ENGLISH SPEAKERS' PERCEPTION OF BILINGUAL SPANISH-ENGLISH SPEECH: ANALYZING THREE PHONETIC DIFFERENCES BETWEEN ENGLISH AND SPANISH

by

CAROLINE SCHNEIDER

(Under the Direction of Margaret E.L. Renwick and Timothy Gupton)

ABSTRACT

The debate on how bilingual and monolingual speech differs has been ongoing since the creation of linguistics as a scientific field. In this thesis, Spanish-English bilingual and English monolingual speech will be investigated regarding three phonetic differences that exist between English and Spanish. The second experiment of this study includes a perception task that investigates how monolinguals process the two groups' speech and their attitudes towards both groups. Results of the two experiments indicate that VOT differences in the two languages it the most salient feature of the three investigated. Additionally, while naïve listeners can accurately distinguish monolingual speech, the same cannot be said with bilingual speech. Finally, listeners harbor positive and negative attitudes towards both groups, revealing a more complex picture of listener attitudes than what previous literature concluded.

INDEX WORDS: Perception, Attitudes, Language Acquisition, Hispanic Linguistics, Phonology

ENGLISH SPEAKERS' PERCEPTION OF BILINGUAL SPANISH-ENGLISH SPEECH: ANALYZING THREE PHONETIC DIFFERENCES BETWEEN ENGLISH AND SPANISH

by

CAROLINE SCHNEIDER

B.A., The University of North Carolina at Chapel Hill, 2022

A Thesis Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment

of the Requirements for the Degree

MASTER OF ARTS

ATHENS, GEORGIA

2024

ENGLISH SPEAKERS' PERCEPTION OF BILINGUAL SPANISH-ENGLISH SPEECH: ANALYZING THREE PHONETIC DIFFERENCES BETWEEN ENGLISH AND SPANISH

by

CAROLINE SCHNEIDER

Major Professors: Margaret E.L. Renwick
Timothy Gupton

Committee: Chad Howe
Jon Forrest

Electronic Version Approved:

Ron Walcott
Vice Provost for Graduate Education and Dean of the Graduate School
The University of Georgia
May 2024

ACKNOWLEDGEMENTS

I would like to express my gratitude to all the members of my committee, but especially my co-advisors, Dr. Margaret Renwick and Dr. Timothy Gupton. Without them I would be unable to even start such a laborious project, let alone finish. Both provided me with invaluable support and advice through the creation of this thesis. I am extremely grateful for all I was able to learn and accomplish as your advisee.

I would also like to acknowledge the friends and colleagues who supported me through this time. I am truly grateful to have met such wonderful people in Athens and I am excited to see how we all grow as people and academics.

Last, but certainly not least, I must thank my family. Maggie, Abbey, and Erin, you are models of success that I will always strive to reach. I can always count on you to make me laugh and to give the best advice. I can truly say that I have the best and most amazing sisters of anyone I know. Mom and Dad, I could not have moved to Georgia and completed this degree without you and your support. I am forever grateful for everything that you have sacrificed for us. You are my biggest motivators and supporters.

TABLE OF CONTENTS

Page

ACKNOWLEDGEMENTS iv

LIST OF TABLES vii

LIST OF FIGURES viii

CHAPTER

1 INTRODUCTION & BACKGROUND 1

Introduction 1

Study objectives 2

Research questions 2

2 RELEVANT LITERATURE 3

Comparing Spanish and English phonologies 3

Perception: An exemplar-based approach 11

Language and Attitudes 17

3 EXPERIMENT 1 22

Introduction 22

Methods and materials 22

Experiment hypotheses 26

Descriptive results 26

Quantitative analysis 30

Conclusion 32

4 EXPERIMENT 2 34

Introduction 34

Methods and materials 34

Experiment hypotheses 37

Descriptive results 37

Quantitative analysis 45

Qualitative analysis 50

5 DISCUSSION AND CONCLUSIONS 52

Summary 52

Limitations 55

Conclusions 57

REFERENCES 60

APPENDICES

A EXPERIMENT 1 STIMULI 68

B EXPERIMENT 2 QUESTIONS 69

LIST OF TABLES

Page

Table 3.1: Example Experiment 1 stimuli 23

Table 3.2: /i/ and /ɪ/ T-tests 30

Table 3.3: /l/ T-tests 31

Table 3.4: /t/ and /d/ voice onset time T-tests 32

Table 4.1: Rate of accuracy in nativeness judgments 42

Table 4.2: Nativeness judgments linear ordinal regression 44

Table 4.3: Nativeness judgments linear ordinal regression – Threshold estimates 44

Table 4.4: Wealth linear ordinal regression 46

Table 4.5: Wealth linear ordinal regression – Threshold estimates 46

Table 4.6: Education linear ordinal regression 47

Table 4.7: Education linear ordinal regression – Threshold estimates 47

Table 4.8: Politeness linear ordinal regression 48

Table 4.9: Politeness linear ordinal regression – Threshold estimates 48

Table 4.10: Open-ended question comment themes 51

LIST OF FIGURES

Page

Figure 3.1: Spectrogram image - Can you say “Pete”? 24

Figure 3.2: Spectrogram Image - Can you say “late?” 25

Figure 3.3: Spectrogram Image - “I like to text” 25

Figure 3.4 Monolingual /i/ formants 27

Figure 3.5: Bilingual /i/ formants....27

Figure 3.6: Monolingual /ɪ/ formants....27

Figure 3.7: Bilingual /ɪ/ formants....27

Figure 3.8: Bilingual Vowel Duration 28

Figure 3.9: Monolingual Vowel Duration 28

Figure 3.10: [l] formant values 29

Figure 3.11: [ɫ] formant values 29

Figure 3.12: /d/ and /t/ voice onset times....29

Figure 4.1: Languages Studied 35

Figure 4.2: Highest level of foreign language courses take....35

Figure 4.3: Participant age 35

Figure 4.4: Nativeness Judgments towards speech of monolingual and bilinguals....38

Figure 4.5: Nativeness judgments by feature: /i/ and / ɪ/ 39

Figure 4.6: Nativeness judgments by feature: [l] and [ɫ] 39

Figure 4.7: Nativeness judgments by feature: /d/ and /t/ 39

Figure 4.8: Social Judgment ratings: Education 40

Figure 4.9: Social Judgment ratings: Politeness 40

Figure 4.10: Social Judgment ratings: Wealth 40

Figure 4.11: Nativeness Judgments by audio condition: Unmodified audio 41

Figure 4.12: Nativeness Judgments by audio condition: Modified audio 41

Figure 4.13: Accuracy of Nativeness Judgments 42

Figure 4.14: Code in RStudio package *ordinal*: Nativeness judgments regression.......................42

Figure 4.15: Code in RStudio package *ordinal*: Wealth judgments regression............................46

Figure 4.16: Code in RStudio package *ordinal*: Education judgments regression 47

Figure 4.17: Code in RStudio package *ordinal*: Politeness judgments regression 48

Figure 4.18: Comment Themes...50

CHAPTER 1

INTRODUCTION

1.1 Introduction

A monolingual nation in a multilingual world

In its current state, the United States is a rarity when it comes to the linguistic backgrounds of its citizens. Most Americans are monolingual, with many bilinguals learning their second language in adulthood rather than childhood. According to a 2022 census, approximately 83% of US citizens over the age of 18 only speak English (U.S. Census Bureau 2022). Additionally, only 20% of K-12 students study a language other than English (Baker & Wright 2021, 2). Also, despite the rapid globalization of the world, there has been a decrease in foreign language programs at four-year universities (Baker & Wright 2021, 2). This may not stick out as a problem for some; if you can live your life entirely in one language, why learn more? However, with an increasingly connected world, there are some issues with this. First, it is more cognitively demanding to learn languages as one grows older (Rahman et al. 2017; Hartshorne et al. 2018). Therefore, second language education may decrease in its efficacy the older the student is. Next, the US is a global powerhouse, attracting immigrants from all over the world. Due to the status of English as the only majority language, immigrants are all but forced to learn it. Additionally, native speakers may not be understanding towards these learners, as they have never had to navigate their lives in another language. Rejection and ridicule of immigrants learning English both perpetuates and is exasperated by power structures in which the majority language (in this case, English) has more social, and at times legal, prestige than minority languages (Fowler et al. 1979). To researchers, it begs the question: how prejudiced are monolinguals towards bilinguals? Can prejudices be remedied or mitigated? Certainly, bilinguals are not only judged for their status as a native or nonnative speaker; race, gender, physical appearance, and much more contribute to the judgments we make towards people daily (McBride 2015). Therefore, as linguists, we can narrow the question: how prejudiced are monolinguals towards bilingual speech production?

1.2 Study objectives

The following study aims to contribute to the existing knowledge on monolingual and bilingual production, as well as monolingual perception. In all, the two experiments will address existing questions related to exemplar models. The first experiment will illuminate similarities and differences in bilingual-monolingual speech. This experiment will focus on three phonological contrasts that do not exist in both languages. These include the English tense-lax vowel distinction, /l/ allophonic variation, and /t/ and /d/ voice onset times. All three are distinctions that L2 English speakers are anticipated to have difficulty acquiring. The second experiment aims to answer questions on how American monolinguals perceive and judge the speech of their bilingual peers. In this experiment, participants will listen to recordings from the first experiment and respond to questions about the speakers in the recordings.

1.3 Research questions:

1. Which of the three phonetic differences is the most salient between monolingual and bilingual English speakers?
2. Can monolingual English speakers differentiate between monolingual English speakers and bilingual Spanish-English speakers accurately when only using acoustic cues?
3. Based on English monolinguals' ability to differentiate between the two groups, what attitudes do they express to both groups of speakers?

CHAPTER 2

RELEVANT LITERATURE

2.1 Comparing Spanish & English phonologies

Prior to introducing the phonetic features under investigation, it is important to discuss major acquisition models that attempt to explain L2 phonology acquisition. First, the Speech Learning Model introduced by Flege postulates that learners may use their L1 phonological system to learn their L2 phonological system (1995). This means that learners may position new L2 sounds in relation to the L1 sound inventory. The SLM postulates a perceptual basis for the difficulties learners may experience learning the L2. Although theoretically learners' phonetic categories can evolve to accommodate both the L1 and L2, their phonological systems may remain as one. On a similar note, the Perceptual Magnet Effect (PME) also proposes that perception plays a significant role in L2 phonology acquisition (Kuhl & Iverson 1995). This effect argues that exposure to the L1 from infancy changes how a person perceives distances between different sounds. When an adult native speaker hears a sound, they judge this in comparison to a prototype sound. Sounds that share phonetic properties of prototypes are "drawn" to them, perceptually shortening the distance between the variants and the prototypes. These perceived distances can become problematic when learning an L2, where the perceived and real distances between sounds are different than the L1. The Perceptual Assimilation Model (PAM) predicts that phonological acquisition is effected by how similar or dissimilar phonological contrasts are in the L1 versus the L2 (Best 1991). Under this model, L2 phonological contrasts assimilate into the L1 phonological space. L2 phonemes are likely to be acquired successfully if they are gesturally similar to L1 phonemes. In comparison, L2 phones will be more difficult to acquire when a phonological contrast exists in the L2 that does not exist in the L1. Lastly, the Contrastive Analysis Hypothesis postulates that differences between the L1 and L2 (phonological or otherwise) will cause problems for the learner (Wardhaugh 1970). Some linguists have suggested that L1 phonology exists as a constraint on an L2 learner's pronunciation (Eckman 2004). In this case, contrastive analysis can be most predictive of challenges learners will face when it comes to acquiring L2 phonology. Contrastive analysis would entail analyzing the phonological inventories of both languages and noting the differences and similarities between the two. This analysis predicts that when assigning allophones in the L1 to different phonemes in the L2, learners will have significant difficulty.

Additionally, salience can cause challenges for L2 learners, as differences between the L1 and L2 may not be noticeable enough for learners to acquire, but very noticeable by native listeners. Previous work indicates that salience is relative; meaning that a segment or feature is more noticeable due to differences in pitch, prosody, loudness, and more compared to the other segments around it (Drager & Kirtley 2016). In regards to second language acquisition, salience can also appear as unexpectedness, meaning that a segment is produced in a way that a learner does not anticipate (Boswijk & Coler 2020). For instance, an L1 English speaker learning Spanish may notice trilled [r] more than tap [ɾ] because [r] does not exist as an allophone nor phoneme in English. Meanwhile, [ɾ] may be harder for learners to notice, therefore harder to produce. Additionally, as tap exists as an allophone of US English /d/, learners may assimilate it to that category, rather than to a rhotic category (Boomershine 2007). Even though English speakers can produce tap, the different allophonic distribution it has in the two languages can negatively affect how they acquire it in the L2. Therefore, both salience and allophonic distribution have significant impact on the successful acquisition of phonological contrasts in the

L2.. The three differences in this study range in their proposed salience, as to capture both types of challenge.

This study aims to focus on three specific phonological differences between Spanish and English: 1) /i/ and /ɪ/ 2) [l] and [ɫ], and 3) /t/ and /d/ Voice Onset Times (VOTs). All three are differences that the SLM, PME, PAM, and the Contrastive Analysis Hypothesis would anticipate learners to have difficulty acquiring. Additionally, the three vary regarding their level of salience to both native speakers and to listeners. This study aims to address the question of which difference is acoustically most salient and which affects listeners' judgments most.

2.1.1 Tense-Lax Vowels

The first difference is the tense-lax vowel distinction that exists in English, but not in Spanish. This study will focus on /i/ and /ɪ/. Although this distinction is very salient to native English speakers, it can be difficult for English learners to both perceive the difference and produce it. This is especially the case since some words can be produced with either phoneme and be intelligible. For instance, the word India could be pronounced as either [ɪndiə] or [indiə] and be understood as referring to the Southeast Asian country. Since it can be difficult for learners coming from a language without tense and lax vowels to hear the difference, I propose that it will be a challenge for them to produce it as well.

Previous literature has indicated that English uses multiple cues to distinguish these vowels; both spectral features and duration are involved in the tense-lax distinction (Fox et al. 1995). English speakers tend to use multiple cues in both perception and production. Fox et al.'s study found that English speakers typically employed three variables to distinguish tense and lax vowels. These included cues such as high/low, front/black, central/noncentral, and duration. Meanwhile, Spanish-English bilinguals employed only two variables to distinguish the vowels.

English proficiency was found to be a significant factor, indicating that higher proficiency led bilingual participants to distinguish the vowels in a similar manner as the English monolingual participants. The researchers concluded that vowel distinction is highly reliant on listeners' L1 vowel spaces. When the listener has an L1 without a tense-lax distinction (such as Spanish), fewer acoustic cues are employed to perceive this distinction. As listeners become more experienced with a language with a tense-lax distinction, they start to utilize more acoustic cues to distinguish vowels.

A study in L1-Spanish L2-English speakers' perception of /i/ and /ɪ/ found that these speakers distinguish tenseness and laxness with a variety of acoustic cues. Morrison (2009) hypothesizes a developmental process where L1 Spanish-L2 English speakers utilize all available acoustic cues from the start. This goes against proposed processes of Bohn and Escudero (1995) and Boersma (2004), who proposed that duration is used in the first stage of acquisition, then spectral features are used in the second stage of acquisition. Unlike other proposals, Morrison's does not consider the perception of monolinguals in the creation of his proposed developmental process. Morrison's results also indicated that L1 dialect was not a significant factor in the acquisition of the tense/lax distinction; participants from different dialects demonstrated the same behaviors when reacting to /i/ and /ɪ/.

Shortly later, a perception study by Kondaurova & Francis (2010) found that English speakers used spectral features, particularly the first and second formants, to determine whether a vowel was tense or lax. Meanwhile, Spanish speakers attempted to use length as the distinguishing factor. As expected, English speakers were very accurate in judging the distinction, while Spanish speakers' accuracy depended on their proficiency in the language. As this was a perception study, the question remains of how Spanish speakers learning English produce the distinction. It can be predicted that as length is used to detect the distinction, it could also be used to produce (or at least attempt to produce) the distinction.

A recent study from Cebrian et al. (2021) found that Spanish/Catalan L1-English L2 learners have difficulty perceiving the difference between /i/ and /ɪ/, even when they have high English proficiency. Participants with lower English proficiency typically grouped /ɪ/ along with /e/ when asked to assign it to a category. Participants with higher English proficiency were more likely to assign /ɪ/ to /ɪ/, but also assigned it to /i/ frequently. Additionally, in a production task, /ɪ/ was the least accurately produced. Cebrian et al. concluded that pedagogical practices could be at fault for these learners' vowel categories; many of the participants had learned that tense /i/ was long and lax /ɪ/ was short. This led to the participants distinguishing the two sounds using duration only, rather than also focusing on spectral differences.

Difficulties acquiring tense-lax vowel distinction are not limited to those learning English; Navarra et al.'s (2005) study investigated the difference between /e/ and /ɛ/, which exist as phonemes in Catalan, but not in Spanish. The researchers found that among Catalan-Spanish bilinguals, Spanish dominant bilinguals had significantly more difficulty perceiving a difference between the two phonemes compared to their Catalan-dominant counterparts. This study again reinforces the idea that phonemic distinction, such as tense-lax distinction, is difficult to perceive if a speaker's dominant language does not have such a distinction.

In conclusion, tense-lax vowel distinctions can be difficult for learners to acquire when they come from an L1 background without this distinction. This includes problems in both perception and production of the vowels. These numerous challenges may be due to salience of the feature and/or pedagogical practices. Speakers of a language with this distinction will employ multiple acoustic cues to perceive and produce the vowels, but learners do not, causing them to categorize the vowels differently than the target categorizations. It can be predicted that the lax vowel /ɪ/ will be harder for the bilingual group to produce, leading to its formants being different than the formants of the monolingual group.

2.1.2 Laterals

The second difference evaluated was the allophonic distribution of [l] and velarized (or dark) [ɫ]. While in United States English, [l] is expected in syllable initial positions and [ɫ] expected in syllable final positions, Spanish does not have this distribution. Rather, most varieties of Spanish only have [l] as an allophone of /l/. In fact, dark /l/ is found far less cross-linguistically than light /l/ (Recasens 2012). Typically, clear /l/ is expected to have higher frequencies than dark /l/. The acoustic split between clear and dark appears to be around 1300-1400HZ (F2) for /i/ contexts and 1000HZ (F2) for /a/ contexts. Frequency range is greater for clear /l/ than dark /l/. F2 values are typically higher in initial positions than in final positions. Languages with a larger F2 difference between the initial and final positions are those that differentiate between the two allophones.

This is supported by studies such as Barlow et al. (2013) and Barlow (2014) which studied the production of /l/ by English monolinguals and Spanish-English bilinguals. These studies revealed that both groups have the same distribution of /l/ when speaking English, however that [l] and [ɫ] were acoustically different between the two groups. Formant values for /l/ were overall higher for Spanish-English bilinguals compared to the English speakers. The investigators concluded that the bilinguals have two distinct phonemic inventories: one English inventory and one Spanish. These two inventories allow the speakers to adhere to the expected allophonic variation of /l/ in both languages. Despite this, bilinguals still produced acoustically distinct laterals when speaking in English compared to their monolingual counterparts, indicating transfer from Spanish to English. The studies also found that the younger the bilinguals were introduced to both languages, the closer their realizations of /l/ were to the monolinguals. As Spanish does not have the same allophonic distribution of /l/, it can be predicted that the late bilinguals are more likely to realize /l/ as [l] in all environments. Additionally, these studies indicate that Spanish-English bilinguals in the current study will have different acoustic realizations of /l/, particularly that they will have higher formant values than the English-speaking participants.

Unlike tense-lax vowel distinctions, the literature does not indicate that pedagogical practices have direct influence on the acquisition of English laterals. It is possible that language instructors are not trained to know about the allophonic distinction of /l/, therefore it is not something that they teach to learners. Additionally, most native English speakers are not aware of this distinction, meaning it has low salience for most speakers. This, along with the fact that Spanish does not have this allophonic distribution, can contribute to an L2 learner failing to acquire [ɫ] and to know when and when not to use it. The inclusion of this low-salience allophonic pair is intended to allow for comparisons between it and other distinctions of higher-salience (i.e. tense-lax vowel distinction). Although most native speakers are unaware of this allophonic pair, the current study will investigate if it can impact speech perception, even unconsciously.

2.1.3 Voice Onset Time

The last difference investigated in this study is Voice Onset Time (VOT) in Spanish and English. VOT is the time between a plosive burst and the start of vocal fold vibration (Cho & Ladefoged 1999; Abramson & Whalen 2017). VOT allows for languages to have differences in voicing for segments such as stops and affricates. Although voicing is usually described as a binary feature in phonological feature theory, VOT exists as an acoustic continuum (Flege and Eefting 1995). The voicing binary that exists in many languages is determined by arbitrary cut-offs in this continuum. Many languages can be described as “short lag” language, meaning the cut-off between voiced and voiceless segments is temporally short. The opposite would be described as a “long lag” language.

When studying L1 Brazilian Portuguese L2 English speakers’ production of English voiced and voiceless stops, this group of speakers produced voiceless stop with short-lag VOT values, similar to the VOT values found in Brazilian Portuguese (Major 1986). These speakers would need to considerably modify their VOT patterns to be considered proficient at producing English voiceless stops. In a second task, L1 English listeners rated how “accented” each speaker sounded. There was a correlation between “accented” judgment and VOT values; the closer a speaker’s values were to target VOT values, the less accented they were judged.

In their 1995 study, Flege and Eefting investigated the VOTs of voiceless plosives /p t k/ and voiced plosives /b d g/ by native English speakers and native Spanish speakers learning English. Accordingly, Flege and Eefting found that the Spanish speakers produced plosives with shorter VOTs, even when speaking in English. This indicates that L2 learners of a language will transfer features such as VOT to their L2.

The last difference this study evaluates is the voice onset time (VOT) of /d/ and /t/ between English and Spanish. While acoustically VOT exists on a continuum, every language has different phonological cut-offs that influence the perception the speakers have of voicing. VOTs in English plosives are typically longer than those of Spanish plosives. For instance, voice onset time of 0-25 milliseconds will likely be perceived as a voiceless stop consonant in Spanish, but a voiced stop consonant in English (Wig and García-Sierra, 2021). Additionally, they found that when prompted to listen to /ta/ and /da/ in both Spanish and English contexts, Spanish-English bilinguals were able to more accurately judge which phoneme was being produced according to the language context it appeared in. Meanwhile, English monolinguals judged the phonemes according to the expected VOTs found in English in both English and Spanish contexts. These results indicate that language contexts are important to experienced listeners (in this case, Spanish-English bilinguals) but had no impact for naive listeners (English monolinguals).

Since phonemes for /d/ and /t/ both exist in Spanish, learners may not put extensive effort into learning the English acoustics for these two sounds. Rather, they may assimilate the English phonemes into their Spanish categories. Past studies indicate that L2 speakers of English in the current study will have shorter VOTs for /d/ and /t/, more appropriate for Spanish, when speaking in English. It can be predicted that Spanish speakers in the current study will have shorter VOTs on account of Spanish influence compared to their English-speaking counterparts. In particular, VOTs for syllable initial /t/ are predicted to be significantly shorter when produced by the bilingual group in comparison to the monolingual group.

2.2 Perception: An exemplar-based approach

Exemplar models illustrate well how listeners process and build archetypes of linguistic nature using both linguistic and non-linguistic information (Goldrick & Cole 2023). Originally proposed as a psychological theory, exemplar models have been adopted by linguists to explain both speech perception and production (Goldinger 1996; Drager & Kirtley 2016; Pierrehumbert 2000;2001). These models exist like intricate webs that store any and all necessary information to help listeners normalize the speech of others (Ambridge 2019). Exemplar models work by collecting and storing both episodic and generalizable information. That is to say, these webs contain information on each experience hearing a token, its phonological structure along with information about individual speakers. As children receive input in their L1, generalizable knowledge from the exemplars is combined to aid in the acquisition of the L1's phonological, morphological, and syntactic structures. Even as adults, listeners continue to store information in their exemplar clouds, creating enriched webs of information that generalize and aid their understanding of the L1. Over time, listeners create perceptual categories that Johnson describes as "the set of all experienced instances of the category," meaning, a listener will store every instance they hear of a word, the meaning, who said it, and in what context (1997, 146). Though these models accommodate for certain weaknesses that generative models cannot (sociolinguistic information, etc), it is unknown precisely what weight each piece of linguistic or non-linguistic information has.

A common challenge for exemplar models is how much memory an individual listener is supposed to store. To address this, I turn to connectionist exemplar models, which argue that not all information is stored, at least not to the same degree (Kruschke 1992, Johnson 1997). Each phonological exemplar is connected to nodes that include information about the exemplar, such as acoustic information and social information about the speaker. As listeners gain more language input, the connections between the exemplar and its corresponding information are either fortified or weakened. Over time, the listener learns what information is important to store, giving weight to different types of information for each exemplar.

Previous research has indicated that salient segments are more likely to be weighted heavier compared to less salient segments (Drager & Kirtley 2016). For instance, using double modals "might could" may have a heavy weight, strongly indicating that the speaker is from the American South. While salience appears to have an important role in exemplar models, it is unknown how nonlinguistic and linguistic information weigh compared to each other. Is someone's perceived race more important than whether they use double modals? The question that remains, then, is how can researchers detangle this web in order to investigate specific linguistic features? Additionally, what can be done to target the weight of these features?

The three phonological contrasts in question are all considered to be salient to native English speakers (Kondaurova and Francis 2010). Additionally, PAM, PME, and SLM would all predict that nonnative speakers would have difficulty perceiving and producing the three. Therefore, it can be predicted that monolinguals and bilinguals will produce the three significantly differently. Furthermore, listeners should be able to identify the two groups of speakers by listening to audio alone.

The current study aims to address the following: how salience of the three phonetic features in question affects speech perception, how information stored in the exemplar can aid listeners in identifying different sociolinguistic groups, and what attitudes are stored in the listeners' exemplars towards the two different groups. In order to fully investigate this, it is important to first discuss the linguistic and non-linguistic information that affect speech perception according to an exemplar-based approach.

2.2.1 Non-linguistic factors

Perception has been the subject of many studies, with researchers investigating how listeners perceive speech from various languages and dialects. In this case, perception relates to whether listeners can use linguistic and non-linguistic features to create judgments on speakers, how they do this, and to what extent it occurs. Judgments that listeners make can include judgments about the speaker's first language, race, gender, socioeconomic background, and geographic background, among others. Studies ranging from the early 1990s to the current day have investigated the role non-linguistic features have had when studying speech perception, finding that certain features impact perception significantly (Rubin 1992; Hay et al. 2006; Staum-Casasanto 2008; Hay & Drager 2010).

When shown pictures of individuals along with audio, North American English speaking undergraduates rated Asian “speakers” as being less intelligible and less qualified to teach compared to White “speakers” (Rubin 1992). A picture of an Asian person was enough to elicit perceptions of a foreign accent, despite the speech being the same regardless of the picture shown. When the students believed that the instructor's race was the same as their own, they rated them more positively when it came to intelligibility, qualifications to teach, and teaching skills. The participants were chosen because they were native English speakers and had few encounters with nonnative accented speech in their lives. This could indicate that students without culturally and linguistically diverse backgrounds could be prone to having stereotypical judgments when encountering new situations. Additionally, perceived homophily between the listener and speaker has a significant effect on listeners’ judgments of the speaker.

Niedzielski (1999) found that listeners’ categorization of vowels was significantly impacted by the information given to them about the speaker, namely their nationality. In her study, listeners were either told that the speaker was from Canada or Michigan. Both groups of listeners heard the same speaker, who was from Detroit. Results found that the listeners categorized vowels differently based on the nationality they were told the speaker was. This indicates that listeners’ phonological spaces are not aligned solely using phonetic information; rather, stereotypes about speakers are included in these mental representations.

A later study indicated that listeners were more likely to hear more extreme realizations (i.e. higher, fronter) of /ɪ/ when ‘Australian’ was written on their answer sheet, and more likely to hear a centralized /ɪ/ if “New Zealander” was written on it (Hay et al. 2006). However, listeners all correctly identified the speaker to be a New Zealander. Multiple subsequent studies found that exposing participants to stuffed kangaroos, koalas, or kiwis affected whether the participants believed the speech came from an Australian or New Zealander speaker (Hay & Drager 2010, Walker et al. 2019, Hurring et al. 2022). This suggests that a large variety of information is used by listeners to process speech. These results can be explained by exemplar models; in these models, listeners store both linguistic and non-linguistic data that contributes to how they process and perceive speech.

2.2.2 Linguistic factors

As discussed in the previous section, exemplar models can explain well why non-linguistic information affects perception. While this is useful, it also illuminates the challenges that researchers have when studying speech perception. Researchers studying speech perception must take into account the effect non-linguistic factors have. The question still remains how much influence these non-linguistic factors have in comparison to linguistic factors. The current conclusion of researchers is that listeners *can* use linguistic factors to their advantage, especially if they are familiar with the variety or varieties involved (Campbell-Kibler 2007; McBride 2015; Newman & Wu 2011; Purnell et al. 1995; Schmidt 2022).

In Purnell et al.’s famous 1995 study, participants were tasked with calling landlords and asking for an appointment to tour apartments. Participants spoke using Standard American English (SAE), African American English (AAE), or Chicano English (CE). The landlords could not see the speakers and all speakers used the same script. Therefore, participants were forced to only use the speakers’ voice to make judgments. Despite this, participants speaking AAE or CE were significantly less likely to get an in-person appointment compared to their SAE speaking counterparts. The different treatment by the landlords of the participants indicated that the landlords could reliably discern between the three varieties without having extra-linguistic information on the speakers.

On a similar trend, results of a 2011 study investigating the speech of Asian Americans found both familiarity with the varieties in question and the salience of features in the varieties play a significant role in perception (Newman & Wu 2011). In this study, listeners of various racial and ethnic backgrounds were more accurate at judging whether a speaker was black or Hispanic than Asian or white. They had the most mixed reactions listening to Asian American speakers. Phonetic analyses demonstrated that while the Asian American speakers did have certain features in their speech that differed from the other groups, the speakers varied in their usage of said features. The authors conclude that the salience of the features played a large role in the listeners' judgments. The features in a so-called "Asian American" accent are perceived to be less salient compared to African American English and Chicano English. When analyzing responses to the three other groups, judgments were most accurate when the speaker's race matched their own. This contributes to the conclusion that listeners are most accurate when faced with speech with which they are familiar, however salience of the features in the varieties also plays a significant role in this familiarity.

Other studies have investigated perception across different languages, with the general conclusion being that more familiarity with the languages under investigation improves listeners' accuracy discriminating between varieties (McBride 2015; Schmidt 2022; Tsang 2020). In her 2022 study, Schmidt found that when second language learners of Spanish become more proficient, they can distinguish dialectal features in native speakers' speech accurately. She concludes that this is due to two main factors: awareness of different dialects and experience with native speakers. This indicates that beginning and intermediate learners typically cannot distinguish different dialects, and when they can, they are not conscious of what features are different between dialects. Additionally, learners across proficiency levels correctly identified both Peninsular Spanish speakers and identifying features more accurately than the other varieties. It can be proposed that features of Peninsular Spanish are more salient to L2 learners of Spanish, allowing them to identify speakers from that country. Tsang (2020) found that L2 learners of English were more accurate at judging nativeness when listening to English accents with which they had previous experience. Additionally, they were more likely to assign negative characteristics to speakers who had non-native accents. These results are in line with those of Newman & Wu (2011); both level of familiarity with the varieties in question and salience of features can improve the discrimination between them.

2.3 Language & Attitudes

As seen in the previously cited studies, speech perception contributes to attitudes held towards speakers. There is clear evidence that familiarity with linguistic varieties can aid listeners in correctly judging the language background of a speaker, but regardless of whether listeners are correct, they can still make positive, negative, or neutral judgments towards speakers. These judgments include thoughts on the speaker's education level, personality, income, appearance, among others.

Studies such as Rubin's and Purnell et al.'s indicate that when a listener believes the speaker is of a certain race or nationality, they may display attitudes related to common stereotypes about said race or nationality. This may be especially the case when a listener is unfamiliar with certain social groups compared to groups they have extensive familiarity. In the former case, listeners do not have enriched exemplar clouds towards people of those groups, so they may rely on postive or negative stereotypes. In comparison, more exposure to a group of individuals allows for a detailed exemplar cloud to form, which allows the speaker to make or hold more nuanced beliefs about that group. This is especially the case when the listener is unfamiliar with people of that race or nationality. A more recent study by Kang and Rubin (2009) found that native speakers hear what they expect based on speakers' perceived social identity. That is to say that listeners' speech perception can be highly influenced by their stereotypes of the social identities in question, rather than the acoustic information alone. Purnell et al.'s (1995) results show that listeners' negative attitudes may contribute to prejudiced behavior towards the speakers. These studies investigated perception and attitudes of native English speakers towards other English speakers; this begs the question, how are these attitudes different or similar in regard to second language speakers and bilinguals?

Vaughn (2019) found in a matched-guise style perception task that listeners' speech comprehension was affected by the guise they received. In the task, participants listened to an L2 speaker of English and were told one of the following about the speaker: 1) the speaker was an L2 speaker of English, 2) the speaker was an L1 speaker of English, or 3) they were told nothing about the speaker. Results of her first experiment showed that the guises may have helped the participants adapt to the speech, indicating that any information on the speaker (even false information) can help speech processing. A second experiment showed that participants were able to understand the speech better when they believed the speaker was an L1 speaker. Overall, Vaughn concludes that information *about* speakers is critical to speech processing, not just the speech itself. Not only that, but listeners' comprehension is affected by their beliefs about speakers.

In her 2015 study, McBride investigated the attitudes of Spanish natives towards Spanish learners' speech. McBride found that natives were able to accurately judge the level of proficiency of the learners and that those judgments aligned with comprehensibility judgments. More interestingly, comprehensibility judgments were positively correlated with sociolinguistic judgments. McBride reported that speakers who were rated as "easy to understand" were also rated as "attractive" and "educated." The opposite was true was well; speakers who were rated as "difficult to understand" were rated as "ugly" and "uneducated." This indicates that comprehensibility may be a predictor of attitudes towards speakers, particularly those who are L2 speakers.

In a similar vein, attitudes about nonnative accents from both native and nonnative English speakers revealed that nonnative accents are associated with a weaker feeling that one belongs in the US (Gluszek & Dovidio 2010). Additionally, problems in communication and stigma consciousness were associated with lower sense of belonging in the US on part of the nonnative speakers. Nonnative speakers were perceived by native speakers as having more problems in communication, leading to a perceived hardship on the interlocutor. Meanwhile, nonnative speakers reported having difficulty in communication, but not to the extent that nonnative speakers perceived them to have. Not all accents were treated equally, however. Asian and Latino accents were found to have a significant difference in perceived stigmatization from European accents, with native speakers assigning more positive attributes to European accents, one of which was intelligibility. Ultimately, perceived intelligibility and communication problems directly correlated with the sense of belonging a speaker has in the US. Both nonnative and native speakers believed those with more accented speech belonged less to the US, while those with less accented speech belonged to it more.

When investigating American English accents, Campbell-Kibler found that there is an association between using [ɪn] for <ing>, perceived Southernness, and lack of education (2007). Participants listened to speakers of urban and rural backgrounds, identified where they believed the speakers were from, and assigned attributes to the speakers. Although listeners were able to identify the geographical background of the Southern speakers, they were unable to identify the geographical background of the urban speakers. The urban speakers were considered to be "anywhere" speakers that could be from anywhere in the US. There was an association between "anywhere" speaker and high levels of education. Speakers whose accents were perceived to be closer to a central norm had overall positive attributes assigned to them. This could indicate that listeners can identify the range of features that exist for different varieties and assign "lack of accent" based on location to the central norm. Additionally, this indicates that standard varieties, or those that are perceived to be standard, are overall preferred by listeners. According to Campbell-Kibler's results, listeners associate standard varieties with positive characteristics, while non-standard varieties can carry negative stereotypes. Most importantly, whether the listeners were correct in their geographical identification of the speakers, their judgments towards them corresponded with the geographical identification they made.

It can be argued, then, that listeners' exemplar models not only allow them to process speech, but to also make subsequent judgments about speakers. Not only that, but sociolinguistic judgments are stored along with linguistic information and non-linguistic information about the speaker and the utterances they say. With an understanding of how many factors contribute to speech perception and listener attitudes, researchers can begin to piece together the exemplar model, in order to see how each factor contributes to perception and how they affect each other.

In total, listeners utilize their entire range of linguistic experience to produce and perceive speech. In order to narrow down what linguistic information is most impactful in speech perception, this study includes two experiments. Experiment 1 targets the three aforementioned differences between English and Spanish in a production task. Experiment 2 continues investigating these differences in a perception task where listeners do not have visual cues related to the speakers. These two experiments together investigate how salience and lack of visual information will impact listeners' judgments towards speech. This contributes to current understandings of exemplar models and what attitudes exist towards bilinguals and monolinguals in the United States.

CHAPTER 3

EXPERIMENT 1

3.1 Introduction

Considering the previous literature regarding Spanish and English phonologies, as well as L2 phonological acquisition, the first course of action is to investigate how Spanish-English bilinguals and English monolinguals produce the three variables in question. In this production task, speakers were recorded saying sentences that targeted each variable. The recordings were acoustically analyzed to determine which of the three was most salient, if any. Experiment 1 focuses on how the two speaker groups' production of these variables is similar or dissimilar. This experiment is the first stage to answering the ultimate question of how listeners perceive monolingual and bilingual speech.

3.2 Methods and Materials

Upon receiving approval from UGA's Institutional Review Board, eight total participants were recruited for the production task. All participants were recruited from UGA and were female. Participants filled out a language background survey to determine their eligibility in the study. Four of the participants were labeled under the "Monolingual English" group (M) and the other four under the "Bilingual English-Spanish" group (B). To be considered for the M group there were two requirements: 1) the speaker's first language must be English and 2) the speaker must have three or fewer semesters of foreign language courses at an undergraduate level. Three of the participants in this group had studied three semesters of Spanish, and one had studied three semesters of French. The goal of these requirements was to eliminate any participants with language experience that could affect their English pronunciation. The B group also had two requirements: 1) the speaker's first language must be Spanish and 2) the speaker must have acquired English concurrently or after acquiring Spanish. Out of the four recruited, one was introduced to English at 6 years, one at 9 years, and one at 13 years of age. The last participant was introduced to both English and Spanish from birth. All participants have resided in the United States for at least five years and completed the *Bilingual Language Profile* developed by Birdsong et al. (2012). Possible dominance scores of the BLP range from -280 (total dominance in Spanish) to 280 (total dominance in English). The four bilingual participants' dominance scores ranged from -64.56 to 45.05. Their scores were in line with age of introduction to English; the participant introduced to English from birth scored the highest (English dominant) and the participant introduced to English at 13 years of age scored the lowest (Spanish dominant).

The production task consisted of the participants reading aloud sentences prepared by the researcher and being audio recorded. All participants were recorded in a sound attenuated booth at the University of Georgia. All sentences were recorded with a sampling rate of 44kHz. The sentences focused on one of the three phonetic differences. In total, there were four types of sentences: 1) sentences testing /l/ formants, 2) sentences testing /i/ vs /ɪ/ distinction, 3) sentences testing /d t/ VOTs, and 4) distractor sentences. All recorded stimuli can be found in Appendix A. Examples of the sentences can be found below:

Table 3.1 - Example Experiment 1 stimuli

Set	Sentence	Target segment
1	Can you say "late"?	Late - [l]
1	Can you say "tale"?	Tale - [ɫ]
2	Can you say "pit"?	Pit - /ɪ/
2	Can you say "Pete"?	Pete - /i/
3	I like to text.	Text /t/
3	I like to date.	Date /d/
4	I have a car.	N/A

4	I want a vacation.	N/A

There were six sentences each in sets (1)-(3) and eight in set (4), for a total of 18 target sentences (26 total). All participants were asked to repeat the whole list twice, producing a total of 32 tokens per participant, and 242 tokens total. One participant had a total of 16 tokens, due to audio issues during the second recording. Text Grids were created which denoted where each targeted segment was in each recording. Tokens were analyzed by hand via Praat (Boersma and Weenink 2023). Formant values of the first two formants were collected for tokens testing the tense/lax distinction and lateral distinction. Formant measurements were taken at the approximate middle point of the lateral and vowel sounds (based on duration). Formant tracking was set for a maximum of 5500 Hz and to track 5 formants for all participants. Voice onset times were collected for tokens testing /t/ & /d/ distinction. VOTs were measured from the plosive burst to the onset of voicing. There were no negative VOTs found in the data at hand. Examples of tokens analyzed in Praat can be seen below:

Figure 3.1 – Spectrogram Image - Can you say "Pete"? (participant 7, monolingual) - highlighted section on [i] of *Pete.* F1 and F2 collected from this section.

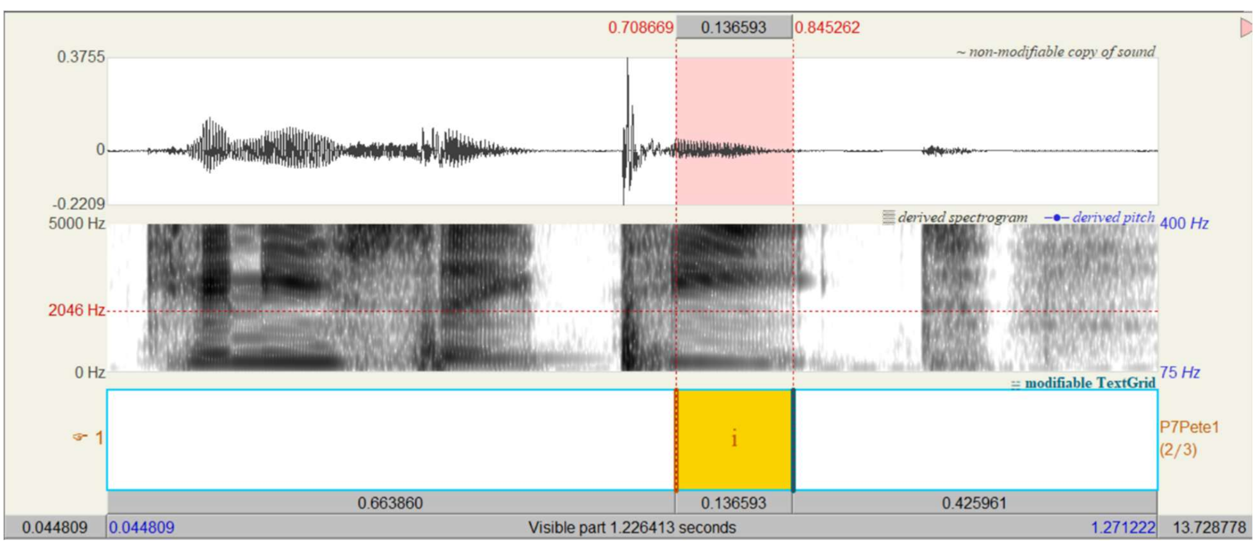


Figure 3.2 – Spectrogram Image - Can you say “late?” Can you say “late”? (Participant 4, monolingual) - highlighted section on [l] of *late*. F1 and F2 values collected from this section.

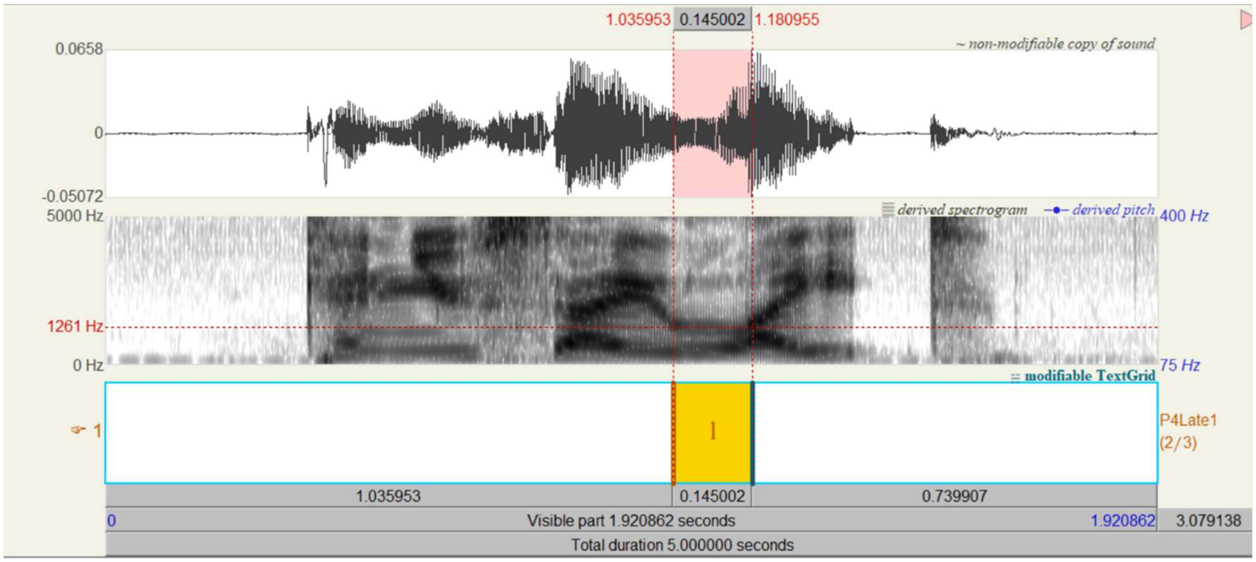
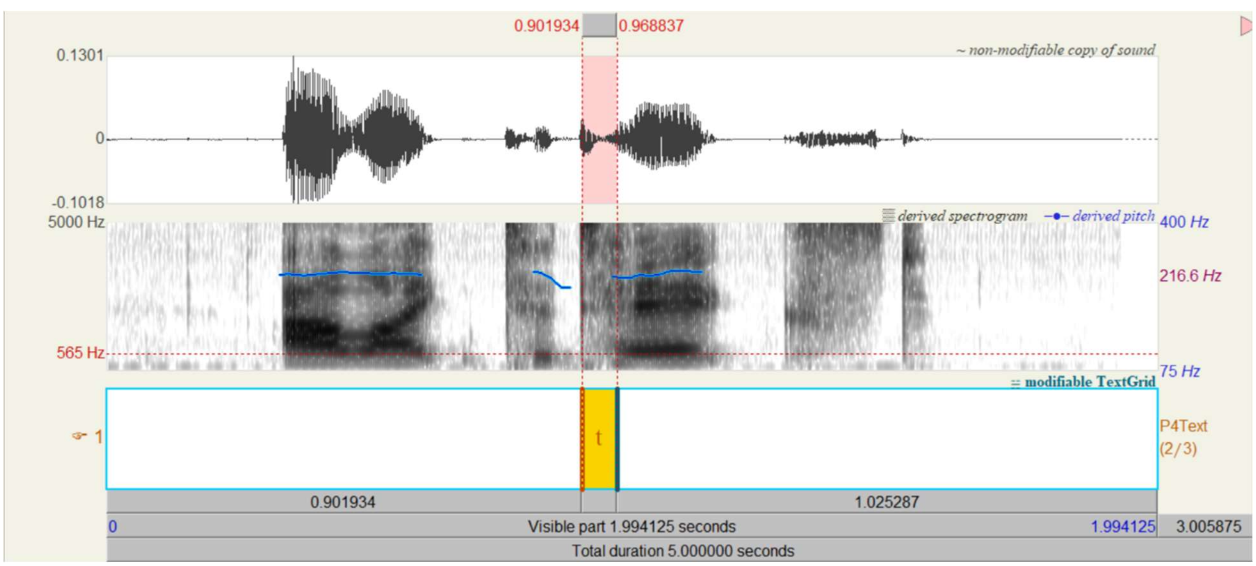


Figure 3.3 - Spectrogram Image - “I like to text” (participant 4, monolingual) - highlighted section on distance between plosive burst and voice onset of /t/ in *text*



3.3 Experiment Hypotheses

Hypotheses regarding the three phonological contrasts are proposed based on the previous literature on L2 phonological acquisition. They are as follows:

H1 – The F2-F1 of /ɪ/ will be larger when produced by bilinguals than by monolinguals.

H2 - The F2-F1 of /ɪ/ will be larger when produced by bilinguals than by monolinguals.
H3 -The voice onset times of /t/ will be longer when produced by monolinguals than when produced by bilinguals.

3.4 Descriptive Analysis

The goal of the acoustic analysis was to evaluate any differences between the production of the monolingual group and the bilingual group. In order to compare the two groups, visualizations of the data were created, and t-tests were conducted. The following section will discuss a descriptive analysis of the acoustic data.

3.4.1 Tense/Lax distinction

The first difference evaluated is the tense-lax vowel distinction that exists in English, but not Spanish. The first hypothesis predicts that the formant values of /ɪ/ will be significantly different between the two groups, as /ɪ/ does not exist as a phoneme in Spanish. Visualizations indicate that while /i/ formants are similar, /ɪ/ formants are not. This trend will be evaluated via t-tests in the following section. The following scatterplots indicate the F1 and F2 values of /ɪ/ and /i/ for both groups:

Figures 3.4-3.7 - /i/ and /ɪ/ formants

Fig. 3.4 Monolingual /i/ formants

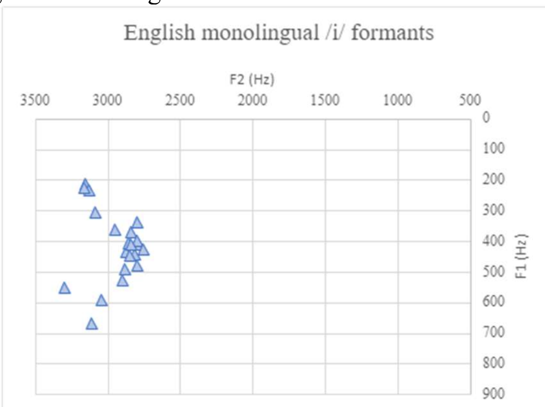


Fig. 3.5 Bilingual /i/ formants

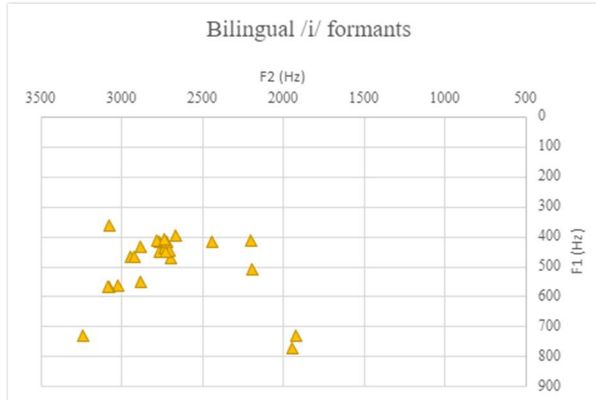


Fig. 3.6 Monolingual /ɪ/ formants

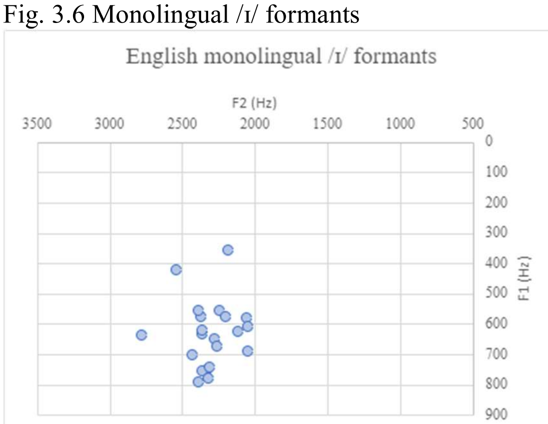
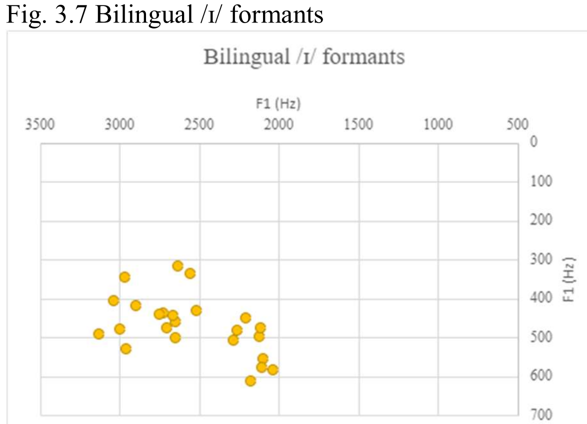


Fig. 3.7 Bilingual /ɪ/ formants



Although prosodic features were not the focus of the current experiment, duration of the vowels was measured in order to see if and how the speaker groups differed in this aspect. The duration was measured along with the following sound: 1) voiceless, 2) voiced, and 3) nasal. Past literature has indicated that /i/ will have a longer duration in English than /ɪ/ in all three environments (Kondaurova & Francis 2010). Descriptive visualizations of the duration of the two were created to find that the pattern of /i/ being longer than /ɪ/ was true between the two groups in all three environments.

Fig. 3.8 Bilingual Vowel Duration

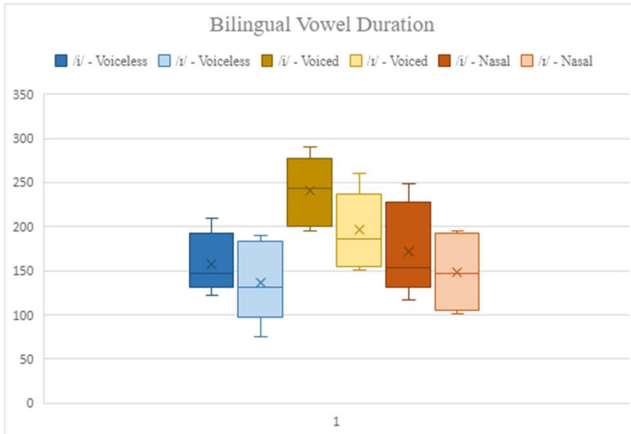
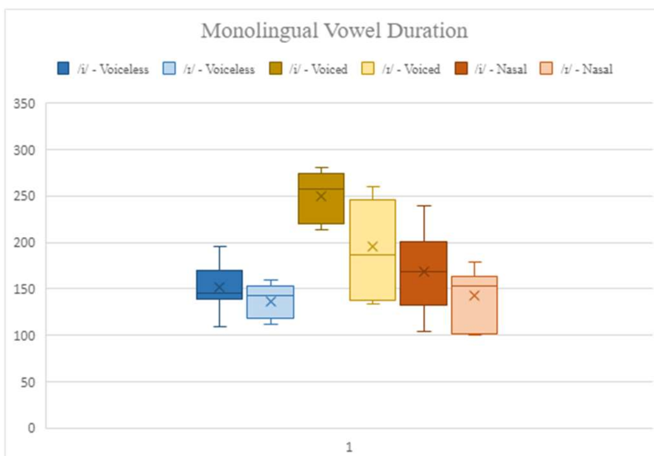


Fig. 3.9 Monolingual Vowel Duration



3.4.2 Lateral Distinction

The next difference evaluated was that of the distribution of /l/ between the two groups. The main difference expected is that the formant values for [ɫ] will be different between monolinguals and bilinguals. Unlike the previous difference, visualizations for this difference do not confirm or deny the second hypothesis. Therefore, it can be predicted that the differences between the realization of [ɫ] are smaller or not statistically significant at all.

Figures 3.10 and 3.11 - /l/ formants

Fig. 3.10 [l] formants

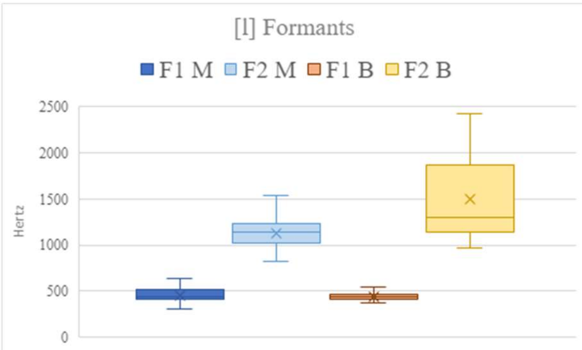
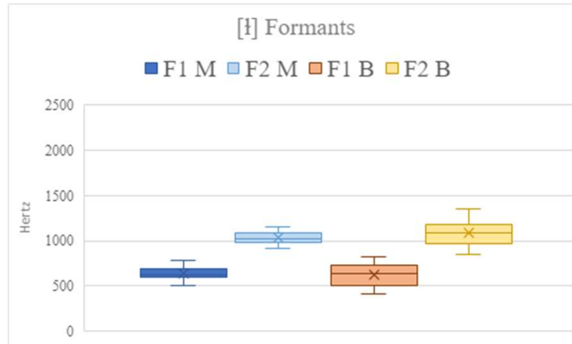


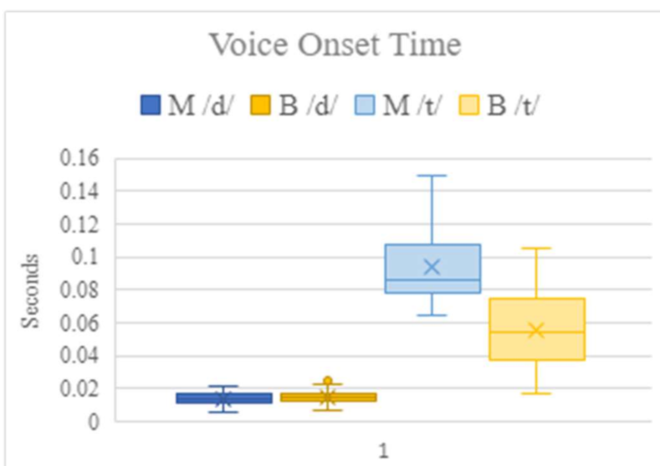
Fig. 3.11 [ɫ] formants



3.3.3 VOT Distinction

The last difference analyzed was the voice onset time of /t/ and /d/. The visualization of this difference is quite promising; /t/ appears to have very different VOTs between the two groups. Meanwhile, /d/ does not seem to differ between the two. This indicates that the third hypothesis could be supported, which will be further evaluated in the next section.

3.12 – Voice Onset Time /d/ and /t/



3.5 Quantitative Analysis

3.5.1 Tense/Lax distinction

As one can see in the above visualizations, the vowel space of both /i/ and /ɪ/ appear to be tighter for the monolingual group. In comparison, the bilingual group appears to have more variation in their realizations of the vowels, in particular with /ɪ/. This is in line with the study predictions; /i/ is part of the Spanish vowel inventory, so learners' acquisition of the vowel is more stable compared to /ɪ/, which is not part of the Spanish vowel inventory. In addition, the formant values of the monolingual and bilingual group appear to be more similar with /i/. T-tests comparing the formant values of the vowels between the two groups were conducted to test hypothesis #1. The results are as follows:

Table 3.2 - /i/ and /ɪ/ T-tests. Each row represents a comparison of bilingual vs monolingual speakers' formant values.

	t-value	**DF**	**p-value**	**Mean of B**	**Mean of M**
/ɪ/ F1 values	-5.791	34.401	1.545e-06***	471.70	636.01
/ɪ/ F2 values	3.0736	34.369	0.004124**	2543.86	2296.25
/ɪ/ F2-F1	4.5245	34.932	6.708e-05***	2072.15	1660.23
/i/ F1 values	2.7541	42.911	0.008597*	493.72	405.51
/i/ F2 values	-1.3239	41.147	0.1928	2712.60	2855.71
/i/ F2-F1	-1.9246	41.435	0.06116	2218.87	2450.20

As predicted, both F1 and F2 values of /ɪ/ were significantly different between the monolingual and bilingual group. In comparison, /i/ is only different in F1 values, while F2 values were not significantly different. Therefore, H1is supported.

H_1 – The F2-F1 of /ɪ/ will be larger when produced by bilinguals than by monolinguals. - **Supported**

3.4.2 Lateral Distinction

It can be seen in table 3.3 that though there is a significant difference in the F1 values of [ɫ], there is not a significant difference in the F2 values. More specifically, the F2 values of [l] were significantly different between the two groups, but the F2 values of [ɫ] are not significantly different between the two groups. It cannot be concluded that the realizations of one lateral is more (or less) salient than the other. Rather, the two speaker groups both differ in their realizations of both laterals, which was unexpected.

Table 3.3 - /l/ T-tests. Each row represents a comparison of bilingual vs monolingual speakers' formant values.

	t-value	**DF**	**p-value**	**Mean of B**	**Mean of M**
[l] F1 values	0.32316	41.00	0.7482	464.94	457.50
[l] F2 values	3.6087	30.14	0.0011**	1488.36	1133.91
[l] F2-F1	3.5315	29.67	0.0013**	1023.42	676.41
[ɫ] F1 values	-4.5432	39.90	5.031e-05***	476.29	619.88
[ɫ] F2 values	1.7718	31.69	0.08604	1187.26	1087.504
[ɫ] F2-F1	4.5251	35.065	6.649e-05***	710.96	475.9824

H_2 - The F2-F1 of /ɪ/ will be larger when produced by bilinguals than by monolinguals. - **Supported**

3.4.3 VOT Distinction

As predicted, VOTs for /t/ are indeed significantly different, while /d/ is not. This means that the alternative hypothesis is supported. The data from the current participants matches that of previous studies on these segments, such as Flege and Eefting (1995) and Wig and García-Sierra (2021), which both found English voiceless plosives to have longer VOTs than their Spanish counterparts.

Table 3.4 - /t/ and /d/ VOT T-tests. Each row represents a comparison of bilingual vs monolingual speakers' VOT values, in seconds.

	t-value	DF	p-value	Mean of B	Mean of M
/t/ VOT values	-5.4874	41.058	2.303e-06***	0.0553	0.0940
/d/ VOT values	0.79099	41.78	0.4334	0.0148	0.0139

> **H3** -The voice onset times of /t/ will be longer when produced by monolinguals than when produced by bilinguals. – **Supported**

3.6 Conclusion

Results of Experiment 1 indicate that all three phonological contrasts in question are realized acoustically differently between the bilingual and monolingual groups. T-tests regarding each feature support the three hypotheses. Although all three are supported, unexpected results appeared regarding the lateral sounds. Previous literature indicated that dark [ɫ] would be a difficult phone for the L2 English speakers to acquire, but not [l]. However, formants of both [ɫ] and [l] were different between the two speaker groups. Though this difference is surprising, Recasens (2012) can shed some light. In-line with their results, the mean F2 values of [l] are higher for the bilingual group than the monolingual group. This is in-line with their results, which found the average F2 values of prevocalic /l/ to be 1500-1800Hz in Spanish and 892-1272Hz in English (373). In both the current study and this previous study, the F2 values of prevocalic /l/ were produced lower by the English monolingual group. Additionally, Recasens found that postvocalic /l/ had a higher F1 value in English than in Spanish, which was also the case in the current study. However, it remains unclear why the participants in Experiment 1 did not differ in F2 values of [ɫ]. This reveals a more complex picture of L2 acquisition of English laterals. That being said, the experiment one results point to an answer to the first research question: "Which of the three phonetic differences is the most salient between monolingual and bilingual English speakers?" None of the three contrasts appears to be more or less salient than the other, at least acoustically. In order to continue investigating this question and the remaining two research questions, we will turn to Experiment 2.

CHAPTER 4

EXPERIMENT 2

4.1 Introduction

In order to address how listeners perceive monolingual and bilingual speech, a second experiment was motivated. In Experiment 2, monolingual listeners were tasked with listening to audio from Experiment 1 and making numerous judgments about the speakers in each recording. This task aims to answer questions of how well listeners can identify the two groups of speakers and what attitudes they have towards them. This will contribute to current understandings of how exemplar models can explain how listeners perceive speech of groups that have similar and dissimilar linguistic backgrounds.

4.2 Methods & Materials

Upon receiving approval from UGA's Institutional Review Board, a total of 175 participants were recruited via social media, word-of-mouth, and more to participate in the perception task. To be eligible for the task, participants had to be native English speakers and must not have had fluency in a language besides English. To screen eligibility, the following cutoffs were created: participants must not speak a language other than English at home nor must they have taken 4 or more semesters of a foreign language at a university level. These cutoffs were created to screen for participants who may be heritage speakers of another language or those who have extensively studied another language. Of the 175 participants, 89 completed the survey and met the language eligibility criteria. Figures 4.1-4.3 show breakdowns of participant age and language background:

Figures 4.1-4.3: Experiment 2 Participant Breakdown

Fig. 4.1 Foreign language studies

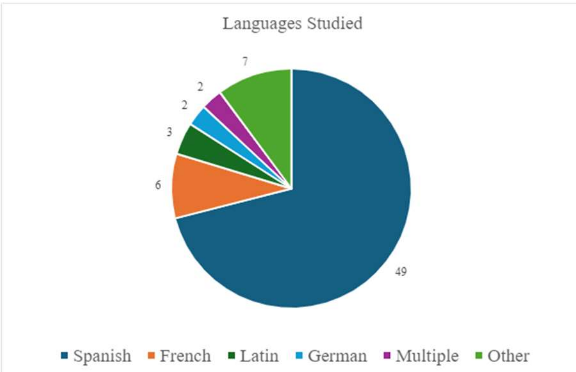


Fig. 4.3 Highest level of foreign language course taken

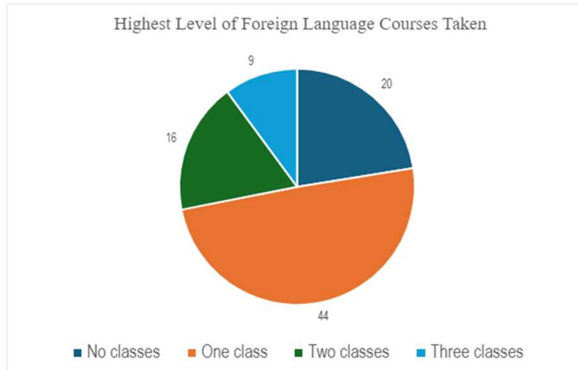
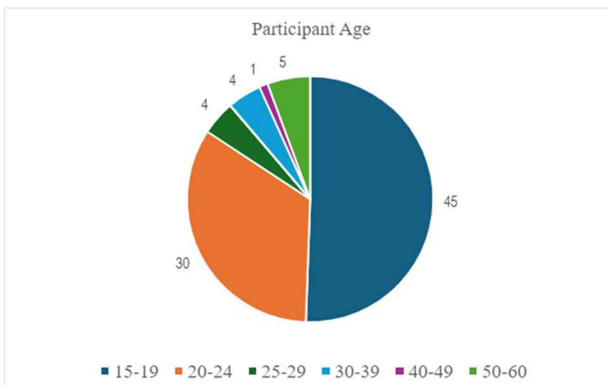


Fig. 4.3 Participant age



The perception task took place via Qualtrics. Participants were asked to fill out a screening and demographic questionnaire that recorded their language background. After the questionnaire, they were instructed to listen to 15 total recordings from the speakers who completed the production task. Of the 15 recordings, participants listened to 6 recordings produced by the bilingual group, 6 produced by the monolingual group, and 3 distractor sentences. Each recording targeted one of the phonetic distinctions in question and included the entire carrier sentence along with the targeted word. The participants were explicitly told at the beginning of the survey that the speakers they would be listening to both Native English speakers and Non-native Bilingual Spanish-English speakers. They were also told that they would have to differentiate between speakers of the two groups. This was decided due to pilot results that indicated that listeners had significant difficulty differentiating the groups when they were not given this information. As this study investigates attitudes towards Spanish-English bilinguals, it is key that participants know that is who they are listening to. After each recording, they responded to Likert-scale questions about the individual in the recording. The proxy question "Rate how native the speaker in the recording sounds, with 1 being Native English speaker and 5 being Non-native English speaker" was used to gauge the participants' judgments of a speaker being monolingual or bilingual. Additional questions asked the participants to make social judgments towards the speaker. These included judgments about their education, personality, and economic class. Lastly, an open-ended question was optional for each recording, where participants could list any features of the speaker that impacted their answers.

Pilot results indicated that prosody and inflection had a significant impact for some participants in their judgments. Therefore, the final survey controlled for the two by having two possible surveys that participants could receive: 1) survey with unedited audio and 2) survey with edited audio. The second survey included recordings for which the F0s were flattened, using the Praat Vocal Toolkit plug-in (Corretge 2023). Each F0 was flattened to the average F0 of their respective recording. By including monotonized recordings, half of the participant pool was not provided with intonation cues to make their judgments towards the speakers. In the current data, 44 participants received a survey with unedited audio and 45 received a survey with edited audio. Qualtrics randomly assigned each participant one of the two surveys.

4.3 Experiment Hypotheses

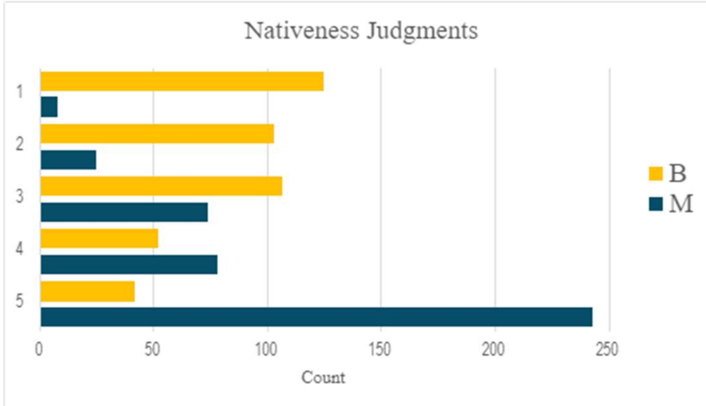
Based on results from Experiment 1, a pilot survey, and previous literature, four hypotheses are proposed regarding Experiment 2. They are as follows:

> **H_1** - Respondents' nativeness judgments of monolingual speech will be more accurate compared to their nativeness judgments of bilingual speech.
> **H_2 -** Respondents' nativeness judgments will be significantly impacted by the targeted segments [ɫ] /ɪ/ and /t/. This hypothesis is broken into three sub-hypotheses for each segment:
>
> > **H_{2A}** – Stimuli including [ɫ] will have a significant impact on nativeness ratings
> > **H_{2B}** – Stimuli including /ɪ/ will have a significant impact on nativeness ratings
> > **H_{2C}** – Stimuli including /t/ will have a significant impact on nativeness ratings
>
> **H_3** - Respondents will give significantly different social judgments towards those they judge as "Native" and "Non-Native".
> **H_4** – Respondents' nativeness judgments towards unedited audio and edited audio will be significantly different.

4.4 Descriptive Analysis

Descriptive analyses indicate that participants are more accurate in their nativeness judgments when reacting to monolingual speech than bilingual speech. While listeners overwhelmingly respond with "5 - Native English speaker" when listening to monolingual speech, they respond with a range of answers when listening to bilingual speech. The most common rating for monolinguals was "5 – Native English speaker," meanwhile the most common rating for bilinguals was "1 – Non-Native English speaker." Descriptive visualizations hint that H_1 may be supported. This can be seen in Figure 4.4:

Figure 4.4 Nativeness Judgments towards speech of monolingual and bilinguals



1 = Likely Non-Native English speaker, 5 = Likely Native English speaker

H_1 - Respondents' nativeness judgments towards monolingual speech will be more accurate compared to their nativeness judgments towards bilingual speech. - **Supported**

The second hypothesis predicts that the nativeness judgments will be more or less accurate based on the acoustic differences between the monolingual and bilingual groups of experiment 1. To review, t-tests from Experiment 1 found that F1 and F2 differences for /ɪ/ and [l] were both significantly different between the two groups. Additionally, the voice onset time of /t/ was significantly different. Therefore, H_2 predicts that these three segments will be more salient to respondents, increasing their accuracy in native judgments. Figures 4.5-4.7 show nativeness judgments separated by targeted segment (not separated by audio condition).

Figures 4.5-4.7 Nativeness judgments by feature

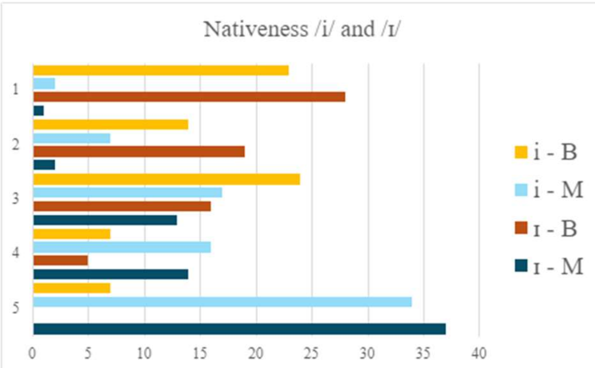


Fig. 4.5 Nativeness judgments towards stimuli targeting tense-lax vowels

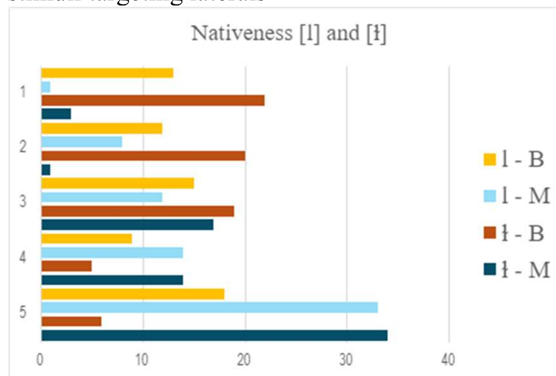
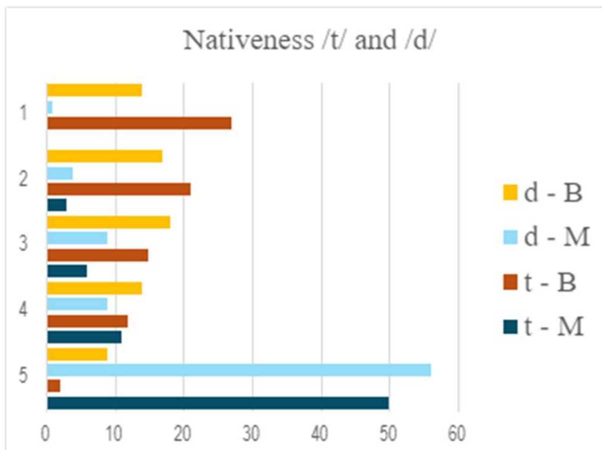


Fig. 4.6 Nativeness judgments towards stimuli targeting laterals

Fig. 4.7 Nativeness judgments towards stimuli targeting /t/ and /d/



H_2 - Respondents' nativeness judgments will be significantly impacted by the targeted segments [ɫ], /ɪ/, and /t/. This hypothesis is broken into three sub-hypotheses for each segment:

H_{2A} – Stimuli including [ɫ] will have a significant impact on nativeness ratings
H_{2B} – Stimuli including / ɪ/ will have a significant impact on nativeness ratings
H_{2C} – Stimuli including /t/ will have a significant impact on nativeness ratings

- Tentatively Supported

The figures indicate that there are different trends in the ratings of monolingual speech and bilingual speech. This could indicate that H_2 is supported.

In order to visualize results relating to H_3, social judgments were only considered when they corresponded to a Nativeness rating of 5 (native English speaker) or 1 (non-native English speaker). This is regardless of whether the speaker in the audio was monolingual or not; rather, this analysis aimed to discover the attitudes respondents had to speakers they *believed* to be native or non-native English speakers. As can be seen in the following visualizations, sociolinguistic ratings followed the same general trend for both groups. This indicates that H_3 cannot be supported.

Figures 4.8-4.10 – Social Judgment ratings

Fig. 4.8 Education ratings separated by bilingual and monolingual audio

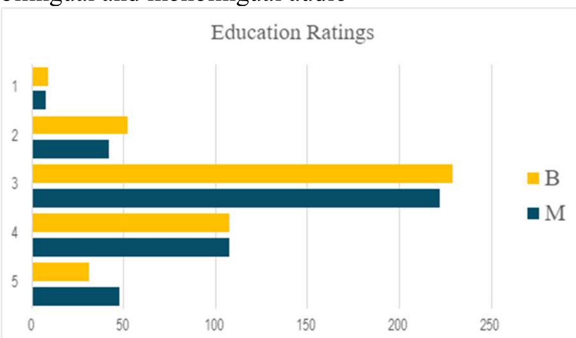


Fig. 4.9 Politeness ratings separated by bilingual and monolingual audio

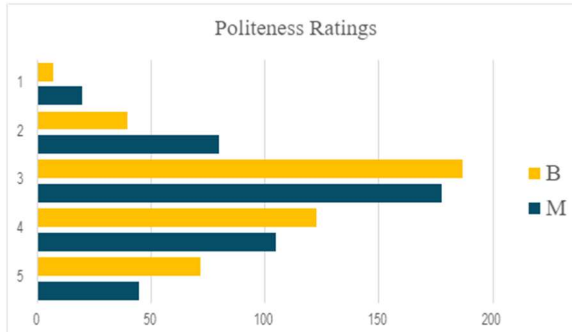
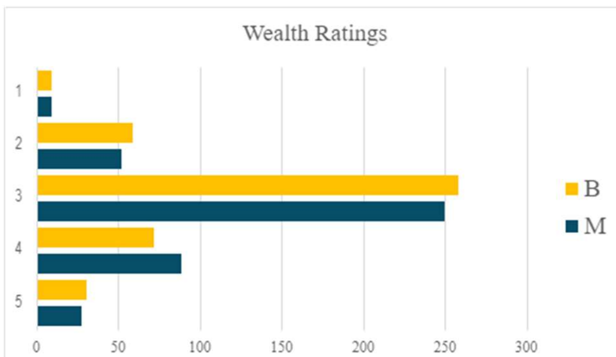


Fig. 10 Wealth ratings separated by bilingual and monolingual audio



4.8 Education, 1 = uneducated, 5 = highly educated

4.9 Politeness, 1 = rude, 5 = very polite

4.10 Wealth 1 = poor, 5 = very wealthy

H_3 - Respondents will give significantly different sociolinguistic ratings towards those they judge as “Native” and “Non-Native”. - **Not supported**

The fourth and final hypothesis predicts that F0 flattening will impact listeners’ ability to discriminate between monolingual and bilingual speech. The following visualizations show counts of the nativeness judgments for both speaker groups and survey types. Nativeness judgments do not appear to be different when listening to monolingual audio; the general trend being that participants assign a “5 – Native English Speaker” rating to the speaker the most. The nativeness judgments towards bilinguals appear to be more complicated. Participants are still less accurate at identifying bilingual speech compared to monolingual speech, however it is unclear what impact editing F0 had on these judgments. Nonetheless, it appears there is an interaction between the audio modification and the speaker in the recording. This will be further investigated with a linear ordinal regression. The fourth hypothesis is supported.

Figures 4.11 and 4.12 – Nativeness Judgments by audio condition

Fig. 4.11 – Nativeness judgments to unmodified audio

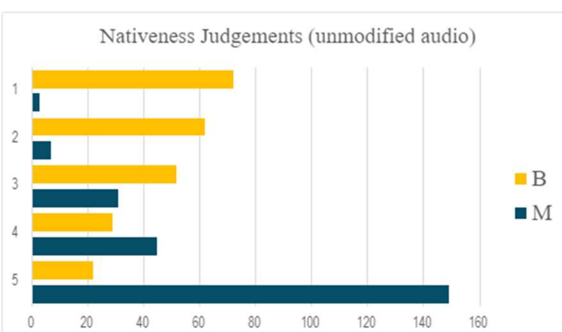
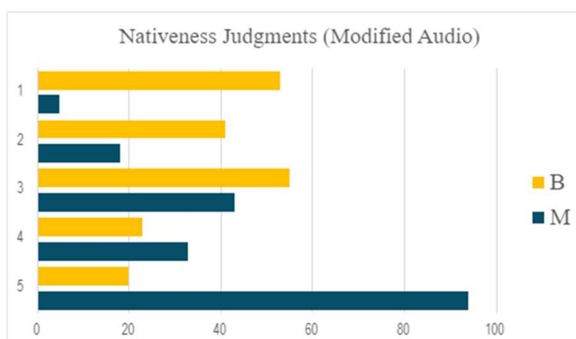


Fig. 4.12 Nativeness judgments to modified audio



H_4 – Respondents' nativeness judgments towards unedited audio and edited audio will be significantly different. - Supported

4.5 Quantitative Analysis

In order to analyze H1, the accuracy of nativeness judgments must be operationalized. To do so, the speaker in the recording will be compared to the nativeness ratings assigned to each recording. As previously mentioned, the nativeness ratings were made on a 5-point Likert scale, where 1 meant "non-native" and 5 mean "native." To measure accuracy, this scale was collapsed to three categories: *Correct, Incorrec*t, and *Intermediary*. All ratings of 3 were considered *Intermediary* since that rating was the middle point of the scale and therefore, it is impossible to these ratings as "correct" or "incorrect". Ratings towards monolingual recordings were coded as *Correct* if the participant rated the speaker 4 or 5 and *Incorrect* if they rated the speaker 1 or 2. Ratings towards the bilingual recordings were coded as *Correct* if the participated rated the speaker 1 or 2 and *Incorrect* if they rated the speaker 4 or 5. The following figure shows the accuracy rates when reacting to both monolingual and bilingual recordings:

Figure 4.13 – Accuracy of Nativeness Judgments

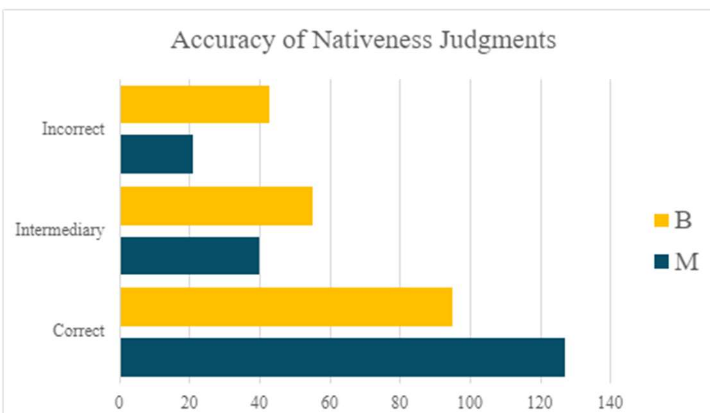


Table 4.1 - Rate of accuracy in nativeness judgments

	Bilingual Recording	Monolingual Recording
Incorrect	22.3 %	11.2%
Intermediary	28.4%	21.2%
Correct	49.2%	67.6%

As can be seen in the figure and table, participants were much more accurate when reacting to monolingual speech than bilingual speech. Especially interesting is that the *Correct* percentage when reacting to bilingual recordings is 49.2%, slightly less than chance. Therefore, H_1 can be supported.

H_1 - Respondents' nativeness judgments towards monolingual speech will be more accurate compared to their nativeness judgments towards bilingual speech. – Supported

In order to concretely analyze the next hypotheses, multiple statistical tests and models must be used. As experiment 2 utilizes a Likert scale, it must be accounted for that 1-5 is not equidistant for each participant. Therefore, the main model that will be used to address the hypotheses is a linear ordinal regression, which can evaluate the distance between each rating for each participant (Veríssimo 2021).The following outlines the proposed statistical analysis for each hypothesis:

First, we will create a linear ordinal regression with Native Judgments as the dependent variable using the clm() function from the *ordinal* package in RStudio (Christensen 2019). Independent variables included: speaker (bilingual/monolingual), modified audio (yes/no), target segment (/i/ /ɪ/ [l] [ɫ] /t/ /d/). Additionally, the interaction between speaker and audio condition was included as an independent variable.

Figure 4.14 - Code in Rstudio package *ordinal* – Nativeness Judgments regression

```
Cumulative Link Mixed Model fitted with the Laplace approximation

formula: Nativelike ~ MB * Modified + MB * Target + (1 | Speaker) + (1 |      Participant)
data:    task2

 link  threshold nobs logLik   AIC     niter      max.grad cond.H
 logit flexible  856  -1079.76 2197.53 2270(8901) 6.98e-04 6.0e+02

Random effects:
 Groups      Name        Variance Std.Dev.
 Participant (Intercept) 0.6513   0.8070
 Speaker     (Intercept) 0.8355   0.9141
```

Table 4.2 – Nativeness Judgments Linear Ordinal Regression

	Estimate	**Std. Error**	**Z value**	**Pr (>|z|)**
Speaker - Mono	3.8735	0.7720	5.017	75.24e-07 ***
Modified - Yes	0.1826	0.2555	0.715	0.4748
Target - /ɹ/	0.2777	0.3711	0.747	0.4549
Target - [l]	1.6608	0.3665	4.532	5.85e-06 ***
Target - [ɫ]	1.5716	0.4012	3.928	8.58e-06***
Target - /t/	1.6533	0.3740	4.420	9.86e-06***
Target - /d/	0.7665	0.3634	2.081	0.0374*
Speaker(Mono): Modified(Yes)	-1.0785	0.2715	-3.971	87.15e-05 ***
Speaker(Mono): /ɹ/	0.2671	0.4936	0.541	0.5884
Speaker(Mono): [l]	-1.462	0.6018	-2.430	0.0151*
Speaker(Mono): [ɫ]	-1.341	0.5693	-2.356	0.0185*
Speaker(Mono): /t/	-0.0524	0.5338	-0.098	0.9218
Speaker(Mono): /d/	0.3867	0.5640	0.686	0.4930

Table 4.3 – Nativeness Judgments Linear Ordinal Regression. Estimates of threshold values for ratings "1 – nonnative English Speaker" to "5 – Native English Speaker"

Threshold	**Estimate**	**Std Error**	**Z value**
1|2	-0.1405	0.5449	-0.258
2|3	1.1892	0.5472	2.173
3|4	2.7613	0.5583	4.946
4|5	3.852	0.5667	6.785

The model indicates that the monolingual-bilingual status of the speaker had the most significant effect on nativeness ratings, with respondents being likely to respond 3.8735 points higher on the nativeness rating when responding to a monolingual compared to a bilingual. Additionally, segments targeting [ɫ] and /t/ were likely to produce a higher nativeness rating, although this variable is less significant than speaker status. When introducing an interaction effect between target segment and speaker in the audio, monolingual audio targeting [ɫ] was found to receive a rating 1.5716 points lower than bilingual audio with the same target. This was unexpected, as this segment was predicted to cue that the speaker was bilingual, not monolingual. Finally, the model finds that there is a significant interaction between the monolingual-bilingual speaker status and the audio condition. Participants are likely to rate the nativeness 1.0785 points lower when the speaker is monolingual, and the audio is edited.

H_2 - Respondents' nativeness judgments will be significantly impacted by the targeted segments [ɫ] /ɪ/ and /t/.

> **H_{2A}** – Stimuli including [ɫ] will have a significant impact on nativeness ratings - **Supported**
>
> **H_{2B}** – Stimuli including /ɪ/ will have a significant impact on nativeness ratings - **Rejected**
>
> **H_{2C}** – Stimuli including /t/ will have a significant impact on nativeness ratings - **Supported**

H_4 – Respondents' nativeness judgments towards unedited audio and edited audio will be significantly different. - **Supported**

Unlike the preliminary visualizations that showed extreme responses (1 & 5), the linear ordinal regressions modelling social judgments included all responses. In total, there are three models, where wealth, education, and politeness ratings are the dependent variables. Nativeness judgments were the independent variable in all three. The default level for all three is a rating of 1. The results of these regressions can be seen in the tables below:

Figure 4.15 - Code in Rstudio package *ordinal* – Wealth Judgments regression

```
Cumulative Link Mixed Model fitted with the Laplace approximation

formula: Wealthy ~ Nativelike + Modified * MB + Target + (1 | Speaker) +      (1 | Participant)
data:    task2

 link  threshold nobs logLik  AIC     niter       max.grad cond.H
 logit flexible  855  -874.10 1784.19 2308(20607) 3.28e-04 9.8e+01

Random effects:
 Groups      Name        Variance Std.Dev.
 Participant (Intercept) 2.0376   1.4274
 Speaker     (Intercept) 0.1481   0.3849
```

Table 4.4 – Wealth Ratings linear ordinal regression

	Estimate	**Std. Error**	**Z value**	**Pr (>|z|)**
Nativelike2	-0.0094	0.2848	0.033	0.9737
Nativelike3	-0.1047	0.2941	-0.356	0.7217
Nativelike4	0.1514	0.3132	-0.356	0.7217
Nativelike5	0.7795	0.3064	2.543	0.0110*
Modified- Yes	-0.4207	0.3749	-1.122	0.2618
Speaker - Mono	-0.3870	0.3614	-1.071	0.2843
Target - /ɪ/	0.1636	0.2554	0.641	0.5218
Target - [l]	-0.0155	0.2858	-0.54	0.9566
Target - [ɫ]	-0.6518	0.2840	-2.295	0.0217 *
Target - /t/	0.5067	0.2719	1.863	0.0625
Target - /d/	0.5648	0.2612	2.162	0.0306 *
Speaker(Mono):Modified(Yes)	0.1937	0.288020	0.673	0.5011

Table 4.5 – Wealth Ratings Linear Ordinal Regression. Estimates of threshold values for ratings "1 – Poor to "5 – Very wealthy"

Threshold	**Estimate**	**Std Error**	**Z value**
1|2	-4.8173	0.4870	-9.891
2|3	-2.2427	0.4131	-5.429
3|4	1.15371	0.4062	3.784
4|5	3.6234	0.4314	8.399

Figure 4.16 - Code in Rstudio package *ordinal* – Education Judgments regression

```
Cumulative Link Mixed Model fitted with the Laplace approximation

formula: Educated ~ Nativelike + Modified * MB + Target + (1 | Speaker) +      (1 | Participant)
data:    task2

 link  threshold nobs logLik  AIC     niter      max.grad cond.H
 logit flexible  855  -918.28 1872.56 1882(9396) 2.01e-04 1.4e+02

Random effects:
 Groups      Name        Variance Std.Dev.
 Participant (Intercept) 2.1073   1.4517
 Speaker     (Intercept) 0.4309   0.6564
```

Table 4.6 – Education Ratings linear ordinal regression

	Estimate	**Std. Error**	**Z value**	**Pr (>|z|)**
Nativelike2	0.4661	0.2764	1.686	0.0917
Nativelike3	0.3181	0.2839	1.120	0.2625
Nativelike4	0.2737	0.3087	0.887	0.3753
Nativelike5	0.8954	0.3020	2.965	0.00303**
Modified- Yes	-0.4331	0.3774	-1.147	0.2511
Speaker - Mono	-0.5026	0.5203	-0.966	0.3340
Target - /ɪ/	-0.1719	0.2540	-0.677	0.4985
Target - [l]	0.2530	0.3004	0.842	0.3996
Target - [ɫ]	-0.5110	0.2822	-1.811	0.0701
Target - /t/	0.3551	0.2688	1.321	0.1865
Target - /d/	1.1573	0.2664	4.345	1.04e-05***
Speaker(Mono):Modified(Yes)	0.3893	0.2820	1.380	0.1674

Table 4.7 – Education Ratings Linear Ordinal Regression. Estimates of threshold values for ratings "1 – uneducated" to "5 – highly educated"

Threshold	**Estimate**	**Std Error**	**Z value**
1|2	-4.5047	0.5474	-8.229
2|3	-2.2137	0.4909	-4.510
3|4	1.2983	0.4837	2.684
4|5	3.6736	0.5054	7.269

Figure 4.17 - Code in Rstudio package *ordinal* – Politeness Judgments regression

```
Cumulative Link Mixed Model fitted with the Laplace approximation

formula: Polite ~ Nativelike + Modified * MB + Target + (1 | Speaker) +      (1 | Participant)
data:    task2

 link  threshold nobs logLik   AIC     niter       max.grad cond.H
 logit flexible  856  -1057.91 2151.81 2106(10592) 4.10e-04 1.6e+02

Random effects:
 Groups      Name        Variance Std.Dev.
 Participant (Intercept) 2.2578   1.503
 Speaker     (Intercept) 0.2591   0.509
```

Table 4.8 – Politeness Ratings linear ordinal regression

	Estimate	Std. Error	Z value	Pr (>|z|)
Nativelike2	-0.3537	0.2635	-1.342	0.1795
Nativelike3	-0.3064	0.2720	-1.127	0.0299*
Nativelike4	-0.3665	0.2965	-1.236	0.2164
Nativelike5	-0.37622	0.2900	-1.297	0.1946
Modified- Yes	-0.1096	0.3803	-0.288	0.7732
Speaker - Mono	-1.0255	0.4239	02.419	0.0156*
Target - /ɪ/	0.1475	0.2421	0.061	0.9514
Target - [l]	0.2994	0.3088	0.970	0.3322
Target - [ɫ]	-0.0191	0.2811	-0.068	0.9456
Target - /t/	0.0148	0.2527	0.059	0.9531
Target - /d/	0.4831	0.2527	1.905	0.0568
Speaker(Mono):Modified(Yes)	0.67837	0.26988	2.514	0.0119*

Table 4.9 – Politeness Ratings Linear Ordinal Regression. Estimates of threshold values for ratings “1 – Rude” to “5 – Very polite”

Threshold	Estimate	Std Error	Z value
1|2	-4.6965	0.4818	-9.748
2|3	-2.5894	0.4424	-5.852
3|4	-0.0050	0.4316	-0.012
4|5	2.0892	0.4417	4.730

The first two models (wealth and education) predict that a nativeness rating of 5 has a significant positive effect on ratings of perceived speaker wealth and education. However, the last model predicts that a nativeness rating of 3 has a significant negative effect on politeness ratings. As a rating of 3 was the middle point between “monolingual” and “bilingual,” this could indicate that listeners judge speakers as less polite when they cannot determine each speaker’s respective linguistic background. Additionally, the speaker in the audio was a significant predictor for politeness ratings, with monolingual speakers generally receiving a rating that was 1.0255 points lower than bilingual speakers. Previous literature has found that social judgments are typically higher when participants believe the speaker is native compared to when they believe the speaker is nonnative (McBride 2015, Tsang 2020). The results of these three models indicate a more complicated picture: although speakers rated as “native English speakers” were positively rated on their wealth and education, that is not the case when it comes to politeness. This would indicate that future studies should include a variety of characteristics and traits for participants to judge. Furthermore, there should be research into what types of traits are associated with native and non-native speakers, rather than a general negative-positive trait binary that has been previously investigated. As nativeness ratings of 5 had a significant impact in all the social judgments, the third hypothesis is supported.

> **H_3** - Respondents will give significantly different sociolinguistic ratings towards those they judge as “Native” and “Non-Native”. - **Supported**

4.6 Qualitative Analysis

Along with the Likert scale questions that participants received, participants had the option to respond to an open-ended question for each recording they heard. There was a total of 48 responses to the open-ended question "Were there any noticeable features about the recording that influenced your answers? (Optional)". Of these 48, multiple comment themes appeared. Comments were categorized as related to the following: 1) accent, 2) (meta)linguistic, 3) robotic, 4) attitude, 5) intelligibility. Any comments unrelated to those five were put into the 6) Other category. The most common type of comment was "accent," where participants predicted where the speaker was from or what type of accent they had. The second most common type of comment (apart from "other") was "metalinguistic" where participants reported something about the phonetics of each speaker. The most common type of response noted that syllable stress and/or lack of vowel reduction contributed to participants' judgments.

Figure 4.18 – Comment Themes

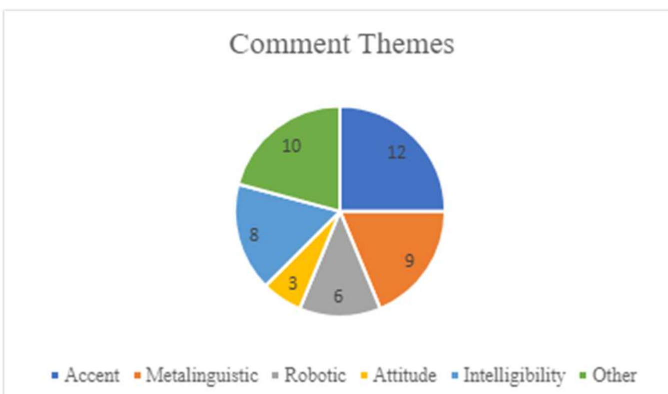


Table 4.10 – Open-ended question comment themes

Comment example	Comment type	Survey type	Target segment and speaker in the recording
"She sounded like she was from California I think"	Accent	Unedited audio	Bilingual - /d/
"ending of the word like and beginning of teach ran together"	(Meta)linguistic	F0 modified	Monolingual - /t/
"Sounded like a robot"	Robotic	F0 modified	Monolingual - /t/
"I refer to people who speak English under a foreign accent to be highly educated as they most likely can speak multiple languages."	Attitude	F0 modified	Bilingual - /ɪ/
"I don't understand what is being said"	Intelligibility	F0 modified	Bilingual - [l]
"The speaker was very soft spoken"	Other	Unedited audio	Bilingual - Filler

CHAPTER 5

DISCUSSION AND CONCLUSIONS

5.1 Summary

The results of Experiment 1 indicate that all three phonological contrasts caused difficulty for the bilingual group. That is to say, that all three were acoustically significantly different between the monolingual and bilingual groups. This finding is in line with previous studies on the three (Flege & Eefting 1995, Barlow et al. 2013, Fox et al. 1995). The bilinguals' production also supported predictions from the Speech Learning Model, the Perceptual Magnet Effect, and the Perceptual Assimilation Model (Flege 1995, Kuhl & Iverson 1995, Best 1995). As all three propose that exposure to the L2, differences in L1 and L2 allophonic distribution, and salience of contrasts will impact L2 acquisition, it was hypothesized that the bilingual groups would produce the three contrasts differently than their monolingual counterparts. The results of Experiment 1's acoustic analysis found this to be true. From the current data, there is no concrete evidence that one contrast was significantly more salient than another. However, stimuli targeting laterals revealed an interesting picture: although it was hypothesized that [ɫ] would be different between the two groups (instead of [l]), both were significantly different.

Results of both experiments indicated that though the /i/, /t/, and [ɫ] were acoustically different between the two speaker groups, only /t/ had an impact on nativeness judgments. This is a surprising result, as previous literature found that salience is a significant factor in speech perception, therefore we would expect all to have a significant impact on these ratings (Drager & Kirtley 2016). The following explanations can be proposed as to why this study's results differed from previous studies: feature salience is not as important as originally thought, feature salience works in tandem with visual information, or other cues that were not controlled for impacted nativeness judgments. Of the three, I believe the first explanation to be least likely, considering the mountain of evidence in previous literature that shows salience is impactful. However, it is possible that the other two explanations hold some weight. From an exemplar-based approach, visual information about speakers is connected with all other information about the utterance. It is possible that, by taking away all visual information about the speakers, the listeners had a difficult time interpreting the speech they heard. The last explanation is the most plausible; given the amount of data in this study and the limitations (which will be discussed shortly) it is likely that speakers relied on other features that were not originally controlled for. In order to postulate why this surprising result occurred, it would be wise to re-create the current study with more participants and other phonological contrasts that exist between English and Spanish. It is possible that including more non-laboratory speech with multiple acoustic cues, such as in Purnell et al's (1999) study, would positively impact participants' ability to discern the speech of different groups without visual cues.

Experiment 2 also revealed that monolingual listeners rated monolingual recordings with more extreme values in regard to nativeness. That is to say, that participants overwhelmingly rated monolingual recordings as a "5 - native English speaker." In comparison, bilingual recordings elicited a variety of nativeness ratings, despite the fact that within the Experiment 1 bilingual group, acoustic measurements were not significantly different. This could indicate that monolingual speakers are more confident rating other monolinguals as native English speakers. Additionally, the reactions to each speaker group could indicate that monolingual listeners have more detailed exemplar models of monolingual speech. This makes sense, as the task 2 participants reported coming from monolingual households. Therefore, these listeners may have more monolingual speech input than bilingual, causing their exemplars of monolingual speech to be more concrete.

The next finding of interest from Experiment 2 is that F0 modification appears to have an effect on nativeness ratings, when interacting with the speaker in the recording. It is initially unclear why participants rated modified audio lower in nativeness, but turning to qualitative results can paint a complete picture. When a participant received the survey with modified audio, the most common response to the open-ended question was that the speaker sounded like a robot or like AI. One participant remarked that "The (very obviously synthesized) speaker sounds like a pre-teen white boy" indicating that the participant was not sure if the speech was natural human speech at all. Another wrote "This one sounded more like a real person and not robotic, so it was clear that this was an actual person" when responding to a filler recording. Both reactions are enlightening, as they indicate that at least some participants with modified audio did not know if the speakers were indeed people or machine generated voices. This could contribute to why they rated nativeness as low, as listeners likely associate native speaker status with being a human. Other responses to the open-ended question revealed that participants found it more difficult to make judgments about the speaker when they believed the audio to be from a device, rather than a human. However, a recent 2023 study reported that listeners display perceptual learning when exposed to computer-generated speech, indicating that listeners adapt and generalize speech from humans and devices in the same way (Zellou et al. 2023). It would be beneficial for future studies investigating human and device guises to explore the impact of nativeness on listeners' speech perception.

The last significant finding was that listeners were more likely to rate the speakers in the recordings as wealthier and more educated if they also rated them "5 - native English speaker." This indicates that speakers store social information about speakers in their mental phonological representations. Specifically, there is no evidence that listeners connect negative traits with bilingual speech; rather they connect positive traits with monolingual speech. Additionally, listeners are more likely to rate a speaker lower in politeness if they rate them a 3 (the middle rating) on nativeness or if the speaker in the recording is monolingual. As a rating of 3 is intermediary, it is possible that listeners assigned speakers with that rating when they simply did not know the linguistic background of the speakers. When listeners have less connection to the speaker in the recording (in this case, because they cannot identify their background) they may be likely to rate them lower on certain characteristics. Additionally, they may view monolingual speakers as "prestige" speakers, and view this particular group as being less polite compared to speakers of lower prestige. The fact that the social ratings are impacted by different nativeness ratings is surprising, as previous relevant literature has had a binary positive-negative view of listener attitudes (McBride 2015; Tsang 2020). The results of Experiment 2 indicate that listeners do not harbor solely positive or negative judgments towards native or non-native speaker status; rather, listeners associate different positive characteristics with the two.

5.2 Limitations

There were a number of limitations in the design of experiment one. The first of which is no surprise; the speech recorded was not naturalistic or conversational speech. In order to target certain features, prepared stimuli were necessary. However, it is important to note that the speech is far from how the participants would speak normally. Therefore, it is possible that results of this study are not entirely indicative of how people produce and perceive speech in their daily lives. Additionally, of the four bilingual participants, two spoke Mexican varieties of Spanish and other two spoke Peninsular varieties of Spanish. The two Spanish speakers were from Catalunya, and although they reported not having proficiency in Catalan, it is possible that exposure to it affected their speech. This is especially noteworthy, as stimuli targeting laterals yielded unexpected results. Unlike Spanish, Catalan has both [l] and [ɫ] (Recasens 2012). Therefore, it is possible that those speakers produce [ɫ] more similarly to an English native speaker that the average Spanish-English bilingual.

Experiment two included its own set of limitations, the first of which being that participants had to self report proficiency in any language besides English. Therefore, it is possible (though unexpected) that participants underestimate their own bi- or multi-lingualism, which could affect participant screening. This is especially the case for speakers who speak a heritage language, but do not consider themselves "bilingual". To address this, the survey did ask if they speak other languages at home or with friends, but it is possible that participants did not accurately report their language background. Additionally, participants were not asked about any travel or study experiences abroad. Although most university study abroad requirements involve foreign language classes beyond the cut-off of this study (3 classes or less), it is still possible that participants have spent time abroad. This could influence their responses, as attitudes on culture and society can be affected by experiences in other countries (Schmidt 2020). Next, the design of the survey assumes that participants have access to fair-good quality listening devices. It is possible that participants listened to the recordings using listening devices of poor quality, which could possibly negatively affect the quality of the recording.

Lastly, while the audio stimuli controlled for certain aspects (i.e. targeting certain feature, flattening F0, etc.), other features were not controlled for or were unable to be measured with the available equipment. For example, all stimuli included multisyllabic words, with a stressed and unstressed syllable. Additionally, the entire carrier sentence was included in the recordings in Experiment 2. Therefore, participants may have used other features, such as prosody, to make their conclusions about the speakers. Previous studies, such as Magen (1998) found that listeners may be more sensitive to suprasegmental features when evaluating accentedness and nativeness. In line with Magen's results, the current study's qualitative results indicated that syllable stress was an influential factor for some participants. Finally, Spanish and English do have additional differences regarding /t/ and /d/, specifically place of articulation. While these sounds are alveolar in English, they are produced dentally in Spanish. Due to limitations with available equipment, the articulatory differences of these phones were not investigated. However, it would be recommended for future studies to include analysis of this articulatory difference. IN all, it can be questioned what exact features the listeners used to make their judgments, and if these judgments are holistic. Future studies are encouraged to investigate how sensitive monolinguals are at evaluating accent based on segmental features versus suprasegmental features.

5.3 Conclusion and future studies

The current study focused on the speech perception and judgments of monolinguals towards two groups of speakers. Results of the two experiments indicate that bilingual Spanish-English speakers do pronounce certain phones differently than English monolinguals, despite high proficiency in the language and living for more than five years in areas where English is predominantly spoken. The second experiment indicates that nativeness judgments are much more accurate when listeners hear a salient acoustic difference compared to those that are less salient. However, listeners are much less accurate at identifying bilingual speakers compared to identifying monolingual speakers overall. This contributes to the existing knowledge on connectionist exemplar models; listeners are more accurate at judging speech they have more experience - and therefore, exemplars - of (Kruschke 1992). Without visual stimuli, listeners must rely on an exemplar cloud that is impoverished, which may cause problems in accurately identifying speakers. The third hypothesis of Experiment 2 predicts that listeners would assign negative attributes to speech they perceived as bilingual. Descriptive visualizations did not support this hypothesis, as the listeners gave bilinguals and monolinguals the same ratings for education, politeness, and wealth judgments. However, ordinal linear regressions indicated that nativeness judgments do play a significant role in these social judgments. Not only that, but audio condition and targeted segment interact with the actual native speaker status of the speaker in the recording. This indicates that listeners are tapping into different exemplar clouds for these two groups of speakers. When they are primed that they will hear native English speakers and bilingual Spanish-English speakers, they already have the information about the two groups at hand. If the listeners were not utilizing different exemplar clouds, we would expect them to answer very similarly to both groups of speakers. However, the fact that there are interactions between the independent variables in all three sociolinguistic ordinal regressions points to the connections between information that exist in the various exemplar clouds. A variable may not have a connection (and therefore, interaction) to another variable in a listener's exemplar cloud on "monolingual speech," but may have that connection in the listener's exemplar could on "bilingual speech." This would explain why the social judgments are not made according to a general positive-negative binary; rather, some positive attributes could be assigned to monolinguals and others to bilinguals. These argue in favor of future research into the attributes that are assigned to these two groups.

Due to the limitations of the study, there are various next steps that future research can take to answer still-existing questions. First, although this study used laboratory speech, spontaneous speech should be used in perception tasks, because spontaneous speech can more accurately reflect the speech listeners encounter in real life. Next, it would be ideal to have bilingual speakers who speak the same variety of Spanish, to control for any dialectal differences that could impact their English pronunciation. Due to recruitment limitations, the participants of Experiment 1 did not have a wide range in their age of introduction to English. Future studies should include a wider range of ages of introduction to L2 acquisition, and overall, more participants to ensure a more representative sample, results of which should prove to be more generalizable. This study focused on three differences in the phonological inventories of Spanish and English, however many more remain. Next steps could include looking at other differences between the two, or other languages besides English and Spanish, in order to make more generalizable conclusions. Finally, future experiments should compare the effects of non-linguistic (e.g. visual stimuli) and linguistic (e.g. salient phonetic) features in perception tasks. This will allow linguists to separate individual parts of speakers' exemplar models and determine what factors have greater weight in speech perception.

REFERENCES

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

Ambridge, B. (2020). Against stored abstractions: A radical exemplar model of language acquisition. *First Language*, *40*(5-6), 509-559. https://doi.org/10.1177/0142723719869731

Barlow, J.A. (2014). Age of acquisition and allophony in Spanish-English bilinguals. *Frontiers in Psychology. 5*, 1-14. Doi: 10.3389/fpsyg.2014.00288

Barlow, J.A., Branson, P.E., & Nip, I.S.B. (2013) Phonetic equivalence in the acquisition of /l/ by Spanish-English bilingual children. *Bilingualism: Language and Cognition. 16*(1), 68-85. Doi: 10.1017/S1366728912000235

Best, C. T. (1991). The emergence of native-language phonological influences in infants: A Perceptual Assimilation Model. Haskins Laboratories Status Report on Speech Research. *108,* 1-30

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

Boomershine, Amanda, Kathleen Currie Hall, Elizabeth Hume & Keith Johnson. 2008. The impact of allophony versus contrast on speech perception. Contrasts in phonology: Theory, perception, acquisition 145–171.

Boersma, P. & Weenink, D. (2023). Praat: Doing phonetics by computer [Computer program], Version 6.4. http://www.praat.org.

Bohn, O.S. (1995). Cross-language speech perception in adults: First language transfer doesn't tell it all. In W. Strange (Ed.), Speech perception and linguistic experience: Issues in cross-language research (pp 279-304). Timonium, MD: York Press

Boswijk, V. and Coler, M. (2020) What is Salience? *Open Linguistics. 6* https://doi.org/10.1515/opli-2020-0042

Campbell-Kibler, K. (2007). Accent, (ing), and the social logic of listener perceptions. American Speech. 82(1), 32-64. Doi: 10.1215/00031283-2007-002

Cebrian, J., Gorba, C., & Gavalda, N. (2021) When the easy becomes difficult: Factors affecting the acquisition of the English /iː/-/ɪ / contrast. *Frontiers in Communication. 6.* https://doi.org/10.3389/fcomm.2021.660917.

Cho, T. & Ladefoged P. (1999) Variation and universals in VOT: evidence from 18 languages. Journal of Phonetics. 27(2) 207-229. https://doi.org/10.1006/jpho.1999.0094

Christensen, R. H. B. (2019). ordinal - Regression models for ordinal data. R Package Version 2019.12-10.

Corretge, Ramon (2012-2023). Praat Vocal Toolkit. https://www.praatvocaltoolkit.com

Drager, Katie & M. Joelle Kirtley. 2016. Awareness, salience, and stereotypes in exemplar-based models of speech production and perception. In Anna Babel (ed.), Awareness and control in sociolinguistic research, 1–24. Cambridge: Cambridge University Press.

Eckman, F.R. (2004) From phonemic differences to constraint rankings: Research on second language phonology. *Studies in Second Language Acquisition. 26*, 513-149. DOI: 10+10170S027226310404001X.

Escudero, P., & Boersma, P. (2004). Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition, 26*, 551–585.

Flege, J. E. (1995) Second language speech learning: Theory, findings, and problems. In Winifred Strange (editor) *Speech perception and Linguistic Experience: Issues in Cross-Language Research.* York Press. 233-277

Flege, J. E. & O.-S. Bohn. (2005). The perception of English vowels by native speakers of Spanish. *The Journal of the Acoustical Society of America 85*(S1). S85. https://doi.org/10.1121/1.2027177.

Flege, J.E., & Eefting, W. (1987). Production and perception of English stops by native Spanish speakers. *Journal of Phonetics. 15*(1), 67-83. Doi: 10.1016/S0095-4470(19)30538-8

Fowler, R., Hodge, B., Kress, G., & Trew, T. (1979). *Language and control*. Routledge. https://doi.org/10.4324/9780429436215

Fox, R.A., J.E. Flege & M.J. Munro. (1995). The perception of English and Spanish vowels by native English and Spanish listeners: A multidimensional scaling analysis. Journal of the Acoustical Society of America 97. 2540–2551.

Gluszek, A. & Dovidio, J.F. (2010) Speaking with a nonnative accent: Perceptions of bias, communication difficulties, and belonging in the United States. *Journal of Language and Social Psychology. 2*9(2), 224-234. Doi: 10.1177/0261927X09359590

Goldinger, S.D. 1996. Words and voices: episodic traces in spoken word identification and recognition memory. *Journal of experimental psychology: Learning, memory, and cognition*. American Psychological Association 22(5). 1166.

Goldrick, M. & Cole, J.. (2023). Advancement of phonetics in the 21st century: Exemplar models of speech production. *Journal of Phonetics 99*. 101254. https://doi.org/10.1016/j.wocn.2023.101254.

Hartshorne, J.K., Tenenbaum, J.B, & Pinker, S. (2018) A critical period for second language acquisition: Evidence from 2/3 million English speakers. *Cognition*. *177*. 263-277. https://doi.org/10.1016/j.cognition.2018.04.007

Hay, J. & Drager, K. (2010). Stuffed toys and speech perception. *Linguistics. De Gruyter Mouton 48*(4). 865–892. https://doi.org/10.1515/ling.2010.027.

Hay, J., Nolan, A., & Drager, K. 2006. From fush to feesh: Exemplar priming in speech perception. The Linguistic Review 23(3). 351–379.

Hurring, G., Hay, J., Drager, K., Podlubny, R., Manhire, L., and Ellis, A. (2022). Social Priming in Speech Perception: Revisiting Kangaroo/Kiwi Priming in New Zealand English. *Brain Sciences. Multidisciplinary Digital Publishing Institute 12*(6). 684. https://doi.org/10.3390/brainsci12060684.

Johnson, K. (1997a). Speech perception without speaker normalization: an exemplar model. In K. Johnson, & J. W. Mullennix (Eds.), Talker variability in speech processing (pp. 145–166). San Diego: Academic Press.

Johnson, K. (2005). Speaker normalization. In R. Remez, & D. B. Pisoni (Eds.), The Handbook of Speech Perception

Kang, O. and Rubin, D.L. (2009) Reverse Linguistic stereotyping: Measuring the effect of listener expectations on speech evaluation. *Journal of Language and Social Psychology. 28*(4) 441-456. 10.1177/0261927X09341950

Kondaurova, M.V. & Francis, A.L. (2010) The role of selective attention in the acquisition of English tense and lax vowels by native Spanish listeners; Comparison of three training methods. *Journal of Phonetics. 38*(4), 569-587. Doi:10.1016/j.wocn.2010.08.003

Kuhl, P. K. and Iverson, P. (1995) Linguistic experience and the "Perceptual Magnet Effect" In Winifred Strange (editor) *Speech perception and Linguistic Experience: Issues in Cross-Language Research.* York Press. 121-154.

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review, 99*, 22–44.

Magen, Harriet S. 1998. The perception of foreign-accented speech. *Journal of Phonetics 26(*4). 381–400. https://doi.org/10.1006/jpho.1998.0081.

Major, R. C. (1986) English voiceless stop production by speakers of Brazilian Portuguese. *Journal of Phonetics. 15*(2), 197-202. Doi:10.1016/S0095-4470(19)30560-1

Mcbride, K. (2015). Which features of Spanish learners' pronunciation most impact listener evaluations? *Hispania. 98*(1), 14-30. Doi: 10.1353/hpn.2015.0001

Morrison, G.S. (2009) L1-Spanish speakers' acquisition of the English /i/-/ɪ/ contrast II: Perception of Vowel Inherent Spectral Change. *Language and Speech. 52*(4), 437-462. doi: 10.1177/0023830909336583

Navarra, J., Sebastian-Galles, N., & Soto-Faraco, S. (2005). The perception of second language sounds in early bilinguals: New evidence from an implicit measure. *Journal of Experimental Psychology: Human Perception and Performance. 31*(5), 912-918. Doi:10.1037/0096-1523.31.5.912

Niedzielski, N. (1999). The Effect of Social Information on the Perception of Sociolinguistic Variables. Journal of Language and Social Psychology 18(1). 62–85. https://doi.org/10.1177/0261927X99018001005.

Newman, M. & Wu, A. (2011) Do you sound Asian when you speak English? Racial identification and voice in Chinese and Korean Americans' English. *American Speech. 86*(2), 152-178. Doi: 10.1215/00031283-1336992

Pierrehumbert, J.B. (2000) Exemplar dynamics: Word frequency, lenition and contrast. In Bybee J.L. & Hopper P.J. Frequency and the Emergency of Linguistic Structure. John Benjamins Publishing. https://doi.org/10.1075/tsl.45

Pierrehumbert, Janet (2001). Exemplar dynamics, word frequency, lenition, and contrast. In Bybee and Hopper (eds). Frequency Effects and the Emergence of Linguistic Structure,137–157. Amsterdam: John Benjamins.

Purnell, T., Idsardi, W., & Baugh, J. (1999) Perceptual and phonetic experiments on American English dialect identification. *Journal of Language and Social Psychology. 18*(1), 10-30.

Rahman, M.M., Pandian, A., Karim, A., & Shahed, F.H. (2017) Effect of age in second language acquisition: A critical review from perspective of Critical Period Hypothesis and Ultimate Attainment. *International Journal of English Linguistics. 7*(5) http://doi.org/10.5539/ijel.v7n5p1

Recasens, D. (2012). A cross-language acoustic study of initial and final allophones of /l/. *Speech Communication. 54*(3), 368-383. doi:10.1016/j.specom.2011.10.001

Rubin, D.L. (1992). Nonlanguage factors affecting undergraduates' judgments of nonnative English-speaking teaching assistants. *Research in Higher Education. 33*(4), 511-531. Doi: 10.1007/BF00973770

Schmidt, L.B. (2020). Role of developing language attitudes in a study abroad context on adoption of dialectal pronunciations. Foreign Language Annals, 53(1), 785-806. https://doi.org/10.1111/flan.12489

Schmidt, L.B. (2022) Second language development of dialect awareness in Spanish. *Hispania*. 05(2), 267-284. Doi: 10.1353/hpn.2022.0055

Staum Casasanto, L. (2008). Does social information influence sentence processing? *Proceedings of the 30th annual meeting of the Cognitive Science Society (CogSci 2008), 30*(30), 799–804.

Tsang, A. (2020) Why English accents and pronunciation 'still' matter for teachers nowadays: a mixed-methods study on learners' perceptions, *Journal of Multilingual and Multicultural Development, 41*(2), 140-156, DOI: 10.1080/01434632.2019.1600528

U.S. Census Bureau.(2022) "Language Spoken at Home." *American Community Survey, ACS 1-Year Estimates Subject Tables, Table S1601*, https://data.census.gov/table/ACSST1Y2022.S1601?t=Language Spoken at Home.

Vaughn, C. R. 2019. Expectations about the source of a speaker's accent affect accent adaptation. *The Journal of the Acoustical Society of America. Acoustical Society of America 145*(5). 3218–3232. https://doi.org/10.1121/1.5108831

Veríssimo, J. (2021). Analysis of rating scales: A pervasive problem in bilingualism research and a solution with Bayesian ordinal models. *Bilingualism: Language and Cognition*, *24*(5), 842–848. doi:10.1017/S1366728921000316

Wardhaugh, R. (1970). The Constrative Analysis Hypothesis. *TESOL Quarterly.* 123-130.

Walker, M., Szakay, A., and Cox, F. (2019). Can kiwis and koalas as cultural primes induce perceptual bias in Australian English speaking listeners? Laboratory *Phonology: Journal* *of the Association for Laboratory Phonology 10*(1). 7. https://doi.org/10.5334/labphon.90.

Wig, N. & Garcia-Sierra, A. (2021) Matching the mismatch: The interaction between perceptual and conceptual cues in bilinguals' speech perception. *Bilingualism: Language and Cognition. 24*(3), 467-480. Doi: 10.1017/S1366728920000553

Zellou, G., Cohn, M. & Pycha, A. 2023. Listener beliefs and perceptual learning: Differences between device and human guises. *Language*. Linguistic Society of America (1). https://muse.jhu.edu/pub/24/article/913403.

APPENDICES

Appendix A: Experiment 1 Stimuli

- Can you say “bid”?
- Can you say “bead”?
- Can you say “pit”?
- Can you say “Pete”?
- Can you say “tin”?
- Can you say “teen”?

- I have a car
- I have a truck
- I have a flower
- I have a doll

- Can you say “late”?
- Can you say “law”?
- Can you say “lit”?
- Can you say “tale”?
- Can you say “tall”?
- Can you say “till”?

- I want a cat
- I want a vacation
- I want a diploma
- I want a house

- I like to text
- I like to tan
- I like to teach
- I like to dance
- I like to date
- I like to daydream

Appendix B: Experiment 2 questions

On a scale from 1-5, how educated is the speaker in the recording, with 1 being uneducated and 5 being highly educated.

Uneducated Educated

1 2 3 4 5

On a scale from 1-5, how wealthy is the speaker in the recording, with 1 being not wealthy and 5 being wealthy.

Not wealthy Wealthy

1 2 3 4 5

On a scale from 1-5, how polite is the speaker in the recording, with 1 being rude and 5 being polite.

Rude Polite

1 2 3 4 5

On a scale from 1-5, how native-like does the speaker in the recording sound, with 1 being non-native speaker of English and 5 being native speaker of English.

Non-native English speaker Native English speaker

1 2 3 4 5

On a scale from 1-5, how easy is it to understand the speaker in the recording, with 1 being difficult and 5 being easy.

Difficult Easy

1 2 3 4 5

On a scale from 1-5, how likely do you believe it is for the speaker to have been born in the US, with 1 being unlikely (not born in the US) and 5 being likely (born in the US).

Unlikely Likely

1 2 3 4 5

Were there any noticeable features about the recording that influenced your answers? (Optional)