[THE PHONOLOGY OF THE INITIAL STATE OF L3 ACQUISITION]

By

[KYLE PARRISH]

A Dissertation submitted to the

School of Graduate Studies

Rutgers, the State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

Graduate Program in [Second Language Acquisition and Bilingualism]

Written under the direction of

[Joseph V. Casillas]

And approved by

New Brunswick, New Jersey

[May] [2023]

ABSTRACT OF THE DISSERTATION

[THE PHONOLOGY OF THE INITIAL STATE OF L3 ACQUISITION]

by

[KYLE PARRISH]

Dissertation Director:

[Joseph V. Casillas]

This is where the body of your abstract goes. The abstract should summarize your work. The abstract for a dissertation or document may be longer than one page; word count is more important than page length in this section.

ACKNOWLEDGEMENTS

Insert your acknowledgements here.

DEDICATION

Insert your dedication here (optional).

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1: Introduction

1.1 Introduction

The difficulty of learning a new language in adulthood has a well documented history. This is especially true in the case of phonological acquisition. For example, many studies over the course of the past five decades have demonstrated that adult second language (L2) learners often produce and perceive the sounds of the target language in a non-native manner (J. Flege, 1991; J. Flege & Eefting, 1988). This difficulty arises, in part, because bilinguals often navigate complex communicative situations in which they produce and perceive speech from both of their languages in real time. Much less is known about the acquisition of a third language (L3A), particularly with regard to L3 production, perception, and phonological learning. The present dissertation explores the production patterns and perceptual categorization routines of adult bilinguals during the initial stages of L3 acquisition.

Empirical studies in L3 phonological acquisition have found evidence of multi-directional influence in language production (Llama & Cardoso, 2018; Llama, Cardoso, & Collins, 2010), and have found that the impact of previously learned languages varied in their relative influence on the third language. In some cases, L2 influence on L3 production has been found, where in others, L1 influence or simultaneous influence of both languages on L3 production has been reported. It is unclear what factors could be at the root of these varied findings.

At large, empirical studies in L3 acquisition have used low samples and their results have been used as the basis of predictive models. An issue with the use of few participants to make generalizations is that it increases the likelihood of false positive and false negative results due to an insufficient quantity of samples from a hypothetical population distribution to reliably infer information about that group. In addition to low samples, L3 studies typically use analyses such as an ANOVA, t-tests, non-parametric tests against zero, such as the Kruskal Wallis test. An issue with these choices of analyses is the lack of evidence for the null hypothesis within the larger umbrella of the frequentist approach to inferential statistics. That is, in the case of L3 acquisition, where the interest is which of two source languages impacts a third, the researcher is interested in the possibility of providing evidence that L3 performance is either L1 or L2 like. Often, the lack of a so-called "significant" p-value in frequentist analysis is taken as evidence of similar performance between groups or within subjects. This conclusion is not well justified, since it is unclear if null results are due to low samples with wide confidence intervals and how much noise is expected in the data. That is, the lack of a significant p-value alone is not evidence for the null hypothesis. When only a p-value is reported, and not detailed descriptive statistics, it is difficult to establish objective and precise criteria for evidence for practical equivalence.

Additionally, factors such as L2 and L3 proficiency, language dominance and choices of methods and analysis may play a role in the conclusions drawn by these studies. L3 proficiency has varied in the body of literature, with some studies examining beginners, while others study intermediate and advanced learners. Language dominance and use has not always been considered in L3 studies.

The present dissertation aims to address the issue of low sample sizes and varied statistical analyses by, firstly, recruiting bilinguals who do not yet know a third language. Secondly, more fitting statistical analyses are used to evaluate equivalence for within and between group comparisons, such as Bayesian Regression with a region of practical equivalence. In terms of sample size, it is probable that low samples in L3 research have occurred as a result of the difficulty in finding subjects with similar enough backgrounds in three languages. To work around the issue of low availability in participants, bilinguals' first exposure to a third language is studied in the present dissertation. It is likely that relatively homogeneous populations of bilinguals are in greater supply than trilingual populations, particularly when suggested methodological practices are to be used, such as the use of mirror-image groups (groups with the same L3 but opposite order of acquisition). With the increase in sample size, more precise observations and conclusions may be drawn about the very starting point of L3 phonological acquisition which are far less likely to be explained by statistical limitations such as sampling error.

Additionally, first exposure to the L3 can be very valuable information when it comes to modeling third language acquisition. Broadly, Third Language Acquisition (L3A) refers to the acquisition of a third language by an individual who already speaks to languages. Current models of L3A do not always account for potential individual differences in CLI, and typically are restrictive in their predictions. By investigating first exposure, it is possible to gather evidence regarding whether there is a default state of L1 and L2 CLI at first exposure to an L3, and whether this relative language influence is variable at the individual level. If the goal of L3 models is to predict the trajectory of L3 acquisition, then it seems pivotal to uncover the potential variability of its starting point.

To deal with the potential impact of proficiency and dominance on L3 production and perception, L2 proficiency was measured using the LexTALE in English (Izura, Cuetos, & Brysbaert, 2016; Lemhöfer & Broersma, 2016) and Spanish (Izura et al., 2016). The Bilingual Language Profile (BLP) was used to measure language dominance and background (Birdsong, Gertken, & Amengual, 2012). Finally, participant groups will be exposed to either German or French, to examine how phonetic similarity across languages may play a role in the first exposure to an L3.

1.2 Cross-lingusitic language features

In order to gain insights into the relative influence of a first and second language the present dissertation makes use of both (relative) voice-onset time (VOT) in stop production and perception and spectral qualities of vowels, specifically the first and second formants, in vowel perception and production. In the following sections, the vocalic systems of Spanish, English, French, and German, as well as their respective uses of voice-onset time, are overviewed.

1.2.1 Voice-Onset Time

Voice-onset time (VOT) is a phonetic measure in milliseconds of the release of the closure and its relationship with the onset of the vibration of the vocal folds (Lisker & Abramson, 1964). In particular, when the vocal folds begin to vibrate prior to the release, this is described as pre-voicing and is characterized by a negative VOT value. On the other hand, when the release occurs prior to the onset of voicing, a positive VOT is measured. Stops with a positive VOT are considered phonetically voiceless, where a negative VOT is phonetically voiced. VOT has been found to be a primary acoustic cue in the distinction of stop consonants in many of the world's languages. However, the use of VOT to make distinctions between sounds is language specific. For example, Spanish and French belong to a classification of languages referred to as true-voicing languages (Lisker & Abramson, 1964). This label suggests that the difference between a stop consonant that is phonemically voiced, such as /b/, is also phonetically voiced (negative VOT) and its counterpart that is phonemically voiceless, such as /p/, is also phonetically voiceless (positive VOT). True voicing languages stand in contrast to aspirating languages, such as German and English, in that the phonemic categories /p/ and /b/ exist within these languages, but are distinct in the way that they make use of VOT. That is, in aspirating languages, /p/ typically has a long-lag VOT (a longer duration in milliseconds) than the true-voicing languages' /p/. In the latter case, /p/ also has positive VOT, but typically with a much shorter duration. The phoneme /b/, on the other hand, is often realized with a short-lag positive VOT in aspirating languages, in which it likely has acoustic/phonetic overlap with the true-voicing language /p/. In other words, true voicing languages, like Spanish and French, use lead (phonetically voiced) and short-lag (phonetically voiceless) VOT to distinguish stop consonants, where aspirating languages (English and German) use short-lag and long-lag VOT. Overall, these cross-linguistic differences in VOT provide a useful continuous measure that can provide insights into cross-linguistic influence.

1.2.2 Vocalic systems

Another useful language feature to examine cross-linguistic influence is the vowel inventory of each language. Studies examining the articulatory space of vowels typically involve spectral measurements such as formant frequencies. In these studies, the frequency of the at least first and second formants are measured and compared for each vowel. The first formant (F1) is an acoustic correlate for vowel height, with higher values corresponding to lower vowels. The second formant (F2) represents vowel frontness, with higher values corresponding to a more fronted vowel. The chosen languages also vary in the size of their vowel inventories. The Spanish vowel space is the smallest of the present study and consists of 5 distinct vowel monophthong categorizations (Bradlow, 1995). English, French and German have larger relative vowel spaces, where English has 11 monophthongs, and German has 14 (Bradlow, 1995; Jongman, Fourakis, & Sereno, 1989). French has as many as 16 phonemic vowel sounds.

Cross-linguistic differences in vowel spaces are important in part due to evidence that language specific L1 categories seem to matter during L2 learning. That is, previous research has identified patterns in learning novel vowel contrasts, where some are easier than others and this ease is driven by cross-linguistic similarity and vowel inventory size between languages. (Escudero & Vasiliev, 2011; Leussen & Escudero, 2015).

The main question of the present dissertation is how cross-linguistic acoustic similarity and language status effects interact in L3 acquisition. In other words, it will be examined whether L1 and L2 phonological systems have equal impact on the L3 in perception and production in that they are equally affected by cross-linguistic acoustic similarity, regardless of the most similar acoustic sound to an L3 sound is an L1 or L2 sound. Alternatively, it is possible that a privileged status of either the L1 or L2 could cause perception and production errors, in which L3 learners assimilate L3 sounds to a sound that is not the best acoustic match.

Chapter 2: Literature Review

2.1 Models of third language acquisition

Research in third language acquisition has attempted to model the interplay between L1 and L2 language systems and their cumulative influence in the process of the acquisition of a third. Among questions asked by third language models is whether the L1 or L2, or a combination of both languages, serves as the basis in L3 acquisition. This question is complicated in the context of multilingualism due to the widespread diversity in bilingual populations that include wide variation in ultimate attainment in adult L2 learners, and, in the case of phonological acquisition, wide variation in the production patterns of L2 segments. To date, the models of L3 acquisition have largely focused on morphosyntax.

2.1.1 The Cumulative Enhancement Model

The Cumulative Enhancement Model (CEM) (Flynn, Foley, & Vinnitskaya, 2004) was one of the first formal models of L3 acquisition in adulthood. The authors argued that L3 acquisition can provide unique insights about language learning that is not possible in L1 or L2 research alone. The key question in the founding of the CEM was whether the properties of the L1 maintain a privileged status in L3 acquisition, or if L3 acquisition is cumulative process, in which all grammatical properties of previously known languages impact subsequent language acquisition, and, importantly, are helpful. Flynn et al. (2004) conducted a study to provide evidence for the newly proposed CEM. The study examined the production of restrictive relative clauses in adults and children who spoke L3 English (L1 Kazakh-L2 Russian). The results revealed similar performance on the experimental task by adults and children, and that the production of relative clauses was influenced by the participants' L2 Russian. This influence was facilitative, since this syntactic structure is common in Russian and English, but not Kazakh. The authors note, however, that the results of this study could also be explained by a special status for the L2, rather than the L1.

Berkes & Flynn (2012) conducted a further study to empirically test the predictions of the CEM. German L1 and Hungarian L1-German L2 speakers were tested in English (their L2 and L3). The authors found evidence that the L3 group performed better than the L2 group when the syntactic structure in question was L1-like, contrary to Flynn et al. (2004), which found facilitation when L2 and L3 elements were similar. The authors argue that the results of these two studies taken together suggest that both the L1 and the L2 may influence the L3 grammar, and that these languages' influence help to produce target-like L3 productions, when the L3 has a common feature with either the L1 or the L2. A proposed issue with this model is that it can not explain non-facilitative influence from previous languages in the L3 (Rothman, 2011). Non-facilitation refers to when L3 performance is not target like and resembles performance in the L1 or L2. Several empirical studies have found evidence for non-facilitation in L3 tasks and would underly the predictions of further L3 models.

2.1.2 The L2 Status Factor Model

In contrast to the CEM, the L2 Status Factor Model (L2SF) predicts that the L2 will influence the L3 by default. This prediction stems from the proposed cognitive similarity between the L2 and the L3. These proposed cognitive similarities stem from the Declarative-Procedural model, which posits that the grammar of late learned languages are largely subserved by the declarative memory system, whereas early learned languages are subserved by procedural memory. Procedural and declarative memory are long term memory systems which serve distinct general cognitive functions. The declarative memory is largely used to store explicitly known knowledge, such as factual information. The procedural memory, on the other hand, subserves implicit knowledge and procedures, such as riding a bike. During first language acquisition, learning is argued to be largely implicit and is seen as a procedure in this view. Alternatively, L2 learning is associated with declarative memory and explicit learning of a grammar, at least to a larger extent than L1 learning (Paradis, 2009) As a result, the L2 Status Factor does not make predictions in L3 learning for simultaneous bilinguals, since it is argued that both of their languages are procedural. In summary, the L2 Status Factor predicts that, due to the cognitive similarity between a late-learned L2 and L3, the L2 will influence the L3 by default. Likewise, unlike the CEM, this influence will not always be facilitative, and is predicted to block access to the L1.

Several studies have found L2 influence in L3 tasks. In an early study Bardel & Falk (2007) examined two sets of participants learning Swedish as an L3 (n = 5), and either Dutch or Swedish as the L3 (n = 4). The participants varied in their L1 and L2, and sometimes spoke three languages (which the authors classified as 2 L2s). This design was chosen to vary the order of acquisition of V2 and non-V2 languages, which differ in their placement of negation. An inspection of individual language background by the participants in Bardel & Falk (2007) (p 471-472) reveals that 4 of the participants spoke a non-V2 L1 and a V2 L2, and the remaining 5 spoke a V2 L1 and non-V2 L2. The researchers were investigating whether cross-linguistic similarity (the presence or absence of V2 in the L1 or L2) affected the production of V2 in the L3. The results provided evidence that that the group with the V2 L2 outperformed the non-V2 L2 group. The authors took this result as evidence for a privileged status of the L2 in L3 production, despite its lack of facilitation. These results contradict the predictions of the CEM, which suggested that L3 learning is a cumulative process that takes advantage both L1 and L2 grammars in L3 learning.

The L2 Status Factor model has seen further empirical support in studies examining L3 syntax. Bayona (2009) examined the acquisition of middle and impersonal passive constructions in L2 and L3 Spanish. The L2 group consisted on mostly English L1 speakers, and the L3 group consisted of English L1 and French L2 speakers. The author found that the L3 Spanish group was more accurate in rejection or acceptance of target-like use of middle and passive constructions in L3 Spanish than the L2 group. These results suggest that the L2 influenced that L3, and provided evidence for both the L2 Status Factor and the Cumulative Enhancement Model. In another early study, Leung (2005) investigated the acquisition of determiner phrases and adjective word order in French as an L2 (L1 Vietnamese) and L3 (L1 Cantonese-L2 English). Using a battery of tests, the authors concluded that the L3 group experienced influence from their L2 in both determiner phrases, which was facilitative in this case, and adjective order, which resulted in non-target like pre-nominal French adjectives. In this case, the results could arguably be accounted for by the L2 Status Factor, while the CEM cannot explain non-facilitation. However, the author's narrative conclusion of these results did not claim that there was sole access to the L2, or that the L1 was blocked. Rather, they suggest that L3 learning is not simply another case of L2 learning, in which the L1 is the sole influence.

This result was replicated in other studies. For instance, Bohnacker (2006) found that L1 Swedish-L2 English-L3 German speakers showed evidence of English like word order in L3 German V2 constructions, rather than facilitative L1 Swedish, Additionally, Falk & Bardel (2011) found that 44 L3 German learners with L1 English-L2 French or L1 French-L2 English behaved differently in tasks involving object pronoun placement. This distinct behavior was likened to L2 influence and suggested that order of acquisition was more deterministic in predicting CLI than typology. Interestingly, later studies found that L1 access was possible when learners had higher metalinguistic awareness of cross-linguistic similarity between the L1 and the L3. In a study of L3 learners of a V2 language who spoke a V2 L1 and non-V2 L2, Falk, Lindqvist, & Bardel (2015), participants were given a survey which measured their awareness of cross-linguistic features in their known languages. The results revealed that participants who scored higher on the metalinguistic survey also showed evidence of facilitative influence of their L1. This outcome led the authors to revise the predictions of the L2SF to include the potential for L1 influence when metalinguistic awareness is high.

2.1.3 The Typological Primacy Model

Later studies in L3 morphosyntax would first aim to demonstrate that influence of previously known languages in L3A is not always facilitative, and would then suggest an alternative explanation and methodology for the results found in previous studies supporting L2 influence in L3 acquisition. In one such study, Rothman & Cabrelli

Amaro (2010) aimed to demonstrate that the influence of previously known languages is not always helpful, as the CEM predicts, by examining null subjects in L3 learners. The study included two groups of Spanish-English bilinguals who were learning either L3 French or L3 Italian. Importantly, all participants spoke L2 Spanish, which allows for null subjects, and spoke either L3 French (obligatory overt subjects) or L3 Italian (optional null subjects). As a result, dropping a subject in L3 French would result in non-target like production, where the opposite would be true in L3 Italian. The authors found that the data from this study provided evidence that the L2 was the influence on L3 production in both L3 groups, and provided evidence against the CEM, since the French group experimented non-facilitative effects from a previously learned languages, and the L2 Status Factor, since neither group showed evidence of L1 influence. Despite these results, the authors posited that it was possible that the perception of cross-linguistic structural similarity (Kellerman, 1983) could also be a potential explanation for the results obtained in (Rothman & Cabrelli Amaro, 2010).

Eventually, this perspective was formalized as The Typological Primacy Model (Rothman, 2010, 2011, 2013, 2015) (TPM). TPM, like the L2SF, predicts that a single language will influence the L3, but differs in that this language may be either the L1 or the L2, and that which language transfers to the L3 is determined by psychotypology after exposure to input. Rothman (2015) explains that L3 input is parsed in a hierarchical manners in which the lexicon is the primary cue parsed to determine cross-linguistic similarity between the L3 and the languages known by the L3 learner, followed by phonological or phonotactic cues, functional morphology and finally by syntactic structure. Importantly, the transfer of one grammar holistically is predicted to occur during the initial stages of acquisition. This idea of initial stages is described by Rothman as "very early in the L3 process" Rothman (2015, p. 180). At what the TPM refers to the "absolute initial state", it suggests that both languages are in principle available to the L3 learner, but does not elaborate or make predictions as to how features of the L1 and L2 will affect absolute initial state perception and production. The present dissertation refers to the "absolute initial state" as "first exposure" to avoid potential misunderstandings regarding the amount of exposure that L3 learners have prior to the experimental tasks done in the present dissertation. As a result, the TPM does not make clear predictions in the context of the present dissertation, since the learners examined here are intended to perceive and produce the L3 at their very first exposure. Nonetheless, the TPM would best be able to account for results in which L3 speakers produce and perceive the L3 in a practically equivalent manner to their L1 or L2 in a cross-sectional design, rather than solely the L2.

Several published studies include conclusions in which the TPM is supported. In the discussion of Rothman & Cabrelli Amaro (2010), the authors suggested that it could not be determined whether L3 performance was due to L2 effects or psychotypology. As a result, the authors suggested the use of language combinations in which the order of acquisition of L1 and L2 are reversed. These groups, sometimes referred to as mirror-image groups, would include groups such as L1 Spanish-L2 English and L1 English-L2 Spanish speakers who learned the same L3. In this view, between-group performance should be similar if transfer is driven by psychotypology, and groups should behave differently if L2 status is deterministic. For example, if mirror-image groups had been used in Rothman & Cabrelli Amaro (2010) with L1 Spanish-L2 English and L1 English-L2 Spanish groups, the acceptance or rejection of L3 null subjects would allow for the examination of the predictive power of L2 status and psychotypology relative to one another. In a future study using this methodology, Rothman (2011) aimed to investigate whether L3 Brazilian Portuguese would be influenced by Spanish both when it was and was not the L2. The study examined adjective placement and meaning in the L3. The first group consisted of L1 Italian-L2 English speakers learning L3 Spanish, while the second group spoke L1 English and L2 Spanish and was learning L3 Brazilian Portuguese (BP). Based on the absence of a main effect for group in a one-way ANOVA, Rothman (2011) concluded that there was similar performance between groups and used this as evidence for the basis of the TPM. These results suggested that both groups performed in a target-like manner on the experimental task and also supported the Cumulative Enhancement Model.

Following the formal introduction of the TPM, further studies included mirror-image groups in their design and aimed to test the predictions of the model relative to the L2 Status Factor. Several of these studies involve the acquisition of L3 Brazilian Portuguese (BP) by Spanish-English bilinguals in both orders of acquisition. Montrul, Dias, & Santos (2011) examined the production of clitic and object expressions in L3 BP by mirror-image groups of Spanish-English bilinguals. Using an oral production task and a written Acceptability Judgment Task, (AJT), it was found that L3 BP was influenced by Spanish whether it was the L1 or the L2 of the participants. Examining the same language combination and using similar mirror-image groups, Giancaspro, Halloran, & Iverson (2015) also found investigated the use of differential object marking (DOM) in L3 BP. The results suggested that Spanish influenced L3 BP whether it was the L1 or the L2. Like Montrul et al. (2011), Parma (2017) also investigated L3 BP clitic development and expanded upon previous studies by including both a perception and production task. Another novelty of Parma (2017) was the inclusion of an online measure, a self-paced reading task, to measure comprehension, while the study also included a story-telling task to measure production. The results of the comprehension task did not find a reliable difference between experimental conditions. The production task, on the other hand, found evidence of errors in clitic production in L3 BP by both L1 and L2 Spanish speakers that could be likened to Spanish influence. The results of these studies provided evidence contrary to the predictions of the L2 Status Factor and suggest that Spanish influences L3 BP.

Following these initial results in the language combination of L3 BP with English and Spanish, other studies aimed to test whether either the L1 or the L2 could influence performance on L3 tasks. The evidence outside of the BP/Spanish/English triad has been more limited and controversial. Some accounts in favor of the TPM's predictive power have aimed to accumulate evidence in its favor. For instance, in a recent systematic review of 92 studies in L3 acquisition, Puig-Mayenco, González Alonso, & Rothman (2020) concluded that, in support of the TPM, that either the

L1 or L2 influenced L3 performance in 59 out of 92 studies. On the other hand, 29 of the studies out of 92 found that the L2 influenced the L3. Importantly, the findings were not coded in a mutually exclusive manner, since the authors coded 25 total studies as being explained both by L2 status and typology, meaning that the results of these studies reported that the L2 transferred to the L3, but could not rule out the possibility that psycho-typological transfer could also explain the results, since the studies did not use mirror image groups (i.e., L3 groups with the same languages, but the opposite order of acquisition). A closer examination of this systematic review reveals, however, that the coding procedure did not follow a clear objective criterion. That is, there are instances in which the coding provided by the authors in the appendix contradicts the narrative conclusions of the cited studies, and it is unclear how the authors of the systematic review determined when to depart from the conclusions of authors of studies and when they decided not to.

Limitations of the TPM in addition to the lack of robust and cross-linguistic support include overstated predictive power and over-interpretation of results. The potential problem when it comes to the predictive power of the TPM is the vagueness associated with the term “initial stages” of L3 acquisition. Puig-Mayenco et al. (2020) describes this as some 20-25 hours of instruction, before which time access to either language is possible. This vague criterion makes it quite difficult to derive the predictions of the TPM in the case of learners who first encounter an L3. The problem of over-interpretation can be seen in the use of single language properties and inappropriate statistical tests to justify narrative conclusions. For example, the TPM is rather explicit in its prediction that languages transfer holistically to the L3, rather than on a property-by-property basis that the CEM would suggest. While, some of the evidence gathered to date provides counter-evidence to sole L2 influence on the L3, it is unclear how these studies support holistic transfer. In other words, it does not seem that the body of evidence can rule out the influence of both languages within the property examined, since small samples which provide inconclusive results (non-significant p-values) do not entail equivalence within or between subjects. Such small samples and choices of methods do not allow for potential gradient and small effects of co-activation to be observed. This is due to the idea that, on one hand, it is not possible to determine whether group results have to do with sampling issues, and, on the other, clear criteria for co-activation of both languages known by a bilingual was not explicitly included in predictions and design of these studies. As a result, in the case of the L3 BP studies, it cannot be determined based on the results of studies to date that Spanish is solely activated and that English is not activated. Telling this would require much larger samples and clear criteria for interpreting how results support models prior to data collection, which could include potential for non-binary interpretation of data.

2.1.4 The Linguistic Proximity Model

The prediction that a sole language will influence L3 production and perception is not shared by all models. More recent models of L3 acquisition, like the CEM, predict that both the L1 and the L2 are available to influence an L3. For example, the Linguistic Proximity Model (LPM) predicts that there is full transfer potential (FTP) of either linguistic system, but that this occurs in a gradient fashion and on a property-by-property basis (Westergaard, 2021; Westergaard, Mitrofanova, Mykhaylyk, & Rodina, 2017). In the founding article of the LPM, Westergaard et al. (2017) posed several research questions. In addition to investigating whether a sole language influences an L3, they also examined whether CLI comes from the more typologically similar language, and whether this influence is facilitative. To provide evidence that could aid in answering these questions, the authors recruited 22 Norwegian-Russian simultaneous bilinguals who spoke English as an L3. The participants completed a binary Grammaticality Judgment Task related to verb movement. The results of the tasks indicated that, compared to L1 Norwegian speaking children, bilingual children were able to benefit from their knowledge of Russian when learning L3 English. At the same time, the bilingual participants did not perform as well as L1 Russian children learning English. The authors interpreted these results, taken together, as influence from both the L1 and the L2 in the L3 English of the Russian-Norwegian bilingual children. That is, the intermediate score of the Russian-Norwegian simultaneous bilingual children learning L3 English relative to the comparison groups was taken as evidence of the co-occurrence of facilitative influence from their Russian and non-facilitative influence from their Norwegian.

The LPM, like the TPM and CEM, predicts that similarities between languages plays a major role in L3 acquisition, rather than order of acquisition, as the L2 Status Factor predicts. The LPM departs from the TPM in that it predicts that abstract structural properties causes CLI, rather than the general perceived typological proximity. In other words, it suggests that cross-linguistic influence is decided on a feature-by-feature basis, rather than generalizing whole language predictions. Additionally, the LPM predicts that all languages are available to the L3 learner throughout the learning process, unlike the TPM.

A methodological consideration advocated for by the LPM is the use of subtractive groups. Unlike mirror image groups, which seek to compare two trilingual groups, subtractive groups compare L3 learners to L2 learners. In the case of the studies to date, L3 learners of English who speak Russian and Norwegian were compared to Russian L1/English L2 and Norwegian L1/English L2 groups. It has been argued that this design allows for gradient effects of both the L1 and L2 to be observed. That is, intermediate values in L3 performance on experimental tasks relative to the L2 performance of the comparison groups is taken as evidence of co-activation of both languages. On the other hand, if no difference could be found between L2 and L3 learners of the same language, the LPM would consider this evidence for the influence of a single language on the L3.

Following the introduction of the model, several studies have tested the predictions of the LPM. In order to test whether bilinguals have access to two languages during L3 learning, Mitrofanova & Westergaard (2018) conducted a study on Norwegian-Russian bilinguals and Norwegian monolinguals in which they taught them some of an artificial language which was designed to contain Norwegian-like lexical items, but also Russian like case marking to distinguish subjects from objects. The results revealed that the bilingual group performed better than the monolingual group on a sentence/picture matching task, in which the monolinguals relied solely on word order when making decisions. These results suggested that the bilingual group was able to make use of both of their languages at first exposure. In a series of studies, Stadt, Hulk, & Sleeman (2016) examined L1 Dutch - L2 English - L3 French speakers in both English immersion and general Dutch-dominant educational settings. In the first study, the results of a Grammaticality Judgment Task indicated that third year the immersion group showed signs of heavy L2 influence in the L3, where the regular group showed evidence of CLI from both the L1 and the L2.

In a follow up study, the authors recruited 4th year students from the same educational background to investigate whether L2 influence would differ between 3rd and 4th year students (Stadt, Hulk, & Sleeman, 2018b). The results revealed that both 3rd and 4th year immersion groups behaved similarly and showed evidence of primarily English influence on their L3 judgments, though evidence for L1 influence was also found. In a second follow up study, the authors gave similar materials to the an L3 German group (Stadt, Hulk, & Sleeman, 2020). In this case, L3 German judgments were concluded to have been influenced by the more typologically similar L1 Dutch, rather than L2 English as found in the previous cases. Taken together, the results of these studies suggest that both typology and L2 Status may play a role in modeling the patterns of cross linguistic influence in an L3. Finally, these authors also conducted longitudinal study (Stadt, Hulk, & Sleeman, 2018a), again with L1 Dutch-L2 English-L3 French speakers. In this study, they found that the L1 was more influential in the beginning stages of L3 development, but that the L1 also maintained influence as L3 proficiency increased. In a more recent study, Jensen et al. (2021) investigated similar groups to Westergaard et al. (2017), in which simultaneous Russian-Norwegian bilinguals were compared to subtractive L1 Russian-L2 English and L1 Norwegian-L2 English groups, but included seven total linguistic properties. The results indicated that the L3 group experienced facilitative influence in some cases, but non-facilitative influence in others. Additionally, the sources of this influence could be likened to either Russian or Norwegian. The results of studies, taken together, suggest that both languages are active in L3 learning, and it also appears that they remain active throughout L3 development.

Like other L3 models, the LPM has received criticism. Scholars have argued that the LPM makes vague predictions that create a problem in modeling L3 transfer acquisition (Bardel & Falk, 2021; Wrembel, 2021), since it is unclear when transfer of a particular structure occurs and when it does not. Likewise researchers have argued that, unlike the TPM and the L2SF, the LPM is not easily falsifiable (Bardel & Falk, 2021). However, Westergaard (2021) argue that the LPM does make specific predictions, since it is predicted that accuracy should fall between L1 and L2 values on experimental tasks. The authors state that, when using subtractive groups, if the L3 group performance falls above or below the one of the two L2 groups, then then model would be falsified.

Additionally, it is notable that the empirical studies that serve as the basis of the LPM (Westergaard, 2021; Westergaard et al., 2017) utilize groups of simultaneous bilinguals, which do not allow for the examination of potential L2 status effects. Following this idea, Westergaard et al. (2017) suggest that an optimal design to examine both the individual contributions of languages, as well as potential language status effects, would be the so-called fully combined design. This suggestion entails the use of both mirror-image groups, as seen in the TPM studies (e.g. L1 Spanish-L2 English-L3 BP and L1 English-L2 Spanish-L3 BP), and subtractive groups (e.g. L1 Spanish-L2 English-L3 BP, L1 Spanish-L2 BP and L1 English-L2 BP) used in the LPM studies. As a result, a fully combined design would result in the use of 6 total groups, in which a mirror image design contains L1 speakers of both background languages who learn the L3 as an L2.

2.1.5 The Scalpel Model

The Scalpel Model (Slabakova, 2017) is an additional model of L3A which overlaps in many of its predictions with the LPM. Like the LPM, the Scalpel Model rejects the idea of wholesale transfer of either L1 or L2 proposed by the TPM. Rather, it suggests that, given L3 input, the L1 and L2 combined grammar should successfully extract facilitative features from the input. In this view, non-facilitation is driven by misleading, processing complexity, and frequency of a construction. That is, more complex constructions and less frequent constructions are predicted to be associated with non-facilitative CLI.

The major divergence of the Scalpel Model from other L3 modes lies in its focus on the role of input in driving non-facilitation. Evidence for this claim comes from 2 empirical studies. Both studies examined three groups of participants L1 Basque-L2 Spanish-L3 English, L1 Spanish-L2 Basque-L3 English, and L1 Spanish-L2 English (García Mayo & Slabakova, 2015; Slabakova & García Mayo, 2015). García Mayo & Slabakova (2015) tested the use of null objects, which are not allowed in English, they are in Basque and sometimes in Spanish. Using speaker judgments, the results showed that null objects were often correctly rejected in L3 English. The second study, examining the same groups, focused on topicalization (Slabakova & García Mayo, 2015). In this case, L3 English speakers did not show evidence of target like behavior on grammaticality judgment tasks. The results of these studies suggest that one feature, topicalization, is difficult to acquire, where null objects was easier to acquire in L3 English. The author suggests that this difficulty could be explained the likely lower frequency of topicalization in L3 English input relative to the lack of null objects. Overall, the predictions of the Scalpel Model at first exposure might suggest that facilitation would occur in the event that the L3 stimulus has a clear bias to the L1 or the L2. That is, the Scalpel Model and the CEM should share similar predictions for the behavior of L3 learners at first exposure, but differ as the L3 develops. The Scalpel Model does not seem to predict or concur with the prediction of the LPM that L3 performance on experimental task should fall between L1 and L2 performance.

2.2 Models of L2 phonological acquisition

Notably, much of the evidence for the L3 models to date come from studies which examine the acquisition of (morpho)syntactic features. The models do not always spell out specific predictions when it comes to L3 phonology. In the case of the TPM, full transfer could be taken to mean the transfer of an entire language system to the L3, including phonology, but this stipulation is not explicitly spelled out in the recent articles articulating the predictions and motivations of the TPM (Rothman, 2010, 2011, 2013, 2015). In the LPM, phonology is also not directly addressed, but there does not seem to be any reason why the same predictions should not apply to phonology as syntax. Specifically, the LPM should predict that L3 performance should be intermediate and fall between L1 and L2 performance on experimental tasks. Though a model of L3 phonological acquisition has yet to be proposed, the expansion of L2 models of phonological acquisition has been proposed (Wrembel, Marecka, & Kopečková, 2019).

Many accounts for L2 phonological acquisition exist. In general, these models all have in common that language-specific L1 categories drive L2 speech development. It is still not well known how the predictions of these models apply to L3 speech development when L2 categories are also theoretically available to the L3. In the section that follows, a brief overview of each model, along with its evidence, will be covered. Following the model introductions, its relationship and proposed expansion to L3 phonology will be discussed.

2.2.1 The Speech Learning Model

The original SLM (J. Flege, 1995) focuses on the acquisition of segments in the L2 and emphasizes the importance of cross-linguistic similarity in L2 speech learning. In this view, new segments are predicted to be easier to learn, where segments which have close matches in the L2 will be much harder to acquire at a native-like level. Essentially, the SLM proposes that L1 and L2 sounds are linked through a process referred to as “interlingual identification”, in that the L2 sound encountered as first exposure are seen as either good or bad phonetic variants of a native category (J. Flege, 1995). In this view, acquisition of L2 segments occurs after increased exposure to these segments, where segmental complexity modulates the rate of learning (J. Flege, Aoyama, & Bohn, 2021). Additionally, the SLM posits that allophonic variation and segment position matter, as opposed to the learning of a phoneme generalizing across positions. For instance, Iverson, Hazan, & Bannister (2005) conducted a study in which they successfully trained Japanese speakers to better identify /r/ and /l/ in word initial position, but that this training was not effective for word-medial or word-initial clusters containing liquids.

The original SLM also suggested that the age of first exposure (the younger the better), and L2 experience were important predictors to L2 learning success. The SLM makes a distinction between identification as opposed to categorization. The importance of this distinction involves the learning of phonetic variants during the learning process, which lead to new L2 phonetic category formations. For example, using a two-alternative forced choice task, Bohn & Flege (1992) asked Spanish monolinguals to identify English stops and provided them with pre-voiced, short lag and long lag stops as tokens. Importantly, Spanish monolinguals identified English stops as long lag /t/ as /t/, despite the lack of a long-lag category in their native inventory.

2.2.1.1 The revised SLM

The SLM was recently revised, and has updated some assumptions and predictions. The revised model maintained many of the same assumptions as the original, and importantly focuses on sequential bilinguals, where speech learning begins when the phonetic categories of the L1 have been established. According to J. Flege et al. (2021), the updated tenets of the Revised Speech Learning Model (SLM-r) are that, first, phonetic categories are formed based on their statistical regularities in the input, such that greater exposure to a particular phonetic cue should be correlated with the re-tuning of that cue towards the input. Second, all learners make use of the same learning mechanisms in L2 learning that they do in native language learning. That is, the SLM-r refutes the idea of a critical period for speech learning, and suggests that non-native production and perception is due to differences the quality and quantity of input between native speakers and L2 learners and L1 effects. Another important update to the SLM-r is the idea that perception and production co-evolve. In other words, the SLM-r does not predict that perception will precede production.

The revised model has not yet received empirical support, due to its recent revision. It is outside the scope of the present dissertation to evaluate the predictions of the SLM-r when it comes to statistically driven re-tuning of phonetic categories. However, the phonetic categories that are in place in a bilingual at the first exposure to L3 learning could be treated as an integrated inventory when it comes to L3 learning. Based on the idea that the same mechanisms are in place in L1 and L2 phonetic category formation, the same mechanisms should be in place for L3 phonetic category formation. In this view, the acoustic similarity of an L3 segment to an L1 or L2 segment should predict how difficult learning an L3 segment will be. In this view, there is not reason to predict a blocking of one language system (L1 or L2) provided that these systems have well-established phonetic categories.

2.2.2 The Perceptual Assimilation Model

Relative to the SLM, the Perceptual Assimilation Model (SLM) focuses on the perception of sound contrasts by L2 learners (Best & Tyler, 2007). The PAM involves various scenarios in which the cross-linguistic inventories of specific languages predict how easily L2 sounds will be to acquire. For instance, if a L2 learner's native language contains a contrast that also exists in the L2, then this contrast will be easily perceived in the L2 (two-category assimilation). In the event that two native language categories correspond the a single L2 category (single category assimilation), discrimination is predicted to be intermediate. In the event that two non-native sounds are equally good exemplars on an L1 category (uncategorized assimilation) discrimination is predicted to be poorer.

Evidence for the predictions of the PAM has been found in studies which involve the presentation of sounds to naive learners of an L2 (first exposure), or to more experienced L2 learners. Experiments typically present the subjects with native-language vowel categories in written form and auditory stimuli of L2 vowel sounds and participants are tasked with choosing the closet matching native-language vowel category given the options, and to rate the goodness of fit of this decision.

This methodological paradigm has found evidence that two category assimilation is difficult for L2 learners. For example, Escudero & Chládková (2010) found that Spanish L1 speakers assimilated SSBE /ae/ and /a/ to Spanish /a/ (that discrimination of this contrast was difficult). Additionally, Escudero, Sisinni, & Grimaldi (2014) provided evidence that Salento Italian L1 speakers assimilated SSBE /ae/, /a/ and /ˆ/ to Salento Italian /a/, while Escudero & Vasiliev (2011) Spanish speakers assimilated Canadian English /ae/ and /e/ to Spanish /a/. Finally, Escudero & Williams (2011) found that Spanish listeners assimilated the two categories in Dutch /a/ and /a:/ to their single Spanish category /a/.

The present dissertation tests whether assimilation acquired during second language acquisition apply to third language sound perception. In other words, it tests whether L3 perception mirrors L2 perception in similar ways that has been observed in L3 production. Specifically, it tests whether L3 speakers categorize L3 sounds similarly to L2 sounds, and whether phonetic discrimination of sounds that would be phonemic in the L1 is accessible in the perception of L3 words at first exposure, or whether there is an initial blocking or the L1 or L2 bias effect.

2.2.3 The Second Language Linguistic Perception Model

The Second Language Linguistic Perception Model (L2LP) is a computational model of L2 speech learning and is similar to the PAM in that it focuses on sound contrasts as the basis for L2 speech learning, rather than single segments (Leussen & Escudero, 2015). In this view, it is difficult for L2 learners to make contrasts which are not present in their L1. In their revision of the L2LP, van Leussen & Escudero support this claim with several empirical studies. Namely, that these studies have provided evidence that L2 learners experience difficulty with the contrasts of /r/ and /l/ in Japanese, (Aoyama, Flege, Guion, Akahane-Yamada, & Yamada, 2004) “beat” and “bit” in Spanish and Portuguese (Jim. Flege, Frieda, & Nozawa, 1997) “bet” and “bat” in Dutch (Broersma, 2005). The authors argue that these results suggest that linguistic experience is at the heart of L2 learning. Specifically, cross-linguistic comparisons of L1 categories and L2 categories are thought to predict the ease of L2 category learning. The L2LP, unlike the PAM and the SLM, also aims to model the entire learning process, rather than the beginning stages of L2 speech learning. This entire learning process is predicted computationally and is based on Stochastic Optimality Theory (Boersma, 1998).

An important postulate of the L2LP is the optimal perception hypothesis. Essentially, this proposal suggests that the initial perception of L2 sounds is the result of L1 acquisition. The developmental of L2 learners is predicted similarly to PAM, where a single category assimilation from PAM, in which a native category must be split, is called a new category scenario in the L2LP. This scenario is predicted by both models to be difficult for the L2 learner. On the other hand, the PAM and the L2LP predict that the case when two L1 sounds correspond well to two L2 sounds that discrimination of these sounds will be relatively easier. This is referred to as a single category assimilation in PAM and a similar scenario in the L2LP. The final scenario of the L2LP is the subset scenario. In this case, a single L2 phoneme is perceived as two L1 categories. The same case is referred to as uncategorized or categorized-uncategorized in the PAM, and both models predict that discrimination of these sounds will be better than the case of new scenarios, but not as good as discrimination of subset scenarios.

Empirical support has been found for these predictions. In two-category assimilation/similar scenario, Escudero & Boersma (2004), native Dutch speakers assimilated the Spanish /i/ to their native Dutch /i/, and the Spanish /e/ to their Dutch /I/. However, the L2LP predicts that the similar (two-category assimilation) scenario may lead to inappropriate lexical contrasts, and argues that pre-lexical and lexical contrasts should be taken into account when it comes to sound discrimination. Evidence for this claim stems from studies which found that L2 learners could not perceive a contrast in lexical items that they could discriminate outside of lexical items (Curtin, Goad, & Pater, 1998). Other studies show that some lexical items can also be reliably distinguished by L2 learners that could not be told apart pre-lexically (Cutler, Weber, & Otake, 2006; Weber & Cutler, 2004). As a result, an important tenet of the revised L2LP (Leussen & Escudero, 2015) is that meaning-driven learning predicts the developmental path of L2 phoneme perception. The L2LP simulates the entire trajectory of L2 learning based on the various learning scenarios. In the revised L2LP, Leussen & Escudero (2015) suggest that category reduction is possible when it is driven by meaning based learning.

The predictions of the L2LP which are important for the present dissertation are the Full Copying hypothesis. If L3 phonological learning is another instance of L2 learning, then the L2LP and the TPM should share the prediction that the initial state of L3 learning is the end state of either L1 or L2 learning, but not both. In this view, bilingual participants who are first exposed to an L3 should produce and categorize L3 sounds as similarly to a single language, rather than producing or perceiving some sounds as L1-like and others as L2-like. Additionally, these participants' behavior on the experimental tasks should a) resemble either their own L2 behavior or L1 behavior or b) in the case of L1 influence, resemble a monolingual comparison group who is first exposed to an L2 (e.g. Spanish L1, English L2, exposed to German, and Spanish L1 exposed to German should behave similarly if Spanish is language which is "fully copied" at first exposure.)

2.3 Previous literature in L3 phonology

In the research to date, L3 models have focused largely on empirical results stemming from experiments in morpho-syntax to derive predictions about how the L1 and L2 affect L3 acquisition. Fewer studies have been conducted in L3 phonology, and it remains unclear whether L3 models make specific predictions for phonology and morphosyntax. Despite this lack of specificity in the models when it comes to phonology, some trends have emerged in the body of research. Specifically, it appears that the L1 and L2 likely both affect L3 perception and production, with the L2 often providing a greater level of influence in the beginning stages of acquisition, even when it is not facilitative. This trend cannot readily be accounted for by the TPM, the L2SF, the CEM nor the Scalpel Model, leaving only the LPM which can predict co-activation of all languages known by a bilingual in L3 acquisition. The following sections cover the empirical studies that have been done in L3 production studies across L3 proficiency levels, and the fewer studies carried out in L3 perception, as well as their relevance to L3 models.

2.3.1 L3 Production Studies

The findings in empirical studies of L3 phonological cross-linguistic influence have varied. One of the first studies to examine progressive influence of the L1 or L2 on L3 production was the seminal case study of Williams & Hammarberg (1998). This study elicited the production of an adult L1 British English, L2 German, and L3 Swedish speaker in the L2 and L3 upon her arrival to Sweden. The speech samples were rated for native-likeness by native speakers of German and Swedish respectively, with low ratings (i.e. non-nativeness) being elaborated upon. In the event of non-native speech, the raters guessed where the speaker in the recording might be from. The informant was rated as having near native productions in her L2 German, while her L3 Swedish was rated as being non-native like and to be German-accented. The experiment was repeated after 6 months in Sweden, however, and the Swedish raters then judged the informant's Swedish to be British English accented.

This study constituted evidence of an L2 status effect in the initial stages of L3 phonological acquisition, in which the second learned language influence L3 production, but also provided evidence that this effect diminishes as L3 proficiency increases. This notion has been called the 'foreign language effect' (Meisel, 1983), which refers to the idea that speakers who learn a second non-native language are biased to sound unlike a native speaker of their native language.

The default L2 status effect has received some empirical support in the L3 phonology literature. In a study of global accent production, heavier L2 influence in L3 productions was found by L1 Polish, L2 German and L3 English speakers based on ratings of EFL instructors (Wrembel, 2010). Similar findings have also been reported in vowel production (Kamiyama, 2007) and vowel reduction and speech rhythm (Gut, 2010). L2 influence has also been found in VOT productions. Llama et al. (2010) examined L3 Spanish VOT production by French-English mirror-image bilingual groups and found that both groups had L2-like productions of the L3.

Other findings in L3 production, however, have yielded mixed results. Several studies have found that acoustic properties of the participants' productions fall between L1 and L3 values, suggesting that both the L1 and the L2 have some influence on L3 productions, rather than solely one language. For instance, (Wrembel, 2014) measured VOT and aspiration in all languages of participants with two different language combinations: L1 Polish, L2 English, and L3 French; (2) L1 Polish, L2 English, and L3 German. The results showed that each language had a specific stop-value, and that the L3 VOT productions were intermediate, falling between the L1 and L2 values. Similarly, (Wrembel, 2011) examined thirty-two learners of L3 French with L1 Polish and L2 English who were recorded reading lists of words in carrier phrases. As in previous studies (Wrembel, 2014), combined transfer from the L1 and the L2 in VOT productions was found. Findings of combined L1 and L2 influence in VOT productions were also reported by Wunder (2010) in L3 Spanish speakers, and by Blank & Zimmer (2009) in L3 English speakers who spoke L1 Brazilian Portuguese and L2 French. Other studies have found an L1 influence on production despite L3 proficiency (Wrembel, 2012), or in advanced L3 learners (Llama & Cardoso, 2018). Importantly, these studies report L3 VOT values which fall between L1 and L2 values. Following the suspicion that intermediate values might have to do with either sampling issues or proficiency effects, Parrish (2021) examined Mexican Spanish-English bilinguals who produced voiceless stop-initial French words in isolation at first exposure to the language. The results found that the relative VOT of the L3 fell between their own L1 and L2 values, in line with previous research, and that suggested that intermediate values were less likely to have been seen in previous studies as a result of small samples or proficiency effects. However, a subsequent analysis of the data suggested that wide individual variation existed, in which some participants produced L3 French as L1 Spanish like, and other produced intermediate, L2-like values. This result suggests that higher samples could reveal group trends and provide better insights into individual variation in crosslinguistic influence, as opposed to assuming that a single group trend exists.

2.3.2 L3 Perception studies

Overall, studies in L3 perception have been much more scarce relative to studies in production. These few studies have often used methods found in L2 speech acquisition research, such as studies testing the predictions of the Perceptual Assimilation Model. For instance, Wrembel et al. (2019) examined the categorization and discrimination of L3 vowels by 10 young trilinguals who spoke L1 German-L2 English-L3 Polish. To test categorization, a cross-linguistic similarity task was used in which participants heard minimal pairs of sounds and had to rate how similar sounds were on a 1-7 Likert scale. The results showed evidence that participants assimilated L3 sounds to both L1 and L2 categories, but preferred the L2. In a second experiment, and AX discrimination task was given to participants to evaluate whether retroflex and palato-alveolar silibant discrimination, a feature of Polish, could be accessed in L3 words. The results revealed that discrimination of the L1 Polish contrast was very good (84% accuracy), suggesting that L3 learners retain access to L1 sound contrasts in L3 words. Additionally, this language specific phonetic discrimination was attended to by even L3 beginners. Balas (2018) also used the PAM as a perceptual framework to work in and adapted it to L3 learners. The study recruited three groups of Polish L1 speakers, including two L3 groups (L1 Polish-L2 English-L3 Dutch and L1 Polish-L2 English-L3 Dutch). The third group spoke only English as an L2. All three groups were listened to Dutch vowels and were asked to categorize them given Polish vowel categories. Unfortunately, the L3 groups were not given L2 English categories as options during this task, so the results of this study cannot directly provide evidence that L3 learners categorize L3 sounds using both the L1 and L2. The same study also conducted an AXB discrimination task of 8 Dutch vowels and found that discrimination was at ceiling for all vowels involved.

An additional line of research in L3 perception studies has pondered whether L3 learners have a general perceptual advantage over their L2 learning counterparts. In an early study, Werker (1986) investigated mono, bi and multilingual participants' discrimination of non-native perceptual contrasts and did not find any evidence that the bi or multilingual group had any perceptual advantage. Patihis, Oh, & Mogilner (2015) examined Korean stops by naive mono and multilingual listeners and found that phoneme discrimination was feature dependent, rather than bilingualism providing a general advantage. Antoniou, Best, & Tyler (2013) investigated the contrast of Ma'di stops by early English-Greek bilinguals, English monolinguals and Greek monolinguals. The researchers wanted to know whether language model affected categorization and discrimination. They found that categorization is modulated by language mode, but discrimination of novel sounds was not. In particular, the Greek monolingual group was most successful in the discrimination of the novel contrast, where the English monolingual group was the least accurate. The bilingual group, regardless of language mode, displayed intermediate ability to discriminate the novel sounds. In a further study, Antoniou, Liang, Ettlinger, & Wong (2015) taught mono and bilingual participants an artificial language based on either English or Mandarin. The results revealed that bilingual participants outlearned the monolingual groups overall, but universally more difficult unfamiliar L3 segments do no seem to be learned more easily by bilinguals. Wrembel, Gut, Kopečková, & Balas (2020) also found that cross-linguistic influence is structure dependent and varies among individuals. Onishi (2016) reported a bilingual advantage, but, rather than feature dependent, suggested that this advantage was global. The study involved speakers of Korean who spoke L2 English and L3 Japanese. The author found a correlation between successful discrimination of L3 contrasts and L2 experience, and argued that this correlation is evidence for a bilingual advantage.

These findings suggest that crosslinguistic similarity of both a bilingual's languages play a role in the ease of acquiring a novel, L3 sound. Additionally, research focused on a bilingual advantage in phonetic learning can inform L3 models, since it implies that bilinguals have access to the phonetic categories in both languages that they speak, even at first exposure, whether or not the advantages of bilingualism on phonetic learning are global or feature dependent. Few studies have investigated the role of VOT in L3 perception. On such study is Z. Liu, Gorba, & Cebrian (2019), which examined the perceptual boundary of a VOT continuum in L1 Mandarin, L2 English and L3 Spanish. Using a /pi-bi/ continuum, the authors found that the perceptual boundaries of each language were 28ms for Chinese, 24.6ms for English and 23ms for Spanish. Despite the goal of the study being to examine regressive transfer and to compare the L3 speakers to monolinguals, descriptive evidence of L2 effects can be inferred from the reported means. Additionally, J. Liu & Lin (2021) examined perception and production of word-initial voiced and voiceless stops in speakers of L1 Mandarin, L2 English and L3 Russian or Japanese. The authors found that voiced stops were effectively perceived, but not reliably produced, where voiceless stops were both successfully perceived and produced.

2.4 Methods and anylsis in previous work

The widely varied findings in previous work do not have a clear plausible correlate, but may be related to issues related to sampling issues combined with methodological choices. In this subsection, an overview of sample sizes in previous work will be given, followed by potential issues related to small samples. Subsequently, common methods in the body of research used to analyze L3 data will be discussed, followed by their potential shortcomings. Finally, the manner in which the present dissertation addresses these issues will be covered.

Overall, it is unclear whether sufficient sample sizes have been used in the body of research to date. Of the empirical studies reported in the present dissertation, none used a power analysis or otherwise justified their sample size per group. Unfortunately, the tradition in L3 research has involved model building with small samples. The use of small samples are associated with higher sampling error, and, as a result, a higher risk of type 1 error (Brysbaert, 2020). In other words, a single study with a small sample cannot rule out the possibility that their results are due to sampling error, or a non-representative pool of participants from an assumed population distribution. As Brysbaert (2020) argues, low samples lead to low statistical power, and in turn provide a metaphorically blurred picture of our desired outcome.

In addition to issues associated with low sample size, L3 research to date has used statistical methods which provide dubious evidence for their claims. Among these issues is the interpretation of the results of various inferential statistical methods, such as the lack of a main effect in an ANOVA, as a basis for determining whether groups or individuals perform experiment tasks in a practically equivalent manner rather than a statistical Test of Equivalence (Lakens, 2017). At the heart of this issue is a criticism which may apply to frequentist methods of statistical analysis in general; testing against the null hypothesis. If the null hypothesis is rejected, evidence is provided that the difference between or within groups is non-zero. On the other hand, if a non-zero difference is not found, there is not evidence for practical equivalence. Such as assumption has been made in the L3 literature, and in particular in L3 model building.

For example, in his seminal article introducing the Typological Primacy Model, Rothman (2011) concluded that two groups of L3 learners did not perform differently on the L3 Brazilian Portuguese acceptability judgment task and took this as evidence for similar performance between the groups, and evidence of typological similarity effects in L3 judgments. There are two possible issues with this conclusion. Firstly, the lack of a power analysis does not rule out sampling error. It cannot be argued that the sample used in this study was high enough to reliably detect an effect (or lacktherof). Secondly, a wide confidence interval from the low sample size would make providing evidence for statistical equivalence in a test of equivalence likely impossible. Unfortunately, the reporting of results in many L3 studies to date do not allow for post-hoc power analyses to be run, since means and standard deviations are seldom reported. It remains unclear whether studies in L3 research are sufficiently powered to be able to generalize, and, by extension, to build models.

Despite the lack of empirical data available to evaluate the statistical power of L3 studies to date, it is probable that the sample sizes used are not statistically powered. Brysbaert (2020) argues that, for between group comparisons, samples of at least 80 per group are often necessary, but a power analysis should be carried out to justify sample size. To the present author's knowledge, no L3 study has been able to recruit this many participants per group. For instance, the seminal study motivating the TPM (Rothman, 2011) recruited just 11 and 15 in two groups. Plonsky (2015) suggests that combating issues associated with low sample size is possible in ways other than simply increasing sample size. For instance, he suggests that the use of descriptive statistics, including effect sizes and confidence intervals, would be an improvement in L2 research in general. This advice is in line with the idea that frequentist analysis, and linguistic research, has relied on p-values to determine the presence of a so-called significant statistical difference. Plonsky (2015), along with others, have argued that the use of p-value alone to make real-world inferences in problematic due to issues associated with sampling error and the presence of the magnitude of an effect.

In order to address these potential issues, the present dissertation recruited bilinguals, rather than trilinguals, at first exposure to a third language in order to pull from a likely higher and more homogeneous population of participants. This higher sample, coupled with the use of Bayesian inference allow for both a categorical and gradient interpretation of the data. In doing so, less reliance is put on a narrative interpretation of the results, and the results lend themselves to a more objective outcome.

2.4.1 Bringing L2 speech models and L3 models together and evaluating predictions

By using the methods used to test the prediction of models of L2 phonological acquisition (The SLM, the PAM, and the L2LP), more nuanced evidence may be obtained to evaluate the predictive power of L3 models. With the revision of the SLM, at least three L2 speech models advocate for the study of naive or beginning learners in L2 speech learning research. The present dissertation adopts this point of view and applies it in a third-language context in perception and production. By measuring the perception and production patterns of the first (or, at least, very early) exposure to a third language, coupled with measurements in each language, and the variation of cross-linguistic influence can be observed/studied in L3 perception and production. If, as the SLM-r predicts, category formation is driven by input and retuning of L1 categories in L2 acquisition, and the same mechanisms that are used in L1 phonetic category formation in L2 phonetic category formation, then it is reasonable to suppose that the SLM would predict that those same mechanisms are at play in L3 phonetic category formation, and that this process is guided by L3 input.

Following this logic, the question becomes whether L1 or L2 categories, or a combination of both, are retuned to L3 categories. Additionally, it is unclear what conditions determine which language category, whether L1 or L2, is initially chosen and its rate of retuning to become more target (L3) like as L3 input exposure increases. The present dissertation focuses on which category is chosen as the initial L3 category, while the rate of change in these categories is left for future research. One possibility which may influence whether an L1 or L2 category influences an L3 is the acoustic similarity of the L3 segment relative to an individuals' L1 and L2 phonetic categories. The TPM predicts that phonetic cues do play a role in L3 input parsing, in that they are parsed and used in order to make a decision of which language system to holistically transfer. However, the TPM would (likely) not predict that two segments would be assimilated to two distinct language categories. Additionally, if one language holistically impacts the acquisition of a third at first exposure, then the behavior of

L3 learners should resemble L2 learners, provided the L1 is the source of influence in L3 acquisition, or should be practically equivalent in a within-subject comparison. In the case of the predicted behavior of L2 influence on L3 productions, L2 production and perception should be practically equivalent in a within-subject comparison. The predictions of the Linguistic Proximity Model would be able to account for the same L3 sound being categorized differently by the same subject, and by different subjects, and for two different L3 sounds being categorized as an L1 and L2 category respectively.

Chapter 3: Perception of non-native vowels by Spanish-English bilinguals

3.1 Introduction

The body of research on phonetic and phonological learning has uncovered a crucial role for the proximity of the native language of a given speaker and the language they are learning. For example, the empirical basis for two major models of L2 phonology, the Speech Learning Model and the Perceptual Assimilation Model, posit key roles for the native phonological structure of a speaker when predicting the ease or difficulty of acquiring specific L2 sounds. For example, it has been shown that speakers of languages with a long-short lag VOT distinction in stop consonants have difficulty acquiring native-like VOT values in their L2 if it is a true voicing language (Nagle, 2019; Reeder, 1998).
Additionally, it is difficult for speakers to perceive differences in two distinct L2 categories that would both be good exemplars on a native category, such as Spanish speakers who have difficulty acquiring the /i/-/ɪ/ contrast (J. Flege, Bohn, & Jang, 1997). Both of these difficulties arise, presumably, due to the native phonology of the learner.

These models do not address whether, in addition to one's native phonological structure, additional (L2) categories can also impact the subsequent learning of sounds. In other words, it is not clear whether both the native and L2 phonological systems may both be possible sources of influence during new language learning. The present chapter tested whether both the L1 and L2 of a speaker are used in the very initial stage of learning a new, third language. In the first experiment, German and French vowels were categorized given both English and Spanish categories. The perceptual assimilation task was designed to determine whether bilingual speakers would categorize new language sounds using both English and Spanish, or whether would be biased to one language or the other. In experiment 2, the discrimination of English vowel contrasts in L3 words were tested. This experiment avoids categorization and measures discrimination of sounds with the aim of showing whether phonetic discrimination is possible in a new language when the sound contrast exists in either the L1 or the L2. If speakers categorized and discriminated new language sounds using both Spanish and English phonology, this would suggest that it is not simply native language phonology that predicts new language learning, but that all previously known languages may play a role in the learning of a new language sound system.

3.1.1 Perception/perceptual assimilation in L2 and L3 speech

Previous research has found that the ease of learning sounds in a new language depends upon your language background and experience (Best & Tyler, 2007; Escudero & Chládková, 2010; Escudero et al., 2014; Escudero & Vasiliev, 2011). In particular, it has been found that the phonemic and phonetic qualities of ones native language can predict the ease or difficulty of sound contrast perception in both accented L1 and L2 speech. For example, when they were asked to categorize vowel sounds of Standard Southern British English (SSBE), speakers of Salento Italian categorized three SSBE vowel sounds as their native /a/ (/æ/, /ɑ/ and /ʌ/), and also did not discriminate these three sounds efficiently (Escudero et al., 2014). This finding suggests that experience with Italian, a five vowel system with a quantity distinction, makes the discrimination of languages with larger vowel inventories more difficult than it would be for a speaker of a language with a similarly large vowel system. Other language pairings also show evidence for this phenomenon, such as Spanish listeners categorized Dutch /a/ and /a:/ as their Spanish /a/ (Escudero & Vasiliev, 2011), and Spanish speakers categorizing Canadian English /æ/ and /ɛ/ to Spanish /a/ (Escudero & Vasiliev, 2011), where speakers of languages with a smaller vowel inventory would categorize more than one vowel sound in another language as one in their native language. In other cases, speakers have matched two novel-language vowels to two of their native categories. Escudero & Chládková (2010) found that Spanish speakers assimilated English vowel sounds to their Spanish /a/ and /e/, although they assimilated British /æ/ to their Spanish /a/ and English /æ/ to their Spanish /e/, while matching /a/ in both dialects to their Spanish /a/. Taken together, these studies suggest that native language categories impact the categorization of sounds in a new language, although they did not explicitly investigate whether second language sounds may have a similar effect.

Best and Tyler formalized the idea that the ease of the perception of any new language is driven by the phonemic and phonetic differences between the inventories of the speaker's native language and an unknown language. Originally proposed to account for non-native (foreign accented) speech perception, the Perceptual Assimilation Model (*PAM*; Best & Tyler (2007)) has also been extended to L2 learning (PAM-L2). The model predicts that when languages (the one known and the one being learned) are more similar in terms of how sounds contrast, then they will be easier to effectively perceive. For instance, in what the PAM calls Two-Category Assimilation, two L1 categories would be hypothetically matched to two L2 categories, and discrimination of that L2 contrast would be predicted to be very good. On the other hand, in the event that one L1 category corresponds to two L2 categories, discrimination is predicted to be poor (Single-category Assimilation).

Despite the general empirical support for the model, the PAM has not explicitly discussed whether bilingual language learners approach learning new-language sounds differently than monolinguals. In their 2007 chapter, Best and Tyler state that "the relative ease or difficulty of a given contrast varies according to the listener's native language" (p. 16), but make no mention of bilingual listeners. It is not clear whether, if given the option, speakers could be guided by both their native and subsequently learned languages when asked to categorize new language sounds. Very little research exists testing whether this is the case. Wrembel et al. (2019) examined the categorization of L3 German vowels by 10 multilingual L1 Polish, L2 English, L3 German speakers, given L1 and L2 categories from which they could choose. They found that the speakers assimilated L3 sounds to both their L1 and L2, with a preference for L2 sounds, suggesting that subsequently learned languages might guide subsequent language perception in addition to the native language. Onishi (2016) examined the impact of L2 experience on L3 contrast perception ability in speakers of L1 Korean, L2 English and L3 Japanese and found that better experience with specific L2 contrasts correlated to better perception of those same contrasts in the L3 contrasts, but only in some cases. These studies provide evidence that speakers may draw upon more than simply their native language when accounting for new language sounds. However, since the @Wrembel et al. (2019) and Onishi (2016) both examined the perceptual assimilation of L3 sounds to L1 and L2 sounds, it is more difficult to disentangle whether their assimilation choices might be a result of L3 acquisition or previous language influence. For example, the target like contrast performance by the L3 Japanese speakers, while correlated to L2 performance contrast, does not provide evidence that these speakers began their development with an advantage due to their L2 knowledge.

The present chapter adds to the scarce work done in L3 perception by recruiting naive bilingual speakers, rather than actual multilingual speakers. This approach offers valuable information about the absolute starting point in L3 phonological acquisition by investigating whether novel language sounds are assimilated to both L1 and L2 categories when learning begins, rather than later in development. This distinction is important in the context of L3 models, which debate whether speakers maintain access to both their known languages throughout development.

3.1.2 Overview of experiments

The goal of the experiments was to examine how participants categorized and discriminated vowel sounds in a new language. For this reason, two experiments were created. The first experiment, designed to measure categorization, was a perceptual assimilation task. During the perceptual assimilation task, participants heard a single new-language vowel sound and their task was to match the played sound to the most similar vowel sound. The methodology resembles studies carried out in L2 phonology to test the predictions of the PAM-L2. Differently from the L2 studies, the present study included both L1 and L2 categories and tested bilinguals rather than functional monolinguals.

The second experiment was an AX discrimination task. It examined how all groups discriminated English vowel contrasts when they are in new language words.

Specifically, several seven-step vowel quality continua ranging from /i/ to /ɪ/ and /a/ to /æ/–/ɑ/ was created. During a trial, participants heard two tokens drawn from the continuum back-to-back and had to determine whether the sounds they heard were the same or different. This task allowed for the examination of perception in a purely phonetic, rather than phonemic, context, since categorization was not explicitly involved in this task.

3.2 Experiment 1 - L3 Perceptual Assimilation Task

Experiment one examined the initial state of categorization of the vowels of two unknown languages, French and German, by Spanish-English bilinguals, Spanish monolinguals and English monolinguals. Each participant completed a background questionnaire and a perceptual assimilation task.[1]

3.2.1 Participants

A total of 199 participants took part in the present study. The data of 10 total participants were removed who self-reported being bilingual yet not feeling comfortable speaking or understanding their second language. The present sample includes L1 Mexican Spanish - L2 English speakers (n = 59) and L1 American English - L2 Spanish speakers (n = 55). The monolingual participants included 29 Mexican Spanish monolinguals and 56 American English speakers. All participants were recruited on prolific and were pre-screened according to criteria detailed below. In addition to filters in place from Prolific.co, the participants were screened further using an adapted version of the Bilingual Language Profile (Birdsong et al., 2012). All participants who answered ‘no’ to the question “Do you speak a language other than English and Spanish” were permitted to continue the experiment.

---

[1]The results of Experiment 1 of this dissertation have been published in the peer-reviewed journal *Languages* https://doi.org/10.3390/languages7030226

3.2.1.1 Bilinguals

Screening data and experiment-initial questionnaires were used to find bilingual participants who began learning their L2 later in life and reported not having learned a language aside from English and Spanish. The English L1 group came from all over the United States while Spanish L1 group came from Mexico. Each groups' mean age, L2 use, self-reported oral and perceptive proficiency are seen in Figure 3.1. As can be seen from the figures, the English L1-Spanish L2 group began L2 learning later on average, while they also felt comfortable in their L2 at a later age than the Spanish L1-English L2 group. The participants also rated their L2 proficiency. They were given a 0-6 Likert-type scale in which they answered the questions "How well do you speak [their L2]?" and "How well do you understand [their L2]?". "0" corresponded to "not very well at all", where "6" corresponded to "very well" (Figure 3.2). On average, the Spanish L1 group rated themselves as more proficient in both speaking and listening than the English L1 group.

3.2.2 Materials

3.2.2.1 Target phrases/conditions

The participants heard and categorized a total of 4 vowel conditions per language given seven total carrier words to choose from. The four vowel conditions included the /i/, /y/, /o/, and /ʌ/. The four sounds were embedded in both a fricative /fVf/ and bilabial /pVp/ or /pVf/ frames and played a total of fice times each. Thus, each participant categorized 40 tokens per language (five repetitions x two frames x four vowel conditions), given seven carrier words from which they chose. The seven word choices included three English carrier word choices intended to represent the phonemes /i/, /ʌ/, /u/ and /a/ (*feel*, *fun*, *fool* and *thought*). The remaining 3 choices were Spanish carrier words intended to represent the phonemes /i/, /u/ and /o/ (*fin*, *su*, *son*). Figure 3.3 shows an example of the trial screen that the participants saw.

The vowel sounds included in both experiments were intended to bring about four

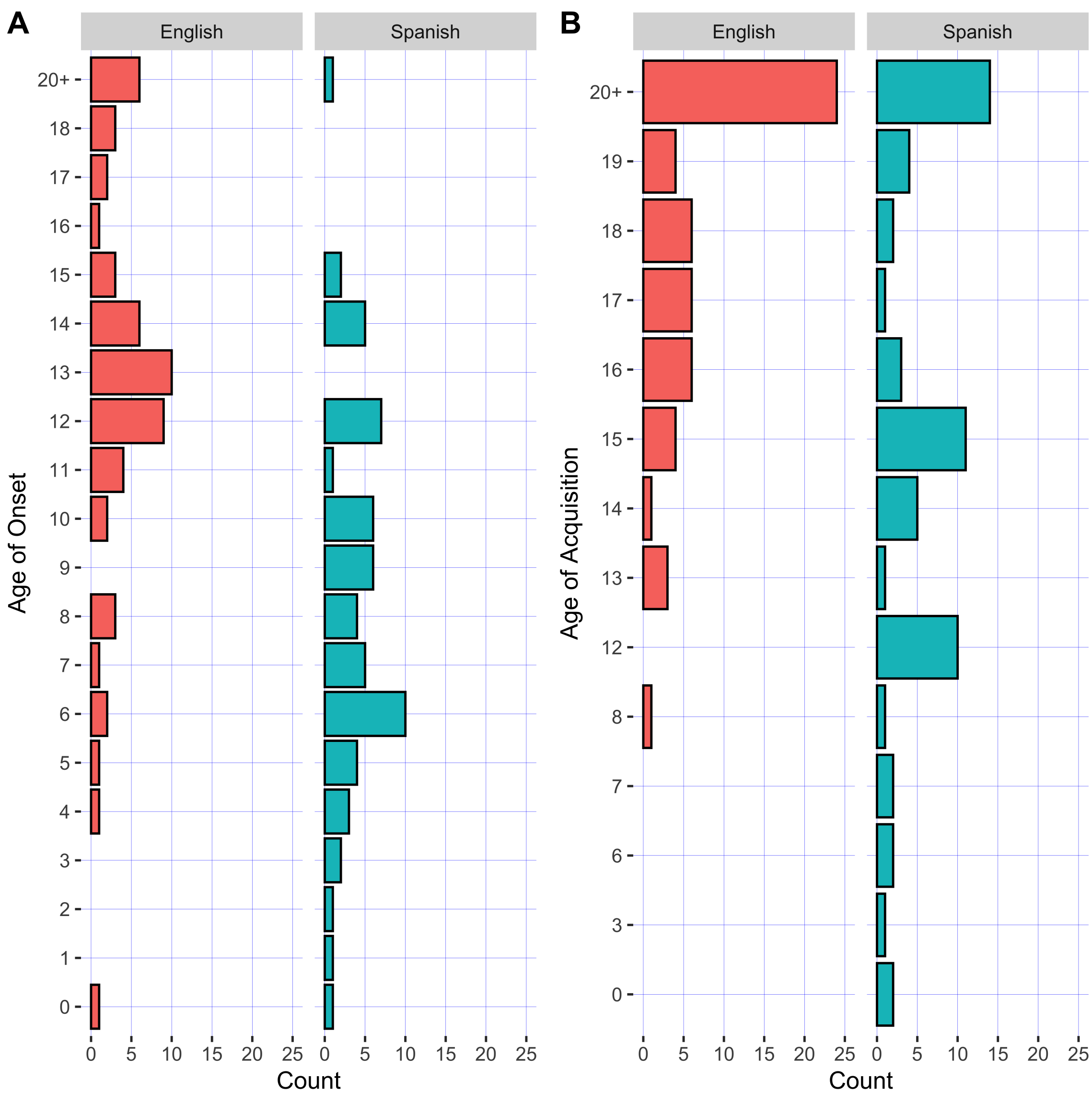


Figure 3.1: Age of Onset and Age of Acquisition in each bilingual group

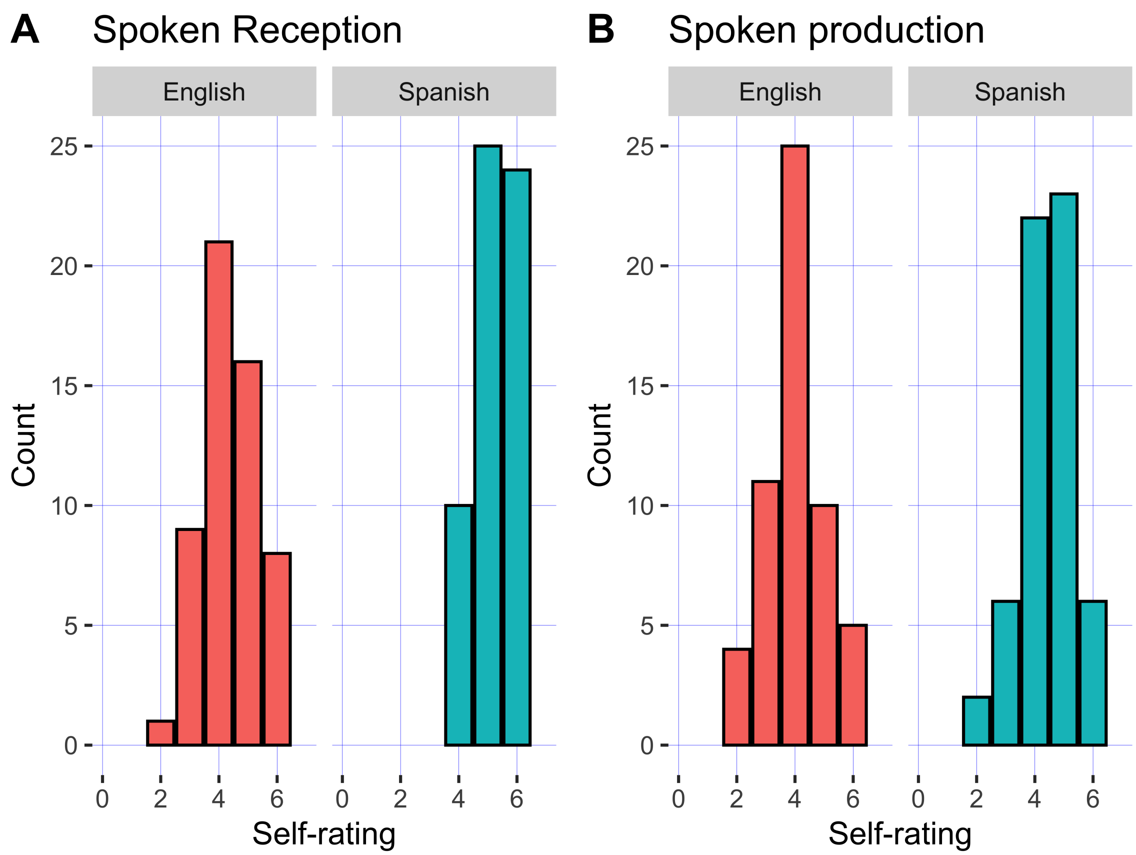


Figure 3.2: Self-rated 1-6 proficiency in production and perception

distinct cross-linguistic situations. First, the L3 phoneme /i/ was included to create a conflict in which both source languages, Spanish and English, have a similar sound /i/. The phoneme /i/ was given in the Spanish word *fin* and the English word *feel*. Next, a centralized vowel, /ʌ/, was included in an attempt to bias the selection of English. This condition was intended to be assimilated to the English choice *fun*. Third, the phoneme /o/ was included to bias the same Spanish category, where a rounded /o/ does not exist in American English. The intended choice in this case was the Spanish word *son*, but the English word *thought* was also provided as an alternative. Finally, the phoneme /y/ was added to explore how a sound that is not present in either language will be categorized. Given that /y/ is a high-front vowel, it is possible that it could be assimilated to other high vowels, either /i/ as in *feel* or *fin*, or /u/, as provided in *fool* or *su*. Additionally, after making each selection, the participants then rated their pick for goodness of fit by clicking a 1-5 continuous Likert scale (3.4.

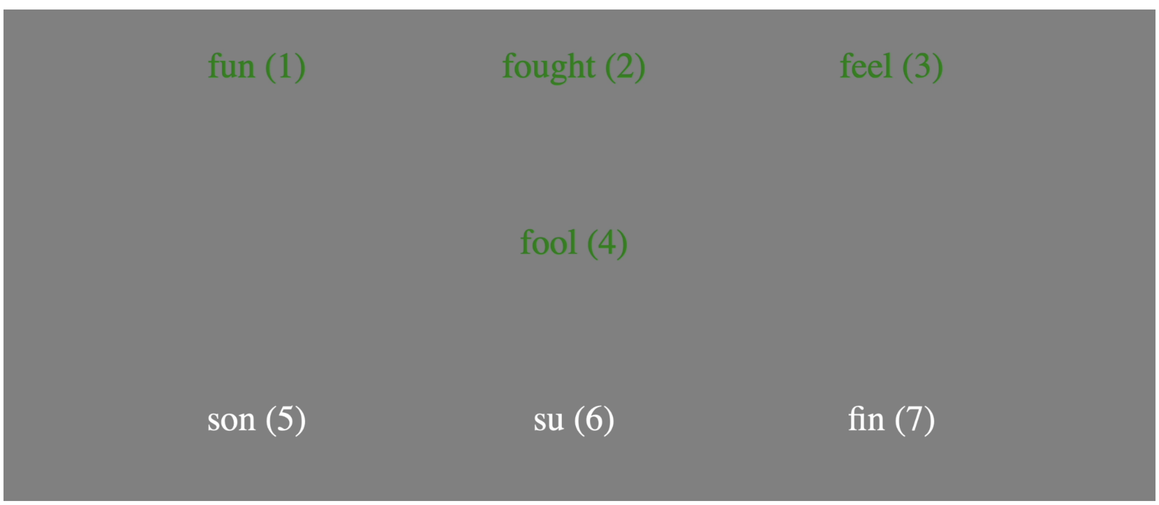


Figure 3.3: Example screen of the perceptual assimilation task

3.2.2.2 Stimuli

The stimuli were recorded by adult, female L1 speakers of French and German, respectively, and were also collected online via Prolific. The speakers were given each vowel in a word or non-word in both a fricative and bilabial frame. In the event a non-word was provided, a real word containing that vowel sound was included to

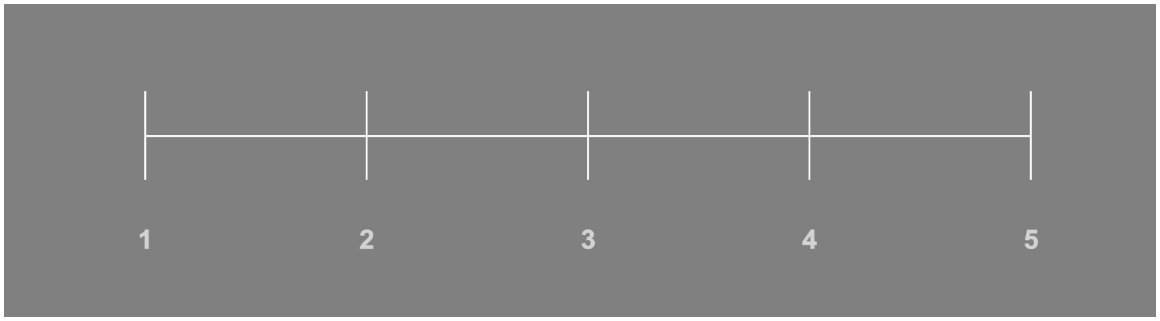


Figure 3.4: Example of the likert style rating after each selection

aid the informant in producing the intended pronunciation of the vowel. Once the stimuli were recorded, one of the two tokens provided by the speaker for each vowel was selected and spliced adding the appropriate onset and coda. In total, eight stimuli were created per language. Figure 3.5 shows the formant values of the included stimuli in German and French in comparison to similar sounds in English and Spanish. For the purpose of this Figure, an adult female speaker of Madrid Spanish and an adult female American English speaker provided vowel tokens of the answer choices in the present study by producing the carrier words while being recorded in PRAAT (*son*, *su* and *fin* in Spanish and *fought*, *feel*, *fool*, and *fun* in English).

Table 3.1: Written stimuli used to elicit the auditory stimuli from the native German and French speakers.

Written stimulus	Language	Intended Phoneme
Pief, Fief	German	/i/
Pof, Fof	German	/o/
Püf, Füf	German	/y/
Puff, Fuff	German	/ʌ/
Pif, Fif	French	/i/
Pof, Fof	French	/o/
Puf, Fuf	French	/y/
Puff, Fuff	French	/ʌ/

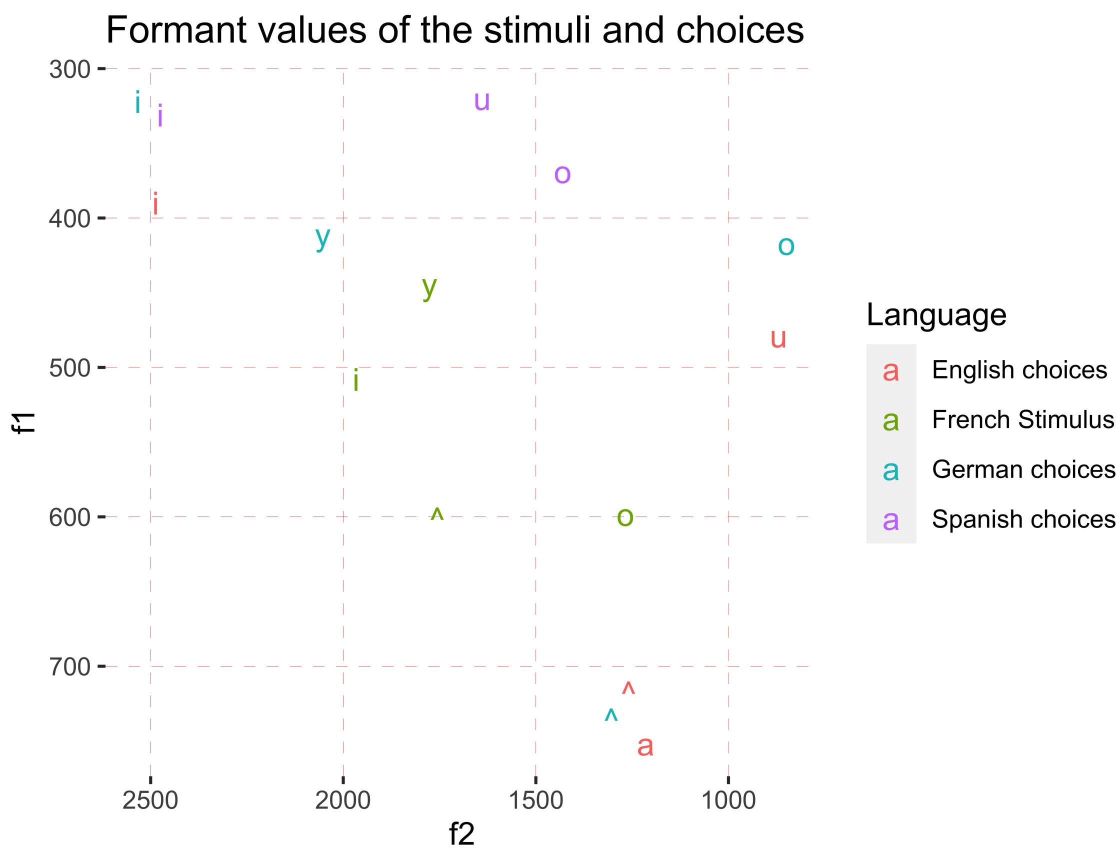


Figure 3.5: Formant values of the model speakers

3.2.3 Procedure

First, all participants first completed the adapted Bilingual Language Profile (Birdsong et al., 2012) online. An English and Spanish version of the questionnaire was adapted and given to the participants based on their L1. All participants who answered "no" to the question "Besides English and Spanish, do you speak a third language?" were invited to take part in the experimental task. The perceptual assimilation task was given to participants online. An English and Spanish version of this task was also created, in which all instructions were given in either English or Spanish. During the task, all participants heard French first, followed by a brief pause, and then heard German sounds. The order of the stimuli was counterbalanced and the two tasks were given in a single session with a brief pause between them. The experiments were programmed in Psychopy (Peirce et al., 2019) and made available online via Pavlovia.

3.2.4 Statistical Analysis

The data were analyzed using a series of Bayesian multilevel multinomial logistic regression models in R @[R-base]. The models were fit using the R package `brms` (Bürkner, 2017). A model was run for each of 4 groups: L1 Spanish, L1 English, monolingual English and monolingual Spanish. In each model, the outcome variable was word choice. In the bilingual groups, this consisted of sevel total options (three Spanish words: *fin*, *su*, *son* and four English words: *fun*, *fought*, *feel*, and *fool*.) Thus, the bilingual model estimates the log odds of choosing one of the seven choices, and would sum to 1 when converted to probability. The fixed effect predictors of the bilingual models were phoneme (/i/, /ʌ/, /y/ and /o/) and stimulus language (French or German). Random effects included a random intercept for participant to take into account the nested structure of the data.

The monolingual models analyzed word choice as a function of phoneme and stimulus language, again with a random intercept for participant. In this case, language choice was more limited in each group, with the Spanish monolingual group only having 3 options: *fin*, *su*, *son*, while the English group had 4 word choices: *fun*, *fought*, *feel*, and *fool*. All models included regularizing, weakly informative priors (Gelman, Simpson, & Betancourt, 2017), which were normally distributed and centered at 0 with a standard deviation of eight for all population-level parameters. The region of practical equivalence (ROPE) was set to 0.18, as the outcome variable was in log-odds (see Kruschke (2018)). All models were fit with 2000 iterations (1000 warm-up). Markov-chain Monte-Carlo sampling was carried out with six chains distributed between six processing cores.

3.2.5 Results of the L3 Perceptual Assimilation Task

3.2.5.1 English speaking group comparison

Table 3.2 shows the overall percentage of each word choice (out of 4 possible in English), given each of the 4 phonemes in both French and Spanish by the English monolingual group. The bold numbers are cases in which a word received at least 33 percent of choices.

Table 3.2: The percentage of categorizations of phonemes in the English monolinguals group..

Stimulus Language	choice	i	o	ʌ	y
German	feel	**0.93**	0.01	0.01	0.06
German	fool	0.02	**0.45**	-	**0.78**
German	fought	0.05	**0.52**	**0.42**	0.11
German	fun	-	0.02	**0.57**	0.05
French	feel	**0.90**	-	0.14	0.04
French	fool	0.04	0.07	0.03	**0.69**
French	fought	0.04	**0.38**	0.31	0.21
French	fun	0.02	**0.54**	**0.52**	0.06

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \tag{3.1}$$

Figure 3.6 shows the probability of each word choice given a language and phoneme in the monolingual group and compares it to the conditional probability of the same conditions in the bilingual group. These estimates were derived from the parameter estimates of the Bayesian Multinomial model, in which log odds were converted to probability. The bilinguals' conditional probability was calculated by dividing the probability from the full model by all categories of that language and shows the probability of a particular selection given that its language has been chosen. This calculation allows for the direct comparability between the monolingual and bilingual groups, since the bilinguals had both English and Spanish categories to choose from. In other words, the calculation of conditional probability tells us whether bilinguals behave similarly to monolinguals when they choose their L1 to categorize L3 sounds. The equation 3.1 illustrates the formula used to calculate the conditional probability of a word choice given that its language has been chosen.

The triangular shaped points represent the monolingual group and the circular points represent the conditional probability of the bilingual group in the event they chose and English category. The color of each point represents each word choice. In each panel of the figure, it is of interest whether the same color points of different shapes are close to one another, as this would indicate that the bilingual and monolingual groups are assimilating new language sounds to English similarly.

Tables 3.3, 3.4, 3.5 and 3.6 show the most probable responses of each group given a particular phoneme and language in both the bilingual and monolingual groups. The 2.5% and 97.5% columns represent the extremes of the 95% Highest Density Interval (HDI), in which 95% of the most probable parameter estimates fell. There were some differences between the monolingual and bilingual groups, while other categorizations were rather consistent. As inspection of Table 3.3 indicates, the phoneme /i/ was categorized similarly (as *feel*) by monolinguals and bilinguals when they chose an English category, regardless of L3. On the other hand, Table 3.4 shows that responses to /o/ saw some variation between groups. In particular, both *fought* and *fool* were probable responses by the monolingual group in German, while *fun* and *fought* were probable responses in French. The bilingual group, however, preferred *fool* in German, and *fun* in French. Responses to /ʌ/ also varied. In German, the monolingual group

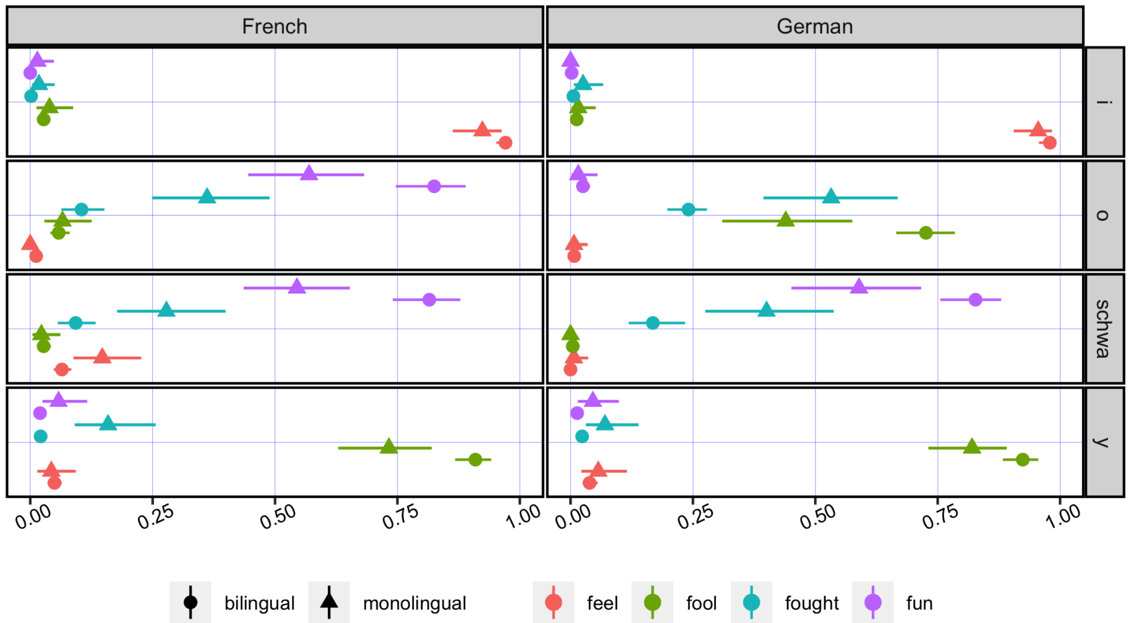


Figure 3.6: A comparison of the probability and conditional probability of each word choice per stimulus per language in the English monolingual group and English L1 bilingual group.

preferred the choice *fun*, but *fought* was also a probable response. In French, the monolingual group preferred *fun*. The bilingual group preferred *fun* in both German and French. Finally, categorizations of /y/ were, like /i/, consistent between the groups and languages. Both groups assimilated /y/ to *fool* in both French and German.

Table 3.3: Most probable responses given /i/ in French and German by L1 English monolinguals and L1 English bilinguals.

Phoneme	Choice	Stimulus Language	Group	Probability	2.5%	97.5%
i	feel	German	monolingual	0.96	0.91	0.98
i	feel	French	monolingual	0.92	0.86	0.96
i	feel	German	bilingual	0.98	0.96	0.99
i	feel	French	bilingual	0.97	0.95	0.98

In sum, the English bilingual and monolingual groups had distinct preferences for some L3 vowels, but not others. This result suggests, at least, that learning a

Table 3.4: Most probable responses given /o/ in French and German by L1 English monolinguals and L1 English bilinguals.

Phoneme	Choice	Stimulus Language	Group	Probability	2.5%	97.5%
o	fought	German	monolingual	0.53	0.39	0.67
o	fool	German	monolingual	0.44	0.31	0.58
o	fun	French	monolingual	0.57	0.45	0.68
o	fought	French	monolingual	0.36	0.25	0.49
o	fool	German	bilingual	0.73	0.67	0.79
o	fun	French	bilingual	0.83	0.75	0.89

Table 3.5: Most probable responses given /ʌ/ in French and German by L1 English monolinguals and L1 English bilinguals.

Phoneme	Choice	Stimulus Language	Group	Probability	2.5%	97.5%
/ʌ/	fun	German	monolingual	0.59	0.45	0.72
/ʌ/	fought	German	monolingual	0.40	0.28	0.54
/ʌ/	fun	French	monolingual	0.54	0.44	0.65
/ʌ/	fun	German	bilingual	0.83	0.76	0.88
/ʌ/	fun	French	bilingual	0.82	0.74	0.88

second language could have an impact one's categorization of third language sounds in the event participants are only given single-language categories, as has been done in previous studies.

3.2.5.2 Monolingual Spanish assimilations

Unlike the English groups, the Spanish monolingual group and Spanish bilingual group had similar categorization patterns overall when the Spanish bilingual group picked a Spanish category to categorize. Table 3.7 shows the percentage of choices given a phoneme in the Spanish monolingual group in both French and German. For all four phonemes, both groups chose the same word the highest percentage of the time. In particular, given /i/, both groups chose *fin* the most, while both /o/

Table 3.6: Most probable responses given /y/ in French and German by L1 English monolinguals and L1 English bilinguals.

Phoneme	Choice	Stimulus Language	Group	Probability	2.5%	97.5%
y	fool	German	monolingual	0.82	0.73	0.89
y	fool	French	monolingual	0.73	0.63	0.82
y	fool	German	bilingual	0.92	0.88	0.96
y	fool	French	bilingual	0.91	0.87	0.94

and the /ʌ/ were matched to *son*, and /y/ was assimilated to *su* by both groups in both languages. Figure 3.7 visualizes the probabilities generated by the Bayesian Multinomial regression model. In all cases, the bilingual and monolingual participants appear to categorize French and German sounds similarly.

Table 3.7: The percentage of categorizations of phonemes in the Spanish monolingual group.

Stimulus Language	choice	i	o	/ʌ/	y
German	fin	**0.96**	0.09	0.28	0.16
German	son	0.01	**0.59**	**0.56**	0.08
German	su	0.03	0.32	0.17	**0.76**
French	fin	**0.83**	0.13	0.25	0.20
French	son	0.03	**0.61**	**0.47**	0.06
French	su	0.13	0.26	0.29	**0.74**

3.2.5.3 English L1 group

Figure 3.8 shows the categorization data of the English L1 group of each phoneme in both languages. The shaded bars represent the rating for goodness of fit, where a lighter shade represents a higher average rating. Tables 3.12 and 3.13 show the percentage of each word choice given a particular phoneme in French and German in the English L1 group. The choice with the highest percentage per for each phoneme is in bold. In the event that there were two choices that were above 33%, they are both in bold.

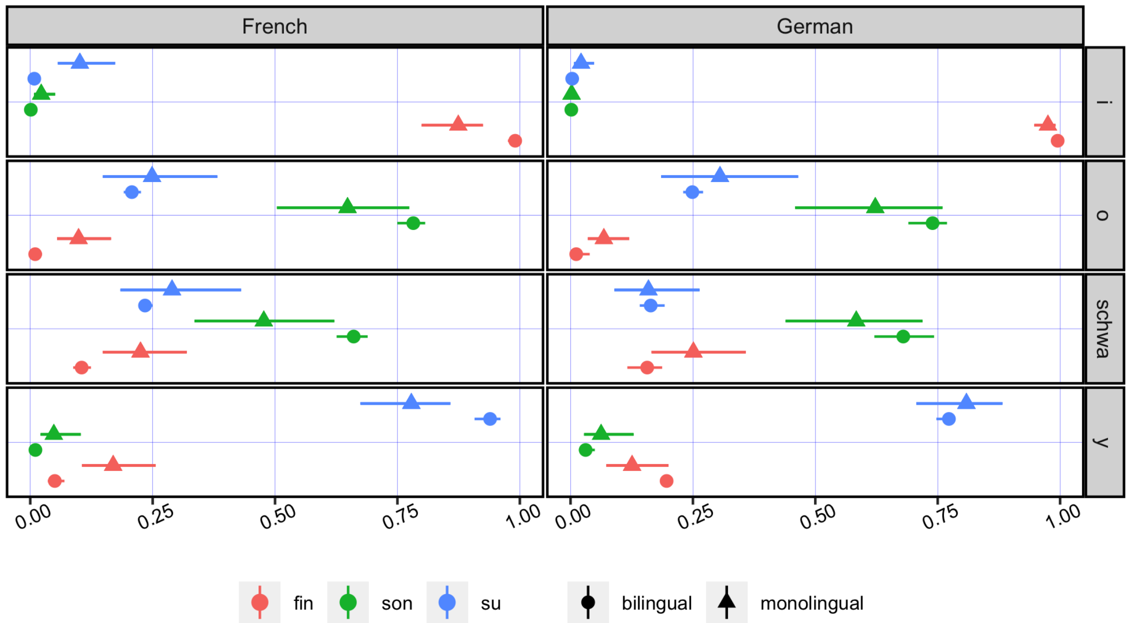


Figure 3.7: A comparison of the probability and conditional probability of each word choice per stimulus per language in the Spanish monolingual group and Spanish L1 bilingual group.

As can be seen in Tables 3.12 and 3.13, the English L1 group, given the L3 French phoneme /i/, chose both their English category /i/ provided in the choice *feel* and their Spanish category /i/ provided in fin *fin*. German /i/ was categorized similarly. For /o/, this group chose *fun* the most often in both cases in French, while the most chosen word in German for /o/ was *fool*. /ʌ/ in both French in German was most often assimilated to the intended category *fun* Finally, both the L3 French and German phonemes /y/ resulted in choices of *fool* and *su*, the English and Spanish categories for /u/.

3.2.5.4 Spanish L1 group

Figure 3.9 shows the percentage that each answer choice was chosen given a particular phoneme in each language. Table 3.14 and Table 3.15 similarly report the percentage of each answer choice given a particular phoneme in each language in the Spanish L1 group. The Spanish L1 group had a similar preference to the English L1 group in

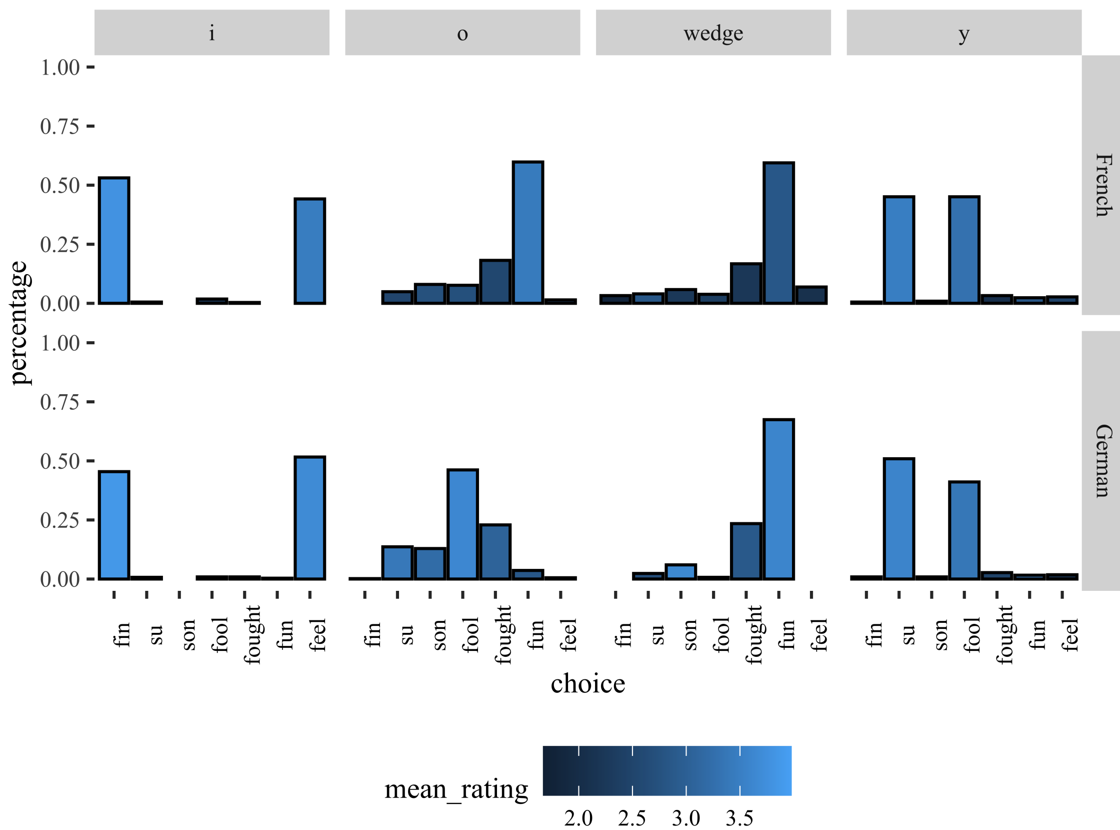


Figure 3.8: Percentage of each word choice given a phoneme in French and German in the L1 English-L2 Spanish group and Self-reported spoken production

Table 3.8: Most probable responses given /i/ in French and German by L1 Spanish monolinguals and L1 Spanish bilinguals

Phoneme	Choice	Stimulus Language	Group	Probability	2.5%	97.5%
i	fin	German	monolingual	0.98	0.95	0.99
i	fin	French	monolingual	0.87	0.80	0.92
i	fin	German	bilingual	0.99	.98	1.00
i	fin	French	bilingual	0.99	.98	1.00

Table 3.9: Most probable responses given /o/ in French and German by L1 Spanish monolinguals and L1 Spanish bilinguals

Phoneme	Choice	Stimulus Language	Group	Probability	2.5%	97.5%
o	son	German	monolingual	0.62	0.46	0.76
o	son	French	monolingual	0.65	0.50	0.77
o	son	German	bilingual	0.74	0.69	0.77
o	son	French	bilingual	0.78	0.75	0.81

their categorization of /i/, where both English and Spanish categories were chosen. However, the Spanish L1 group chose the English category given a German stimulus more often than the Spanish category. In both French and German, when both /o/ and /ʌ/ were played, the most chosen answer was *fought*. Finally, given the phoneme /y/, the Spanish L1 group chose they English *fool* most often in German, and the Spanish *su* most often in French.

3.2.5.5 Results of the models

The output of each model was converted from log-odds to probability using a combination of the `conditional_effects` and `make_conditions` functions in the R package `brms` (Bürkner, 2017). Thus, for each phoneme in each language, the sum of the probability of choosing the categories combined is 1. Figure 3.10 shows the probability of each choice in French and German by both groups when /i/ was the phoneme. Figure 3.11, shows the same set of probabilities when /o/ is played. Figures 3.12

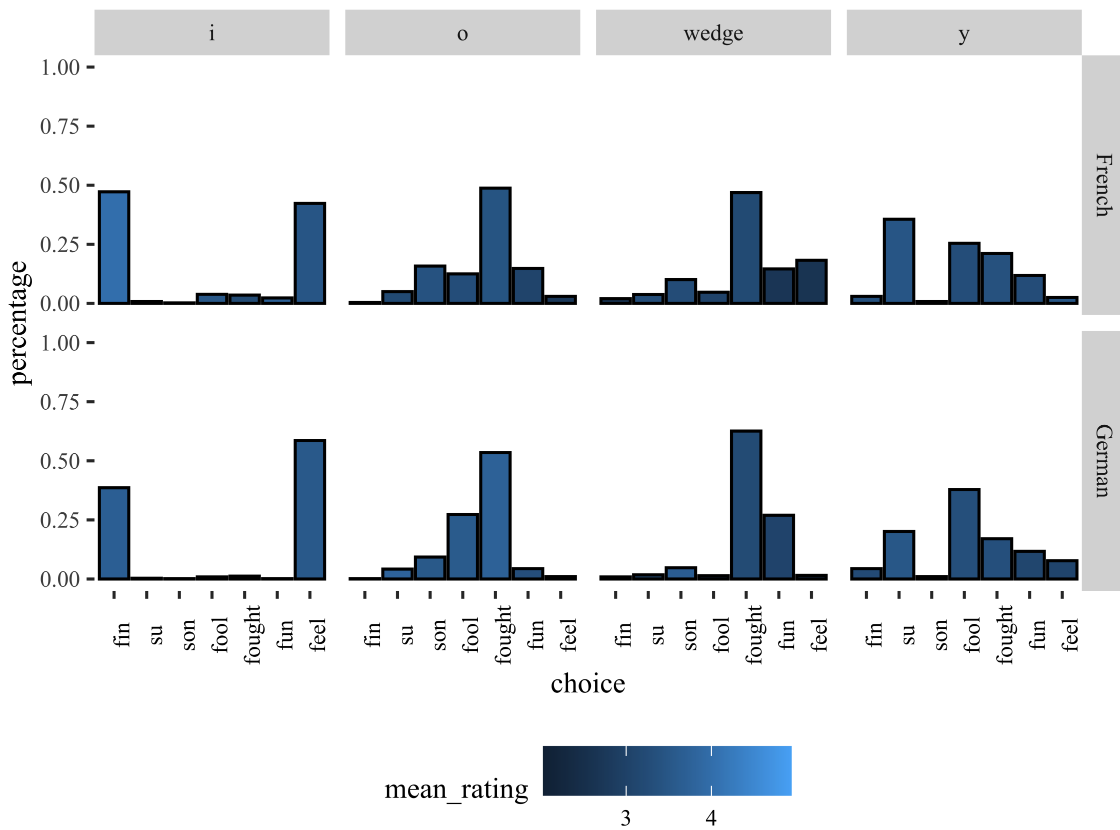


Figure 3.9: Percentage of each word choice given a phoneme in French and German in the L1 Spanish-English L2 group.

Table 3.10: Most probable responses given /ʌ/ in French and German by L1 Spanish monolinguals and L1 Spanish bilinguals

Phoneme	Choice	Stimulus Language	Group	Probability	2.5%	97.5%
ʌ	son	German	monolingual	0.58	0.44	0.72
ʌ	son	French	monolingual	0.48	0.34	0.62
ʌ	son	German	bilingual	0.68	0.62	0.74
ʌ	son	French	bilingual	0.66	0.63	0.69

Table 3.11: Most probable responses given /y/ in French and German by L1 Spanish monolinguals and L1 Spanish bilinguals

Phoneme	Choice	Stimulus Language	Group	Probability	2.5%	97.5%
y	su	German	monolingual	0.81	0.71	0.88
y	su	French	monolingual	0.78	0.67	0.86
y	su	German	bilingual	0.77	0.75	0.78
y	su	French	bilingual	0.94	0.91	0.96

and 3.13 represent the probabilities of each response when /ʌ/ and /y/ are played respectively.

Figure 3.10 corroborates the trends observed in the descriptive statistics when the participants in both groups were asked to categorize the phoneme /i/ in French and German. In particular, for the English L1 group in both French and German and the Spanish L1 group in French, portions of the credible parameter estimates from the posterior distribution overlap with a probability of .5, suggesting that compelling evidence is not present for a preference for the English or Spanish category. Differently, the Spanish L1 group did show evidence in German of a preference for the English category *feel*, in which the probability of picking *feel* when the stimulus was German was 0.61 (HDI = 0.53 - 0.68). On the other hand, when phoneme was /i/ in French the probability of choosing *feel* was 0.46 (HDI = 0.37 - 0.55).

After hearing the stimulus /o/, the English L1 group preferred the choice *fun* in

Table 3.12: The percentage of categorizations of French phonemes in the L1 English group.

choice	i	o	/ʌ/	y
feel	**0.44**	0.01	0.07	0.03
fin	**0.53**	-	0.03	0.01
fool	0.02	0.08	0.04	**0.45**
fought	-	0.18	0.17	0.03
fun	-	**0.60**	**0.59**	0.02
son	-	0.08	0.06	0.01
su	0.01	0.05	0.04	**0.45**

Table 3.13: The percentage of categorizations of German phonemes in the L1 English group.

choice	i	o	/ʌ/	y
feel	**0.52**	0.01	0.02	
fin	**0.45**	0.00	0.01	
fool	0.01	**0.46**	**0.41**	0.01
fought	0.01	0.23	0.03	0.23
fun	0.00	0.04	0.02	**0.67**
son	-	0.13	0.01	0.06
su	0.01	0.14	**0.51**	0.02

French (0.78 (HDI = 0.67 - 0.87), and *fool* (0.57 (HDI = 0.48 - 0.64) in German, and the Spanish L1 group preferred *fought* in both languages. /ʌ/ was categorized as the choice *fun* in both French (0.78 [HDI = 0.67 - 0.87] and German (0.79 [HDI = 0.67 - 0.87]. The Spanish L1 group preferred *fought* in both languages. (French: 0.52 [HDI = 0.44 - 0.6]; German: 0.68 [HDI = 0.61 - 0.74].

The English L1 group assimilated /y/ in both French and German to their English and Spanish /u/ (choices *su* and *fool*), without a clear preference for either. In particular, the probability of the English L1 group's choice of *su* given the phoneme /y/ and the stimulus language German was 0.53 [HDI = 0.44 - 0.61], while the probability of choosing *fool* was 0.42 [HDI = 0.34 - 0.51]. In French, the probability of

Table 3.14: The percentage of categorizations of French phonemes in the L1 Spanish group.

choice	i	o	/ʌ/	y
feel	**0.42**	0.03	0.18	0.02
fin	**0.47**	0.00	0.02	0.03
fool	0.04	0.12	0.05	0.25
fought	0.04	**0.49**	**0.47**	0.21
fun	0.02	0.15	0.15	0.12
son	0.00	0.16	0.10	0.01
su	0.01	0.05	0.04	**0.36**

Table 3.15: The percentage of categorizations of German phonemes in the L1 Spanish group.

choice	i	o	/ʌ/	y
feel	**0.59**	0.01	0.02	0.08
fin	**0.39**	0.00	0.01	0.04
fool	0.01	0.27	0.01	**0.38**
fought	0.01	**0.54**	**0.63**	0.17
fun	0.00	0.04	0.27	0.12
son	0.00	0.09	0.05	0.01
su	0.00	0.04	0.02	0.20

choosing *su* given the phoneme /y/ was 0.47 [HDI = 0.36 - 0.58], and the probability of *fool* was 0.48 [HDI = 0.37 - 0.58].

The Spanish L1 group, on the other hand, assimilated French /y/ to the Spanish *su*, and the German /y/ to English *fool*. The probability of the Spanish L1 group's choice of *su* given the phoneme /y/ and the stimulus language French was 0.35 [HDI = 0.24 - 0.46]. The probability of the Spanish L1 group's choice of *fool* given the phoneme /y/ and the stimulus language German was 0.43 [HDI = 0.35 - 0.5].

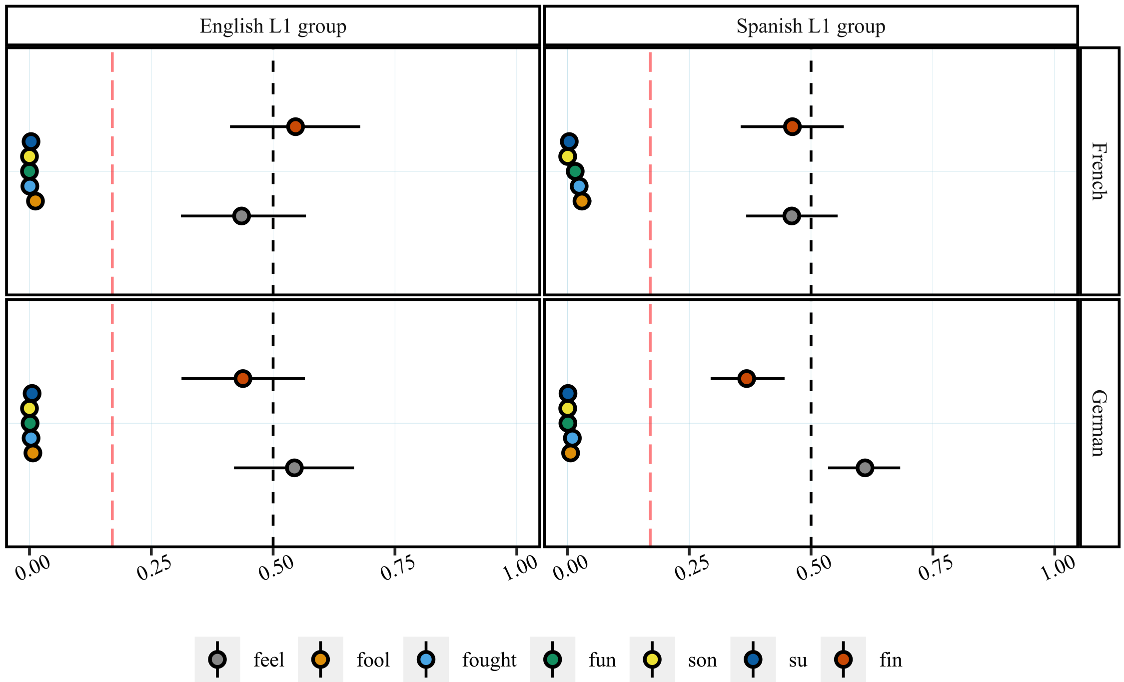


Figure 3.10: Probability of each choice given /i/

3.3 Experiment 2 - AX Discrimination Task

Experiment two examined the discrimination of two English vowel contrasts using a resynthesized continua. In particular, the discrimination of the /i/ to /ɪ/ and /æ/–/ɑ/ were investigated. The purpose of this experiment was to determine whether access to English-like vowel contrasts in L3 words in groups in which English was their L1 and when English was the L2.

3.3.1 Participants

A total of 134 participants took part in experiment 2. The data of eight participants was removed, who, despite self-reporting being bilingual, also answered that they were not yet comfortable speaking their second language during the BLP. The remaining 142 participants made up three total groups: L1 Mexican Spanish - L2 English speakers (n = 59), L1 English - L2 Spanish speakers (n = 34) and American English monolinguals

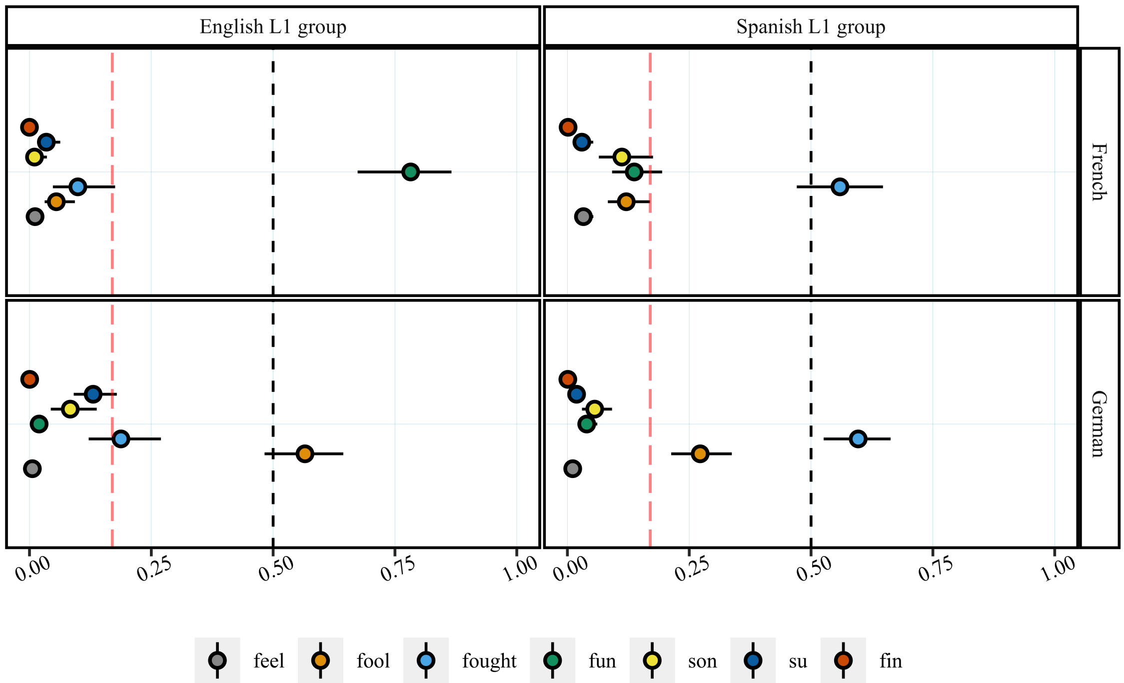


Figure 3.11: Probability of each choice given /o/

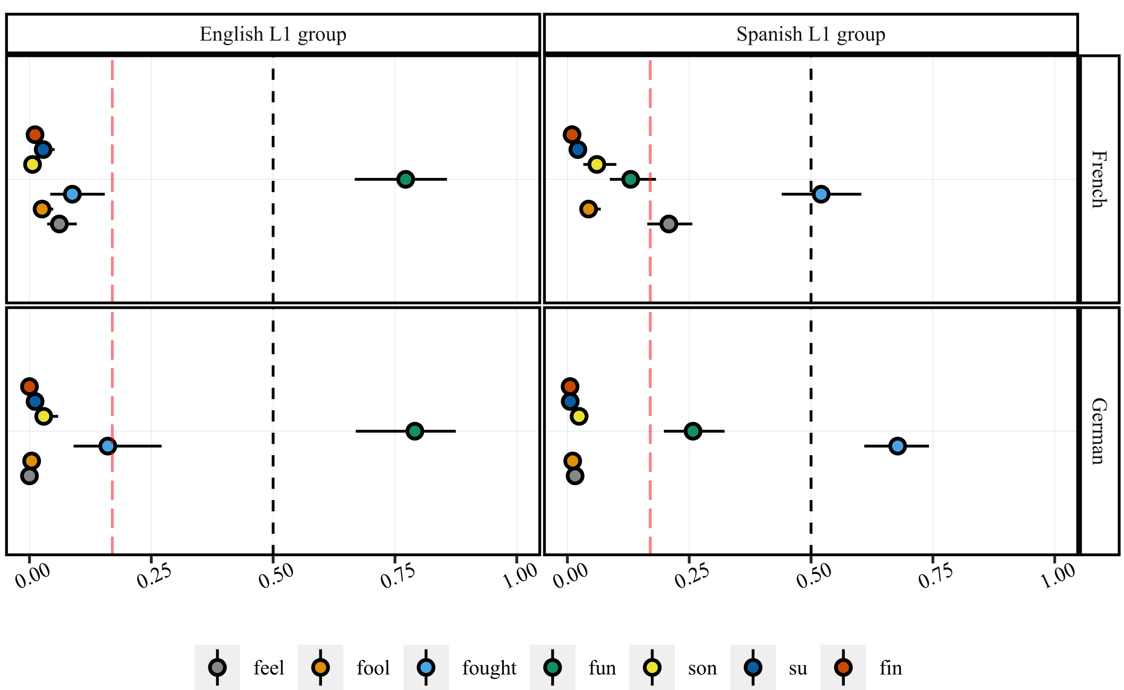


Figure 3.12: Probability of each choice given schwa

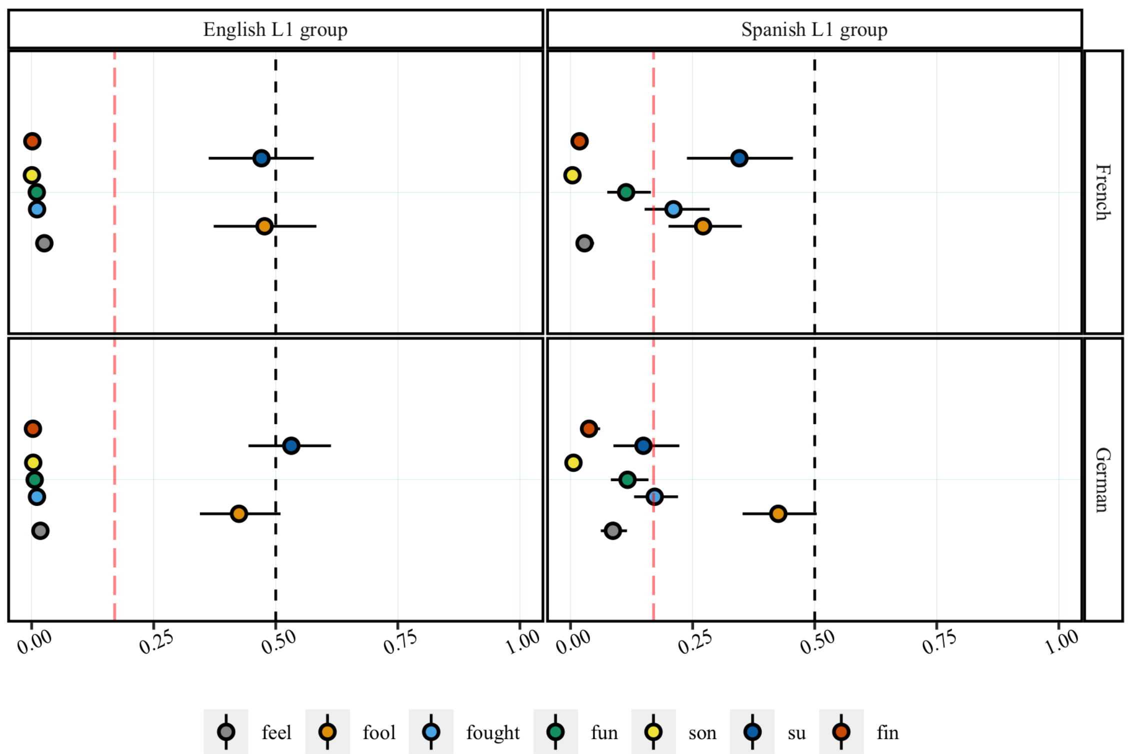


Figure 3.13: Probability of each choice given /y/

49. Many of the L1 Spanish speakers returned for the second experiment, while very few L1 English speakers returned. In total, eight L1 English-L2 Spanish participants from experiment 1 completed experiment 2, while 44 L1 Spanish-L2 English participants completed both experiment 2. Again, all participants were recruited on prolific and were pre-screened according to criteria detailed below. In addition to filters in place from Prolific.co, the participants were screened further using an adapted version of the Bilingual Language Profile (Birdsong et al., 2012). All participants who answered 'no' to the question "Do you speak a language other than English and Spanish" were permitted to continue the experiment.

3.3.1.1 Bilinguals

Exactly as in experiment 1, screening data and experiment-initial questionnaires were used to find bilingual participants who began learning their L2 later in life and reported not having learned a language aside from English and Spanish. The English L1 group came from all over the United States while Spanish L1 group came from Mexico. Each groups mean age, L2 use, self-reported oral and perceptive proficiency are seen in figure 3.14. As can be seen from the figures, The English L1-Spanish L2 group began L2 learning later on average, while they also felt comfortable in their L2 at a later age than the Spanish L1-English L2 group. The participants also rated their L2 proficiency. They were given a 0-6 Likert-type scale in which they answered the questions "How well do you speak [their L2]?" and "How well do you understand [their L2]?". "0" corresponded to "not very well at all", where "6" corresponded to "very well" (Figure 3.15). The L1 Spanish group rated themselves higher on average in their L2 spoken proficiency 4.4 (sd = 0.9) than the English group 4.1 (sd =1.1). The same trend was observed in perception. The L1 Spanish group rated their proficiency in perception as 5.2 (sd = 0.7) and English group rated their L2 perception as 4.3 (sd = 1).

3.3.2 Materials

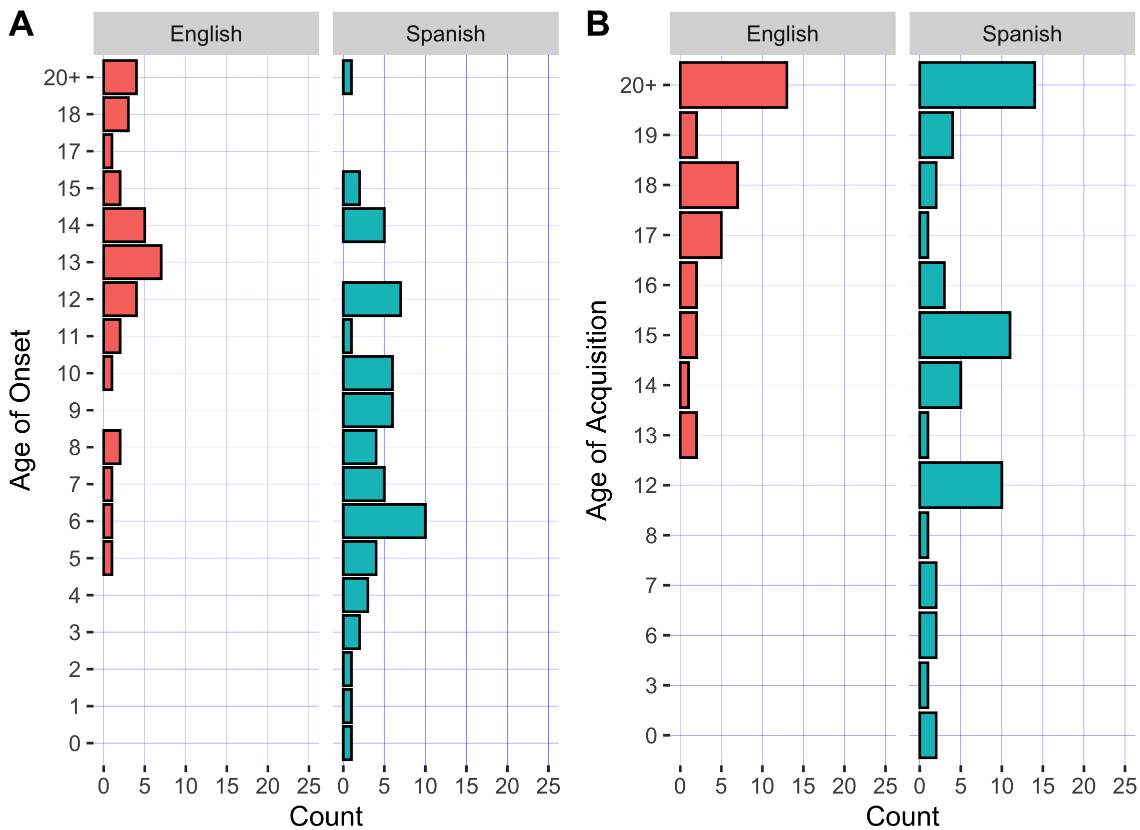


Figure 3.14: Age of Onset and Age of Acquisition in each bilingual group (AX)

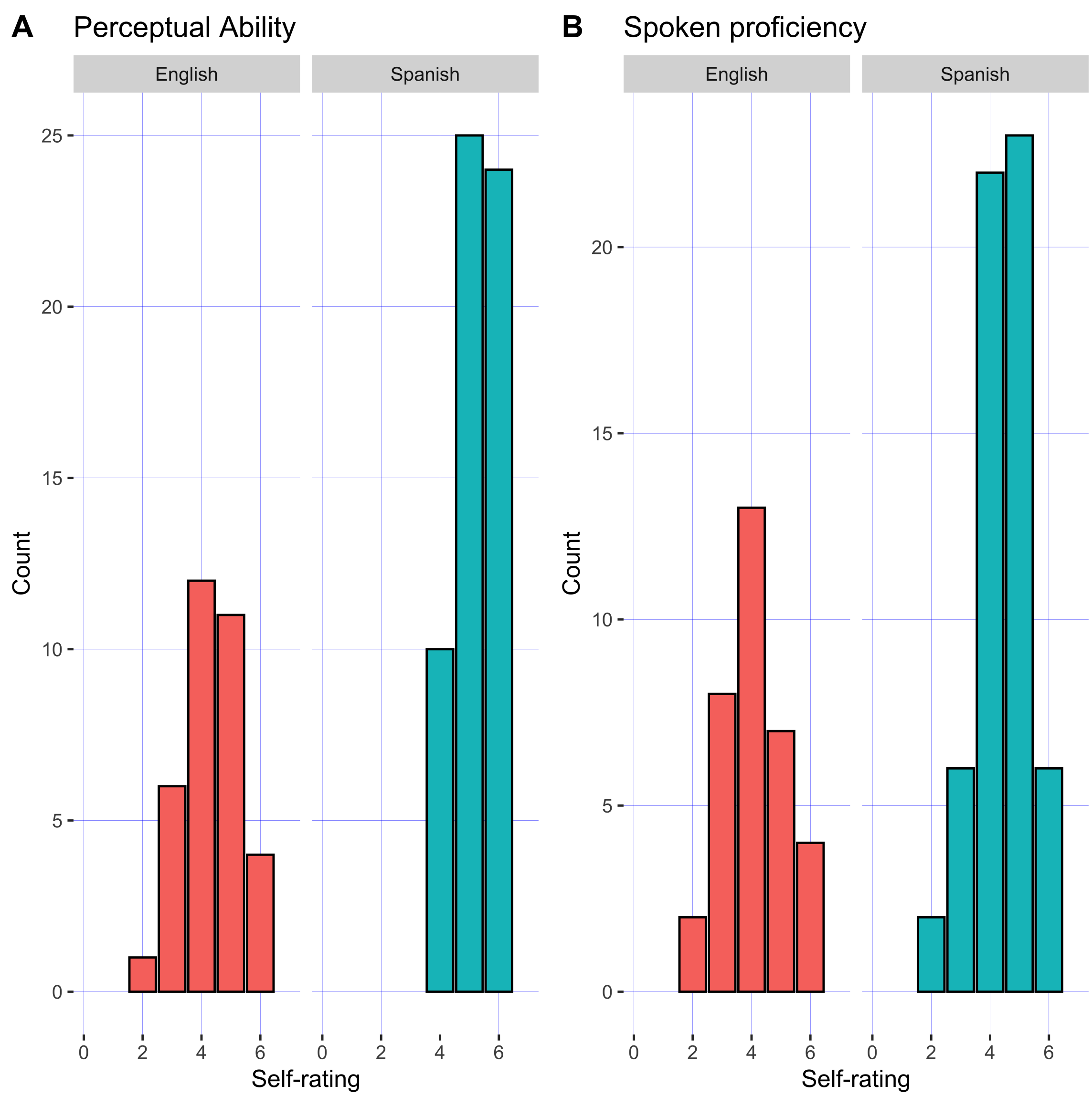


Figure 3.15: Self-rated 1-6 proficiency in production and perception (AX)

3.3.2.1 Acoustic continua

Participants completed two AX discrimination tasks, one in French and one in German. During the task, the participants heard two words or psuedowords played back to back and were asked whether they heard a difference between the two words. These words only differed in a single vowel sound. This vowel was drawn from a seven step continuum for each vowel contrast. Each task contained 149 total trials and was designed to measure the discrimination of two English vowel contrasts: /i/ to /ɪ/ and /æ/–/ɑ/ and was embedded into a French or German carrier word.

Eight total 7-step vowel continua were created for this experiment to measure two vowel contrasts in "French" and "German". It it important to note that, although French and German do not necessarily make use of these English vowel contrasts, this is not important since these speakers do not yet speak either French or German. As a result, the use of English-like vowels in what we refer to as "French" or "German", while they may not be authentic, serve the purpose of identifying whether or not these speakers have access to their language-specific discrimination routines in what they believe is a new language, and whether this ability varies as a function of which language they believe that they are hearing. In each case, a German or French word-pair was produced by a native English speaker by producing the word on both ends of the continuum by producing, for example, /i/ or /ɪ/ in the given word. Following these productions, the vowel sounds were extracted and a single onset and offset was saved per word. Then, using a PRAAT script, the first 3 formants were manipulated to create the continuum, and the vowels were spliced with the saved onsets and offsets, to re-create whole words. The carrier words can be seen in Table 3.16. Due to an error, the German carrier word were three /i/ continua and one /a/ continuum. This error was caused by an erroneous entry in the conditions file of the psychopy experiment used to collect the ends of the continua and was not discovered until after data collection had taken place. However, this does not appear to have impacted the results, since discrimination was very poor for every contrast.

Table 3.16: Carrier words used in the AX discimination task.

language	frame	word	word type
French	fricative 1	fat	non-word
French	fricative 2	fit	non-word
French	bilabial 1	pat	non-word
French	bilabial 2	pic	non-word
German	fricative 1	fisch	fish
German	bilabial 1	pake	non-word
German	bilabial 2	pike	non-word
German	bilabial 3	pisch	non-word

3.3.2.2 Procedure

During the AX discrimination task, participants heard two sounds from a given continuum back to back. Their task was to choose whether or not the sounds they heard were the same (by pressing 1) or different (by pressing 0). Each participant heard a total of 149 tokens per language, which were randomly sampled from the four possible continua per language. German and French were given in individual sessions with a brief break between them. All of the step interval combinations were possible, in which a given step one could also be presented with any other step in the continuum.

3.3.3 Statistical Analysis

In order to measure sensitivity, d prime (d') (McNicol, 2005) was calculated using formula 3.2. D prime is a tool used to measure a subject's ability to detect a signal, and is generally accepted as being more robust as a measure of sensitivity than accuracy alone (Banks, 1970). The formula uses z(H) and z(F), which are the z-transforms of the hit rate and the false alarm rate. The hit rate is any case in which a difference between continuum steps is correctly identified. The false alarm rate refers to choosing that there was difference when there was none. D prime ranges from 0 (chance) to +/- 4.65 (near-perfect discrimination).

$$d' = z(H) - z(F) \tag{3.2}$$

In order to determine whether meaningful differences existed between the groups, a Bayesian multilevel model was run. The outcome variable of the model was d' and the fixed effect predictors were group (L1 English or L1 Spanish), language (French or German), and continuum type (labeled "a" or "i") and self-rated proficiency (continuous 1-7). A random intercept for participant was included to take into account the nested structure of the data. Model priors were the default in `brms`, a student's T distribution with 3 degree of freedom. All models were fit with 4000 iterations (1000 warm-up). Hamiltonian Monte-Carlo sampling was carried out with six chains distributed between eight processing cores.

3.3.4 Results

Figures 3.16 and 3.17 show d' as a function of language and group. The boxplots suggest that the English groups slightly outperformed the Spanish group. 3.17 reports the mean d' and standard deviation in each group for each token. Overall, the discrimination of every sound contrast was above chance for the L1 English group, but only around chance for the L1 Spanish group. Figures 3.18 and 3.19 shows the total quantity of each of the four possible outcomes of the a given trial: false alarm, hit, miss, and correct rejection. Both figures show that false alarms were slightly more frequent than correct rejections, suggesting that discrimination was close to chance overall.

Table 3.17: Mean D' scores and standard deviations for each group and token

Token	Group	language	Mean	SD
fat	English	French	0.60	0.97
fat	English monolingual	French	0.54	0.82

Token	Group	language	Mean	SD
fat	Spanish	French	0.04	0.83
fisch	English	German	0.49	0.80
fisch	English monolingual	German	0.28	0.78
fisch	Spanish	German	0.14	0.78
fit	English	French	0.26	0.94
fit	English monolingual	French	0.41	0.84
fit	Spanish	French	0.05	0.75
pake	English	German	0.40	0.86
pake	English monolingual	German	0.02	0.76
pake	Spanish	German	-0.05	0.73
pat	English	French	0.38	0.90
pat	English monolingual	French	0.29	0.91
pat	Spanish	French	0.15	0.75
pic	English	French	0.38	0.87
pic	English monolingual	French	0.20	0.92
pic	Spanish	French	0.32	0.77
pike	English	German	0.26	0.73
pike	English monolingual	German	0.16	0.75
pike	Spanish	German	0.00	0.80
pisch	English	German	0.42	0.80
pisch	English monolingual	German	0.18	0.80
pisch	Spanish	German	-0.13	0.68

Despite the overall near-chance performance, figure 3.20 shows an edge for the L1 English participants. Specifically, the model estimated that d' for the English group in French was 0.39 [95% HDI = `r`report_ax$hdi_lo[1]` - 0.8] and 0.37 [95% HDI = -0.04 - 0.75] in German. On the other hand, the Spanish group was less sensitive to the contrasts in both French 0.02 [95% HDI = -0.4 - 0.38] and German -0.14 [95%

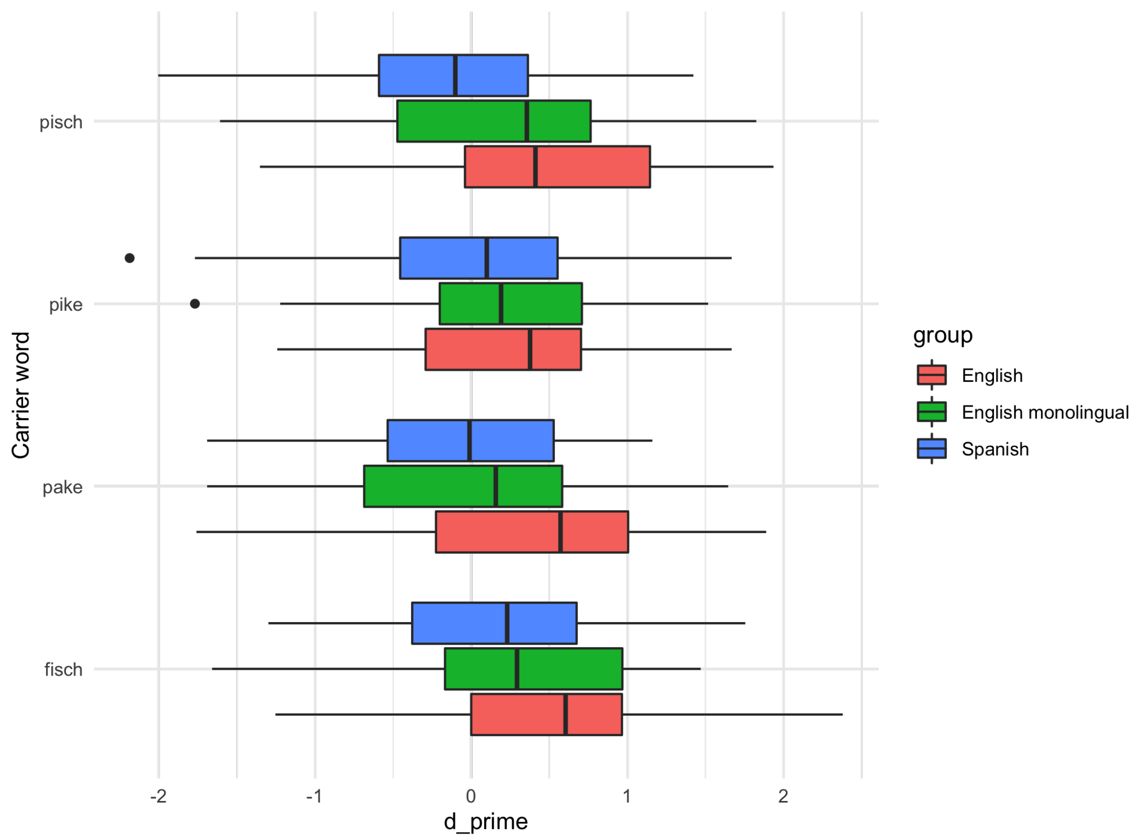


Figure 3.16: D' in German carrier words per group

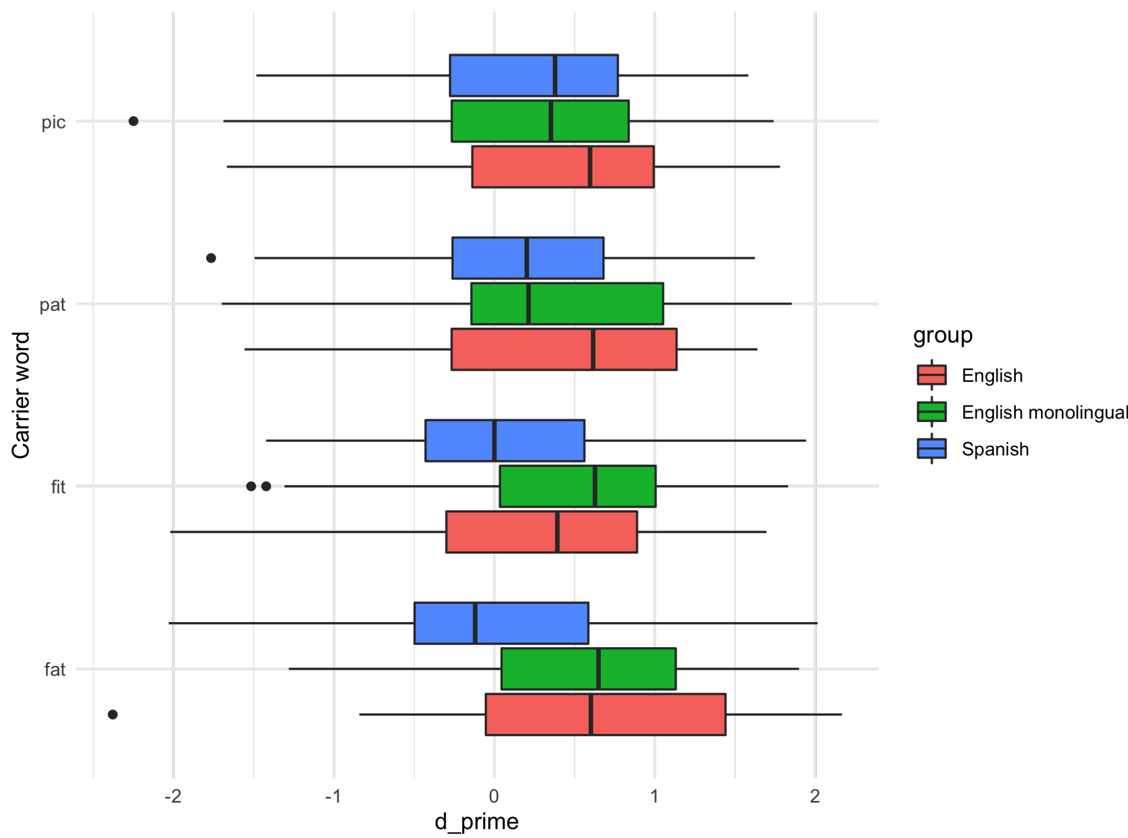


Figure 3.17: D' in French carrier words per group

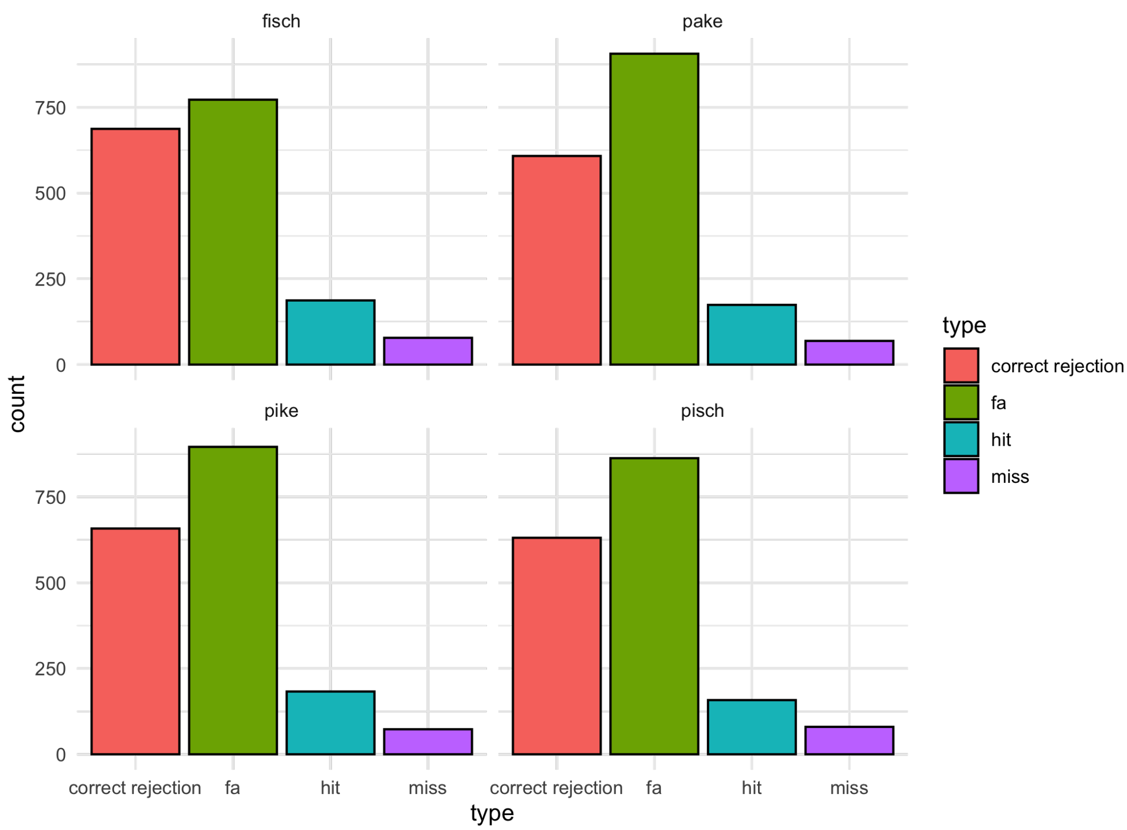


Figure 3.18: Quantity of repsonse types per carrier word in German

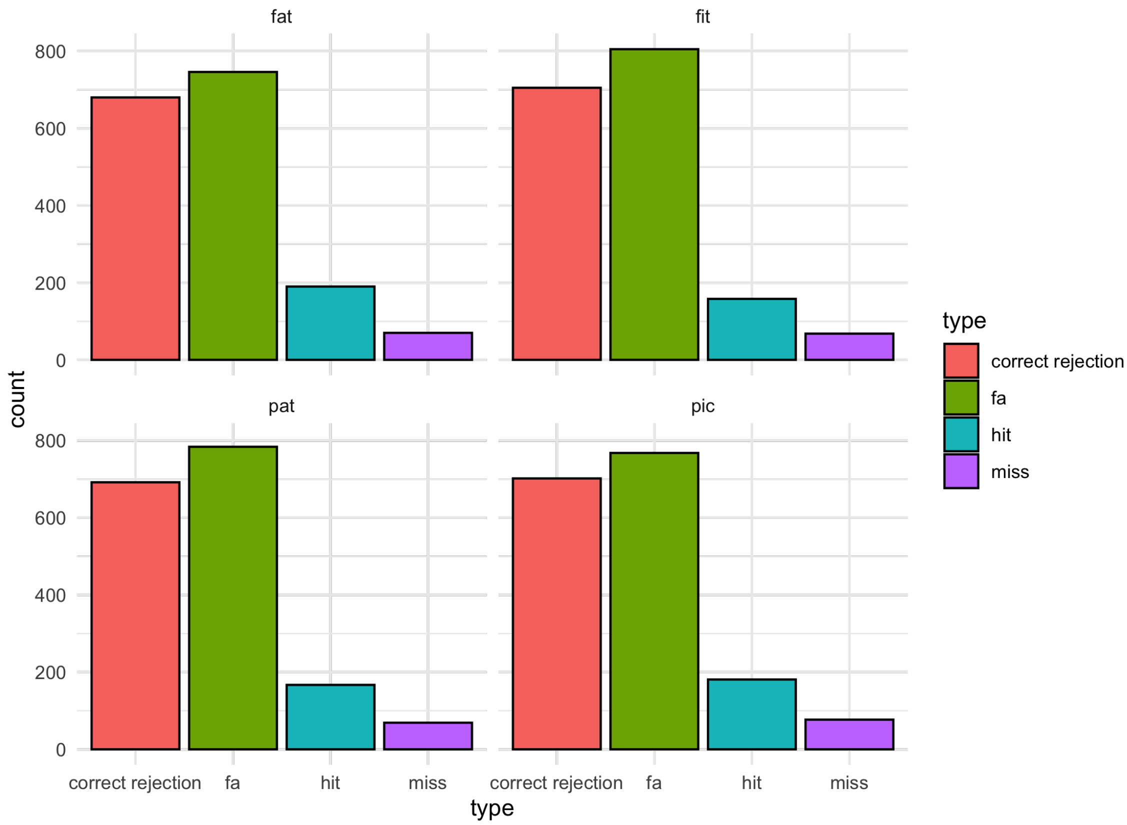


Figure 3.19: Quantity of repsonse types per carrier word in French

HDI = -0.57 - 0.2]. Table @ref(tab: ax-tab-mod) shows the full model output.

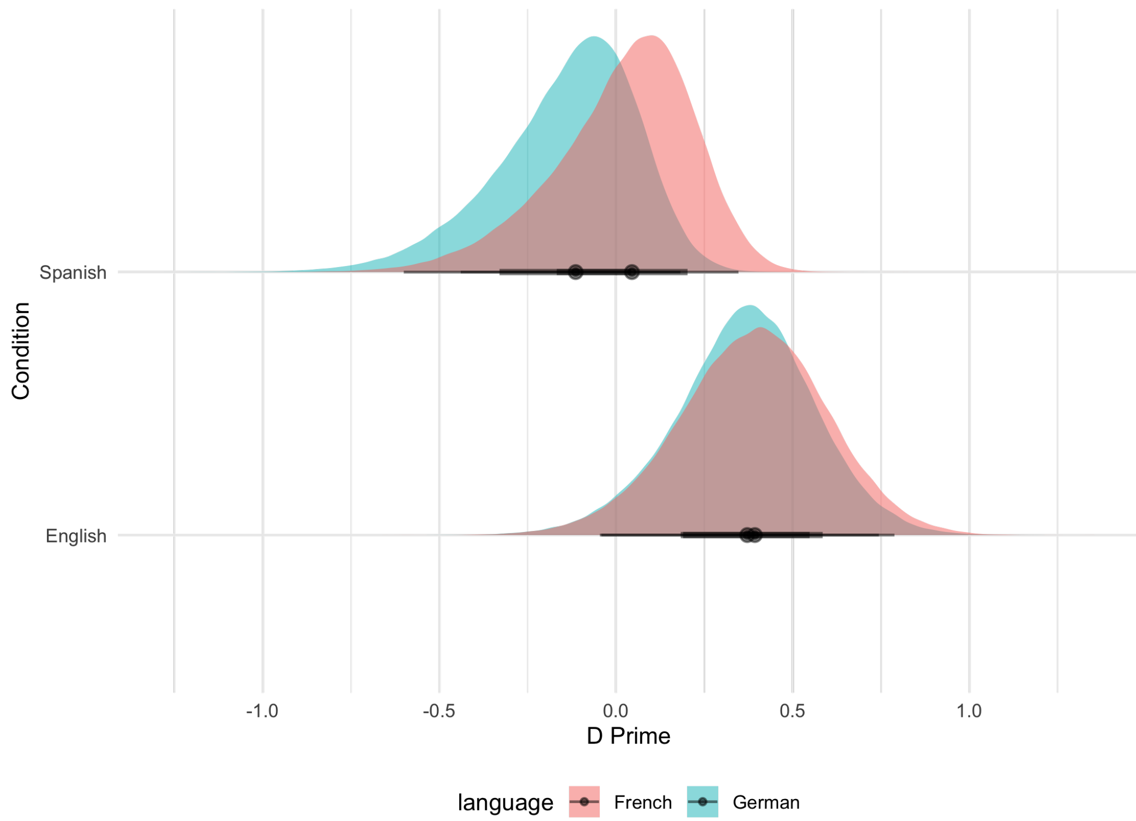


Figure 3.20: D-prime as a function of group and language for the AX discrimination task

Table 3.18: Full model of the AX discrimination task

Parameter	Median	HDI	X..in.ROPE	MPE	Rhat	ESS
Intercept	0.11	[-0.45, 0.68]	0.26	0.65	1	6830
L1 Spanish	-0.50	[-0.80, -0.19]	0.00	1.00	1	6758
German	-0.10	[-0.39, 0.19]	0.44	0.76	1	8272
i	-0.17	[-0.41, 0.07]	0.27	0.92	1	8740
Proficiency	0.09	[-0.03, 0.21]	0.56	0.93	1	6352
L1 Spanish:German	-0.04	[-0.41, 0.33]	0.42	0.59	1	8250
L1 Spanish:i	0.26	[-0.05, 0.55]	0.14	0.95	1	8726

Parameter	Median	HDI	X..in.ROPE	MPE	Rhat	ESS
German:o	0.16	[-0.20, 0.52]	0.32	0.81	1	7134
L1 Spanish:German:i	-0.19	[-0.65, 0.27]	0.26	0.80	1	6997

3.4 Discussion

The present study examined the categorization and discrimination of vowel sounds in German and French languages by Spanish-English bilinguals in both orders of acquisition. In several instances, L3 sounds had a similar probability of being assimilated to English or Spanish sound, such as in 3 out of 4 cases with the phoneme /i/. The results were mixed in terms of whether the L3 that the participants heard impacted their preference for English or Spanish categories. In total, there were three cases in which the L3 did result in the preference for an English or Spanish category out of 16 possible. In particular, the Spanish L1 group showed a slight preference for the English /i/ (the choice *feel*) when they heard /i/ in German, but not in French. Additionally, the Spanish L1 group also categorized the phoneme /y/ differently when they heard it in L3 French than when they heard it in L3 German. In particular, the L3 French /y/ was assimilated to Spanish /u/ (*su*) and the L3 German /y/ was assimilated to English /u/ (*fool*). Although, in categorization, there was an overall preference for English, the results suggested that bilinguals do have access to both their L1 and L2 during L3 perception and categorization, a finding which single-language access models, such as the TPM and L2SF cannot account for.

These results are best explained by the Linguistic Proximity Model (Westergaard, 2021; Westergaard et al., 2017), which suggests that L3 learners have access to categories in both their L1 and L2 during L3 acquisition. The results of the perceptual assimilation task provided evidence that both groups of bilinguals categorized French and German with both English and Spanish categories. Only the LPM can account for this result, since other models predict that only one language influences the L3. In these cases, one would expect to see L3 vowels categorized using a single language, rather than both of them. On the other hand, The Typological Primacy Model (Rothman, 2010, 2011, 2013, 2015) suggests that, while bilinguals have access to both their languages at first exposure, this access is lost at some point during what the model refers to as the "initial stages" of L3 acquisition. While the present data does not necessarily provide counter-evidence to the TPM, it provides a starting point for future research by providing a picture of the categorization patterns of L3 phonemes in two languages by Spanish-English bilinguals which could be examined in a longitudinal design. If the TPM is correct, a distinct change in categorization would be expected to reflect a bias for a single language's category. The present study provides a basis of comparison, and evidence that bilinguals do have access to both the L1 and L2 at first exposure. However, in six out of eight cases, global typological similarity (English and German vs. French and Spanish), did not result in any obvious bias in categorization.

It is unclear why the Spanish L1 group seemed to be sensitive to which L3 they heard, where the English L1 group was not. It is worth noting that there are differences between the groups in self-rated proficiency and age of onset and acquisition; the Spanish L1 group rated themselves as more proficient on average in spoken and perceptual abilities and had an earlier age of onset and acquisition on average. L3 models have suggested that differences in L2 proficiency impact its level of influence on the L3. For example, the TPM has stated that the L2 must be sufficiently proficient to be a source of influence. In this view, a beginning learner of a second languages is not thought to be able to transfer their L2 representations to their L3 initial state and, in many of the TPM studies to date, this has been taken to mean "advanced proficiency". Though the present study was not interested in the effect of L2 proficiency on L3 categorizations, it is worth noting that the group differences in proficiency might have impacted the results, although it is difficult to say exactly how, since the cutoff for "advanced" proficiency is subjective and arbitrary. This notion of the impact of L2 proficiency on CLI could be further investigated, in which the more traditional categorical grouping of proficiency (e.g., novice, intermediate, advanced) could also be examined in a continuous fashion and corroborated across measures. In other words, it's unclear for the purpose of the present study, and likely with many proficiency measures, where the TPM's cutoff for "advanced" should be.

In discrimination, the L1 English group outperformed the L1 Spanish group, although discrimination was not excellent by either group. It is possible that the nature of online data collection was associated with this poor performance across groups. For instance, it may be that inconsistent speaker quality, varied speaker attention levels and inconsistent use of headphones all contributed to poor performance. Future experiments for AX discrimination should take place in a lab setting in which these variables can be properly controlled. Despite these potential limitations, the result that the L1 English speakers outperformed the L1 Spanish speakers can still be inform L3 perception. The contrasts in the AX discrimination task were English contrasts and, although both groups speak English (as a first or second language), it appears that native language discrimination may play a role in L3 processing. This result is surprising based on the predictions of current L3 models, since none of them posit a privileged status for the native language. To the author's knowledge, only the work by Hermas (2010) has suggested that a speaker's L1 contrains L3 development, and that this view has not been formally modeled.

The present study had additional limitations. Firstly, as briefly mentioned in reference to the AX discrimination task, online studies offers unique challenges and limitations. One potential issue is the lack of the ability to control the participant's environment outside of the instructions and experiment itself. As a result, language mode effects (Joseph V. Casillas & Simonet, 2018; Grosjean, 1998) cannot be completely ruled out. Additionally, headphone quality, speaker quality, background noise and volume level are all variables within the participant's control that could not be reasonably controlled for in the context of online data collection. Second, self-reported proficiency, while convenient and fast, is a subjective measure of language ability that would likely be improved by a more objective proficiency measure such as the LexTALE (Lemhöfer & Broersma, 2012). Future research could aim to test categorization and discrimination at at L3 first exposure in a better controlled lab setting.

3.5 Conclusion

This chapter presented two experiments that analyzed the categorization and discrimination of novel language sounds by bilinguals. In particular, this chapter was concerned with how Spanish-English bilinguals categorized French and German vowels given both of their language categories, and whether discrimination of English-like vowels would be successful in the new languages. Primarily, the purpose of these experiments was to determine whether bilinguals would show bias toward either their L1 or L2 when they perceive new language vowels based on typology or language status. The experiments showed that the categorization of both French and German vowels were matched to the most appropriate English or Spanish category, and there was not evidence of typological or language status effects blocking access to a given language. The results for discrimination were very poor and are considered inconclusive, and likely associated with the experimental setting or design.

Chapter 4: The Production of the new language vowels by Spanish-English bilinguals

4.1 Introduction

The present chapter examines the impact of two previously known languages on the production of a novel language. In particular, it examines how Spanish-English bilinguals in both orders of acquisition produce German and French vowels and stop consonants. Models of L3 acquisition vary in their prediction regarding how previously known languages influence a third and suggest that typology, language status (L1 or L2) or language activation predict how either language will play a role in L3 development. All of these views have received empirical support to some degree, and there is still debate as to why these conflicting findings exist.

The typological view, for example (see e.g. Rothman, 2011), suggests that overall crosslinguistic similarity will result in whole-language influence. For example, a Spanish-English bilingual who learns Portuguese would be predicted to be influenced by their Spanish, given the close historical relationship between Spanish and Portuguese. While this view has been supported in syntax, it has not seen much empirical testing or support in the acquisition of phonology. In reality, an increasing number of studies are finding that both languages impact third language phonological development Llama & Cardoso (2018). However, it is still not well understood what impacts the relative degree of influence of the L1 or L2 and whether, while they may not cause whole language influence, the factors of language status and typology may still be at play.

In order to test the relative impact of the L1 and L2 on the L3, the present study used a mirror-image design of participants and gave them words from two “third” languages produce (French and German). The mirror image design refers to two groups of speakers who speak the same two languages but acquired them in opposite order (e.g., L1 Spanish-L2 English and L1 English-L2 Spanish). The mirror image design is intended to reveal whether language status (order of acquisition) affects performance on a particular task, since these speakers theoretically have the same linguistic resources. If the groups perform differently, it could provide evidence that order of acquisition impacts access to linguistic representations in new language learning, while similar performance would be predicted by the typological view. The use of two "third" languages at first exposure was chosen to examine whether the same individual would be impacted by the same source language no matter what language they learn, or whether both the source languages known and the language being learned are important. Typological views of L3 acquisition would suggest that, given the historical relationship between French and Spanish and English and German respectively, that these languages would be produced more similarly. Language status views would predict that influence should not vary as a function of which L3 is being learned. Finally, views that suggest both languages impact L3 learning have not yet developed enough explanatory power to derive clear predictions related to Spanish-English bilinguals learning French and German.

4.1.1 L3 Production Studies

The findings in empirical studies of cross-linguistic influence in L3 production have varied. One of the first studies to examine progressive influence of the L1 or L2 on L3 production was the seminal case study of Williams & Hammarberg (1998). This study elicited the production of an adult L1 British English, L2 German, and L3 Swedish speaker in the L2 and L3 upon her arrival to Sweden. The speech samples were rated for native-likeness by native speakers of German and Swedish, respectively, with low ratings (i.e. non-nativeness) being elaborated upon. In the event of non-native speech, the raters guessed where the speaker in the recording might be from. The informant was rated as having near native productions in her L2 German, while her L3 Swedish was rated as being non-native like and to be German-accented. The experiment was repeated after 6 months in Sweden, however, and the Swedish raters then judged the informant's Swedish to be British English accented.

This study constituted evidence of an L2 status effect in the initial stages of L3 phonological acquisition, in which the second learned language influenced L3 production, but also provided evidence that this effect diminishes as L3 proficiency increases. This notion has been called the ‘foreign language effect’ (Meisel, 1983), which refers to the idea that speakers who learn a second non-native language are biased to sound unlike a native speaker of their native language.

The default L2 status effect has received some empirical support in the L3 phonology literature. In a study of global accent production, heavier L2 influence in L3 productions was found by L1 Polish, L2 German and L3 English speakers based on ratings of EFL instructors (Wrembel, 2010). Similar findings have also been reported in vowel production (Kamiyama, 2007) and vowel reduction and speech rhythm (Gut, 2010). L2 influence has also been found in VOT productions. Llama et al. (2010) examined L3 Spanish VOT production by French-English mirror-image bilingual groups and found that both groups had L2-like productions of the L3.

Other findings in L3 production, however, have yielded mixed results. Several studies have found that acoustic properties of the participants’ productions fall between L1 and L3 values, suggesting that both the L1 and the L2 have some influence on L3 productions, rather than solely one language. For instance, Wrembel (2014) measured VOT and aspiration in all languages of participants with two different language combinations: L1 Polish, L2 English, and L3 French; (2) L1 Polish, L2 English, and L3 German. The results showed that each language had a specific stop-value, and that the L3 VOT productions were intermediate, falling between the L1 and L2 values. Similarly, Wrembel (2011) examined thirty-two learners of L3 French with L1 Polish and L2 English who were recorded reading lists of words in carrier phrases. As in previous studies (Wrembel, 2014), combined transfer from the L1 and the L2 in VOT productions was found. Findings of combined L1 and L2 influence in VOT productions were also reported by Wunder (2010) in L3 Spanish speakers, and by Blank & Zimmer (2009) in L3 English speakers who spoke L1 Brazilian Portuguese and L2 French. Other studies have found an L1 influence on production despite L3 proficiency (Wrembel, 2012), or in advanced L3 learners (Llama & Cardoso, 2018). Importantly, these studies report L3 VOT values which fall between L1 and L2 values. Following the suspicion that intermediate values might have to do with either sampling issues or proficiency effects, Parrish (2021) examined Mexican Spanish-English bilinguals who produced voiceless stop-initial French words in isolation at first exposure to the language. The results revealed that the relative VOT of the L3 fell between their own L1 and L2 values, in line with previous research, and suggested that intermediate values were less likely to have been seen in previous studies as a result of small samples or proficiency effects. However, a subsequent analysis of the data suggested that wide individual variation existed, in which some participants produced L3 French as L1 Spanish-like and others produced intermediate, L2-like values. This result suggests that higher samples could reveal group trends and provide better insights into individual variation in crosslinguistic influence, as opposed to assuming that a single group trend exists.

4.1.2 The present study

In order to examine the impact of order of acquisiton and language similarity on cross linguistic influence, the present study utilized pseduoword reading and pseudoword repetition tasks in order to elicit productions of vowels and stop consonants in 4 total languages in each participant (or 3 in the case of monolingual comparisons).

In particular, the present chapter is guided by the following research questions:

RQ1: When Spanish-English bilinguals produce French and German words at first exposure, to what degree will they be influenced by their known languages?

RQ2: Will source language impact vary as a function of segment (vowel or stop)?

The first research question was examined by extracting formant values from each of the vowels and measuring the VOT of each consonant in each language. These acoustic measures were recorded and are assumed to provide insight into a particular segment's representation. That is, by comparing the acoustics of each sound, the present study extrapolates acoustics to phonological representation in each of the languages involved. In regard to RQ1, it was predicted that both French and German productions would fall between Spanish and English points of comparison, which has previously been found in VOT (Parrish, 2022). However, it was predicted that German VOT would be more English-like than French VOT, and that French VOT would be more Spanish-like than German VOT. Other than both L3 vowels and stops being intermediate of the L1 and L2, there were not specific predictions for RQ2, given the lack of studies done in the acquisition of L3 vowels.

4.2 Vowel Production

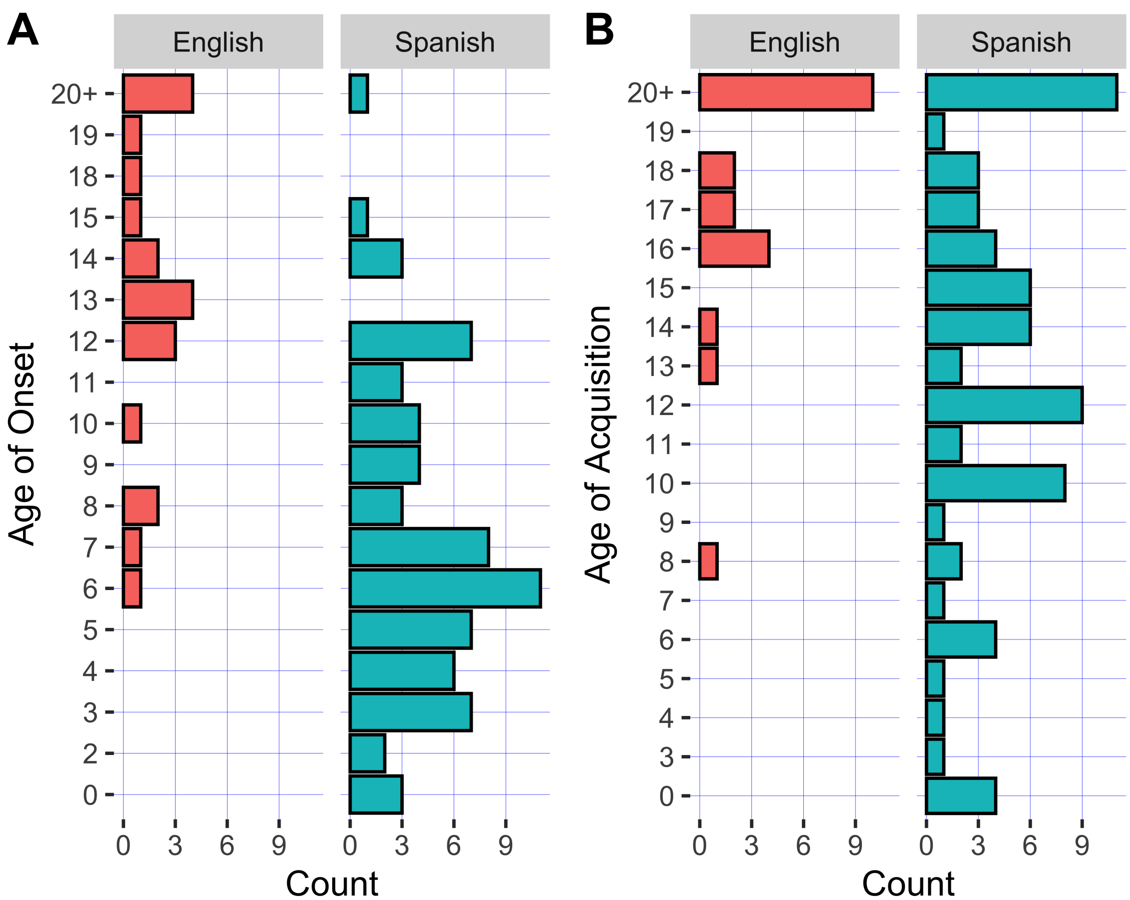
4.2.1 Participants

A total of 172 participants participated in the present experiments and made up a total of 4 groups: L1 English bilingual (n = 20), L1 Spanish bilingual (n = 66), L1 English monolingual (n = 58) and L1 Spanish monolingual (n = 28). Just as in the perception experiments, the all participants completed the Bilingual Language Profile (Birdsong et al., 2012). Figure **??** shows the age of Onset (A) and age of acquisition (B) by the L1 English and L1 Spanish bilingual groups. Figure 4.1 shows the self-rated proficiency of the same groups. All participants who answered 'no' to the question "Do you speak a language other than English and Spanish" were permitted to continue the experiment.

4.2.2 Materials

4.2.2.1 Pseudoword Reading Task

All participants completed a psuedoword reading task in Spanish and English in the event that they knew both languages. The pseudowords were comprised of the same vowel sounds tested in Chapter 3 in bilabial and fricative frames (/pVf/, /fVf/) and repeated 3 times each per frame. As a result, each participant produced 24 tokens in English and Spanish (4 vowels x 2 frames x 3 repetitions). In the same screen, audio was submitted using a widget, which contains a button to record, playback, and submit. The playback button allowed for participants to verify that they were submitting working audio. The English vowels /a/, /i/, /o/, /ʌ/ were solicited via


A
English
Spanish
Age of Onset
20+
19
18
15
14
13
12
11
10
9
8
7
6
5
4
3
2
0
0
3
6
9
Count
B
English
Spanish
Age of Acquisition
20+
19
18
17
16
15
14
13
12
11
10
9
8
7
6
5
4
3
0
Count

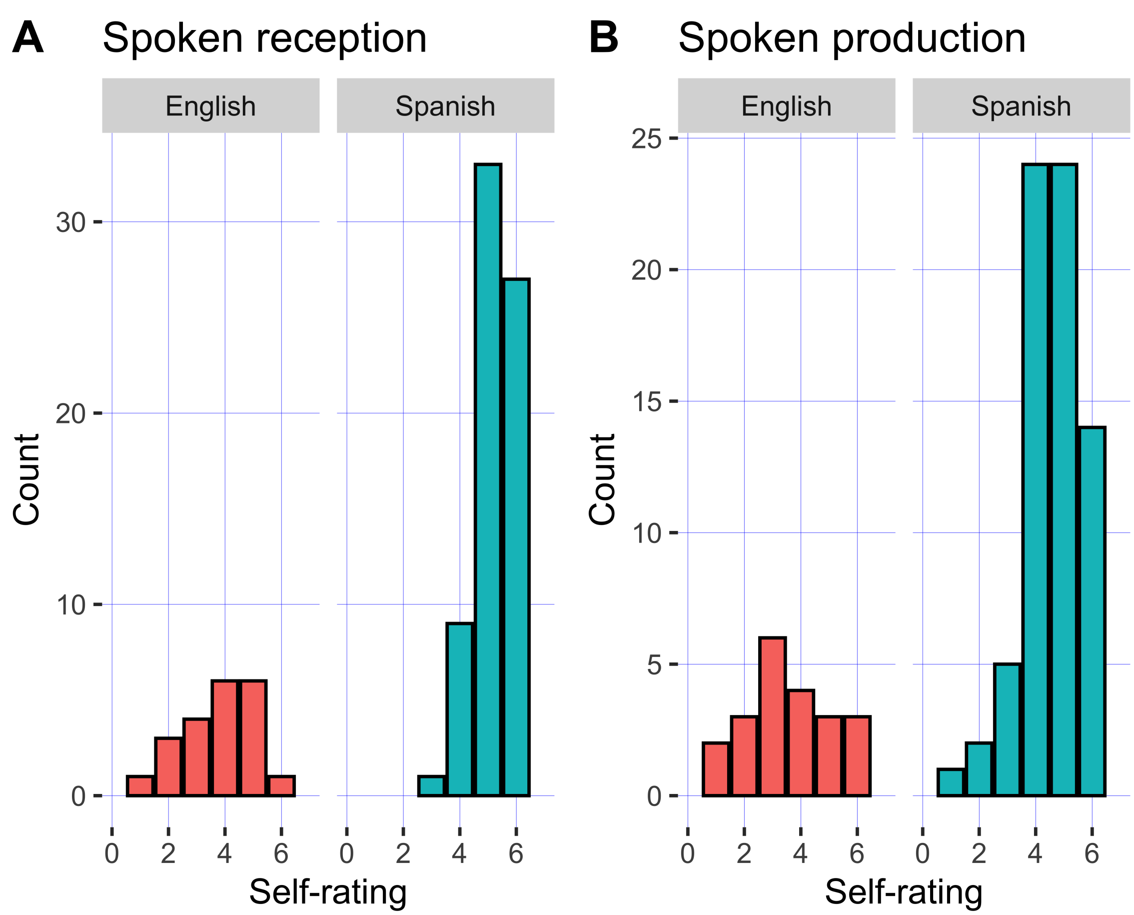


Figure 4.1: Self-rated spoken and eceptive proficiency by the L1 English and L1 Spanish bilingual groups

the following orthographic prompts: faff, fiff, foff, fuff. The Spanish vowels /a/, /i/, /o/, /u/ were solicited using faf, fif, fof, fuf (or p-initial minimal pairs for each of the pseudo words in both languages).

4.2.2.2 Word Repetition Task

In order to elicit French and German psuedowords, the same tokens used in the elicited production task in Chapter 3 (described in section 3.2.2.2) were provided to participants to repeat. During a given trail, participants saw an orthographic representation of a pseudoword on their screen, together with a similar audio widget to the pseudoword reading task which allowed the participants to play the audio production of the given word at will. Participants could play this audio as many times as they wished. In the same screen, audio was submitted using a similar widget, which contains a button to record, playback, and submit. The playback button allowed for participants to verify that they were submitting working audio. Again, four vowel contexts per condition were produced in two frames : /pVf/, /fVf/. As a result, the French vowels /ʌ/, /i/, /o/, and /y/ were represented in the psuedocarrier words: faf, fif, fof, fuf, paf, pif, pof, puf. The German words German: /ʌ/, /i/, /o/, /y/ were elicited in the pseudocarrier words: faf, fif, fof, fuf, paf, pif, pof, puf.

4.2.3 Procedure

Language order was randomized for each participant, and all languages were produced in the same session and administered using Labvanced in browser via prolific.co. Participants were only allowed to participate via laptop or desktop computer and were instructed to use headphones or high quality speakers with a microphone in a quiet, distraction free environment. Participants completed the Bilingual Language Profile in a separate session.

4.2.4 Statistical Analysis

A series of Bayesian multilevel regression models were run to determine whether there were differences in vowel height (F1) an and frontness (F2) between groups. In each model, the formant was analyzed as a function of language (4 levels for bilingual groups, 3 levels for monolingual groups), and phoneme (4 levels). Random intercepts for token and participant were also included to take into account the nested structure of the data. Additionally, varying slopes by participant were included in the model. Model priors were the default in `brms`, a student's T distribution with 3 degree of freedom. All models were fit with 4000 iterations (1000 warm-up). Hamiltonian Monte-Carlo sampling was carried out with six chains distributed between si processing cores.

4.2.5 Results

4.2.5.1 Forced Alignment

A total of 15019 data points were force aligned in the initial analysis using Webmaus basic (Schiel, 1999). In order to estimate the accuracy of the forced-alignment, about 3 percent of the bilingual data was randomly subset and hand-corrected (380 tokens). Subsetting was done using an R script. Figure 4.2 shows the quantity of each phoneme in each language that was hand-corrected in the random subset.

Figure 4.3 shows the formant values of hand-corrected tokens and force-aligned tokens. An inspection of the figure suggests that /a/, /ʌ/ and /o/ were mostly accurately aligned. On the other hand, the tokens for /i/ and /u/ benefited from hand-correction.

Figure 4.4 shows the root mean square deviation between the hand-corrected values and force-aligned values. Root mean square deviation is an measurement of the difference between predicted and actual values. In the present study, these numbers represent Hertz for F1 and F2. Each numerical label in the figure represents the root mean square deviation, while the colors represent phonemes. The phonemes /i/ and /u/ have the largest root mean square deviation, showing, again, that they were the

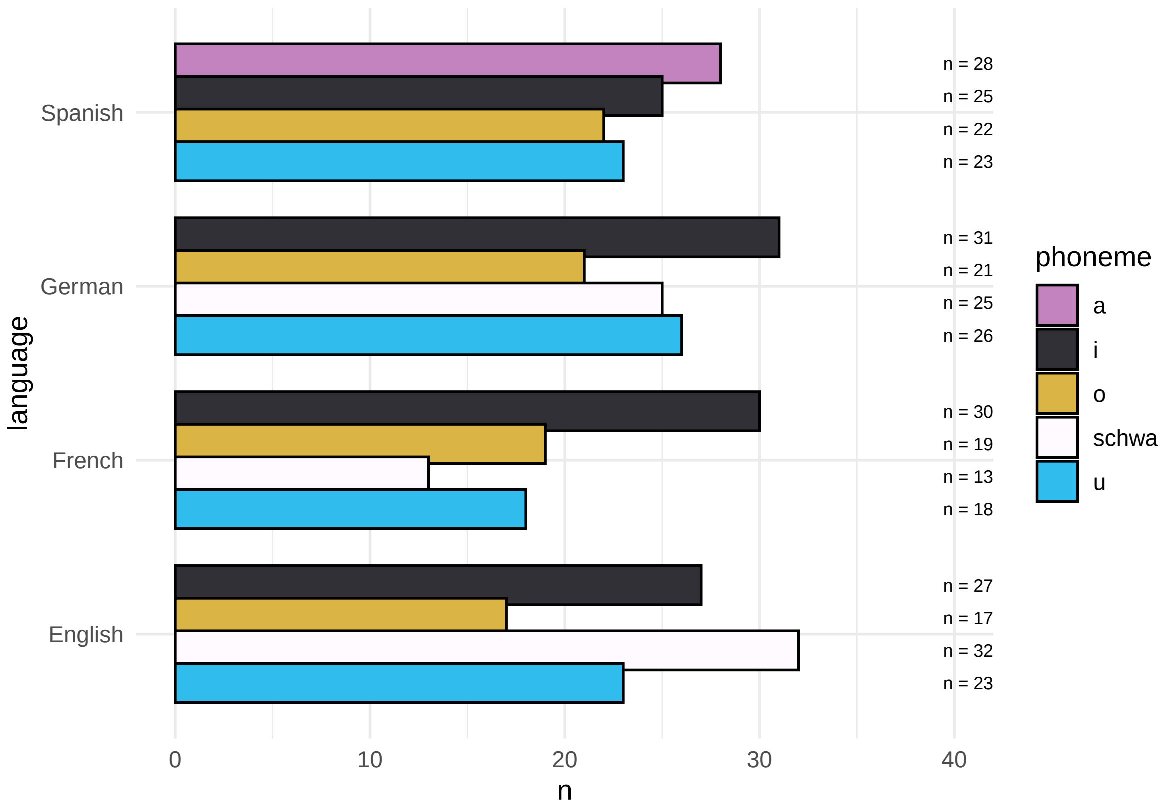


Figure 4.2: Quantity of Hand-correced tokens per phoneme per language

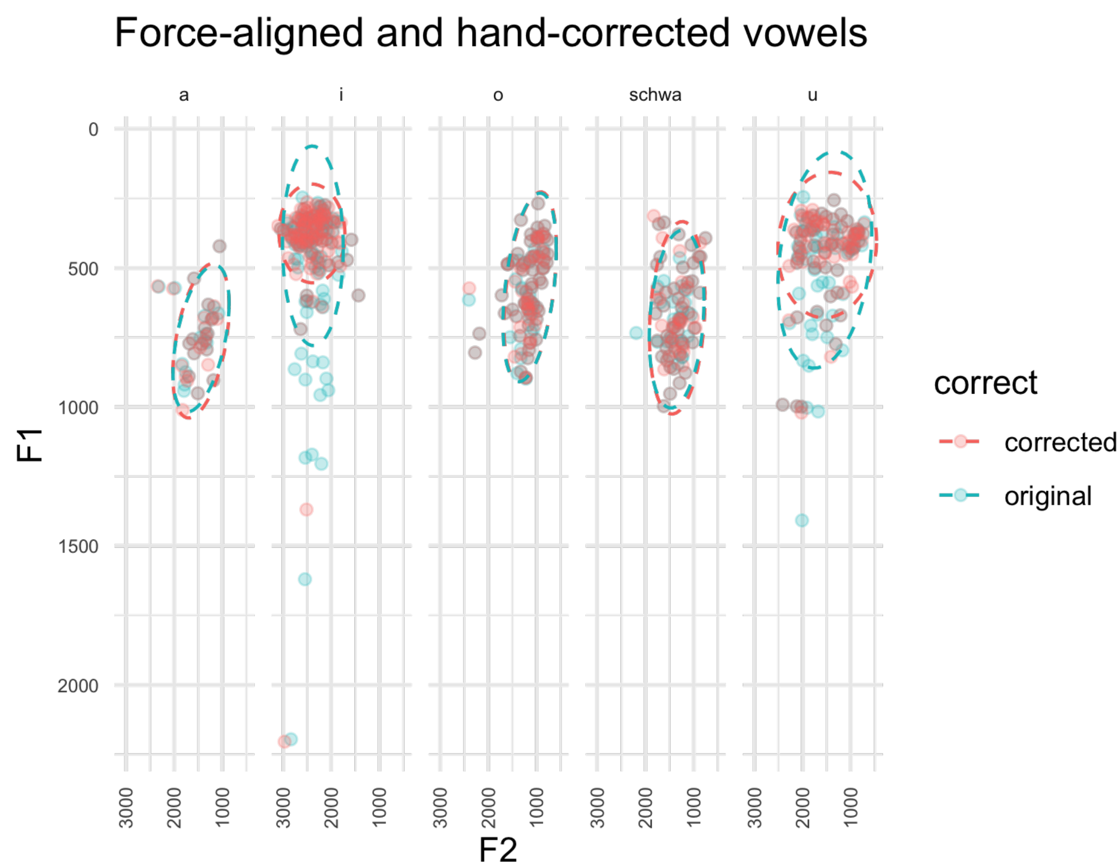


Figure 4.3: Hand Corrected versus automated formant values of a subset of the data

least accurately force-aligned. In order to account for poorly aligned tokens, outliers were identified and removed from the data set.

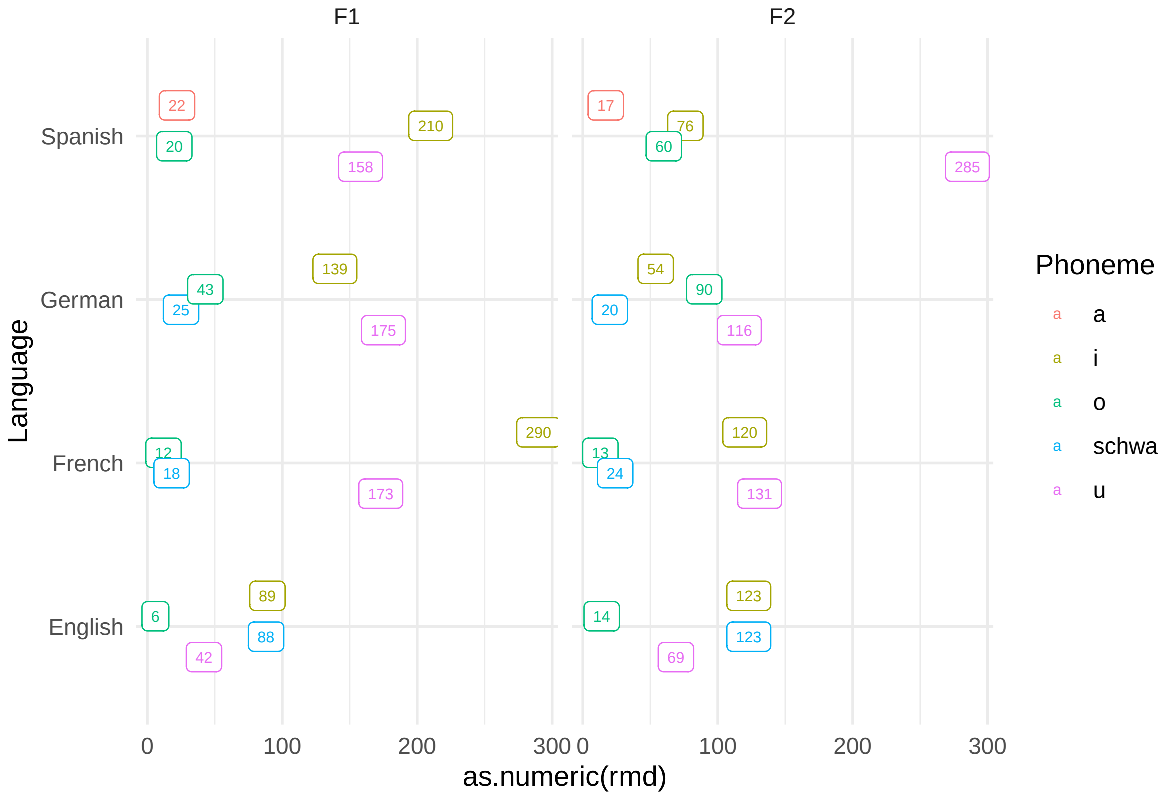


Figure 4.4: Root Mean Square Deviation for F1 and F2

A total of 2005 tokens were removed from the dataset. A token was considered an outlier when it was more than plus or minus two standard deviations of the mean of the hand-corrected subset. Table A.1 in the appendix shows the mean and standard deviations of each phoneme in each language used to identify outliers in the main dataset. Figure 4.5 shows the number of removed tokens compared to the root mean square deviation. The most removed phoneme was German /o/, in which 371 tokens were removed as outliers. Table 4.1 summarizes the total number of each phoneme removed. Figure 4.6 shows the the final number of tokens analyzed after removal of outliers (n = 13014) broken down by language and phoneme.

Table 4.1: Total Number of each removed phoneme

phoneme	n
a	54
i	496
o	676
schwa	278
u	501

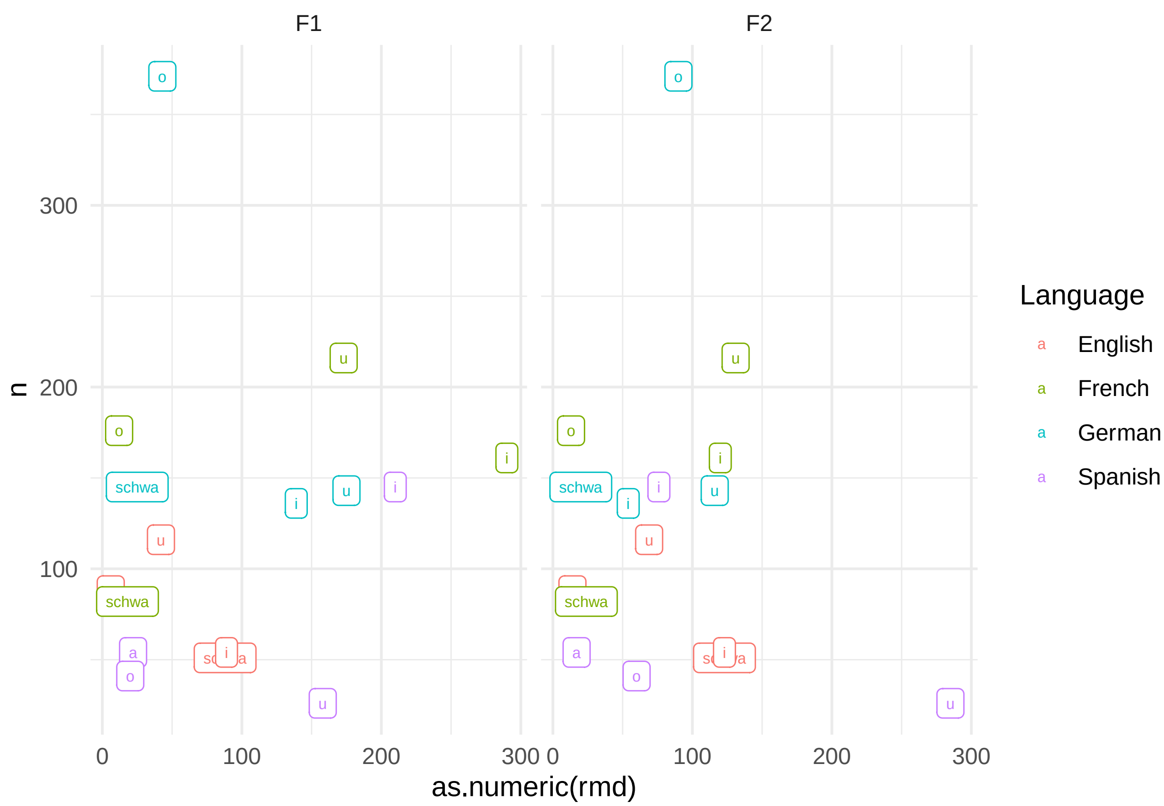


Figure 4.5: Root Mean Square Deviation for F1 and F2 versus total removed tokens

4.2.5.2 Analysis of the included data

Figure 4.7 shows the productions of each of the four groups for the phonemes /u/ (in Spanish and English) and /y/ (in French and German). The raw data points are

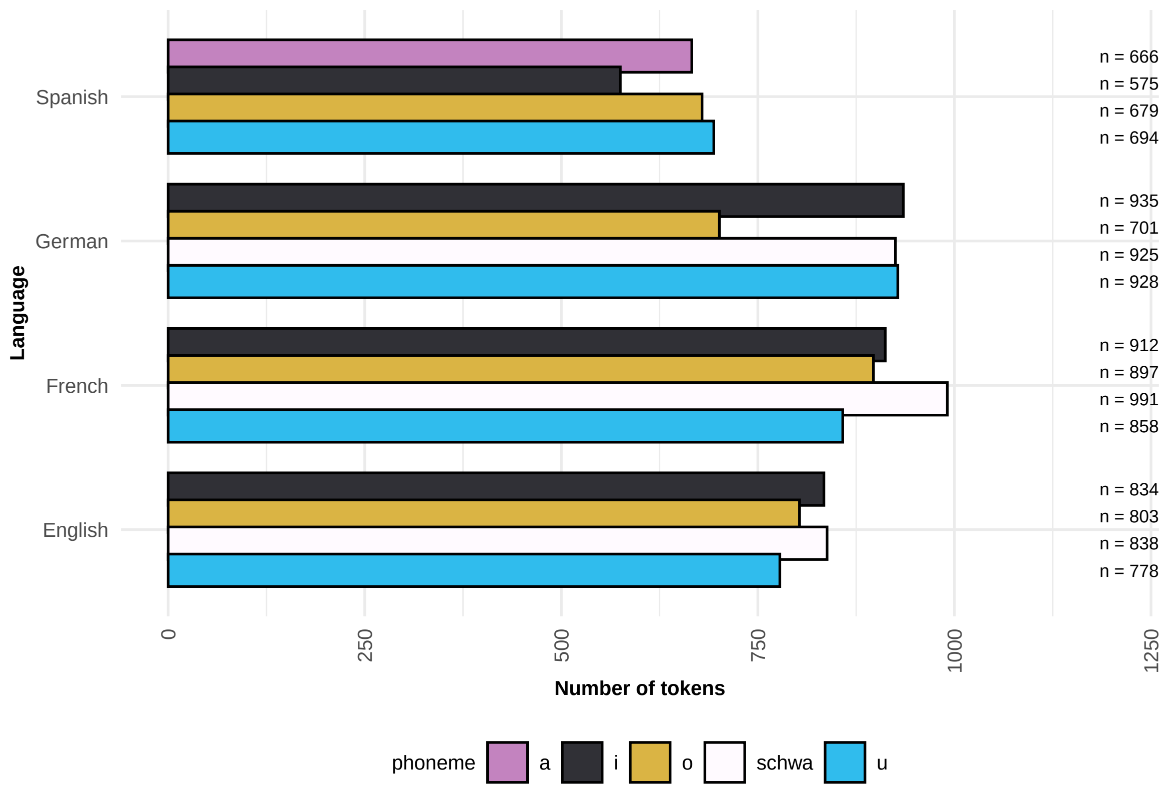


Figure 4.6: Total Number of Tokens per phoneme and language after removal of outliers

plotted together with the dashed circular shapes that represent a 95 percent confidence interval ellipse generated by the ggplot function in R. This condition was intended to explore how Spanish-English bilinguals would produce /y/, which is not present in either Spanish or English. An inspection of this figure suggests that /y/ was effectively imitated by all groups. In addition, it does not appear that there is much meaningful difference between the French and German imitations of the /y/ sound.

The models further supported the effective imitation of French and German /y/ by both bilingual groups. Figure 4.8 shows the conditional effects of the F1 and F2 models for /u/ in English and Spanish, and /y/ in French and German. Overall, the results of the models suggest that both bilingual groups performed similarly and could effectively imitate the French and German /y/. The models provided evidence that both bilingual groups had more fronted production of French (L1 English: 2353 [95% HDI `round(f2_eng_bil$lower__[3])`- 2442], L1 Spanish: 2261 [95% HDI `round(f2_eng_bil$lower__[4])`- 2315]) and German /y/ (L1 English: 2453 [95% HDI `round(f2_eng_bil$lower__[5])`- 2539], L1 Spanish: 2330 [95% HDI `round(f2_eng_bil$lower__[6])`- 2384]) relative to their Spanish /u/ (L1 English: 2349 [95% HDI `round(f2_eng_bil$lower__[7])`- 2439], L1 Spanish: 2300 [95% HDI `round(f2_eng_bil$lower__[8])`- 2353]) and English /u/ (L1 English: 2463 [95% HDI `round(f2_eng_bil$lower__[1])`- 2552], L1 Spanish: 2347 [95% HDI `round(f2_eng_bil$lower__[2])`- 2401]). Additionally, although the English /u/ was slightly more fronted than Spanish /u/ by both groups, there was compelling evidence that productions of English and Spanish /u/ and German and French /y/ were distinct in terms of frontness, since there was no overlap in the posterior distributions.

Figure 4.9 shows the productions of /o/ by all groups in Spanish, French and German, while the English sound was /ɑ/. Although this condition was intended to bias the use of Spanish, this visualization does not imply that this was successful, nor that there were many meaningful differences between groups. It appears that all groups distinguish and similarly imitate the French and German /o/ sounds, regardless of whether they are bilingual or monolingual, or whether they differently produce the

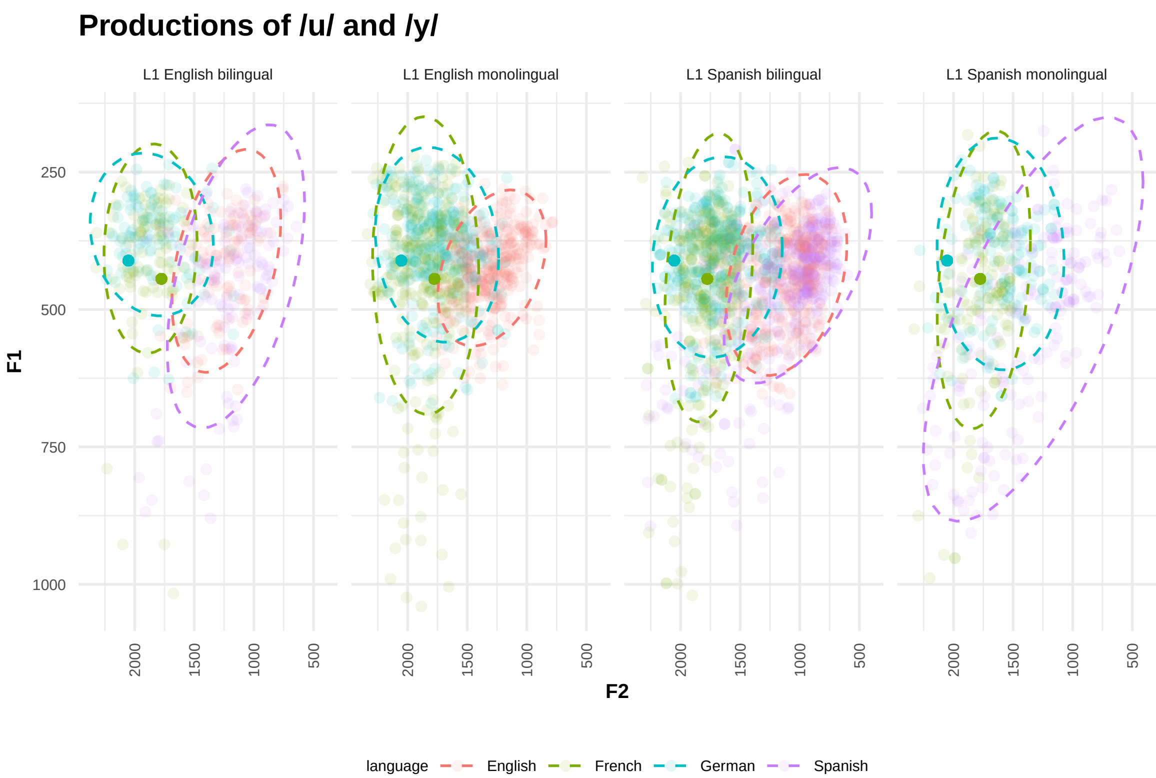


Figure 4.7: Productions of /u/ in each language by each group

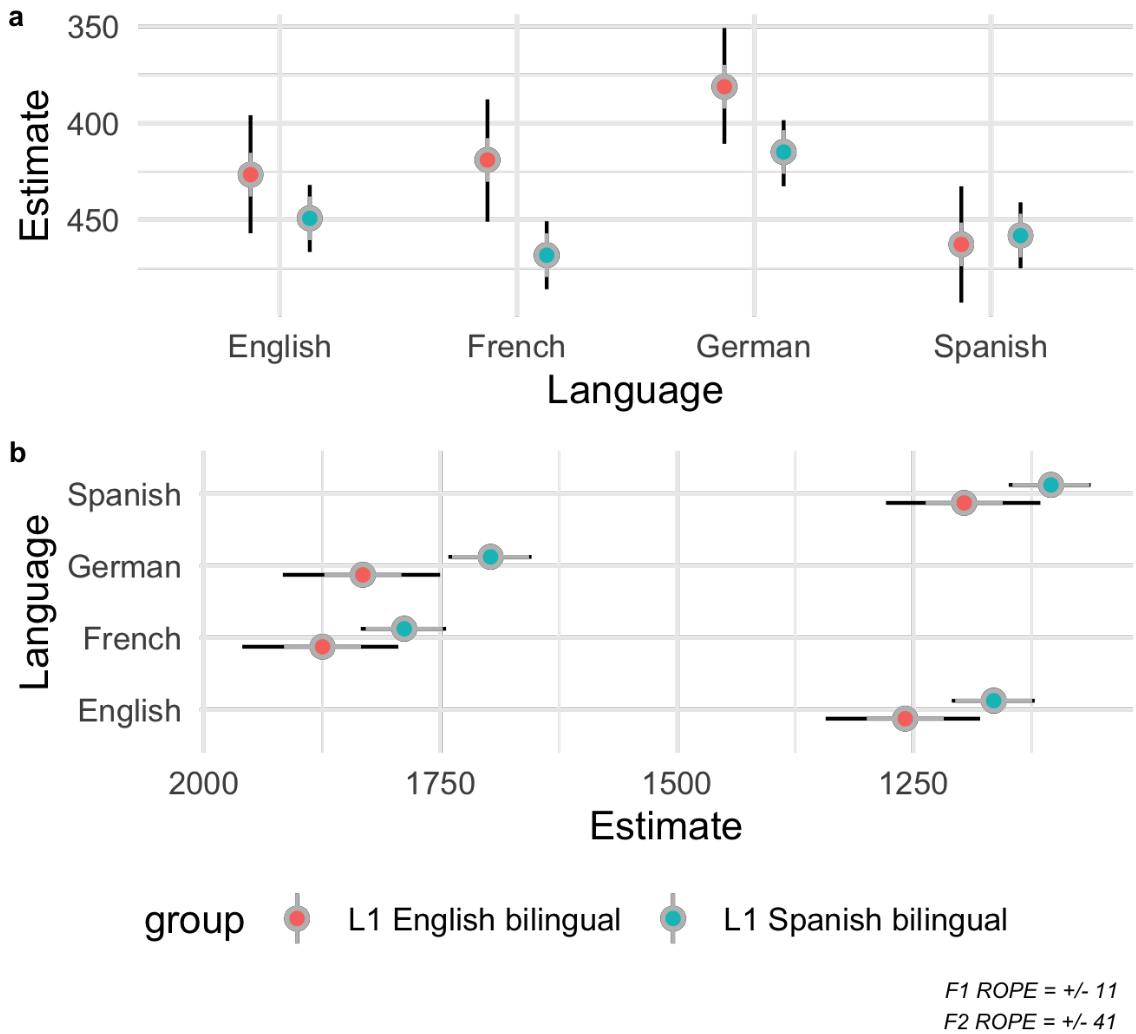


Figure 4.8: Conditional effects plot of /u/ for F1 (a) and F2 (b) in the bilingual groups

Spanish /o/ from the English /ɑ/. Figure 4.10 again shows the conditional effects of the Bayesian models for F1 and F2. The models shows that both bilingual groups produced English /ɑ/ and French /o/ as more fronted than Spanish and German /o/ (panel b), while height varied by both groups as a function of language (panel a).

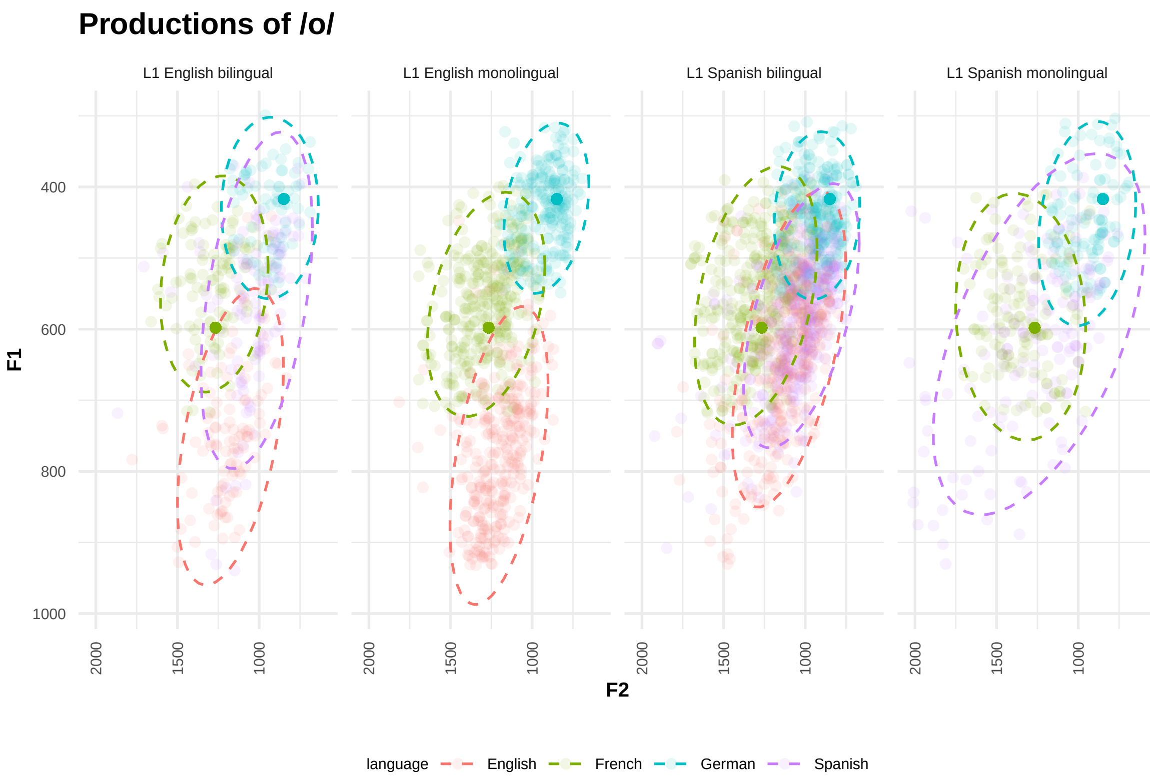


Figure 4.9: Productions of /o/ in each language by each group

Figure 4.11 shows the productions of /i/ in all languages and all groups. This condition was intended to examine how Spanish-English bilinguals produced novel language sounds that were present in both their L1 and L2. The figure does suggest that any of the groups produced /i/ distinctly as a function of language. The Bayesian analysis, however, revealed that the English L1 group had a more fronted /i/ in German and English than in Spanish and French (4.12, panel a). The Spanish L1 group showed less variation in /i/ fronting as a function of language (4.12, panel b). 4.12 also shows that both /i/ was slightly higher in German and Spanish by both groups.

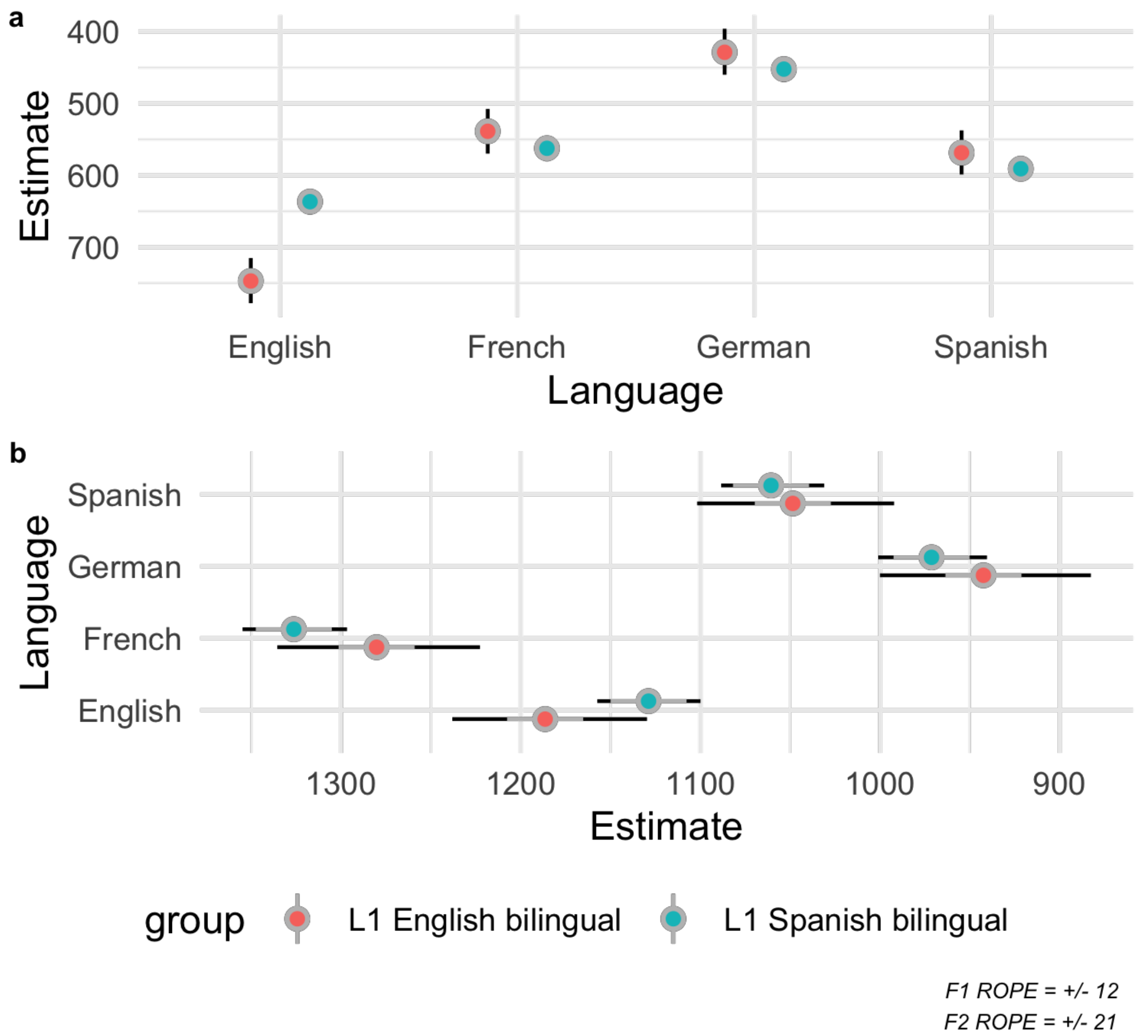


Figure 4.10: Conditional effects plot of /o/ for F1 (a) and F2 (b) in the bilingual groups

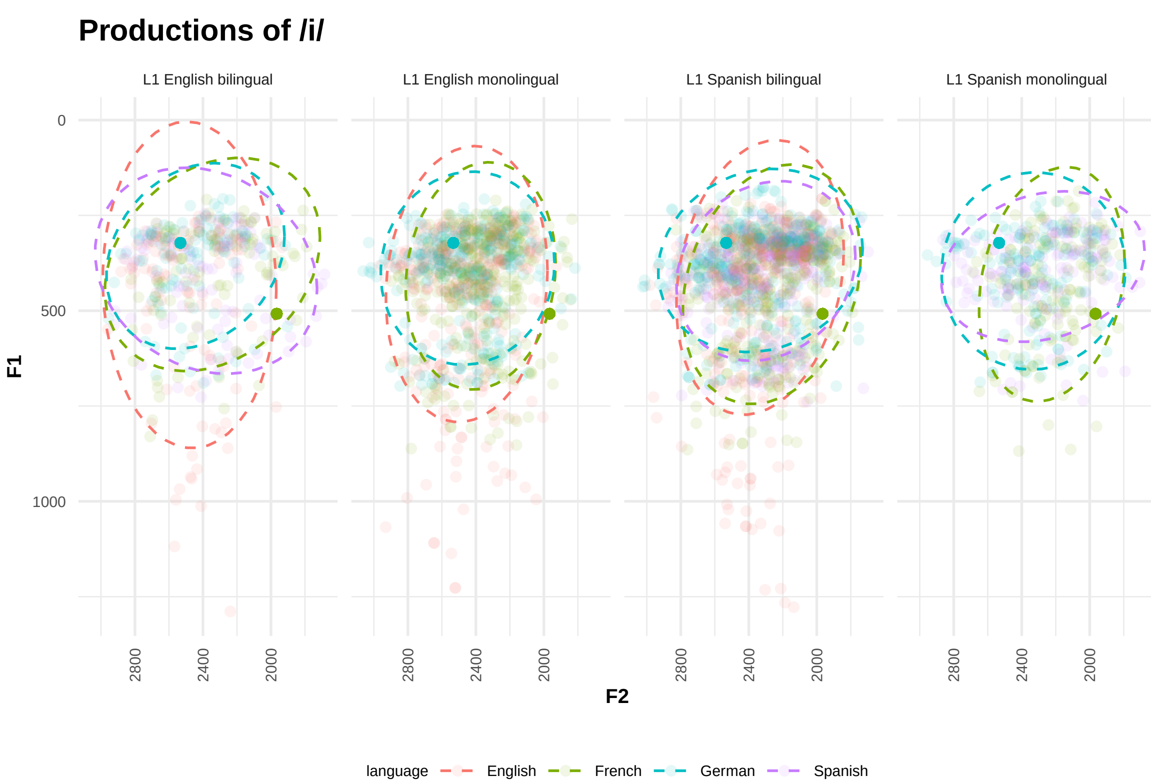


Figure 4.11: Productions of /i/ in each language by each group

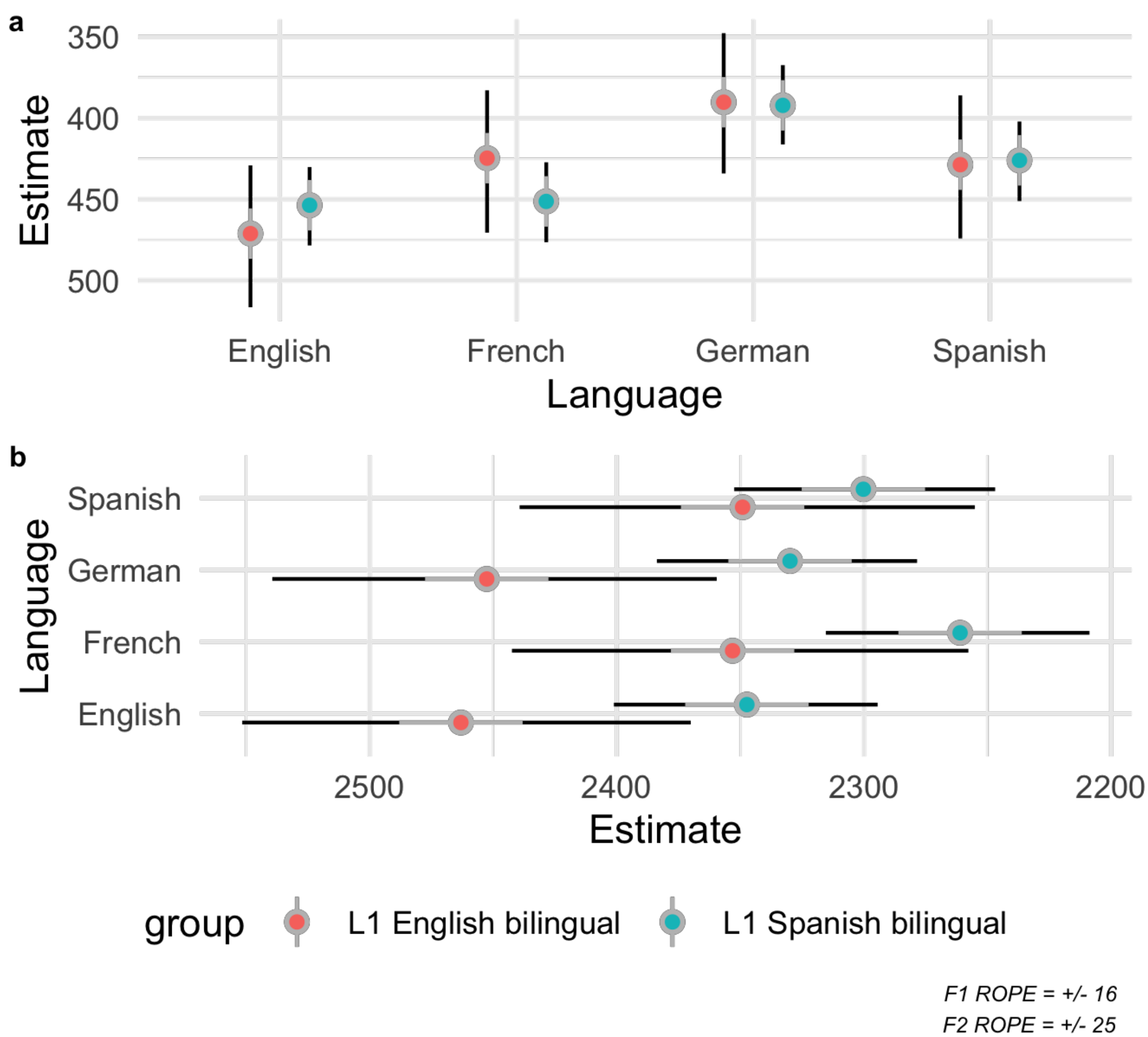


Figure 4.12: Conditional effects plot of /i/ for F1 (a) and F2 (b) in the bilingual groups

However, figure 4.13 shows the production of the /ʌ/ in English, French and German and /a/ in Spanish. This condition was intended to bias the production of English-like L3 vowels, with the Spanish /a/ for comparison. The results in this case are not consistent across groups. The L1 English bilingual group appeared to produce German /ʌ/, English /ʌ/, and the Spanish /a/ similarly. The L1 Spanish group produced German and English /ʌ/ similarly, but diverged slightly in their production of /a/. All groups produced French /ʌ/ similarly to one another, and dissimilarly from their own productions of other languages. 4.14 shows that both bilingual groups produced vowels similarly in a particular language when they produced /ʌ/ or /a/ (in the case of Spanish).

Figure 4.13: Productions of /a/ and the wedge in each language by each group

4.2.6 Interim Discussion

Based on the data of the present experiment, vowel imitation production appears to be relatively easy for bilingual and monolingual participants alike. When these speakers

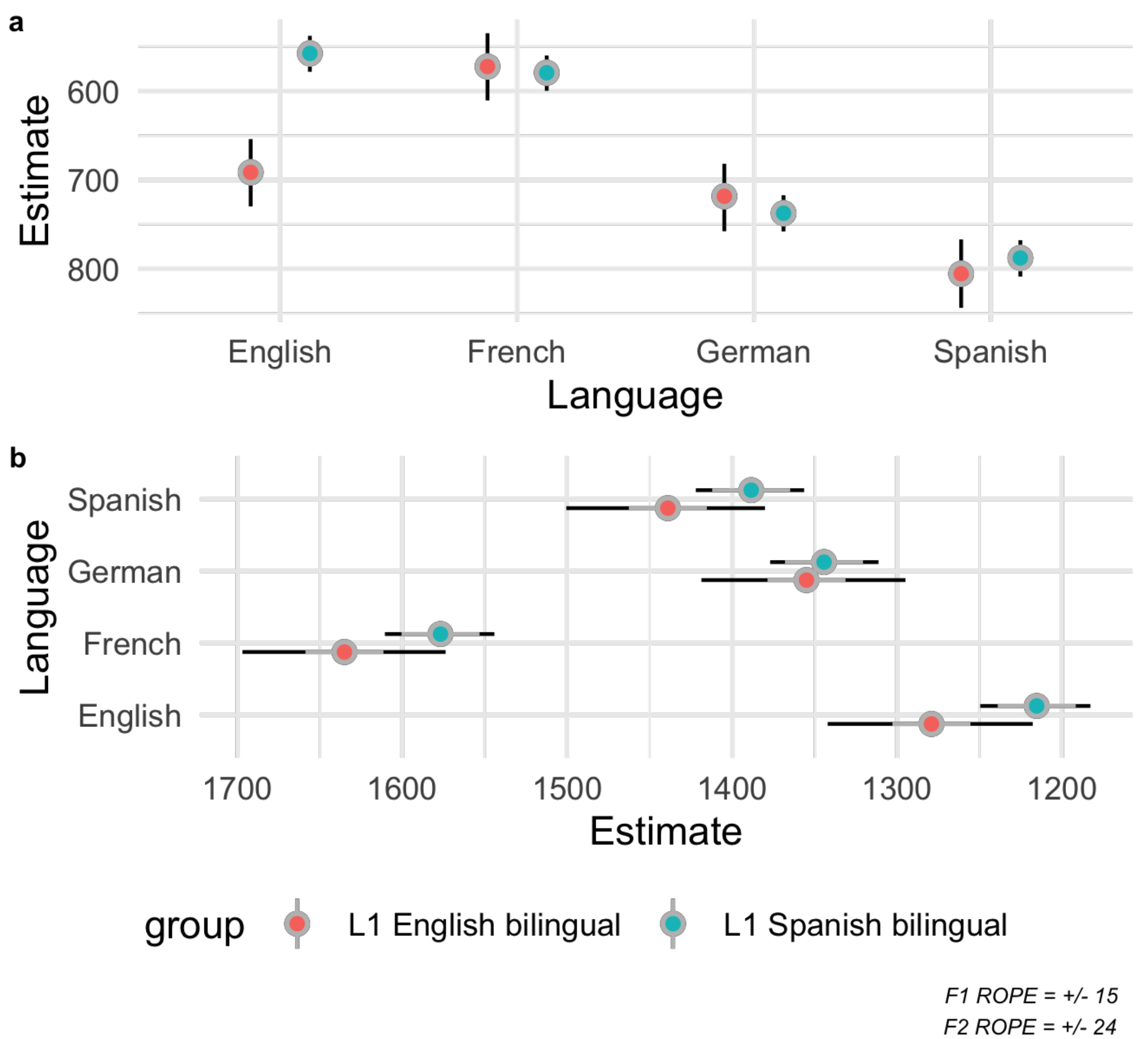


Figure 4.14: Conditional effects plot of the wedge for F1 (a) and F2 (b) in the bilingual groups

encounter a new language sound that is not phonemic in their L1 or L2, at least with the German and French /y/, they appear to have no problems imitating this sound, and do not assimilate the sound to their /u/ or /i/ categories in production.

These results have implications for L3 models, which largely cannot account for the results reported here. The TPM is largely not supported by the preset data, since it would predict that in the case of a new sound like /y/, the initial state of L3 acquisition is the final state of either a participants L1 or L2, and thus should be guided by L1 or L2 categories. This was clearly not the case, since participants appear to effectively imitate /y/ at first exposure, regardless of whether it is contained in a German or French sound. That is, in terms of vowels, there appears to be no effect of global structure similarity or typology. L2 status also cannot account for these results, since there were no meaningful differences found between bilingual groups and because the L3 sounds did not appear to be assimilated to L2 sounds. Finally, the LPM also has trouble explaining successful production of /y/ at first exposure. The LPM does not make specific predictions regarding sounds which are not present in the L1 or L2 in the L3, but would likely not predict successful imitation at first exposure, given that L1 and L2 categories are thought to impact the L3.

4.3 VOT Production

In addition to the analysis of vowels, the p-initial tokens from the pseudoword reading and pseudoword repetition tasks were analyzed for VOT. Half of the tokens for each participant began with /p/ and totaled 12 per language (4 vocalic contexts x 3 repetitions). Unlike vowel production, a subset of the stop consonants were force-aligned and subsequently hand corrected.

Previous research in multilingualism has found that it is difficult to find a group of multilingual participants who produce sounds between languages distinctly enough to determine which language they could have come from, although they may have high proficiency. For example, when examining L3 French production by Spanish-

English bilinguals, Parrish (2022) found that only a subset of participants produced English and Spanish distinctly and were included in the study. The inclusion of a given participant in Parrish (2022) was determined by segmenting all data from every participant and running a t-test with a participants Spanish and English. When this t-test returned a significant result, the participant was included.

Utilizing the context and insights of this previous research, the present dissertation subset participants in an effort to save resources in a replicable way. Rather than segmenting and subsequently discarding the acoustic data of discluded participants, the present dissertation used blinded judgment data by the author to determine which participants were producing English and Spanish that were distinct enough to be informative when compared to their French and German productions. In particular, an R script was used to find all sound files that began with /p/ and were produced in each bilingual participant's L2. From there, these files were copied to new directory and used to create a 2-alternative forced choice task in Psychopy. In total, 774 tokens were included, where 86 total participants (66 Spanish L1, 20 English L1) included 9 of their total 12 tokens each.

4.3.1 Subset Procedure

One at a time, each sound was played, and "English-like" and "Spanish-like" were displayed on the left and right sides of the screen. After the sound was played, the researcher chose whether the segment /p/ sounded like an English-like or Spanish-like production of that sound, not knowing whether the speaker was an English or Spanish speaker. Between trials, a cross was displayed for 500ms in the center of the screen.

4.3.1.1 Results

Participants were included when more than half of their utterances were judged correctly. A judgment was considered correct when it was judged as sounding English-like or Spanish-like for an L2 English or L2 Spanish speaker. For example, if a given production from an L1 English-L2 Spanish speaker was judged as Spanish-like, then it was coded (1) as correct. In total, 33 participants fit this criteria and had all of their data subsequently segmented (mean correct = 7.2 (SD = 1.5)), of which only 3 were L1 English speakers. Overall, 3 of the participants were L1 English speakers and 30 were L1 Spanish speakers. The data of an additional 5 participants were also removed due to poor audio quality, which made segmenting the voice onset time impossible or very difficult to do reliably. As a result, the final dataset contains 28 bilingual participants (25 L1 Spanish and 3 L1 English speakers)

4.3.2 Statistical Analysis

For the L1 Spanish group, a Bayesian multilevel regression model was run in which VOT was the outcome variable. In the model, language was the sole fixed factor (4 levels: Spanish, English, French and German). Random intercepts for token and participant were also included to take into account the nested structure of the data. Additionally, varying slopes by participant were included in the model. Due to the low number of L1 English participants, a single model was fit to each of them. That is, 3 additional models were fit where VOT was the continuous outcome variable and the fixed effect was language. In this case, only a random intercept for token was included. All model priors were the default in `brms`, a student's T distribution with 3 degree of freedom. All models were fit with 4000 iterations (1000 warm-up). Hamiltonian Monte-Carlo sampling was carried out with six chains distributed between six processing cores.

4.3.3 Results

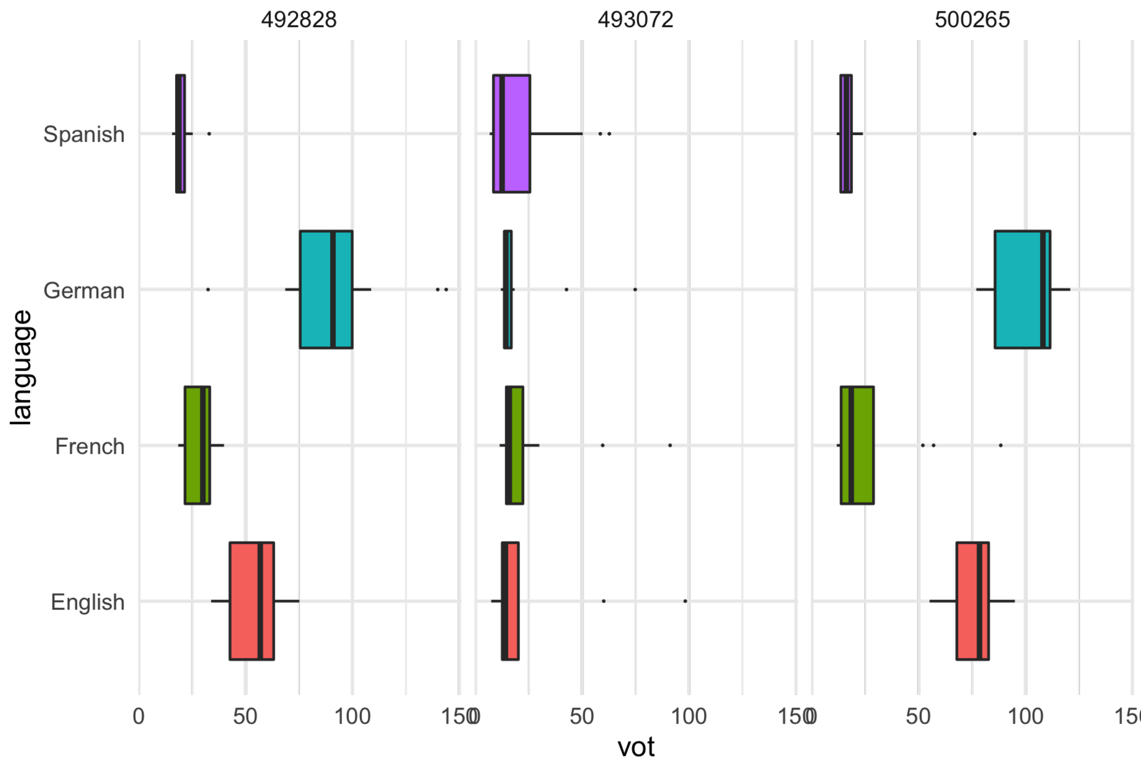
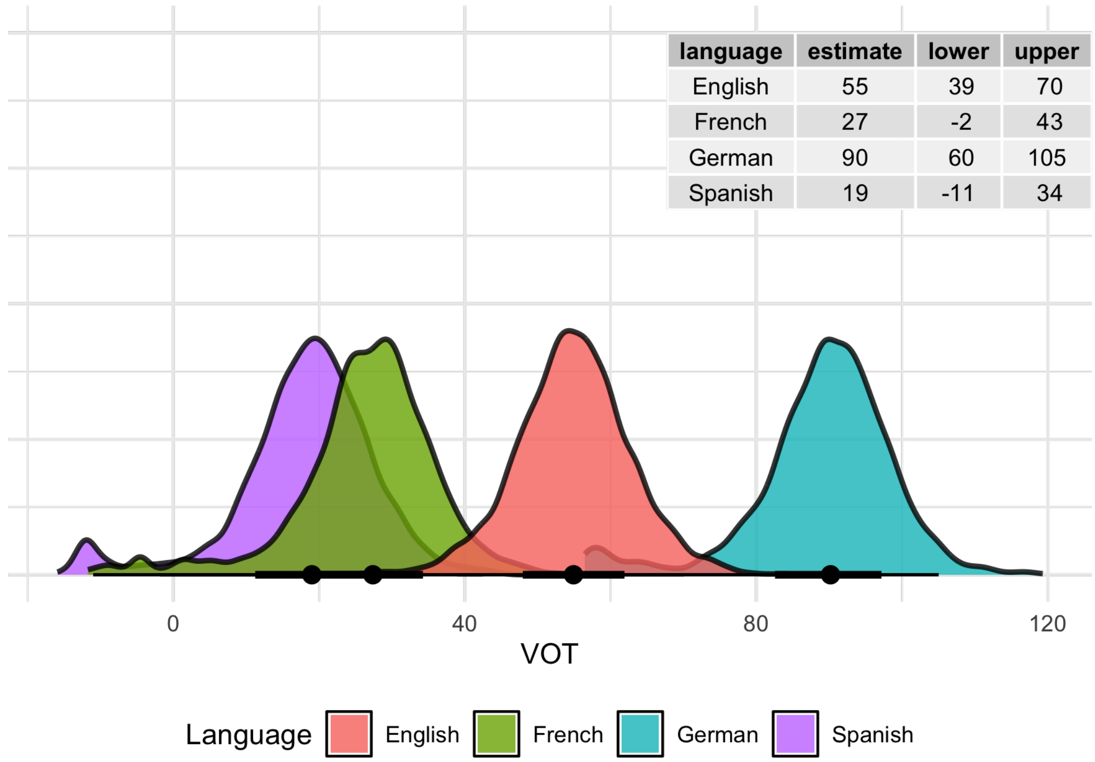


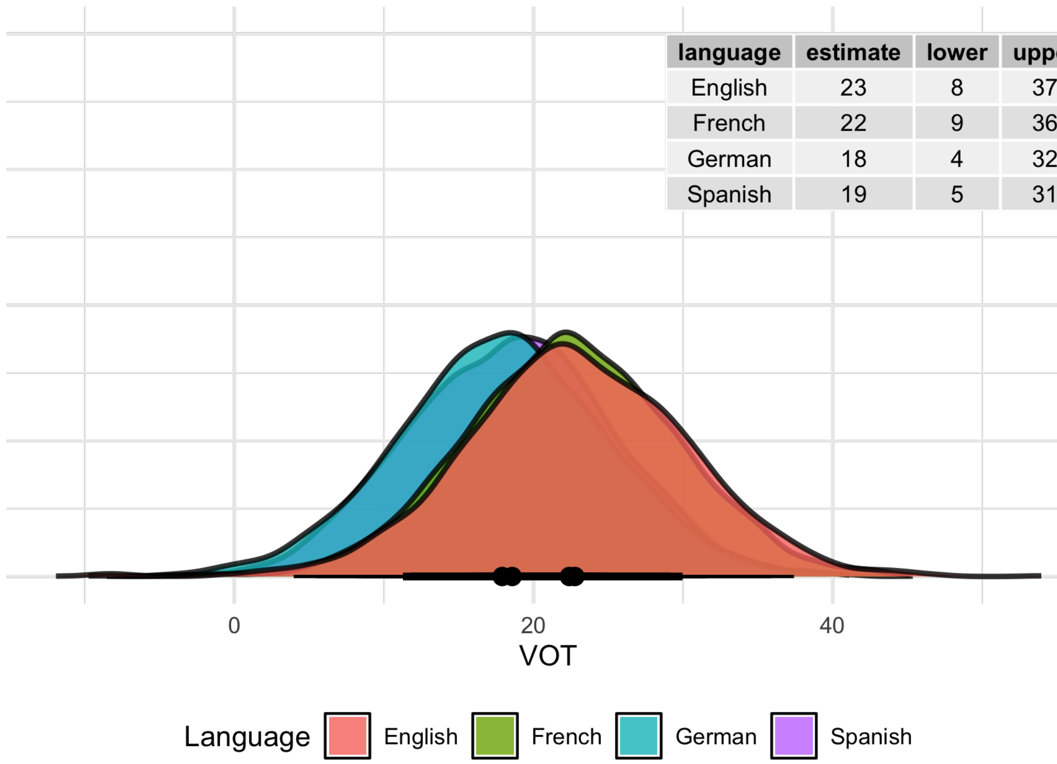
Figure 4.15: The VOT of /p/ for each individual L1 English-L2 Spanish speaker

4.3.3.1 L1 English Speakers



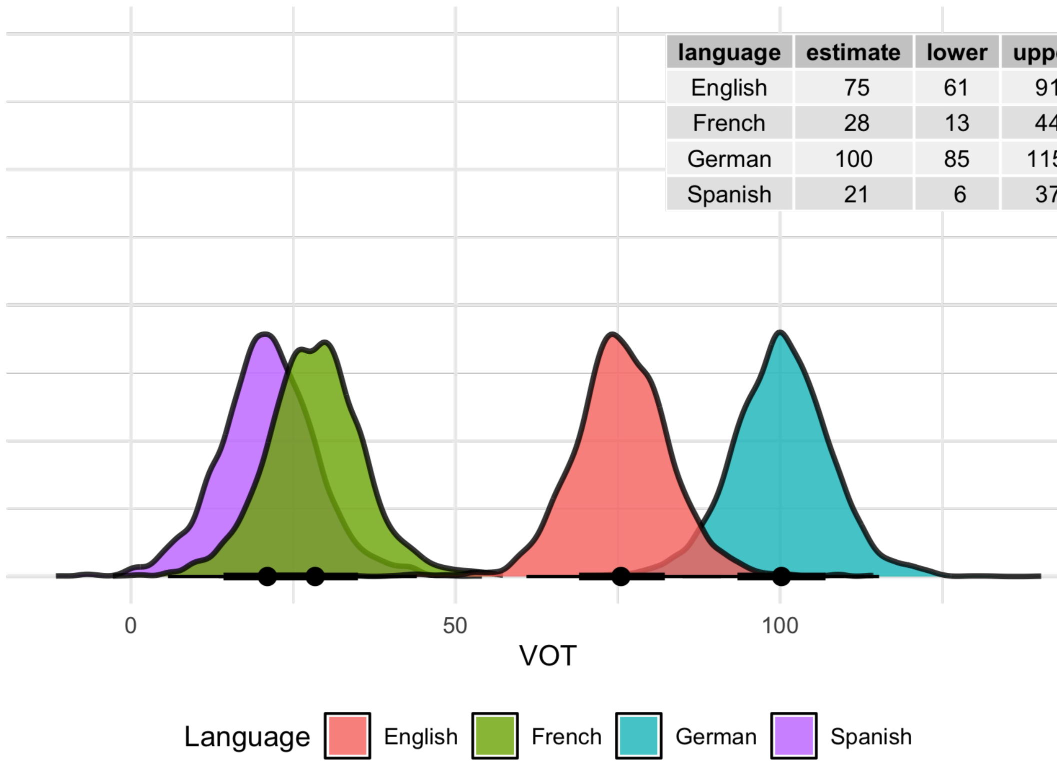
\begin{figure}

\caption{95% credible intervals of all plausible values of VOT as a function of Language for participant 492828} \end{figure}



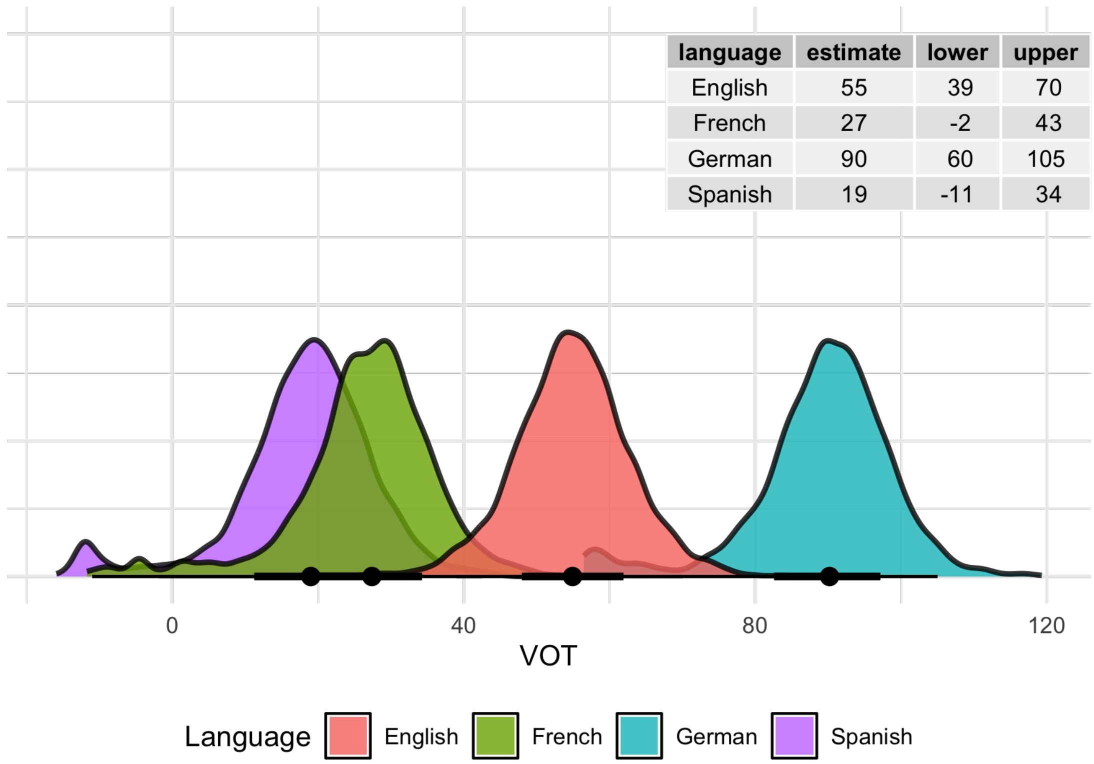
\begin{figure}

\caption{95% credible intervals of all plausible values of VOT as a function of Language for participant 493072} \end{figure}

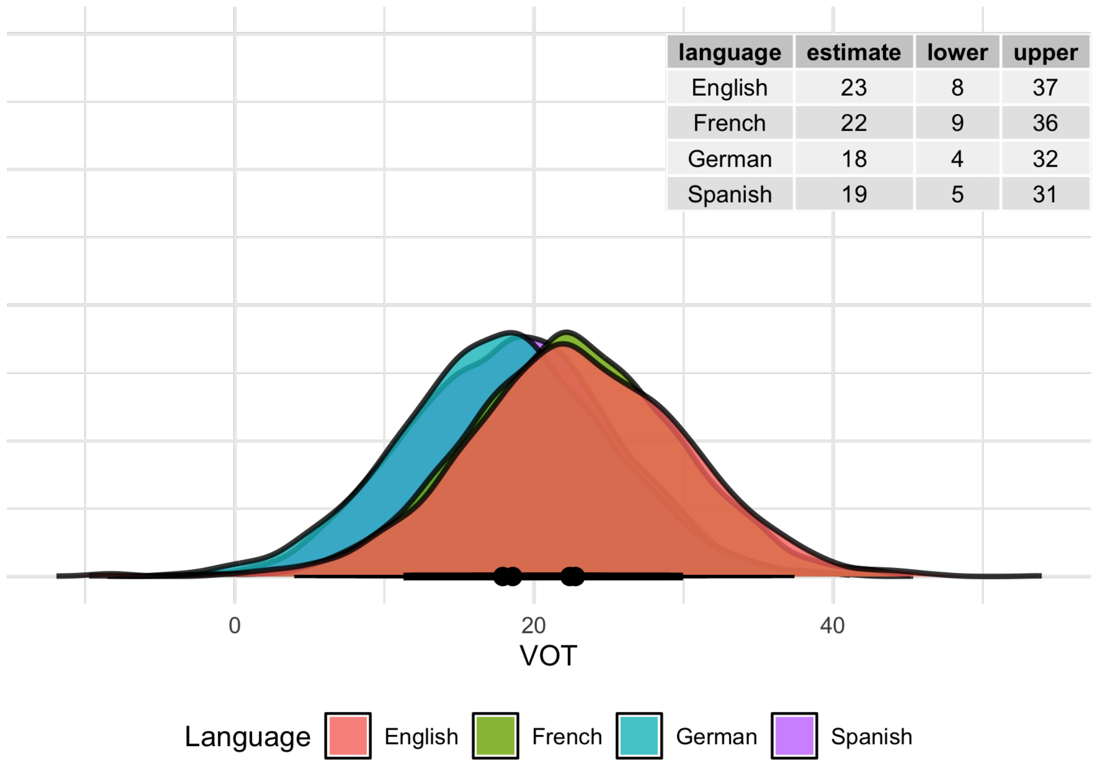


\begin{figure}

\caption{95% credible intervals of all plausible values of VOT as a function of Languagefor participant 500265} \end{figure}



language	estimate	lower	upper
English	55	39	70
French	27	-2	43
German	90	60	105
Spanish	19	-11	34


language estimate lower upper
English 23 8 37
French 22 9 36
German 18 4 32
Spanish 19 5 31
0
20
40
VOT
Language
English
French
German
Spanish

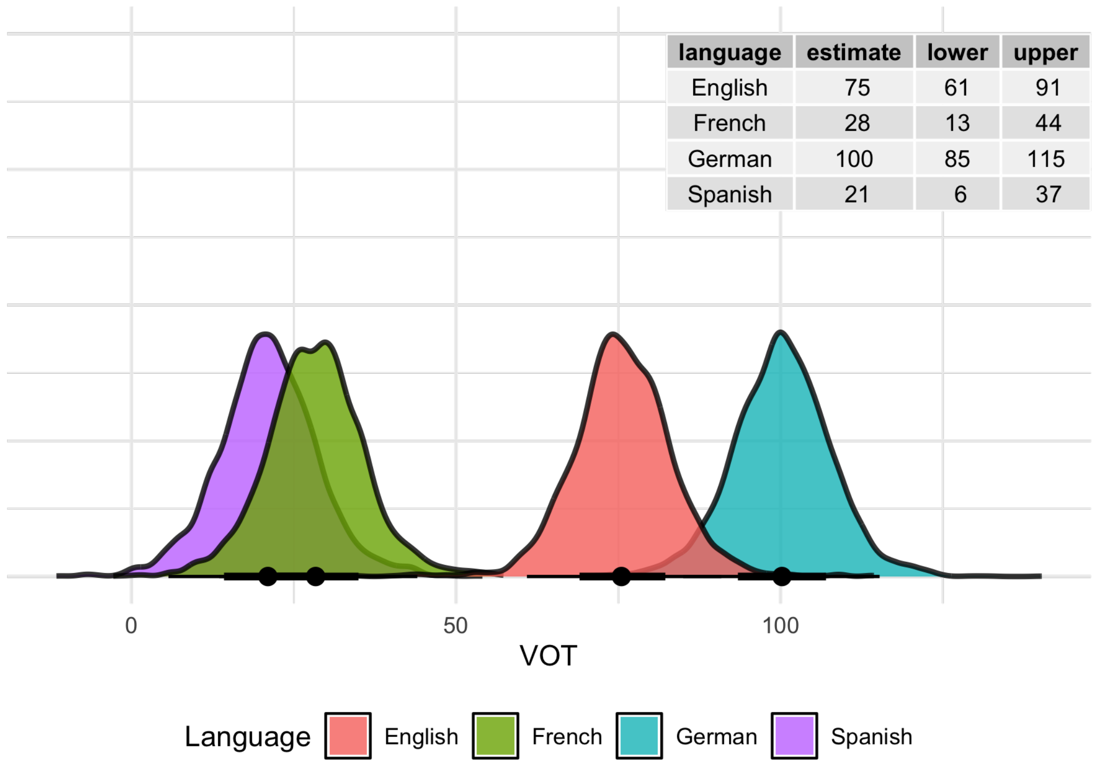
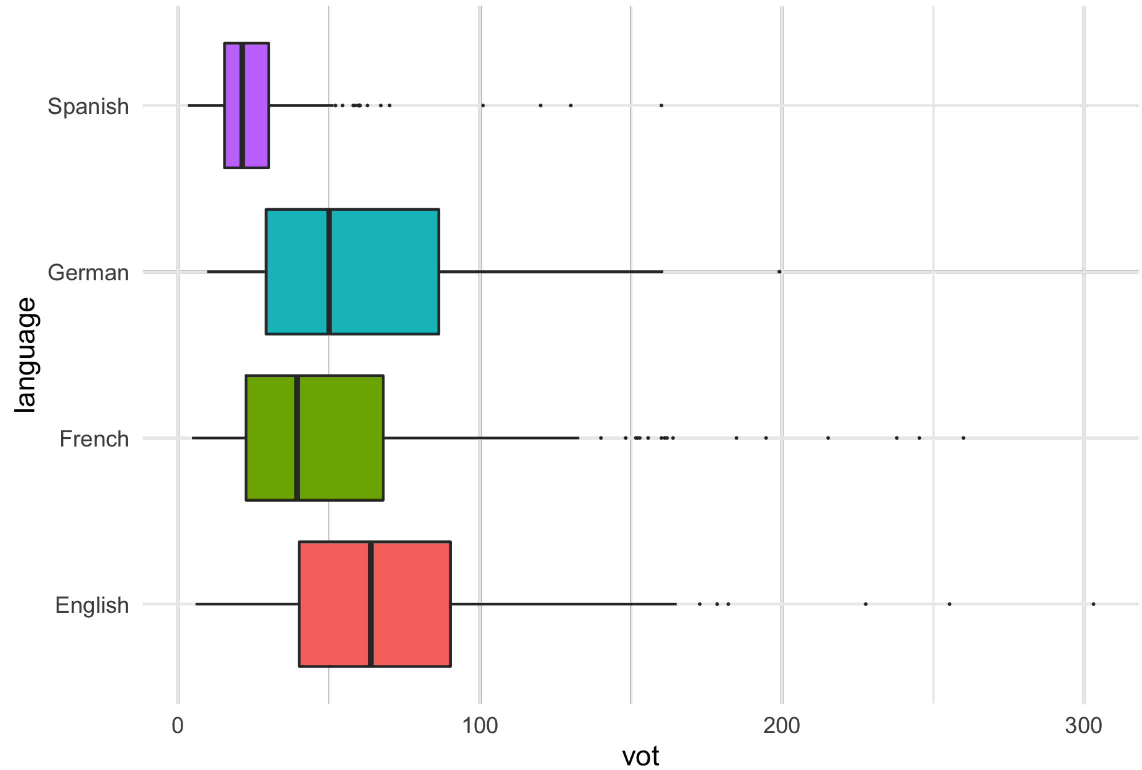


Figure 4.15 shows the VOT of each of the 3 included L1 English participants in all 4 languages. Figures 4.3.3.1, 4.3.3.1 and 4.3.3.1 show the 95% credible intervals for each L1 English speaker. In figure 4.3.3.1 it appears that the trend for participant 492828 is for French (27 [95% HDI -2 - 43]) productions to be more Spanish-like (19) [95% HDI -11 - 34]), while their German (90) [95% HDI 60 - 105]) productions were more English-like (55 [95% HDI 39 - 70]). Participant 500265 (figure 4.3.3.1) shows a similar trend. Their German (100 [95% HDI 85 - 115]) was closer to to their English productions (75[95% HDI 61 - 91]) and their Spanish was (21 [95% HDI 6 - 37]) closer to their French were (28 [95% HDI 13 - 44]). On the other hand, figure 4.3.3.1 suggests that participant 493072 did not produce distinct VOT as a function of language: their productions of English (23 [95% HDI 8 - 37]), Spanish, (19 [95% HDI 5 - 31]) French (22 [95% HDI 9 - 36]) and German (18 [95% HDI 4 - 32]) were all similar.

4.3.3.2 L1 Spanish Speakers

Figure 4.16 shows the VOT of each language at the group level for the L1 Spanish speakers.



4.3.3.2 shows the posterior distributions for each language in the L1 Spanish group. The range of plausible estimates for both French (54 [95% HDI 39 - 69]) and German (62 [95% HDI 47 - 77]) fell between English (71 [95% HDI 57 - 87]) and Spanish (25 [95% HDI 14 - 36]), with German being closer to English than French.

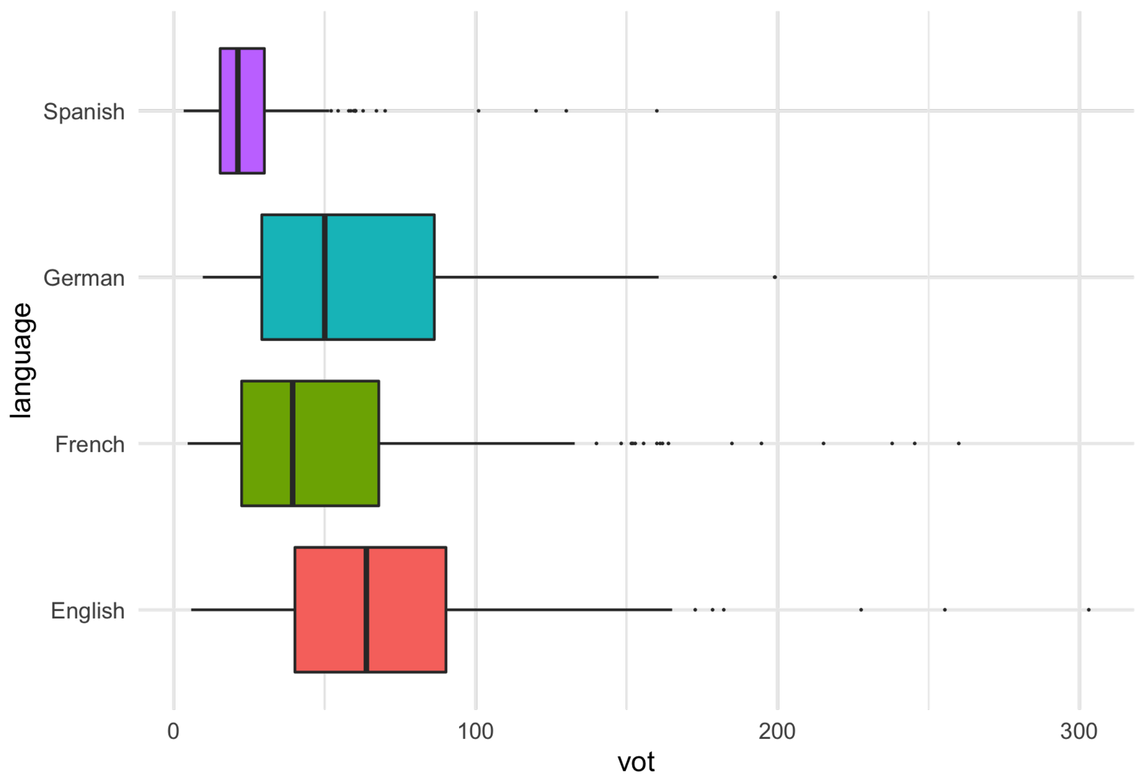
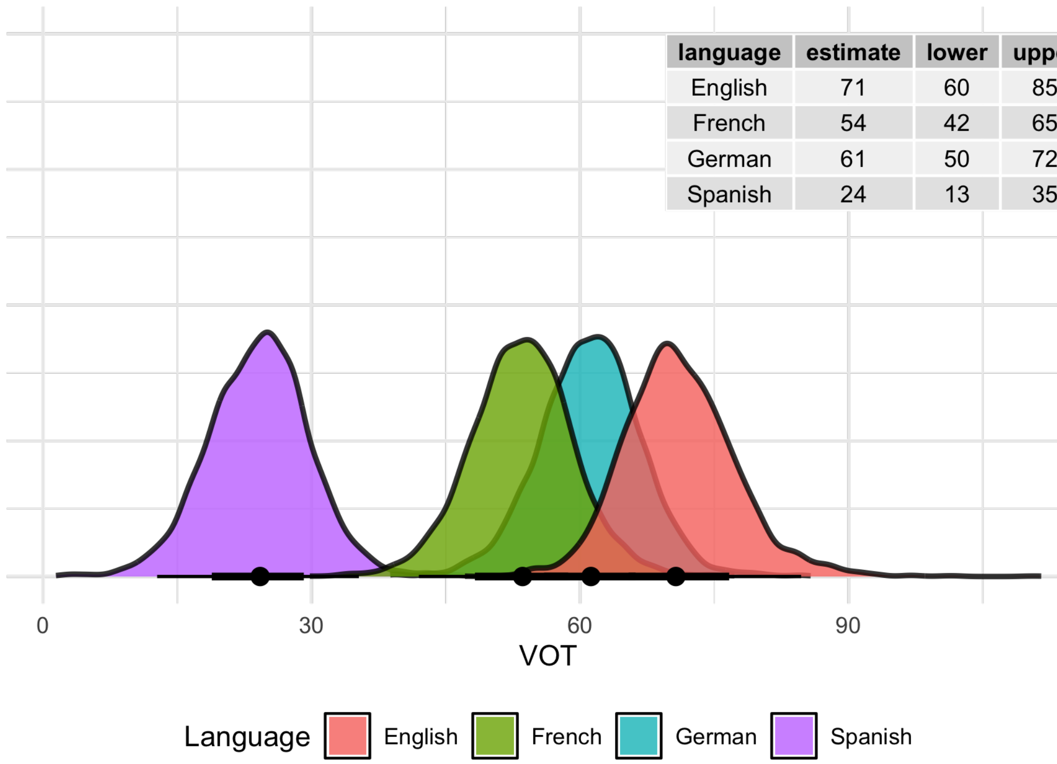


Figure 4.16: The VOT of /p/ for L1 Spanish-L2 English group



\begin{figure}

\caption{95% credible intervals of all plausible values of VOT as a function of Language} \end{figure}

4.4 Discussion

These results replicated previous work (Parrish 2022), which found that L3 French VOT was intermediate of L1 and L2 VOT when those languages are Spanish and English. Unlike previous research, the present work gave initial exposure learners two "L3s" to repeat, making a within subjects comparison possible. The German VOT results (62 [95% HDI 47 - 77]) were higher on average than French, (54 [95% HDI 39 - 69]) but still fell intermediate of L1 (25 [95% HDI 14 - 36]) and L2 (71 [95% HDI 57 - 87]) values in the L1 Spanish group. Taken together, the results in VOT suggest that, while both the L1 and L2 influence L3 production at first exposure, the phonetic properties of the L3 that is being learned should also be taken into account.

At the same time, the results for VOT and vowels were distinct. Largely, successful imitation was obtained for novel vowel formants, while VOT was far less successful. There are at least three possible reasons for this finding. First, it is possible, since the VOT of stop consonants is less salient than vowel quality, that VOT distinctions fall below the level of consciousness and are therefor harder to imitate at first exposure. Secondly, the combination of task demands and differences in salience of the chosen segments may have impacted the results. There is evidence that more salient segments are more easily shadowed than less salient ones (Podlipsky & Simácková, 2015), and that increased task demands can negatively impact performance on an experimental task (Declerck & Kormos, 2012). For example, Declerck & Kormos (2012) found that, when comparing single and dual task demand conditions in L2 speech production, that there was a negative impact on accuracy and lexical selection for the dual task demand condition. The authors argue that this occurs due to an attentional "bottleneck" effect, in which attentional resources are finite and impacted by task demands. It is possible that a similar situation occurred in the present study, where the participants had to attend to both a novel vowel and stop consonant in an unfamiliar language in the same word. In this situation, given that it is generally agreed that vowels are more salient than voiceless stop consonants, it is possible that more attentional resources were paid to the more salient feature of the novel words. In other words, if VOT were the focus and the vocalic context was held constant, it is plausible that more target-like VOT would be produced. Just as with vowels, the present data on stops informs L3 models, and adds to the body of research that partially supports the the claim of the LPM that L3 learners maintain access to both their L1 and L2. Several previous studies have reported L3 VOT values that were intermediate of the L1 and the L2 (Blank & Zimmer, 2009; Wrembel, 2011, 2014; Wunder, 2010) and involve a range of language combinations and ranges of proficiency. The present study adds initial exposure and Spanish-English bilinguals learning both French and German to the body of work that has been done. In doing so, the present work strengthens the idea that L3 learners begin access to their L1 and L2, and that access to a particular language does not appear to be completely inhibited for stop consonant representations. As a result, the TPM and L2 Status Factor cannot readily explain intermediate VOT that has been found here or in previous research, since these models would predict whole language influence of either the typologically closer language or the second learned language respectively.

Taken together, the results of the same participants' productions in stops and vowels suggest that L1 and L2 influence are not consistent across segments. This conclusion, however, should be interpreted with caution. The participants of the present study simply heard and imitated a new language sound, and were not actually in the process of acquiring a third language. The purpose of including these participants was to examine the starting point of L3 phonology. Future research could benefit from using a longitudinal design and multiple segments to more fully support that idea that L3 learners are differently impacted by the L1 and L2 depending upon the segment and how this may change during L3 development. Past research has reported a difference in L1 or L2 influence on L3 production as a function of test time. In their seminal case study, Williams & Hammarberg (1998) reported heavy L2-influence in L3 production by their speaker, but, upon re-examination after 6 months in the L3 country the speaker's L3 speech was rated as heavily L1-accented. Using a similar design to Williams & Hammarberg (1998), but using multiple segments and more subjects would create an ideal situation to examine L3 development as it relates to influence from previously learned languages. This type of design would also help to reduce the uncertainty brought about by conflicting findings in L3 research, such as those reported in VOT studies, since it would utilize the same participants and the same design at two different points in time. As a result, if differences in source language influence were observed, it could not be due to differences between experimental design or elicitation techniques, which could be the case between two different cross-sectional experiments.

4.5 Conclusion

The present chapter examined the production of vowels and stop consonants in German and French by Spanish-English bilinguals in both orders of acquisition. Importantly, in order to examine the initial state of typologically distinct “third” languages, these subjects did not speak French and German and produced the target words using a pseudoword imitation task. The results of the analyses of stops and vowels revealed distinct performance. In particular, vowels were often successfully imitated while stops were not. French and German stops fell between L1 and L2 values, presumably due to L1 and L2 influence. Overall, these results suggest that L3 phonology cannot accurately be accounted for in the current models of L3 acquisition. Future work should consider the role of the segment, stage of development, and typological similarity of the involved languages as subjects that may impact L3 production.

Chapter 5: The perception-production link in the intitial stage of L3 acquisition

5.1 Introduction

Historically, it has been debated whether L2 perceptual learning occurs prior to production. Previous theories of L2 acquisition of phonetics and phonology have debated the relationship between production and perception in new language learning. Some theories, such as Motor Theory (Liberman & Mattingly, 1985) and Direct Realist Theory (Fowler, 1986), predict that perception is a prerequisite for production The PAM and PAM-L2 are based (Best & Tyler, 2007) on DRT and also thus suggests that the perception of sound contrasts predicts the acquisition of new-language sounds. In a new revision, The SLM-R (J. Flege et al., 2021) now predicts that non-native sounds develop in tandem, rather than perception preceding production.

The perception-production link can also inform L3 acquisition. L3 models differ as to whether full language transfer or property-by-property influence occurs. If perception and production develop at different rates, as has been widely reported in the L2 literature, then counter-evidence would be provided for full transfer models. On the other hand, property-by-property models would more readily explain a difference between perception and production.

The present chapter analyzes the perception and production data from Chapters 3 and 4 in tandem with the primary goal of examining whether L1 and L2 influence occurs differently in perception and production. While the primary goal of this analysis was to inform L3 models, these results also have implications for the perception-production link, and studies in L2 phonetic and phonology largely impact the interpretation of the results of the present study.

5.1.1 Earlier research in the perception-production link

Early research in the perception-production link largely found that perception preceded production in development. For example, J. Flege (1993) found that proficient L1

Chinese - L2 English speakers perceived word final differences between /d/ and /t/ better than they were able to produce it. In a study examining the perception and production of the English /r/-/l/ contrast, Borden, Gerber, & Milsark (1983) found that perceptual training or higher performance on identification tasks predicted better production of the same contrasts in L1 Korean-L2 English speakers. Similar trends were reported in several additional studies, including more recent ones (J. Flege et al., 1997; Kissling, 2014, *inter alia*).

Other research, however, found that production precedes perception or that the two modalities are not related. Goto (1971) tested Japanese learners of English on their perception and production of English words containing /r/ and /l/ sounds, and found that even participants who achieved relatively high production accuracy still exhibited poor discrimination." Also examining /r/-/l/ in L1 Japanese-L2 English, Sheldon & Strange (1982), reported similar results to Goto (1971), where the participants were more accurate in production than they were in perception. This trend has also been observed in additional language pairings. For example, L1 Dutch-L2 English speakers have been found to improve to a higher degree in the production of VOT than their perception (J. Flege & Eefting, 1987). Additionally, J. Flege et al. (1997) reported better production than perception of /æ/ by L1 German-L2 English speakers.

Finally, it has also been found that the relationship between production and perception is segment-dependent or are autonomously represented altogether. For example, Hao & Jong (2016) found that L1 Korean-L2 English speakers were better at perception than production for fricatives, but the opposite trend was observed in stops. On the other hand, Hattori & Iverson (2010) concluded that production and perception are autonomously represented, based on the finding that L1 Japanese L2 English speakers' /r/-/l/ production and perception was weakly correlated across tasks.

5.1.2 Recent reserach in the perception-production link

Given the history of contradictory results regarding whether the development of production is dependent upon production, more recent research has been conducted with one principle aim to identify rectify issues with previous research that may have contributed to the conflicting results. The trend in recent research seems to be that perception precedes production, although the conflicting results have remained. For example, in a study of L1 English speakers of L2 Japanese, Okuno & Hardison (2016) studied whether perception training would lead to gains in production. For one training condition, the authors found that they also improved in production, and argued that this result supported the production-perception link. Joseph V. Casillas (2020) also found that gains in perception preceded gains in production in VOT for L2 Spanish speakers over a 7 week domestic immersion program. Finally, Melnik-Leroy, Turnbull, & Peperkamp (2022) found that perception preceded production of the /y/-/u/ contrast in L1 English-L2 French speakers in both prelexical and lexical tasks.

Other studies have found an unclear relationship between perception and production. Another recent longitudinal study, Nagle (2018) also examined L2 Spanish stop production over 4 semesters of non-immersion Spanish courses. The results in this case were mixed showed a high degree of variability. While some participants showed evidence of developing perception and production contemporaneous (in the same session), others showed evidence of time-lagged improvement, where improvement in production occurred only in the session after improved perception was occurred. Additionally, other participants showed evidence of an asymptotic relationship between perception in production, meaning that perception had to reach a high level of accuracy prior to production improving. Kim & Han (2022) analyzed the relationship between perception and production in L1 Korean-L2 English speakers across processing levels (lexical and prelexical) using a myriad of tasks. The authors concluded, based on the lack of a clear trend in the data, that perception and production are independent at both the lexical and prelexical level.

5.1.3 L3 reserach in the perception-production link

Only one study examines the perception-production link in L3 acquisition (Wrembel et al., 2022). The study examined the longitudinal acquisition of rhotic sounds by participants who speak English, German, Polish, with English being their L2, and German and Polish their L1 or L3. The participants were tested a two testing times - 5 months of L3 learning and 10 month of L3 learning. To measure perception, the authors used a forced-choice goodness task, in which participants heard two versions of a target word within a frame and chose which was a better exemplar. Production was measured using a delayed repetition task, where target words were repeated after a 1000ms delay and prompt in all 3 languages. The results revealed that, for many participants, L3 perception and production accuracy was low, and a clear link could not be established.

5.1.4 The Present study

To summarize the literature to date on the perception-production link, there is a trend toward production preceding perception, but this relationship appears to also potentially vary by individual, segment and processing level. Additionally, there is very little research on the production-perception link in L3 learners. The present study compares the vowel production of initial exposure French and German from two prelexical tasks involving categorization in perception and pseudoword repetition in production, and provides a new language combination to the multilingual literature on the perception-production link. In particular, this work is guided by the following research question:

RQ1: Do Spanish-English speakers show meaningful differences in perception in production at their initial exposure to a new language?

In this case "meaningful differences" will be evaluated using the overlap of the posterior distribution and overall difference in probability between perception and production.

This study has implications for the perception-production link, as well as L3 acquisition. If perception appears to develop prior to production, then it would stand as evidence that subsequent language development (after an L2) follows a similar pattern to the L2 development reported in many studies. Of course, this would also stand in contrast to studies which found that production develops prior to or independently of perception. L3 models also benefit from these results, since some models predict that full language transfer occurs. Regardless of whether perception precedes production or vice-versa, this would not support full transfer models, such as the TPM or L2 Status Factor.

5.1.5 Participants

In total, 37 participants completed both the perception and production experiments (English L1 n = 9; Spanish L1 n = 28). The participants were late bilinguals who completed both the perception and production experiments of the present dissertation.

5.2 Analysis

Since the perception data was categorical forced-choice categorization data and the production data was continuous, the production data was converted from continuous to categorical using a method devised by the author for the purpose of the present analysis using an R script. Specifically, the production data of vowels was transformed such that it matched the perception data and was binary (whether a given French or German token was more like English or Spanish). The data from the perception portion of the present chapter was reported in the vowel categorization task described in section 3.2. In this task, French and German vowel sounds were categorized given the choice of seven total carrier words, of which 4 English and 3 Spanish words were included. Thus, for each vowel token, either a Spanish or English category was chosen. The production data was taken from the vowel production in Chapter 4.2.5 of these is continuous in two dimensions (F1 and F2). A pipeline was devised to convert the continuous production data into binary data, essentially simulated a Native speaker forced-choice judgment of English-like or Spanish like on all French and German data points. For each data point, the Euclidian distance was measured to both the spectral centroid of corresponding Spanish phoneme and English phoneme. These distances were used to calculate relative distance in the two dimensional space from a given data point to both centroids using the following equation.

$$probability_s = distance_e/(distance_e + distance_s) \tag{5.1}$$

“Distance_e” refers to the Euclidian distance from a point to the English spectral centroid, and “distance_s” is the distance from that same point to the Spanish centroid. The equation produces “probability_s”, which in this case is the probability of that token being categorized as Spanish-like. For example, if the Euclidian distance from a data point to the English centroid was 2, and the distance from the Spanish centroid was 1, then the probability of choosing a Spanish category 2/2+1 = .66. To find the probability of an English category, one can subtract (1 - probability_s). From here, using the probability derived from the Euclidian distance of each data point, Spanish (coded a 1) and English (coded as 0) “choices” were probabilistically simulated using a function in R. The data sets from the vowel categorization and pseudoword repetition task were then combined.

Figure 5.1 shows an example for a single data point for how Euclidian distance was converted to probability. The logic to this formula is that the closest corresponding centroid to a given data point should be proportionally more likely to be chosen than one that is further away. The corresponding centroids were /i/ in all languages, /y/ and /u/ in French and German and Spanish and English, /o/ in French, German and Spanish to English /ɑ/, and /ʌ/ in French German and English to Spanish /a/. Table 5.1 shows each corresponding phoneme and its label in the data. Figure 5.1 shows a hypothetical example of how a given vowel production was probabilistically converted to either a Spanish or English production. In the example, the Spanish centroid is closer (distance of 1) than the English centroid (distance of 2) to the data point. The use of this formula provides an probabilistic alternative to simply choosing the closer centroid as an all-or-nothing choice. Figure 5.1 shows all raw French and German data points relative to their respective centroids.

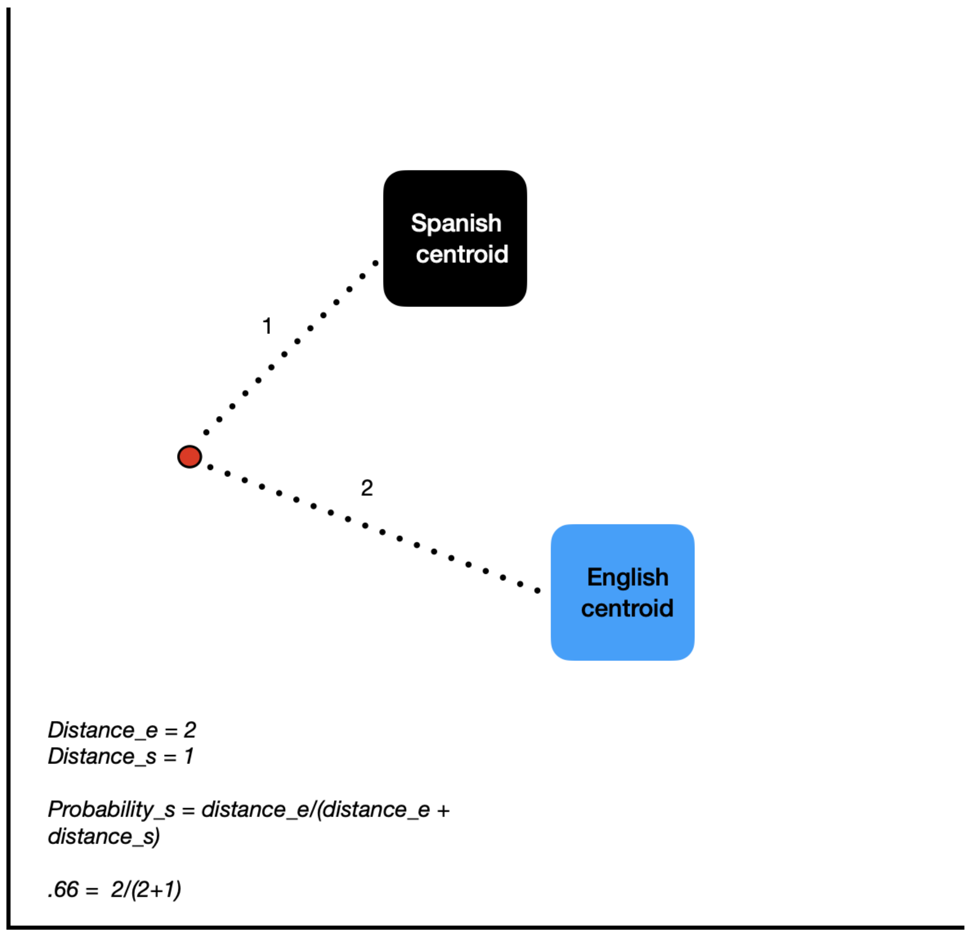


Figure 5.1: A hypothetical example of the coverstion from Euclidian Distance to Probability for a single data point

To analyze the newly formed data set, Bayesian logistic regression models were run for each the Spanish L1 and English L1 group. In each case, the outcome variable was binary: choosing a Spanish category (coded as 1) or an English category (coded as 0). The fixed effect predictors were mode (perception or production) and phoneme (4 levels: “i”, “u”, “a” or “o”.) The phoneme in this case was labeled in the data

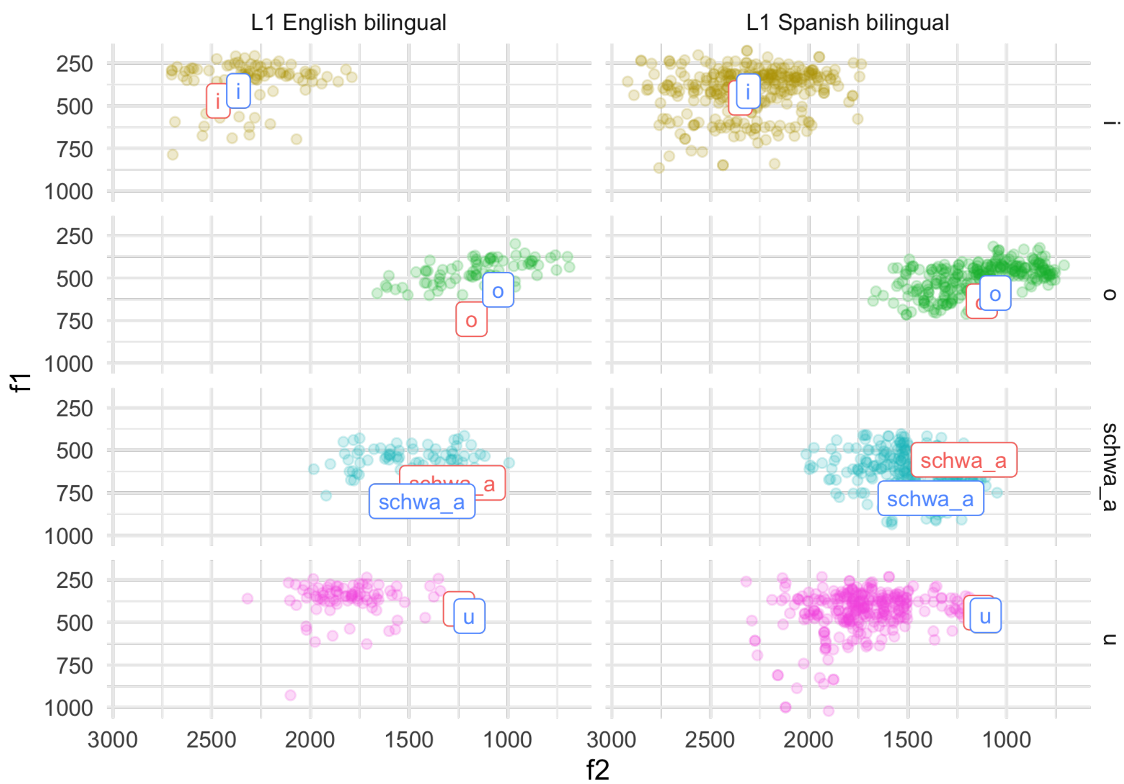


Figure 5.2: Raw French and German data points relative to each Spanish and English centroid

by their corresponding grapheme so that the model could compare the intended corresponding phonemes across languages. The models both included the default brms priors: Student's T distribution with 3 degrees of freedom. The model was run using with 4000 iterations of Hamiltonian Monte-Carlo sampling (1000 warm up), across 4 chains and 8 processing cores.

Table 5.1: Corresponding Phonemes across languages and their label in the data.

English	Spanish	French	German	data label
/i/	/i/	/i/	/i/	"i"
/u/	/u/	/y/	/y/	"u"
/ɑ/	/o/	/o/	/o/	"o"
/ʌ/	/a/	/ʌ/	/ʌ/	"a"

5.3 Results

Figure 5.3 shows the total count of choices for Spanish and English categories per phoneme and language (French or German) by Spanish L1 group. Figure 5.4 shows the same information by the English L1 group. In both figures, the red bar represents the total quantity of German and French tokens categorized using an English or Spanish category. For the L1 Spanish group, across all 4 phonemes, the quantity of English categorizations was much higher than Spanish categorizations in perception. In production, there was not a clear preference for English or Spanish in any phoneme. For the L1 English group, the trends were similar. An overall bias for English was evident for /i/, /o/ and /ʌ/, but the production of /y/ was assimilated to /u/ in both languages similarly. Similarly to the Spanish L1 group, there was not a clear trend in terms of language preference in production in the English L1 group.

Figure 5.5 shows the posterior distribution of the logistic model converted to probability (for the log-odds output see Table 5.3. The model corroborates the trends observed in the descriptive data, since it shows that the probability of a Spanish segment is around .5 for all segments in production, but much lower in perception

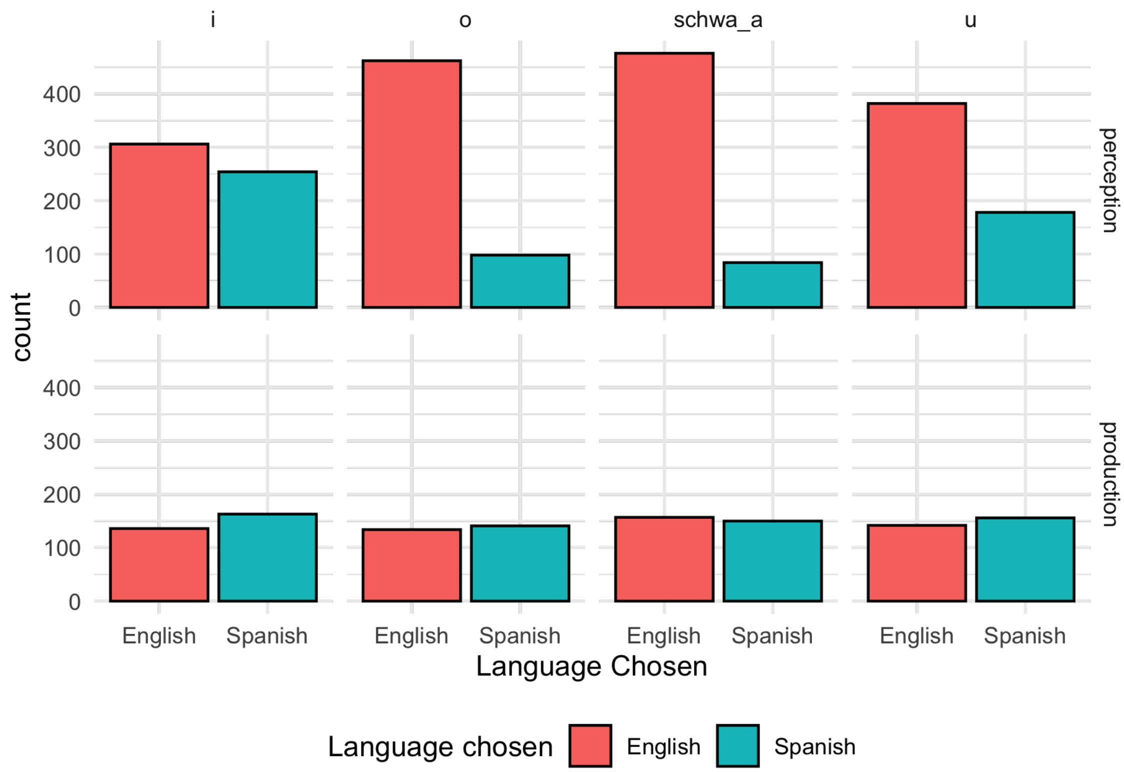


Figure 5.3: The quantity English and Spanish categories by the L1 Spanish group as a function of mode and segment

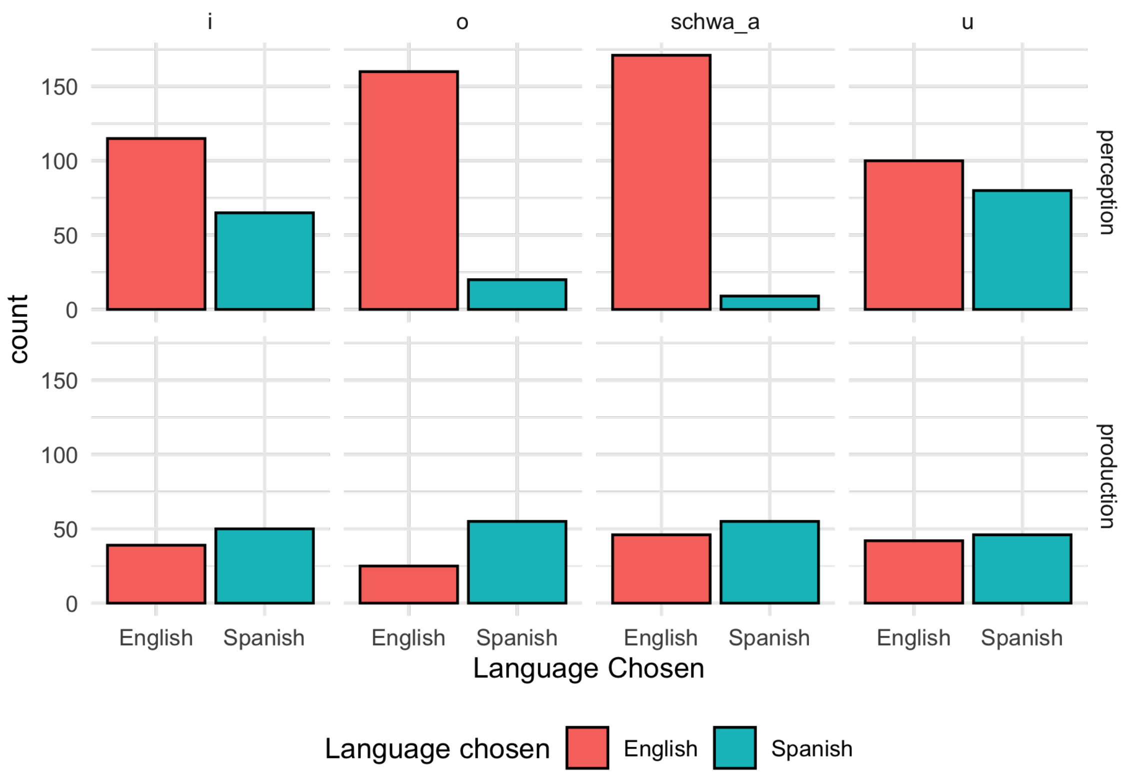


Figure 5.4: The quantity English and Spanish categories by the L1 English group as a function of mode and segment

overall. Figure 5.6 (and table 5.2 shows the same information for the L1 English group. This model again mirrors the descriptive data, where there was generally around a .5 probability of a choice of English or Spanish in production, but a preference for an English category in perception.

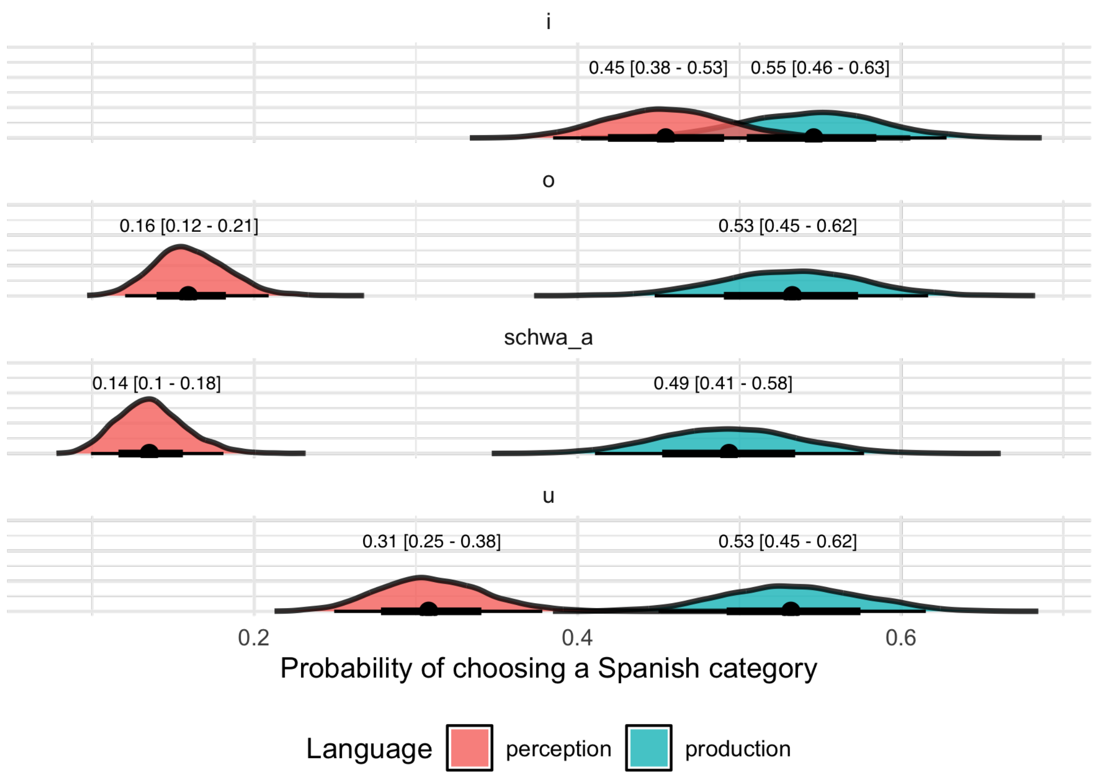


Figure 5.5: The probability of choosing a Spanish category by the L1 Spanish group as a function of mode and segment

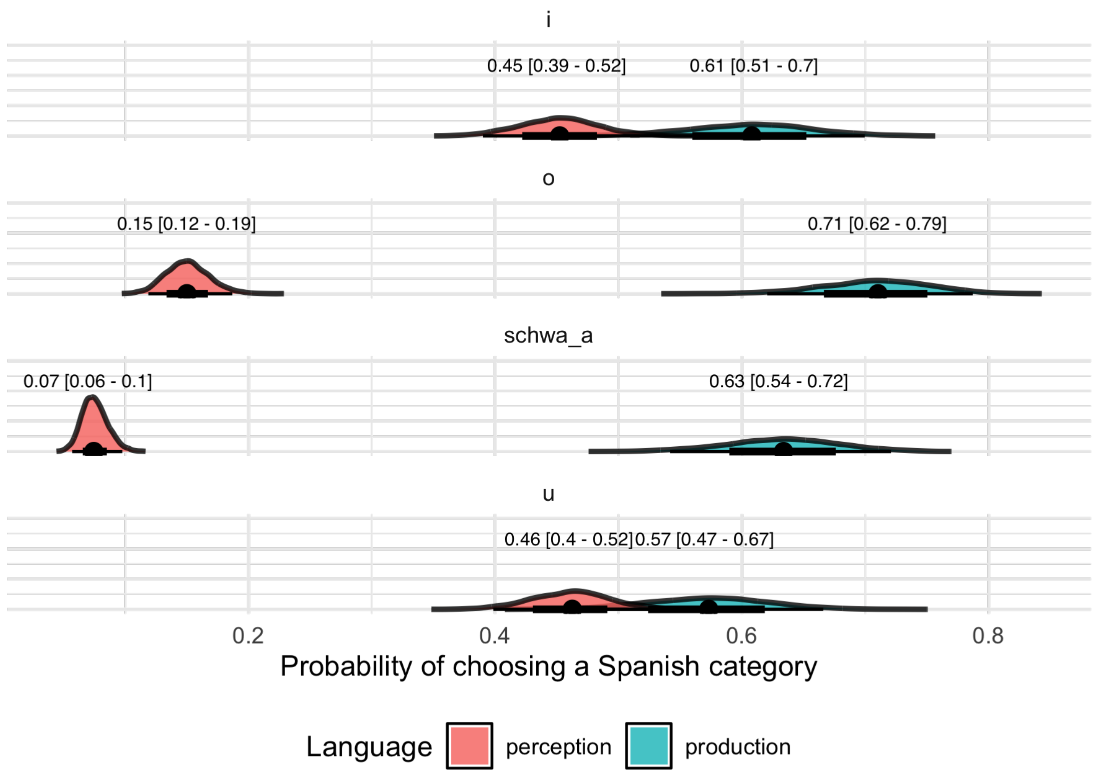


Figure 5.6: The probability of choosing a Spanish category by the L1 English group as a function of mode and segment

Table 5.2: Summary of the English L1 group posterior distribution modeling language choice as a phoneme and mode. The table includes posterior medians, the 95% HDI, the percentage of the HDI within the ROPE, and the maximum probability of effect (MPE).

Parameter	Median	HDI	% in ROPE	MPE	Rhat	ESS
Intercept	-0.19	[-0.45, 0.07]	0.21	0.93	1.00	536
/o/	-1.54	[-1.73, -1.36]	0.00	1.00	1.00	3702
schwa, /a/	-2.32	[-2.55, -2.11]	0.00	1.00	1.00	3980
/u/, /y/	0.04	[-0.12, 0.20]	0.76	0.68	1.00	3649
Production	0.63	[0.25, 1.01]	0.00	1.00	1.00	2037
/o/: Production	2.01	[1.53, 2.47]	0.00	1.00	1.00	2882
schwa, /a/: Production	2.43	[1.98, 2.90]	0.00	1.00	1.00	2880
/u/, /y/: Production	-0.18	[-0.62, 0.26]	0.26	0.79	1.00	2630

Table 5.3: Summary of the Spanish L1 group posterior distribution modeling language choice as a phoneme and mode. The table includes posterior medians, the 95% HDI, the percentage of the HDI within the ROPE, and the maximum probability of effect (MPE).

Parameter	Median	HDI	% in ROPE	MPE	Rhat	ESS
Intercept	-0.18	[-0.47, 0.11]	0.27	0.89	1.00	911
/o/	-1.48	[-1.76, -1.20]	0.00	1.00	1.00	2883
schwa, /a/	-1.67	[-1.98, -1.38]	0.00	1.00	1.00	2985
/u/, /y/	-0.62	[-0.88, -0.37]	0.00	1.00	1.00	2688
Production	0.36	[0.08, 0.66]	0.01	0.99	1.00	1949
/o/: Production	1.43	[0.98, 1.85]	0.00	1.00	1.00	2414

Parameter	Median	HDI	% in ROPE	MPE	Rhat	ESS
schwa, /a/: Production	1.46	[1.02, 1.92]	0.00	1.00	1.00	2333
/u/, /y/: Production	0.57	[0.14, 0.99]	0.00	1.00	1.00	2297

Overall, these results point to a difference in perception and production. In particular, there is evidence that, regardless of order of acquisition, there was a preference for English categories in perception in most cases, although this varied by segment. For instance, /i/, which is present in all languages was closer to a 50-50 split than other segments. The phoneme /y/ in French and German was also categorized as closer to evenly given Spanish and English /u/, with slight biases to either English or Spanish in perception. This differed from production, where both groups showed a nearly equal probability of an English or Spanish category, suggesting that no clear trends towards English or Spanish emerge in production.

5.4 Discussion

5.4.1 Implications for L3 models

Full transfer L3 models, in their current form, cannot account for different rates of development in L3 perception and production. These models, including the Typological Primacy Model (Rothman, 2011) and the L2 Status factor (Bardel & Falk, 2007), predict that full language transfer occurs during L3 development or either the more typologically similar language to the L3 or the most recent late learned language. This transfer is predicted to occur all at once, rather than incrementally. The present chapter provides counter-evidence for this claim, since the influence of English on the L3 was heavier in perception than in production.

Other L3 models, which predict both the L1 and L2 can influence the L3, such as the Linguistic Proximity Model [LPM; Westergaard et al. (2017)] and the Scalpel Model (Slabakova, 2017), can more readily explain the results obtained in the present chapter. Although these models do not explicitly predict a difference in language influence between perception and production, they suggest that there is Full Transfer Potential (FTP) from a given language representation. A post-hoc explanation from the LPM might be that modality (perception versus production) has an impact on this full transfer potential, where access to a phonetic category in perception is initially easier than in production.

5.4.2 Implications for the perception-production link

The present study also provides another instance of perceptual development preceding production, in this case at the prelexical level. Importantly, these results should not be taken as evidence that all of perception precedes all of production. This distinction is important given that previous research has found that successful perception and production at the prelexical level does not necessarily transfer to the lexical level (Llompart & Reinisch, 2019). That is, since these tasks did not involve lexical items or processing, the present data cannot provide evidence that (L3) lexical perception precedes its production.

It is important to note, also, that the idea that perception precedes production is an interpretation by the present author of the differences obtained in the data. In previous research of the production-perception link, the prelexical tasks in perception are typically AX discrimination tasks based on accuracy of the discrimination of a contrast. That accuracy is then compared to the production of that contrast in a pseudoword or real word (although this is a case where prelexical and lexical tasks are compared). In the present case, differences in perception and production, rather than differences in accuracy, were of import, since the primary goal of this chapter was to examine previous language influence on perception relative to production rather than whether one of the two develops first. In other words, the nature of the tasks involved did not allow for accuracy to be measured, since it provided imperfect matches between the stimuli and answer choices. Nevertheless, production was quite close to chance, while perception largely showed English bias. This result is taken as evidence that previous languages have more of a controlling influence over perception initially.

Whether this influence eventually impact production is a question for future research, and would further inform the production-perception link in a multilingual context.

5.4.3 Limitations and future directions

The present study was not without limitations. Firstly, it is important to point out that the use of imitation in general in not agreed to be a measure of pure production. For instance, Hao & Jong (2016) argue that an imitation task can be viewed either as a production task given an auditory prompt, or, oppositely, as a perception task with an oral response. They gathered evidence for this view using L1 English–L2 Mandarin and L1 Korean-L2 English speakers whose accuracy in L2 imitation was not associated with their accuracy in identification tasks.

Additional limitations include the low sample in the English speaking production task. One issue with a smaller sample size is the increased probability of a false negative finding. While this may be these case here, it is argued that the trend observed in the L1 Spanish group and in the English speakers' perception follow similar trend and have greater statistical power. That is, while one still cannot be certain that small differences do not exist between groups, it would not be predicted in a future, higher sample study, given the trends observed in the similar groups.

Finally, the use of the pipeline to convert the continuous vowel production data to a binary choice was devised relative to this particular experiment, and could likely be improved with time. For instance, the same logic of taking a given point and deriving a probability based on its proximity to centroids could be expanded beyond the corresponding phonemes to include all centroids. Such an approach might reveal more nuance in future experiments.

5.5 Conclusion

The present chapter examined the the link between perception and production of new language sounds by Spanish-English bilinguals. French and German vowel sounds were categorized and repeated by L1 English - L2 Spanish and L1 Spanish L2 English speakers and were categorized in both modalities as either English-like or Spanish-like. The results revealed a difference in perception and production. In particular, it was found that an overall English bias was present in perception, while no clear bias for English or Spanish was found in production. This result is interpreted as evidence that L3 perception is influenced by source languages to a higher degree than production, and may develop sooner as a result.

Chapter 6: Conclusion

6.1 Introduction

The present work investigated the initial state of L3 perception and production by Spanish-English bilinguals. The primary goal of the experiments used in the present dissertation were to determine how the known languages of a bilingual influence the perception and production of sounds in a new, third, language. Specifically, the chapters of the present work examined the production of vowel sounds and stop consonants of French and German by the same bilingual individual. Additionally, the discrimination and categorization of vowel sounds were tested. This chapter summarizes the results of the previous chapters and discusses its contribution to the field of third language acquisition. Additionally, the implications for models of L2 phonetics and phonology are discussed, followed by recommendations for the future.

6.2 Summary of main findings

The present dissertation provides evidence that bilinguals have access to both their known languages at near first exposure to a new language. This relative access to the L1 and L2 varies by individual, segment, modality and which third language is being learned. Taken together, the evidence provided by each chapter of the present dissertation suggest that L3 phonology should continue to test new hypotheses in order to progress and uncover what factors might be associated with the large degree of individual variability present in the data.

Chapter 3 examined the categorization and discrimination of vowel sounds in a third language. The analysis revealed that bilingual speakers categorize new language sounds using categories from both their first and second languages, but that discrimination of English sounds in L3 words was poor. The categorization found no evidence of a language bias when both Spanish and English contained the same L3 phoneme. Additionally, the English-like vowel sounds in the L3 were more often categorized using an English category, but so was the Spanish like vowel sound /o/, indicating an overall bias for English categories in both vowel sounds designed to elicit English and Spanish categories.

Chapter 4 studied the production of both vowel sounds and stop consonants. For English and Spanish, pseudoword tokens were elicited using a pseudoword reading task. For French and German, a pseudoword reptition task was used. The results revealed that most participants were able to effectively imitate vowel tokens, but not stop consonants. Specifically, the vowel /y/, which is present in French and German but not Spanish or English, was effectively imitated by bilinguals and monolinguals alike.

Chapter 5 analyzed perception and production of vowels together to determine whether there were differences in the relative influence of the bilingual's known languages on a new language between the two modalities. Using a formula, the continuous production data was probabilistically converted to binary data.
The results showed that the English bias present for some segments in perception did not transfer to production, where the probability of a more English-like or a more Spanish-like production was close to chance.

6.3 Implications for L3 models

The present dissertation provides several contributions to L3 models. First, it provides evidence that bilinguals have access to both languages during the L3 initial state. This finding complicates the predictions of full transfer models, such as the L2 Status Factor (L2SF) and the Typological Primacy Model (TPM). The L2SF would predict that the L2 of participants would influence the L3 by default, where the TPM would choose the more typologically close source language to influence the L3. The L2SF argues that L2 status arises from the cognitive similarity of the L2 and L3 in late learners of languages. The main argument for full transfer in the TPM is cognitive economy. In other words, the TPM suggests that unilingual processing is less cognitively taxing than bilingual processing, and the that the mind subconsciously chooses only one source language when processing L3 input. Neither of these views holds up to the perception nor production experiments of the present dissertation. In perception, both bilingual groups categorized French and German sounds using both Spanish and English categories. In production, vowels were effectively imitated and stops were typically intermediate of L1 and L2 categories.

Additionally, the results of the experiments suggest that the similarity between known languages does play role in cross-language influence at the initial state. Primary evidence for this conclusion can be seen in the analysis of VOT production in Chapter 4. The German tokens were produced with longer VOT than French tokens, which cannot be accounted for by language status. Additionally, while a typological account of L3 acquisition would predict that the more similar language would impact the L3, it cannot account for intermediate productions, which were observed in both French and German tokens. Finally, the disparity between perception and production suggest that the relative cross-linguistic influence of the L1 and the L2 on the L3 vary by these modalities. Full transfer models do not have a clear explanation for this case, while partial transfer models also do not explicitly make predictions for production and perception on new language phonology.

The Linguistic proximity model best explains the results, but did not predict them. Overall, these results point to the need for the development of a phonology-specific model of third language acquisition, which must take into account the relationship between perception and production, language typology. In addition to these factors, future work should consider what other factors might impact language access during new language learning. The current body of literature and the factors examined in the present work do not adequately account for the wide individual variation that was observed.

6.4 Implications for L2 phonetics and phonology

Although the primary goal of the present dissertation was to inform L3 acquisition, the results have important implications for models of L2 phonological acquisition as well. Firstly, the results of the vowel categorization task from Chapter 3 suggest that bilinguals use all of their full inventory of phonemes across all languages that they know, rather than treating an additional late learned language as another instance of L2 learning, in which novel sounds are categorized according to the native phonology. The PAM and PAM-L2 (Best & Tyler, 2007), which specifically predict that that L2 sound discrimination is predictable based on the *native* phonology could be revised to include L3 learning. That is, it could be noted that subsequent language learning may not be the same in the case of L2 learning and L3/Ln learning, since it appears that more than the native phonology impacts language learning.

In addition to the PAM and PAM-L2, the newly revised Speech Learning Model (SLM-R) can also benefit from the results obtained here. In particular, the results of the analysis of perception and production in Chapter 5 suggest that L3 perception and production develop at different rates, at least for some segments. Unlike the original Speech Learning Model, which suggests that perception develops prior to production, the SLM-R predicts that the two systems co-evolve. The results of Chapter 5 support the original SLM, but not the SLM-R. Additionally, like the PAM and PAM-L2 , the SLM and SLM-R do not make explicit conditions for L3 learning.

6.5 Future Research

Although the present dissertation produced informative findings, many of them should be interpreted with caution when it comes to L3 acquisition. It is important to note that the "L3" learners here were not actually in the process of acquiring a third language. Rather, they were bilingual speakers who did not yet speak French or German, and they were recruited to explore what the starting point is for L3 acquisition. Based on the present data, the starting point of L3 acquisition is not the same as L2 acquisition. In order to gain a deeper understanding of how source languages influence a third language over time, longitudinal designs should be encouraged.

Despite its benefits, longitudinal research arguably poses several logistical issues. One potential issue is participant attrition over the course of a study, leading to a low sample size and an potentially underpowered research. In L3 acquisition, where providing evidence for equivalence can be as important as finding differences, high power becomes important, since the probability of a false negative finding increases as power goes down. A possible compromise is to increase repeated measures in studies with few participants to increase the reliability of the estimates obtained between participants.

In addition to practices in individual studies, I argue that L3 theory should move forward by collaborating with other fields. Specifically, an integration of current L2 models of phonetics and phonology and existing models of L3 acquisition with sound analysis and research methodology would be an optimal path forward. The insights from the L2 models consider how a phonological inventory impacts L2 learning, while L3 models posit that previously learned languages may not always be straightforwardly available. These theories combined, together with proper interpretations and us statistical tests can improve the chance that research is cumulative and informative.

6.6 Conclusion

The present dissertation investigated the initial state of L3 acquisition by Spanish-English bilingual speakers. The studies carried out in this work regarding the perception and production of both stops and vowels add to our understanding of the starting point of L3 acquisition, which appears to differ from L2 acquisition. The data suggest that perception precedes production in L3 acquisition and that bilinguals begin the process of acquiring a novel language with access to both their previously acquired systems, contrary to some previous research. Additionally, previous language influence appears to vary by individual and segment. Future research should be open to all potential explanations and draw insights from other focuses within linguistics or other closely related fields.

Appendix A: Sample Appendix

Table A.1: Mean and standard deviation of each phoneme in the hand-corrected subset

phoneme	language	mean_f1	sd_f1	mean_f2	sd_f2
a	Spanish	746	130	1487	293
i	English	476	409	2453	268
i	French	472	203	2343	279
i	German	421	140	2419	345
i	Spanish	428	158	2319	318
o	English	681	126	1206	307
o	French	553	83	1306	200
o	German	424	63	940	128
o	Spanish	658	144	1306	366
schwa	English	622	181	1301	286
schwa	French	630	113	1592	248
schwa	German	736	101	1307	164
u	English	437	109	1256	275
u	French	558	265	1920	209
u	German	404	135	1724	290
u	Spanish	542	189	1298	496

REFERENCES

Antoniou, M., Best, C. T., & Tyler, M. D. (2013). Focusing the lens of language experience: Perception of Ma'di stops by Greek and English bilinguals and monolinguals. *The Journal of the Acoustical Society of America*, *133*(4), 2397–2411. http://doi.org/10.1121/1.4792358

Antoniou, M., Liang, E., Ettlinger, M., & Wong, P. C. M. (2015). The bilingual advantage in phonetic learning. *Bilingualism: Language and Cognition*, *18*(4), 683–695. http://doi.org/10.1017/S1366728914000777

Aoyama, K., Flege, J., Guion, S. G., Akahane-Yamada, R., & Yamada, T. (2004). Perceived phonetic dissimilarity and L2 speech learning: The case of Japanese /r/ and English /l/ and /r/. *Journal of Phonetics*, *32*(2), 233–250. http://doi.org/10.1016/S0095-4470(03)00036-6

Balas, A. (2018). *Non-native vowel perception: The interplay of categories and features = Percepcja samogłosek w języku nienatywnym: Oddziaływanie kategorii i cech* (Wydanie I). Poznań: Wydawnictwo Naukowe UAM.

Banks, W. P. (1970). Signal detection theory and human memory. *Psychological Bulletin*, *74*(2), 81.

Bardel, C., & Falk, Y. (2007). The role of the second language in third language acquisition: The case of Germanic syntax. *Second Language Research*, *23*(4), 459–484. http://doi.org/10.1177/0267658307080557

Bardel, C., & Falk, Y. (2021). L1, L2 and L3: Same or different? *Second Language Research*, *37*(3), 459–464. http://doi.org/10.1177/0267658320941033

Bayona, P. (2009). The Acquisition of Spanish Middle and Impersonal Passive Constructions from SLA and TLA Perspectives. In *Third language acquisition and universal grammar*. Bristol, UK; Buffalo: Multilingual Matters.

Berkes, É., & Flynn, S. (2012). Further Evidence in Support of the Cumulative-Enhancement Model: CP Structure Development. In *Third Language Acquisition in Adulthood*. Amsterdam; Philadelphia: John Benjamins.

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O.-S. Bohn & M. J. Munro (Eds.), *Language Learning & Language Teaching* (Vol. 17, pp. 13–34). Amsterdam: John Benjamins Publishing Company. http://doi.org/10.1075/lllt.17.07bes

Birdsong, D., Gertken, L. M., & Amengual, M. (2012). Bilingual Language Profile: An Easy-to-Use Instrument to Assess Bilingualism. University of Texas at Austin: COERLL. Retrieved from %3Chttps://sites.la.utexas.edu/bilingual/%3E.

Blank, C. A., & Zimmer, M. C. (2009). A transferência fonético-fonológica L2 (francês) – L3 (inglês): Um estudo de caso. *Revista de Estudos Da Linguagem*, *17*(1), 207–233. http://doi.org/10.17851/2237-2083.17.1.207-233

Bohn, O.-S., & Flege, J. (1992). The Production of New and Similar Vowels by Adult German Learners of English. *Studies in Second Language Acquisition*, *14*(2), 131–158. http://doi.org/10.1017/S0272263100010792

Bohnacker, U. (2006). When Swedes begin to learn German: From V2 to V2. *Second Language Research*, *22*(4), 443–486. http://doi.org/10.1191/0267658306sr275oa

Borden, G., Gerber, A., & Milsark, G. (1983). Production and perception of the/r/-/l/contrast in korean adults learning english. *Language Learning*, *33*(4), 499–526.

Bradlow, A. R. (1995). A comparative acoustic study of English and Spanish vowels. *The Journal of the Acoustical Society of America*, *97*(3), 1916–1924. http://doi.org/10.1121/1.412064

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

Brysbaert, M. (2020). Power considerations in bilingualism research: Time to step up our game. *Bilingualism: Language and Cognition*, 1–6. http://doi.org/10.1017/S1366728920000437

Bürkner, P.-C. (2017). **Brms** : An *r* Package for Bayesian Multilevel Models Using *stan*. *Journal of Statistical Software*, *80*(1). http://doi.org/10.18637/jss.v080.i01

Casillas, Joseph V. (2020). Phonetic category formation is perceptually driven during the early stages of adult L2 development. *Language and Speech*, *63*(3), 550–581.

Casillas, Joseph V., & Simonet, M. (2018). Perceptual categorization and bilingual language modes: Assessing the double phonemic boundary in early and late bilinguals. *Journal of Phonetics*, *71*, 51–64. http://doi.org/10.1016/j.wocn.2018.07.002

Curtin, S., Goad, H., & Pater, J. V. (1998). Phonological transfer and levels of representation: The perceptual acquisition of Thai voice and aspiration by English and French speakers. *Second Language Research*, *14*(4), 389–405. http://doi.org/10.1191/026765898674095369

Cutler, A., Weber, A., & Otake, T. (2006). Asymmetric mapping from phonetic to lexical representations in second-language listening. *Journal of Phonetics*, *34*(2), 269–284. http://doi.org/10.1016/j.wocn.2005.06.002

Declerck, M., & Kormos, J. (2012). The effect of dual task demands and proficiency on second language speech production. *Bilingualism: Language and Cognition*, *15*(4), 782–796.

Escudero, P., & Boersma, P. (2004). BRIDGING THE GAP BETWEEN L2 SPEECH PERCEPTION RESEARCH AND PHONOLOGICAL THEORY. *Studies in Second Language Acquisition*, *26*(04). http://doi.org/10.1017/S0272263104040021

Escudero, P., & Chládková, K. (2010). Spanish listeners' perception of American and Southern British English vowels. *The Journal of the Acoustical Society of America*, *128*(5), EL254–EL260. http://doi.org/10.1121/1.3488794

Escudero, P., Sisinni, B., & Grimaldi, M. (2014). The effect of vowel inventory and acoustic properties in Salento Italian learners of Southern British English vowels. *The Journal of the Acoustical Society of America*, *135*(3), 1577–1584. http://doi.org/10.1121/1.4864477

Escudero, P., & Vasiliev, P. (2011). Cross-language acoustic similarity predicts perceptual assimilation of Canadian English and Canadian French vowels. *The Journal of the Acoustical Society of America*, *130*(5), EL277–EL283. http://doi.org/10.1121/1.3632043

Escudero, P., & Williams, D. (2011). Perceptual assimilation of Dutch vowels by Peruvian Spanish listeners. *The Journal of the Acoustical Society of America*, *129*(1), EL1–EL7. http://doi.org/10.1121/1.3525042

Falk, Y., & Bardel, C. (2011). Object pronouns in German L3 syntax: Evidence for the L2 status factor. *Second Language Research*, *27*(1), 59–82. http://doi.org/10.1177/0267658310386647

Falk, Y., Lindqvist, C., & Bardel, C. (2015). The role of L1 explicit metalinguistic knowledge in L3 oral production at the initial state. *Bilingualism: Language and Cognition*, *18*(2), 227–235. http://doi.org/10.1017/S1366728913000552

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

Flege, J. (1993). Production and perception of a novel, second-language phonetic contrast. *The Journal of the Acoustical Society of America*, *93*(3), 1589–1608.

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

Flege, J., Aoyama, K., & Bohn, O.-S. (2021). The Revised Speech Learning Model (SLM-r) Applied. In R. Wayland (Ed.), *Second Language Speech Learning* (1st ed., pp. 84–118). Cambridge University Press. http://doi.org/10.1017/9781108886901.003

Flege, J., Bohn, O.-S., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of english vowels. *Journal of Phonetics*, *25*(4), 437–470.

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

Flege, J., & Eefting, W. (1988). Imitation of a VOT continuum by native speakers of English and Spanish: Evidence for phonetic category formation. *The Journal of the Acoustical Society of America*, *83*(2), 729–740. http://doi.org/10.1121/1.396115

Flege, Jim., Frieda, E. M., & Nozawa, T. (1997). Amount of native-language (L1) use affects the pronunciation of an L2. *Journal of Phonetics*, *25*(2), 169–186. http://doi.org/10.1006/jpho.1996.0040

Flynn, S., Foley, C., & Vinnitskaya, I. (2004). The Cumulative-Enhancement Model for Language Acquisition: Comparing Adults' and Children's Patterns of Development in First, Second and Third Language Acquisition of Relative Clauses. *International Journal of Multilingualism*, *1*(1), 3–16. http://doi.org/10.1080/14790710408668175

Fowler, C. A. (1986). An event approach to the study of speech perception from a direct–realist perspective. *Journal of Phonetics*, *14*(1), 3–28.

García Mayo, M. del P., & Slabakova, R. (2015). Object drop in L3 acquisition. *International Journal of Bilingualism*, *19*(5), 483–498. http://doi.org/10.1177/1367006914524643

Gelman, A., Simpson, D., & Betancourt, M. (2017). The Prior Can Often Only Be Understood in the Context of the Likelihood. *Entropy*, *19*(10), 555. http://doi.org/10.3390/e19100555

Giancaspro, D., Halloran, B., & Iverson, M. (2015). Transfer at the initial stages of L3 Brazilian Portuguese: A look at three groups of English/Spanish bilinguals. *Bilingualism: Language and Cognition*, *18*(2), 191–207. http://doi.org/10.1017/S1366728914000339

Goto, H. (1971). Auditory perception by normal japanese adults of the sounds" l" and" r.". *Neuropsychologia*.

Grosjean, F. (1998). Transfer and language mode. *Bilingualism: Language and Cognition*, *1*(3), 175–176.

Gut, U. (2010). Cross-linguistic influence in L3 phonological acquisition. *International Journal of Multilingualism*, *7*(1), 19–38.

http://doi.org/10.1080/14790710902972248

Hao, Y.-C., & Jong, K. de. (2016). Imitation of second language sounds in relation to L2 perception and production. *Journal of Phonetics*, *54*, 151–168.

Hattori, K., & Iverson, P. (2010). Examination of the relationship between L2 perception and production: An investigation of english/r/-/l/perception and production by adult japanese speakers. In *Second language studies: Acquisition, learning, education and technology*.

Hermas, A. (2010). Language acquisition as computational resetting: Verb movement in L3 initial state. *International Journal of Multilingualism*, *7*(4), 343–362.

Iverson, P., Hazan, V., & Bannister, K. (2005). Phonetic training with acoustic cue manipulations: A comparison of methods for teaching English /r/-/l/ to Japanese adults. *The Journal of the Acoustical Society of America*, *118*(5), 3267–3278. http://doi.org/10.1121/1.2062307

Izura, C., Cuetos, F., & Brysbaert, M. (2016). *Lexical Test for Advanced Learners of Spanish*. American Psychological Association. Retrieved from http://doi.apa.org/getdoi.cfm?doi=10.1037/t47086-000

Jensen, I. N., Mitrofanova, N., Anderssen, M., Rodina, Y., Slabakova, R., & Westergaard, M. (2021). Crosslinguistic influence in L3 acquisition across linguistic modules. *International Journal of Multilingualism*, 1–18. http://doi.org/10.1080/14790718.2021.1985127

Jongman, A., Fourakis, M., & Sereno, J. A. (1989). The Acoustic Vowel Space of Modern Greek and German. *Language and Speech*, *32*(3), 221–248. http://doi.org/10.1177/002383098903200303

Kamiyama, T. (2007). Acquisition of French vowels by Japanese-speaking learners: Close and close-mid rounded vowels. Freiburg, Germany.

Kellerman, E. (1983). Now you see it, now you don't. In *Language Transfer in Language Learning. Issues in Second Language Research.* (pp. 112–134). Rowley, MA: Newbury House Publishers.

Kim, S. Y., & Han, J.-I. (2022). The relationship between perception and production of illusory vowels in a second language. *Second Language Research*, 02676583221135185.

Kissling, E. M. (2014). What predicts the effectiveness of foreign-language pronunciation instruction? Investigating the role of perception and other individual differences. *Canadian Modern Language Review*, *70*(4), 532–558.

Kruschke, J. (2018). Rejecting or Accepting Parameter Values in Bayesian Estimation. *Advances in Methods and Practices in Psychological Sciences*, *1*(2), 270–280. http://doi.org/https://doi.org/10.1177/2515245918771304

Lakens, D. (2017). Equivalence Tests: A Practical Primer for *t* Tests, Correlations, and Meta-Analyses. *Social Psychological and Personality Science*, *8*(4), 355–362. http://doi.org/10.1177/1948550617697177

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

Lemhöfer, K., & Broersma, M. (2016). *Lexical Test for Advanced Learners of English*. American Psychological Association. Retrieved from http://doi.apa.org/getdoi.cfm?doi=10.1037/t47085-000

Leung, Y.-K. I. (2005). L2 vs. L3 initial state: A comparative study of the acquisition of French DPs by Vietnamese monolinguals and Cantonese–English bilinguals. *Bilingualism: Language and Cognition*, *8*(1), 39–61. http://doi.org/10.1017/S1366728904002044

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

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, *21*(1), 1–36.

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

Liu, J., & Lin, J. (2021). A Cross-Linguistic Study of L3 Phonological Acquisition of Stop Contrasts. *SAGE Open*, *11*(1), 215824402098551.

http://doi.org/10.1177/2158244020985510

Liu, Z., Gorba, C., & Cebrian, J. (2019). EFFECTS OF LEARNING AN ADDITIONAL LANGUAGE ON VOT PERCEPTION. In *19th International Congress of Phonetic Sciences*. Melbourne, Australia: Australasian Speech Science; Technology Association Inc.

Llama, R., & Cardoso, W. (2018). Revisiting (Non-)Native Influence in VOT Production: Insights from Advanced L3 Spanish. *Languages*, *3*(3), 30. http://doi.org/10.3390/languages3030030

Llama, R., Cardoso, W., & Collins, L. (2010). The influence of language distance and language status on the acquisition of L3 phonology. *International Journal of Multilingualism*, *7*(1), 39–57. http://doi.org/10.1080/14790710902972255

Llompart, M., & Reinisch, E. (2019). Imitation in a second language relies on phonological categories but does not reflect the productive usage of difficult sound contrasts. *Language and Speech*, *62*(3), 594–622.

McNicol, D. (2005). *A primer of signal detection theory*. Psychology Press.

Meisel, J. M. (1983). Transfer as a second-language strategy. *Language & Communication*, *3*(1), 11–46. http://doi.org/10.1016/0271-5309(83)90018-6

Melnik-Leroy, G. A., Turnbull, R., & Peperkamp, S. (2022). On the relationship between perception and production of L2 sounds: Evidence from anglophones' processing of the french/u/–/y/contrast. *Second Language Research*, *38*(3), 581–605.

Mitrofanova, N., & Westergaard, M. (2018). Wholesale vs. Property-by-property transfer: Acquisition of morphological case in an artificial L3. Braunschweig.

Montrul, S., Dias, R., & Santos, H. (2011). Clitics and object expression in the L3 acquisition of Brazilian Portuguese: Structural similarity matters for transfer. *Second Language Research*, *27*(1), 21–58. http://doi.org/10.1177/0267658310386649

Nagle, C. L. (2018). Examining the temporal structure of the perception–production link in second language acquisition: A longitudinal study. *Language Learning*, *68*(1), 234–270.

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

Okuno, T., & Hardison, D. M. (2016). Perception-production link in L2 japanese vowel duration: Training with technology.

Onishi, H. (2016). The effects of L2 experience on L3 perception. *International Journal of Multilingualism*, *13*(4), 459–475. http://doi.org/10.1080/14790718.2016.1217604

Paradis, M. (2009). *Declarative and Procedural Determinants of Second Languages* (Vol. 40). Amsterdam: John Benjamins Publishing Company. http://doi.org/10.1075/sibil.40

Parma, A. (2017). Cross-Linguistic Transfer of Object Clitic Structure: A Case of L3 Brazilian Portuguese. *Languages*, *2*(3), 14. http://doi.org/10.3390/languages2030014

Parrish, K. (2021). The relative VOT productions of L3 French by Spanish-English bilinguals at first exposure. In *Multiple language acquisition, processing, and development*. Poznań, Poland: Adam Mickiewicz University.

Parrish, K. (2022). The production of L3 stop-initial words by Spanish/English bilinguals. *International Journal of Multilingualism*, 1–18. http://doi.org/10.1080/14790718.2022.2060234

Patihis, L., Oh, J. S., & Mogilner, T. (2015). Phoneme discrimination of an unrelated language: Evidence for a narrow transfer but not a broad-based bilingual advantage. *International Journal of Bilingualism*, *19*(1), 3–16. http://doi.org/10.1177/1367006913476768

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., … Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. http://doi.org/10.3758/s13428-018-01193-y

Plonsky, L. (Ed.). (2015). *Advancing quantitative methods in second language research*. New York: Routledge.

Podlipsky, V. J., & Simácková, S. (2015). Phonetic imitation is not conditioned by preservation of phonological contrast but by perceptual salience. In *ICPhS*.

Puig-Mayenco, E., González Alonso, J., & Rothman, J. (2020). A systematic review of transfer studies in third language acquisition. *Second Language Research*, *36*(1), 31–64. http://doi.org/10.1177/0267658318809147

Reeder, J. T. (1998). English speakers' acquisition of voiceless stops and trills in L2 spanish. *Texas Papers in Foreign Language Education*, *3*(3), 101–18.

Rothman, J. (2010). On the typological economy of syntactic transfer: Word order and relative clause high/low attachment preference in L3 Brazilian Portuguese. *IRAL - International Review of Applied Linguistics in Language Teaching*, *48*(2-3). http://doi.org/10.1515/iral.2010.011

Rothman, J. (2011). L3 syntactic transfer selectivity and typological determinacy: The typological primacy model. *Second Language Research*, *27*(1), 107–127. http://doi.org/10.1177/0267658310386439

Rothman, J. (2013). Cognitive economy, non-redundancy and typological primacy in L3 acquisition: Initial stages of L3 Romance and beyond. In S. Baauw, F. Drijkoningen, L. Meroni, & M. Pinto (Eds.), *Romance Languages and Linguistic Theory* (Vol. 5, pp. 217–248). Amsterdam: John Benjamins Publishing Company. http://doi.org/10.1075/rllt.5.11rot

Rothman, J. (2015). Linguistic and cognitive motivations for the Typological Primacy Model (TPM) of third language (L3) transfer: Timing of acquisition and proficiency considered. *Bilingualism: Language and Cognition*, *18*(2), 179–190. http://doi.org/10.1017/S136672891300059X

Rothman, J., & Cabrelli Amaro, J. (2010). What variables condition syntactic transfer? A look at the L3 initial state. *Second Language Research*, *26*(2), 189–218. http://doi.org/10.1177/0267658309349410

Schiel, F. (1999). Automatic phonetic transcription of non-prompted speech.

Sheldon, A., & Strange, W. (1982). The acquisition of/r/and/l/by japanese learners of english: Evidence that speech production can precede speech perception. *Applied Psycholinguistics*, *3*(3), 243–261.

Slabakova, R. (2017). The scalpel model of third language acquisition. *International Journal of Bilingualism*, *21*(6), 651–665. http://doi.org/10.1177/1367006916655413

Slabakova, R., & García Mayo, M. D. P. (2015). The L3 syntax–discourse interface. *Bilingualism: Language and Cognition*, *18*(2), 208–226. http://doi.org/10.1017/S1366728913000369

Stadt, R., Hulk, A., & Sleeman, P. (2016). The influence of L2 English and immersion education on L3 French in the Netherlands. *Linguistics in the Netherlands*, *33*, 152–165. http://doi.org/10.1075/avt.33.11sta

Stadt, R., Hulk, A., & Sleeman, P. (2018a). The influence of L1 Dutch and L2 English on L3 French: A longitudinal study. *Journal of the European Second Language Association*, *2*(1), 63. http://doi.org/10.22599/jesla.42

Stadt, R., Hulk, A., & Sleeman, P. (2018b). The role of L2 exposure in L3A: A comparative study between third and fourth year secondary school students in the Netherlands. In J. Berns, H. Jacobs, & D. Nouveau (Eds.), *Romance Languages and Linguistic Theory* (Vol. 13, pp. 279–296). Amsterdam: John Benjamins Publishing Company. http://doi.org/10.1075/rllt.13.17sta

Stadt, R., Hulk, A., & Sleeman, P. (2020). L2 Influence in L3 Acquisition: The Role of the L3. In A. Trotzke & T. Kupisch (Eds.), *Formal Linguistics and Language Education* (Vol. 43, pp. 203–221). Cham: Springer International Publishing. http://doi.org/10.1007/978-3-030-39257-4_11

Weber, A., & Cutler, A. (2004). Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language*, *50*(1), 1–25. http://doi.org/10.1016/S0749-596X(03)00105-0

Werker, J. F. (1986). The effect of multilingualism on phonetic perceptual flexibility. *Applied Psycholinguistics*, *7*(2), 141–155. http://doi.org/10.1017/S0142716400007360

Westergaard, M. (2021). Microvariation in multilingual situations: The importance of property-by-property acquisition. *Second Language Research*, *37*(3), 379–407. http://doi.org/10.1177/0267658319884116

Westergaard, M., Mitrofanova, N., Mykhaylyk, R., & Rodina, Y. (2017). Crosslinguistic influence in the acquisition of a third language: The Linguistic Proximity Model. *International Journal of Bilingualism*, *21*(6), 666–682. http://doi.org/10.1177/1367006916648859

Williams, S., & Hammarberg, B. (1998). Language Switches in L3 Production: Implications for a Polyglot Speaking Model. *Applied Linguistics*, *19*(3), 295–333. http://doi.org/10.1093/applin/19.3.295

Wrembel, M. (2010). L2-accented speech in L3 production. *International Journal of Multilingualism*, *7*(1), 75–90. http://doi.org/10.1080/14790710902972263

Wrembel, M. (2011). Cross-linguistic Influence in Third Language Acquisition of Voice Onset Time. In *Proceedings of the 17th International Congress of Phonetic Sciences* (pp. 2157–2160). Hong Kong: City University of Hong Kong.

Wrembel, M. (2012). Foreign accentedness in third language acquisition: The case of L3 English. In J. Cabrelli Amaro, S. Flynn, & J. Rothman (Eds.), *Studies in Bilingualism* (Vol. 46, pp. 281–310). Amsterdam: John Benjamins Publishing Company. http://doi.org/10.1075/sibil.46.16wre

Wrembel, M. (2014). VOT Patterns in the Acquisition of Third Language Phonology. In *Proceedings of the International Symposium on the Acquisition of Second Language Speech* (Vol. 5). Concordia Working Papers in Applied Linguistics,.

Wrembel, M. (2021). Multilingual acquisition property by property: A view from a wider perspective. *Second Language Research*, *37*(3), 441–447. http://doi.org/10.1177/0267658320935167

Wrembel, M., Gut, U., Kopečková, R., & Balas, A. (2020). Cross-Linguistic Interactions in Third Language Acquisition: Evidence from Multi-Feature Analysis of Speech Perception. *Languages*, *5*(4), 52. http://doi.org/10.3390/languages5040052

Wrembel, M., Gut, U., Kopečková, R., & Balas, A. (2022). The relationship between the perception and production of L2 and L3 rhotics in young multilinguals; an exploratory cross-linguistic study. *International Journal of Multilingualism*, 1–20.

Wrembel, M., Marecka, M., & Kopečková, R. (2019). Extending perceptual assimilation model to L3 phonological acquisition. *International Journal of Multilingualism*, *16*(4), 513–533. http://doi.org/10.1080/14790718.2019.1583233

Wunder, E.-M. (2010). Phonological cross-linguistic influence in third or additional language acquisition. In *Proceedings of the 6th International Symposium on the Acquisition of Second Language Speech* (pp. 566–571). Poznań, Poland: Adam Mickiewicz University.