The L2 Perception of Word-Initial Voicing Stop Contrasts in Spanish[1]

Guangbo Chen

A thesis submitted in partial fulfilment of the requirements for the degree of Bachelor of Arts in Linguistics

Bryn Mawr College

December 2024

Abstract

The present study examines the perception of voicing stop contrasts (/p-b/, /t-d/, /k-g/) in Spanish at the word-initial position by non-native learners. Collected by an experiment in the format of the AXB discrimination test, the results support a positive correlation between Spanish learning experiences and general performance in the perceptual test. The positive correlation also applies to the perception of individual contrast pairs /p-b/ and /t-d/ but does not extend to the perception the /k-g/ contrast. Other factors (including immersion and frequency of using and listening to Spanish) do not show support for any significant correlation with the performance in the perception test.

[1] It would have been impossible for me to complete this thesis without help from many people. I want to express my gratitude to Cassadra and Deman, who helped me record the audios for the experiment and to my thesis advisor, Prof. Noah Elkins, for his invaluable guidance throughout the process. I am also grateful to Prof. Rikker Dockum and my student readers, Kit and Michaela, for their enlightening suggestions and feedback. A special thanks goes to my friends, especially Jie, who accompanied me during the drafting process and helped me prepare for presentations. Finally, I would like to thank my mother for her continued support throughout my life journey.

Table of Contents

1. Introduction 3
2. Background 4
2.1. L2 Speech Perception in General 4
2.2 L2 Perception of Spanish in Particular 7
3. Methods 8
3.1 Present Study 8
3.2 Experiment Mechanics 10
4. Results 11
4.1 General Results 11
4.2 Accuracy of Perception Given Spanish Background 12
4.2.1 Learning Experiences of Spanish (Years of Learning Spanish) 13
4.2.2 Immersion Experiences 14
4.2.3 Frequency of Using Spanish 15
4.2.4 Frequency of Listening to Spanish 16
4.3 Interpretations 17
5. Discussions 18
5.1. Implication 18
5.2. Limitations 19
6. Conclusion 20
7. Appendix 21
7.1. Demographic Questions 21
7.2. Main Questions 22
8. References 26

1. Introduction

Learning an L2 language is usually not an easy but challenging task for adult learners, and the most common difficulties are in the perception and production of certain sound segments (Patience, 2022, p.2). For many L2 learners of Spanish, perceiving the voicing stop contrasts at the word-initial position is important because they are central to distinguishing meanings. For example, while the difference between /p/ and /b/ is hard for many L2 learners of Spanish to detect, /pata/ (‘leg’) and /bata/ (‘robe’) refers to 2 different objects. However, the L2 perception performance varies across the population, and learning experiences are expected to play a role in shaping perceptual accuracy. The present study expects to test any correlation between the learning experiences in Spanish and perceptual results for the research question on “How do L2 learners of Spanish perceive voicing stop contrasts at the word-initial position.”

To find the patterns behind the L2 perception of voicing stop contrasts in Spanish, the perceptual performance of L2 Spanish learners is tested with an experiment in the format of an AXB discrimination test. Including testing questions made of both minimal pairs and distraction pairs, the results show support for the correlation between learning experiences in Spanish and the general performance in the perception test, and the same case also applies to the perception of the specific samples on the /p-b/ and /t-d/ contrast. On the other hand, the perception results of the /k-g/ contrast do not provide support for the hypothesis with its large p-value ($p = 0.165 > 0.05$, $r = 0.204$). In addition, there is no significant correlation founded between other factors (including immersion, frequency of using Spanish, and frequency of listening to Spanish) and performance in the perception test.

More details of this study are covered in the following sections. Section 2 discusses the L2 speech perception in general and factors behind the L2 perception of Spanish, examining existing frameworks that contribute to the set-up of the current thesis. Section 3 outlines the research goal of the present study and experiment design. Section 4 presents the results and analysis with correlation and ANOVA (analysis of variance). Section 5 mentions the limitations existing in the present study and implications for future studies to take on.

2. Background

2.1. L2 Speech Perception in General

The study of sounds is an integral part of learning any language, but also a more challenging task for many adult L2 learners. In this context, L1 refers to native languages acquired during childhood, while L2 refers to any non-native language learned in later life. For learning any language, mastering a native-like production of sounds is expected for successful learners, yet it can't be achieved without an accurate perception of the new sounds. Because production and perception are correlated, accurate perception is expected to be reached before having satisfactory production; in reverse, failure in perceptual accuracy will cause troubles for production later (González-Bueno, 2019, p.254). Second language acquisition thus requires precise perception of contrasts in the input in the first place (Brown, 2000, p.7).

For adult L2 learners, there have been demonstrations of difficulties producing certain sounds and sound contrasts in L2 (Simonet, p.730). This also applies to the case of perception, as adult L2 learners usually face difficulties in the perception of non-native sounds, especially those that are not used in their native languages (Barrios et al., 2016; Kuhl, p.104). Some widely known examples are the difficulties of Japanese learners of English in perceiving the /ɹ/-/l/ contrasts (Miyawaki, 1975; Flege, 1995). Other experiments have presented examples such as the difficulties faced by L1 Spanish speakers in perceiving the /i/-/ɪ/ contrast in English and L1 Korean speakers in perceiving the /e/-/æ/ contrast in English (Paola & Boersma, 2004; Kartushina et al., 2015). In the above examples, although the above examples vary in the target sounds for acquisition, they all show that difficulties in perceiving sounds in L2 is a cross-linguistic pattern, implying that shared reasons could be possible for explaining the difficulties in the perception of non-native sounds.

Many previous studies have attributed the difficulties that exist in L2 perception to the sound structure in L1, which claims that the phonological system in L1 contributes to the acquisition of sounds in L2 (Kartushina et al., 2015). The perceptual magnet effect describes certain L1 sounds as magnets that make listeners perceive all similar sounds as the same, which then distorts the perception of the contrasts in their L2 (Kuhl, 1999, p.106). It argues that as people start to compile more experience during language learning, clusters are formed by individual exemplars of different segments. For example, certain sounds within a phonetic category would pull other nearby sounds closer to them perceptually, reducing the ability to discriminate between similar sounds (Lotto et al., 1998, p.3648). Hence, it is easier for L2 learners to distinguish the “between categories” that cross categorical boundaries but harder to determine the “within-category” sounds (Zsiga, 2014, p.455-456). Previous language experiences are therefore viewed as a powerful factor in the perception of L2 sounds (Levy & Strange, 2017). Because one of the biggest differences between L1 and L2 acquisition is the age when they start to learn a language (Perani et al., 1998), unlike acquiring L1, adult L2 learners already know a language (in some cases they may speak multiple first languages) and possess a mature phonological system (Fabra, 2005, p.76). Therefore, many adult L2 learners tend to replace a certain L2 sound with its closest L1 counterpart in perception, leading to perceptual inaccuracy (Oliveira, 2020, p.5). In addition, the speech contrasts in L1 convey meaning to listeners, whereas they may not in their L2 (Polka, 1992, p.37). For instance, since Spanish does not have the /ɪ/ sound, Spanish learners of English often perceive the English /ɪ/ as the closest Spanish equivalent, /i/, which causes the Spanish learners of English to perceive the word ‘bit’ (/bɪt/) as /bit/ (Flege, 1991; Brown, 2000). As a result, the cross-linguistic interaction affects speech perception in L2 (Oliveira, 2020, p.6).

When addressing the relative level of difficulty in perceiving non-native sound contrasts, the degree of acoustic similarity also plays a significant role in theoretical frameworks on L2 speech perception in addition to L1 influence. (Barrios et al., 2016, p.368; Escuerdo et al., 2011, p.280). The Speech Learning Model (SLM), for example, predicts perceptual accuracy based on the similarity between L1 and L2 sounds (Flege, 1995). It proposes that L2 learners’ ability to acquire a new sound is based on their accuracy in perceiving the sound, which is already shaped by their pre-existing L1 sound categories. If learners can discern some phonetic difference between L1 and L2 sounds, they can establish a new category for a respective L2 sound that differs phonologically from the closet L1 sound (SLM-H2, Flege 1995); if the L2 sound and its closet L1 counterpart have greater dissimilarity, it’s more likely for listeners to discern their phonetic differences (SLM-H3, Flege 1995) (Colatoni & Steele, 2008, p.492).

With regards to the degree of acoustic similarity, Flege classifies the sounds into 3 main categories based on the perceived differences between L2 sounds and their L1 counterparts, which are “identical,” “new,” and “similar.” The “similar” sounds are represented by the same IPA symbol but differ acoustically from the nearest L1 sound, while the “new” sound is represented by different IPA symbols and is acoustically different from its closest L1 counterpart (Barrios et al., p.369). It's argued that neither the "identical" or "new" L2 sounds would be difficult for L2 learners to perceive because they face little interference from L1. On the contrary, the "similar" sounds are viewed as the most difficult for L2 perception because of the negative interference brought by the similarity between the L2 sounds and its closest L1 counterparts (Flege, 1995).

Although other models are analyzed using different approaches, they also stay aligned with SLM on the impact of cross-linguistic similarity on perception, underscoring the role of L1 categories in shaping L2 perception, particularly when learners have to differentiate between similar sounds in both languages. To be more specific, a new L2 sound category can be established if distinguishable from the closest L1 category (Eger, 2019, p.180-181). As an example, Best's Perceptual Assimilation Model (PAM) proposes that the assimilation of non-native speech sounds into native categories depends on the phonological similarities between the L1 and the L2. According to PAM, perception is predicted to be easier when non-native contrasts are assimilated into two distinct native categories. In contrast, difficulties increase when non-native sounds are perceived to be a single native category (Broersma, 2005, p. 3890). To give an illustration, while it is challenging for Spanish learners of English to distinguish the [i] and [ɪ] ([i] exists in Spanish sound inventory but [ɪ] does not), it will be much easier for them to perceive the clicks in Zulu since there are no click sounds in Spanish (Zsiga, 2014, p.458).

In conclusion, the influence of L1 sound categories and the degree of cross-linguistic acoustic similarity both play an important role in L2 perception. Adult L2 learners often struggle with perceiving and producing sounds that are not present in their native language, particularly when the sounds are acoustically similar to those in L1. Theoretical frameworks such as the Speech Learning Model (SLM) and the Perceptual Assimilation Model (PAM) offer useful but diverging approaches to analyze the relationship between similarity and assimilation as well as difficulty and acquisition. Both models emphasize the role of L1 inference in perceiving L2 sounds, in which the more similar sounds will cause more difficulties, while the more distinct sounds between L1 and L2 will be easier to perceive and learn. However, while PAM focuses more on how L2 sounds are assimilated into existing L1 categories, it does not emphasize the role of learning experiences on perception over time. SLM, on the other hand, considers the effect of learning, suggesting that with increasing exposure and learning experience, learners can form new categories for L2 sounds and overcome some interferences from their L1.

2.2 L2 Perception of Spanish in Particular

Globally, Spanish is spoken by more than 400 million people as their 1st or 2nd language (Lipski, 2012). According to the annual report by Instituto Cervantes, over 22 million people are learning Spanish as a foreign language (Instituto Cervantes, 2021). In the US, Spanish is also one of the most spoken languages, with about 43.3 million speakers, and ranked as the most studied second language. Almost any higher institution offers Spanish as a second language, and more than 8 million students have taken it as their L2 (US Bureau of Census, 2023).

Compared with English or French, Spanish is more orthographically shallow and transparent, which means there is a strong correspondence between how it is spelled and pronounced (Lipski, 1994). Most letters in Spanish represent one single sound, which helps learners predict the pronunciation more easily based on the spellings. Spanish has 18 consonants and 5 vowels. Of the consonants, there are 6 stops, which are /p/, /b/, /t/, /d/, /k/, and /g/, which all belong to oral stops and are articulated when the airstream is blocked in both the mouth and the nasal cavity (Ladefoged & Johnson, 2014, p.15). Among them, /p t k/ are the voiceless stops while /b d g/ are the voiced stops. Voicing distinctions, as one of the most common phonetic contrasts, are the differences between voiced and voiceless sounds (Wiliams, 1977, p.169).

For L2 learners of Spanish, it is crucial to perceive and master the voicing stop contrasts, especially for those that appear at the word-initial position, which serves as the main cue to distinguish one from another in a minimal pair. For example, in the minimal pair /pata/ ('leg') and /bata/ ('robe'), the two are different in the word-initial voicing contrasts (the /p-b/ contrast). Since the phonemes /p/ and /b/ sound very similar and are likely to get confused, it is challenging for L2 learners to discern the differences.

The difficulty in perceiving voicing contrast for many L2 Spanish learners is caused by the need to respond to distinctions in Voiced Onset Time, also known as VOT (Mayr, 2014, p.1007). VOT, which refers to the interval between the release of a consonant closure and the start of the voicing, is an important acoustic cue in the perception of voicing contrasts (Ladefoged & Johnson, 2014, p.159). The voiceless sounds have a positive VOT with the absence of pre-voicing during intervals of articulatory closures; the voiced sounds are in reverse, which have a much shorter VOT and pre-voicing during intervals of articulatory closures (Vicente, 1964, p.109). There are no significant cross-dialectal differences in VOT within different Spanish varieties (Williams, 1977, p.169). However, in some varieties such as those spoken in Cuba and the Canary Islands, voiceless stops are realized as voiced. For example, within the word ‘campana’, the /p/ at the internal position get voiced ([kam**p**ana] → [kan**b**ana]). (Campos-Astorkiza, 2012, p.94).

Not only marked by the voicing contrasts, VOT is also the primary percept in distinguishing between the stop contrasts in English and Spanish. While aspiration is a percept in the voicing contrast in English, prevoicing serves the same function in Spanish. Prevoicing, which refers to a negative value of VOT, enables most Spanish speakers to separate the voiced initial stops from their voiceless counterparts. In English, although voiced stops have much shorter VOTs than voiceless stops, they rarely exhibit prevoicing. In contrast, Spanish voiced stops typically have negative VOTs due to prevoicing.

According to Flege’s classification of the L2 sounds based on their perceived similarities from those in L1, the voicing contrasts in Spanish and English can be considered as “similiar”, the most difficult for L2 perception. Although voicing contrast exists in both English and Spanish, it varies cross-linguistically. In terms of phonetic realization, the English /b/ is more similar to the Spanish /p/. For example, the mean VOT of a word-initial /p/ in English lasts around 39 ms, but its Spanish equivalent is below 20 ms; a voiced stop in English has a mean VOT much shorter than the corresponding voiceless stops, but their Spanish equivalents have already dropped below zero and become prevoiced (Szabó, 2020; McCarthy *et al*, 2013; Rosner *et al*, 2000; Mayr, 2014). VOT adjusts to multiple factors, not only cross-linguistically, but also according to the place of articulation. Due to the difference in air pressure, the further back a stop is produced in the oral cavity, the larger the VOT would be, as shown by the increasing VOT from /p/ to /b/ in Spanish and English (Mayr, 2014, p.1008).

3. Methods

3.1 Present Study

The present study focuses on the perception of Spanish voicing stop contrasts at the word-initial position by L2 learners of Spanish, who are either L1 or L2 English speakers due to the selection range of the participants. The target contrasts are /p/ vs /b/, /t/ vs /d/, and /k/ vs /g/. As told in the prior section, these voicing contrasts exist in both Spanish and English but have evident differences in VOT, and the VOT length of a Spanish stop is usually shorter than an equivalent in English. The intervocalic and word-final positions were not incorporated in this study. In the intervocalic position, lenition, which refers to the weakening of consonants, is common (Zsiga, p.233; Hualde et al., 2011, p.303). For example, the voiced stops in Spanish would become frictives at the intervocalic position (e.g. /b d g/ → [β ð ɣ] / V_V). Likewise, it is difficult to find minimal pairs at the word-final position since Spanish has a limited set of consonants that can occur at the end of words, with most examples being loanwords (e.g. robot, Internet). Interested in finding any cross-linguistic or voicing impacts on perception results, my current research question is:

1. RQ: How do L2 learners of Spanish perceive voicing stop contrasts /p/ vs /b/, /t/ vs /d/, and /k/ vs /g/ at the word-initial position? Is there any correlation between experiences of learning Spanish and perception results?

Learning a new sound also means forming a new L2 category, but L2 formation is almost as long as L1 (Flege & Bohn, 2021, p.20). As the exposure to L2 and input increases, learners are expected to gradually acquire new sounds (Flege & Bohn, p.14; Best & Tyler, 2007, p.15). As a result of this, the hypothesis is presented below.

2. H1: Participants with more experience of learning Spanish are expected to discriminate the target sounds more accurately.

   And here is the null hypothesis:

3. H0: There is no significant correlation between experience of Spanish and perception.

Despite the important role learning experiences have in the perception of sound contrasts, increasing experience in L2 does not necessarily lead to better perception results. The Speech Learning Model (Flege, 1995) proposed that L2 learners would be able to establish a new phonological category for a sound in L2 if they could discern some phonetic differences between an L2 sound and the closest L1 sound (Colatoni & Steele, 2008, p.492). Therefore, the similarity between L1 and L2's phonological structures matters here. If the target sound in L2 and its closest sound in L1 have more dissimilarities, it is more likely to discern their phonetic differences (Broersma, 2005, p.3890). Age (which is related to a critical age theory in 2nd language acquisition) and phonological categories are two other potential factors in perception differences (Best & Tyler, p.16).

An experiment is designed to check which hypothesis is more persuasive. The experiment will be explained in detail in the next section.

3.2 Experiment Mechanics

Mostly inspired by Patience's study (2022) on the perception of contrasts between Spanish phonemes and Levy & Strange's (2017) on American English speakers' perception of French vowel contrasts, both of which applied AXB discrimination tests to test participants' perception of L2 sounds, the current study was also conducted in the form of an AXB discrimination test, in which the X served as the stimuli to be identified and for A or B to compare to. For example, a series of recordings would be played in the order of:

A. /kama/ X. /kama/ B. /gama/

In this test trial, B was the one different from X, so the participants would be expected to choose B and they would receive a point. They would not be given the respective credit if they chose A in this scenario. Besides A or B, the participants could also choose "Neither" if they thought neither A nor B was different from X. There were 18 minimal pairs in this experiment, 6 for each stop contrast. Besides them, 6 distraction pairs were added, in which all 3 tokens in a trail would be the same. At the end of the test, a summative score (with a maximum of 24 points) would be calculated based on the participant's accuracy rate. The results were not judged, and their only function was to test the research question's hypothesis.

The testing stimuli were recorded by a native Mexican Spanish speaker in her twenties, who also spoke English as her L2. The average VOT of the target contrasts generally match with the previous measurements, with the VOTs of /p/, /t/, and /k/lasting less than 25 ms, and those of /b/, /d/, /g/ all far below 0 ms (McCarthy *et al*, 2013). The speaker was compensated with the thesis funding granted by the Tri-Co Linguistics Department. As instructed, she recorded each stimulus twice at a normal speed, volume, and pitch, so any distracting factor such as the changing intonation within the contrast pairs was minimized. The audio files were transferred to me in MP3 format, and I then applied Audacity to eliminate the background noise so the audibility of all stimuli could be ensured.

The experiment was conducted on a Google Form and open to the Tri-Co community. It consisted of 2 parts: the demographic questions and the main questions on the perception of voicing contrasts. The audio files were not labeled in the test and the questions were randomized. The test did not have a time limit for completion, and the participants could play the audio as many times as needed. More details on the demographic and main questions can be found in the appendix.

4. Results

The experiment was concluded after 6 weeks of trial. After the experiment was concluded, the results were first downloaded. The data was then reorganized into separate Excel spreadsheets based on the factors for analysis. Later, they were exported into JASP, a statistical analysis software program. Statistical elements such as mean, range, standard deviation, accuracy, and p-value were calculated with Excel and JASP.

4.1 General Results

The experiment received a total of 48 responses, and all participants were recruited from the Tri-Co community (M age = 19.92). The main recruiting mediums were a link shared among the Tri-Co Linguistics email list and the Spanish department, and distributed flyers at the campus. Participation was totally voluntary, and participants were given extra credits on their coursework (for LING H115) and snacks (at the on-campus recruiting event) as incentives.

Perception accuracy was high, with an average score of 19.54/24, a median score of 21/24, and individual results ranging from 0 to 24.

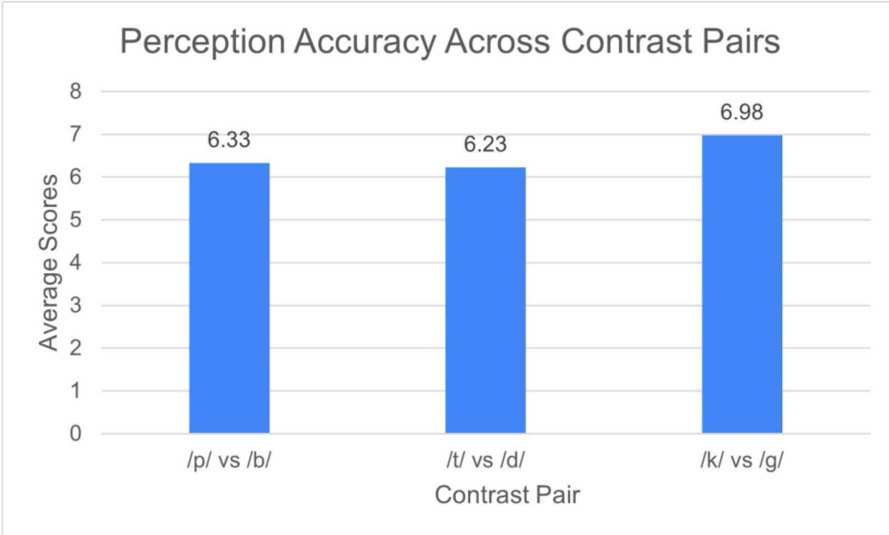

Figure 1 Perception Accuracy Across Contrast Pairs

The 3 contrast pairs present distinct patterns in perceptual accuracy. The /p-b/ contrast has an average score of 6.33/8, with a standard deviation of 1.83. Among its 8 testing samples, the distraction pair /bena/ vs /bena/ receives the highest average scores (M = 0.94). The minimal pair /pata/ vs. /bata/ receives the lowest average score (M = 0.69). The /t-d/ contrast shows similar results (M = 6.23, SD = 2.09). Among the 8 testing samples of the /t-d/ contrast, the minimal pair /tuba/ vs. /duba/ receives the highest average scores (M = 0.88). The minimal pair /tala/ vs. /dala/ receives the lowest average score (M = 0.60). Compared with the 2 other pairs, the /k-g/ contrast presents a higher perceptual result (M = 6.98, SD = 1.60). Among the 8 testing samples of the /k-g/ contrast, the distraction pair /kola/ vs. /kola/ and the minimal pair /kasa/ vs /gasa/ receive the highest average scores (M = 0.96). The minimal pair /kaɾo/ vs. /gaɾo/ receives the lowest average score (M = 0.69).

4.2 Accuracy of Perception Given Spanish Background

To test the correlation between L2 learners' perception of word-initial voicing stop contrasts in Spanish and learning experiences of Spanish, several variables, such as the years of learning Spanish and frequency of using Spanish in daily life, are included.

Since the experiment is conducted on a test-based questionnaire, the perceptual results, represented by the total test scores and performance on individual contrast pairs, serve as the dependent variable. The independent variables include numerical (e.g., years of learning Spanish, age when starting to learn Spanish) and categorical between-subject variables (e.g., immersion experience, frequency of language use, etc.). All dependent variables are numerical, so correlation analysis and ANOVA (Analysis of Variance) are applied to identify possible statistically significant patterns (as shown in Figure 2).

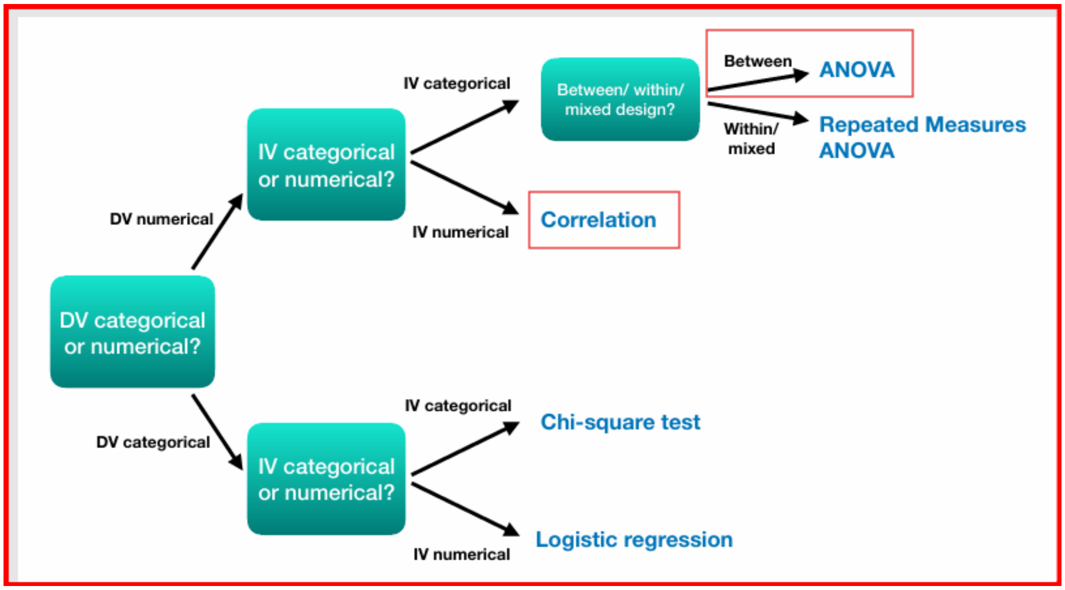

Figure 2 Klapper, A. (2024, June). Statistical Test Decision Chart

4.2.1 Learning Experiences of Spanish (Years of Learning Spanish)

The primary independent variable analyzed is the years of learning Spanish (representing the learning experiences), and its corresponding dependent variables are the total scores in the perception test and partial scores gained in each voicing contrast. Since both independent and dependent variables are numerical, the correlation analysis is applied.

Correlation ▼

Pearson's Correlations

			Pearson's r	p
Years of Learning Spanish	-	Total Scores	0.368	0.010

Figure 3 Years of Learning Spanish vs Total Scores-Correlation (p-value)

Since the p-value is 0.010 ($p < 0.05$, $r = 0.368$), there is a statistically significant positive correlation between the learning experiences of Spanish (as indicated by years of learning Spanish in the questionnaire) and the total scores in the perception test, providing support for H1.

Pearson's Correlations

			Pearson's r	p
NYSL	-	SUM-Scores	0.288	0.047

Figure 4 Years of Learning Spanish vs Partial Scores (p-b contrast)-Correlation

Pearson's Correlations

			Pearson's r	p
NYSL	-	SUM-Scores	0.442	0.002

Figure 5 Years of Learning Spanish vs Partial Scores (t-d contrast)-Correlation

The same analytical method is also applied to individual voicing contrasts. For the /p-b/ contrast, there is a relatively weak positive correlation between learning experiences of Spanish and perceptual performance in these two contrast pairs (p=0.047, r=0.288). For the /t-d/ contrasts, there is a stronger statistically significant relationship (p=0.002, r=0.442), and therefore providing support for H1.

Pearson's Correlations

			Pearson's r	p
NYSL	-	SUM-Scores	0.204	0.165

Figure 6 Years of Learning Spanish vs Partial Scores (k-g contrast)-Correlation

Unlike the /p-b/ or /t-d/ contrast, there is no support for H1 regarding the relationship between learning experiences in Spanish and the perceptual results of the /k-g/ contrast. This is indicated by its weak correlation (r=0.204) and large p-value, up to 0.165 and way higher than 0.05. Therefore, there is no statistically significant relationship between learning experiences in Spanish and the perception of the /k-g/ contrast. The potential reasons are discussed in *Section 4.3 Interpretations*.

4.2.2 Immersion Experiences

Unlike learning languages in an academic setting, which is usually acquired as a L2, immersion refers to language acquisition in a more natural setting, usually in a home or surrounding environment (Day & Shapson, 1991). In the present experiment, the immersion experience is self-reported by participants in the demographic- questions and entered as a "Yes/No" question. An ANOVA (analysis of Variance) is conducted with the immersion experiences of Spanish (Yes/No) as the independent variable and the total scores in the perception test as the dependent variable. The results show no significant difference between the frequency of using Spanish in daily life, F(1, 46) = 0.02, p = 0.892.

ANOVA - Total Scores

Cases	Sum of Squares	df	Mean Square	F	p
Immersion(Yes/No)	0.438	1	0.438	0.018	0.892
Residuals	1091.478	46	23.728		

Note. Type III Sum of Squares

Figure 7 Immersion vs Total Scores-ANOVA

Specifically, the total scores in the perception test are slightly higher in the group that have immersion experiences of Spanish (M = 20.00, SD = 5.66) compared to the group that do not have immersion experiences of Spanish (M = 19.52, SD = 4.85).

Descriptives

Descriptives - Total Scores

Immersion(Yes/No)	N	Mean	SD	SE	Coefficient of variation
No	46	19.522	4.852	0.715	0.249
Yes	2	20.000	5.657	4.000	0.283

Figure 8 Immersion vs Total Scores-Descriptives

4.2.3 Frequency of Using Spanish

The frequency of using Spanish (as indicated by 'FUSDL' in the related charts) is analyzed as an independent variable and measured at four different levels (never, occasionally, frequently, daily) based on participants' self-reported responses. The ANOVA is conducted with the total scores in the perception test as the dependent variable. The results show no significant difference between the frequency of using Spanish in daily life, F(2, 45) = 0.94, p = 0.398.

ANOVA - Total Scores ▼

Cases	Sum of Squares	df	Mean Square	F	p
FUSDL	43.767	2	21.884	0.940	0.398
Residuals	1048.150	45	23.292		

Note. Type III Sum of Squares

Figure 9 FUSDL vs Total Scores-ANOVA

Specifically, the total scores in the perception test are lowest in the group that never use Spanish in daily life (M = 18.92, SD = 4.39) and highest in the group that frequently use Spanish in their daily life (M = 23.50, SD = 0.71).

Descriptives ▼

Descriptives - Total Scores

FUSDL	N	Mean	SD	SE	Coefficient of variation
Occasionally	21	19.905	5.412	1.181	0.272
Never	25	18.920	4.387	0.877	0.232
Frequently	2	23.500	0.707	0.500	0.030

Figure 10 FUSDL vs Total Scores-Descriptives

4.2.4 Frequency of Listening to Spanish

Similar to the frequency of using Spanish, the frequency of listening to Spanish is also a categorical independent variable and measured at four different levels (never, occasionally, frequently, daily). However, the ANOVA cannot be conducted due to the unequal distribution between groups, since at least two observations in each group are required by JASP to perform any analysis of variance.

Descriptive Statistics

	Total Scores			
	Occasionally	Never	Frequently	Daily
Valid	22	14	11	1
Missing	0	0	0	0
Mean	19.182	17.857	22.000	24.000
Std. Deviation	3.527	7.156	1.897	
Minimum	13.000	0.000	18.000	24.000
Maximum	24.000	24.000	24.000	24.000

Figure 11 FLS vs Total Scores-Descriptive Statistics

Specifically, the total scores in the perception test are the highest in the group who listen to Spanish daily (M = 24.00), the lowest in the group who never listen to Spanish (M = 17.86, SD = 7.16) The groups who frequently (M = 22.00, SD = 1.90) and occasionally (M = 19.18, SD = 3.53) listen to Spanish rank as the 2nd and 3rd in the perceptual results.

4.3 Interpretations

For the relationship between learning experiences in Spanish and general scores in the perception test, the p-value is smaller than 0.05. In addition, the results and measurements are mostly reliable, and it is likely that there will be a meaningful difference or a significant correlation based on the learning experiences in Spanish. However, since the details of the independent variables are collected in an online questionnaire and self-reported, there may be concerns about the validity of the results. Similar cases also apply to the perception of the /t-d/ contrast (p = 0.002, r = 0.442), but have a weaker positive correlation for the /p-b/ contrast (p = 0.047, r = 0.288). In contrast, there is no support for any significant correlation between the learning experiences in Spanish and the performance in the perception of the /k-g/ contrast pair, which is represented by its high p-value (p = 0.165, r = 0.204). The striking difference in perceiving the tested samples of the /k-g/ contrast pair may be caused by the relatively closer difference in VOT between an English /k/ and a Spanish /k/ compared with the difference between an English /p/ and a Spanish /p/. Another potential reason may come from the irrelevant acoustic cue the participants picked for differentiating the sound contrasts, even though the background noises have been eliminated to the largest extent.

For the other independent variables (immersion experience, frequency of using Spanish, frequency of listening to Spanish), the p-values are much larger than 0.05, hence the perceptual results do not differ significantly between the groups. The other reasons may be the insufficient sample size and the sampling that is not representative enough. Using an exact test run by G Power Analysis, the appropriate sample size for detecting a significant correlation (effect size=0.3, α=0.05, 1-β=0.8, two-tailed) is 84. The ideal sample sizes for the between-subject comparison are even larger, which are 128 for 2 group comparison (effect size=0.25, α=0.05, 1-β=0.8, two groups) and 180 for 4 groups (effect size=0.25, α=0.05, 1-β=0.8, four groups) respectively. The sample size in this experiment is far below. Since participation in the experiment is completely voluntary, the participants who choose to take the test may have been more interested in the research topic or more skilled at perceiving contrast pairs, potentially leading to a biased sample.

5. Discussions

The previous section presents the results of the study and provides an analysis based on statistical models, and this section discusses the broader implications of these findings and addresses the study's limitations.

5.1. Implication

The present study is inspired by previous research on L2 perception (Levy & Strange, 2017; Patience, 2022), seeking any correlation between learning experiences and perceptual results. It follows the existing framework of Flege's SLM and Best's PAM on the influence of L1 on L2 perception of the voicing stop contrasts (Flege & Bohn, 2001; Best & Tyler, 2007), and the VOT is applied as a percept to explain factors that could influence cross-linguistic perception. The results show a preference for the SLM, as reflected by the statistically significant correlation between general discrimination performance and learning experiences. SLM emphasizes the role of learning in shaping perceptual outcomes, whereas PAM does not highlight the potential impact of language learning on the discrimination of voicing contrasts.

The present study adds a new perspective to the existing research by focusing on the word-initial voicing stop contrasts in Spanish, which is a challenging area for many L2 Spanish learners but has received little notice in previous research. Using minimal pairs and distraction pairs, participants' perceptual accuracy across different voicing contrasts is analyzed. The perceptual results of individual samples also inspire my further investigation into the reasons behind the notably high performance on the /k-g/ contrast and relatively lower perceptual accuracy on the minimal pairs /pata/ vs /bata/ and /tala/ vs /dala/. In addition, more innovative instructional strategies in L2 learning should also be designed to address the perception and production of challenging contrast pairs.

5.2. Limitations

One major limitation in this study is the exclusion of complete linguistic background of the backgrounds, while participants might speak additional languages as their L1 or L2 besides Spanish and English, potentially impacting the perceptual results of the target language.

Onishi's study on the impacts of L2 experience on L3 perception of phonological contrasts (Onshi, 2016), in which L1 speakers of Korean who already know English as their L2 were tested their knowledge of Japanese as their L3, tests that the level of L2 proficiency is generally related to the perception of phonological contrasts in L3 but limited to a few contrasts. Onishi argues that the perceptual results of L3 are more likely to be influenced by general learning experiences of languages rather than patterns associated with a particular language since multiple features of a single language can lead to different perceptual outcomes. For example, the geminate and singleton stops in Japanese are distinguished by their closure duration, which is not a percept in English. However, they are considered less difficult for Korean speakers as the same percept is applied in distinguishing the 2 different types of stops (Onishi, p.463). On the other hand, the voicing stop contrast in Japanese is difficult for Korean speakers because the voiceless and voiced stops are assimilated into the same category as Korean, but less difficult for English speakers as the voicing stop contrasts also exist in English (p.464).

The striking results of Onishi's study show that the impacts of additional languages on the perception are multifaceted, highlighting the need to standardize the comparison groups in future studies. Future research may extend the present work in several ways. Firstly, incorporating more phonotactic contexts helps present a more complete picture of participants' perceptual results. Secondly, comparing monolingual and bilingual speakers with their perception of contrasts in the target language is expected to provide insights into how language background influences perceptual abilities.

The other major limitations lay behind the fact that all responses to the demographic questions were self-reported, which led to the unequal distribution in each group and might bring a sampling that was not representative enough. For instance, the effects of years of learning Spanish might be diminished, as some participants might have longer learning experiences but didn't actively practice the language for a length of time, while some other participants might only actively learn the language for a short time but in an intensive way. Such predictions implied the limitations caused by the self-reported responses and were worthy of designing future studies to further eliminate the limitations.

Researchers interested in conducting similar studies can address the above limitations through several approaches. For example, the demographic questions can be enhanced by asking for a more complete linguistic background of the participants. Therefore, the perceptual performance alongside the similarity between the additional languages spoken by the participants can be compared and potentially address the impacts on perception based on the cross-linguistic similarity of the target sounds. Secondly, more detailed illustrations should be provided for demographic questions, and therefore lower the likelihood of miscommunication, which would potentially lead to biased results.

6. Conclusion

The present study focuses on the L2 perception of voicing stop contrasts at the initial position in Spanish, with the research question aiming to examine how do L2 learners of Spanish perceive voicing stop contrasts /p/ vs /b/, /t/ vs /d/, and /k/ vs /g/ at the word-initial position, and if there is any correlation between experiences of learning Spanish and perception results. The experiment, which is conducted in the format of an AXB discrimination test, provides results that are further analyzed using correlation and ANOVA based on different variables.

The general results of the perception test provide support for H1 (Participants with more experience of learning Spanish are expected to discriminate the target sounds more accurately), which has a p-value of 0.010 ($p < 0.05$, $r=0.368$) and therefore supports the hypothesis that there are positive correlational effects of learning experiences in the accurate discrimination of the target sounds. The hypothesis also applies to the partial scores in the perception of the /p-b/ ($p = 0.047$, $r=0.288$) and /t-d/ contrast ($p = 0.002$, $r=0.442$), but not for that of the /k-g/ contrast ($p = 0.165 > 0.05$, $r=0.204$). In addition, there is no significant correlation founded between other factors (including immersion, frequency of using Spanish, and frequency of listening to Spanish) and performance in the perception test. The results help reveal the underexplored areas in the existing research and call for more attention to the innovation of teaching strategies for the challenging contrast pairs. Future research may extend this work by exploring L2 perception under other phonotactic contexts such as the intervocalic and word-final position. The limitations of the present study can also be mitigated by incorporating a direct comparison of additional languages to different L2 learners who are learning the same target language in the perception test. For instance, L2 perception of voicing stop contrasts in Spanish can be further understood by comparing bilingual and monolingual speakers' perception of these contrasts in a linguistically distinct language (e.g., one from a different language family). Likewise, the experiment can also be enhanced by manipulating the VOT of the target sounds and creating a continuum, which benefits the identification of the threshold of categorical perception. Additionally, because the testing platform (Google Forms) lacked control over interstimulus intervals in the AXB discrimination test, future studies could address this limitation by standardizing the timing of each question.

7. Appendix

7.1. Demographic Questions

All responses to this survey will be kept strictly confidential, and no personal information will be shared or disclosed. The data collected will be used solely for an unpublished thesis, and an IRB exemption has been granted. Participation in this survey is completely voluntary, and you are free to withdraw at any stage without any consequences or obligations. If you have any concern, feel free to contact Guangbo Chen '25 through email (**schen2@brynmawr.edu**). Thank you so much for your participation!

1. What is your age?
   _____(fill in blank)

2. Which campus do you go to?
   a. Bryn Mawr b. Haverford c. Swarthmore

3. How many years have you been learning Spanish? (enter ‘0’ if you never learn Spanish)

______(fill in blank)

4. At what age did you start learning Spanish? (optional)
   _______(fill in blank)

5. Have you ever lived in an environment where Spanish is predominately spoken?
   a. Yes b. No

6. (Following Question 5) If yes, for how long?
   _______(fill in blank)

7. How often do you use Spanish in daily life?
a. Never b. Occasionally c. Frequently d. Daily

8. How often do you listen to Spanish?
b. Never b. Occasionally c. Frequently d. Daily

9. Do you have any known hearing impairments or issues that affect your ability to perceive sounds?
   b. Yes b. No

7.2. Main Questions

In this test, you will hear three sounds in each trial: **X**, **A**, and **B**. Your task is to determine which of the two sounds (**A** or **B**) is different from **X**. If you perceive A as the different one from X, please select A; if you perceive B as the different one from X, please select B; if neither of them is perceived as different from X, select "Neither".

For example, in a series of trial: [A: muda X: nuda B: nuda], you may select A as it is different from X.

The file names are only used for naming the audios, and they do not have any correlation with the test itself.

Minimal Pairs

cama vs gama (Apple)

/kama/ /gama/
'bed' 'range'

A=cama, X=cama, B=gama → Correct answer: B

casa vs gasa (Banana)

/kasa/ /gasa/
'house' 'gauze'
A = gasa, X = casa, B = casa → Correct answer: A

cana vs gana (Cantaloupe)

/kana/ /gana/
'gray hair' '3sg ganar'
A = cana, X = cana, B = gana → Correct answer: B

codo vs godo (Date)

/koðo/ /goðo/
'elbow' 'Gothic'
A = godo, X = codo, B = codo → Correct answer: A

caro vs garo (Elderberry)

/kaɾo/ /gaɾo/
'expensive' 'garum'
A = caro, X = garo, B = garo → Correct answer: A

cola vs gola (Fig)

/kola/ /gola/
'tail' 'ruff'
A = gola, X = gola, B = cola → Correct answer: B

pala vs bala (Grape)

/pala/ /bala/
'shorel' 'bullet'
A = bala, X = pala, B = pala → Correct answer: A

pata vs bata (Honeydew)

/pata/ /bata/
'leg' 'robe'
A = pata, X = bata, B = bata → Correct answer: A

pena vs vena (Kiwi)

/pena/ /bena/
'pitty, shame' 'vein'
A = vena, X = pena, B = pena → Correct answer: A

pesa vs besa (Lemon)

/pesa/ /besa/
'weighs' 'kisses'
A=besa X=pesa B=pesa → Correct answer: A

puso vs buso (Mango)

/puso/ /buso/
'3sg poner' 'hole'
A = buso, X = buso, B = puso → Correct answer: B

puro vs buro (Nectarine)

/puɾo/ /buɾo/
'pure' 'fulbr's earth'
A = buro, X = puro, B = puro → Correct answer: A

tala vs dala (Orange)

/tala/ /dala/
'felling' 'pump dale'
A = dala, X = dala, B = tala → Correct answer: B

toma vs doma (Papaya)

/toma/ /doma/
'3sg tomar' '3 sg domar'
A=doma, X=toma, B=toma → Correct answer: A

temo vs demo (Quince)

/temo/	/demo/
‘1sg temer’	‘computing music’

A=demo, X=demo, B=temo → Correct answer: B

tela vs dela (Raspberry)

/tela/	/dela/
‘fabric’	‘contraction of ‘de la’ ‘

A = dela, X = dela, B = tela → Correct answer: B

tuna vs duna (Strawberry)

/tuna/	/duna/
‘band’	‘dune’

A=tuna, X=tuna, B=duna → Correct answer: B

tuba vs duba (Tangerine)

/tuβa/	/duβa/
‘instrument’	‘bank’

A = duba, X = tuba, B = tuba → Correct answer: A

Distraction Pairs (All 3 Audios Are the Same):

cama vs gama (Ugli)

/kama/	/gama/
‘bed’	‘range’

A = cama, X = cama, B = cama → Correct answer: Neither

cola vs gola (Vanilla)

/kola/	/gola/
‘tail’	‘ruff’

A = cola, X = cola, B = cola → Correct answer: Neither pena vs vena (Watermelon)

/pena/ /bena/
'pitty, shame' 'vein'
A = vena, X = vena, B = vena → Correct answer: Neither

puso vs buso (Plum)

/puso/ /buso/
'3sg poner' 'hole'
A = puso, X = puso, B = puso → Correct answer: Neither

tala vs dala (Pineapple)

/tala/ /dala/
'felling' 'pump dale'
A = tala, X = tala, B = tala → Correct answer: Neither

tuba vs duba (Coconut)

/tuβa/ /duβa/
'instrument' 'bank'
A = duba, X = duba, B = duba → Correct answer: Neither

8. References

Barrios, S., Jiang, N., & Idsardi, W. J. (2016). Similarity in L2 phonology: Evidence from L1 Spanish late-learners' perception and lexical representation of English vowel contrasts. *Second Language Research*, *32*(3), 367–395. https://doi.org/10.1177/0267658316630784

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception. *Language Learning & Language Teaching*, 13–34. https://doi.org/10.1075/lllt.17.07bes

Broersma, M. (2005). Perception of familiar contrasts in unfamiliar positions. *The Journal of the Acoustical Society of America*, *117*(6), 3890–3901. https://doi.org/10.1121/1.1906060

Brown, C. (1997). The interrelation between speech perception and phonological acquisition from infant to adult. In *Second Language Acquisition and Linguistic Theory* (pp. 4–63). essay, Blackwell.

COLANTONI, L., & STEELE, J. (2008). Integrating articulatory constraints into models of second language phonological acquisition. *Applied Psycholinguistics*, *29*(3), 489–534. https://doi.org/10.1017/s0142716408080223

Day, E. M., & Shapson, S. M. (1991). Integrating formal and functional approaches to language teaching in French immersion: An experimental study*. *Language Learning*, *41*(1), 25–58. https://doi.org/10.1111/j.1467-1770.1991.tb00675.x

de Oliveira , D. M. (2020). *Auditory selective attention and performance in high variability phonetic training: The perception of Portuguese stops by Chinese L2 learners* (dissertation).

Del Saz, M. (2019). Native and nonnative perception of western Andalusian Spanish /s/ aspiration in quiet and noise. *Studies in Second Language Acquisition*, *41*(04), 673–694. https://doi.org/10.1017/s0272263119000056

Escudero, P., & Boersma, P. (2004). Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition*, *26*(4), 551–585. https://doi.org/10.1017/S0272263104040021

Escudero, P., Simon, E., & Mitterer, H. (2012). The perception of English front vowels by North Holland and Flemish listeners: Acoustic similarity predicts and explains cross-linguistic and L2 perception. *Journal of Phonetics*, *40*(2), 280–288. https://doi.org/10.1016/j.wocn.2011.11.004

Fabra, L. R. (2005). Predicting ease of acquisition of L2 speech sounds: a perceived dissimilarity test. *Vial-Vigo International Journal of Applied Linguistics*, 75–92.

Flege, J. E. (1991). Age of learning affects the authenticity of voice onset time (VOT) in stop consonants produced in a second language. *Journal of the Acoustical Society of America*, *89*(1), 395–411. https://doi.org/10.1121/1.400473

Flege, J. E. (1995). Second language speech learning: Theory, findings and problems. In *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233–277). essay, York Press.

Flege, J. E., & Bohn, O.-S. (2021). The revised speech learning model (SLM-R). *Second Language Speech Learning*, 3–83. https://doi.org/10.1017/9781108886901.002

Freeman, M. R., Blumenfeld, H. K., Carlson, M. T., & Marian, V. (2021). First-language influence on second language speech perception depends on task demands. *Language and Speech*, *65*(1), 28–51. https://doi.org/10.1177/0023830920983368

García Montero, L., Noguero Galilea, C., Pastor Villalba, C., López-Vega, M., & Lázaro Gozalo, A. (2021). *El Español en el mundo 2021: Anuario del Instituto Cervantes*. Instituto Cervantes ; Bala Perdida.

González-Bueno, M. (2019). Suggestions for teaching Spanish voiced stops /b, D, g/ and their lenited allophones [β̞, ð, ɣ]. Key Issues in the Teaching of Spanish Pronunciation, 60–83. https://doi.org/10.4324/9781315666839-4

Hernandez, S. D. and E. (2022, December 13). *Nearly 68 million people spoke a language other than English at home in 2019*. Census.gov.
https://www.census.gov/library/stories/2022/-

12/languages-we-speak-in-united-states.html

Kartushina, N., Hervais-Adelman, A., Frauenfelder, U. H., & Golestani, N. (2015). The effect of phonetic production training with visual feedback on the perception and production of foreign speech sounds. *The Journal of the Acoustical Society of America*, *138*(2), 817–832. https://doi.org/10.1121/1.4926561

Kissling, E. M. (2014). Phonetics instruction improves learners' perception of L2 sounds. *Language Teaching Research*, *19*(3), 254–275. https://doi.org/10.1177/1362168814541735

Kuhl, P. K. (1999). The Role of Experience in Early Language Development: Linguistic Experience Alters the Perception and Production of Speech. *The Role of Early Experience in Infant Development*, 101–125.

Ladefoged, P., & Johnson, K. (2014). *A course in phonetics* (7th ed.). Cengage Learning.

Levy, E. S., & Strange, W. (2008). Perception of French vowels by American English adults with and without French language experience. *Journal of Phonetics*, *36*(1), 141–157. https://doi.org/10.1016/j.wocn.2007.03.001

Lipski, J. M. (1994). *Latin American Spanish*. Longman.

Lipski, J. M. (2012). *Assessing linguistic proficiency in Spanish-English bilinguals: Current findings and methodological advances*. In K. Geeslin & M. Díaz-Campos (Eds.), *The Handbook of Spanish Second Language Acquisition* (pp. 130-147). Wiley-Blackwell.

MAYR, R., & MONTANARI, S. (2014). Cross-linguistic interaction in trilingual phonological development: The role of the input in the acquisition of the voicing contrast. *Journal of Child Language*, *42*(5), 1006–1035. https://doi.org/10.1017/s0305000914000592

McCarthy, K. M., Evans, B. G., & Mahon, M. (2013). Acquiring a second language in an immigrant community: The production of Sylheti and English stops and vowels by London-Bengali speakers. *Journal of Phonetics*, *41*(5), 344–358. https://doi.org/10.1016/j.wocn.2013.03.006

Patience, M. (2022). *Articulatory Difficulty in L2 Spanish* (dissertation).

Perani, D., Paulesu, E., Galles, N. S., Dupoux, E., Dehaene, S., Bettinardi, V., Fazio, F., Mehler, J. (1998). The bilingual brain: Proficiency and age of acquisition of the second language. *Brain: A Journal of Neurology*, *121*(10), 1841–1852. https://doi.org/10.1093/brain/121.10.1841

Polka, L. (1992). Characterizing the influence of native language experience on adult speech perception. *Perception & Psychophysics*, *52*(1), 37–52. https://doi.org/10.3758/bf03206758

Szabó , I. E. (2020). *Representational limitations and consequences of phonetic accommodation: English and Hungarian speakers' imitation of word-initial voiced and voiceless stops* (dissertation).

Zsiga, E. C. (2014). *The sounds of language: An introduction to phonetics and phonology*. Wiley-Blackwell.