and higher frequency lexical items, most participants were able to identify and accurately produce the stimuli on the first attempt. Participants making multiple attempts to produce the correct word was not a major issue during data collection.

[14]While these instructions are provided here in English, code switching was used during the verbal explanation in the lab. This was done to create an environment were the subjects were exposed to both languages since the start to elicit the same behavior, especially in the code switching activities.

their order were intended to keep all subjects conscious about their language usage, in order to lead them to constant speech monitoring, as this is the most speech-attention activity among the two code switching tasks for this study.

During the code switching tasks, participants were informed that, given that this is a code switching study, it was important for them to alternate languages during their conversations in order to yield an appropriate amount of data that can be analyzed. However, they were then asked to switch languages only as it feels natural, asking them to perform only as they would in other natural contexts. Accordingly, they had the freedom to switch whenever and wherever they saw fit. Furthermore, the researcher actively used code switching during the sociolinguistic interview and the puzzle task. This leading approach was meant to prime the subjects and elicit a similar behavior in their conversations with the researcher.

3.5 Analysis

As a matter of course, this study was not designed to investigate the effect of language alternation on VOT productions. In fact, the lack of monolingual data in each language from all participants prevents us from making such a determination. Rather, this study focuses on the effect of methodological, linguistic, and social factors on VOT productions within a language alternation context. Thus, in this sense, the present study differs from the previous literature since prior studies have focused on determining how language alternation (either in naturalistic contexts or experimental settings) impacts speech patterns compared to monolingual settings, whereas this study focuses on how different language alternation tasks impact speech patterns. The dependent variable under investigation in this research project is the VOT measurements associated with voiceless stop-initial words in Spanish and English proceeding a language switch. The independent variable for the methodological aspect is experimental task–cued switching: word list and passage reading; and code switching: puzzle completion and casual interview. The linguistic independent variables are: POA and speaking rate. And the social variables are divided into two categories: speaker-internal and environmental social factors. Overall, two different statistical models were performed: one for the methodological and linguistic variables and another one for the social variables. Below is a more detailed description on the data and their processing, followed by the analysis performed for the two models.

Data processing

While the cued switching tasks essentially produced the same amount of data per participant, regardless of activity completion time (i.e., recording duration), the code switching tasks had a much greater amount of variability in the duration of the task, and as such the amount of data produced by speaker. The 60 recordings for the puzzle activity had an average duration of 13:31 min (SD=3:10 min; Min=8:03 min; Max=22:16 min), and the recordings for the interview activity had an average duration of 38:26 min (SD=9:44 min; Min=22:11 min;

Max=59:11 min). Overall, the puzzle activity yielded 13 hours, 31 minutes, and 11 seconds. The interview activity, on the other hand, yielded 38 hours, 26 minutes, and 17 seconds. Among the cued switching tasks, we find that the word list recordings had an average duration of 5:51 min (SD=1:05 min; Min=4:36 min; Max=10:45 min), whereas the passage reading task recordings had an average duration of 5:32 min (SD=1:03 min; Min=2:36 min; Max=11:05 min). In total, the word list activity yielded 5 hours, 33 minutes, and 46 seconds. Meanwhile, the passage activity yielded 5 hours, 20 minutes, and 34 seconds.

The audio for the word list and passage reading tasks were annotated by hand, given the simplicity of the task for the content was known/expected. The productions for the puzzle and interview tasks were spontaneous and thus required careful transcriptions that were sensitive to language alternation. Consequently, with the support of four research assistants, a TextGrid file was created for each puzzle and interview audio recording. The TextGrid files were then segmented into tiers with Spanish speech and English speech. Transcriptions for each language were then automated with OpenAI's Whisper automatic speech recognition software, relying on the most powerful model available at the time of data processing (i.e., `large-v2`; Radford et al., 2022). From the transcriptions, ten percent of the data were manually inspected to ensure accuracy; the model performed exceptionally, with minimal errors.[15] All data for both languages were then processed with the Montreal Forced Aligner for phonetic alignment; each language was processed individually with a language-specific dictionary (McAuliffe et al., 2017). This process made use of the pre-trained acoustic models for acoustic alignments for US English (McAuliffe & Sonderegger, 2023b) and Latin American Spanish (McAuliffe & Sonderegger, 2022b) as well as the pre-trained grapheme-to-phoneme models for generating pronunciation dictionaries of the orthographic transcriptions for US English (McAuliffe & Sonderegger, 2023a) and Latin American Spanish (McAuliffe & Sonderegger, 2022a), all pre-trained and maintained by the Montreal Forced Aligner team. Finally, VOT measurements for voiceless stop-initial words were obtained in both languages using AutoVOT's software through the VOT-CP wrapper (Keshet et al., 2014; Gutiérrez Topete, 2021).

The word list recordings yielded a total of 7,200 tokens (20 tokens per place of articulation, for two language, for 60 participants; that is, $20 \times 3 \times 2 \times 60 = 7{,}200$). However, 306 tokens were lost as a result of corrupted (portions of) files, needed to be thrown out because the incorrect word was produced, or were never produced at all–in the case the the subject did not recognized the word on the screen. That left us with a total of 6,894 tokens for the word list task. The passages produced a total of 3,600 tokens (1 token per place of articulation, for 10 passages, for two languages, for 60 participants; that is, $1 \times 3 \times 10 \times 2 \times 60 = 3{,}600$). However, 200 tokens were lost for the same reasons reported above. The passage reading tasks produced a total of 3,400 tokens. The puzzle task yielded a total of 2,340 tokens of interest, while the interview tasks produced a total of 1,935 critical tokens. All in all, 14,569 tokens were collected and analyzed for the present study. Table 3.7 shows the number of

[15] Among the most common errors produced, the model omitted filler words such as *uh* and *uhm* as well as the discourse marker *like*. However, since these items are irrelevant for this study, such errors were ignored.

	English			Spanish		
	p	t	k	p	t	k
Passages	527	593	583	569	533	595
Word list	1199	1141	1138	1088	1154	1174
Puzzle	348	266	333	349	397	647
Interview	45	89	52	770	273	706

Table 3.7: Distribution of critical tokens per task, language, and place of articulation

tokens produced in each task, divided by language and place of articulation. Note that the puzzle activity returned many more tokens than the interview activity, especially among English data. This is due to the fact that the image used in the puzzle activity was exclusively designed to elicit voiceless stop-initial lexical items, as the majority of the objects and actions presented in the image were /p t k/-initial words. As such, it was expected that a significant number of utterances in the puzzle activity would begin with a voiceless stop consonant and would, thus, be included in the analysis. However, due to the high percentage of relative pronouns in Spanish that are voiceless stop-initial, the Spanish interview data managed to accumulate many tokens. More information on this will be presented at the beginning of the following chapter.

Methodological and linguistic variables

A mixed effects linear regression model was performed in R via the lmer package (Kuznetsova et al., 2017). The model included VOT duration per token as the dependent variable as well as language, task, POA and speech rate as independent variables, with the first three in a 3-way interaction.[16] Word and subject were included as random intercepts. The model was later past through the `mixed()` function of the afex package to compute the *p*-values for each individual fixed effect and the *p*-values for each interaction within the mixed model (Singmann et al., 2016). Recall that word order for the cued switching tasks was not analyzed as an independent variable, but a random intercept of word was used as a proxy for such an analysis.

Speaker-dependent and environmental social variables

In addition to the standard linear regression model performed and analyzed for the methodological and linguistic factors, which was described above, a second model was performed for the social variables under examination. The second model also analyzed VOT productions per token as the dependent variable and a wide array of social factors as independent

[16]Based on the reported literature, there was no justified motivation for including speech rate in an interaction with any other independent variable; therefore, it was added to the model as a separate effect for model simplicity purposes (Theodore et al., 2009).

variables. The social variables under examination fall under one of the following categories: language use, language attitudes, linguistic proficiency, and linguistic history, directly taken from the BLP survey; the answer to each question was included as an individual variable. Furthermore, the model includes the dominance score obtained by the BLP questionnaire. Finally, categorical variables were converted into binary indicator variables (or dummy variables) through one-hot encoding to convert the values into numeric input data for analysis, which means that those with more than two levels expanded the data set into several more columns, even after dropping the first dummy variable column from the expanded data set. When using one-hot encoding, it is common practice to drop one of the binary variables to avoid a situation known as the "dummy variable trap." The dummy variable trap occurs when there is perfect multicollinearity between the binary variables, which means that one of the binary variables can be predicted perfectly from the others.[17]

Furthermore, during the BLP demographic/linguistic survey, participants were additionally asked to report any "accented" speech varieties they were in constant contact with; in particular, they were asked to specify those varieties that are influenced by a different language. For example, if they often heard Japanese-accented English or American English-accented Spanish, they were asked to report that. Those data were later standardized (for instance, some participants reported Indian-accented or Hindu-accented English; both responses were reprocess as the same response). After the standardization took place, all unique responses were converted to dummy variables as well; that is, each category became a single column, and those who reported hearing it often received a value of 1 for said column or a value of 0 otherwise.

This process undoubtedly led to the creation of a wide data set with many variables to be analyzed, given that some of these factors needed to be repeated for each language. In total, 75 columns were procured from the BLP's 60 total individual questions. To address the issue of a wide data set (i.e., having many more variable columns than is typical in linguistic research), the data were analyzed with a Lasso (Least Absolute Shrinkage and Selection Operator) regularization method. Lasso regularization, or L1 regularization, is a technique used in linear regression analysis to prevent overfitting of a model by adding a penalty term to the loss function (Tibshirani, 1996). The penalty term is proportional to the absolute values of the coefficients of the regression variables. This statistical algorithm works by adding a penalty term to the regression cost function, which is the sum of squared errors between the predicted values and actual values of the dependent variable. The penalty term is the sum of the absolute values of the regression coefficients multiplied by a tuning parameter, which is used to control the amount of regularization applied. The effect of the penalty term is to shrink the coefficients of the regression variables towards zero, resulting in a simpler model that is less prone to overfitting. The advantage of Lasso regularization is that it can be used to perform feature selection, as the penalty term has the effect of setting the coefficients of less important variables to zero. This can result in a more interpretable and efficient model (for an exhaustive description of lasso, refer to James et al., 2013).

[17] More details on the one-hot encoding process are provided in chapter 4.

Once the one-hot encoding technique was applied to the data, they were later standardized to achieve a zero-mean and 1-standard deviation form in all continuous variables (that is, those that are not binary such as the dummy variables). The standardization process was completed for the following purposes:

1. Scaling: Standardizing a data set scales the data to have zero mean and unit variance. This can help to avoid issues with different scales and units of measurement for different variables, which can affect the performance of some machine learning algorithms that rely on distance-based calculations.

2. Improved convergence: Some optimization algorithms converge faster on standardized data, as the optimization path becomes smoother and more predictable.

3. Improved model performance: Some machine learning algorithms, such as logistic regression, assume that the data are standardized. Standardizing the data can lead to improved performance and more stable model estimates.

4. Facilitating interpretation: Standardizing the data can facilitate interpretation of the model coefficients. When the data are standardized, the coefficients of the model reflect the change in the dependent variable associated with a one standard deviation change in the independent variable.

To reiterate, standardization was performed to the data set to improve the performance, convergence, and interpretation of the machine learning model. It was also intended to help avoid issues with different scales and units of measurement for different variables, which can affect the performance of some machine learning algorithms.

Since the model was fed a high number of variable columns, the data set was further split into three sets: training, validation, and test sets, with 60%, 20%, and 20% of the data in these sets, respectively. This data splitting technique was performed following standard practice in the field of machine learning. This technique led to the following benefits:

1. Model selection: The training set is used to train the model, while the validation set is used to evaluate the performance of different models and select the best one. By comparing the performance of different models on the validation set, we can choose the model with the best generalization performance and avoid overfitting.

2. Hyperparameter tuning: Hyperparameters are parameters of a model that are not learned during training, such as the learning rate or regularization strength. The validation set can be used to tune these hyperparameters and find the optimal values that result in the best performance.

3. Performance estimation: Once the final model is selected and hyperparameters are tuned, the test set is used to estimate the performance of the model on unseen data. This provides a more realistic estimate of the model's performance and helps avoid overestimating the performance of the model.

4. Avoiding data leakage: By keeping the test set separate from the training and validation sets, we can ensure that the model has not "seen" the test data during training or tuning. This helps avoid data leakage and provides a more unbiased estimate of the model's performance when tested on truly unseen data.

Data splitting is a crucial step in developing and evaluating machine learning models, as it helps ensure that the model performs well and allows for generalization beyond the training data set, as was the case with the unseen test data set. In the model performed here, three hyperparameters were tuned via GridSearchCV on scikit-learn: (1) penalty (range: 0.1 to 1; the L1 term that controls the regularization strength), (2) maximum number of iterations (range: 500 to 2000; how many iterations to run before determining if the model did not converge), and (3) selection type ("cyclic" versus "random;" whether we move through variables orderly or randomly). The one-hot encoding, scaling, data splitting, hyperparameter tuning, and model training and testing for the social variables were all conducted using the scikit-learn (sklearn) software (Pedregosa et al., 2011).

While the current section provided a brief introduction of each of these terms and their benefits, chapter 4 will revisit these concepts and provided more detailed descriptions of how they were implemented in the second model (i.e., the social factors). This dissertation places additional emphasis on the Lasso model due to the fact that this statistical algorithm is not commonly used in linguistic analyses, despite the fact that it provides accurate and generalizable results, and that is has been in existence since the mid 90's.

Finally, the social variables were all split into two categories, one for speaker internal factors and another for environmental factors. The classification into either category was made following my general judgment of each variable–whether the factor is an inherent characteristic of the speaker (e.g., age of acquisition or attitude towards a given language) or a factor that may depend on social norms or expectations (e.g., whether a social environment or interlocutors may influence language usage practices).[18] The latter is often out of the control of the speaker themselves (e.g., what language to use in certain contexts such as school or the workplace). Refer to appendix F for a complete list of the social factors divided into the two categories.

[18]While the model was trained and tested using data sets that contained both types of social factors, after the results were returned from the model, each category was analyzed and interpreted individually.

Chapter 4

Results and discussion

Note that, in general, speakers produced more voiceless stop-initial tokens in Spanish than English, primarily due to the higher number of tokens in the Spanish interview portion, compared to the English portion (refer back to table 3.7). In total, 6,314 English tokens were produced, compared to a total of 8,255 tokens in Spanish. Table 4.1 displays the total count, mean, standard deviation, the minimum value, the one-, two-, and three-quarter percentiles, and the maximum value for the data in each language. The difference was in particular noticeable in the free-response interview task, in which participants used a remarkably higher number of non-cognate, voiceless stop-initial words in Spanish than in English. This is in part due to the fact that Spanish has a higher number of high frequency relative pronounce that are non-cognate, voiceless stop-initial. For instance, on the one hand, the top three most common words in Spanish were *pero* ("but"; 434), *que* ("that"; 426), and *como* ("how" or "like"; 350), totaling 1,210 tokens. On the other hand, the top three most frequent words in English were *people* (303), *talking* (179), and *taking* (123), totaling 605 tokens. Nonetheless, the data for each language was somewhat normally distributed, as show in figure 4.1. While there is some overlap in the VOT measurements obtained for both languages, the data are nevertheless centered around each language's expected mean, according to the values reported in chapter 2.

Tables 4.2 and 4.3 further detail the frequency distribution of produced lexical items in English and Spanish by displaying the top ten most frequent words for each language, respectively. Notice that in English, many items from the top 10 list are part of the stimulus lists for the word list reading task or the passage reading tasks. This outcome suggests that, unlike Spanish data, English productions for this corpus are composed of a wider spread of lower frequency lexical items. However, a quick look at the corpus reveals that across all four activities, English has a total of 333 unique non-cognate lexical items found in utterance initial position after a switch. On the other hand, Spanish has a total of 332 lexical items that fit the same description. Thus, the data show a virtually identical spread of unique lexical item between the two languages, with the only difference being the higher word frequency peak for the Spanish productions driven by the voiceless stop-initial relative pronouns in Spanish, as indicated above.

	English	**Spanish**
Count	6314	8255
Mean	69.94 ms	24.64 ms
Standard deviation	25.22 ms	10.58 ms
Minimum	15 ms	4 ms
25% percentile	51 ms	16 ms
50% percentile	68 ms	22 ms
75% percentile	87 ms	31 ms
Maximum	130 ms	52 ms

Table 4.1: Descriptive statistics for English and Spanish VOT data

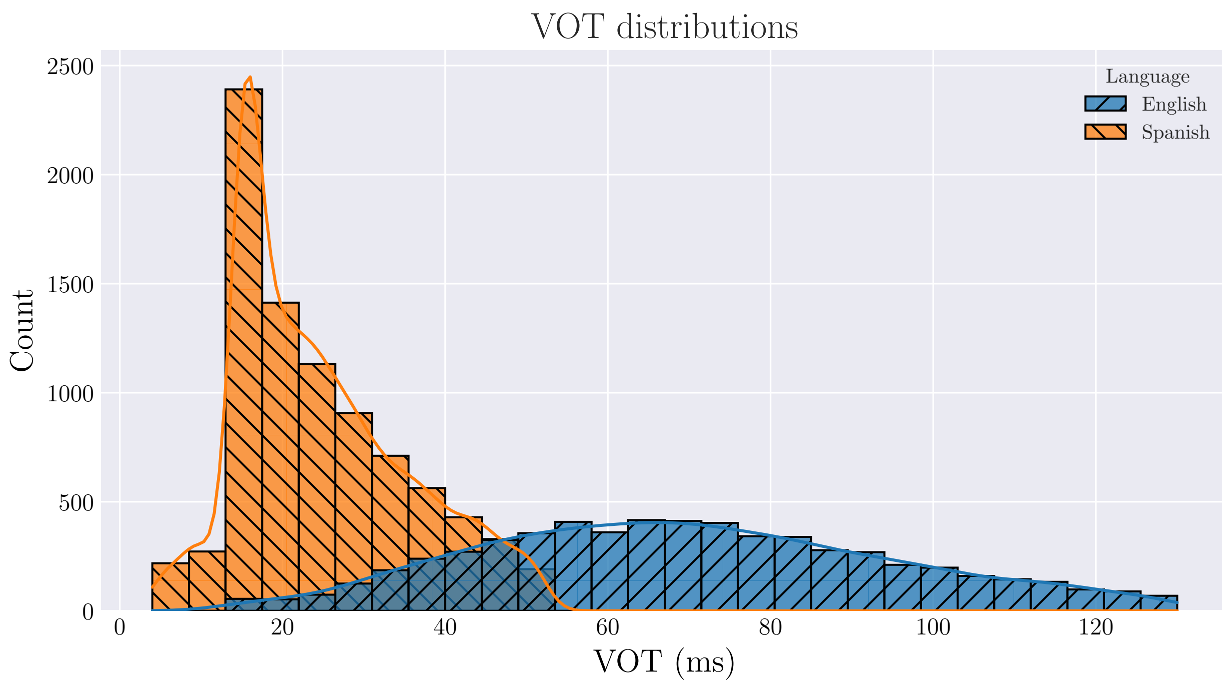


Figure 4.1: Distribution of VOT measurements for English and Spanish data

The current chapter is divided into two sections: (1) methodological (i.e., research task) and linguistic factors (i.e., language, POA, and speech rate) and (2) social factors (i.e., speaker-inherent and environmental variables) from the BLP survey. Each of these two sections will begin with the results from the statistical models, followed by a discussion of those results. Such an organization was followed to maximize clarity of this chapter as a whole. Note that an additional sub-subsection was added to the methodological and linguistic section in order to address a concern regarding the possibility of cross-linguistic

Word	Frequency
people	303
talking	179
taking	123
coming	122
kids	120
to	95
keyboard	88
two	76
pool	66
cool	65

Table 4.2: Top 10 English words

Word	Frequency
pero	433
que	425
como	348
también	296
para	245
cuando	236
con	222
cada	184
tiene	175
porque	164

Table 4.3: Top 10 Spanish words

resyllabification in the passage data. More details on this will be provided within the section itself.

4.1 Methodological and linguistic factors

As a reminder, the statistical analyses for the methodological and linguistic factors were performed jointly using R. For visualization purposes, the model performed with the lmerTest package is provided below:

– `lmer(VOT ~ Lang * Task * POA + Sp rate + (1|subject) + (1|word))`

I will now proceed to presenting the results returned by the statistical model above.

Results

After the aforementioned statistical model was generated, the model itself was fed to the `mixed()` function of the afex package[1] using the Satterthwaite approximation in order to obtain the p-values of each fixed effect and interaction. In addition, the calculations returned the numerator and denominator degrees of freedom (i.e., number of parameters used for the effect and effective residual degrees of freedom for testing the effect, respectively) and the scaling of F-statistic computation. The results from this calculation are presented in an anova-style format in table 4.4.

[1]As described in the documentation of the package, "mixed() fits mixed models using lme4::lmer() and computes p-values for all fixed effects using either Kenward-Roger or Satterthwaite approximation for degrees of freedom (LMM only), parametric bootstrap (LMMs and GLMMs), or likelihood ratio tests (LMMs and GLMMs)" (Singmann et al., 2016, p. 3).

Effect	**df**	**F**	***p*-value**
Language	1, 379.16	2016.32	<0.001
Task	3, 619.42	8.76	<0.001
POA	2, 316.41	28.08	<0.001
Speech rate	1, 3536.93	325.10	<0.001
Language:Task	3, 603.78	8.56	<0.001
Language:POA	2, 439.93	7.17	<0.001
Task:POA	6, 576.18	3.79	0.001
Language:Task:POA	6, 586.17	1.89	0.081

Table 4.4: `Mixed()` results for methodological and linguistic fixed effects

In short, these results indicate that each of the individual independent variables (i.e., language, task, POA, and speech rate) is statistically significant. In other words, English VOT measurements differed from Spanish in that English productions yielded higher VOT values than Spanish (ndf = 1, ddf = 379.16, F = 2016.32, p-value < 0.001; refer back to figure 4.1 for a visualization of the data distribution). Moreover, these results also indicate that VOT values for at least one task differ from the rest (ndf = 3, ddf = 619.42, F = 8.76, p-value < 0.001), a result that is also present for POA (ndf = 2, ddf = 316.41, F = 28.08, p-value < 0.001), meaning that at least one POA category differs from the rest; both of these results will be unpacked and addressed shortly. Finally, speech rate, as a single main effect, was also found to be significant (ndf = 1, ddf = 3536.93, F = 325.10, p-value < 0.001), indicating that there is a positive relationship between speech rate and VOT productions; as one variable increases, so does the other. Refer to figure 4.2 for a visualization of this relationship. Interactions between language and task (ndf = 3, ddf = 603.78, F = 8.56, p-value < 0.001), language and POA (ndf = 2, ddf = 439.93, F = 7.17, p-value < 0.001), and task and POA (ndf = 6, ddf = 576.18, F = 3.79, p-value = 0.001) were also found by the statistical model. While the 3-way interaction approached significance, it did not meet the threshold p-value of 0.05 (ndf = 6, ddf = 586.17, F = 1.89, p-value = 0.081).

The summary from the lmer statistical model with comparisons across variable levels is provided in table 4.5, including variable level names, estimates, standard error, degrees of freedom, t-values, and p-values; note that the baseline levels are language: English; task: interview; and POA (or stop): k.

Considering that both the task and POA variables have more than two levels and each of them is found to interact with language, additional post-hoc analyses were in order. The post-hoc tests were conducted using emmeans (Lenth, 2023). As such, emmeans tests with an adjusted p-value calculations were performed. The pairwise comparison with a tukey adjustment between task and language returned a consistent difference between language across all four tasks: interview productions yielded longer VOT in English than Spanish (coeff = 36.837, SE = 1.577, z ratio = 23.354, p-value < 0.0001), word list productions

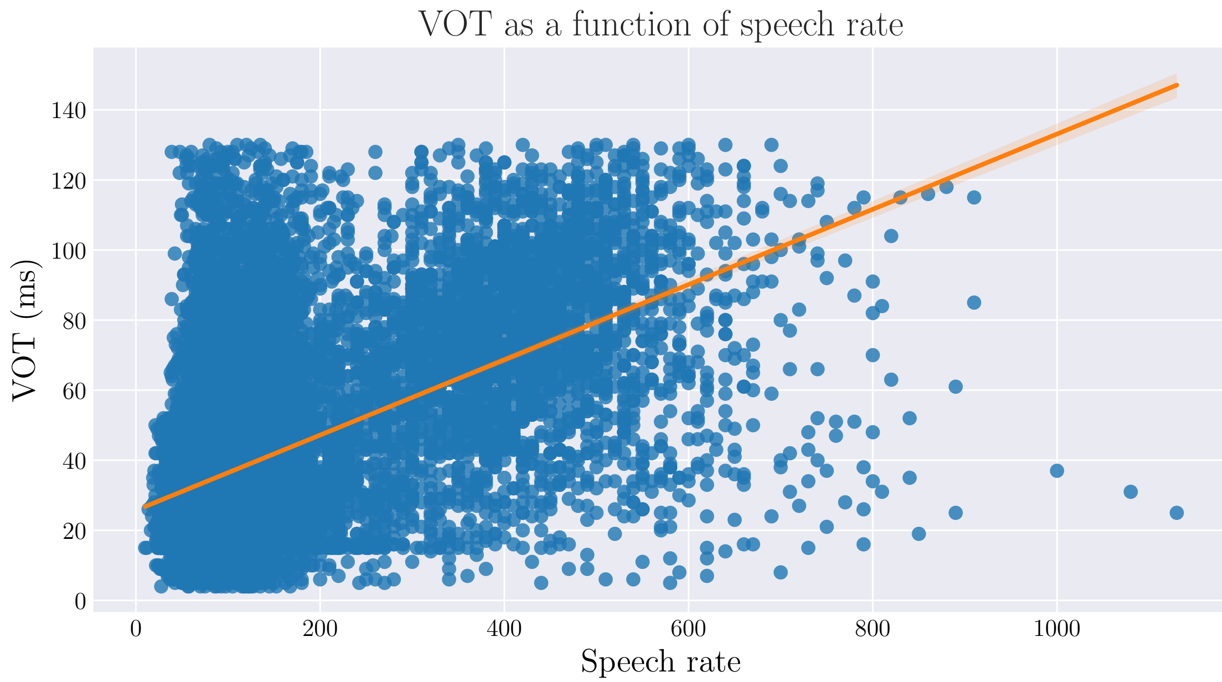


Figure 4.2: VOT values across speech rate

yielded longer VOT in English than Spanish (coeff = 40.814, SE = 1.100, z ratio = 37.109, p-value < 0.0001), passage productions yielded longer VOT in English than Spanish (coeff = 33.340, SE = 1.504, z ratio = 22.161, p-value < 0.0001), and puzzle productions yielded longer VOT in English than Spanish (coeff = 41.198, SE = 1.105, z ratio = 37.294, p-value < 0.0001).

Given that a language distinction in VOT productions has been consistently reported in the literature and that such distinctive pattern was corroborated in these results, the statistical language difference for each task just reported is unsurprising. However, the within language comparisons reveal the source of the interaction between language and task. For example, on the one hand, for Spanish data, we find no statistical difference across tasks: interview versus word list (coeff = -2.039, SE = 0.949, z ratio = -2.148, p-value = 0.3839), interview versus passage (coeff = -2.142, SE = 0.807, z ratio = -2.653, p-value = 0.1372), interview versus puzzle (coeff = -0.810, SE = 0.695, z ratio =-1.166 , p-value = 0.9416), word list versus passage (coeff = -0.103, SE = 1.048, z ratio = -0.098, p-value = 1.0), word list versus puzzle (coeff = 1.229, SE = 0.974, z ratio = 1.263, p-value = 0.9126), and passage versus puzzle (coeff = 1.332, SE = 0.873, z ratio = 1.527, p-value = 0.7933). On the other hand, English data do present a task difference: interview productions yielded shorter VOT than word list (coeff = -6.017, SE = 1.589, z ratio = -3.787, p-value = 0.0038), interview productions did not differ from passage (coeff = 1.355, SE = 1.775, z ratio = 0.763, p-value = 0.9949), interview productions yielded shorter VOT than puzzle (coeff = -5.171, SE =

	Estimate	**SE**	**df**	***t*-value**	***p*-value**	
(Intercept)	0.5668	2.561	0.0044	22.130	$<$ 2e-16	***
Lang_Spa	-0.3413	2.725	0.0032	-12.524	$<$ 2e-16	***
Task_List	0.1303	2.739	0.0021	4.757	2.09e-06	***
Task_Pass	4.553	3.072	0.0949	1.482	0.1386	
Task_Puzz	8.481	2.659	0.0051	3.190	0.0014	**
Stop_p	-3.793	3.618	0.0042	-1.048	0.29455	
Stop_t	4.246	3.247	0.0043	1.308	0.19098	
Spaeech rate	0.0342	0.0019	0.0034	18.052	$<$ 2e-16	***
Lang_Spa:Task_List	-8.082	3.067	0.0016	-2.636	0.0085	**
Lang_Spa:Task_Pass	-1.341	3.365	0.0012	-0.398	0.69035	
Lang_Spa:Task_Puzz	-7.776	2.808	0.0058	-2.769	0.00564	**
Lang_Spa:Stop_p	0.0573	3.984	0.0028	0.014	0.9885	
Lang_Spa:Stop_t	-8.190	3.684	0.0028	-2.223	0.0263	*
Task_List:Stop_p	-0.1005	4.017	0.0017	-2.501	0.01246	*
Task_Pass:Stop_p	-0.1096	4.553	0.0857	-2.407	0.0163	*
Task_Puzz:Stop_p	-2.827	3.899	0.004	-0.725	0.46845	
Task_List:Stop_t	-0.1099	3.668	0.0015	-2.998	0.00277	**
Task_Pass:Stop_t	-6.764	4.104	0.0965	-1.648	0.09965	.
Task_Puzz:Stop_t	-7.101	3.442	0.0057	-2.063	0.0392	*
Lang_Spa:Task_List:Stop_p	5.766	4.602	0.0011	1.253	0.21047	
Lang_Spa:Task_Pass:Stop_p	9.519	4.897	0.001	1.944	0.05220	.
Lang_Spa:Task_Puzz:Stop_p	2.034	4.173	0.0046	0.487	0.62595	
Lang_Spa:Task_List:Stop_t	6.549	4.271	0.0012	1.533	0.12542	
Lang_Spa:Task_Pass:Stop_t	4.995	4.611	0.0012	1.083	0.27886	
Lang_Spa:Task_Puzz:Stop_t	8.210	3.842	0.0068	2.137	0.03266	*

Table 4.5: lmer statistical model summary

1.496, z ratio = -3.456, p-value = 0.0128), word list productions yielded longer VOT than passage (coeff = 7.372, SE = 1.498, z ratio = 4.922, p-value $<$ 0.0001), word list productions did not differ from puzzle (coeff = 0.846, SE = 1.027, z ratio = 0.823, p-value = 0.9919), and passage productions yielded shorter VOT than puzzle (coeff = -6.526, SE = 1.411, z ratio = -4.626, p-value = 0.0001). The results for the language-task interactions for the VOT measurements are visualized in figure 4.3.

To recapitulate, while no differences were revealed by the post-hoc analysis for Spanish data, English data show a statistical difference between the list and interview/passage activities and between the puzzle interview/passage activities. The post-hoc model did not return a difference between the list and puzzle activities or between the interview and passage activities. A summary for the language-task post-hoc analysis model results for the

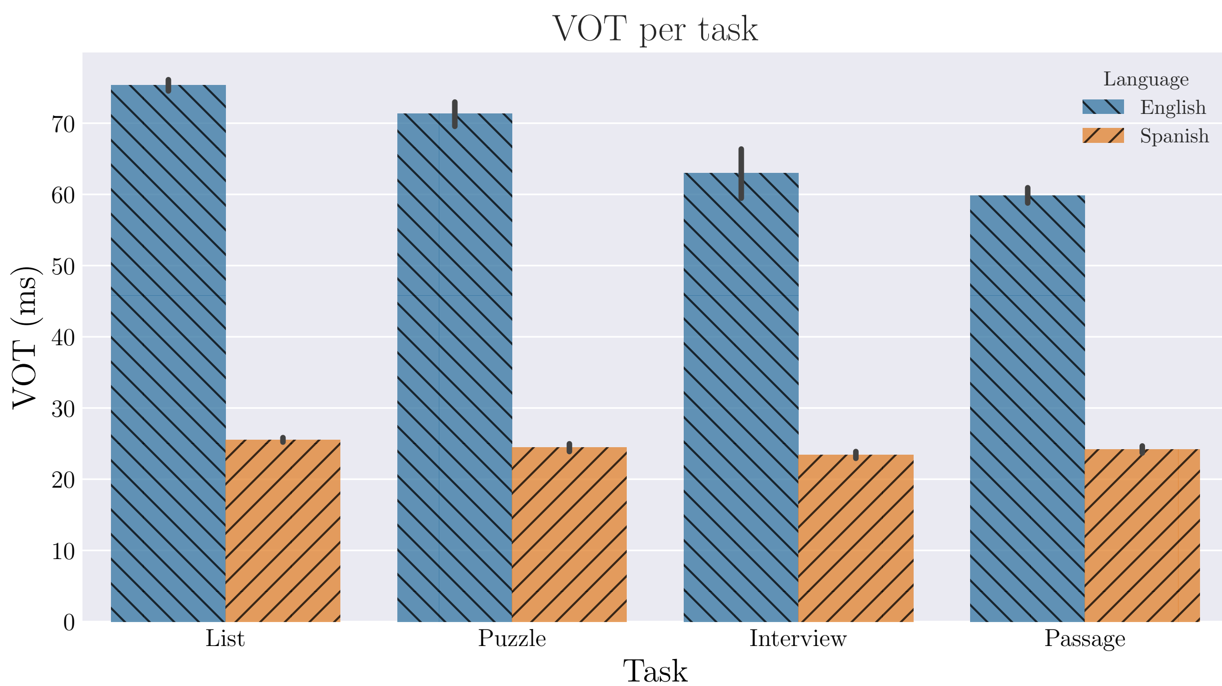


Figure 4.3: VOT averages per task for English and Spanish data

English data is provided in table 4.6 and a corresponding summary for the Spanish data is provided in table 4.7.

	Estimate	**SE**	**z ratio**	***p*-value**	
Interview-List	-6.017	1.589	-3.787	0.0038	**
Interview-Passage	1.355	1.775	0.763	0.9949	
Interview-Puzzle	-5.171	1.496	-3.456	0.0128	*
List-Passage	7.372	1.498	4.922	<.0001	***
List-Puzzle	0.846	1.027	0.823	0.9919	
Passage-Puzzle	-6.526	1.411	-4.626	0.0001	***

Table 4.6: English language-task post-hoc summary

Moreover, the pairwise comparison with a tukey adjustment between place of articulation and language also returned a consistent difference between language across the three POA's: English bilabial productions yielded longer VOT than Spanish bilabial (coeff = 34.039, SE = 1.43, z ratio = 23.743, p-value < 0.0001), English coronal productions yielded longer VOT than Spanish coronal (coeff = 41.677, SE = 1.37, z ratio = 30.439, p-value < 0.0001), and English velar productions yielded longer VOT than Spanish velar (coeff = 38.426, SE = 1.34,

	Estimate	SE	z ratio	*p*-value
Interview-List	-2.039	0.949	-2.148	0.3839
Interview-Passage	-2.142	0.807	-2.653	0.1372
Interview-Puzzle	-0.810	0.695	-1.166	0.9416
List-Passage	-0.103	1.048	-0.098	1.0000
List-Puzzle	1.229	0.974	1.263	0.9126
Passage-Puzzle	1.332	0.873	1.527	0.7933

Table 4.7: Spanish language-task post-hoc summary

z ratio = 28.585, p-value = < 0.0001), as expected, indicating that the language difference in VOT production patterns exists across POA.

Nonetheless, as a source of the language-POA interaction, the model results reveal that there are some language-specific differences in the ways POA patterns. For example, on the one hand, among English data we find that bilabial stops yielded shorter VOT than coronal stops (coeff = -7.784, SE = 1.46, z ratio = -5.313, p-value < 0.0001) and than velar stops (coeff = 9.752, SE = 1.47, z ratio = 6.616, p-value < 0.0001). However, coronal and velar stops did not display a statistical difference (coeff = 1.968, SE = 1.38, z ratio = 1.428, p-value = 0.7100). On the other hand, we find a reverse pairing among Spanish data, where bilabial stops yielded shorter VOT than velar stops (coeff = 5.365, SE = 1.28, z ratio = 4.188, p-value = 0.0004) but did not differ from coronal stops (coeff = -0.145, SE = 1.31, z ratio = -0.111, p-value = 1.0); coronal stops, in this case, do differ from velar stops (coeff = 5.220, SE = 1.29, z ratio = 4.040, p-value = 0.0008), though. A joint summary of the English and Spanish POA post-hoc comparisons can be found in table 4.8, and the results for the POA comparisons for both languages are illustrated in figure 4.4

	Estimate	SE	z ratio	*p*-value	
English p - Spanish p	34.039	1.43	23.743	<0.0001	***
English t - Spanish t	41.677	1.37	30.439	<0.0001	***
English k - Spanish k	38.426	1.34	28.585	<0.0001	***
English p - English t	-7.784	1.46	-5.313	<0.0001	***
English k - English p	9.752	1.47	6.616	<0.0001	***
English k - English t	1.968	1.38	1.428	0.7100	
Spanish p - Spanish t	-0.145	1.31	-0.111	1.0000	
Spanish k - Spanish p	5.365	1.28	4.188	0.0004	***
Spanish k - Spanish t	5.220	1.29	4.040	0.0008	***

Table 4.8: English and Spanish POA post-hoc summary

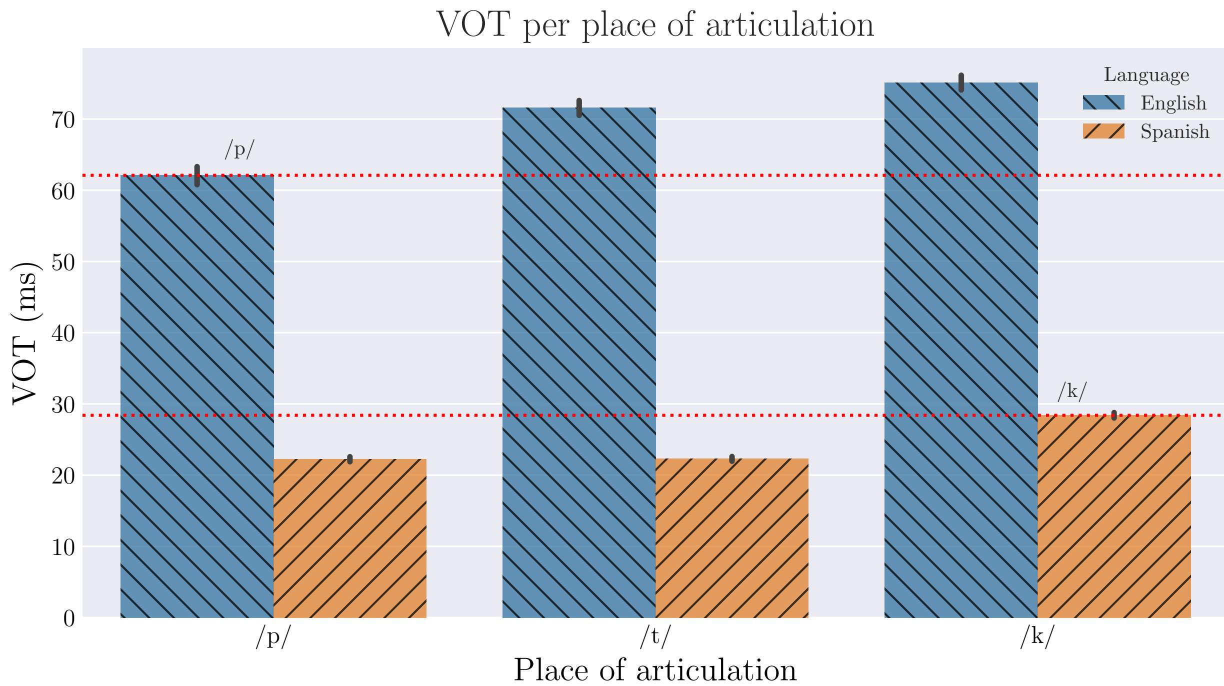


Figure 4.4: VOT distributions per place of articulation for English and Spanish data

Because a significant difference was not found for the 3-way interaction between language, task, and POA,[2] a post-hoc analysis was not performed for this comparison. We will now move on to the discussion of the above-reported statistical results for the methodological and linguistic factors mediating VOT productions in language alternation.

Discussion

To summarize the contents of the previous section, the analysis for the methodological and linguistic factors used a mixed-effects linear regression model implemented in R, specifically using the lmerTest package. This model included several independent variables: language, task, POA, and speech rate, with the first three in a 3-way interaction while speech rate remained a standalone main effect. The results of the statistical model indicate that each of the independent variables (language, task, POA, and speech rate) has a statistically significant effect on VOT measurements. Specifically:

1. Language: English VOT measurements differed significantly from Spanish. English productions had higher VOT values than Spanish (p-value < 0.001).

[2]The results came near to but did not reach the p-value threshold of 0.05.

2. Task: VOT values for at least one task differed significantly from the rest (p-value $<$ 0.001).

3. POA: VOT values for at least one POA category differed significantly from the rest (p-value < 0.001).

4. Speech rate: There was a significant positive relationship between speech rate and VOT productions. As speech rate increased, VOT values also increased (p-value $<$ 0.001).

Additionally, the statistical model revealed significant interactions between the independent variables. Given that language interacted with task and POA, these two variables were not analyzed as standalone main effects in post-hoc analyses; only post-hoc analyses of the interactions were performed to further examine the differences between variable levels while separating languages:

1. Language and task: There was a significant interaction between language and task (p-value < 0.001), indicating that the effect of language on VOT measurements varied depending on the task being performed. For the language-task interaction, English data showed significant differences between the list and interview/passage activities and between the puzzle interview/passage activities. In other words, the word list reading and puzzle completion activities yielded longer VOT than interview and passage reading activities. No differences were found for Spanish data.

2. Language and POA: There was a significant interaction between language and POA (p-value < 0.001), indicating that the effect of language on VOT measurements varied depending on POA. For the language-POA interaction, there were inconsistent POA pairings in English and Spanish for the three POA categories: bilabial, coronal, and velar. While English showed that /p/ $<$ /t k/, Spanish showed that /p t/ $<$ /k/ in VOT measurements.

3. Task and POA: There was a significant interaction between task and POA (p-value $= 0.001$), indicating that the effect of task on VOT measurements varied depending on the place of constriction. However, since these results are averaged across language and language differences constitute the crux of findings, no substantive interpretation of this result was pursued.

4. Language, task, and POA: There was a 3-way interaction between language, task, and POA, which approached significance but did not meet the threshold p-value of 0.05 (p-value $= 0.081$).

Therefore, the results indicate that there is a difference in VOT productions between word list and puzzle activities compared to interview and passage tasks for English data, as shown in figure 4.5. Moreover, figure 4.6 exhibits the larger divergence of VOT means for each of the four tasks in English compared to the English-general mean for all activities, compared to the smaller divergence observed in the VOT means for each of the four tasks in Spanish compared to the Spanish-general mean for all activities, as seen in figure 4.7. These figures illustrate the statistical results stated above.

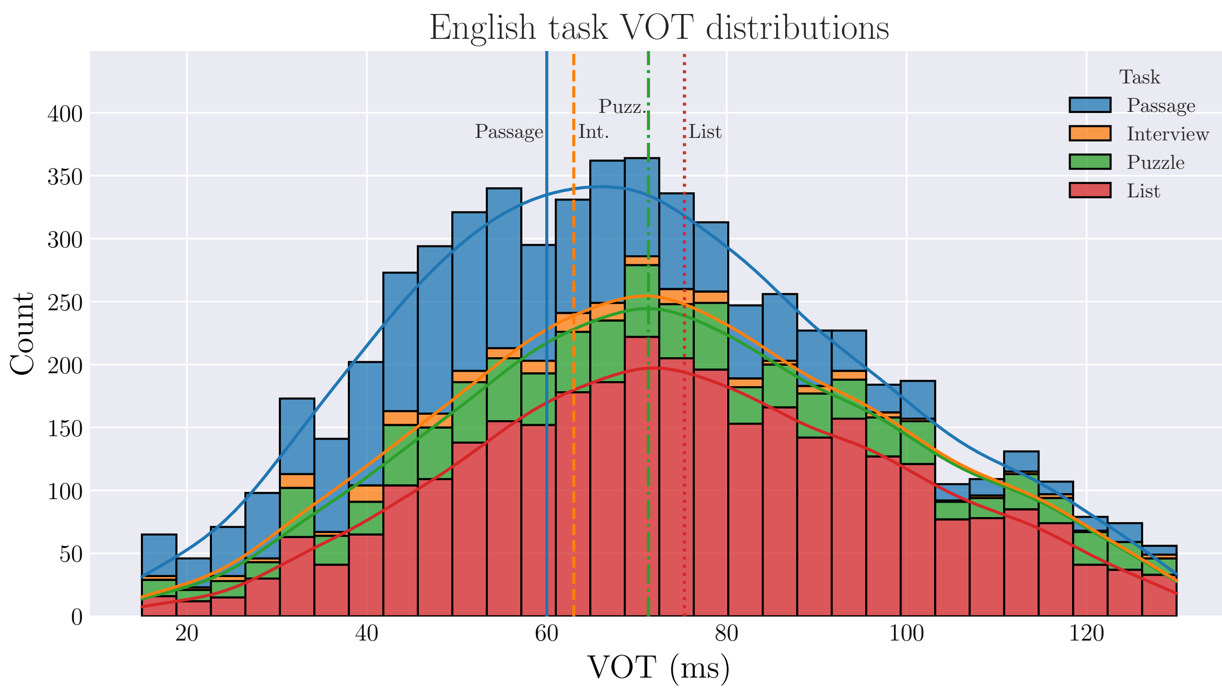


Figure 4.5: VOT distributions and means per task for English data

Now, the questions that arise are:

1. Why is this effect only observed in English data?
2. Why do the passage reading and interview tasks display higher levels of convergence (i.e., the VOT ranges between languages are closer together) compared to the puzzle completion and word list reading tasks?

To address the first question, recall that in cases of unidirectional convergence, it is more common for English (or the long-lag language) to be the converging language (Antoniou et al., 2011; Balukas & Koops, 2015; Bullock et al., 2006; Olson, 2016). This is precisely the pattern that is predicted by the phonetic latitude hypothesis proposed by Bullock and Toribio (2009c). In other words, given the acoustic distance in the aspiration range between English voiceless stops and (a) their voiced counterparts and (b) their Spanish voiceless peers, Spanish-English bilinguals have more leeway to shift their phonetic productions of this category without causing phonological ambiguity, even while alternating languages (also

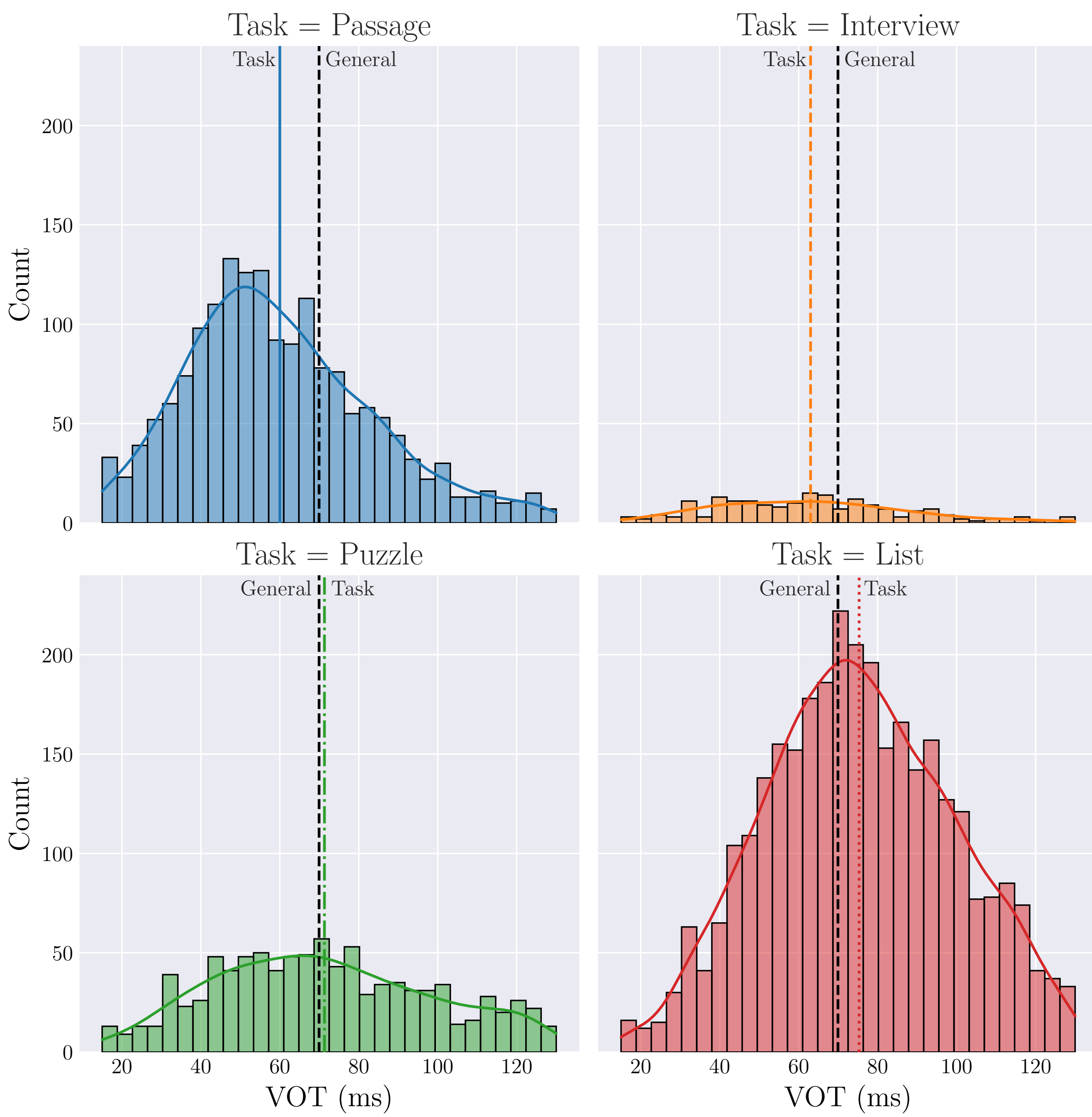


Figure 4.6: VOT general and task means for English data

see Olson, in press). In short, the results from the current study corroborate the results previously reported in the literature, as well as the notion of phonetic latitude mediating convergence patterns in language alternation speech patterns.

Now we turn to the second–and much more complex–question. First, let's revisit the theoretical frameworks described in chapter 2. First, Communication Accommodation Theory

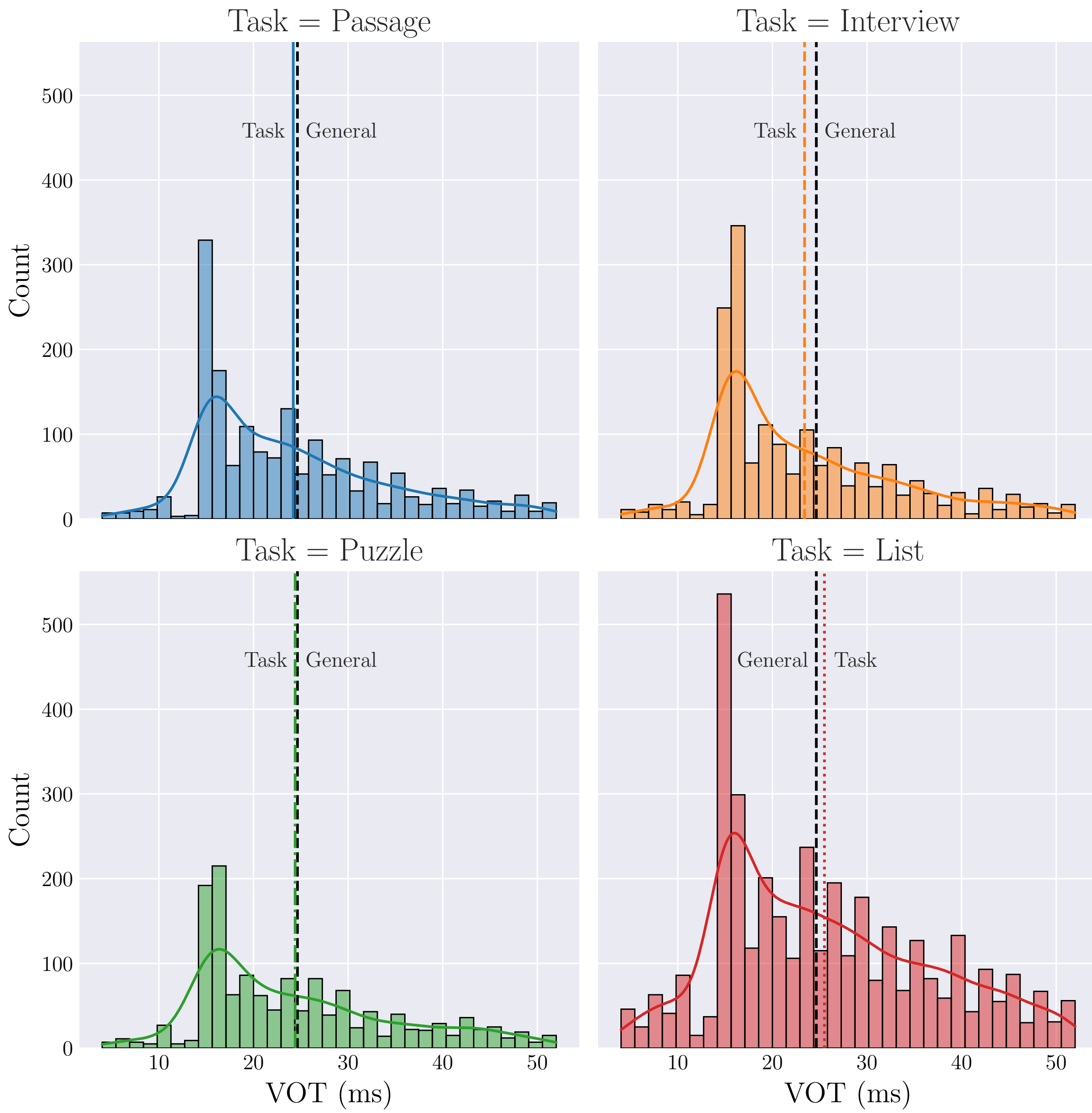


Figure 4.7: VOT general and task means for Spanish data

(CAT; Giles, 1973; Giles and Ogay, 2007), examines how individuals adjust their communication to align with or differentiate themselves from their interlocutors. According to this theory, people engage in two main accommodation strategies: convergence and divergence. Convergence involves adapting speech patterns to match the other person's style, fostering social harmony and similarity. Divergence, on the other hand, shows a drifting of speech behaviors emphasizes distinctiveness and differences. Accommodation can occur in various aspects of communication, including speech rate, accent, vocabulary, and nonverbal behaviors. The theory suggests that accommodation is influenced by social identity, power dynamics, and communicative goals, ultimately affecting social perceptions and relationship outcomes.

Similarly, audience design (Bell, 1984) proposes considering the stylistic dimension in linguistic variation in relation to the attributes of the hearers instead of the speaker. It suggests that speakers primarily respond to other individuals and consider the audience when planning their speech. The characteristics of the first person (the speaker) account for differences in speech between speakers, but the style is tailored for the audience. For a single speaker, differences can be attributed to the influence of the second person (the addressee) and some third persons, who collectively form the audience for the speaker's utterances.

If we are to incorporate the audience design and CAT frameworks as an explanation for the linguistic patterns observed in the methodological results (that is, the results that involve task-based productions), specifically, we would expect to see, first of all, equal–or equivalent–patterning for both languages, an expectation that does not materialize in this corpus. To put it in other words, because all subjects interacted with an interlocutor–the researcher–for the duration of the data collection session in both languages, it would be expected that all subject would accommodate to the interlocutor in the same way in both languages, whether it is by converging their speech to decrease their social distance or vice versa. This expectation arises from the models' postulations that the speakers evaluate their relationship to the interlocutor and act accordingly. In this case, because all subjects and the researcher were competent Spanish-English bilinguals with a shared experience of living and interacting in both languages inside the United States, it is not anticipated that subjects would behave differently in one language or the other in a contradictory manner. As such, these theoretical frameworks are unable to account for the distinct linguistic behaviors observed in the two languages under examinations.

In addition, we also find unexpected pairings of experimental task in the results of this study. Namely, the passage reading and the casual interview tasks are found to display more convergence than the puzzle completion and the word list reading tasks. Nonetheless, the interview and puzzle activities are the only ones that required and facilitated an interaction between the subjects and the researcher. Therefore, if interlocutor accommodations were driving VOT convergence behavior in language alternation, we would expect for the interview and puzzle activities to pair up together and not to yield a statistical difference in the results of these two tasks. Yet, we find results that undermine the argument that speaker-interlocutor accommodation leads to convergence or divergence of VOT productions in this particular corpus.

All in all, the CAT and audience design theoretical frameworks are unable to account for the VOT convergence patterns presented above. While these models are valuable in understanding language variation and the influence of social factors, they appear to have limitations when it comes to explaining phonetic convergence effects observed in language alternation data. These theories primarily focus on the social and stylistic dimensions of language variation and the role of speakers in adapting their speech to accommodate their audience. However, with regard to language alternation, phonetic convergence observed in certain in methodological and linguistic contexts is seemingly more dependent on cognitive factors that propel this effect.

Following the discussion above regarding VOT convergence in language alternation and CAT and audience design, we now turn to the "attention to speech" factor (Labov, 2011) in search of a model that can provide an explanation for the aforementioned linguistic patterns. Labov's attention to speech factor refers to the idea that individuals' level of attention and awareness of their speech affects their linguistic behavior. According to Labov, when individuals are more conscious of their speech, such as in formal or monitored situations, they tend to exhibit an increase in more standard or prestigious linguistic forms. In contrast, in casual or less monitored settings, their speech may present an increased rate of more vernacular/non-standard features. In this particular case for this study, we expect the most formal or monitored situations (i.e., reading the word list–shorter text–followed by reading the passages–longer text) to yield lower levels of convergence in a way that VOT productions would maintain the separate categories expected–and reported–in monolingual speech, even by multilingual speakers. Conversely, situations that would give rise to the least amount of self-monitoring (i.e., completing the puzzle task followed by participating in the casual interview) are expected to bring about more convergence, as speakers would be less vigilant about maintaining the separate VOT categories for the voiceless stops. In short, the attention to speech factor emphasizes the role of attention and social context in shaping linguistic variation and highlights the dynamic nature of language usage.

This attention to speech factor sheds light on the dynamic nature of language variation and the influence of social context on linguistic behavior. It suggests that language use is not static but rather adapts to the perceived expectations and scrutiny by the audience. Labov's research on attention to speech highlights the role of awareness and social factors in shaping linguistic variation, providing valuable insights into the relationship between language, social context, and individual consciousness. Nonetheless, this concepts differs from CAT and audience design in their underlying principles. On the one hand, the attention to speech factor highlights the influence of social context and the level of scrutiny individuals perceive, with speech becoming more monitored and conforming to standard norms in formal situations. On the other hand, CAT posits that individuals may converge or diverge their speech patterns based on social motivations such as affiliation or differentiation. According to CAT, speakers may adapt their speech to be more similar to their interlocutors to create rapport or emphasize social identity. Audience design suggests that speakers tailor their speech style to their audience, taking into account the characteristics, expectations, and needs of the listener(s). This theory recognizes that speakers consider the social attributes of the hearers and design their speech accordingly. Thus, the former framework focuses on individuals' awareness of their own speech and their subsequent phonetic adjustments as a result of their own monitoring; the latter two postulate that convergence/divergence may be driven by factors such as communication efficiency, social solidarity, or the desire to establish a shared identity. While all three concepts consider social factors in language variation, they differ in their specific emphasis and perspective on the motivations for a speaker's speech adjustments–facilitating communication, emphasizing social distance, or adhering to social expectations based on formality level.

Below is Labov's (2011) attention to speech scale presented in chapter 2, with activities in descending order based on the amount of attention to speech; that is, those activities further down the list yield more attention to speech and thus more monitoring and adjustment.

1. casual speech
2. careful speech
3. group
4. elicited
5. reading text
6. word lists
7. minimal pairs

Following this order, we can rank the amount of attention to speech, in descending order, for the four activities completed in this study as follows:

1. puzzle
2. interview
3. passage
4. word list

Note that speakers are expected to have monitored their own speech less closely in the puzzle activity than the casual interview due to the added cognitive load that results from the puzzle completion itself while speaking. As such we could expect to see more convergence higher up in this second list and less convergence lower down in the list. Nevertheless, the passage reading and interview activities displayed the highest levels of convergence, compared with the puzzle and word list reading activities.

Hence, in these results we find a contradiction to the attention to speech factor ranking in that we expected the puzzle activity to show more convergence than the interview given that it is expected to induce less monitoring of speech due to the added cognitive load of the activity itself. Between the passage and word list reading activities, we do find the expected results, with more convergence in the passage reading task given the less awareness of individual linguistic segments during speech productions expected in longer scripted stimuli. Furthermore, the attention to speech factor predicts that any form of reading activity, including word list and passage reading, should produce lower levels of convergence as a consequence of the increased inherent monitoring levels associated with read speech, a task ranking that does not materialize in this corpus, with a swap of the puzzle and passage activities in the quantitative results.

Thus, taken as a whole, this project produced overall convergence paring and ranking results that are unable to be fully accounted for by the attention to speech rankings. However, unlike the CAT and audience design theories, the attention to speech factor is able to explain one portion of the results: more VOT convergence in the passage reading task than the word list. So the question that arises now is, why do these results not adhere to the rankings of the attention to speech ranking? To answer this question, we need to focus on the major difference that distinguishes the data used for generating the original ranking and the corpus used in this study: the monolingual versus bilingual speech mode. While the data that produced the ranking reported in Labov (2011) and the data used by many other researchers to corroborate the ranking used monolingual speech–primarily in English–my study uses Spanish and English speech produced in language alternation. As such speech mode is a core distinction between the data used for the development of the set ranking and the data from the present study.

As mentioned earlier, language alternation is a speech mode that incorporates additional or distinct processing mechanisms, compared to monolingual speech. With that in mind, my next step is not to diminish the validity of the attention to speech ranking or suggest modifications. Rather, language alternation requires a new ruler. The results from this study that compares the speech in four distinct experimental tasks are but one corpus in one study, and as such they fall short of what is necessary to produce a full-fledged theory of rankings of attention to speech during language alternation. Nonetheless, it provides us with materials to take the first step towards the development of such a theory. Firstly, these results tell us that, among the reading activities, language alternation follows the expected pattern in attention to speech, and thus production of more careful speech, in shorter segments of text, as is the case with the word list task. Put differently, the word list elicitation task resulted in more careful speech as indicated by the lower level of convergence when compared to the passage reading task. However, the level of convergence observed in the passage reading activity surpasses the level reported for both of the code switching tasks, a result that was unexpected. This suggests that, unlike in monolingual speech, in bilingual speech cued switching activities may yield an exponential amount of convergence as attention to speech decreases, rather than following a linear pattern. In other words, as text stimuli get longer in reading tasks (that is, subjects are paying less attention to individual phonetic segments), convergence effects may increase exponentially in bilingual speech. As a result, the convergence effect seen in the passage activity surpasses the magnitude of the effect seen in the puzzle and interview activity.

Secondly, the results for the puzzle and interview tasks are unexpected because, given the additional cognitive load supplied to the subjects by the puzzle activity, the interview task returned a higher level of convergence than the puzzle activity did. This suggests that, while code switching in naturalistic, spontaneous speech, multilingual speakers have higher levels of monitoring of their speech in the puzzle activity, a notion that is contradictory to the idea of cognitive overload created by the puzzle itself. One potential explanation for such an inverse outcome, compared to monolingual speech, is that the increased amount of inhibition observed in bilingual speech is countered by an increase of language processing resources among multilingual speakers. That is, when completing a puzzle in monolingual mode, the activity adds a cognitive load which inhibits a speaker's ability to perform the same amount of self-monitoring of speech production that is present in the casual speech produced during an interview task. However, when completing a puzzle in bilingual mode, the activity adds the cognitive load of the puzzle plus an additional cognitive load from managing two active languages at once. Such an elevated amount of cognitive load in working memory may reach some threshold that initiates the process for expanding the cognitive resources available during speech production for multilingual speakers. Put differently, the more working memory load a multilingual speaker experiences, the more processing resources they may make available during speech processing to an extent that greatly surpasses the amount of processing resources during monolingual speech. That increase could allow speakers to more closely monitor their own phonetic productions during the puzzle activity, compared to the casual interview of the present study.

Finally, this study also analyzed linguistic factors that could mediate the convergence effect observed during language alternation. First of all, speech rate was found to be a significant factor, which corroborates the numerous studies that have reported that as speech becomes slower, VOT productions increase. This indicates that speech rate patterns observed and reported previously also manifest themselves in cued switching or code switching tasks in bilingual mode. The magnitude and robustness of this effect were illustrated in table 4.4 and figure 4.2.

Furthermore, we find that there are language-specific pairings for POA productions in English and Spanish in that in English /t k/ are shown to be produced with higher VOT values than /p/, whereas in Spanish only /k/ is shown to be produced with higher VOT values than /p t/, a pattern that is exhibited in figure 4.8. Given that the literature has reported a /p/ < /t/ < /k/ VOT production pattern in monolingual speech cross-linguistically (Cho & Ladefoged, 1999), this suggests that the pattern seen in this study is the result of an interaction between POA and (bilingual) language mode. That is to say, during language alternation, other factors influence the production of VOT across POA to the point that the VOT length order is interrupted and leads to the patterns seen here. One potential explanation has to do with the exact POA for coronal stops, specifically. Recall that Spanish and English have dental and alveolar stops, respectively (Hualde, 2005; Hammond, 1999). That could suggest that the closer POA between Spanish /p/ and /t/ is the reason for the lack of a difference in the VOT measurements of these two phonetic categories. Nevertheless, there are two complications with that explanation. For once, Cho and Ladefoged (1999) report that the place-dependent VOT values are found cross-linguistically, even in languages with dental articulations. Moreover, while the closer POA between Spanish /p/ and /t/ could potentially explain the pairing of these two sounds in Spanish, it is unable to account for the pairing of /t/ and /k/ in the English data. Thus, a new explanation is necessary to account for the shift of /t/ in both English and Spanish, albeit in different directions.

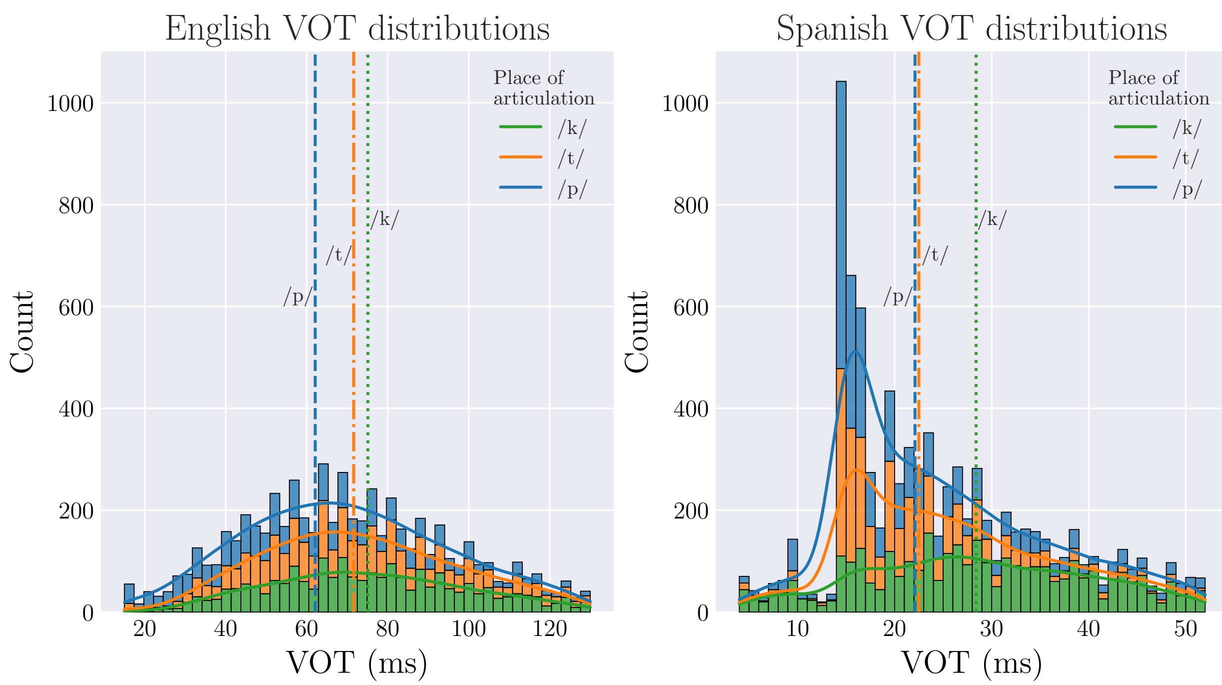


Figure 4.8: VOT distributions and means per POA for English and Spanish data

In a study that investigates the role of social salience in shaping the outcomes of linguistic contact, Erker (2022) aims to understand how salience–or lack thereof–influences linguistic patterns among Spanish-speaking communities in a new linguistic environment. The participants' language use patterns were examined to determine the impact of social salience on language outcomes. The author reports that immigrants who arrived as children and were raised in the U.S., who wish to maintain their heritage dialects, were able to produce comparable linguistic levels of variable features for only the features that are stereotypical of their specific dialects (i.e., socially salient), when compared to the speech of adult arrivals from the same region of origin. Other features that do not benefit from the same level of social salience displayed more difference between the child and adult arrivals. This is consistent with the assertion that "speakers are more aware of some variables than others" (Trudgill, 1986, p. 11), a position long held in sociolinguistics (Labov, 1972; Rácz, 2013; Barnes, 2015; Nycz, 2018). Erker concludes by saying that "[less salient] variables are dim social lights whereas [a socially salient variable] is a beacon, illuminating linguistic identities in ways that are strong and clear" (Erker, 2022, p. 157).

In the present study, the comparable VOT measurements between English /t/ and /k/ indicate that the coronal stop is undergoing a lesser amount of convergence, compared to its velar counterpart. According to these results, following Erker's social beacon analogy, /t/ is seemingly seen by Spanish-English bilingual speakers as the standard of typical English aspiration in voiceless stops–in other words, English /t/ is the beacon of English VOT. As such, the social salience of VOT productions has intervened in the amount of convergence observed in this individual phonetic category. This argument is supported by data from a different ongoing project of mine in which a distinct, albeit smaller, set of Spanish-English bilinguals displayed a similar linguistic behavior, where English /t/ was produced with comparable VOT measurements to English /k/. While this study cannot conclusively make that determination, due to the lack of monolingual speech data from this bilingual speaker pool, the patterns observed here in combination with the preliminary results from my other project in progress appear to point in that direction.

In conclusion, if we were to take the aforementioned argument at face value, a general explanation for the divergent behavior of Spanish and English /t/ in these data may be driven by the social salience of this phonetic category. In English, the higher VOT measurements could be the result of bilingual speakers' attempt to maintain an English-typical VOT production, whereas in Spanish, the reduced VOT measurements could be the result of the same speakers' emphasis on a phonetic contrast as it relates to the amount of aspiration that is produced/perceived in this specific sound. Nonetheless, it remains unclear from these data why only the coronal sound could benefit from this social salience while its bilabial and velar peers remained in the periphery of social salience.

Resyllabification post-analysis

During the presentation of these results to a linguistic audience, I introduced examples of intra-speaker variation that provide a visual representation of the wave form, as is the case with figures 4.9 and 4.10. The first figure represents a production of the English word "ten" by one speaker during the word list task with a rather shorter VOT: 28 ms. The second figure provides a representation of the same word by a different speaker in the same activity with a rather longer VOT: 53 ms. I then followed with quintessential examples of inter-speaker variation, especially as it occurs across activities, as is the case with figure 4.11, a production of "cabin" in the word list activity by one speaker with rather longer VOT, and figure 4.12, a production of "coaches" in the passages activity by the same speaker with rather shorter VOT–84 ms and 36 ms, respectively.

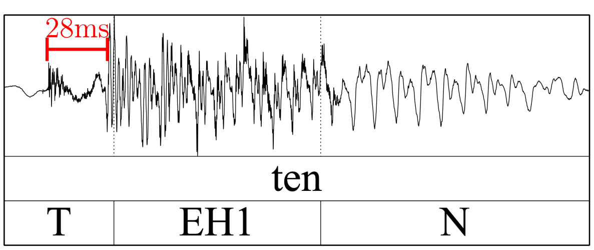


Figure 4.9: English /tɛn/ with shorter VOT

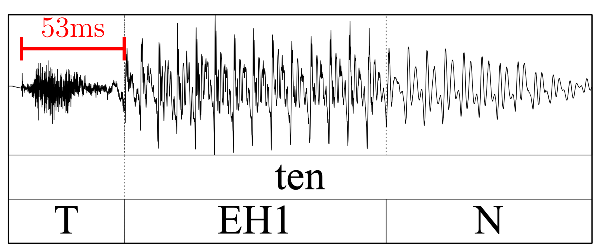


Figure 4.10: English /tɛn/ with longer VOT

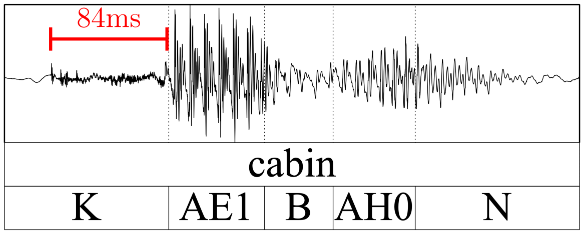


Figure 4.11: English /kæbɪn/ in word list

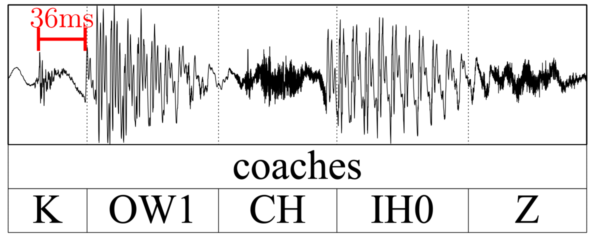


Figure 4.12: English /koʊtʃɪz/ in passages

Given that the English word "coaches" /koʊtʃɪz/ shares many phonetic similarities with the Spanish word "coches" (*cars*) /kotʃes/, I demonstrated that this was, in fact, a token of the English word produced immediately after a language switch, by providing a visualization of the linguistic context as shown in figure 4.13. Accordingly, an audience member asked a very fair question regarding a potential confound of linguistic context. Specifically, she asked if the reduced VOT in tokens such as "coaches" could be the result of a process of resyllabification of the Spanish coda /s/ leading to an onset cluster formation in the English word following a Spanish alveolar sibilant.[3] In English, it has been known for a long time that voiceless stops following a fricative /s/ are produced with short VOT (Lisker & Abramson, 1964), and that the short VOT in this linguistic context yields measurements similar to the typical range of the voiced counterparts /b d g/ (Cho et al., 2014). Since the suppression of VOT in English /s/-voiceless stop clusters is well known, this concern merited further investigation of this potential confound.

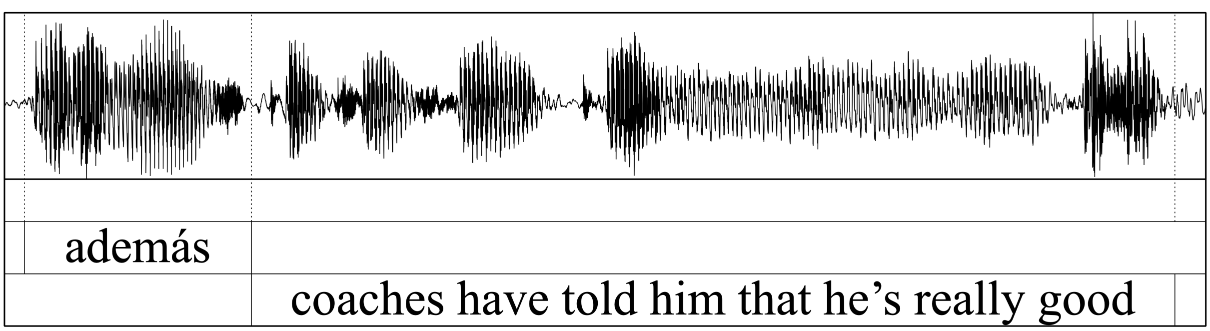


Figure 4.13: English "coaches" in passages with its linguistic context

After a search of the published literature and consultations with advisors and colleagues, I was unable to identify any research that points to the possibility of a resyllabification

[3]Thanks to Nicole Holliday for raising this concern.

process across languages around switch sites in language alternation. Nonetheless, I gathered a subsection of the corpus that included only English passage data for further statistical analysis. A mixed effect linear regression model that analyzed VOT measurements as a function of linguistic context with word and subject as random effects was performed, as visualized below:

– `lmer(VOT ∼ Boundary_Type + (1|subject) + (1|word))`

Three linguistic contexts were coded in the data for analysis:

1. /V#C/ – A vowel in Spanish coda position followed by an English voiceless stop after the switch.
2. /s#C/ – An alveolar fricative in Spanish coda position followed by an English voiceless stop after the switch.
3. 'Other' – A lateral /l/ or tap /ɾ/ in Spanish coda position followed by an English voiceless stop after the switch.[4]

The `mixed()` function applied to the statistical model described above indicated that there were no differences across the levels within the boundary type variable, as shown in table 4.9. These results are visualized in figure 4.14, which shows average VOT measurements for each of the three coded boundary categories.

Effect	**df**	**F**	***p*-value**
Boundary Type	2, 37.05	0.04	0.960

Table 4.9: `Mixed()` results for boundary type fixed effect

As the results above indicated, there is no evidence to suggest that Spanish-English bilingual speakers engage in a process of resyllabification that crosses switch points during language alternation. As such, the higher level of convergence observed in the passage activity, compared to other tasks, is indicative of an effect that is amplified in passage reading compared to other experimental tasks. In other words, the results reported in this section corroborate the previous assertion that English VOT productions in voiceless stop-initial, non-cognate words after a switch site are more susceptible to shortening due to a convergence effect in a language alternation speech mode. However, it remains unclear whether or not this conclusion can or should be generalized to other potential resyllabification contexts/boundaries with different phonemic candidates across languages in language alternation speech mode. Future research should further investigate this question.

[4]These two contexts were combined into a single boundary category for two reasons: (1) to my knowledge, there is no evidence to suggest they would or should behave differently from one another, and (2) there were fewer examples of these two boundary types, thus combining them allowed for more robust statistical results.

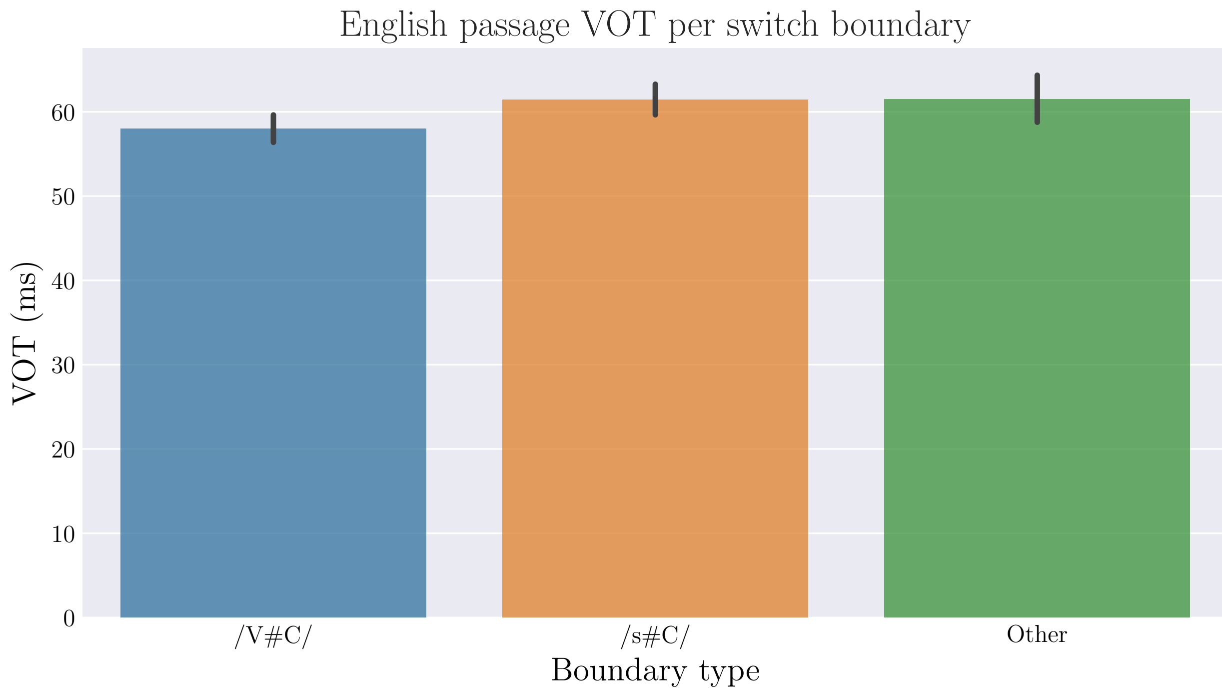


Figure 4.14: VOT distributions per linguistic context for English passage data

4.2 Social factors

This section will now provide a description of the second statistical model used in this research study and details on its implementation. Recall that the statistical model for the social factors employed in this study was constructed using the lasso regularization algorithm in Python using the sklearn package (Pedregosa et al., 2011). The utilization of the lasso regularization technique provides a means to handle complex statistical problems and facilitate robust model selection, an essential approach to deal with the wide data set that resulted from including the answer to each question in the BLP survey as its own variable. All the steps in this second analysis were implemented following discipline standards (James et al., 2013).

Lasso incorporates a regularization term in order to enhance the model's predictive accuracy and interpretability. It achieves this by imposing a penalty on the absolute values of the coefficients, thereby encouraging sparse solutions where some coefficients are shrunk to zero. This feature allows for automatic feature selection, identifying the most relevant predictors and disregarding those with little to no impact on the response variable, and thus deriving models that strike a balance between predictive accuracy and interpretability. The regularization process aids in mitigating overfitting, which occurs when a model is excessively complex and performs poorly on new, unseen data. The lasso algorithm effectively reduces the dimensionality of the problem by shrinking less relevant coefficients to zero, thereby enhancing the generalizability and robustness of the model. The use of the lasso regularization algorithm in the current section was a deliberate choice made to address the statistical challenges at hand. By leveraging the capabilities of Python and its libraries, I was able to construct a statistical model that balances accuracy and interpretability while mitigating overfitting. This approach facilitates insightful analysis and aids in the identification of the most influential predictors, thereby enhancing our understanding of the underlying linguistic behaviors.

However, before the statistical model was performed, the following steps were taken to preprocess the data in order to ensure accurate but generalizable model results:

1. Creation of dummy variables (referring to one-hot encoding)
2. Splitting data into training, validation, and test sets
3. Scaling numerical data (in all three sets)

First, to incorporate categorical information into the model and account for its impact on the response variable, the statistical model employed in the analysis utilized one-hot encoding (or dummy variables). This approach is a common technique used in statistical modeling to represent categorical variables in a quantitative manner. These variables do not have a natural numerical order or magnitude, and therefore cannot be directly included in statistical algorithms; the one-hot encoding technique helps overcome this limitation. In this encoding scheme, each category is represented by a binary vector with a value of 1 in the corresponding position and 0 in all other positions. Recall that one column from each variable was dropped in order to avoid multicollinearity issues resulting from the dummy variable trap.[5] While this approach widened the data set even further, the utilization of the lasso regularization algorithm was able to handle the data without an issue, as this technique is designed for this shape of data (i.e., wide–containing many variables).

Second, to assess the performance and generalization of the predictive model, the data were split into three distinct subsets: the training set, the validation set, and the test set. Note that equal amounts of data from each subject were directed to each of the three sets to ensure balanced sets. The training set constituted the largest portion of the data (60%) and was used to train the model. During this phase, the model learned the underlying patterns and relationships between the input variables and the corresponding output or target variable (i.e., the dependent variable). The validation set, made up of another 20% of the data, was employed to fine-tune the model's hyperparameters–further discussed of hyperparameters is

[5]The dummy variable trap occurs when two variable columns are split from a single categorical variable and thus provide the same exact information. For example, in a binary gender system, if we were to split a category variable "gender" that contains the levels "male" and "female" into two columns, for which the first column had a label of 1 for a participant who identified as male and 0 for female, and vice versa, then both columns are telling us the same exact information. As such, one of the two columns is dropped since only one of these columns is enough to give us information about gender identification.

provided below–and evaluate its performance. It serves as an unbiased evaluation set to assess how well the model generalizes to unseen data. The validation set was used to iteratively adjust the model's hyperparameters, such as regularization parameters or learning rates, to optimize its performance. Once the model had been trained and fine-tuned using the training and validation sets, it was evaluated using the test set. The test set, composed of the remaining 20% of the data, represented an independent and unseen data set, simulating real-world scenarios. The model's performance on the test set provides an unbiased estimate of its predictive ability and generalization to new, unseen data. Train-validate-test data splitting ensures that the model's performance is assessed on data that were not used during training or validation, thus minimizing the risk of overfitting or overestimating the model's performance.

Third, to remove any systematic bias in the data and ensure that positive and negative deviations from the mean are treated equally, the numerical data used in the analysis were scaled in order to standardize the data set to a zero mean and unit variance. This process was done to each set after data splitting, as is common practice in machine learning, because the test set functions as a set of new and unseen data, thereby not being intended for accessibility during the training phase. Any utilization of information derived from the test set before or during training introduces a potential bias in the evaluation of the model's performance. Scaling numerical data aims to standardize the variables and bring them to a comparable scale. When working with numerical variables that have different units or scales, such as age, income, or population size, it becomes challenging to compare their magnitudes directly. This discrepancy in scales can potentially lead to biased or misleading results in statistical analysis. To address this issue, scaling the data to a mean of zero and a variance of one is often employed. This involves subtracting the mean value of each variable from its individual observations and dividing all the values by the standard deviation. This process was done automatically for this project using sklearn's `StandardScaler()` function. The now centered and normally distributed data remove any inherent bias in the variables and enables fair comparisons between them, preventing variables with larger numerical scales–regardless of unit of measurement differences–from obtaining an unfair advantage in the statistical analysis. The scaling process does not alter the relative relationships between the variables; it simply adjusts their scales to be more comparable to one another. Consequently, the interpretation of the coefficients and effects in the statistical model remain unchanged. Scaling the data to a zero mean is particularly useful in models that involve regularization techniques, such as ridge regression or lasso regression, as it prevents variables with larger scales from dominating the regularization process. Ultimately, by removing the bias introduced by different scales, it helps prevent certain variables from unduly influencing the analysis. This technique contributes to robust statistical modeling and accurate interpretation of the results.

Note that both the speaker-internal and -external factors were analyzed together in the second statistical model; more information on this will be provided below. The current section proceeds with the results of the model for both of the social factor categories, followed by a discussion of those results.

Results

Once the data were preprocessed as described above, the following steps were taken during the training, validation, and testing phases of the model performance:

1. Performing hyperparameter tuning on training and validation sets
2. Applying best model to test set
3. Evaluating test set model performance

Hyperparameter tuning refers to the process of optimizing the hyperparameters of a machine learning or statistical model. Hyperparameters are adjustable parameters that are set prior to the training of a model and determine its behavior and performance. Unlike the model's internal parameters, which are learned from the training data, hyperparameters are typically defined by the user or researcher. The goal of hyperparameter tuning is to search for the optimal combination of hyperparameter values that results in the best performance of the model. This involves systematically exploring different combinations of hyperparameters and evaluating the model's performance using the Root Mean Square Error (RMSE), further discussed below, as a predefined evaluation metric. The process of hyperparameter tuning was approached through the `GridSearchCV()` function from sklearn. Grid search involves exhaustively trying all possible combinations of hyperparameter values within a predefined range.

Hyperparameter tuning is a crucial step in developing robust and accurate models. It helps in optimizing the model's performance, improving generalization, and avoiding overfitting. By systematically exploring different hyperparameter settings, we can fine-tune models to achieve the best trade-off between bias and variance, resulting in better predictive power and reliability. In short, by systematically exploring and optimizing adjustable parameters, this process plays a vital role in enhancing the model's predictive accuracy and ensuring its suitability for specific tasks and data sets. In this particular case, three hyperparameters were investigated: alpha, maximum number of iterations, and selection type, all of which are described below.

1. Alpha, denoted as α, serves as the regularization parameter or penalty term in lasso regression, exerting control over the degree of regularization employed. It delineates the equilibrium between model fitting to the training data and the attenuation of coefficient magnitudes. Elevating the alpha value amplifies the extent of coefficient shrinkage, inducing a higher number of coefficients to approach zero, potentially resulting in a sparser model representation. In contrast, diminishing the alpha value curtails the level of shrinkage, permitting larger coefficients to persist. The selection of an optimal alpha value assumes paramount importance to avoid the issues of overfitting or underfitting the data, thereby ensuring the model's accuracy and (appropriate) simplicity.

2. Maximum iterations denotes the user-specified parameter that delineates the upper limit of iterations or optimization steps allowed during the convergence of the model. Lasso regression, employing an iterative algorithm to estimate the coefficients by minimizing the objective function, relies on the maximum number of iterations parameter to determine the point at which the algorithm terminates, irrespective of whether convergence has been achieved. It is essential to set a suitably high value for maximum iterations to ensure that the algorithm has an ample number of iterations to converge adequately, particularly in scenarios characterized by sluggish convergence or intricate data structures.

3. The selection type hyperparameter is responsible for determining the feature selection algorithm employed in lasso regression. It offers two distinct options: 'cyclic' and 'random'. The 'cyclic' selection method updates the coefficients sequentially in a predetermined cyclic order, while the 'random' selection method involves the random selection of a coefficient to update during each iteration. The selection parameter significantly impacts both the computational efficiency and the stability of coefficient estimates.

For alpha, the range provided to the grid search function was 0.1 to 1, with a step value of 0.1 (i.e., increments of 0.1 to the alpha value in each combination); the options for the maximum number of iterations were 500, 1,000, or 2,000; and the selection options were 'cyclic' and 'random.' The sklearn function returned the best combination of hyperparameters as an alpha of 0.5, a maximum iteration of 500, and a selection type of 'cyclic.' With these optimal parameters determined by grid search, a lasso regression was performed on the unseen test data using these values. Both the evaluation of the best model setting during the training phase and the evaluation of the final model on the test set were performed using the RMSE value.

Root Mean Square Error (RMSE) is a statistical metric utilized to assess the accuracy and precision of a predictive model by quantifying the average magnitude of the residuals or errors between predicted and observed values. It represents the square root of the mean of the squared differences between predicted and observed values, thereby encapsulating the overall discrepancy in a single numerical value. The RMSE calculation involves several key steps. First, the individual errors are obtained by subtracting the predicted values from the corresponding observed values. These errors reflect the deviations or disparities between the model's predictions and the actual data. Second, the errors are squared to ensure non-negativity and to emphasize larger errors due to the quadratic nature of the operation. By squaring the errors, both positive and negative deviations are treated uniformly, and larger errors contribute more significantly to the overall measure. Third, the mean of the squared errors is computed by summing the squared errors and dividing the sum by the total number of predictions. This step provides an average value that represents the central tendency of the squared errors. Finally, the square root of the mean squared error is taken, yielding the RMSE value. This square root operation returns the metric to the original scale of the observed values, facilitating meaningful interpretation and comparison.

RMSE serves as an objective criterion to evaluate the accuracy and goodness-of-fit of predictive models in machine learning–and across various other domains, including regression analysis and time series forecasting. A lower RMSE value indicates superior model performance, as it signifies reduced overall deviation between the predicted and observed values. Conversely, a higher RMSE value signifies increased dispersion and larger errors, suggesting diminished predictive accuracy and precision. By utilizing RMSE, we can effectively compare and assess the performance of different models or variations within the same model class. This metric enables rigorous quantitative evaluation, allowing for informed decisions regarding model selection, parameter tuning, and overall model improvement. Furthermore, the RMSE value can be interpreted in the same units as the observed values–in this case, ms of VOT–facilitating practical understanding of the magnitude and significance of the prediction errors.

In other words, RMSE provides a concise numerical measure of the average magnitude of errors between predicted and observed values. It enables objective evaluation of predictive model accuracy and precision, with lower RMSE values indicating better alignment between predictions and observed data. Ultimately, RMSE plays a vital role in model assessment, selection, and refinement.

The hyperparameter tuning process returned a RMSE value of 16.96 for the validation predictions in the best model combination. In the test set, the same lasso regression settings returned a RMSE of 16.52 for the unseen data. While the evaluation of the test set returns a slightly lower value in the unseen data, the difference is very small, practically negligible. The comparable error value in both models indicates that the lasso regression settings are, indeed, generalizable to new data sets while yielding a similar fit, suggesting no overfitting or underfitting of the training data.

The lasso regression model on the test data returned a total of 15 relevant factors from the over 70 variables fed to the model. This means that only 15 variables returned a non-zero coefficient value, while the remaining 60 variables were shrunk to zero, as they did not reach the threshold placed by the penalty term of the regularization algorithm. The total coefficient values returned in the model are presented in table 4.10. Coefficients are correlated with the dependent variable, in this case VOT; that is, positive coefficients indicate a positive relationship between the dependent variable and the independent variable at hand and vice versa. After the model was performed and the coefficients were obtained, the results were split into the pre-determined speaker-inherent and environmental social factors, as exhibited in appendix F. From this division, we find that 11 of the 15 non-zero coefficients belong to speaker-inherent factors, whose names and coefficient values are illustrated in figure 4.15. The remaining four non-zero coefficients belong to environmental social factors, whose names and coefficient values are exhibited in figure 4.16.

The results indicate that subjects' higher self-reported values for their level of comprehension of Spanish were correlated with lower VOT productions. Moreover, self-reported answers that indicated a higher likelihood to identify with an English-speaking culture, and

Variable name	**Coefficient**
Comprehension: Spanish	-1.234173
Identity: English-speaking culture	0.759501
Writing: Spanish	0.672962
Lived in Spanish-speaking country	0.596903
Reading: English	0.554616
Counting: English	0.551257
Worked in English-speaking environment	0.449051
Speaking: English	-0.409049
Age of acquisition: Spanish	-0.304043
Feels like oneself in Spanish	-0.277145
Comprehension: English	-0.252217
Attitude: Important to use Spanish	-0.247748
Percentage of English with friends	0.191996
Attitude: Associated with Spanish	-0.062063
Exposed to Spanish with American English accent	0.043502

Table 4.10: Relevant variables and coefficient values in order of magnitude

answers that report higher competency in writing in Spanish, and reading in English, as well as a higher likelihood of counting in English are correlated with higher VOT values. Lastly, subjects who reported higher competency in speaking in Spanish, a higher age of acquisition of Spanish, a higher likelihood to feel like oneself when speaking in Spanish, a higher competency of English comprehension, and a higher value for the importance they place on the Spanish language and the need to be associated with a Spanish-speaking culture were correlated with lower VOT measurements.

Finally, the relevant environmental social factors indicate that individuals who reported lower lengths of residence in Spanish-speaking countries, higher percentage of speaking English in the workplace and with friends, and higher exposure to Spanish with an American English accent are more likely to produce higher VOT values while speaking English and Spanish in bilingual mode.

Discussion

The machine learning model described and applied in this section provides the opportunity to explore a much wider set of variables that were obtained during the data collection session of the present study, and it does it via several techniques. First, the utilization of lasso regularization, one-hot encoding, and zero mean scaling in linguistic research studies–as it was the case in this study–offers various benefits and advancements to the field. Lasso regularization, also known as L1 regularization, introduces a penalty term that controls the strength of regularization in regression models. By encouraging sparsity in the coefficient estimates,

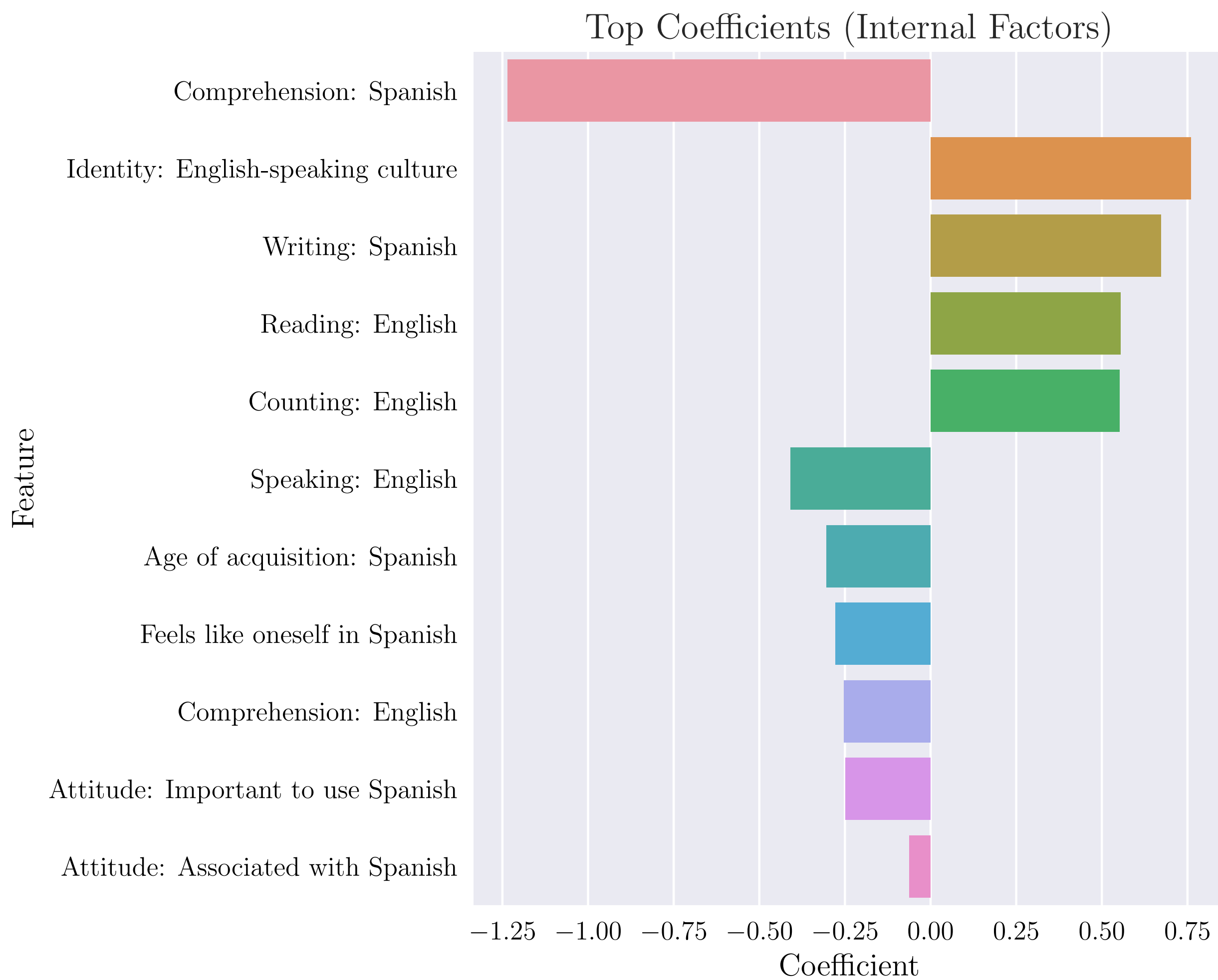


Figure 4.15: The 11 relevant speaker-inherent social factors that mediate VOT productions

lasso regularization facilitates feature selection and provides interpretable and parsimonious models. In linguistic research, this technique can help identify the most relevant linguistic features for predicting certain outcomes, enabling researchers to uncover key linguistic patterns and relationships, without having to make predetermined simplified model that potentially exclude critical variables. Accordingly, researchers are able to explore effects in large (especially wide) data sets while returning accurate results and mitigating the potential for overfitting (that is, poor generalizability). Thus, we can be sure to obtain results that capture a better/more informative glimpse of reality.

Second, one-hot encoding is a method for representing categorical variables as binary vectors, where each category is assigned a distinct binary attribute. This encoding scheme allows for the inclusion of categorical data in machine learning algorithms that typically operate on numerical data, as is the case with lasso regularization. It enables researchers to

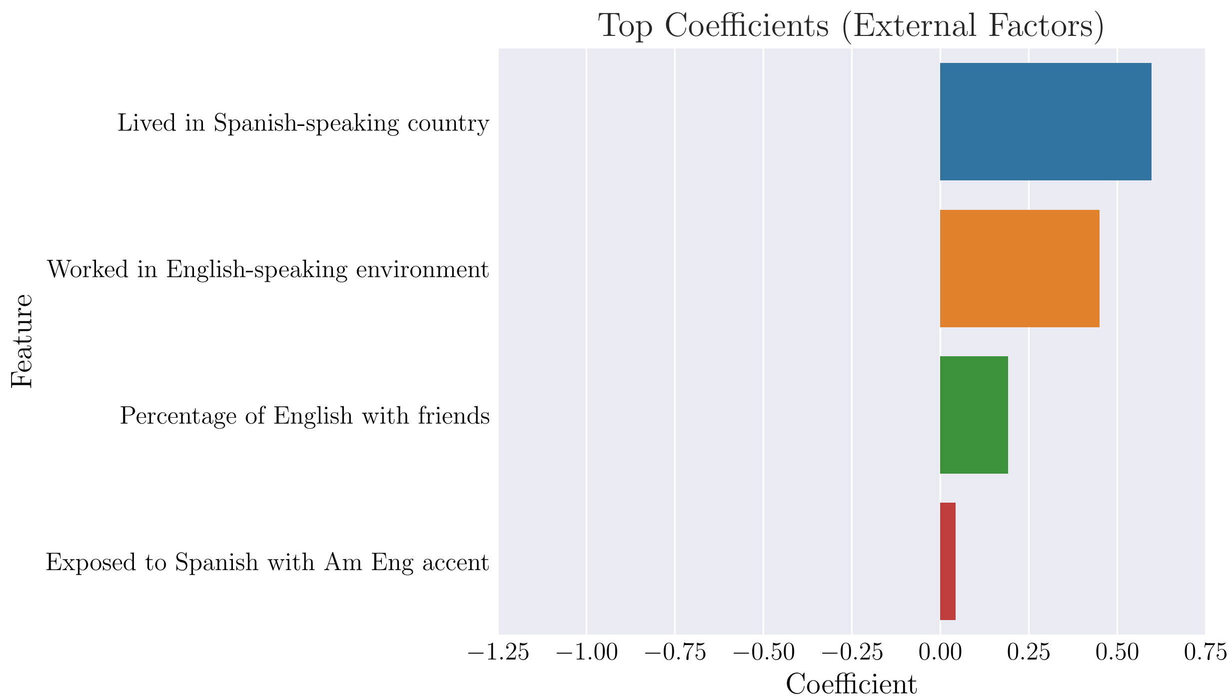


Figure 4.16: The 4 relevant environmental social factors that mediate VOT productions

capture the categorical nature of linguistic and social phenomena or characteristics, facilitating the exploration of relationships between linguistic/social variables and the prediction of linguistic behaviors or outcomes. Third, the division of data sets into training, validation, and test sets offers several notable advantages in the data analysis and model development. For instance, (a) the training set serves as the foundation for model training and parameter estimation. By exposing the model to a large portion of the data, it learns patterns, relationships, and underlying structures, thus enabling the model to make accurate predictions or classifications; (b) the validation set plays a crucial role in model evaluation and hyperparameter tuning. It serves as an independent data set that is not used during model training; and (c) the test set serves as a final assessment of the model's performance and generalization ability. It provides an unbiased evaluation of the model's predictive power on unseen data.

Fourth, standardization via zero mean and unit variance involves transforming numerical variables to have a mean of zero and a standard deviation of one. This scaling technique brings variables to a comparable range, allowing for fair comparisons and preventing any single variable from dominating the analysis. For this project, zero mean scaling was applied to various numerical features, such as age (e.g., general age, age of language acquisition, etc.), self-reported linguistic competency scores (i.e., speaking, comprehension, reading, and writing), and self-reported percentages of language usage (e.g., with family, with friends, at work, etc.). This normalization facilitates the identification of subtle linguistic patterns and the detection of meaningful differences among different linguistic groups or conditions while removing bias from uneven ranges across units of measurement. Ultimately, the slightly lower–albeit comparable–RMSE value in the unseen test data, compared with the lowest error value for the best validation set during the training phase, indicates that the model settings identified by this training process, which applied the machine learning techniques summarized above, returned algorithmic parameters that result in an accurate and generalizable statistical model.

Before discussing the significant results, it is important to discuss (two of) the non-significant variables in this study. First of all, the model did not find a difference in VOT productions as a result of participants' age or gender. While age-related VOT differences have not been attested in the literature during monolingual mode, gender-based differences have been reported. Nevertheless, there is a big difference between the present study and prior ones: previous results make social comparisons in monolingual speech, whereas the results from the present study make social comparisons across experimental tasks, all in bilingual mode. Thus, the lack of a statistical difference within the two aforementioned social categories does not indicate that there are no age- or gender-related VOT differences in bilingual speech compared monolingual speech. Rather, these results suggest that subgroups within these two social categories do not vary their VOT phonetic productions in bilingual mode as a function of experimental task.

As we transition into discussing the significant results, it is important to remember that coefficient values reported earlier–and discussed here–are correlated with VOT productions. That is, higher coefficient values suggests a given variable's likelihood to increase VOT productions in bilingual speech. In other words, a positive coefficient is indicative of lower levels of convergence in English, in particular, given that Spanish showed little variability at all, across all four tasks. English, on the other hand, showed a higher propensity to convergence and a much wider range of variation, as shown in figure 4.1, which displays general VOT distributions in English and Spanish data. As such, the direction of the magnitude of each coefficient will be used as an indication for convergence in English, specifically, where a positive value will be interpreted as little or no convergence towards Spanish, compared to a negative value, which will be associated with higher convergence in English VOT productions towards the Spanish ranges. Lastly, keep in mind that this study does not analyze the distributions of self-reported scores across social variables among Spanish-English bilinguals, but rather it analyzes how these distributions mediate VOT productions in bilingual mode across a series of research tasks.

Now, let's turn our attention to the results that the social model yielded in this section. Recall that participants who reported (a) higher Spanish and English comprehension ability scores, (b) higher English speaking ability scores (c) a later age of acquisition of Spanish, and (d) more positive attitudes towards Spanish (i.e., feeling like oneself in Spanish, finding it important to use Spanish, and finding it important to be associated with the Spanish language) had lower VOT productions–primarily in English, since that is the most accommodating/converging language, as it was discussed above–as indicated by the negative coefficients associated with these factors in the lasso regression model. By contrast, participants who reported (a) a closer association with an English-speaking culture, (b) a higher percentage of English usage when conducting simply mathematical tasks like counting, and (c) higher Spanish writing and English reading abilities had higher VOT productions–also in English speech–per the positive coefficient values returned by the statistical model. With these results in mind, we now turn to a recapitulation of the relevant literature on the pertinent factors.

In discussing the speaker-inherent variables, it is important to remember from chapter 2, firstly, that linguistic attitudes are influential in language variation and change, providing insights into social norms and values. For example, social evaluations of languages and subjective linguistic attitudes are connected to the identities of the social groups speaking those languages, and strong associations between a language and its speakers tend to make linguistic attitudes more prominent. Secondly, some languages or language varieties are associated with specific contexts or situations, such as Spanish in the US being linked to the home environment. For instance, Carranza and Ryan (1975) conducted a study on linguistic evaluations of Spanish and English in home and school contexts. Mexican-American bilingual speakers generally rated English more favorably in solidarity and status scales, while Spanish received higher evaluations in the home context and English in the school context. This could be attributed to the internalization of higher value associated with English due to living in an English-dominant society. Moreover, linguistic behaviors like language choice within bilingual communities may reflect linguistic attitudes but are influenced by various social factors. Language choice can be influenced by person-oriented motivations as described in the decision tree and accommodation theoretical frameworks. The decision tree model involves a hierarchical set of binary choices based on factors like interlocutor ethnicity, style, and conversation topic. However, this model does not account for code switching or the use of two languages simultaneously. Recall that Sankoff (1972) proposed modifications to allow for interpretive choices beyond the expected or unmarked choice.

Thirdly, there are influences on linguistic evaluations deriving from age-related shifts in linguistic attitudes, the dominance of English among bilingual teenagers, attitudes towards Spanish in academic and casual settings, perceptions of code switching, and the influence of bilingual proficiency, as described in the literature. For instance, L. Miller (2017) conducted a study on linguistic attitudes among school-age children and found that younger participants had comparable ratings for Spanish and English, but as they grew older, they displayed a preference for English through higher positive ratings. However, there was no negative view of Spanish found in a matched-guise task. Furthermore, Galindo (1995) analyzed the language attitudes of Spanish-English bilingual teenagers and found that they self-identified as English-dominant. They associated Spanish with a first-generation immigrant status, leading to a preference for English. Moreover, Achugar and Pessoa (2009) studied the linguistic attitudes towards Spanish in an academic community and found positive views of Spanish use and bilingualism in academic settings. However, negative views were expressed towards local varieties of Spanish, especially when spoken by monolingual speakers–very often first-generation immigrants in the US. In addition, Toribio (2002) explored language attitudes towards code switching among Spanish-English bilinguals and found a generally low prestige associated with code switching. Covert prestige drove the practice as a marker of identity, but not all bilinguals engaged in it due to stigma and stereotypes. Finally, Anderson and Toribio (2007) examined linguistic attitudes through an experimental study and found that single-noun insertions were more positively evaluated than more complex code switching. Bilingual proficiency and the recognition of English's influence on Spanish in the US played a role in participants' ratings.

The review above provides a refresher for the social backdrop in which to contextualize the findings in the present study. The variable coefficients returned from the model on the social factors illustrates the complexity of language identity and linguistic attitudes, as exemplified in the literature review. Nonetheless, these complex results also corroborate clear patterns attested in previous studies. For instance, self-association with an English-speaking culture among the participants resulted in a strong relationship with little or no convergence of English VOT measurements. Recall that Toribio (2002) found covert prestige associated with a positive view of code switching, as it was a marker of their ethnic identity. While it is not a direct one-to-one mapping, in this study, we can nonetheless use association with a Spanish-speaking culture as a proxy for association with a Latinx ethnic identity. Thus, lack of convergence may be the result from speakers' distancing from an identity linked to a Spanish-speaking culture. Two social phenomena appear to be at play here: (1) lack of convergence driven by a deviation from this ethnic identity, which results in a higher association with an English-speaker culture and/or (2) a lack of convergence driven by non-Latinx speakers who do not use variants (in this case shortened VOT productions) linked to Spanish-speaking culture either to avoid usage of linguistic forms typical of other social groups or because infrequent usage of language alternation has resulted in lower levels of convergence–in other words, the phonetic adjustments (very often, phonetic convergence towards the other language) that are very often seen in bilingual mode have not become commonplace in the speech of these bilingual speakers.

The other three relevant factors related to linguistic attitudes returned by the model suggest that the first social phenomenon above (i.e., bilinguals are deviating from an ethnic or Spanish-speaking identity) is better able to account for the speech behaviors observed here. When we look at the coefficients for "feeling like oneself when speaking Spanish," "finding it important to use Spanish," and "finding it important to be associated with a Spanish-speaking culture," those who reported higher scores for these questions also displayed higher levels of convergence, indicated by the negative coefficients among these three variables. Therefore, it is more likely that the linguistic attitudes that my lead to lower levels of convergence are (1) non-Latinx individuals avoiding inaccurate association with a Latinx/Spanish-speaking culture and/or (2) non-Latinx individuals' lack of *muscle memory*, for lack of a better term, given infrequent engagement with code switching or avoidance of the practice.[6] Even though

[6]Refer to Olson (2017) for a discussion on the effects of a switching cost associated with language alternation, as well as Olson (in press) for a discussion on the short-term and long-term effects of bilingual speech behaviors.

the BLP survey lacks questions that directly get to this matter in order to quantify its usage, the interview task included a question on this particular topic. For instance, a white participant provided the following response when asked if he uses *Spanglish*, a term used to describe the linguistic practice of code switching between Spanish and English:

> "It depends because, like I have said, I speak Spanish you know primarily with my professors and teachers, so I do try to avoid it with them because I guess like sometimes I feel like there's a little bit of a stigma associated with Spanglish. But sometimes it's just easiest. You know there are certain things I know how to say in Spanish and not English and certain things, you know, that I know how to say in English and not Spanish. So like I'm sure you've noticed, something I do a lot is I speak in Spanish, but there are certain nouns that I simply don't know how to say them in Spanish, so I say everything in Spanish except for a noun in English, simply because I'm not sure what the Spanish word is. And I don't really... So I try to avoid it because you know I'm trying to speak with professors, and I want to sound like, you know, I'm only speaking in Spanish."[7]

The response above encapsulates the general attitude shared by non-Latinx subjects: usage of Spanish that is often limited to the classroom where usage of Spanglish–i.e., Spanish-English code switching–is dispreferred. On the other hand, subjects of Latinx heritage cited in their interview responses frequent usage of Spanglish with family members and friends. As such, this study indicates that the lack of convergence in English VOT shown among participants who reported a closer identity with an English-speaking culture is driven by bilingual speakers who are trying to avoid an association with the stigmas and stereotypes linked to this linguistic practice, especially non-Latinx speakers who want to demonstrate a dominance of the Spanish language in academic settings. On the other hand, a convergence of English VOT productions appears to be driven by participants who reported more positive attitudes towards Spanish and more closely identify with a Spanish-speaking identity, as evidenced by their admission of higher rates of Spanglish usage with their friends and families.

In regard to linguistic competence or proficiency, on the one hand, subjects who reported higher levels of comprehension of both English and Spanish as well as higher levels of speaking ability in English displayed a higher propensity for convergence. On the other hand, subjects wo reported higher levels of writing in Spanish and reading in English demonstrated lower levels of convergence. First of all, while the trend of more convergence associated with higher levels of comprehension in *both* English and Spanish appears to be contradictory, in and of itself, zooming out and looking at the bigger picture provides a clearer view of these trends. To do this, rather than analyzing language-specific proficiency reports, it is important to identify the trends associated with the modes of communication themselves. In other words, we find that higher self-reported scores of speaking and comprehension abilities, regardless of language, are linked to more convergence (i.e., these factors returned negative coefficients in

[7]While this passage was produced while the subject was code switching, only the English version of the passage is provided here for convenience to the reader.

the statistical model of social factors). Conversely, we see lower propensity for English VOT convergence among bilingual speakers who reported higher proficiency scores in reading and writing, regardless of language. These trends are seemingly related to the language attitude results. Namely, speakers who report higher proficiency in modes of communication typical of academic settings (i.e., reading and writing) appear to be more concerned with the acquisition and usage of more standard/formal speech patterns, which in turn manifest themselves as lower levels of convergence of VOT productions, even in bilingual speech. This is contrasted by the trends observed by the speakers who reported higher scores for modes of communication that are more general (i.e., speaking and comprehension), regardless of the language.

Furthermore, some of these factors corroborate the general trends that have already been reported in the literature, when contextualized. For instance, those who report higher scores in their abilities to write in Spanish are those who have studied Spanish formally, either as a major or minor, in their college education. Like Achugar and Pessoa (2009) found, usage of Spanish and bilingualism among academics receives positive social evaluations in formal settings. However, local, less formal varieties receive negative social evaluations. As such, subjects who study Spanish formally and attribute higher scores to their writing abilities in this language appear to reduce the rate of less standard productions of VOT, resulting in less VOT convergence in their speech when alternating languages.

Similarly, situational context appears to align with the idea that convergence magnitude is linked to an identity that is more closely aligned to one language or the other. That is, subjects who report typically resorting to English for counting also have higher productions of VOT, indicating less convergence. Thus, in line with linguistic attitude and proficiency results, speakers who use English as the language for solving mathematical calculations also display less convergence, providing further support for the idea of language association-based convergence effects that has developed throughout this section.

Finally, the results from the social factors model indicate that as age of acquisition of Spanish increases, VOT predictions decrease, suggesting that Spanish-L2 speakers are more likely to converge their VOT categories in their English speech. This suggestion seems contradictory to all other results on speaker-inherent social factors that have been discussed so far–namely that non-Latinx, Spanish-L2 speakers want to maintain separate voiceless stop categories in an attempt to acquire and use standard/formal speech patterns in both their L1 and L2. While the creation of L2 sound categories–that is, voiceless stops–are expected to influence L1 categories, as Flege (1995) and Flege and Bohn (2021) suggest, these results appear to be contradictory. Nonetheless, this topic will be covered further below, after the presentation of additional information (i.e., environmental factors) that will help contextualize this factor.

Regarding the environmental social factors, the lasso regression model found this category less relevant for predicting VOT productions in language alternation, as only four variables in this category contained non-zero coefficients, namely, living in a Spanish-speaking country, working in an English-speaking workplace, percentage of time using English with friends, and amount of exposure to Spanish spoken with an American English accent. To forecast the analysis of this second portion of the current section, all four variables align closely to the pattern observed in the first portion–speaker-inherent social variables–in which (a) preference for or higher exposure to English or (b) higher levels of formal education in Spanish resulted in lower levels of VOT convergence among the Spanish-English bilingual speakers in this study.

First, length of residence in a Spanish-speaking country is difficult to analyze given subjects' complex interpretation. On the one hand, some individuals reported longer residence times in Spanish-speaking countries because they participated in study abroad programs. While the BLP does not ask about topics such as study abroad periods, thus making it difficult to quantify this matter, conversations during the interview portion revealed that many students who are Spanish majors took advantage of study abroad opportunities by studying in Spain or other Latin American countries for one semester or longer. Therefore, under this analysis, the results would align with language proficiency results reported earlier, where speakers who study Spanish formally are more likely to produce higher rates of formal language, which includes distinct (and stable) VOT categories in each of the two languages, even during language alternation. On the other hand, however, during informal discussions after the completion of the data collection session, many subjects revealed that they counted the US as a Spanish-speaking country, as they resided in communities where Spanish is very prevalent and often utilized in public spaces, commercial contexts, and government offices. As such, untangling the two contradicting realities associated with the factor at hand complicates any accurate interpretation of how it assists in the prediction of linguistic behavior among bilingual speakers. Accordingly, additional research is necessary in order to determine the context under which Spanish-English bilinguals are spending extended lengths of residence in Spanish-speaking countries and identify how these experiences are able to account for the convergence effects seen in their linguistic patterns.

Second, higher amounts of English spoken in the workplace or with friends are positively correlated with longer VOT productions (i.e., less convergence in English) in bilingual mode across all four tasks. This pattern is in line with previously reported results on speaker-inherent social variables from this study, which indicate that subjects who identified more closely with an English-speaking culture had longer VOT productions. Given that speakers who identify more closely with an English-speaking culture are assumed to also be the ones who use more English in environment like the workplace and with friends (and vice versa), the two variables at hand fit in with the variables analyzed earlier in this section. Thus, higher exposure to and/or preference for English, as it relates to environmental social variables, also appears to minimize convergence effects in language alternation.

Third, of all types of accents that subjects reported having frequent and constant exposure to, only Spanish spoken with an American English accent was found relevant in the predictions of VOT measurements in the language alternation data. Albeit with a smaller magnitude than other factors, exposure to Spanish spoken with an American English accent, as a predictor, indicates to us that bilingual speakers are influenced by their linguistic input, as it relates to accented speech. As a reminder, Flege and Bohn (2021) argue that the quality of input in L2 speech research has been overlooked despite its potential influence leading to the speech production differences seen between L2 learners and native speakers. They also exemplify its importance. For instance, input quality, whether native-like or accented, plays a significant role in the development of new sound categories in the L2, they argue. In monolingual research, the optimal weighting of perceptual cues in the development of phonetic categories requires years of exposure to input. Cue weighting patterns depend on cue reliability and the statistical properties of input distributions during L1 acquisition. These patterns are not rigidly applied by monolinguals and can adapt dynamically based on recent exposure and training. This adaptive cue weighting mechanism also extends to speech production. For example, monolingual speakers can adapt their speech productions based on exposure to manipulated stimuli. The SLM-r predicts that individual differences in cue weighting can be most evident in the production and perception of L2 sounds that are perceptually linked to L1 sounds. The influence of L1 cue weighting patterns is expected to be stronger for L2 sounds that remain linked to an L1 category, while newly formed L2 phonetic categories develop based on the reliability of multiple cues in input distributions. Accordingly, convergence in VOT categories observed in bilingual speakers may result from the association between corresponding sounds in their L1 and L2, as well as recent input they have received.

Lord (2005)–and many others–have indicated that non-native VOT productions of voiceless stops /p t k/ have been one of the phonetic features associated with accented speech in both English and Spanish by native listeners. Even though non-target productions do not interfere with intelligibility or comprehensibility, the non-target production of this feature in either language is nonetheless a clear indicator of foreign accentedness to native speakers (Lord, 2005). Following the SLM-r's premise that L1 sound categories affect L2 categories, and vice versa, exposure to accented speech is expected to influence all speakers' productions, regardless of whether a given language is their L1 or L2. Therefore, exposure to Spanish with an American English accent is expected to influence both Spanish-L1 and English-L1 speakers by influencing their English categories in particular. For individuals in both of these groups who have higher exposure to this variety of Spanish, this exposure is likely to serve as a reminder of the typical long-lag VOT range of English productions. As such, the *Englishness* of the input they are receiving, even in a portion of Spanish input, may strengthen the weight of the perceptual cues for long-lag English VOT that later materialize in longer VOT productions in their own English speech, thus leading to longer VOT productions even in bilingual mode.

This argument takes us back to the speaker-internal results indicating that a later age of acquisition of Spanish leads to higher levels of convergence of English towards Spanish. Recall that, at face value, these results appeared to be contradictory because we expected Spanish-L2 speakers to display lower levels of convergence due to their desire to acquire and maintain more formal speech patterns, especially as it relates to academic settings. However, it is also the case that many of the Spanish-L2 speakers learned Spanish in an academic setting, which included the opportunity to study abroad in a Spanish-speaking country. As such, these participants found themselves in predominantly Spanish-speaking environments–in recent history, since many of them were Spanish majors still working on the completion of their undergraduate degrees–which meant a sudden increase of Spanish input in their daily lives. Accordingly, the higher amount of Spanish input with consistent Spanish-monolingual-like VOT may have affected even their L1 categories due to a cue weighting readjustment process that could have taken place as a result of their experience abroad. That is to say, the prevalence and recency of Spanish-like VOT, like the SLM-r predicts, may have influenced the L1 VOT productions among the Spanish-L2 speakers who had the opportunity to spend some time abroad. However, given the entanglement of the "living in a Spanish-speaking country" variable, without any productive way of quantifying and analyzing VOT productions as a function of study abroad length of residency and language immersion, it is difficult to provide a more definitive answer of whether or not this experience itself is what is contributing to the higher levels of convergence observed in bilingual mode for the late Spanish learners among the participants; that is, this potential account of this speech pattern remains uncertain, at the moment, given the fact that this research study does not study this topic systematically.[8]

Finally, while others have found language dominance–as provided by the dominance score value produced by the BLP survey, in some cases–as a predictor of VOT convergence in language alternation (e.g., Antoniou et al., 2011; Olson, 2013, 2016), when combined with the myriad of other variables from all BLP questions, the present study did not find dominance score to be a strong predictor of convergence. However, this study found that some of the variables that factor into the dominance score calculation did rise through the L1 penalty threshold to contribute to the statistical model predictions. That is, the results from this study suggest that dominance score is too broad of a variable and the reality is more nuanced, where only a few demographic and sociolinguistic characteristics actually influence convergence of the feature under investigation. That is, the reality of linguistic behaviors is much more complex than the dominance scores are able to capture and represent; while some variables that contribute to the calculation of such a score may be very informative, others may not provide much information to the statistical predictions. This discernment, thus, is an invitation to other researchers to tease apart the several factors that are used for the calculation of dominance score; this step can be fulfilled with powerful tools such as the regularization technique (i.e., lasso) described in this project. However, a complication with such an approach leads to a description that is at times complex or even contradictory, as indicated throughout this section. In other words, while most variables paint a clear picture of how social factors influence language behaviors, a few of the variables can complicate the interpretation of this bigger picture. As such, I encourage other researchers to adhere to simpler variables such as dominance scores when model simplicity is desired for interpretation purposes and venture out to a more nuanced approach when complexity is essential, especially with larger data sets.

All in all, from all the social factors under investigation in the present study, three main generalizations can be made:

[8]This question can only be answered with future research that specifically tackles this topic in a methodological manner.

1. Social factors associated with Spanish (or a Spanish-speaking culture) result in lower VOT–that is, more convergence in English towards Spanish.

2. Social factors associated with English (or an English-speaking culture) result in higher VOT–that is, less convergence in English towards Spanish.

3. Exposure to Spanish spoken with an American English accent leads to higher VOT–that is, less convergence in English towards Spanish.

Taken as a whole, this section demonstrates that social factors–especially when analyzed quantitatively in a nuanced manner–can be very informative for predicting VOT productions and phonetic convergence effects in bilingual speech. While other studies on language alternation have focused primarily on psycholinguistic/cognitive aspects contributing to phonetic convergence, it is evident that social factors are also at play and can help us better understand and account for the patterns observed in empirical studies.

Chapter 5

Conclusion

Considering the diverse outcomes reported in the phonetics of bilingualism literature, the present study aimed to create a comparative study within a single subject group to analyze the potential effects of research task on the convergence of VOT productions in language alternation research. In addition, the study explored the effects of several linguistic (language, POA, and speaking rate) and social (language history, language usage, language proficiency, and language attitudes) factors on said VOT productions. The study found that there is a pattern of unidirectional convergence of English VOT shifting towards the Spanish production range. In particular, the results indicate that there is a task effect, where VOT productions in word list reading and puzzle tasks display less convergence than the interview and passage reading tasks. Furthermore, we find effects of language (English has longer VOT), POA (English has a /p/ < /t k/ pattern, whereas Spanish has a /p t/ < /k/ pattern), and speaking rate (slower speech leads to longer VOT). Moreover, we also find an effect of (a) language proficiency, with higher ability scores for reading and writing in both English and Spanish reduce the propensity for convergence, compared with higher scores for speaking and comprehension abilities; (b) language history, with a higher age of acquisition of Spanish leading to more convergence; (c) language usage, with higher usage of English with friends and in the workplace leading to less convergence of English VOT; and (d) language attitudes, with a link between a higher association with a Spanish-speaking culture and convergence, whereas a higher association with an English-speaking culture resulted in less convergence. Finally, the study also finds that exposure to Spanish with an American English accent resulting in less propensity to converge VOT productions, compared with exposure to other varieties of accented speech.

The results from this study were put in perspective according to relevant linguistic, cognitive, and social theoretical frameworks. First, with respect to the convergence effects seen in English data, this study makes reference to the language latitude hypothesis proposed by Bullock and Toribio (2009c), in which the authors assert that languages with long-lag VOT patterns are more likely to display convergence effects than short-lag languages, given the leeway provided by the wider and further apart VOT ranges observed in this language class (i.e., long-lag languages). Second, the results of language task difference in convergence magnitude are discussed in relation to the "attention to speech" factor (Labov, 2011) and the potential effects of working memory on language processing mechanisms. Namely, while differences in VOT productions during cued switching read-aloud activities with scripted stimuli were analyzed in reference to the amount of monitoring to individual phonetic segments, differences in VOT productions during spontaneous code switching activities were analyzed in relation to the amount of language processing resources made available during these activities, with those activities placing a higher working memory load receiving substantially more resources. Third, regarding the linguistic factors, most variables (i.e., language and speech rate) behaved as expected, expect for the production of /t/ in the POA variable. While in English /t/ paired up with /k/, in Spanish it paired up with /p/, leading to incongruous groupings across languages. However, this finding is discussed in line with the idea of linguistic salience of "dim social lights" versus "beacons" of linguistic identities proposed by Erker (2022). In other words, if /t/ is perceived as the quintessential marker of VOT differences between Spanish and English, the divergence of its VOT productions are an attempt by bilingual speakers to maintain distinct phonetic categories, even during language alternation.

Fourth, the social factors within the categories of language history, language usage, language proficiency, and language attitudes were all discussed in comparison to other empirical studies in bilingualism research. In particular, we find that Spanish-English bilinguals who study both languages formally (as seen by higher self-reported scores for reading and writing in both languages) are less likely to converge their VOT ranges, similar to previous studies that report higher social valuations for formal language patterns (Achugar & Pessoa, 2009). Nonetheless, speakers who associate more closely with a Spanish-speaking community (many of them expressed during the casual interview a higher inclination for code switching in casual settings) also displayed higher levels of convergence, which may be the result of using this linguistic behavior as a marker of ethnic identity (Toribio, 2002). However, due to complications with the analysis of what it means to "live in a Spanish-speaking country," some aspects of the social factors model remained unclear. Likewise, the reason why speakers who learned Spanish at a later age would show more convergence in English remains uncertain–a topic that requires further research. Finally, we learned that exposure to Spanish with an American English accent decreases the amount of convergence of English. These results are discussed according to the cue weighting adaptation hypothesis that proposes that input characteristics can and will influence perceptual acceptability and production patterns (Flege & Bohn, 2021).

While this research study has advanced our understanding of important factors that may influence speech patterns and the processing mechanisms involved during language alternation, especially with regard to research methodology choices, it is also clear that (a) there are some limitations associated with this project and (b) more research is needed to address those limitations and new questions that arose from these findings. In particular, this study was primarily concern with the comparison of code-switched speech across experimental tasks, which did not require monolingual data. Nonetheless, monolingual speech could further help us understand not only how language alternation speech patterns change across tasks but also across speech modes, with a more systematic and thorough analysis. Furthermore, another limitation related to the puzzle activity, which did not show as much convergence as expected, could be due to the fact that the interlocutor during this activity was the researcher, a methodological decision that could have led the subjects to monitor their own speech even more closely than they would with a peer.

Nonetheless, future research in language alternation on areas such as the salience of English and Spanish /t/, the potential working memory loads associated with a wider array of research tasks, and issues pertaining to the code switching linguistic profile (i.e., engagement in and attitudes towards code switching) of bilingual speakers should be pursued in order to address questions that have arisen from the results presented in this dissertation.

To conclude, the present study provides strong evidence that methodological choices such as the selection of certain experimental tasks can have a strong influence on the amount of phonetic convergence observed in subjects' speech. Accordingly, researchers should control for these differences or analyze them systematically in order to avoid potential confounds in the results obtained in language alternation research as well as to make studies more comparable and better understand the experimental, cognitive, and social influences driving speakers' linguistic behaviors and patterns, including phonetic productions.

Bibliography

Abdelli-Beruh, N. B. (2004). The stop voicing contrast in French sentences: Contextual sensitivity of vowel duration, closure duration, voice onset time, stop release and closure voicing. *Phonetica*, *61*(4), 201–219.

Achugar, M., & Pessoa, S. (2009). Power and place: Language attitudes towards Spanish in a bilingual academic community in Southwest Texas. *Spanish in Context*, *6*(2), 199–223.

Allen, J. S., Miller, J. L., & DeSteno, D. (2003). Individual talker differences in voice-onset-time. *The Journal of the Acoustical Society of America*, *113*(1), 544–552.

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

Amengual, M. (2021). The acoustic realization of language-specific phonological categories despite dynamic cross-linguistic influence in bilingual and trilingual speech. *The Journal of the Acoustical Society of America*, *149*(2), 1271–1284.

Anderson, T. K., & Toribio, A. J. (2007). Attitudes towards lexical borrowing and intra-sentential code-switching among Spanish-English bilinguals. *Spanish in Context*, *4*(2), 217–240.

Antoniou, M., Best, C. T., Tyler, M. D., & Kroos, C. (2011). Inter-language interference in VOT production by L2-dominant bilinguals: Asymmetries in phonetic code-switching. *Journal of Phonetics*, *39*(4), 558–570.

Appel, R., & Muysken, P. (1987). *Language contact and bilingualism*. Edward Arnold Publishing Company.

Appel, R., & Muysken, P. (2005). *Language contact and bilingualism*. Amsterdam University Press.

Aslin, R. N. (2014). Phonetic category learning and its influence on speech production. *Ecological Psychology*, *26*(4), 4–15.

Auer, P. (1984). *Bilingual conversation*. John Benjamins Publishing Company.

Auer, P. (2005). A postscript: Code-switching and social identity. *Journal of Pragmatics*, *37*(3), 403–410.

Baker, R., & Hazan, V. (2011). DiapixUK: Task materials for the elicitation of multiple spontaneous speech dialogs. *Behavior Research Methods*, *43*, 761–770.

Balukas, C., & Koops, C. (2015). Spanish-English bilingual voice onset time in spontaneous code-switching. *International Journal of Bilingualism*, *19*(4), 423–443.

Barnes, S. (2015). Perceptual salience and social categorization of contact features in Asturian Spanish. *Studies in Hispanic and Lusophone Linguistics*, *8*(2), 213–241.

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.

Bell, A. (1984). Language style as audience design. *Language in society*, *13*(2), 145–204.

Birdsong, D., Gertken, L. M., & Amengual, M. (2012). Bilingual language profile: An easy-to-use instrument to assess bilingualism. *COERLL, University of Texas at Austin.*

Bradlow, A. R., & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition*, *106*(2), 707–729.

Bullock, B. E., & Toribio, A. J. (Eds.). (2009a). *The cambridge handbook of linguistic code-switching*. Cambridge University Press.

Bullock, B. E., & Toribio, A. J. (2009b). Themes in the study of code-switching. In B. E. Bullock & A. J. Toribio (Eds.), *The cambridge handbook of linguistic code-switching*.

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

Bullock, B. E., Toribio, A. J., González, V., & Dalola, A. (2006). Language dominance and performance outcomes in bilingual pronunciation. In M. Grantham O'Brien, C. Shea, & J. Archibald (Eds.), *Proceedings of the 8th generative approaches to second language acquisition conference* (pp. 9–16).

Bürki-Cohen, J., Grosjean, F., & Miller, J. L. (1989). Base-language effects on word identification in bilingual speech: Evidence from categorical perception experiments. *Language and Speech*, *32*(4), 355–371.

Byrd, D. (1994). Relations of sex and dialect to reduction. *Speech Communication*, *15*(1-2), 39–54.

Carranza, M. A., & Ryan, E. B. (1975). Evaluative reactions of bilingual Anglo and Mexican American adolescents toward speakers of English and Spanish. *International Journal of the Sociology of Language*, *68*(6), 83–104.

Chandrasekaran, B., Sampath, P. D., & Wong, P. C. (2010). Individual variability in cue-weighting and lexical tone learning. *The Journal of the Acoustical Society of America*, *128*(1), 456–465.

Cho, T., & Ladefoged, P. (1999). Variation and universals in vot: Evidence from 18 languages. *Journal of phonetics*, *27*(2), 207–229.

Cho, T., Lee, Y., & Kim, S. (2014). Prosodic strengthening on the /s/-stop cluster and the phonetic implementation of an allophonic rule in English. *Journal of Phonetics*, *46*, 128–146.

Clarke, C., & Luce, P. (2005). Perceptual adaptation to speaker characteristics: VOT boundaries in stop voicing categorization. *ISCA Workshop on Plasticity in Speech Perception*, 23–26.

Clayards, M. (2018). Differences in cue weights for speech perception are correlated for individuals within and across contrasts. *The Journal of the Acoustical Society of America*, *144*(3), EL172–EL177.

Coetzee, A. W., García-Amaya, L., Henriksen, N., & Wissing, D. (2015). Bilingual speech rhythm: Spanish-Afrikaans in Patagonia. *The Scottish Consortium for ICPhS*.

Coupland, N. (2007). *Style: Language variation and identity*. Cambridge University Press.

Crystal, T. H., & House, A. S. (1982). Segmental durations in connected speech signals: Preliminary results. *The Journal of the Acoustical Society of America*, *72*(3), 705–716.

Diamond, A. (2013). Executive functions. *Annual review of psychology*, *64*, 135–168.

Eckert, P. (2012). Three waves of variation study: The emergence of meaning in the study of sociolinguistic variation. *Annual Review of Anthropology*, *41*, 87–100.

Erker, D. (2022). How social salience can illuminate the outcomes of linguistic contact: Data from Spanish in Boston. *The coherence of linguistic communities* (pp. 145–162). Routledge.

Flege, J. E. (2005). Evidence for plasticity in studies examining second language speech acquisition. *ISCA Workshop on Plasticity in Speech Perception*.

Flege, J. E. (1991). Age of learning affects the authenticity of voice-onset time (VOT) in stop consonants produced in a second language. *The Journal of the Acoustical Society of America*, *89*(1), 395–411.

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

Flege, J. E., & Bohn, O.-S. (2021). The revised speech learning model (SLM-r). In R. Wayland (Ed.), *Second language speech learning: Theoretical and empirical progress* (pp. 3–83). Cambridge University Press.

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., & Munro, M. J. (1994). The word unit in second language speech production and perception. *Studies in Second Language Acquisition*, *16*, 381–411.

Fought, C. (2006). *Language and ethnicity*. Cambridge University Press.

Francis, A. L., Kaganovich, N., & Driscoll-Huber, C. (2008). Cue-specific effects of categorization training on the relative weighting of acoustic cues to consonant voicing in English. *The Journal of the Acoustical Society of America*, *124*(2), 1234–1251.

Fricke, M., Kroll, J. F., & Dussias, P. E. (2016). Phonetic variation in bilingual speech: A lens for studying the production–comprehension link. *Journal of memory and language*, *89*, 110–137.

Galindo, D. L. (1995). Language attitudes toward Spanish and English varieties: A Chicano perspective. *Hispanic Journal of Behavioral Sciences*, *17*(1), 77–99.

Ganley, C. M., Vasilyeva, M., & Dulaney, A. (2014). Spatial ability mediates the gender difference in middle school students' science performance. *Child development*, *85*(4), 1419–1432.

Giles, H. (1973). Accent mobility: A model and some data. *Anthropological linguistics*, *15*(2), 87–105.

Giles, H., & Ogay, T. (2007). Communication accommodation theory. In B. B. Whaley & W. Samter (Eds.), *Explaning communication: Contemporary theories and exemplars*. Taylor & Francis.

Goldinger, S. D. (1998). Echoes of echoes? an episodic theory of lexical access. *Psychological Review*, *105*(2), 251.

Grosjean, F. (1980). Spoken word recognition processes and the gating paradigm. *Perception & Psychophysics*, *28*(4), 267–283.

Grosjean, F., & Miller, J. L. (1994). Going in and out of languages: An example of bilingual flexibility. *Psychological Science*, *5*(4), 201–206.

Gumperz, J. J., & Hernández-Chávez, E. (1971). Bilingualism, bidialectalism, and classroom interaction. *Language in social groups* (pp. 311–339).

Gutiérrez Topete, E. (2021). *Vot-cp: A python program for automatic data codification and calculation of voice onset time* (tech. rep.). Language Flagship Technology Innovation Center.

Hammond, M. (1999). *The Phonology of English: A Prosodic Optimality-Theoretic Approach*. Oxford University Press.

Hardcastle, W. J. (1973). Some observations on the tense-lax distinction in initial stops in korean. *Journal of Phonetics*, *1*(3), 263–272.

Henton, C. G., & Bladon, R. A. (1985). Breathiness in normal female speech: Inefficiency versus desirability. *Language & Communication*, *5*(3), 121–127.

Hewlett, N., & Rendall, M. (1998). Rural versus urban accent as an influence on the rate of speech. *Journal of the International Phonetic Association*, *28*(1-2), 63–71.

Holt, L. L., & Lotto, A. J. (2006). Cue weighting in auditory categorization: Implications for first and second language acquisition. *The Journal of the Acoustical Society of America*, *119*(5), 3059–3071.

Hualde, J. I. (2005). *The Sounds of Spanish*. Cambridge University Press.

Idemaru, K., & Holt, L. L. (2011). Word recognition reflects dimension-based statistical learning. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(6), 1939–1956.

Idemaru, K., Holt, L. L., & Seltman, H. (2012). Individual differences in cue weights are stable across time: The case of Japanese stop lengths. *The Journal of the Acoustical Society of America*, *132*(6), 3950–3964.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning*. Springer.

Kahng, J. (2014). Exploring utterance and cognitive fluency of L1 and L2 English speakers: Temporal measures and stimulated recall. *Language Learning*, *64*(4), 809–854.

Keating, P. A. (1984). Phonetic and phonological representation of stop consonant voicing. *Language*, 286–319.

Keshet, J., Sonderegger, M., & Knowles, T. (2014). AutoVOT: A tool for automatic measurement of voice onset time using discriminative structured prediction [Computer program]. [Version 0.94, retrieved June 2020 from https://github.com/mlml/autovot/.].

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.

Kessinger, R. H., & Blumstein, S. E. (1998). Effects of speaking rate on voice-onset time and vowel production: Some implications for perception studies. *Journal of Phonetics*, *26*(2), 117–128.

Khattab, G. (2000). VOT production in English and Arabic bilingual and monolingual children. *Leeds Working Papers in Linguistics, 8*, 95–122.

Kim, M.-R. (2012). L1-L2 transfer in VOT and f0 production by Korean English learners: L1 sound change and L2 stop production. *Phonetics and Speech Sciences*, *4*(3), 31–41.

Kuehn, D. P., & Moll, K. L. (1976). A cineradiographic study of VC and CV articulatory velocities. *Journal of Phonetics*, *4*(4), 303–320.

Kuhl, P. K., Conboy, B. T., Padden, D., Nelson, T., & Pruitt, J. (2005). Early speech perception and later language development: Implications for the "critical period". *Language Learning and Development*, *1*(3-4), 237–264.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Labov, W. (1972). *Sociolinguistic patterns*. University of Pennsylvania Press.

Labov, W. (2006). *The social stratification of English in New York City*. Cambridge University Press.

Labov, W. (2011). *Principles of linguistic change: Cognitive and cultural factors* (Vol. 3). John Wiley & Sons.

Lee, H., & Jongman, A. (2019). Effects of sound change on the weighting of acoustic cues to the three-way laryngeal stop contrast in Korean: Diachronic and dialectal comparisons. *Language and Speech*, *62*(3), 509–530.

Lee, S. A. S., & Iverson, G. K. (2012). Stop consonant productions of Korean–English bilingual children. *Bilingualism: Language and Cognition*, *15*(2), 275–287.

Lehet, M., & Holt, L. L. (2017). Dimension-based statistical learning affects both speech perception and production. *Cognitive Science*, *41*, 885–912.

Lenth, R. V. (2023). *Emmeans: Estimated marginal means, aka least-squares means* [R package version 1.8.4-1]. https://CRAN.R-project.org/package=emmeans

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

Lisker, L., & Abramson, A. S. (1970). The voicing dimension: Some experiments in comparative phonetics. *Proceedings of the 6th International Congress of Phonetic Sciences*, *563*, 563–567.

Lord, G. (2005). (how) can we teach foreign language pronunciation? on the effects of a spanish phonetics course. *Hispania*, *88*(3), 557–567.

Maddieson, I. (1997). Phonetic universals. In J. Laver & W. J. Hardcastle (Eds.), *The handbook of phonetic sciences* (pp. 619–639). Blackwells.

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.

Matthies, M., Perrier, P., Perkell, J. S., & Zandipour, M. (2001). Variation in anticipatory coarticulation with changes in clarity and rate. *Journal of Speech, Language, and Hearing Research*, *44*(2), 340–353.

McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., & Sonderegger, M. (2017). Montreal forced aligner: Trainable text-speech alignment using kaldi. *Interspeech*, *2017*, 498–502.

McAuliffe, M., & Sonderegger, M. (2022a). *Spanish (Latin America) MFA G2P model v2.0.0a* (tech. rep.). https://mfa-models.readthedocs.io/G2Pmodel/Spanish/Spanish(LatinAmerica)MFAG2Pmodelv2_0_0a.html.

McAuliffe, M., & Sonderegger, M. (2022b). *Spanish MFA acoustic model v2.0.0a* (tech. rep.). https://mfa-models.readthedocs.io/acoustic/Spanish/SpanishMFAacousticmodelv2_0_0a.html.

McAuliffe, M., & Sonderegger, M. (2023a). *English (US) MFA G2P model v2.2.1* (tech. rep.). https://mfa-models.readthedocs.io/G2Pmodel/English/English(US)MFAG2Pmodelv2_2_1.html.

McAuliffe, M., & Sonderegger, M. (2023b). *English MFA acoustic model v2.2.1* (tech. rep.). https://mfa-models.readthedocs.io/acoustic/English/EnglishMFAacousticmodelv2_2_1.html.

Miller, J. L., Green, K. P., & Reeves, A. (1986). Speaking rate and segments: A look at the relation between speech production and speech perception for the voicing contrast. *Phonetica*, *43*(1-3), 106–115.

Miller, J. L., Grosjean, F., & Lomanto, C. (1984). Articulation rate and its variability in spontaneous speech: A reanalysis and some implications. *Phonetica*, *41*(4), 215–225.

Miller, J. L., & Volaitis, L. E. (1989). Effect of speaking rate on the perceptual structure of a phonetic category. *Perception and Psychophysics*, *46*, 505–512.

Miller, L. (2017). The relationship between language proficiency and language attitudes: Evidence from young Spanish-English bilinguals. *Spanish in Context*, *14*(1), 99–123.

Morrongiello, B. A., Robson, R. C., Best, C. T., & Clifton, R. K. (1984). Trading relations in the perception of speech by 5-year-old children. *Journal of Experimental Child Psychology*, *37*(2), 231–250.

Mugari, V. (2014). Code-switching in zimbabwean urban grooves music. *Language Matters*, *45*(2), 224–236.

Nagao, K., & de Jong, K. (2007). Perceptual rate normalization in naturally produced rate-varied speech. *The Journal of the Acoustical Society of America*, *121*(5), 2882–2898.

Neiman, G. S., Klich, R. J., & Shuey, E. M. (1983). Voice onset time in young and 70-year-old women. *Journal of Speech, Language, and Hearing Research*, *26*(1), 118–123.

Ng, S. H., & He, A. (2004). Code-switching in trigenerational family conversations among Chinese immigrants in New Zealand. *Journal of Language and Social Psychology*, *23*(1), 28–48.

Nguyen, T. (2014). *Code switching: A sociolinguistic perspective*. Anchor Academic Publishing (aap_verlag).

Nielsen, K. (2011). Specificity and abstractness of vot imitation. *Journal of Phonetics*, *39*(2), 132–142.

Nittrouer, S. (2004). The role of temporal and dynamic signal components in the perception of syllable-final stop voicing by children and adults. *The Journal of the Acoustical Society of America*, *115*(4), 1777–1790.

Nycz, J. (2018). Stylistic variation among mobile speakers: Using old and new regional variables to construct complex place identity. *Language Variation and Change*, *30*(2), 175–202.

Nygaard, L. C., Sommers, M. S., & Pisoni, D. B. (1994). Speech perception as a talker-contingent process. *Psychological Science*, *5*(1), 42–46.

Olson, D. J. (2013). Bilingual language switching and selection at the phonetic level: Asymmetrical transfer in vot production. *Journal of Phonetics*, *41*(6), 407–420.

Olson, D. J. (2016). The role of code-switching and language context in bilingual phonetic transfer. *Journal of the International Phonetic Association*, *46*(3), 263–285.

Olson, D. J. (2017). Bilingual language switching costs in auditory comprehension. *Language, Cognition and Neuroscience*, *32*(4), 494–513.

Olson, D. J. (2023). Measuring bilingual language dominance: An examination of the reliability of the Bilingual Language Profile. *Language Testing*, 1–27.

Olson, D. J. (in press). Code-switching and language mode effects in the phonetics and phonology of bilinguals. In M. Amengual (Ed.), *Cambridge handbook of bilingual phonetics and phonology* (pp. 1–36).

Pardo, J. S. (2006). On phonetic convergence during conversational interaction. *The Journal of the Acoustical Society of America*, *119*(4), 2382–2393.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Perpiñán, S. (2017). Catalan-Spanish bilingualism continuum: The expression of non-personal Catalan clitics in the adult grammar of early bilinguals. *Linguistic Approaches to Bilingualism*, *7*(5), 477–513.

Piccinini, P., & Arvaniti, A. (2015). Voice onset time in Spanish–English spontaneous code-switching. *Journal of Phonetics*, *52*, 121–137.

Rácz, P. (2013). Salience in sociolinguistics: A quantitative approach.

Radford, A., Kim, J. W., Xu, T., Brockman, G., McLeavey, C., & Sutskever, I. (2022). Robust speech recognition via large-scale weak supervision. *arXiv preprint arXiv:2212.04356*.

Rickford, J. R., & McNair-Knox, F. (1994). Addressee- and topic-influenced style shift: A quantitative sociolinguistic study. In D. Biber & E. Finegan (Eds.), *Sociolinguistic perspectives on register* (pp. 235–276). Oxford University Press.

Roach, P. (1998). Myth 18: Some languages are spoken more quickly than others. In L. Bauer & T. Peter (Eds.), *Language myths* (pp. 150–158). Penguin.

Robb, M., Gilbert, H., & Lerman, J. (2005). Influence of gender and environmental setting on voice onset time. *Folia Phoniatrica et Logopaedica*, *57*(3), 125–133.

Ryalls, J. H., Zipprer, A., & Baldauff, P. (1997). A preliminary investigation of the effects of gender and race on voice onset time. *Journal of Speech, Language, and Hearing Research*, *40*(3), 642–645.

Sankoff, G. (1972). Language use in multilingual societies: Some alternative approaches. In J. B. Pride & J. Holmes (Eds.), *Sociolinguistics: Selected readings* (pp. 33–51).

Schertz, J., Cho, T., Lotto, A., & Warner, N. (2016). Individual differences in perceptual adaptability of foreign sound categories. *Attention, Perception, & Psychophysics*, *78*, 355–367.

Shen, A., Gahl, S., & Johnson, K. (2020). Didn't hear that coming: Effects of withholding phonetic cues to code-switching. *Bilingualism: Language and Cognition*, *23*(5), 1020–1031.

Singmann, H., Bolker, B., Westfall, J., & Aust, F. (2016). *Afex: Analysis of factorial experiments* [R package version 0.16-1]. https://CRAN.R-project.org/package=afex

Stevens, K. N. (1999). *Acoustic phonetics*. MIT Press.

Stölten, K., Abrahamsson, N., & Hyltenstam, K. (2015). Effects of age and speaking rate on voice onset time: The production of voiceless stops by near-native L2 speakers. *Studies in Second Language Acquisition*, *37*(1), 71–100.

Strange, W. (2011). Automatic selective perception (ASP) of first and second language speech: A working model. *Journal of Phonetics*, *39*(4), 456–466.

Summerfield, Q. (1981). Articulatory rate and perceptual constancy in phonetic perception. *Journal of Experimental Psychology: Human Perception and Performance*, *7*(5), 1074.

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

Swartz, B. L. (1992). Gender difference in voice onset time. *Perceptual and Motor Skills*, *75*(3), 983–992.

Sweeting, P. M., & Baken, R. J. (1982). Voice onset time in a normal-aged population. *Journal of Speech, Language, and Hearing Research*, *25*(1), 129–134.

Tendera, A., Rispoli, M., Senthilselvan, A., & Loucks, T. M. (2019). Early speech rate development: A longitudinal study. *Journal of Speech, Language, and Hearing Research*, *62*(12), 4370–4381.

Theodore, R. M., Miller, J. L., & DeSteno, D. (2009). Individual talker differences in voice-onset-time: Contextual influences. *The Journal of the Acoustical Society of America*, *125*(6), 3974–3982.

Thornburgh, D. F., & Ryalls, J. H. (1998). Voice onset time in spanish-english bilinguals: Early versus late learners of english. *Journal of communication disorders*, *31*(3), 215–229.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, *58*(1), 267–288.

Toribio, A. J. (2002). Spanish-English code-switching among US Latinos. *International Journal of the Sociology of Language*, *158*(2002), 89–119.

Toribio, A. J., Bullock, B. E., Botero, C. G., & Davis, K. A. (2005). Perseverative phonetic effects in bilingual code-switching. In R. Gess & E. Rubin (Eds.), *Theoretical and experimental approaches to romance linguistics* (pp. 291–306). John Benjamins Publishing Company.

Trudgill, P. (1986). *Dialects in contact*. Wiley-Blackwell.

Tyler, J., & Feldman, R. (2004). Cognitive demand and self-presentation efforts: The influence of situational importance and interaction goal. *Self and Identity*, *3*(4), 364–377.

Van Engen, K. J., Baese-Berk, M., Baker, R. E., Choi, A., Kim, M., & Bradlow, A. R. (2010). The Wildcat Corpus of native- and foreign-accented English: Communicative efficiency across conversational dyads with varying language alignment profiles. *Language and Speech*, *53*(4), 510–540.

Verhoeven, J., De Pauw, G., & Kloots, H. (2004). Speech rate in a pluricentric language: A comparison between Dutch in Belgium and the Netherlands. *Language and Speech*, *47*(3), 297–308.

Volaitis, L. E., & Miller, J. L. (1992). Phonetic prototypes: Influence of place of articulation and speaking rate on the internal structure of voicing categories. *The Journal of the Acoustical Society of America*, *92*(2), 723–735.

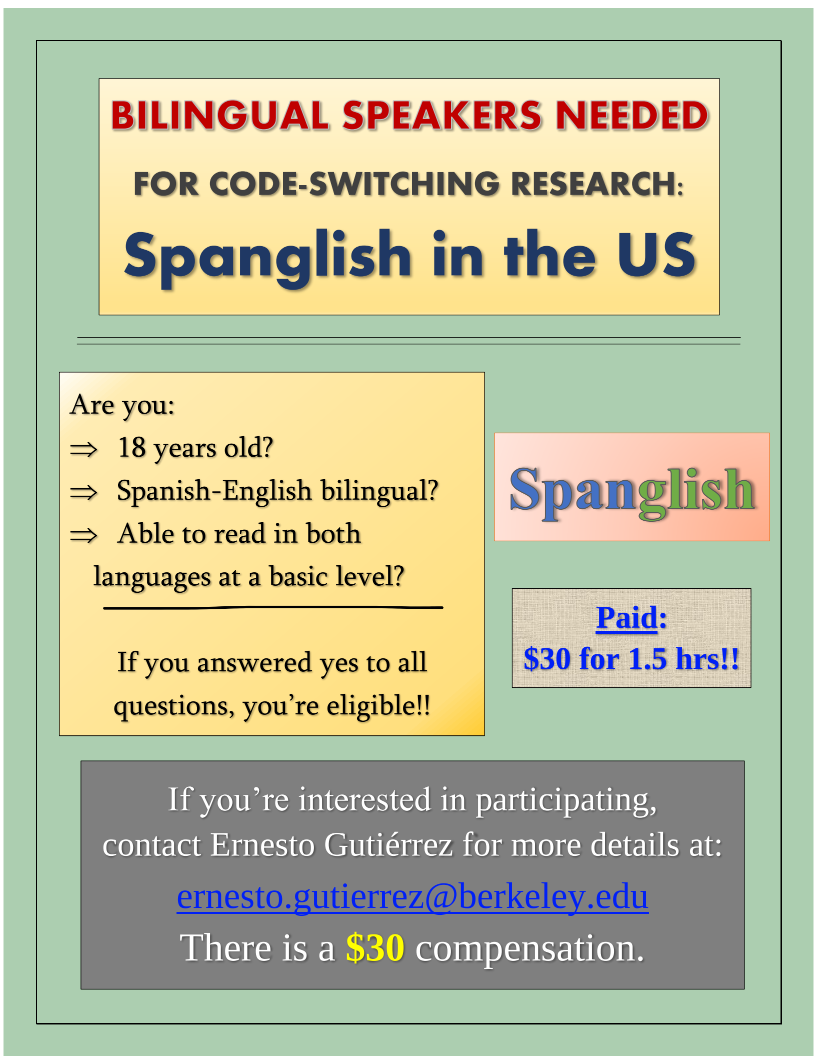
Wei, L. (2005). “how can you tell?”: Towards a common sense explanation of conversational code-switching. *Journal of Pragmatics*, *37*(3), 375–389.

Weismer, G. (1980). Control of the voicing distinction for intervocalic stops and fricatives: Some data and theoretical considerations. *Journal of Phonetics*, *8*(4), 427–438.

Younes, S. A., & Gathercole, V. C. M. (2020). Phonological interaction in Spanish-English bilinguals: Effects of cognate usage on voice onset time of voiced stops. *International Journal of Bilingual Education and Bilingualism*, 1–19.

Appendix A

Recruitment flyer



Appendix B

Word list items

English words

Bilabial	**Coronal**	**Velar**
Path	Tag	Camper
Pass	Tackle	Cat
Package	Task	Car
Pallet	Tally	Cabbage
Pen	Tell	Kernel
Pears	Tent	Kettle
Pet	Ten	Keg
Pest	Test	ketchup
Peel	Teen	Keen
Peaky	T-shirt	Keyword
Peer	Team	Keys
Pea	Tea	Keyboard
Post	Toe	Cold
Poultry	Toast	Coat
Poking	Towing	Comb
Pose	Toad	Coal
Pooch	Tooth	Cool
Poop	Tool	Cooler
Pool	Tomb	Coop
Poo	Toot	Coo

Table B.1: English word list stimuli.

Spanish words

Bilabial	**Coronal**	**Velar**
Párpado	Tanto	Cama
Pato	Talla	Casa
Panza	Taza	Carta
Palo	Tapa	Calle
Pera	Techo	Queso
Peso	Tela	Queja
Perro	Tema	Que
Pena	Tecla	Quema
Pico	Tibio	Quinta
Pila	Tiza	Quicio
Piso	Tina	Quiste
Piña	Tiro	Quince
Pobre	Toro	Comal
Poco	Todo	Codo
Pozo	Torpe	Cofre
Pollo	Torre	Coche
Pulcro	Túnel	Cuyo
Pulga	Tufo	Cupo
Puño	Tunda	Cura
Pulpo	Tuza	Cuna

Table B.2: Spanish word list stimuli.

Appendix C

Passage items

For each of the ten passages presented below, the first language used (either English or Spanish) is visualized here in italics font style, with the following language in monospace font style. The font styles alternate back and forth to indicate language switches. These font styles are used in this appendix to facilitate the identification of language distinctions and switch sites for the reader.

Passage 1

My son has always been a huge soccer fan. `Cuando sale con sus amigos lo único que quiere hacer es jugar y jugar. Además` *coaches have told him that he is really good! Last month* `tuvo su último partido del campeonato` *taking on his rival school for the last time this season. Unfortunately, the end of the game was not what we wanted to see because* `perdieron 4 a 3 y regresaron a casa sin la copa. Ahora el entrenador necesita` *pin down the improvements necessary for next season!*

Passage 2

When I was in middle school, `todos mis amigos vivían lejos de mí, así que para ellos` *coming over to play at my house wasn't always possible.* `Por eso mi mamá a veces iba por ellos a sus casas, pero` *taking them back afterwards was even more challenging.* `Como ella trabajaba en la noche, salía de casa a las 6 de la tarde. Por suerte,` *politely asking my older sister to take my friends home usually worked, though.*

Passage 3

I have always been obsessed with magicians. Imagine, `por ejemplo, poder usar una varita mágica,` *talking to animals whenever you want and even flying!... Unfortunately, that only happens in movies. For me,* `todo comenzó después de un evento de magos.` *People were cheering after every trick, but I was too amazed to even make a sound at all. Anyways,* `cada vez que terminaba un truco nuevo me convencía más y más que la magia era real!` *Can't wait to start practicing magic tricks so I can do that someday!*

Passage 4

Siempre me ha gustado mucho estar al aire libre. Por eso, talking to people while walking outside is one of my favorite things to do. Also *patinar en el parque es algo que no se puede superar.* Can you believe my parents used to be scared of me skating? They used to say *"¡Te vas a caer!" Ellos estaban,* peeved because they thought I was going to hurt myself. Some time after, *cuando finalmente sucedió, recuerdo que mis papás se molestaron demasiado y estuvieron a punto de quitarme mi patineta, pero aun así es mi actividad favorita.*

Passage 5

It is always an adventure trying to introduce my family to others. For example, tener cinco hermanos no es muy común en los Estados Unidos, pero *coming from Latin America, big families are very common.* Cada vez que menciono a mis cinco hermanos, *people tend to look shocked and start asking too many questions like "How do you remember them all?" or "how many rooms are in your house?"* Pero la mejor de todas es cuando les preguntan a mis papás cómo lidiaron con nosotros cuando éramos *teenagers. I honestly could not have done what they did.*

Passage 6

Mis recuerdos favoritos son cuando solíamos visitar places like Disneyland or Six Flags with my family. We used to go *casi todos los veranos y a veces también en ocasiones especiales, por ejemplo* turning 10, and 11, and all birthdays before that. Oh! *También después. De una forma u otra se convirtió en tradición, como usar* costumes during Halloween just to walk around town *pidiendo dulces de casa en casa.*

Passage 7

Nunca me han gustado los vegetales: carrots, spinach, broccoli, none of them. *"Tienes que comer vegetales" me dicen todos los días, pero la verdad es que* putting them in my mouth makes me nauseous. Besides *cada doctor también me ha dicho lo mismo:* "Two veggies a day, that's all you need!" *Puede que lo intente de nuevo algún día.*

Passage 8

My first ever earthquake was a scary experience. Tembló en la ciudad de México cuando tenía 10 años. *People were freaking out because no one had ever felt an earthquake that strong, imagine how I felt that day!* Cuando empezó sentí mucho miedo porque algunos edificios *caved in around me. So you can see why* parecía que era el fin del mundo. Todos estaban corriendo y gritando sin saber qué hacer. Los estudiantes en las escuelas se escondieron debajo de *tables and chairs. Otherwise, it could have been worse.*

Passage 9

My main hobby for the last couple of years has been building Legos. `Para mí construir legos es relajante y reduce el estrés. Lo gracioso es que` *people often think the opposite! I always laugh when* `todos dicen que es estresante y tienes que tener mucha paciencia para poder construirlos. Como yo lo veo es:` *kids can do it and they are made for kids so it's pretty straight forward! I do not know why people don't like them.* `Quisiera que fueran más baratos para poder comprar más, pero por ahora lo único que puedo hacer es` *take my time finishing the last one I have.*

Passage 10

Being an influencer must be so difficult because `tomar fotos cada cinco minutos,` *paying for flights every other week,* `cuidar todo lo que dices y sobre todo` *talking to people online non-stop is exhausting. I really don't understand how they do it.* `Para mí tiene sentido,ya que probablemente sea la carrera del futuro pero aun así es muy complicado ya que los` *kids are going to have a hard time in the future.*

Appendix D

Puzzle image

This image was created by Jacky Duong. The creator has given her permission for other researchers to make use of this instrument for their own research projects, provided it is for non-commercial use.

Appendix E

Interview questions

- Are you a student?
- What do you study?
- Do you like it?
- Which has been your favorite class in your major(s) so far?
- And which has been your least favorite class?
- What type of job would you like after graduation?
- Are you currently employed?
- What do you do for work?
- Where was the last place you went on vacation?
- If money wasn't an issue, where would you like to go on vacation?
- What do you do on a typical day?
- What is your favorite dish?
- Do you like to cook?
- What do you usually cook?
- Who do you typically speak in Spanish with?
- Who do you typically speak in English with?
- Do you use Spanglish with anyone?
- What are the benefits of speaking two languages?
- What are some of the struggles of being bilingual?

Appendix F

Social factor classification

Speaker-inherent social variables:

1. Sex-gender.
2. At what age did you start learning ENGLISH?
3. At what age did you start learning SPANISH?
4. At what age did you start to feel comfortable using ENGLISH?
5. At what age did you start to feel comfortable using SPANISH?
6. When you talk to yourself, how often do you talk to yourself in ENGLISH?
7. When you talk to yourself, how often do you talk to yourself in SPANISH?
8. When you talk to yourself, how often do you talk to yourself in OTHER LANGUAGES?
9. When you count, how often do you count in ENGLISH?
10. When you count, how often do you count in SPANISH?
11. When you count, how often do you count in OTHER LANGUAGES?
12. How well do you speak ENGLISH?
13. How well do you speak SPANISH?
14. How well do you understand ENGLISH?
15. How well do you understand SPANISH?
16. How well do you read ENGLISH?
17. How well do you read SPANISH?

18. How well do you write ENGLISH?
19. How well do you write SPANISH?
20. I feel like myself when I speak ENGLISH.
21. I feel like myself when I speak SPANISH.
22. I identify with an ENGLISH-speaking culture.
23. I identify with a SPANISH-speaking culture.
24. It is important to me to use (or eventually use) ENGLISH like a native speaker.
25. It is important to me to use (or eventually use) SPANISH like a native speaker.
26. I want others to think I am a native speaker of ENGLISH.
27. I want others to think I am a native speaker of SPANISH.
28. BLP-produced dominance score.

Environmental social variables:

1. Highest level of formal education.
2. How many years of classes (grammar, history, math, etc.) have you had in ENGLISH (primary school through university)?
3. How many years of classes (grammar, history, math, etc.) have you had in SPANISH (primary school through university)?
4. How many years have you spent in a country/region where ENGLISH is spoken?
5. How many years have you spent in a country/region where SPANISH is spoken?
6. How many years have you spent in a family where ENGLISH is spoken?
7. How many years have you spent in a family where SPANISH is spoken?
8. How many years have you spent in a work environment where ENGLISH is spoken?
9. How many years have you spent in a work environment where SPANISH is spoken?
10. In an average week, what percentage of the time do you use ENGLISH with friends?
11. In an average week, what percentage of the time do you use SPANISH with friends?
12. In an average week, what percentage of the time do you use OTHER LANGUAGES with friends?

13. In an average week, what percentage of the time do you use ENGLISH with family?
14. In an average week, what percentage of the time do you use SPANISH with family?
15. In an average week, what percentage of the time do you use OTHER LANGUAGES with family?
16. In an average week, what percentage of the time do you use ENGLISH at school/work?
17. In an average week, what percentage of the time do you use SPANISH at school/work?
18. In an average week, what percentage of the time do you use OTHER LANGUAGES at school/work?
19. On a regular basis, what kind of non-native accents do you typically hear?