**Title**

Trilingual Production and Perception of Lexical Stress: Extending the Cue-weighting Transfer Hypothesis to L3 Acquisition



**Author**

Helms, Annie Grey

**Publication Date**

2023

Peer reviewed|Thesis/dissertation

Trilingual Production and Perception of Lexical Stress: Extending the Cue-weighting Transfer Hypothesis to L3 Acquisition

By

Annie Grey Helms

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Romance Languages and Literatures

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Associate Professor Justin Davidson, Chair
Professor Keith Johnson
Associate Professor Mark Amengual
Assistant Professor Isaac Bleaman

Fall 2023

Trilingual Production and Perception of Lexical Stress: Extending the Cue-weighting Transfer Hypothesis to L3 Acquisition

## Abstract

Trilingual Production and Perception of Lexical Stress: Extending the Cue-weighting Transfer Hypothesis to L3 Acquisition

by

Annie Grey Helms

Doctor of Philosophy in Romance Languages and Literatures

University of California, Berkeley

Associate Professor Justin Davidson, Chair

This dissertation analyzes the production and perception of lexical stress in trilinguals' first, second, and third languages (L1, L2, and L3) to evaluate how the cue-weighting transfer hypothesis applies to L3 acquisition. According to this hypothesis, acoustic cues to stress have different weights across languages, and results from both production and perception studies of bilingual speakers indicate that bilinguals transfer cue-weightings from the L1 to the L2 (Chrabaszcz et al., 2014; Ingvalson et al., 2012; Iverson et al., 2003; Tremblay et al., 2021; Zhang & Francis, 2010). Acoustic correlates of lexical stress were analyzed in English, Spanish, and Catalan, as produced and perceived by L1 English-L2 Spanish-L3 Catalan speakers. Productions of stress minimal pairs in the three languages were collected in three separate experimental sessions via a sentence elicitation task, and perceptions of stress were gathered in three separate experimental sessions via a word identification task using nonword stress minimal pairs.

This study focuses on acoustic cue-weightings to word-level prominence in English, Spanish, and Catalan, which all have lexical stress as indicated by the existence of stress minimal pairs and which are reported to belong to different rhythm classes (Prieto et al., 2012). Little, if any, research has been done to investigate how the cue-weighting transfer hypothesis may extend to the L3, which in turn allows for L2 transfer to be better understood (Flynn et al., 2004). Whereas vowel quality is the main correlate to lexical stress in English, duration is the most prominent cue in Spanish, and in Catalan, cue-weighting is vowel-dependent, with duration and vowel quality being prominent cues. The majority of models of multilingual phonetic and phonological acquisition that posit some degree of interaction between acquired language systems have been theorized to extend to L3 acquisition in a similar manner (e.g., Amengual, 2021; Chan & Chang, 2019; de Leeuw & Chang, in press; Escudero et al., 2013; Wrembel et al., 2019). Accordingly, I hypothesize that cue-weightings from previously learned language(s) will transfer into the L3 in similar ways that cue-weightings from the

L1 transfer into the L2. Specifically, I hypothesize that cue-weightings in each language will be mediated by relative language dominance. I additionally predict that transfer will be bidirectional, where cue-weightings in the L2 and the L3 can influence cue-weightings in the L1.

The results of the production and perception tasks indicate that relative language dominance does affect cue-weighting to a different extent in each language in the trilinguals' repertoires. There was additionally evidence for regressive transfer of cue-weighting in both production and perception, indicating that all languages in a trilingual's repertoire are susceptible to crosslinguistic influence. Principal component analysis is shown to be a viable way to extend the Bilingual Language Profile (Birdsong et al., 2012) to obtain relative dominance scores for trilinguals. Lastly, through a comparison of theoretical frameworks of L3 phonetics and phonology, the Attrition & Drift in Access, Production, and Perception Theory (ADAPPT; de Leeuw & Chang, in press) was determined to most closely align with the findings of this study.

To my dad, who has always supported me.

# Contents

# List of Figures

# List of Tables

## Acknowledgments

I want to start by thanking Danny, thank you for all you did over the last five and a half years. You were my biggest supporter as I applied to grad school, made you move to California with me, spent a summer in Barcelona, taught for the first time, hogged the desk during COVID, took my qualifying exams, paid research expenses out of pocket until I got fellowships, and wrote this dissertation. You even let me make puns that were not remotely funny without the linguistic/statistic context that you so graciously let me supply you with.

I also want to thank my dissertation committee, starting with Justin Davidson. Thank you for everything that you have done for me and for linguistics in the department over the last five years. When I started, I didn't have much direction, or much of an idea about what research in linguistics looks like. I first heard about quantitative linguistics from you and wrote my first little lines of code in R in your methods class. Thank you for always repping Catalan, for connecting me with people in Barcelona, for helping me design my first ever study of bilingual phonetics. You challenged me to always keep up side research projects, to submit to conferences before I really felt qualified, and to publish. Your feedback has strengthened my writing so much and has helped me think more critically as a researcher.

Mark Amengual, thank you for your confidence in me and for giving me so many opportunities to learn from you. You went out of your way to make me feel welcomed in the research world after my first solo conference presentation at LSRL in 2020. I admire you for being a researcher that is more passionate about advancing the discipline than promoting your own work. You have been such a great mentor to me, and it is because of your guidance and encouragement to submit something to New Sounds 2022 that I landed on this dissertation topic. Thank you also for giving me the opportunity to work on the handbook with you. It has been so rewarding to gain a holistic knowledge base of bilingual phonetics and phonology and to be able to dialogue with you regularly about a variety of topics.

Keith Johnson, I had no idea what I was in for when I took experimental phonetics with you. I learned that finding out what I don't know can be just as important as knowing what I do, and I sure found out about a lot of things that I didn't know. You have always challenged me as a researcher to think critically about research questions based in theory and to not settle with a known partial solution if I can learn an unknown way to do things better. Thank you for your confidence in me to learn new things, from Python to Emacs to Bayesian modeling. You always demonstrated genuine interest in my research questions and excitement to work with new methodologies. You are truly a life-long learner and inspire me to be the same.

Isaac Bleaman, thank you for being such an approachable resource for collaborative research, the publishing process, and the job market. I learned so much from you about sociolinguistics, statistical models, and getting creative with corpus-based research projects. I really enjoyed collaborating with you and Katie Cugno and am so thankful for that opportunity.

Gemma Repiso-Puigdelliura, you have been a valuable resource for discussing research ideas and providing feedback on this dissertation work. Thank you also for your collaboration and assistance with data collection – I couldn't have done it without you!

Ernesto, you are the sole reason this dissertation was written in LaTeX. My secret wish is that you will leave academia and we'll get jobs at the same company. But in the meantime, I continue to be impressed with your passion for research and helping others in academia. You were the best office-mate I could ever have asked for. Some of my best memories in grad school are sitting in our office, looking up phonological inventories of random languages on Wikipedia and planning some amazing projects we never got to carry out. Thank you for being my friend and always sparking my curiosity with research.

Gabriella, thank you for firstly for being such a good friend and for being willing to talk about life outside of grad school. You felt like my grad school big sister, always paving the way with conferences, courses, academic milestones, internships, jobs, etc. Collaborating on research can be tricky, but I've always enjoyed working with you. It's been great to go through all of this with you!

I also want to thank other great sources of emotional and academic support in Rachel, Anahit, Ben, Rebekah, Alyssa, Zach, Emily, Jordan, and Theo. Thanks for always being willing to grab lunch or coffee.

I've had the privilege of working with some truly wonderful undergraduate research assistants at Berkeley – thank you Da Kyung Lee, Shantal Saloma Gomez, Julian Vargo, Chanhee Lee, Averie Haneul Burke, Erin Krantz, and Anna Knall for all your help with audio annotations. I also want to thank Yenni Cadenas Leon, Maria Castillo Moreno, and Kirk Monzon-Mendez for their help conducting data collection sessions in Spanish.

Thank you Anna Closas, Zach Tronstad, and Ernesto Gutiérrez Topete for providing your voices for my perception experiment. And thank you, Laura Sparkman, for helping me with data collection in Spain. Scott James Perry, thanks for all your work in training an acoustic model for Catalan and for being so generous to share it with me. I also want to thank one of my participants for her enthusiasm in sharing information about this project to several Catalan language social media accounts. I could not have gotten so many participants without you!

*A més, Francina Sorell Monserrat i Ainhoa Marcos González, moltíssimes gràcies per ajudar-me amb la recollida de dades en català i castellà. Molta sort amb tot!*

# Chapter 1

# Introduction

This introductory chapter provides an overview of the field of third language (L3) phonetics and phonology, including typical methodologies, variables of study, and commonly studied languages. A brief overview of theoretical models of second language/bilingual phonetic and phonological acquisition and representation that posit distinct relationships between the first language (L1) and the second language (L2) will be provided, in addition to their proposed extensions to L3 acquisition. Lastly, the chapter situates the current study within the context of this literature and outlines the subsequent chapters of this dissertation.

## 1.1 L3 phonetics and phonology

Many principles of L2 acquisition can and should be carried over to L3 acquisition, though the process of L3 acquisition is inherently more complex given the speaker's heightened metalinguistic knowledge and broadened phonetic system (Wrembel et al., 2020). Therefore, L3 acquisition is not simply treated as a special case of L2 acquisition, but is more nuanced and methodologically more complex to study. As in L2 studies, the majority of L3 studies in the literature focus on production rather than perception (Onishi, 2016), contain an analysis of only one variable, focus on segmental rather than suprasegmental acquisition, and examine the influence of the L1 and L2 on the L3, rather than examining multidirectional influence among the three languages (Amengual, 2023; Cabrelli Amaro & Wrembel, 2016; Domene Moreno, 2021; Wrembel et al., 2020). In addition, most studies involve language triads in the Germanic, Romance, or Slavic language families, and analyses of languages that are not Indo-European are relatively rare (Cabrelli Amaro & Wrembel, 2016).

Studies of L3 acquisition of phonetics and phonology can not only provide evidence for the range of factors which contribute to crosslinguistic influence (henceforth CLI; here inclusive of 'transfer', 'interference', 'borrowing', etc.) and the directionality of this influence, but also can provide further insights into the process of language acquisition that studies of L1 and L2 acquisition alone cannot provide (Flynn et al., 2004). L3 studies can also help uncover parts of L2 acquisition models that should be updated and expanded, such as the relationship between production and perception, enriching the field's understanding of phonetic representation and the connections between language systems. Methodologically, there are a variety of approaches to studying L3 phonetics and phonology, including longitudinal studies (e.g., Kopečková, 2016; Wrembel et al., 2020), an analysis of all three of a trilinguals' languages rather than just the L3 (e.g., Kopečková, 2016; Kopečková et al., 2016; Llama & López-Morelos, 2016), and studying the acquisition of an L3 by heritage speakers and early bilinguals, in addition to sequential bilinguals (e.g., Kopečková, 2016; Kopečková et al., 2016; Llama & López-Morelos, 2016, 2020).

## Common variables of study

Language mode is a variable that has increasingly been recognized as theoretically and methodologically important in studies of L2 production and perception (e.g., Amengual, 2021), and for the same reasons is important for studies of L3 production and perception (Amengual, 2023). Language mode is "the state of activation of the bilingual's languages and language processing mechanisms at a given point in time" (Grosjean, 2011, p. 12), and is conceived of as a continuum of states ranging from bilingual (multilingual) to unilingual. In a multilingual mode there is maximal activation of two or more of the speaker's languages and speakers may exhibit transfer between their languages that is dynamic (temporary, induced by cognitive factors) and static (a reorganization of the language system due to long-term contact). In a unilingual mode, the coactivation of languages is minimized (though not completely absent), thus isolating effects of static transfer from dynamic transfer. Language mode is influenced by the specifics of the communicative context, for example, the language profile of the interlocutor (or researcher conducting the experiment), the base language of interaction, the presence of other languages during an interaction, the function of the interaction, and whether bilinguals know they are being tested as bilinguals (Grosjean, 2001, p. 5).

Language dominance is another important factor in studies of L2 and L3 acquisition. Language dominance can be measured in a variety of ways, but is generally considered to be shaped by factors such as linguistic history, linguistic attitudes, language proficiency, and language use (Gertken et al., 2014). Often relative language dominance and order of acquisition coincide, but this is not usually the case among heritage speakers or among simultaneous bilinguals. A common tool to study language dominance is the Bilingual Language Profile (BLP; Birdsong et al., 2012), which has been used in numerous studies of bilingual speakers and listeners. The relative dominance score for bilinguals is computed as the difference between the dominance score for each language, a technique that cannot be straightforwardly applied to trilinguals. However, components of the relative dominance score obtained from the BLP, such as individuals' attitudes towards each language, have successfully been used as predictors of bilingual perception and production (e.g., Law et al., 2019), and therefore approaches which group these components according to their overlap in explained variance may be successful with trilinguals.

## 1.2 Theoretical frameworks

The first studies of L3 acquisition began in the domains of syntax and the lexicon, though L3 research gradually expanded to include phonetics and phonology (Domene Moreno, 2021, p. 18). These early morphosyntactic studies contributed to L3 models of transfer such as the Cumulative Enhancement Model (CEM; Flynn et al., 2004), the L2 Status Factor Model (L2SF; Bardel & Sánchez, 2017), and the Typological Primacy Model (TPM; Rothman, 2015). According to CEM, all previously learned languages can influence the L3, that is, the L1 does not hold special status. In contrast, L2SF postulates that transfer to the L3 is predominately from the L2 at early stages of learning since the languages are acquired in a more similar manner (e.g., perhaps both were acquired in a classroom setting). Lastly, TPM states that either the L1 or the L2 can influence the L3, but transfer is primarily determined by the typological similarity between languages.

Some studies have pitted the predictions of these frameworks against each other to determine whether these models can be applied to cases of L3 phonetics and phonology (e.g., Chan & Chang, 2019; Patience, 2018, as will be discussed later in this chapter). However, as will be detailed in Section 1.3, it has been shown that CLI is primarily *feature*-dependent, rather than *language*-dependent, where transfer is not uniform, patterning "according to the status of the background languages as units or based on the presence or absence of features in the background languages" (Domene Moreno, 2021, p. 190). Rather, transfer is conditioned by factors specific to the phonetic/phonological features in question. Therefore, studies of L3 phonetics and phonology have primarily focused on extending theoretical models that account for perception and production of individual features in an L1 and/or an L2. The following subsections provide a general overview of such frameworks and ways they have been extended to account for L3 phonetic and phonological acquisition and representation.

### SLM and SLM-r

The Speech Learning Model (SLM; Flege, 1995) and the Revised Speech Learning Model (SLM-r; Flege & Bohn, 2021) postulate that bilingual speakers do not maintain separate phonetic systems for each language, rather, the two systems coexist in a mutual phonetic space and may influence one another. Both the SLM and SLM-r focus on the discrimination of segmental contrasts and state that if differences between an L2 category and an L1 category are perceived, then a unique phonetic category will be established and L1 phonetic categories will shift away from this new category in phonetic space to maximize discriminability. If phonetic differences between an L2 category and an L1 category are *not* perceived by a bilingual individual, the formation of a new category will be blocked. Blockage of category formation results in assimilation and equivalence classification, by which the speaker may produce a merged L1-L2 category (Evans & Iverson, 2007; Kendall & Fridland, 2012). The acoustic properties of this composite category are "defined by the statistical regularities present in the combined distributions of the perceptually linked L1 and L2 sounds" (Flege & Bohn, 2021, p. 41). Phonetic properties of the L1 phone and the L2 phone are activated when a bilingual or L2 learner encounters this phone in spoken or written form, as the category is a composite of the L1 and L2 phonemes, allowing for bidirectional influence between the L1 and L2. The main ways in which the SLM-r departs from the SLM are two-fold. Firstly, the SLM-r no longer holds that accurate perception precedes accurate production. Secondly, the goal of the SLM-r is to account for how the shared phonetic space reorganizes over the individual's lifetime given the quantity and quality of input they receive, whereas the goal of the SLM is to provide an account for age-related limits on target-like L2 production.

In a study of voice onset time (VOT; the length of time between the release burst of a stop and the onset of voicing of the following vowel), Amengual (2021) tested how the predictions of SLM and SLM-r would hold for L3 acquisition. Amengual (2021) examined productions of /k/ in Spanish, English, and Japanese, which are expected to manifest as short lag, long lag, and 'mid' lag, respectively. To analyze the role of language mode and cognate status in bilingual and trilingual production, L1 English-L2 Japanese, L1 Japanese-L2 English speakers, and L1 Spanish-L2 English-L3 Japanese speakers were recorded in a carrier phrase task in separate language mode conditions. A unilingual language mode was experimentally induced in each of the three monolingual sessions, using language-specific instructions and experimental stimuli. In the bilingual and trilingual sessions, bilingual and trilingual language modes were respectively induced by mixing the languages used in the experimental instructions and the sentence elicitation tasks and by including cognates in the stimuli. Amengual (2021) found that each group produced language-specific VOT values, where bilinguals showed dynamic transfer from their L1 to their L2 due to the experimentally induced bilingual language mode. However, the L3 Japanese trilingual speakers produced /k/ in a more native Japanese-like manner than the L2 Japanese bilinguals, suggesting a bilingual advantage during L3 acquisition. No evidence for static transfer across languages was found in the monolingual sessions. Based on the results of the study, Amengual concluded that the SLM and SLM-r can be extended to L3 acquisition, where L3 category formation is dependent on a learner's ability to perceive acoustic differences between the category and similar L1 and L2 sounds. To fully account for the results found in this study, Amengual (2021) concluded that language mode also needs to be accounted for in models of both L2 and L3 acquisition, where phonetic interference is induced in the bilingual and trilingual language modes.

## PAM-L2

The Perceptual Assimilation Model for L2 (PAM-L2; Best & Tyler, 2007) applies the Perceptual Assimilation Model (PAM; Best, 1991) to cases of L2 perceptual learning. The models focus on the discrimination of segmental contrasts, and PAM-L2 compares the discrimination of naïve listeners to L2 listeners. Best and Tyler state that "L1-L2 differences at a gestural, phonetic, or phonological level may each influence the L2 learner's discrimination abilities, separately or together, depending on the context or the perceiver's goals" (Best & Tyler, 2007, p. 25). Whereas the SLM-r focuses on the exposure of L2 input to distinguish production and perception among bilingual speakers, PAM-L2 maintains the Critical Period

Hypothesis, where adult L2 learners are fundamentally different from child L2 learners and cannot attain native-like performance. Whereas PAM-L2 discusses acquisition at both the phonetic and phonological level, the SLM-r has pivoted on claims from the SLM regarding phonology, and views the L1 and L2 as sharing only a common phonetic space.

Wrembel et al. (2019) extended PAM-L2 to L3 acquisition, using perceptual data obtained from L1 German-L2 English-L3 Polish speakers, some who are sequential trilinguals and some who are heritage speakers of Polish for whom Polish is L3 in terms of dominance. A group of early German-Polish bilinguals served as a control group. The researchers predicted that the trilingual speakers would undergo the same speech learning processes described in PAM-L2, but that they would have a perceptual advantage, given their broadened phonetic repertoire. The trilinguals' perception of tense-lax vowel distinctions, which exist in German and English but not in Polish, the /e æ/ distinction which is unique to English, and the contrast of retroflex sibilants in Polish were analyzed. In particular, the researchers were concerned with (1) whether L3 Polish sounds assimilate to similar English and German categories; (2) whether L3 learners have an advantage over L2 learners when discriminating the retroflex and palato-alveolar sibilants in Polish; and (3) and whether heritage speakers perform differently from their sequential learner peers. They hypothesized that (1) assimilation to L1 and L2 sounds will occur given the predictions of PAM-L2; (2) that the bilingual advantage will allow learners to distinguish between Polish sibilants; and (3) that heritage speakers will outperform sequential learners. The results confirmed the first and second hypotheses, in that L3 categories are more readily assimilated to the L2 than the L1, and that within the PAM-L2 framework, all trilingual participants behave like advanced L2 listeners rather than beginner L2 learners. However, the results did not support the third hypothesis that heritage speakers would have an advantage over sequential learners.

Wrembel et al. (2019) suggested that PAM-L2 can be extended to L3 phonological acquisition, where L3 categories assimilate to both L1 and L2 categories, with a possible preference for the latter. This study provides evidence for a bilingual advantage in the process that PAM-L2 describes as single-category assimilation. Whereas L2 learners may not discriminate highly similar sounds in a new language, L3 learners may form separate categories over time. For this reason, the authors equate beginner L3 learners with advanced L2 learners.

## L2LP

In the revised Second Language Linguistic Perception model (L2LP; Van Leussen & Escudero, 2015), previous model assumptions (Escudero, 2005, 2009) were updated to account for L2 acquisition and development being driven by meaning rather than awareness, because learners' ability to infer L2 phonological categories has been shown to be dependent on their ability to understand a speaker's meaning (Van Leussen & Escudero, 2015, p. 3). Like PAM-L2, the SLM, and the SLM-r, L2LP focuses on the perception of segmental contrasts, but unlike both PAM-L2 and the SLM-r, L2LP attempts to account for the entire learning process, from naïve stage (PAM and PAM-L2) to advanced learner (SLM and SLM-r), incorporating "separate but linked representations for *perceptual* and *lexical* contrast"

(Van Leussen & Escudero, 2015, p. 3, emphasis in original). Among the central tenets of the L2LP model is the optimal perception hypothesis, which states that listeners are preferentially attuned to cues which are informative in a given context in a given language (i.e., cue-weighting). Another principle of L2LP is the full copying hypothesis, which states that the initial state of L2 perception is the end-state of their L1 perception, but that language input can shift L2 perception over time.

L2LP has been applied to cases in which suprasegmental features (e.g., pitch variation) in the L1 may affect L2 perception as speakers' initial state of L2 perception is essentially a copy of their L1 perception routine (Escudero et al., 2022). In an extension to L3 acquisition, Escudero et al. (2013) examined the role of L2 and L3 proficiency on L3 word learning and found that when features in the L2 and L3 were perceived as similar, cues that were useful for perceiving contrasts in the L2 (e.g., durational cues) could also be used in the L3. More research is needed, however, to determine the relative weight of the L1 and L2 in the initial state of L3 learning, and factors which may mediate these relative weights.

## ADAPPT

The bidirectional nature of crosslinguistic influence has not been readily accounted for by models of bilingual phonetics and phonology, but Attrition & Drift in Access, Production, and Perception Theory (ADAPPT; de Leeuw & Chang, in press) was theorized specifically to address cases of L1 attrition (long-term effects) and drift (short-term effects). In fact, the model states that the L1 will always be impacted by L2 exposure and acquisition, and that both short- and long-term changes are possible, given proper exposure conditions and experimentally induced language mode. ADAPPT additionally holds that both the L1 and L2 are susceptible to change at any point in an individual's lifespan. Whereas other models (e.g., PAM-L2, L2LP, SLM) predict that accurate perception is a prerequisite for accurate production, ADAPPT does not predict a straightforward link between perception and production. Furthermore, ADAPPT states that all dimensions of L1 and L2 speech in a bilingual's repertoire are candidates for change, including suprasegmental contrasts and processes. Although the theory is relatively new and has not been specifically tested in a unified manner, ADAPPT's predictions are based on a myriad of studies of bilingual speech, which de Leeuw and Chang broadly define as the speech of an individual with two or more languages or dialects in their repertoire. Accordingly, the predictions are not exclusive to L1 and L2 speech, but could conceivably be extended to L3 speech.

## ASP

The Automatic Selective Perception model (ASP; Strange, 2011) states that L1 listeners develop specific perception routines for L1 perception that allow for efficient and automatized processing. These routines may be used in the early stages of L2 perception and processing, but over time, listeners can develop new perception routines for L2 perception with sensitivities to cues that provide for more efficient processing in the L2. This prediction is similar to the optimal perception hypothesis of the L2LP, but places more emphasis on perception of acoustic cues rather than production. Additionally, unlike L2LP, ASP postulates that bidirectional influence between processing routines can occur, in which L2 processing routines can influence L1 processing.

Chan and Chang (2019) studied the early stages of L3 tonal perception (i.e., naïve tonal perception) in Thai and Yoruba by sequential bilingual speakers of different language backgrounds to determine the relative roles of acquisition order and crosslinguistic similarity, testing the relative predictions of CEM, L2SF, TPM, and ASP. To examine acquisition order, mirror-image Mandarin-English bilinguals were studied (i.e., L1 Mandarin-L2 English and L1 English-L2 Mandarin), and to examine crosslinguistic similarity, L1 English bilinguals with a nontonal L2 were also included in the study. Accuracy and reaction time of tonal perception was measured in an oddity discrimination task and a similarity rating task. Chan and Chang (2019) found that bilinguals with Mandarin as an L1 had more accurate perception than those with Mandarin as an L2, and that bilinguals with L2 Mandarin had marginally higher accuracy than bilinguals with a nontonal L2. Rather than provide support for one of the three models developed for L3 morphosyntactic acquisition (i.e., CEM, L2SF, and TPM), the authors' findings align most closely with the ASP, where bilinguals with Mandarin as their L1 were particularly adept at L3 tonal perception, due to their heightened sensitivity to pitch as a correlate of tone in Mandarin. The weak advantage among bilinguals with Mandarin as an L2 suggests either that there is weaker transfer from the L2 to the L3 under the ASP, or that these bilinguals did not have sufficient proficiency or exposure to Mandarin to develop very robust L2 perception routines, and thus they were not particularly attuned to pitch.

## Episodic framework of exemplar models

Exemplar models of language use and cognitive representation (Bybee, 2000; Johnson, 1997; Pierrehumbert, 2001) provide another framework from which to view the nonselective nature of language activation. In the mind of the speaker-hearer, "all phonetic variants of a word are stored in memory and organized into a cluster: exemplars that are more similar are closer to one another than to ones that are dissimilar, and exemplars that occur frequently are stronger than less frequent ones" (Bybee, 2002, p. 271). Exemplars are stored in memory as a set of auditory properties and a set of category labels, where the labels may include any classification important to the perceiver (Johnson, 1997, p. 147), such as the gender of the speaker or the orthographic representation of the sound. Exemplars cluster together in clouds based on similarity and the strength of an exemplar is correlated to its frequency of experience (Bybee, 2002). Although this framework was not originally described in terms of multilingual phonetic and phonological representation, several studies have provided evidence supporting the coactivation of exemplars from different source languages.

Amengual (2012) examined VOT productions in Spanish-English bilinguals across the factor of cognate status and observed longer VOT productions in Spanish for cognate words among bilinguals. He concluded that "[t]he cascaded activation models will need to be revised to include activations at the phonetic level for both languages that must be connected and activated by the non-selected lexical nodes" (Amengual, 2012, p. 527). Additionally, using usage-based models of language representation, cognate representations may be stored in the same exemplar cloud, allowing for the orthographic, phonological, and semantic representation of the cognate to influence production in the other language. A further extension to L3 representation can be imagined by the same rationale.

### Summary of theoretical frameworks

Though progress has been made to extend these theoretical models to account for L3 acquisition, more studies of trilingual speakers are needed to evaluate these models' predictions for bilingual speech and how possible L3 extensions of each model compare. Additionally, due to the importance of language mode and the necessity of analyzing data from all languages in a multilingual's repertoire, future work needs to employ more careful methodologies in order to distinguish static transfer from dynamic transfer and better compare predictions of these theoretical models. Accordingly, the results from this dissertation research are discussed in light of these theoretical models to determine which can most readily be used to explain L3 acquisition for the multilinguals in question and the specific features of this study.

## 1.3 Findings within L3 phonetics and phonology

Various approaches to studying L3 acquisition of phonetics and phonology have been undertaken, with conflicting results. The following review offers a summary of findings of segmental and suprasegmental acquisition, relating to sources and directions of CLI (crosslinguistic influence), as well as the roles of language dominance and language mode.

### Segmental acquisition

#### Morphosyntactic models of L3 acquisition are insufficient

Patience (2018) studied the production of the Spanish tap [ɾ]/trill [r] distinction by L1 Mandarin-L2 English-L3 Spanish speakers. In order to tease apart transfer from the L1 versus the L2, rhotics were analyzed in all three languages, including the English flap [ɾ]. Using predictions from earlier L3 models of morphosyntactic acquisition, and given the acoustic similarity between English [ɾ] and Spanish [ɾ], Patience hypothesized that whether transfer can come from both the L1 and the L2 (following the CEM), predominately from the L2 at early stages of learning (following the L2SF), or whether transfer is dependent on typological similarity (following the TPM), acquisition of the L2 English flap would facilitate acquisition of the L3 Spanish tap. The alternative hypothesis Patience put forward was that [l] substitutions from L1 Mandarin phonology would primarily occur if the L1 takes precedence in transfer. The results are more consistent with the latter hypothesis, though some evidence for L2 transfer was found. Patience (2018) concluded that separate models for L3 acquisition of phonetics and phonology should be developed (see also Cabrelli Amaro & Wrembel, 2016; Chan & Chang, 2019), as results are not fully supported by the CEM, L2SF, or TPM, which were originally developed to account for morphosyntactic transfer.

### CLI is bidirectional

In a study of VOT, Wrembel (2015) analyzed L3 French VOT productions by L1 German-L2 English learners to observe the sources of CLI. English and German are both aspirating languages which differentiate voiceless aspirated stops and voiceless unaspirated stops, whereas French is a voicing language which differentiates voiceless unaspirated stops and voiced unaspirated stops. Regarding VOT, /p t k/ are considered long lag in English, 'mid' lag in German, and short lag in French. Results indicate CLI in the production of L3 French VOT, where stops were produced as long lag, though not as long as either L2 English or L1 German VOT. However, given that the participants' VOTs in the L1 and the L2 were not significantly different, possible influence from the L1 and L2 could not be teased apart. The author interprets the assimilation of L2 English and L1 German VOT production as possible evidence for bidirectional transfer, where properties of the learners' L2 transfer to their L1. However, all three language recordings were obtained in the same experimental session, with a few minutes in between each recording. Additionally, the experimenter conversed with the participants in the language of the subsequent recording task, thereby inducing a multilingual language mode. Based on this methodology, it cannot be determined whether the results are due to static transfer or dynamic transfer induced by the experimental design.

### Role of language use and language dominance

Kopečková et al. (2016) examined the role of language use and context in a study of trilingual vowel productions by heritage Polish speakers in Germany who learned English in school, as well as sequential L1 German-L2 English-L3 Polish learners. Across the three languages, the vowels /i ɪ u ʊ ɛ æ/ have different phonemic and allophonic status and differ according to vowel quality and length. According to the Dynamic Systems Theory (DST; De Bot, 2012; De Bot et al., 2007; Van Geert, 2008), a trilingual's languages are interconnected, marked by variability, and change over time due to a number of internal and external factors. From this framework, the researchers hypothesized that heritage speakers would show different organizations of their vowel systems compared to sequential learners, and that there would be great variability within these groups. Word reading tasks were used for all three language sessions, with delayed repetition tasks also being used in the English and Polish sessions and a story retelling task in the German session. The results showed great variability between and within trilingual profiles, and statistical analyses of vowel formants and duration indicated that language use and context can partially account for variability seen between speaker groups. Kopečková et al. (2016) took these results as evidence for the interconnected nature of the multilinguals' phonological subsystems which are impacted by the external environment.

Llama and López-Morelos (2016) also examined heritage speakers' acquisition of an L3 in a study of VOT productions of voiceless stops by heritage Spanish speakers that use English in school (and are dominant in English) and are learning French as an L3. Whereas English voiceless stops are categorized as long-lag and aspirated, Spanish and French are categorized as short-lag and unaspirated. A unilingual language mode was induced by gathering VOT productions in English, Spanish, and French in three separate experimental sessions. A picture-naming task was used to obtained VOT productions in stressed, word-initial positions. Productions in each language by the target participant group were compared with productions from native speakers of each language to determine both how target-like productions were in each language, and whether separate VOT ranges were employed in each language. Results indicated that speakers maintained separate VOT ranges in Spanish and English, and that these productions were comparable to those of monolingual speakers of each language. However, their L3 French VOT productions were not statistically distinct from their English productions, but were statistically distinct from productions by monolingual French speakers. The authors provide two possible explanations for these results. On one hand, since the trilinguals are more dominant in English than in Spanish, this could be evidence that language dominance plays an important role in determining the source of crosslinguistic influence in the L3. On the other hand, these speakers are predominantly exposed to French spoken by their classmates, who are also dominant speakers of English. Accordingly, their L3 productions could be influenced by the variety of French to which they have been exposed, which is English-accented. In either case, careful observations of language dominance and history of language exposure are critical when studying L3 production.

### L2 status is a stronger predictor of the source of CLI than typology

To tease apart the relative contributions of language status (specifically L2 status) and language typology on L3 production, Llama et al. (2010) examined VOT productions of voiceless stops in English, Spanish, and French by L1 French-L2 English-L3 Spanish speakers and by L1 English-L2 French-L3 Spanish speakers. Whereas Spanish and French voiceless stops are short-lag and unaspirated, English voiceless stops are long-lag and aspirated. Word lists were used to elicit productions of word-initial voiceless stops in participants' L3 (Spanish) and L2 (either French or English). The results point to combined influence from the L1 and the L2 on the L3, though L2 influence seemed to play the biggest role in L3 productions. Because a mirror-image design was used with the two target trilingual groups, the authors suggest that the results point to L2 status having a greater role in determining sources of CLI than typological distance. Although productions from speakers' L3 and L2 were analyzed, an analysis of L1 productions and a comparison of productions to monolingual or bilingual baselines could lend stronger support to the authors' conclusions.

### Phonological Permeability Hypothesis

Cabrelli Amaro (2017) examined the production of word-final unstressed vowels in L3 Brazilian Portuguese by L1 English-L2 Spanish and L1 Spanish-L2 English trilinguals with the goal of analyzing the amount of regressive crosslinguistic influence (rCLI) of the L3 on L1 or L2 Spanish, and therefore, the relative instability of the L2 compared to the L1. According to the Phonological Permeability Hypothesis (PPH; Cabrelli Amaro & Rothman, 2010), language systems acquired later are less stable than systems acquired earlier. Whereas word-final vowels undergo reduction in Brazilian Portuguese, vowels retain their quality irrespective of stress in Spanish (however, there is evidence that monolingual Spanish speakers do in fact reduce unstressed vowels, e.g., Quilis & Esgueva, 1983; Romanelli et al., 2018). Language mode was accounted for, where productions of word-final vowels in Spanish and Brazilian Portuguese were recorded on different days using a delayed repetition task. Formant frequencies, durations, and relative intensities were measured for each word-final vowel, and results suggest that L1 English trilinguals have more back vowel raising in Spanish than L1 Spanish trilinguals and the monolingual Spanish controls, potentially evidencing influence from L3 Brazilian Portuguese on L2 Spanish. However, an analysis on the individual level revealed potential influence from the L3 in the L2 for both trilingual groups in terms of vowel backness, height, and duration. Cabrelli Amaro (2017) took these results as preliminary support for the PPH.

In the same study, Cabrelli Amaro (2017) also examined the perception of word-final vowel reduction in L3 Brazilian Portuguese by L1 English-L2 Spanish and L1 Spanish-L2 English trilinguals. A forced-choice preference task was used to determine whether the Spanish allophone (unreduced) or the Brazilian Portuguese allophone (reduced) sounded most natural in the given language. There was no evidence of L3 influence in the acceptability of L3-influenced L2 mid vowels, that is, listeners did not evidence increased acceptability of word-final vowel reduction in Spanish with increasing L3 proficiency. The author postulates that the Phonological Permeability Hypothesis may be supported in a weak form, in which later systems are only unstable in production, but not perception.

### CLI is feature- and language-dependent

In a longitudinal study, Wrembel et al. (2020) examined the L2 and L3 speech perception of rhotics and final obstruent voicing of L1 Polish-L2 English-L3 German learners. As the researchers point out, the design of this study contributes to the L3 literature, which is deplete in perceptual studies, by examining multiple features over a longer period of time. Rhotics and final obstruent voicing were selected because the three languages differ across these dimensions; Polish has an alveolar trill [r] and obstruent devoicing in syllable-final position, English has the retroflex approximant [ɹ] and no final obstruent devoicing, and German has uvular rhotics, produced either as a trill [ʀ] or a fricative [ʁ] and final obstruent devoicing. Wrembel et al. hypothesized that the degree of CLI will be dependent on the language (L2 or L3) and the feature (rhotic or final obstruent voicing), and that English will have less CLI as the more stable language compared to the L3. Additionally, the researchers hypothesized that CLI will change over time, with different patterns in the L2 and L3. In a forced-choice goodness task, both accuracy and reaction time were measured, with results indicating that the effects of CLI on perception are both feature- and language-dependent and that accuracy in L2 English is higher than in L3 German, as predicted.

Onishi (2016) examined the perception of L2 English contrasts and L3 Japanese contrasts by L1 Korean speakers in order to observe the effect of L2 language learning on L3 perception. In particular, Onishi wanted to identify whether trilingual listeners perform better than bilingual (L1 English-L2 Japanese) listeners because of their experience of having learned a second language (i.e., a bilingual advantage), or because the L3 contrasts were similar to L2 contrasts they had already learned. There were two perception tasks in Japanese – an identification task and an AXB discrimination task, in which a range of segmental contrasts were studied. Some contrasts were expected to be particularly difficult for native English speakers, some particularly difficult for native Korean speakers, and some were expected to be difficult for both groups. The results of the identification task and AXB discrimination task in Japanese indicate that the trilinguals' success in L3 perception is feature-dependent. The last experimental task was an identification task in English, and a correlation analysis between trilinguals' perception in the English and Japanese identification tasks shows a positive correlation between the perception of word-initial stop voicing in English and in Japanese, which could provide evidence for the facilitative influence of L2 phonology since this contrast is fairly easy for native English speakers but fairly difficult for native Korean speakers. Similarly, the correlation analysis between the English identification task and the Japanese AXB discrimination task revealed some potential facilitative influence derived from L2 phonology, but also some global facilitative influence derived from having learned a second language (contrasts that are fairly difficult for native English speakers). Therefore, Onishi (2016) concludes that trilinguals "use the most relevant information from both their background languages when perceiving the target L3 contrasts. These results suggest that perception skill is cumulative and is [facilitated] by both the L1 and the L2" (Onishi, 2016, p. 471).

## Suprasegmental acquisition

### Role of language use

Gut (2010) analyzed the acquisition of vowel reduction and speech rhythm in the L3, where both processes were measured using duration, a suprasegmental feature. Production data in the L2 and L3 of four trilingual speakers with different L1s (Russian, Polish, Spanish, and Hungarian) and with German and English as the L2s or L3s were obtained. All speakers had a high level of proficiency in their L2. Fluency for each speaker in each language was measured as a composite of speakers' mean length of utterances, average articulation rate, and ratio of filled pauses to speaking time. Vowel reduction was measured as the difference in duration of full versus reduced vowels. Speech rhythm was measured as the difference between the duration of full-vowel syllables and reduced-vowel syllables. The author found some evidence in favor of facilitative transfer from the L2 to the L3, where trilinguals that had vowel reduction in their L2 but not their L1 produced reduced vowels in their L3 as well. All speakers had a higher speech rhythm ratio in English than in German, regardless of which language was the L2 and which was the L3, perhaps suggesting a lack of influence between the L2 and L3. Furthermore, there was no difference in L2 and L3 performance between speakers whose L1 was syllable-timed and speakers whose L1 was stress-timed, suggesting a lack of influence between the L1 and the L3. The author postulated the results may be greatly due to the fact that the speakers were all highly proficient in their L2 and used it more than their L3. This may reduce the amount of L1 influence seen in the L3 and prohibit L2 influence in the L3, should they not use the L3 all that much.

As was mentioned earlier in Section 1.2, Chan and Chang (2019) analyzed the early stages of L3 tonal perception within the ASP framework. Listeners with a tonal L1 were more accurate with L3 tonal perception than listeners with a tonal L2 but a nontonal L1. The greater facilitative effect of the L1 than the L2 can be explained through either language dominance or order of acquisition. Under the language dominance hypothesis, listeners are more proficient in their L1 and have greater exposure to the L1, and this higher relative dominance contributes to greater influence from the L1 than the L2 during L3 tonal perception. Alternatively, the authors propose that the L1 as the first learned language holds a special status and therefore has greater influence during L3 perception. Data from switched dominance bilinguals – that is, bilinguals who are more dominance in their L2 than their L1 – are needed in order to tease these competing hypotheses apart.

#### Similarity Convergence Hypothesis

Brown and Chang (2023) studied the nature of regressive CLI (rCLI) among trilinguals to evaluate the factors of order of acquisition and typological similarity. In a picture narration task, the speech rhythm of L1 English-L2 German-L3 Spanish and L1 German-L2 English-L3 Spanish trilinguals was examined using the rhythm metric VarcoV, and L1 English-L2 German and L1 German-L2 English bilinguals served as control groups. Whereas English and German are both stress-timed languages, Spanish is a syllable-timed language. Evidence for rCLI was found in English, which is more typologically similar to Spanish, but only when English was the L1. Accordingly, the authors propose the Similarity Convergence Hypothesis (SCH), in which earlier-acquired and typologically more similar languages are more likely to evidence rCLI from later-acquired languages.

### Summary and issues in L3 studies

The brief review of studies of L3 phonetics and phonology above reveals conflicting findings. Some studies point to a bilingual advantage (Amengual, 2021), in which bilingual learners evidence more native-like production and perception in an L3 than monolingual learners evidence in their L2.[1] Other studies suggest there may be preferential transfer from the L1 to the L3 due to increased language experience and proficiency (Chan & Chang, 2019; Patience, 2018). Bidirectional transfer between the L1 and L2 of trilingual speakers has also been shown (Wrembel, 2010), as well as great variability among trilingual participants, pointing to the complexity of individual differences (Kopečková et al., 2016). Whereas some of these distinctive findings may be due to the language pairings in question or the particular profiles of participants, others may be a result of methodological inconsistencies.

A large theoretical focus in the field has been to tease apart contributions of the L1 and the L2 to L3 production and perception. However, it is now more of an established fact in the field that all of a multilingual's languages are active and available, thus shifting the focus away from identifying sources of influence in CLI. As Domene Moreno (2021) states, "[i]nstead of asking whether a phonological structure is transferred from a particular language and why this transfer surfaces, the aim should be to understand the workings of the interactions that we already know to be there" (p. 191). Theoretical models and research designs accordingly need to shift their focus towards the role of external linguistic factors (such as language mode) and internal linguistic factors (such as crosslinguistic phonetic similarity) on CLI, as well as the cognitive and articulatory impetus for influence given the language grouping and phonetic/phonological profile of the feature of interest.

Further, an inconsistent treatment of language mode has led to unclear findings regarding the effects of CLI and whether they are dynamic or static. Though logistically more complex and time-consuming, studies which include data from all three languages in question are also less common. Robust methodologies are required to investigate the relative roles of factors such as language dominance and language mode. Although L3 phonetics and phonology is gaining traction as a field, perceptual studies and studies of suprasegmental contrasts and processes are outnumbered by those examining production of segmental features.

## 1.4 Goals of this dissertation

The present dissertation will address these gaps in the literature on L3 phonetics and phonology by experimentally inducing a unilingual language mode and gathering production and perception data from all three languages used by English-Spanish-Catalan trilinguals. Further, the prosodic feature of lexical stress will be analyzed to observe the segmental and suprasegmental cues that are employed by trilinguals in each language to distinguish stress minimal pairs. The potential impact of language dominance on production and perception will be specifically observed by analyzing data from all three languages related to language history, use, proficiency, and attitudes.

As will be further detailed in Chapter 2, the ultimate goal of this work is to evaluate whether the cue-weighting transfer hypothesis extends to L3 acquisition of lexical stress,

[1] It is worth noting, however, that individual differences such as motivation and attitude are often not taken into account in studies purporting to find evidence for a bilingual advantage, or otherwise are only briefly mentioned (e.g., Salomé et al., 2022).

and if so, how. Given the differences and similarities in segmental and suprasegmental cue-weightings to stress in each of these languages and the fact that the languages belong to separate rhythm classes (Prieto et al., 2012), how L1 English-L2 Spanish-L3 Catalan speakers perceive and produce lexical stress in each language will provide insight to the way in which segmental and suprasegmental cues are transferred crosslinguistically and how current models of L2 acquisition can be extended to L3 acquisition.

## 1.5 Outline of subsequent chapters

The subsequent chapters are as follows: Chapter 2 introduces the suprasegmental concept of stress, differentiating lexical stress from pitch accent as the variable of focus for this study. This chapter also describes a set of acoustic measurements in both acoustic and articulatory terms which have been shown to correlate to lexical stress. The cue-weighting transfer hypothesis is introduced, and the study of trilingual production and perception of lexical stress is motivated within this context. Additionally, this chapter provides an overview of studies of L1 and L2 cue-weightings of acoustic measures for lexical stress in English, Spanish, and Catalan. This chapter concludes with the broad research question that this dissertation is positioned to address.

Chapters 3 and 4 contain the two empirical studies conducted for this dissertation. Chapter 3 analyzes the production of lexical stress in English, Spanish, and Catalan by L1 English-L2 Spanish-L3 Catalan speakers, and Chapter 4 analyzes the perception of lexical stress in each language by the same group of speakers.

Chapter 5 provides a discussion of both studies in the context of theoretical frameworks of L3 phonetics and phonology. Based on the discussions in Chapters 3 and 4, an extension of the cue-weighting transfer hypothesis is proposed to account for evidence from trilingual speakers.

# Chapter 2

# Lexical stress in English, Spanish, and Catalan

This chapter introduces the concept of stress, differentiating lexical stress from phrasal stress. It describes methodological challenges relating to the isolation of these two suprasegmental features and provides a brief overview of the contrastive potential of lexical stress in language. This chapter defines and explores various acoustic cues that have been shown to correlate with stress and introduces the cue-weighting transfer hypothesis. Specific findings relating to lexical stress in English, Spanish, and Catalan will be covered, and the chapter concludes with theoretical implications of the study of lexical stress among trilinguals.

## 2.1 Stress

Stress refers to the relative prominence of a syllable and is not correlated to absolute acoustic or auditory measures. Rather, it is a "structural feature that is derived from relationships with many different content features" such as duration (Beckman & Edwards, 1994, p. 9). It is a suprasegmental feature of speech because "stress can vary independently while the segmental identity of the syllable remains constant" (Cutler, 2005, p. 265). There are two levels of stress: lexical stress (word stress), in which syllables of a word differ in relative prominence; and phrasal stress (pitch accent), in which words in a phrase differ in relative prominence.

### Lexical stress versus phrasal stress

Research with languages that have both lexical stress and pitch accent requires careful methodology so that properties associated with lexical stress and pitch accent, respectively, can be isolated and described. Often this is achieved by implementing a minimum of four experimental conditions in which the target syllable:

(a) receives lexical stress in an accented context

(b) receives lexical stress in an unaccented context

(c) does not receive lexical stress and is in an accented context

(d) does not receive lexical stress and is in an unaccented context.

Using Catalan declarative sentences (accented) and reporting clauses (unaccented) as an example, these four (simplified) experimental conditions are shown in Table 2.1. Through contrastive analysis across the conditions of stress and accent, acoustic cues to stress and accent in isolation and in combination can be observed.

| | [+accent] declarative | [−accent] reporting |
|---|---|---|
| [+stress] | **(a)** Les INCRIMINA CONFIAT. *(S)he incriminates them with conviction.* | **(c)** "Vostès són les CULPABLES" -les incrimina confiat. *"You are guilty" -she incriminates them with conviction.* |
| [−stress] | **(b)** Les va INCRIMINAR CONFIAT. *(S)he incriminated them with conviction.* | **(d)** "Vostès són les CULPABLES" -les va incriminar confiat. *"You are guilty" -she incriminated them with conviction.* |

Table 2.1: Example experimental conditions to isolate the effects of lexical stress (underlined syllable) and pitch accent (capitalized letters) on the syllable *mi* in the Catalan words 'incrimina' and 'incriminar'.

Studies that have employed such careful methodology have revealed that lexical stress is not a weaker version of pitch accent, but there can be different sets of acoustic properties associated with each suprasegmental feature (Sluijter & van Heuven, 1996b). Accordingly, Ortega-Llebaria et al. (2013) refer to lexical stress as a 'context-sensitive' phenomenon, where different acoustic cues are used in the perception of stress in declarative clauses (i.e., with pitch accent) than in reporting clauses (i.e., with low pitch accent). However, a large portion of the literature on acoustic correlates to lexical stress did not control for pitch accent in such a way (Beckman & Edwards, 1994; Sluijter & van Heuven, 1996a), therefore, the results of these studies need to be interpreted carefully. A notable example of this issue in English is the controversial role of pitch (F0) as a correlate of lexical stress. Many early studies appeared to offer evidence that F0 was one of the main, if not the most prominent, acoustic correlates to primary lexical stress in English (e.g., Fry, 1958). However, others noted that the target stimuli in these studies were words produced in isolation, and thus were produced with a pitch accent. Subsequent research revealed that F0 is a prominent correlate of pitch accent but is relatively unrelated to lexical stress (Sluijter & van Heuven, 1996b). Creaky phonation may also be a correlate of low pitch accent as it has been demonstrated to occur in post-focal vowels (Campbell & Beckman, 1997).

### Lexical stress

Not all languages have lexical stress (Cutler, 2005, p. 270), but languages that do can be categorized as free-stress or fixed-stress languages. In fixed-stress languages, primary stress can occur only in specified locations, but free-stress languages have no such limitations. Some free-stress languages (e.g., Spanish or Catalan) only have primary stress, where all other syllables in a word are otherwise equal. Other free-stress languages, however, have secondary stress in addition to primary stress (e.g., English), resulting in acoustic differences among syllables that don't receive primary stress in a word.

Lexical stress also has the potential to be contrastive in free-stress languages, where word pairs can be distinguished according to stress placement. These word pairs are also known as stress minimal pairs. However, this is relatively rare across free-stress languages (Cutler, 2005) as such contrasts necessarily only occur in words with more than one syllable, and a large portion of multisyllabic words are formed from a stem and a morphological affix that rarely carries stress (Bruggeman et al., 2022).

During spoken-word recognition, input is continuously evaluated and multiple lexical candidates may be activated simultaneously. The presence of stress minimal pairs, however, means that stress may be used in perception during word recognition, decreasing the number of candidates that are activated. Cutler et al. (2004) compared the frequency of lexical embeddings (i.e., words within other words) before and after lexical stress was taken into account in Spanish and English and found that whereas lexical stress greatly reduces the number of embeddings in Spanish (from an average of 2.32 to an average of 0.73), the difference is far smaller in English (from an average of 0.94 to an average of 0.59), indicating the greater functional role of lexical stress during spoken-word recognition in Spanish than in English.

Although lexical stress can be contrastive, lexical stress as a suprasegmental feature of language can be processed both prelexically and postlexcially. Ortín and Simonet (2022) state that "some phonological processing strategies can be deployed even in tasks that do not involve lexical access, suggesting that some aspects of processing are prelexical (or non-lexical)" (p. 246).

## 2.2 Acoustic correlates of lexical stress

All languages have suprasegmental distinctions (such as duration) between stressed and unstressed syllables (Cutler, 2015), though the functional weight of these suprasegmental cues may differ across languages. Some languages additionally use segmental cues (such as vowel quality) to distinguish stressed and unstressed syllables.

In this section, acoustic and articulatory descriptions of properties that have been found to signal lexical stress in various languages are covered, as well as how they are commonly measured. Though the use of these acoustic measurements in specific languages may be discussed, this section is intended to provide the technical foundation that Section 2.4 builds on with discussions specific to cue-weightings in English, Spanish, and Catalan.

## Duration

Duration is the length of time that a sound persists and is usually measured in milliseconds. For languages in which duration is a correlate to lexical stress, duration is a suprasegmental property, meaning that it is above the level of a sound unit (i.e., does not involve phonemic length contrasts). Stressed syllables are produced with longer duration than unstressed syllables. Some studies measure the duration of the syllable (e.g., Ortega-Llebaria et al., 2013; Prieto & Ortega-Llebaria, 2006; Sluijter & van Heuven, 1996a), whereas others measure the duration of the vowel in the syllable nucleus (e.g., Campbell & Beckman, 1997). As an indicator of lexical stress in two-syllable words, duration is mostly operationalized as a ratio of the duration of the stressed to unstressed syllable (or vowel) (Ortega-Llebaria et al., 2013), reflecting the relative (rather than absolute) nature of lexical stress.

## Vowel quality

In languages with strong phonological or phonetic vowel reduction, vowel quality is often a correlate of lexical stress. In languages with phonological vowel reduction, certain vowel contrasts are neutralized in unstressed syllables, affecting the measures of the first two vowel formants, F1 (vowel height) and F2 (vowel backness). In languages with phonetic vowel reduction, vowels tend to centralize (towards [ə]) in unstressed syllables, although phonemic contrasts are maintained. This centralization can be a product of faster or more casual speaking styles and smaller articulatory movements relative to the hyperarticulation of more prominent elements (De Jong, 1995; Nadeu, 2016).

Because the resulting frequency difference in Hertz (Hz) between the full and reduced vowel is vowel-specific, a variety of measures have been used to operationalize changes to vowel quality. One such measure is the difference in Hz of F2 and F1 (*F2-F1*), which is sometimes called the compact-diffuse (*C-D*) statistic, and is correlated to tongue height (Zhang et al., 2008). As such, this measure is employed when high or low vowels are present in the testing materials. High vowels that centralize will have a higher *C-D* value in stressed syllables than in unstressed syllables, whereas low vowels that centralize will have a lower *C-D* value in stressed syllables than in unstressed syllables.

The grave-acute statistic (*G-A*), on the other hand, is correlated with tongue advancement (front/back dimension) and is defined as the arithmetic mean of F1 and F2 ($\frac{F1+F2}{2}$) (Zhang et al., 2008). Accordingly, this measure is employed to describe phonetic reduction of anterior and posterior vowels, where anterior vowels have a smaller value of *G-A* in stressed syllables than in unstressed syllables and posterior vowels have a larger value of *G-A* in stressed syllables than in unstressed syllables.

A measure of vowel quality that is less sensitive to vowel identity is the Euclidean distance from the vowel to [ə] (or the midpoint of the vowel space) across the dimensions of F1 and

F2 (Sluijter & van Heuven, 1996a). The difference between this distance in a stressed and unstressed syllable provides a way to measure the degree of vowel reduction, where a larger difference indicates that more reduction occurs in that context.

### Spectral tilt

Spectral tilt, also referred to as spectral balance, is a measure of the amplitude (intensity) of lower frequencies relative to the amplitude of higher frequencies and is conceived of both as a correlate of vocal effort and of phonation type. Sluijter and van Heuven (1996b) advocate for the use of spectral tilt as a correlate to lexical stress, rather than overall intensity, since spectral tilt "is not easily obscured by environmental factors, so that this operationalization of greater vocal effort seems communicatively more robust than overall intensity" (p. 2472). Generally, the intensity of higher frequencies increases more than the intensity of lower frequencies as a syllable goes from unstressed to stressed; thus, resulting in stressed syllables having lower spectral tilt than unstressed syllables. However, the ranges in Hz that are used as 'higher' frequencies and 'lower' frequencies, respectively, are variously defined in the literature. Campbell and Beckman (1997) limit the analysis to the intensities of the second and first harmonic (properties of the sound source; H2 and H1, respectively), whereas Ortega-Llebaria and Prieto (2007) define 'lower' frequencies as containing the fundamental frequency and the 'higher' frequencies as containing all the formants (properties of the vocal tract). Similarly, Sluijter and van Heuven (1996b) separate the fundamental frequency from the formants, but rather than lump all formants together into one frequency band, they separate the formants into distinct bands.

In articulatory terms, spectral tilt is meant to measure the degree of mouth opening during the production of stressed and unstressed syllables, where stressed syllables can often be hyperarticulated with a larger degree of mouth opening (Sluijter & van Heuven, 1996b). Spectral tilt, however, also serves to indicate phonation type, where creaky phonation (in which the closed phase of the glottal period is longer) is produced with higher intensities in higher frequency bands than modal phonation (Keating et al., 2015). Because unstressed syllables can be produced with more creak than stressed syllables (Davidson, 2021), the directionality of the correlation between spectral tilt and lexical stress is more obscured when phonation type is not clearly measured.

## 2.3 Cue-weighting transfer hypothesis

Previous second language research has shown that for a listener to learn a new phonetic contrast they must be able to restructure their existing knowledge. To do so, "listeners must discover which cues are important in which contexts, and then shift their attention to those cues in those contexts" (Francis et al., 2000, p. 1679). Results from perceptual training studies (e.g., Francis & Nusbaum, 2002; Francis et al., 2000), point to a cognitive mechanism which reanalyzes familiar cues based on feedback. Holt and Lotto (2006) investigated factors that contribute to the weighting of acoustic cues and found that informativeness of the cue only partially accounts for its use by listeners. Listeners showed a preference for specific cues, despite equal informativeness, indicating either innate disposition or learned behavior through exposure. Similarly, Iverson et al. (2003) studied the perception of the English /r/ – /l/ contrast by native speakers of Japanese and German and found that cue-weightings in the L2 were dependent on cue-weightings in the L1. These observations led to the formulation of the cue-weighting transfer hypothesis, in which listeners transfer cue-weightings from the L1 to the L2 (Chrabaszcz et al., 2014; Ingvalson et al., 2012; Iverson et al., 2003; Tremblay et al., 2021; Zhang & Francis, 2010). The hypothesis offers a framework from which to view the influence of the L1 on the L2 in the domain of prosody, an area that is underexplored and not yet fully understood.

## L1 cue-weighting for perception of L2 word boundaries

Tremblay and Spinelli (2014) studied L1 English speakers' identification of word boundaries in French in contexts with and without liaison, a process which creates a misalignment of syllable and word boundaries. Liaison has distributional cues, where [z] is the most frequent liaison consonant and is 2.7 times more likely to occur in liaison than [t]. However, there are 25 times more [t]-initial words than [z]-initial (Tremblay & Spinelli, 2014, pp. 311–312). In addition, there are acoustic cues to liaison, for example, liaison consonants are shorter than nonliaison consonants. Lengthening of word-initial consonants is a phenomenon that has been reported in English, leading the authors to hypothesize that L1 English speakers may be able to use duration as a bottom-up acoustic cue to identify word boundaries. In addition, learners may also be able to use top-down distributional cues, such as consonantal identity. Tremblay and Spinelli (2014) used eye tracking to measure participants' real-time processing of word segmentation in the presence of liaison. The results indicate that both native French listeners and L1 English listeners attend to top-down distributional cues. However, L1 English listeners were more attentive to bottom-up acoustic cues than the native French listeners. Because duration is a strong cue to word boundary in English, the authors point to the cue-weighting transfer hypothesis, suggesting that L1 English listeners transfer the heavily weighted cue of duration in English to their perception of liaison in French.

## L1 cue-weightings for perception of different contrasts in the L2

Zhang and Francis (2010) further extended the theory to observe whether L1 speakers of a tonal language were able to transfer L1 cue-weightings for tone to aid in the perception of a different contrast in the L2. Namely, L1 Mandarin speakers were expected to rely on pitch (F0) in the perception of English lexical stress more than native English speakers, because pitch is the most prominent cue to tone in Mandarin. The results indicate that Mandarin speakers attended simultaneously to vowel quality and pitch to perceive L2 English lexical stress, whereas native English speakers attended separately to these cues. The authors ventured that the simultaneous processing of segmental cues (vowel quality) and suprasegmental cues (pitch) in L2 English by the L1 Mandarin speakers was a pattern that was transferred from their native language since tonal contrasts are phonemic and Mandarin speakers must attend to both vowel quality and pitch during word recognition.

## Shifted L2 cue-weightings

D. Kim et al. (2018) analyzed L1 Korean-L2 English cue-weightings for two English tense-lax vowel contrasts during the first year of residence in the United States in an individual differences approach. The contrasts in question are /i/ −/ɪ/ and /ɛ/ − /ae/, where the tense vowels are slightly longer and are spectrally distinct from their lax counterparts. Whereas native English speakers mainly cues spectral cues to distinguish these vowels, each vowel pairing has been shown to be perceptually assimilated to a single vowel in Korean (/i/ and /ɛ/), thus allowing for the learning of cue-weightings by L2 listeners to be observed. Resynthesized natural speech continua were created for each contrast, where duration and vowel quality varied orthogonally. Participants were presented with a 2AFC task with pictures representing words (e.g., *ship* versus *sheep*). The L2 English speakers made more use of durational contrasts than native English speakers. Participants were also grouped (using hierarchical clustering) according to their model estimates for duration and vowel quality and then these groups were analyzed over time. The L2 English listeners were initially most reliant on durational cues but over time increased their reliance on spectral cues. The developmental pathway (starting point and rate of learning) varied across learners, highlighting the effect of individual differences on developmental trajectories.

Bruggeman et al. (2022) studied L1 English-L2 German listeners to observe whether L2 cue-weightings can shift for increased processing efficiency. Participants were presented with word fragments in German, the first syllable of two syllable words that only differed across suprasegmental cues. Therefore, the first syllables of the stress pairs were phonemically identical, so listeners could not rely on vowel quality, which is weighted heavily in their L1 English. After listening to the first syllable only, listeners had to indicate which word the fragment belonged to – option 1 was a stress-initial word and option 2 was a stress-final word. The L1 English listeners performed similarly to the L1 German listeners, indicating that the L1 English listeners attended to suprasegmental cues despite the minimal utility of suprasegmental cues in English. In a second task, L1 Dutch-L2 English speakers living in Australia were tested with a similar decision task in their L2 English and demonstrated the ability to attend to vowel quality in the absence of durational cues. The authors posit that the extensive daily L2 use allowed for their their L2 cue-weightings to shift, thus optimizing processing efficiency in English.

## Effect of language mode

Yazawa et al. (2020) used a computational simulation of L2LP to examine how language mode may affect cue-weighting in the perception of L2 vowels by L1 Japanese-L2 English speakers. The authors used logit model estimates to determine relative weightings of spectral and durational cues across unilingual Japanese and English language mode settings. For the unilingual Japanese context, listeners relied more on durational cues than spectral cues, but for the unilingual English context, listeners relied more on spectral cues than durational cues. However, there was great individual variability in how cue-weightings shifted between language sessions. Given that cue-weighting was found to shift between the language sessions, the authors concluded that language mode is involved in L2 perception.

### Extensions to L3 cue-weighting

Chan and Chang (2019) measured sequential bilinguals' naïve perception of tonal contrasts in an L3 to determine the degree of facilitative transfer from the L1 and L2 when one or neither of the languages was also tonal. L1 Mandarin-L2 English, L1 English-L2 Mandarin, and L1 English-L2 nontonal language speakers participated in an oddity discrimination task and a similarity rating task with both Yoruba and Thai stimuli. Whereas Mandarin, Yoruba, and Thai are all tone languages, English and the other languages known to participants are nontonal. Mandarin's tonal system is more similar to Thai versus Yoruba, so the degree of typological similarity could be studied, even within the tonal languages. The authors found a strong effect of language experience and preferential transfer from the L1, where learners benefited more from having a tonal L1 than a tonal L2 in L3 tone perception. To my knowledge, this is the only study that has attempted to examine cue-weightings across three languages, although the bilinguals here were naïve listeners of the third language rather than trilinguals.

### Summary

The majority of studies of L2 lexical stress do not examine the production and perception of lexical stress in the learners' L1s. This is problematic especially when the role of individual differences is taken into account. L2 speakers do not behave homogeneously in their L2 or their L1, so methodologies which observe production and perception of lexical stress in all of a speaker's languages are required to better test the cue-weighting transfer hypothesis and the relationship between production and perception in each language. Additionally, most of the studies of L2 cue-weighting reviewed above do not consider how language mode may influence cue-weighting transfer. For example, if L2 perception of stress is experimentally obtained in a setting where the language of instruction and interaction is the L1, it is unclear whether results could point to dynamic or static transfer of cue-weightings. Furthermore, it is unclear whether listeners are engaging in lexical retrieval during these highly repetitive perception tasks, or whether it is merely signal detection that is occurring.

## 2.4 Stress in English, Spanish, and Catalan

English, Spanish, and Catalan all have word stress and are classified as free-stress languages, as indicated by the presence of stress minimal pairs (e.g., English: *permit* [ˈpəɹmɪt] noun, [pəɹˈmɪt] verb; Spanish: *paso* [ˈpaso] 'I pass', *pasó* [paˈso] '(s)he passed'; Catalan: *surti* [ˈsurti] 'leave (command)', *sortir* [surˈti] 'to leave'). All of these languages additionally have a trochaic bias, though to differing degrees (Astruc et al., 2010; Ortega-Llebaria et al., 2013). Because of extensive vowel reduction, English and Catalan additionally have near minimal pairs, or words in which stress and vowel quality covary (English: *object* [ˈɑbʤɛkt] noun, [əbˈʤɛkt] verb; Catalan: *examina* [əɣzəˈminə] '(s)he examines', *examinar* [əɣzəmiˈna] 'to examine'). Given the differences in segmental and suprasegmental cue-weightings to stress in each language, as will be detailed below, how L1 English-L2 Spanish-L3 Catalan speakers are able to perceive and produce lexical stress in each language will provide insight to the way segmental and suprasegmental cues are transferred crosslinguistically.

### Stress and accent in L1 English

In English, words in post-focal clauses are produced with low pitch accent (as shown in Figures 2.1 and 2.2, produced by an L1 and an L3 speaker, respectively). Although English is a free-stress language with contrastive stress, examples of stress minimal pairs are extremely rare (Cutler, 2015, p. 107). Rhythmically, syllables in English tend to alternate across stress, resulting in differences in prominence between syllables without primary stress (Cutler, 2015). Despite being a free-stress language, English does a have a significant tendency towards word-initial stress (trochaic bias), which can bias native English speakers' responses in lexical decision and lexical retrieval tasks (Vitevitch et al., 1997).

Vowel quality is the most prominent cue of primary lexical stress in English (Campbell & Beckman, 1997; Chrabaszcz et al., 2014; Zhang & Francis, 2010), where all stressed syllables must have full vowels and reduced vowels cannot be stressed. There are very few minimal pairs that differ according to lexical stress but not vowel quality, reducing the functional weight of suprasegmental cues such as duration and spectral tilt during lexical recognition. Duration and spectral tilt are also implicated in primary lexical stress production and perception, however, though these cues may be dependent upon cooccurring vowel reduction (Beckman & Edwards, 1994; Campbell & Beckman, 1997). In perception, vowel quality is still the most heavily weighted cue (Zhang & Francis, 2010), and native English speakers have been shown to not use prosodic information in lexical retrieval in the absence of vowel quality cues (Cutler, 1986).

Although pitch and intensity have historically been seen as correlates of stress, both have been shown to be correlates of phrasal prominence, rather than word prominence, when intonational contours are accounted for (Beckman & Edwards, 1994, p. 13). Therefore, this dissertation observes lexical stress production and perception in the context of a low pitch accent, isolating the correlates of word prominence from confounding correlates of phrasal prominence. Isolating lexical stress from pitch accent is not necessarily a straightforward process, however. For example, Campbell and Beckman (1997, p. 68) found a correlation between low pitch accent and creaky phonation in the following intonational pattern:

(2.1) No, it's not JONATHAN Baddle I interviewed, but his brother, Matthew.
H* L-

Since spectral tilt can be a correlate of lexical stress (stressed syllables have low H1-H2 and increased amplitudes of higher frequencies) as well as a way of measuring the degree of creaky versus modal phonation (where creaky phonation has a lower H1-H2 and modal phonation has a higher H1-H2; Davidson, 2021, p. 2), speakers with consistent creaky phonation in low pitch accents need to be analyzed differently. Creaky phonation is also a common feature at the end of an utterance (Davidson, 2021, p. 10), so using sentences with post-focal target words that are phrase-medial – such as (2.1) – can isolate the overall amount of creak to cases of low pitch accent correlates and speaker-specific tendencies.

## Stress and accent in L2 English by L1 Spanish speakers

Rallo Fabra (2015) studied L1 Spanish-L2 English bilinguals' production of English stressed and unstressed vowels across their age of acquisition and length of residence in the United States. Whereas early bilinguals produced more centralized vowels in unstressed syllables, late bilinguals maintained a more peripheral vowel space. However, the author noted that

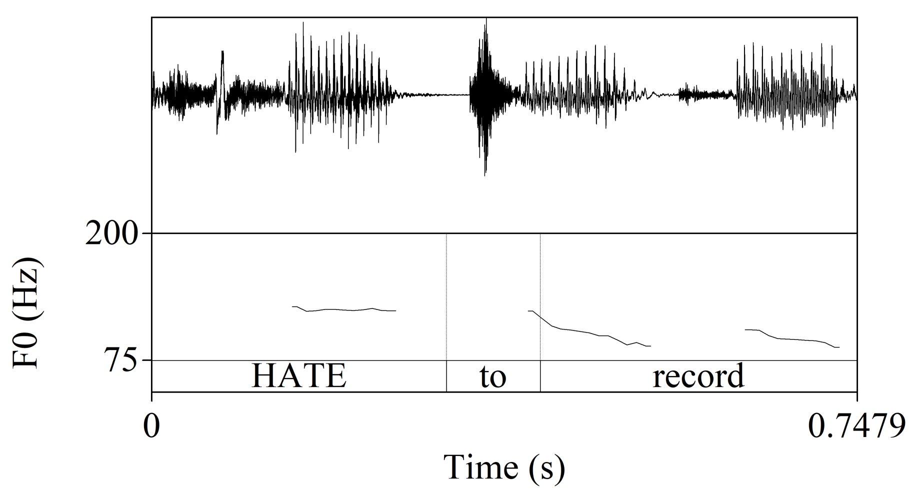



Figure 2.1: Oscillogram with F0 track and segmentation tier of the English post-focal clause *to record*, produced by an L1 speaker. Note the flat and low F0 trajectory on the stressed syllable *cord*.

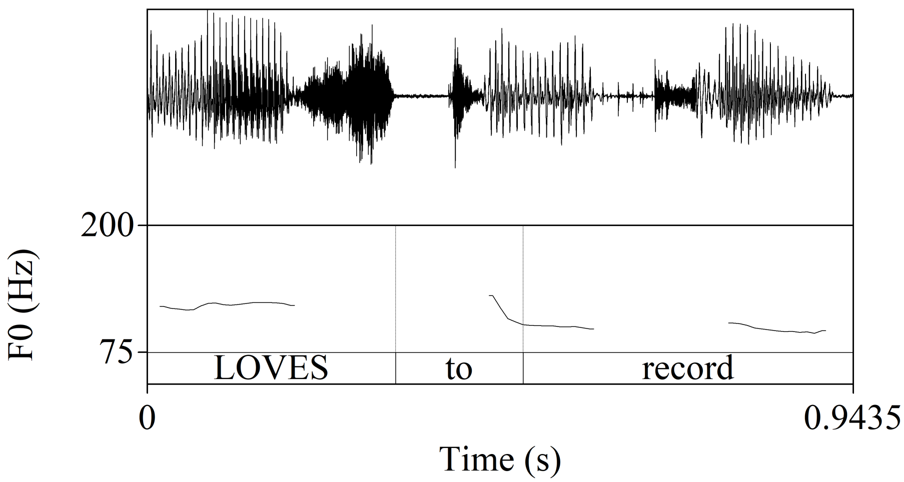



Figure 2.2: Oscillogram with F0 track and segmentation tier of the English post-focal clause *to record*, produced by an L3 speaker. Note the flat and low F0 trajectory on the stressed syllable *cord*.

most of the target words in English had Spanish cognates, possibly inducing more of a bilingual language mode and increasing CLI from Spanish in their English production.

Byers and Yavas (2017) compared productions of unstressed English vowels by English monolinguals, early Spanish-English bilinguals and late Spanish-English bilinguals. Early bilinguals patterned closely to English monolinguals in terms of relative duration of stressed and unstressed vowels, but the late bilinguals did not have as great of a duration ratio. Even though duration is a heavily-weighted cue to lexical stress in Spanish, the authors posit that the phonemic role of duration to distinguish tense and lax vowels in English may contribute to the larger duration ratio employed in English. The Language Experience and Proficiency Questionnaire (LEAP-Q; Marian et al., 2007) was used to evaluate the bilinguals' language dominance, and an analysis of dominance predictors indicated that higher personal attachment to English was significantly correlated to shorter durations for unstressed vowels.

## Stress and accent in L1 Spanish

In Spanish, every content word is expected to bear a pitch accent (Hualde & Prieto, 2015), and Spanish places pitch accents on both old and new information, unlike English (Cruttenden, 1993). However, words in reporting clauses are produced with low pitch accent (as shown in Figures 2.3 and 2.4, produced by an L1 and L2 speaker, respectively). Therefore, lexical stress in Spanish is examined within reporting clauses in order to isolate cues to lexical stress from cues to pitch accent.

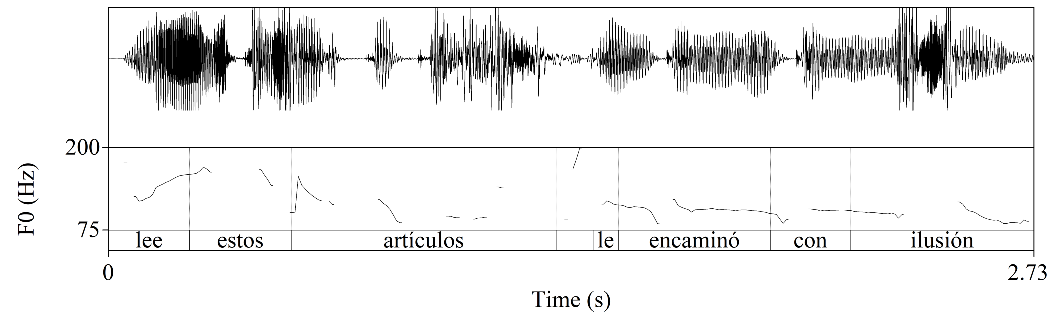



Figure 2.3: Oscillogram with F0 track and segmentation tier of the Spanish reporting clause *le encaminó con ilusión* '(s)he guided him/her eagerly', produced by an L1 speaker. Note the flat and low F0 trajectory on the stressed syllable *nó*.

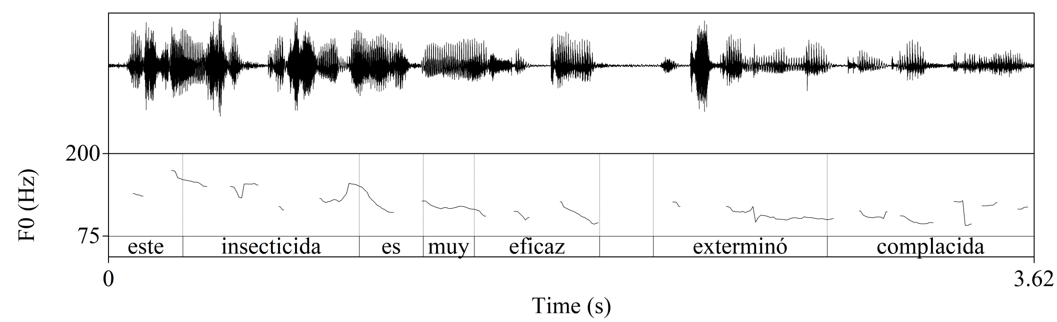



Figure 2.4: Oscillogram with F0 track and segmentation tier of the Spanish reporting clause *exterminó complacida* 'she exterminated happily', produced by an L2 speaker. Note the flat and low F0 trajectory on the stressed syllable *nó*.

In unaccented contexts, duration is the most prominent cue to lexical stress (Ortega-Llebaria & Prieto, 2011; Romanelli et al., 2018), where the relative weight of duration is heavier for word-final vowels (Prieto & Ortega-Llebaria, 2006). Perception studies indicate that L1 Spanish speakers may be able to utilize even relatively small duration ratios of stressed and unstressed syllables to identify lexical stress (Ortega-Llebaria et al., 2007). Although there is no prescriptive phonetic or phonological vowel reduction in Spanish, previous studies have shown that mid vowels may reduce slightly in unstressed contexts (e.g., Quilis & Esgueva, 1983; Romanelli et al., 2018), making vowel quality a weak correlate to stress production mainly for /o e/, though the perceptual salience of this cue may be quite weak (Ortega-Llebaria & Prieto, 2007). Other Spanish vowels such as /i/ may be susceptible to reduction in unstressed syllables (Ronquest, 2013), though the effect has been documented to be more minimal. Spectral tilt has been shown to correlate to stress in production, but it was not utilized by speakers during a perception task (Ortega-Llebaria & Prieto, 2007). Overall intensity may also correlate with lexical stress, where stressed syllables have been shown to be about 2 dB louder than unstressed syllables (Ortega-Llebaria et al., 2013). On the whole, suprasegmental cues, duration in particular, are most heavily weighted in Spanish.

## Stress and accent in L2 Spanish by L1 English speakers

To date, research on L2 Spanish perception of stress by L1 English speakers shows that these learners do not consistently perceive stress in the absence of pitch accent. For example, in a word-stress detection task, Ortega-Llebaria et al. (2013) found that L1 English speakers have a 'context-specific stress deafness,' where they use duration and pitch as cues to stress only in the presence of pitch accents – a position in which these same acoustic measures are used in English. However, the use of these suprasegmental cues demonstrates that the heavy cue-weighting of vowel quality in L1 English does not preclude the use of other cues, rather, processing efficiency can be maximized if other cues are ignored due to the extreme productivity of vowel quality in English. In reporting clauses, English speakers did not use duration as a cue to lexical stress, which Ortega-Llebaria et al. (2013) posit is because the duration ratios of stressed and unstressed syllables in Spanish are smaller than in English, especially in the reporting clause context.

In J. Y. Kim (2016), L1 English speakers did not use pitch, duration, or intensity consistently to cue lexical stress productions in L2 Spanish. Although these speakers did not significantly reduce unstressed vowels, unstressed vowels in paroxytones were lengthened, which may be due to either the increased lexical frequency for these speakers of past tense verbs (oxytones) or final vowel lengthening (J. Y. Kim, 2015). However, Romanelli et al. (2018) showed that L1 English speakers used duration as well as vowel quality to signal stress contrasts. These conflicting results underline the importance of taking into account individual differences and the myriad of factors which may influence production and perception, rather than rely on a generalized speaker category like 'L1 English-L2 Spanish'.

## Stress and accent in L1 Catalan

In Catalan[1], like in Spanish, words in reporting clauses are produced with a low pitch accent (as shown in Figures 2.5 and 2.6, produced by an L1 and an L3 speaker, respectively). Therefore, lexical stress in Catalan is examined within reporting clauses in order to isolate cues to lexical stress from cues to pitch accent. Part of the vowel inventory undergoes phonological reduction in unstressed syllables, where /a e ɛ/ reduce to /ə/, /o ɔ/ reduce to /u/, and /i u/ retain vowel quality (Figure 2.7). Accordingly, cue-weightings for Catalan lexical stress are vowel-specific, where duration and spectral tilt are the most prominent

[1]There are several varieties of Catalan, each with different vowel systems and patterns of vowel reduction. From here on, 'Catalan' specifically refers to the Central Catalan variety, spoken across the provinces of Barcelona, Tarragona, and Girona in the autonomous community of Catalonia.

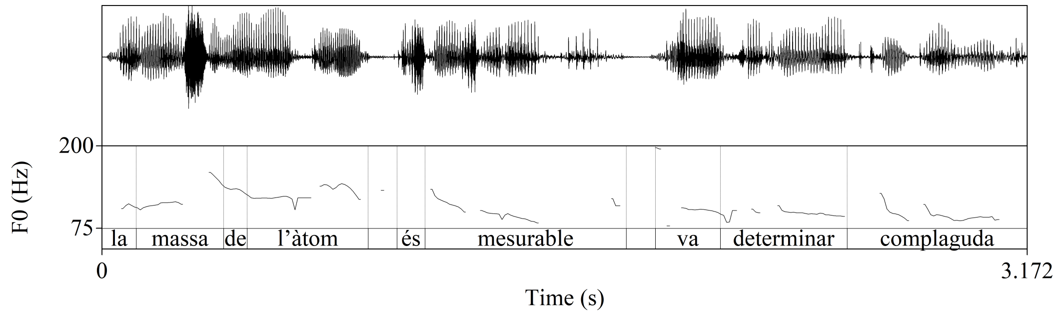



Figure 2.5: Oscillogram with F0 track and segmentation tier of the Catalan reporting clause *va determinar complaguda* 'she established happily', produced by an L1 speaker. Note the flat and low F0 trajectory on the stressed syllable *nar* in *determinar*.

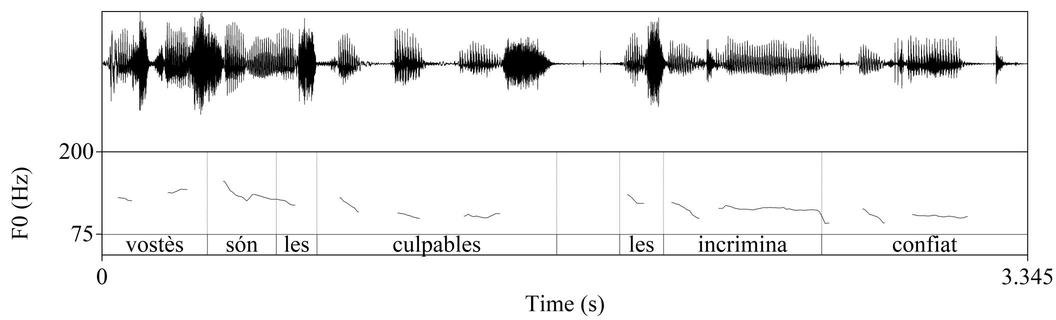



Figure 2.6: Oscillogram with F0 track and segmentation tier of the Catalan reporting clause *les incrimina confiat* 'he incriminates them confidently', produced by an L3 speaker. Note the flat and low F0 trajectory on the stressed syllable *mi* in *incrimina*.

cues for vowels such as /i/, but vowel quality is prominent for vowels such as /a/, along with duration and spectral tilt (Ortega-Llebaria & Prieto, 2009). Whereas Ortega-Llebaria and Prieto (2011) did not observe vowel reduction in /i/, Nadeu (2014) found that unstressed /i/ was produced with a lower F2 than stressed /i/, indicating that some phonetic reduction occurred. Nadeu (2014) ventured that the conflicting results were due to the fact that /i/ in the stimuli used by Ortega-Llebaria and Prieto (2011) was between two nasals, which may have led to formant detection errors and therefore a falsely nonsignificant relationship between vowel quality and stress.

Similarly to Spanish, duration is more strongly correlated to stress for word-final vowels in production (Prieto & Ortega-Llebaria, 2006). Ortega-Llebaria et al. (2010) argue that spectral tilt is actually a consequence of formant movement (vowel reduction) rather than a

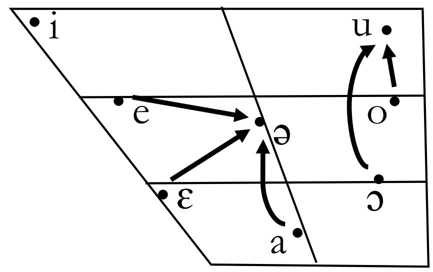



Figure 2.7: Catalan vowel alternation patterns in unstressed syllables.

direct acoustic correlate of stress, an idea that has been previously suggested in the literature (e.g., Sluijter & van Heuven, 1996b).

Despite the differences in vowel reduction patterns between Spanish and Catalan, studies that examine differences in cue-weightings between the two languages (e.g., Ortega-Llebaria & Prieto, 2011; Prieto & Ortega-Llebaria, 2006) exclusively have used results from dominant Spanish speakers and dominant Catalan speakers for the respective languages. To the best of my knowledge, there have been no studies with Catalan that analyze language dominance to take into account the role of L1 influence on L2 production and perception for this language pairing, which would provide insight into how the cue-weighting transfer hypothesis applies to closely related languages that may differ in the informativeness of one acoustic cue (i.e., vowel quality). Similarly, there has been no work analyzing Catalan and English production and perception of lexical stress by bilingual or trilingual speakers. Because vowel quality is heavily weighted in English by native speakers but is only informative in Catalan for a subset of vowels, this language pairing could provide interesting evidence for vowel-dependent aspects of the cue-weighting transfer hypothesis.

## 2.5 Broad research question

This dissertation analyzes the production and perception of lexical stress in trilinguals' first, second, and third languages (L1, L2, and L3) to evaluate how the cue-weighting transfer hypothesis applies to L3 acquisition. Acoustic correlates of lexical stress were analyzed in English, Spanish, and Catalan, which all have lexical stress as indicated by the existence of stress minimal pairs and which are reported to belong to different rhythm classes (Prieto et al., 2012). Little, if any, research has been done to investigate how the cue-weighting transfer hypothesis may extend to L3, which in turn allows for L2 transfer to be better understood (Flynn et al., 2004). Although many languages use suprasegmental features to cue lexical stress and other contrasts, suprasegmental features are not often taken into account in models of phonetic and phonological representation of bilinguals, let alone speakers of more than two languages.

Data were obtained from L1 English-L2 Spanish-L3 Catalan speakers, and early Spanish-Catalan bilinguals that learned English as an L3 and English monolinguals served as baseline groups. Productions of stress minimal pairs in the three languages were collected in three separate experimental sessions via a sentence elicitation task, and perceptions of stress were gathered in the same three separate experimental sessions via a word identification task using nonword stress minimal pairs. Whereas vowel quality is the main correlate to lexical stress in English, duration is the most prominent cue in Spanish, and Catalan cue-weightings are vowel-specific, where both duration and vowel quality can be heavily weighted. Given the body of literature supporting the cue-weighting transfer hypothesis, an analysis of L3 production and perception of lexical stress is positioned to address the following broad research question:

RQ: Does the cue-weighting transfer hypothesis extend to L3 acquisition? If so, how?

Since the majority of models of multilingual phonetic and phonological acquisition that posit some degree of interaction between acquired language systems have been theorized to extend to L3 acquisition in a similar manner (e.g., Amengual, 2021; Chan & Chang, 2019; de Leeuw & Chang, in press; Escudero et al., 2013; Wrembel et al., 2019), I hypothesize that cue-weightings from previously learned language(s) will transfer into the L3 in similar ways that cue-weightings from the L1 transfer into the L2. Specifically, I hypothesize that cue-weightings in each language will be mediated by relative language dominance. I additionally predict that transfer will be bidirectional, where cue-weightings in the L2 and the L3 can influence cue-weightings in the L1.

# Chapter 3

# Study 1: Production of lexical stress

## 3.1 Introduction

In this experiment, trilinguals' lexical stress productions in English, Spanish, and Catalan were obtained in three separate sessions. Each language session was conducted completely in that language by a native speaker to induce as much of a unilingual language mode as was possible. All stress minimal pairs were produced in low pitch accent conditions in order to isolate the acoustic correlates of word-level stress and to address the following research questions:

*RQ1*: How does speakers' relative experience with the three languages affect their cue-weighting in each language?

I hypothesize that with increasing relative dominance in L2 Spanish and L3 Catalan, L1 English speakers will shift cue-weightings towards more nativelike strategies in their L2 and L3 (i.e., to pattern more similarly with early Spanish-Catalan bilinguals), as shifted cue-weightings due to language experience have been evidenced in the literature (e.g., Bruggeman et al., 2022; D. Kim et al., 2018).

*RQ2*: Do L1 English trilinguals use suprasegmental cues to a greater extent than English monolinguals when producing stress contrasts in English?

Despite the demonstrated bidirectional nature of crosslinguistic influence (e.g., Birdsong, 2018; Cook, 2003; Grosjean, 1989), I hypothesize that trilingual speakers will pattern similarly to monolingual English speakers in their predominate use of vowel quality to cue lexical stress shifts (in alignment with J. Y. Kim, 2016). Although trilinguals may use other acoustic properties to cue stress shifts in their L2 and L3, suprasegmental cues have a low functional load in English due to the scarcity of true stress minimal pairs (Cutler, 2015). Since there would not be a functional advantage to employing other acoustic cues to signal stress shifts in English, I do not hypothesize that speakers will do so.

The sections that follow describe the production study designed to provide insight to these research questions, the speakers that participated in this study, and the results.

## 3.2 Methods

### Participants

Twenty-five L1 English-L2 Spanish-L3 Catalan sequential learners (14 female, 10 male, 1 other[1]; average age: 28.1 ± 6.2 years) were recruited to participate in this study. Twenty-five early Spanish-Catalan bilinguals that learned English as an L3 (18 female, 7 male; average age: 25.5 ± 7.3 years) and 20 monolingual English speakers (13 female, 7 male; average age: 40.0 ± 17.7 years) were also recruited to serve as comparison groups. Here, the term 'monolingual' is used to refer to people that are not actively using or learning an L2 and use only their L1 in their everyday life ('funtional monolinguals'; Best & Tyler, 2007). Whereas the baseline trilinguals speak Castilian Spanish, the majority of target trilinguals use either Castilian Spanish or US Spanish. The target trilinguals have been exposed to a range of varieties of Spanish across their years of classroom instruction, and 18 of the 25 target trilinguals have spent at least a year in Spain where they had significant exposure to Castilian Spanish. The target and baseline trilingual speakers were recruited among students from 11 universities in the United States with Catalan language classes,[2] as well as from followers of Catalan language social media accounts. Monolingual English speakers were recruited widely from contacts in the United States and all the participants received compensation for their involvement in the research project.

The English monolinguals completed a short sociodemographic questionnaire to disclose their age, gender, and current place of residence (state or province). The trilinguals completed the Bilingual Language Profile (BLP; Birdsong et al., 2012), adapted to elicit information about all three of their languages, rather than just two. While relatively untested on trilingual speakers, the holistic dominance score obtained with the BLP has been shown to be strongly reliable as a measure of bilingual language dominance (Olson, 2023). Individual components, such as individuals' attitudes towards each language, have also successfully been used as predictors of bilingual perception and production (e.g., Law et al., 2019), and therefore a similar approach is used to examine trilingual dominance in this dissertation. Although homogeneity is not assumed among speakers of the same language profile and an individual-level analysis will be used to look at relative language dominance, Table 3.1 contains the averaged survey responses for L1 English trilinguals and L3 English trilinguals.

Most of the questions in the BLP provide ordinal data; responses fall into discrete bins along a latent continuous variable. For example, the 7-point Likert-style question 'How well do you read in English?' results in discrete bins from 0 to 6, where '0' corresponds to 'not very well' and '6' corresponds to 'very well'. Because participants could give binned responses from an underlying continuum ('reading proficiency'), the data are ordinal. In general, polychoric principal component analysis (pPCA) should be used for dimensional

[1]Participants self-identified their gender from one of these three options.

[2]These institutions are Columbia, Georgetown, Harvard, Indiana University Bloomington, Stanford, University of California Berkeley, University of California Los Angeles, University of Chicago, University of Colorado Boulder, University of Massachusetts Amherst, and the University of Michigan.

| | L1 English | | | L3 English | | |
|---|---|---|---|---|---|---|
| | English | Spanish | Catalan | English | Spanish | Catalan |
| Age of acquisition | Birth | 11.2 (3.7) [**1**] | 18.7 (0.5) [**19**] | 5.8 (2.7) | 0.8 (1.8) | 0.4 (1.0) |
| Age comfortable with language | 0.2 (1.0) | 14.9 (5.1) [**1**] | 19.0 (0.0) [**24**] | 13.8 (3.3) [**4**] | 2.7 (5.3) | 1.6 (4.4) |
| Years of classes | 15.9 (1.6) [**13**] | 7.8 (5.4) [**1**] | 1.4 (1.5) | 10.0 (4.8) [**2**] | 12.1 (5.6) [**5**] | 12.1 (5.9) [**5**] |
| Years in region with language | 19.0 (0.0) [**23**] | 2.9 (2.2) | 1.9 (1.8) | 2.4 (3.6) [**1**] | 15.3 (5.0) [**21**] | 16 (4.6) [**20**] |
| Years in family with language | [**25**] | 1.3 (1.8) [**2**] | 0.5 (0.9) | 0.8 (2.6) | 3.6 (6.8) [**17**] | 0.0 (0.0) [**21**] |
| Years in workplace with language | 7.8 (4.6) [**8**] | 3.4 (3.3) | 1.3 (1.5) | 2.7 (2.8) | 4.5 (4.6) [**5**] | 1.3 (1.5) [**5**] |
| % Used with friends | 57.6 (27.7) | 23.2 (18.1) | 14.0 (15.5) | 31.6 (28.1) | 31.6 (24.4) | 37.6 (32.7) |
| % Used with family | 85.2 (26.8) | 8.0 (16.3) | 3.2 (8.5) | 6.0 (17.1) | 40.0 (44.1) | 58.4 (44.3) |
| % Used at work | 54.0 (31.2) | 22.0 (25.3) | 11.6 (13.7) | 50.8 (38.2) | 24.0 (27.1) | 26.4 (32.4) |
| % Self-talk | 63.2 (21.5) | 19.2 (11.5) | 11.2 (10.9) | 24.8 (23.7) | 28.4 (29.7) | 47.2 (34.9) |
| % Used in counting | 78.4 (18.2) | 12.8 (12.1) | 6.4 (9.9) | 10.0 (15.3) | 29.2 (33.7) | 60.0 (35.2) |
| Speaking (0−6) | 6.0 (0.2) | 5.0 (1.0) | 3.6 (1.4) | 4.9 (1.0) | 5.8 (0.4) | 5.8 (0.5) |
| Listening (0−6) | 6.0 (0.0) | 5.3 (0.9) | 4.3 (1.3) | 5.3 (1.0) | 6.0 (0.0) | 6.0 (0.2) |
| Reading (0−6) | 6.0 (0.0) | 5.3 (0.9) | 4.6 (1.2) | 5.2 (0.9) | 6.0 (0.0) | 5.9 (0.3) |
| Writing (0−6) | 6.0 (0.) | 4.6 (1.3) | 3.2 (1.5) | 4.8 (1.0) | 5.8 (0.6) | 5.5 (0.9) |
| Feel like myself (0−6) | 5.8 (0.7) | 4.7 (1.7) | 4.2 (1.9) | 4.3 (1.5) | 4.4 (2.0) | 5.6 (0.8) |
| Affiliation with language's culture (0−6) | 5.6 (0.9) | 3.5 (1.7) | 3.6 (1.8) | 3.2 (1.3) | 3.8 (2.0) | 5.6 (0.8) |
| Want to sound native (0−6) | 5.8 (0.6) | 5.6 (1.0) | 5.1 (1.5) | 5.3 (1.3) | 5.3 (1.7) | 5.9 (0.4) |
| Want others to think I'm native (0−6) | 5.0 (1.3) | 4.7 (1.6) | 4.2 (1.7) | 4.2 (2.4) | 5.1 (1.8) | 6.0 (0.0) |

Table 3.1: Summary statistics for L1 English trilinguals' and L3 English trilinguals' BLP responses. Because participants could select '20+' as an option for some questions rather than enter a numeric value, some averages reported above are derived from the portion of participants that provided a numeric answer rather than '20+', therefore the true average for these questions is higher than reported. When this is the case, the number of speakers that answered with '20+' is bolded in brackets next to the average.

reduction when there are ordinal features. However, the majority of questions in the survey have a large number of response options (more than 8). When this is the case, these features may be considered to be continuous rather than ordinal, and linear principal component analysis (PCA) can be used (Revelle, 2023). Some questions in the survey elicited a mix of data types. For example, the question 'At what age did you feel comfortable using Spanish?' would normally elicit binned continuous data, but participants could select the responses '20+ years' and 'still not comfortable' in addition to responses of discrete years of age between 0 and 19. As such, the response label '20+' was coded numerically as '20' and the response label 'still not comfortable' was coded numerically as '21' so that all responses were represented numerically.

Data were submitted to PCA using `scikit-learn` (Pedregosa et al., 2011), initially with the number of components that would capture 90% of the variance in the data (15 components). Each computed principal component and its corresponding percent explained variance was plotted (called a Scree plot). The optimal number of components for principal component analysis can be visually determined from a Scree plot using the 'elbow method', in which the value of the point on the x-axis following the steepest decrease in explained variance (at the 'elbow' of the plot) is selected as the maximum number of components that can be used to reduce the dimensionality of the data while still explaining a large portion of the variance. Based on the Scree plot shown in Figure A.1, two components best capture the variance in the BLP responses and simultaneously reduce the dimensionality of the data. Data were again submitted to PCA, this time with the number of components set to two. Table 3.2 shows the BLP questions which correlate strongly to each component (strength of correlation $>| 0.4 |$). Based on these correlations, a large portion of the variance in the BLP data can be represented by the strength of participants' English dominance relative to their Spanish and Catalan dominance (Principal Component 1; henceforth, 'PC1') and the strength of their Catalan dominance relative to their Spanish dominance (Principal Component 2; henceforth, 'PC2'). This result makes sense given that half of the participants are late sequential L1 English-L2 Spanish-L3 Catalan speakers and the other half are early sequential or simultaneous Spanish-Catalan bilinguals that learned English as a third language. As such, the degree of English dominance relative to Spanish and Catalan dominance accounts for the largest variance in the BLP responses (40.36%). Figure 3.1 shows that these two principal components give good separation across the language profile labels of 'L1 English' and 'L3 English', where PC1 mainly accounts for inter-group differences and PC2 mainly accounts for intra-group differences.

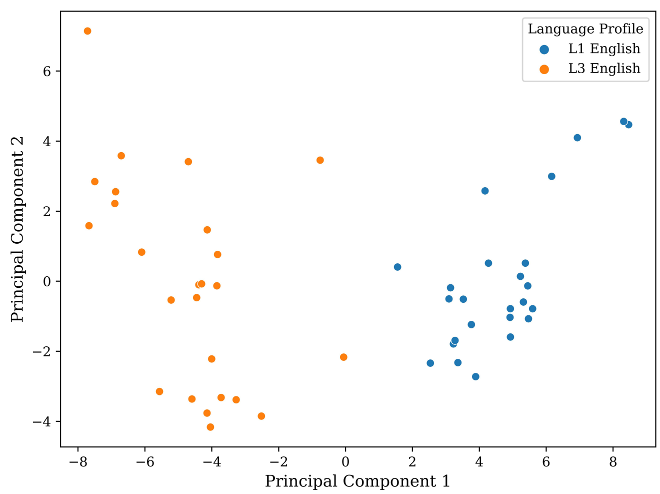



Figure 3.1: Distribution of the 50 trilingual speakers according to their predetermined bins ('L1 English' and 'L3 English') and their loadings for the two principal components.

| Comp. | Feature | Spanish | Catalan | English | Exp. Variance |
|---|---|---|---|---|---|
| 1 | Age learned | 0.83 | 0.94 | -0.82 | 40.36% |
| | Age of comfort | 0.78 | 0.95 | -0.91 | |
| | Years of study | -0.46 | -0.83 | 0.64 | |
| | Years in region | -0.95 | -0.96 | 0.90 | |
| | Years in family | -0.62 | -0.81 | 0.95 | |
| | Years in workplace | -0.48 | -0.66 | 0.63 | |
| | Frequency use with friends | – | -0.61 | 0.60 | |
| | Frequency use with family | – | -0.64 | 0.88 | |
| | Frequency use at work | – | -0.46 | – | |
| | Frequency of self-talk | – | -0.70 | 0.76 | |
| | Frequency in counting | – | -0.77 | 0.95 | |
| | Speaking level | -0.63 | -0.85 | 0.71 | |
| | Listening level | -0.57 | -0.79 | 0.53 | |
| | Reading level | -0.60 | -0.75 | 0.67 | |
| | Writing level | -0.66 | -0.81 | 0.75 | |
| | Feels like oneself | – | -0.60 | 0.63 | |
| | Identifies with culture | – | -0.67 | 0.77 | |
| | Wants to sound native | – | -0.48 | – | |
| | Wants to be perceived as native | – | -0.60 | – | |
| 2 | Frequency use with friends | -0.40 | 0.55 | – | 10.32% |
| | Frequency use with family | -0.62 | 0.46 | – | |
| | Frequency use at work | – | 0.57 | – | |
| | Frequency in self-talk | -0.59 | 0.42 | – | |
| | Frequency in counting | -0.42 | – | – | |
| | Speaking level | -0.47 | – | – | |
| | Reading level | -0.43 | – | – | |
| | Writing level | -0.45 | – | – | |
| | Feels like oneself | -0.84 | – | – | |
| | Identifies with culture | -0.82 | – | – | |
| | Wants to sound native | -0.67 | – | – | |
| | Wants to be perceived as native | -0.61 | – | – | |

Table 3.2: Strength of correlation of features in BLP data to principal components with the percent variance explained by each component; Bartlett's sphericity test $p$-value $< 0.05$.

## Materials

To analyze (supra)segmental cue-weightings in English in the presence and absence of vowel quality information, target stimuli include 10 stress minimal pairs in which vowel quality prescriptively changes and 9 stress minimal pairs in which vowel quality prescriptively is maintained (Table A.1). Similar information can be gathered in Spanish and Catalan by using paroxytone and oxytone verbs ending in -*minar* (Tables A.3 and A.5), as was done by Ortega-Llebaria and Prieto (2011). In Spanish, verbs end in -*mino* (1SG present; [ˈmino]) and -*minó* (3SG preterit; [miˈno]). In Catalan, verbs end in -*mina* (3SG present; [ˈminə]) and -*minar* (infinitive in periphrastic preterit; [miˈna]), where the word-final /r/ is not pronounced in Central Catalan. For both Spanish and Catalan, vowel quality is not expected to change greatly for /i/ across lexical stress, but vowel reduction is expected to affect /o/ in Spanish and /a/ in Catalan, albeit to different degrees (Ortega-Llebaria & Prieto, 2009; Quilis & Esgueva, 1983). To isolate the acoustic correlates of lexical stress from those of pitch accent, target stimuli occur in post-focal context in English and reporting clauses in Spanish and Catalan, phrasal contexts with low pitch accent (Beckman & Edwards, 1994; Ortega-Llebaria & Prieto, 2009, 2011). The Spanish and Catalan stimuli and sentences were adapted from Ortega-Llebaria and Prieto (2011), with permission from the authors. To aid in segmentation of the vowels at the end of the target words in the Spanish and Catalan stimuli, the following segments were always voiceless obstruents /p t k s/, segments that have clear acoustic boundaries in Spanish and Catalan. See Tables A.2, A.4 and A.6 for the sentences used in each language to elicit the target words.

## Procedure

Separate sessions were scheduled for individual participants in Catalan, English, and Spanish (in that order) over Zoom, with at least 48 hours between each scheduled session. Each session was conducted completely in the target language by a research assistant who was a native speaker. The English sessions were conducted by myself and a trained assistant. We both began learning an L2 (Spanish) around the age of 12 and were raised in predominately monolingual regions of the United States. The Spanish sessions were conducted by five different research assistants, two of whom are heritage speakers of Spanish raised in California, and three of whom are early Spanish-Catalan bilinguals raised in Catalonia or the Balearic Islands and who learned English as an L3. The Catalan sessions were conducted by the same three early Spanish-Catalan bilingual research assistants. Prior to the call, participants were instructed to download a professional audio recording app called ShurePlus MOTIV (Shure, Inc., 2022) to their mobile phones. For each experimental session, participants completed the production task (described in this chapter), followed by a perception task (described in Chapter 4). Research assistants shared their screen with participants during the Zoom call and used PsychoPy (Peirce et al., 2019) to present the target sentences in randomized order during the production task. Participants were given written and oral instructions in the language of the session to use a natural and consistent reading voice throughout the recording session. During the call, participants were instructed how to use the recording app to record themselves reading the sentences they saw on their screens and subsequently, how to upload their recordings to a secured Box folder. At the end of the production task, participants were asked to describe what they thought was the focus of study.

## Analysis

The recorded sentences were aligned to their orthographic text using EasyAlign (Goldman, 2011), and Montreal Forced Aligner (McAuliffe et al., 2017) was used to perform phone-by-phone alignment.[3] Phone boundaries in the target words were subsequently hand-corrected by myself and a team of trained undergraduate students. The penultimate and ultimate vowels were annotated according to vowel identity and primary lexical stress assignment. For the English data, the syllabification of the target words was also performed and annotated. With 84 words in the Spanish and Catalan tasks (3 repetitions of 28 words, completed by 50 participants) and 114 words in the English tasks (3 repetitions of 38 words, completed by 70 participants), a total of 16,380 words were hand-corrected and checked for correct stress assignment. Due to improper stress production, word pronunciation, or background noise, 2,392 words were removed from analysis yielding a total of 13,988 words (3,648 Catalan words, 6,879 English words, and 3,461 Spanish words) submitted to further statistical analysis.

VoiceSauce (Shue, 2010) was used to extract a range of acoustic properties at 5 ms intervals from vowels in the ultimate and penultimate syllables of target words. These measures included four different measures of F0 (extracted with Straight, Snack, Praat, and the subharmonic-to-harmonic ratio method from Sun, 2002), the first four formants (estimated with Snack), harmonic-amplitude measures corrected for formant frequency (estimated with Snack; H1-H2, H2-H4, H1-A1, H1-A2, H1-A3), the harmonic-to-noise ratio (HNR) at four different frequency bands (0-500 Hz, 0-1500 Hz, 0-2500 Hz, 0-3500 Hz), the cepstral peak prominence (CPP), the RMS energy, and the subharmonic-to-harmonic ratio (SHR). Before extracting these measures, the number of formants within 6000 Hz that provided the smoothest formant tracking in Praat (and therefore the LPC order) was determined for each speaker and these parameters were used in the Snack formant estimations. The average value for each acoustic measure during each vowel was then calculated with VoiceSauce.

Stress and syllable information was obtained from the annotated TextGrids and vowel and syllable durations were calculated. Formant measures were normalized using ΔF normalization (Johnson, 2020), a formant normalization technique that filters out variation due to vocal tract length differences. In this method, each speaker's ΔF – average distance between formants – is calculated with Equation (3.1) using $i$ number of formants from $j$ number of vowel tokens. The normalized formants are then calculated using Equation (3.2). Since ΔF normalization centers vowels on (1.5, 0.5) in the normalized F1-F2 formant space, normalized F1 and F2 measures were then centered on (0,0) using Equation (3.3).

(3.1)
$$\Delta F = \frac{1}{mn} \sum_{j}^{m} \sum_{i}^{n} \left[ \frac{F_{ij}}{i - 0.5} \right]$$

(3.2)
$$F_i^{'} = \frac{F_i}{\Delta F}$$

[3]English and Spanish grapheme-to-phoneme (G2P) and acoustic models are available in the Montreal Forced Aligner documentation. The G2P model used for Catalan was from the open-source speech synthesis software eSpeak (Silva, 2016). Scott James Perry provided me with the acoustic model for Catalan – also available at https://scottjamesperry.com/catalan-forced-alignment-materials/.

(3.3) $$F'_{i\,(0,0)} = F'_i - (i - 0.5)$$

Because normalized and centered F1 and F2 values were already normalized across speakers, these measures were set aside, and all remaining acoustic measures (vowel duration, syllable duration, H1H2, H2H4, H1A1, H1A2, H1A3 – all corrected for vowel quality in VoiceSauce – cepstral peak prominence [CPP], RMS energy, harmonic-to-noise ratio between 0–500 Hz [HNR05], between 0–1500 Hz [HNR15], between 0–2500 Hz [HNR25], and between 0–3500 Hz [HNR35], subharmonic-to-harmonic ratio [SHR], and four different measures of F0 [SHR method, STRAIGHT, Snack, and Praat]) were scaled within participant and were submitted to PCA. Based on the Scree plot of the eight components needed to explain 90% of the variance (Figure A.2), three components were selected for further analysis. Data were again submitted to PCA, this time with the number of components set to three. Table 3.3 shows the acoustic measures which correlate strongly to each component (strength of correlation $>| 0.4 |$).

| Comp. | Feature | Correlation | Exp. Variance |
|---|---|---|---|
| 1 | H1H2 | 0.53 | 41.70% |
| | H2H4 | 0.46 | |
| | H1A1 | 0.87 | |
| | H1A2 | 0.86 | |
| | H1A3 | 0.72 | |
| | CPP | 0.94 | |
| | HNR05 | 0.92 | |
| | HNR15 | 0.90 | |
| | HNR25 | 0.92 | |
| | HNR35 | 0.93 | |
| | Vowel duration | 0.76 | |
| 2 | F0-STRAIGHT | 0.81 | 15.16% |
| | F0-Snack | 0.80 | |
| | F0-Praat | 0.82 | |
| | F0-SNR | 0.70 | |
| 3 | H2H4 | 0.46 | 8.55% |
| | Syllable duration | 0.80 | |

Table 3.3: Strength of correlation of acoustic measures in production data to principal components with the percent variance explained by each component; Bartlett's sphericity test $p$-value $< 0.05$ and KMO $= 0.84$.

Based on the acoustic features correlated to each component, the first principal component can be interpreted as the degree of modal phonation (i.e., periodicity) and the duration of the vowel (henceforth, 'vowel duration [periodic]'). Harmonic-amplitude measures (used to measure spectral tilt) can correspond to both phonation type and articulatory effort; higher harmonic-amplitude measures have both been shown to correspond to more modal phonation – syllables that are *more* likely to be stressed (Keating et al., 2015) – and to speech produced with less articulatory effort – syllables that are *less* likely to be stressed (Sluijter & van Heuven, 1996b). Because the harmonic-amplitude measures in the first principal component are directly (rather than inversely) correlated to vowel duration (which is in turn expected to correlate directly with lexical stress), here I take these measures to indicate phonation type rather than articulatory effort. The cepstral peak prominence (CPP) corresponds to the periodicity of speech, among other things, where speech with clear harmonic structure has higher CPP and speech that is more aperiodic (e.g., as appears in some types of creak) has lower CPP (Fraile & Godino-Llorente, 2014). The harmonic to noise ratios (HNR) likewise correspond to periodicity, where lower values are perceptually correlated to a rougher voice quality (e.g., creak) (Ferrand, 2002). The interpretation of the second and third principal components are more straightforward given the correlated acoustic features; the second component corresponds to pitch (F0) and the third component to the duration of the syllable (with a slight correlation again with modal phonation). Table 3.4 contains descriptive statistics for each acoustic measure as produced across language, language profile, syllable, and stress. Violin plots of observations of each acoustic measure as produced across language, language profile, syllable, and stress are found in Figures A.3 to A.7.

For the trilingual production data in each language, a Bayesian multilevel model was fitted in R with Equation (3.4) using `brms` (Bürkner, 2017). Bayesian multilevel modeling has been used to model statistical phenomena on multiple levels (e.g., a model with fixed and random effects). Data structures with complicated sets of interaction terms and random slopes can often be fitted using a Bayesian framework, whereas frequentist logistic regression models fitted on the same structure may not converge (Nalborczyk et al., 2019). In Equation (3.4), the interactions between each acoustic measure, BLP principal components, and syllable model how participants with different relative language dominance use each acoustic measure across each syllable of the disyllabic words to cue stress. The by-participant random slopes allow for individual variation in cue-weighting across syllable type. To examine how language profile affects English production, Equation (3.5) was used in a separate Bayesian multilevel model. This formula models how participants with different language profiles use each acoustic measure across each syllable to cue stress. The by-participant random slopes allow for individual variation in cue-weighting across syllable type.

$$\begin{aligned}\text{stressed} \sim\ & \text{vowel duration} * \text{PC1} * \text{syllable} + \text{vowel duration} * \text{PC2} * \text{syllable} + \\ & \text{F0} * \text{PC1} * \text{syllable} + \text{F0} * \text{PC2} * \text{syllable} + \text{syllable duration} * \text{PC1} * \text{syllable} + \\ & \text{syllable duration} * \text{PC2} * \text{syllable} + \text{F1} * \text{PC1} * \text{syllable} + \text{F1} * \text{PC2} * \text{syllable} + \\ & \text{F2} * \text{PC1} * \text{syllable} + \text{F2} * \text{PC2} * \text{syllable} + (\text{vowel duration} * \text{syllable} + \\ & \text{F0} * \text{syllable} + \text{syllable duration} * \text{syllable} + \text{F1} * \text{syllable} + \text{F2} * \text{syllable} \mid \text{participant})\end{aligned} \tag{3.4}$$

| Language | Profile | Syllable | Stress | Vowel duration | Syllable duration | F1 | F2 | F0 |
|---|---|---|---|---|---|---|---|---|
| English | Monolingual | Penultimate | – | -2.357 (1.678) | -0.186 (0.640) | -0.104 (0.100) | 0.057 (0.329) | 0.683 (1.394) |
| | | | + | 0.524 (2.617) | 0.033 (1.142) | 0.045 (0.157) | -0.005 (0.363) | 0.133 (1.539) |
| | | Ultimate | – | 0.511 (2.684) | 0.067 (1.072) | -0.003 (0.132) | 0.149 (0.312) | -0.670 (1.648) |
| | | | + | 1.215 (2.522) | 0.075 (1.166) | 0.018 (0.131) | 0.087 (0.357) | -0.090 (1.343) |
| | L1 English | Penultimate | – | -2.808 (1.462) | 0.032 (0.711) | -0.090 (0.107) | 0.064 (0.320) | -0.060 (1.662) |
| | | | + | -0.299 (2.379) | 0.549 (1.034) | 0.041 (0.143) | 0.005 (0.322) | -0.563 (1.672) |
| | | Ultimate | – | -0.268 (2.448) | 0.425 (1.038) | -0.011 (0.119) | 0.150 (0.280) | -1.212 (1.545) |
| | | | + | 0.314 (2.312) | 0.580 (1.172) | 0.019 (0.113) | 0.080 (0.323) | -0.849 (1.544) |
| | L3 English | Penultimate | – | -1.658 (1.797) | 0.584 (0.748) | -0.108 (0.111) | 0.074 (0.404) | 0.880 (1.922) |
| | | | + | 1.232 (2.853) | 0.890 (1.157) | -0.002 (0.148) | 0.010 (0.427) | 0.376 (1.98) |
| | | Ultimate | – | 0.485 (3.024) | 0.869 (1.064) | -0.055 (0.114) | 0.196 (0.348) | -0.260 (2.975) |
| | | | + | 2.186 (2.793) | 1.063 (1.201) | -0.016 (0.118) | 0.154 (0.405) | -0.001 (1.941) |
| Spanish | L1 English | Penultimate | – | -1.319 (2.237) | -0.553 (0.899) | -0.194 (0.069) | 0.465 (0.409) | 0.076 (0.983) |
| | | | + | -0.382 (2.626) | -0.473 (0.950) | -0.202 (0.063) | 0.482 (0.485) | 0.780 (0.976) |
| | | Ultimate | – | 1.916 (2.046) | -0.810 (1.258) | -0.106 (0.111) | -0.253 (0.297) | 1.467 (1.598) |
| | | | + | 2.531 (1.551) | -0.019 (1.282) | -0.074 (0.109) | -0.323 (0.249) | 0.946 (1.315) |
| | L3 English | Penultimate | – | -2.126 (1.055) | -0.007 (0.489) | -0.202 (0.074) | 0.442 (0.422) | -0.188 (0.872) |
| | | | + | -1.681 (1.657) | 0.057 (0.720) | -0.202 (0.067) | 0.501 (0.401) | 0.228 (0.942) |
| | | Ultimate | – | 0.886 (2.061) | -1.312 (1.334) | -0.105 (0.104) | -0.081 (0.291) | 0.281 (1.710) |
| | | | + | 1.186 (2.376) | -0.889 (1.333) | -0.079 (0.114) | -0.141 (0.284) | -0.170 (1.166) |
| Catalan | L1 English | Penultimate | – | -1.668 (2.143) | -0.432 (0.812) | -0.167 (0.075) | 0.505 (0.378) | 0.201 (0.940) |
| | | | + | -0.777 (2.557) | -0.326 (0.925) | -0.182 (0.066) | 0.569 (0.415) | 0.606 (0.991) |
| | | Ultimate | – | 1.932 (2.060) | -0.672 (1.518) | 0.001 (0.142) | 0.054 (0.232) | 0.910 (1.850) |
| | | | + | 2.664 (1.900) | 0.400 (1.481) | 0.100 (0.159) | 0.025 (0.182) | -0.105 (1.601) |
| | L3 English | Penultimate | – | -2.341 (0.692) | 0.031 (0.434) | -0.192 (0.101) | 0.478 (0.424) | -0.083 (0.909) |
| | | | + | -1.794 (1.591) | 0.023 (0.757) | -0.201 (0.082) | 0.499 (0.486) | 0.127 (0.817) |
| | | Ultimate | – | 1.056 (2.152) | -1.358 (1.402) | -0.018 (0.151) | 0.089 (0.250) | -0.261 (1.311) |
| | | | + | 2.161 (2.305) | -0.524 (1.548) | 0.066 (0.162) | 0.082 (0.189) | -0.968 (1.108) |

Table 3.4: Means and standard deviations for the five acoustic metrics across language, language profile, syllable, and stress. Both F1 and F2 are the centered and normalized values.

$$\begin{aligned}\text{stressed} \sim\ &\text{vowel duration} * \text{language profile} * \text{syllable} + \text{F0} * \text{language profile} * \text{syllable} +\\ &\text{syllable duration} * \text{language profile} * \text{syllable} + \text{F1} * \text{language profile} * \text{syllable} +\\ &\text{F2} * \text{language profile} * \text{syllable} + (\text{vowel duration} * \text{syllable} + \text{F0} * \text{syllable} +\\ &\text{syllable duration} * \text{syllable} + \text{F1} * \text{syllable} + \text{F2} * \text{syllable} \mid \text{participant})\end{aligned} \tag{3.5}$$

## 3.3 Results

### English production compared to monolinguals

Production data from the English sentence elicitation task completed by monolingual and trilingual speakers were submitted to a multilevel Bayesian model with interactions between acoustic measures, syllable, and language profile, and random slopes of these interactions by participant (see Equation (3.5) for model formula). The logodds estimates from this model are shown in Table 3.5. For this statistical model, and all models described in subsequent subsections of this chapter, some post-hoc tests of significant interactions cannot be performed as the interactions include only continuous variables or include more than two variables. However, for these interactions, interpretations are gathered from visualizations of the predicted effects.

As shown in the visual representation of logodds estimates of stressed syllables in Figure 3.2, normalized F1 is the most heavily weighted cue across all speakers ($\beta = 7.62 \pm 0.81$, 95% CI = [6.06, 9.28]), followed by vowel duration (periodic) ($\beta = 0.69 \pm 0.05$, 95% CI = [0.60, 0.79]), normalized F2 ($\beta = 0.64 \pm 0.18$, 95% CI = [0.28, 0.99]), then syllable duration ($\beta = 0.57 \pm 0.11$, 95% CI = [0.36, 0.79]), in decreasing order of cue-weightings.

A post-hoc analysis of F0 and syllable reveal that with an increase in F0, the probability of stress is higher in ultimate syllables than in penultimate syllables ($\beta$ = -0.37, 95% CI = [-0.48, -0.27]). This trend appears to be shared by all speakers, regardless of language profile, and is visualized in Figure 3.3. Additionally, the trend of F0 in penultimate syllables is negative (slope = -0.11, 95% CI = [-0.17, -0.04]), indicating that an increase in F0 in penultimate syllables correlates to a decreased likelihood of stress.

Normalized F1 interacts significantly with syllable, and a post-hoc analysis shows that with an increase in normalized F1 (vowel lowering), the probability of stress is higher in penultimate syllables than in ultimate syllables ($\beta$ = 4.48, 95% CI = [3.40, 5.62]). This interaction is visualized in Figure 3.4. Therefore, when penultimate vowels (the majority of which are low vowels in the target stimuli) are produced with lower F1 (vowel raising towards /ə/), the syllable is more likely to be unstressed.

Likewise, normalized F2 interacts significantly with syllable, and a post-hoc analysis shows that with an increase in normalized F2 (vowel fronting), the probability of stress is higher in penultimate syllables than in ultimate syllables ($\beta$ = 0.89, 95% CI = [0.64, 1.16]). Given that the trend for ultimate syllables is negative, an increase in normalized F2 (vowel fronting) correlates to a decrease of stress likelihood in ultimate syllables (slope = -0.47,

| | Estimate | Error | L-95% CI | U-95% CI |
|---|---|---|---|---|
| Intercept | 1.58 | 0.13 | 1.33 | 1.85 |
| Vowel duration (periodic) | 0.69 | 0.05 | 0.60 | 0.79 |
| L3 English | -0.86 | 0.17 | -1.19 | -0.53 |
| Monolingual | -0.35 | 0.17 | -0.70 | -0.01 |
| Ultimate syllable | -1.29 | 0.15 | -1.60 | -1.01 |
| F0 | -0.09 | 0.06 | -0.21 | 0.02 |
| Syllable duration | 0.57 | 0.11 | 0.36 | 0.79 |
| F1 | 7.62 | 0.81 | 6.06 | 9.28 |
| F2 | 0.64 | 0.18 | 0.28 | 0.99 |
| Vowel duration (periodic) : L3 English | -0.18 | 0.06 | -0.30 | -0.07 |
| Vowel duration (periodic) : Monolingual | -0.13 | 0.07 | -0.26 | -0.00 |
| Vowel duration (periodic) : Ultimate syllable | -0.57 | 0.05 | -0.67 | -0.48 |
| L3 English : Ultimate syllable | 0.66 | 0.21 | 0.26 | 1.06 |
| Monolingual : Ultimate syllable | 0.07 | 0.20 | -0.33 | 0.46 |
| L3 English : F0 | -0.05 | 0.08 | -0.20 | 0.11 |
| Monolingual : F0 | 0.02 | 0.09 | -0.17 | 0.19 |
| Ultimate syllable : F0 | 0.38 | 0.09 | 0.21 | 0.56 |
| L3 English : Syllable duration | -0.43 | 0.15 | -0.73 | -0.13 |
| Monolingual : Syllable duration | -0.36 | 0.16 | -0.68 | -0.05 |
| Ultimate syllable : Syllable duration | -0.59 | 0.12 | -0.84 | -0.36 |
| L3 English : F1 | -2.04 | 1.11 | -4.26 | 0.12 |
| Monolingual : F1 | 1.18 | 1.16 | -1.08 | 3.50 |
| Ultimate syllable : F1 | -4.67 | 0.93 | -6.59 | -2.85 |
| L3 English : F2 | -0.38 | 0.23 | -0.83 | 0.08 |
| Monolingual : F2 | -0.25 | 0.26 | -0.76 | 0.25 |
| Ultimate syllable : F2 | -1.41 | 0.23 | -1.88 | -0.95 |
| Vowel duration (periodic) : L3 English : Ultimate syllable | 0.28 | 0.06 | 0.16 | 0.40 |
| L3 English : Ultimate syllable : Syllable duration | 0.34 | 0.17 | 0.00 | 0.68 |
| L3 English : Ultimate syllable : F2 | 1.04 | 0.31 | 0.43 | 1.63 |

Table 3.5: Coefficients for the Bayesian multilevel model of data from the English production task by trilingual and monolingual speakers. The intercept represents the probability of stressed vowel productions in the penultimate syllable by L1 English trilinguals. To fit the table on one page, three-way interactions with a confidence interval spanning zero are not shown.

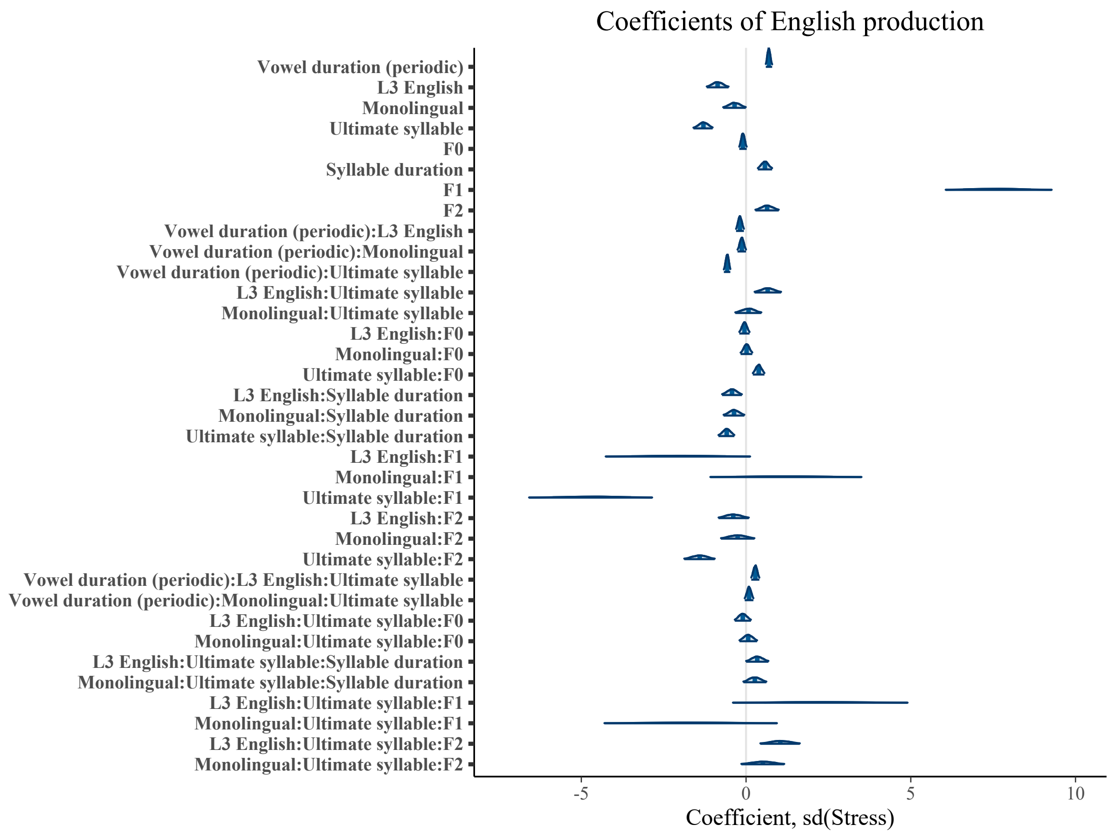



Figure 3.2: Coefficients from the Bayesian multilevel model of English production across language profile. Coefficients with a distribution that does not include $x = 0$ are statistically distinct from 0 within a 95% confidence interval.

95% CI = [-0.63, -0.29]). As most penultimate vowels are front vowels and most ultimate vowels are back vowels, these trends occur as expected by patterns of vowel reduction in English. The three-way interaction between normalized F2, syllable, and language profile is also significant (β = 1.04 ± 0.31, 95% CI = [0.43, 1.63]), where the aforementioned effect of F2 in penultimate and ultimate syllables is greater for L1 English speakers than for L3 English speakers. This interaction is visualized in Figure 3.5.

A post-hoc analysis of the interaction of vowel duration (periodic) and syllable shows that with an increase in vowel duration and degree of modal phonation, the log likelihood of stress is higher in penultimate than ultimate syllables (β = 0.45, 95% CI = [0.39, 0.51]). Additionally, a post-hoc analysis of the interaction of vowel duration (periodic) and language profile shows that with an increase in vowel duration and degree of modal phonation, the log likelihood of stress is higher for L1 English trilinguals than for monolinguals (β = 0.08, 95% CI = [0.00, 0.17]). Further, Figure 3.6 shows the combined effects of language profile, vowel duration (periodic), and syllable on the predicted proportion of stressed vowels. From

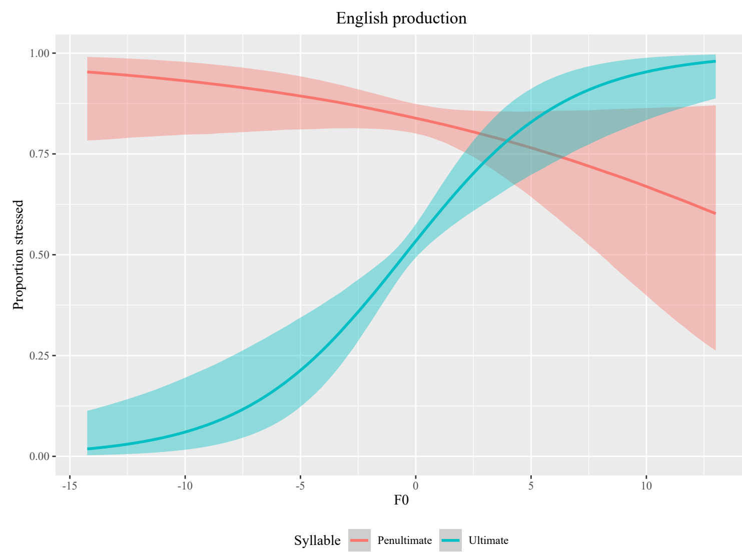



Figure 3.3: Predicted effects of F0 across syllable in English production.

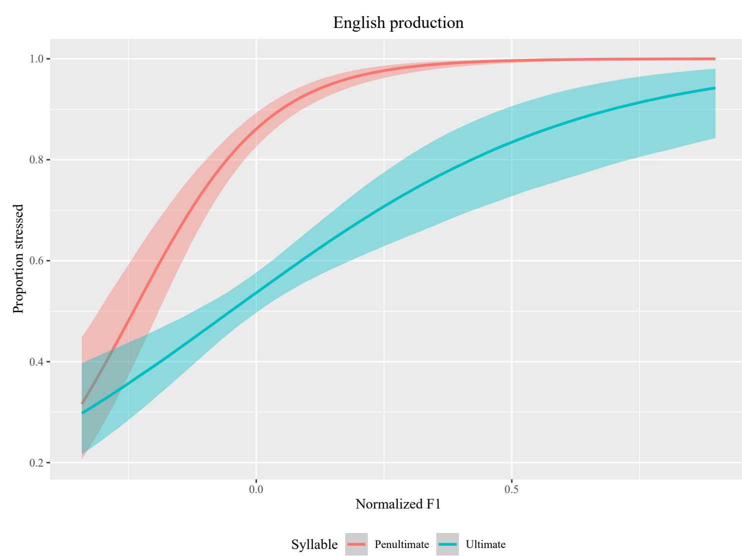



Figure 3.4: Predicted effects of normalized F1 across syllable in English production.

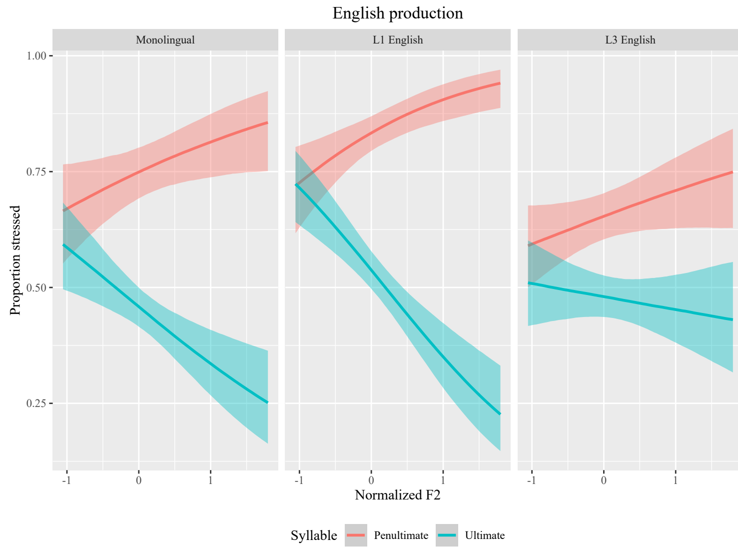



Figure 3.5: Predicted effects of normalized F2 across syllable and language profile in English production.

this figure, it appears that L3 English trilinguals use vowel duration (periodic) to cue stress shifts more strongly in the ultimate syllable than the monolingual speakers and L1 English trilinguals. In turn, the monolingual speakers and L1 English trilinguals appear to use vowel duration (periodic) more strongly as a cue to stress shifts in the penultimate syllable than L3 English trilinguals.

Similarly, syllable duration also interacts with syllable, where with an increase in syllable duration, the log likelihood of stress is higher in penultimate than ultimate syllables ($\beta$ = 0.39, 95% CI = [0.25, 0.52]). Additionally, a post-hoc analysis of the interaction of syllable duration and language profile shows that with an increase in syllable duration, the log likelihood of stress is higher for L1 English trilinguals than for monolinguals ($\beta$ = 0.23, 95% CI = [0.03, 0.43]) and for L3 English trilinguals ($\beta$ = 0.26, 95% CI = [0.07, 0.45]). Further, Figure 3.7 shows the combined effects of language profile, syllable duration, and syllable on the predicted proportion of stressed vowels. From this figure, it appears that L1 English trilinguals have higher cue-weightings for syllable duration in penultimate syllables than either the L3 English trilinguals or the monolingual speakers. Additionally, the L3 English trilinguals and monolingual speakers appear to have an inverse relationship between syllable duration and stress likelihood in ultimate syllables, whereas L1 English trilinguals appear to have no such correlation.

Lastly, the significant interaction between language profile and syllable is a by-product of unequal token counts across speakers. Since all speakers read from the same list of words, in theory the log likelihood of stress in each syllable type (penultimate and ultimate) should be the same across all speakers. However, words which were improperly stressed

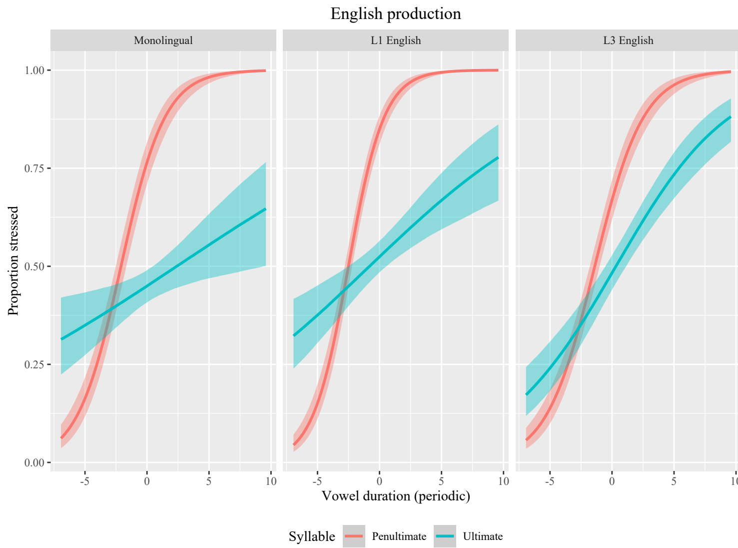



Figure 3.6: Predicted effects of vowel duration (periodic) across syllable and language profile in English production.

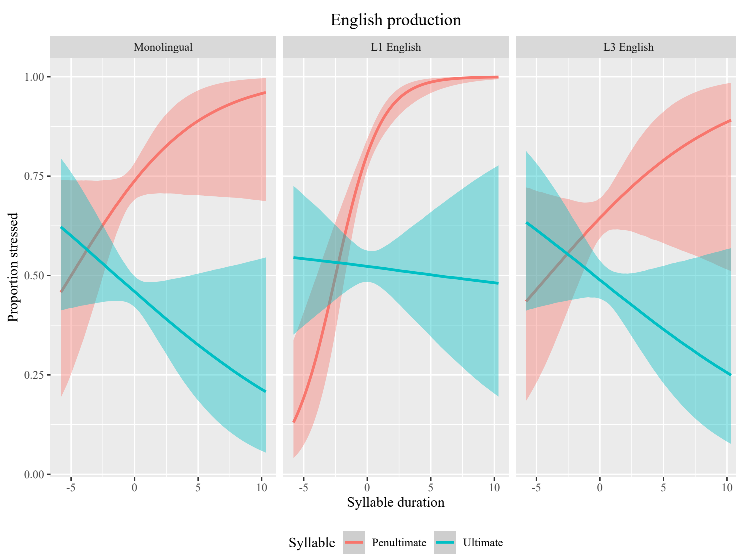



Figure 3.7: Predicted effects of syllable duration across syllable and language profile in English production.

were removed from analysis, resulting in a token imbalance. Whereas monolingual English speakers produced 1,075 stressed penultimate syllables and 974 stressed ultimate syllables, L1 English trilinguals produced 1,345 stressed penultimate syllables and 1,243 stressed ultimate syllables, and L3 English trilinguals produced 1,159 stressed penultimate syllables and 1,087 stressed ultimate syllables.

## English production by trilinguals

English production data by trilingual speakers were isolated and analyzed again, this time using the principal components from the BLP data to analyze production across relative language dominance. The data were submitted to a multilevel Bayesian model with separate interactions between acoustic measures, syllable, and each principal component from the BLP data, with random slopes of these interactions by participant (see Equation (3.4) for model formula). The logodds estimates from this model are shown in Table 3.6.

As shown in the visual representation of logodds estimates of stressed syllables in Figure 3.8, normalized F1 is the most heavily weighted cue across all trilinguals ($\beta = 6.82 \pm 0.60$, 95% CI = [5.68, 8.02]), followed by vowel duration (periodic) ($\beta = 0.61 \pm 0.04$, 95% CI = [0.54, 0.68]), normalized F2 ($\beta = 0.49 \pm 0.12$, 95% CI = [0.26, 0.73]), syllable duration ($\beta = 0.37 \pm 0.08$, 95% CI = [0.22, 0.52]), then F0 ($\beta = -0.12 \pm 0.04$, 95% CI = [-0.21, -0.04]), in order of decreasing logodds estimates.

A post-hoc analysis of F0 and syllable reveal that with an increase in F0, the probability of stress is higher in ultimate syllables than penultimate syllables ($\beta = -0.36$, 95% CI = [-0.49, -0.23]). This trend appears to be shared by all speakers, regardless of relative dominance in the three languages, and is visualized in Figure 3.9. Additionally, with an increase in F0, the probability of stress is higher for speakers with greater Catalan dominance relative to Spanish dominance (higher PC2; $\beta = 0.03 \pm 0.02$, 95% CI = [0.00, 0.07]).

Normalized F1 interacts significantly with syllable, and a post-hoc analysis shows that with an increase in normalized F1 (vowel lowering), the probability of stress is higher in penultimate syllables than in penultimate syllables ($\beta = 3.52$, 95% CI = [2.25, 4.87]). This interaction is visualized in Figure 3.10. Therefore, when penultimate vowels (the majority of which are low vowels) are produced with lower F1 (vowel raising towards /ə/), the syllable is more likely to be unstressed.

Likewise, normalized F2 interacts significantly with syllable, and a post-hoc analysis shows that with an increase in normalized F2 (vowel fronting), the probability of stress is higher in penultimate syllables than in ultimate syllables ($\beta = 0.95$, 95% CI = [0.64, 1.27]). Given that the trend for ultimate syllables is negative, an increase in normalized F2 (vowel fronting) correlates to a decrease of stress likelihood in ultimate syllables (slope = -0.43, 95% CI = [-0.65, -0.22]). As most penultimate vowels are front vowels and most ultimate vowels are back vowels, these trends occur as expected. The three-way interaction between normalized F2, syllable, and PC1 (English dominance relative to Spanish and Catalan dominance) is also significant, where for speakers with higher English dominance relative to Spanish and Catalan dominance, the direct correlation between F2 and stress likelihood in penultimate

| | Estimate | Error | L-95% CI | U-95% CI |
|---|---|---|---|---|
| Intercept | 1.17 | 0.10 | 0.98 | 1.38 |
| Vowel duration (periodic) | 0.61 | 0.04 | 0.54 | 0.68 |
| PC1 | 0.08 | 0.02 | 0.04 | 0.12 |
| Ultimate syllable | -0.97 | 0.12 | -1.21 | -0.74 |
| PC2 | 0.01 | 0.04 | -0.06 | 0.08 |
| F0 | -0.12 | 0.04 | -0.21 | -0.04 |
| Syllable duration | 0.37 | 0.08 | 0.22 | 0.52 |
| F1 | 6.82 | 0.60 | 5.68 | 8.02 |
| F2 | 0.49 | 0.12 | 0.26 | 0.73 |
| Vowel duration (periodic) : PC1 | 0.02 | 0.01 | 0.00 | 0.03 |
| Vowel duration : Ultimate syllable | -0.44 | 0.04 | -0.51 | -0.37 |
| PC1 : Ultimate syllable | -0.05 | 0.02 | -0.10 | -0.01 |
| Vowel duration : PC2 | -0.01 | 0.01 | -0.04 | 0.01 |
| Ultimate syllable : PC2 | -0.07 | 0.05 | -0.15 | 0.02 |
| PC1 : F0 | 0.00 | 0.01 | -0.02 | 0.02 |
| Ultimate syllable : F0 | 0.35 | 0.07 | 0.22 | 0.48 |
| PC2 : F0 | 0.03 | 0.02 | -0.00 | 0.07 |
| PC1 : Syllable duration | 0.05 | 0.02 | 0.02 | 0.08 |
| Ultimate syllable : Syllable duration | -0.43 | 0.09 | -0.60 | -0.27 |
| PC2 : Syllable duration | -0.03 | 0.03 | -0.08 | 0.03 |
| PC1 : F1 | 0.16 | 0.11 | -0.06 | 0.39 |
| Ultimate syllable : F1 | -3.49 | 0.68 | -4.84 | -2.21 |
| PC2 : F1 | 0.36 | 0.24 | -0.12 | 0.83 |
| PC1 : F2 | 0.05 | 0.02 | 0.00 | 0.09 |
| Ultimate syllable : F2 | -0.90 | 0.16 | -1.22 | -0.59 |
| PC2 : F2 | 0.09 | 0.05 | 0.00 | 0.18 |
| Vowel duration (periodic) : PC1 : Ultimate syllable | -0.03 | 0.01 | -0.04 | -0.01 |
| PC1 : Ultimate syllable : Syllable duration | -0.05 | 0.02 | -0.08 | -0.01 |
| PC1 : Ultimate syllable : F2 | -0.11 | 0.03 | -0.18 | -0.05 |

Table 3.6: Coefficients for the Bayesian multilevel model of data from the English production task by trilingual speakers. The intercept represents the probability of stressed vowel productions in the penultimate syllable. To fit the table on one page, three-way interactions with a confidence interval spanning zero are not shown.

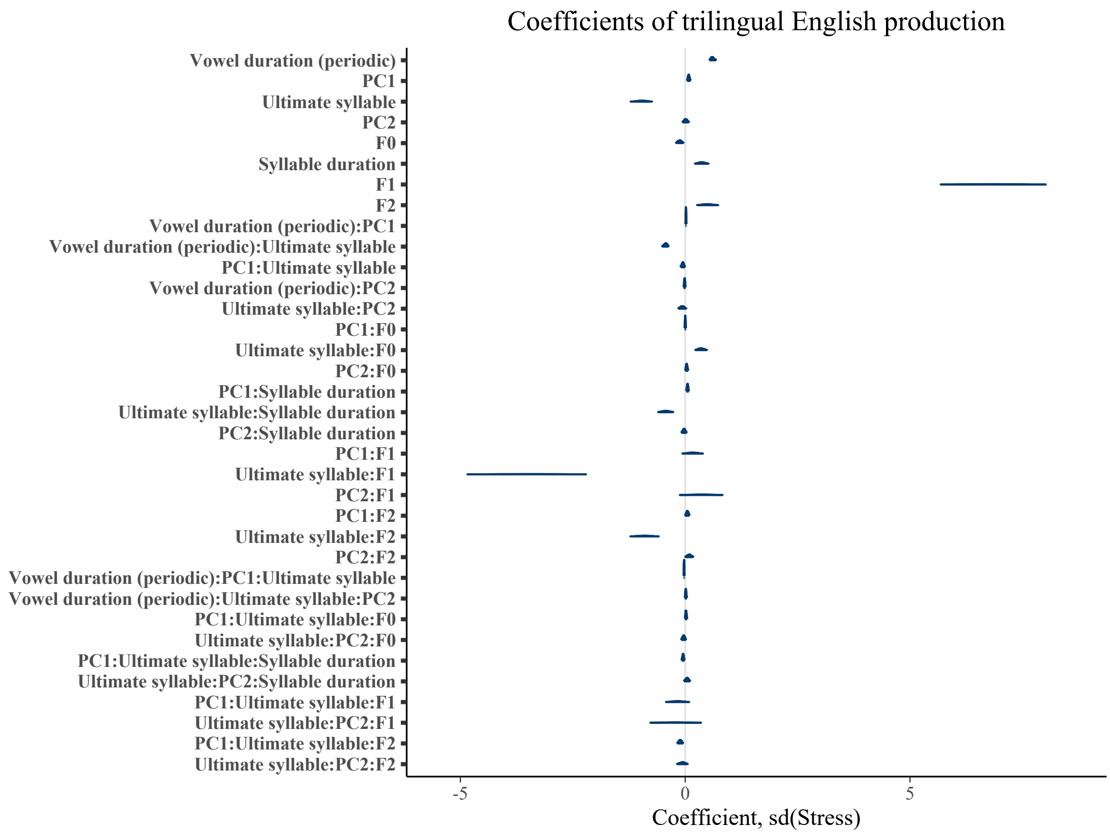



Figure 3.8: Coefficients from the Bayesian multilevel model of English production by trilinguals. Coefficients with a distribution that does not include $x = 0$ are statistically distinct from 0 within a 95% confidence interval.

syllables and the inverse correlation between F2 and stress likelihood in ultimate syllables are even stronger. This interaction is visualized in Figure 3.11. Lastly, the cue-weighting of F2 increases as Catalan dominance relative to Spanish dominance increases (higher PC2; β = 0.09 ± 0.05, 95% CI = [0.00, 0.18]).

The principal component of vowel duration (periodic) interacts significantly with PC1 (English dominance relative to Spanish and Catalan dominance) and syllable. In the visualization of this interaction (Figure 3.12), cue-weightings for vowel duration (periodic) are heavier in the penultimate syllable than in the ultimate syllables (β = 0.44, 95% CI = [0.37, 0.52]). However, increasing English dominance relative to Spanish and Catalan dominance correlates to an increase of cue-weightings of vowel duration (periodic) in penultimate syllables, but a decrease of cue-weightings for vowel duration (periodic) in ultimate syllables (β = -0.03 ± 0.01, 95% CI = [-0.04, -0.01]).

Syllable duration also interacts with PC1 (English dominance relative to Spanish and Catalan dominance) and syllable. In the visualization of this interaction (Figure 3.13), cue-

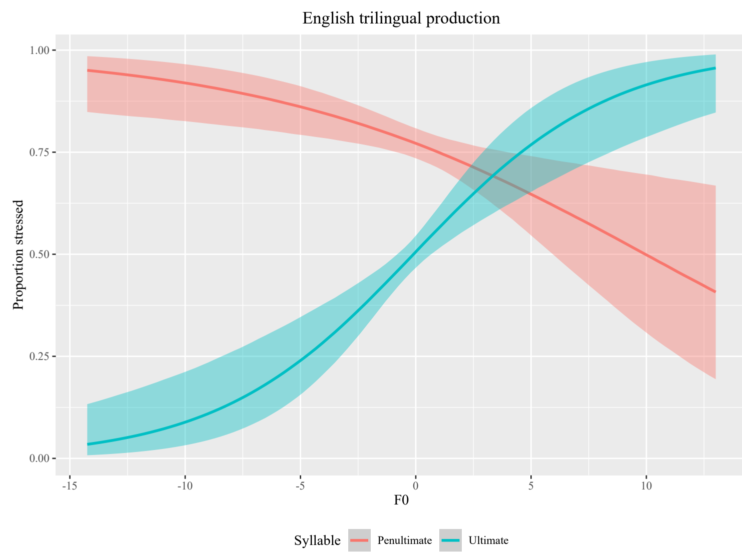



Figure 3.9: Predicted effects of F0 across syllable in English trilingual production.

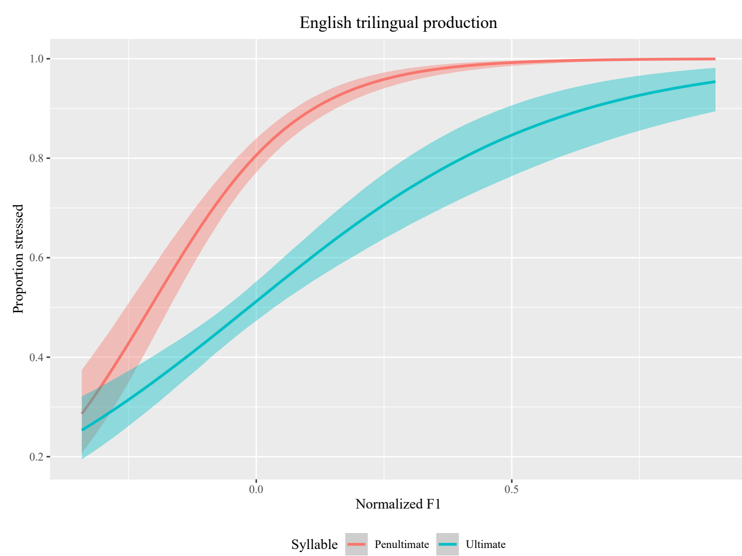



Figure 3.10: Predicted effects of normalized F1 across syllable in Spanish production.

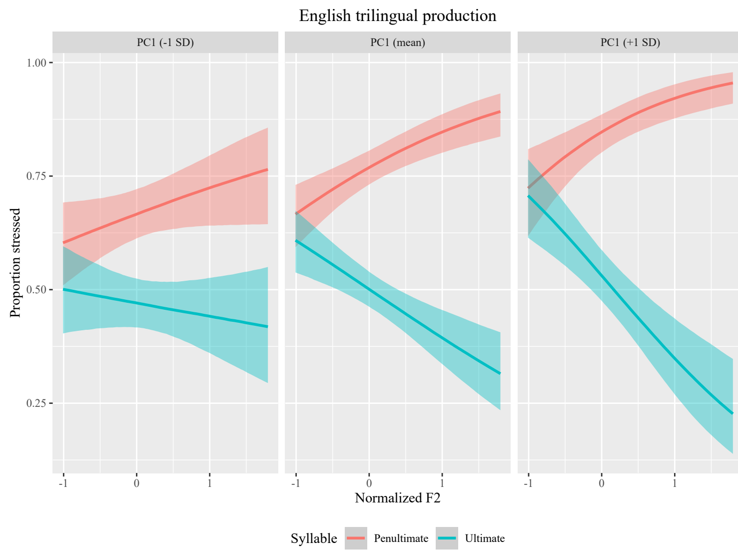



Figure 3.11: Predicted effects of normalized F2 across syllable and PC1 in English production.

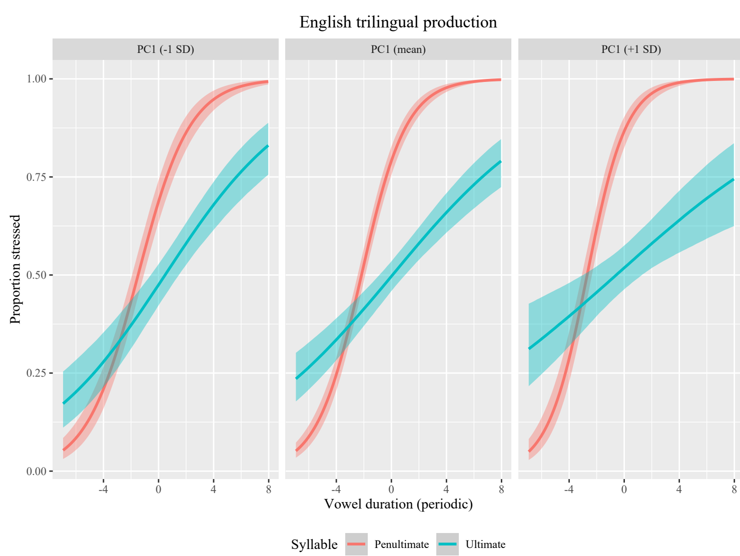



Figure 3.12: Predicted effects of duration (periodic) across syllable and PC1 in Spanish production.

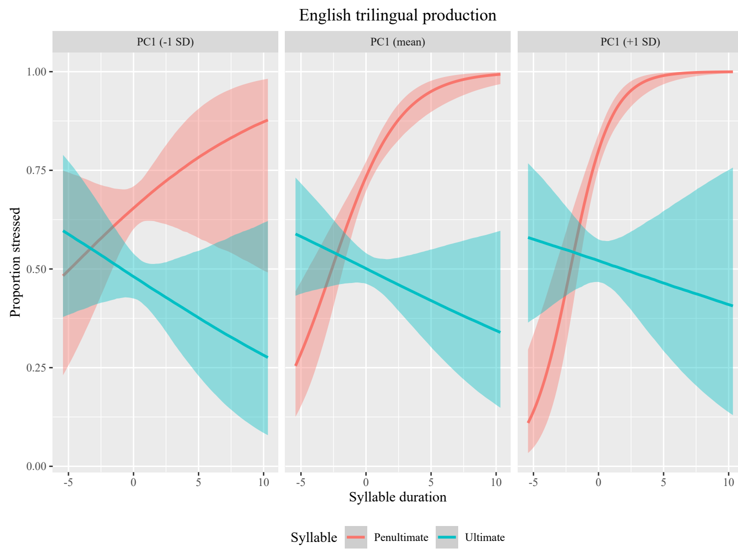



Figure 3.13: Predicted effects of syllable duration across syllable and PC1 in Spanish production.

weightings for syllable duration are heavier in the penultimate syllable than in the ultimate syllables (β = 0.45, 95% CI = [0.29, 0.63]). For speakers with higher English dominance relative to Spanish and Catalan dominance (PC1), cue-weightings for syllable duration in penultimate syllables increase (β = -0.05 ± 0.02, 95% CI = [-0.08, -0.01]).

Lastly, the significant interaction between PC1 (English dominance relative to Spanish and Catalan dominance) and syllable is again a by-product of unequal token counts across speakers. According to the model, an increase in PC1 correlates to a higher probability of stress in penultimate than ultimate syllables (β = 0.09, 95% CI = [0.04, 0.14]). Since all speakers read from the same list of words, in theory the log likelihood of stress in each syllable type (penultimate and ultimate) should be the same across all speakers. However, words which were improperly stressed were removed from analysis, resulting in a token imbalance. Speakers with higher PC1 (greater than 0) had 1,345 stressed penultimate syllables and 1,243 stressed ultimate syllables, whereas speakers with lower PC1 (less than 0) had 1,159 stressed penultimate syllables and 1,087 stressed ultimate syllables.

## Spanish production by trilinguals

Production data from the Spanish elicitation task were submitted to a multilevel Bayesian model with separate interactions between acoustic measures, syllable, and each principal component from the BLP data, with random slopes of these interactions by participant (see Equation (3.4) for model formula). The logodds estimates from this model are shown in Table 3.7. As shown in the visual representation of logodds estimates of stressed syllables in

| | Estimate | Error | L-95% CI | U-95% CI |
|---|---|---|---|---|
| Intercept | -0.08 | 0.30 | -0.66 | 0.52 |
| Vowel duration (periodic) | 0.52 | 0.08 | 0.38 | 0.68 |
| PC1 | -0.17 | 0.06 | -0.29 | -0.04 |
| Ultimate syllable | 0.93 | 0.43 | 0.10 | 1.77 |
| PC2 | -0.12 | 0.12 | -0.37 | 0.13 |
| F0 | 1.05 | 0.18 | 0.71 | 1.41 |
| Syllable duration | 0.68 | 0.13 | 0.43 | 0.94 |
| F1 | -0.64 | 1.13 | -2.80 | 1.64 |
| F2 | 0.37 | 0.21 | -0.03 | 0.81 |
| Vowel duration (periodic) : PC1 | -0.05 | 0.02 | -0.08 | -0.02 |
| Vowel duration (periodic) : Ultimate syllable | -0.35 | 0.08 | -0.52 | -0.20 |
| PC1 : Ultimate syllable | 0.18 | 0.09 | 0.01 | 0.35 |
| Vowel duration : PC2 | -0.07 | 0.03 | -0.12 | -0.01 |
| Ultimate syllable : PC2 | 0.13 | 0.17 | -0.21 | 0.47 |
| PC1 : F0 | 0.02 | 0.04 | -0.04 | 0.10 |
| Ultimate syllable : F0 | -1.56 | 0.26 | -2.08 | -1.07 |
| PC2 : F0 | 0.06 | 0.07 | -0.07 | 0.20 |
| PC1 : Syllable duration | -0.04 | 0.03 | -0.09 | 0.01 |
| Ultimate syllable: Syllable duration | -0.14 | 0.15 | -0.42 | 0.15 |
| PC2 : Syllable duration | -0.12 | 0.05 | -0.22 | -0.02 |
| PC1 : F1 | 0.04 | 0.22 | -0.40 | 0.46 |
| Ultimate syllable : F1 | 3.57 | 1.44 | 0.75 | 6.42 |
| PC2 : F1 | 0.37 | 0.49 | -0.55 | 1.33 |
| PC1 : F2 | -0.09 | 0.05 | -0.18 | -0.01 |
| Ultimate syllable : F2 | -2.16 | 0.39 | -2.99 | -1.44 |
| PC2 : F2 | -0.04 | 0.08 | -0.21 | 0.12 |
| Vowel duration (periodic) : PC1 : Ultimate syllable | 0.06 | 0.02 | 0.03 | 0.09 |
| Vowel duration (periodic) : Ultimate syllable : PC2 | 0.07 | 0.03 | 0.01 | 0.13 |
| PC1 : Ultimate syllable : Syllable duration | 0.08 | 0.03 | 0.03 | 0.14 |
| Ultimate syllable : PC2 : Syllable duration | 0.14 | 0.06 | 0.03 | 0.25 |

Table 3.7: Coefficients for the Bayesian multilevel model of data from the Spanish production task. The intercept represents the probability of stressed vowel productions in the penultimate syllable. To fit the table on one page, three-way interactions with a confidence interval spanning zero are not shown.

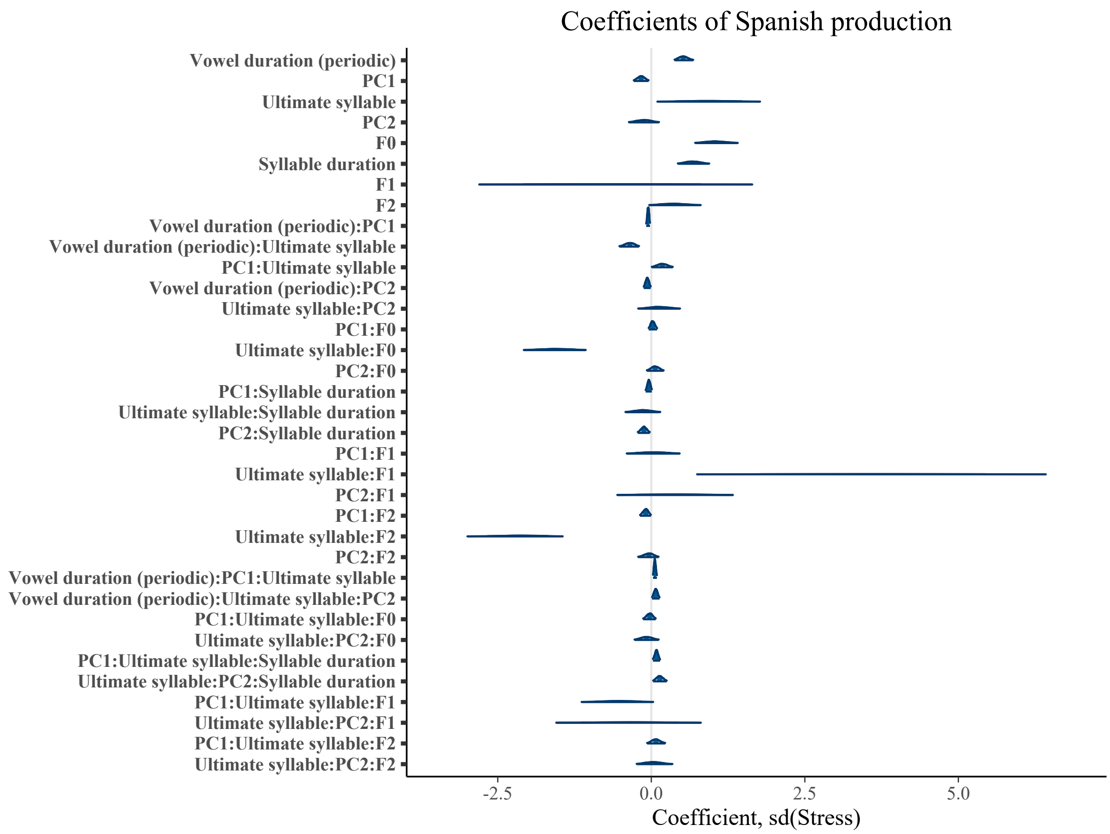



Figure 3.14: Coefficients from the Bayesian multilevel model of Spanish production. Coefficients with a distribution that does not include $x = 0$ are statistically distinct from 0 within a 95% confidence interval.

Figure 3.14, F0 is the most heavily weighted cue across all trilinguals (β = 1.05 ± 0.18, 95% CI = [0.71, 1.41]), followed by syllable duration (β = 0.68 ± 0.13, 95% CI = [0.43, 0.94]), then vowel duration (periodic) (β = 0.52 ± 0.08, 95% CI = [0.38, 0.68]), by decreasing cue-weightings.

A post-hoc analysis of F0 and syllable reveal that with an increase in F0, the probability of stress is higher in penultimate syllables than ultimate syllables (β = 1.53, 95% CI = [1.03, 2.04]). The relationship between F0 and syllable is visualized in Figure 3.15, where F0 is directly correlated to stress in penultimate syllables, but is inversely correlated to stress in ultimate syllables (slope = -0.51, 95% CI = [-0.81, -0.25]).

Normalized F1 interacts significantly with syllable, and a post-hoc analysis shows that with an increase in normalized F1 (vowel lowering), the probability of stress is higher in ultimate syllables (/o/) than in penultimate syllables (/i/; β = -3.93, 95% CI = [-6.65, -1.10]). This interaction is visualized in Figure 3.16. Therefore, when /o/ is produced with lower F1 (vowel raising towards /u/), the syllable is more likely to be unstressed.

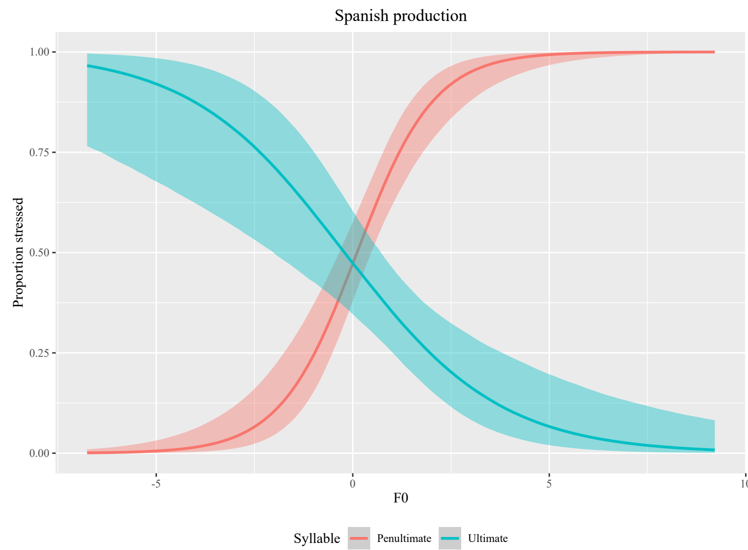



Figure 3.15: Predicted effects of F0 across syllable in Spanish production.

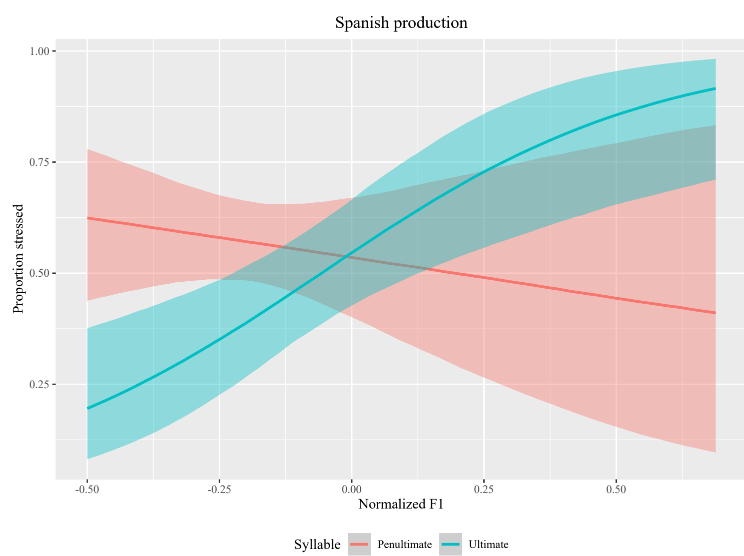



Figure 3.16: Predicted effects of normalized F1 across syllable in Spanish production.

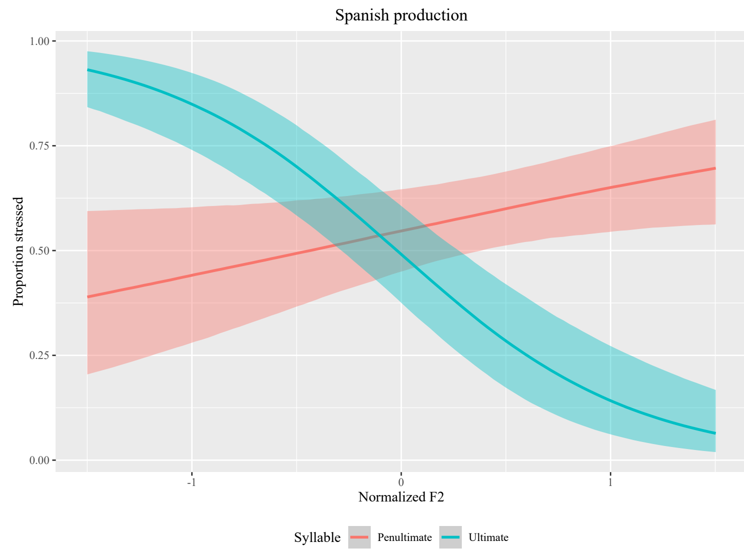



Figure 3.17: Predicted effects of normalized F2 across syllable in Spanish production.

Likewise, normalized F2 interacts significantly with syllable, and a post-hoc analysis shows that with an increase in normalized F2 (vowel fronting), the probability of stress is higher in penultimate syllables (/i/) than in ultimate syllables (/o/; $\beta = 2.19$, 95% CI = [1.47, 2.99]). When /i/ is produced with lower F2 (vowel backing - centralization), the syllable is more likely to be unstressed. This interaction is visualized in Figure 3.17. Additionally, the slope of the effect of normalized F2 in ultimate syllables (/o/) is negative, where increasing F2 (vowel fronting - centralization) is correlated with a lower likelihood of stress (slope = -1.76, 95% CI = [-2.46, -1.16]). Therefore, when /o/ is produced with higher F2 (vowel fronting - centralization), the syllable is more likely to be unstressed.

The principal component of vowel duration (periodic) interacts significantly with PC1 (English dominance relative to Spanish and Catalan dominance) and syllable. In the visualization of this interaction (Figure 3.18), cue-weightings for vowel duration and the degree of modal phonation are heavier in the penultimate syllable than in the ultimate syllables ($\beta = 0.40$, 95% CI = [0.24, 0.57]), however increasing English dominance relative to Spanish and Catalan dominance correlates to a decrease of cue-weighting of vowel duration and the degree of modal phonation in both syllables ($\beta = -0.05 \pm 0.02$, 95% CI = [-0.08, -0.02]), particularly penultimate syllables. Similarly, as PC2 increases (Catalan dominance relative to Spanish dominance), there is a decrease of cue-weighting of vowel duration and the degree of modal phonation in both syllables ($\beta = -0.07 \pm 0.03$, 95% CI = [-0.12, -0.01]), again, particularly in penultimate syllables, as shown in Figure 3.19.

Syllable duration also interacts with PC1 (English dominance relative to Spanish and Catalan dominance) and syllable. In the visualization of this interaction (Figure 3.20), cue-weightings for syllable duration are heavier in the penultimate syllable than in the ultimate syllables for speakers with lower English dominance relative to Spanish and Catalan dominance (PC1) and this trend flips for speakers with higher English dominance relative to Spanish and Catalan dominance ($\beta = 0.08 \pm 0.03$, 95% CI = [0.03, 0.14]).

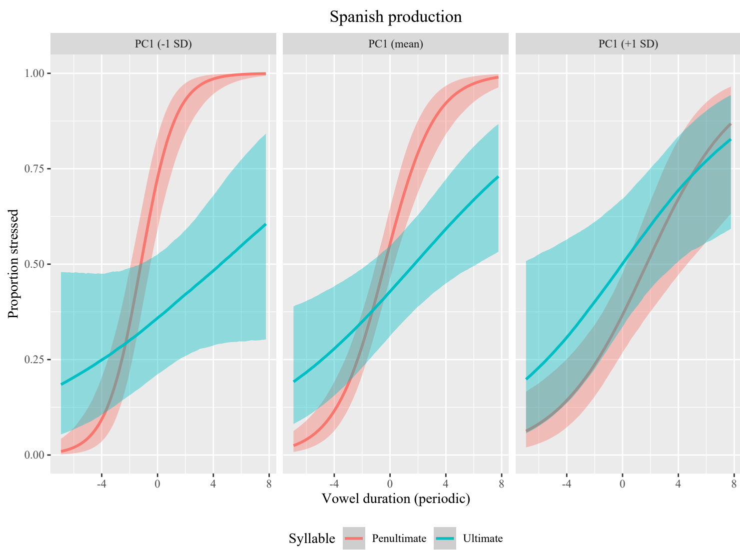



Figure 3.18: Predicted effects of vowel duration (periodic) across syllable and PC1 in Spanish production.

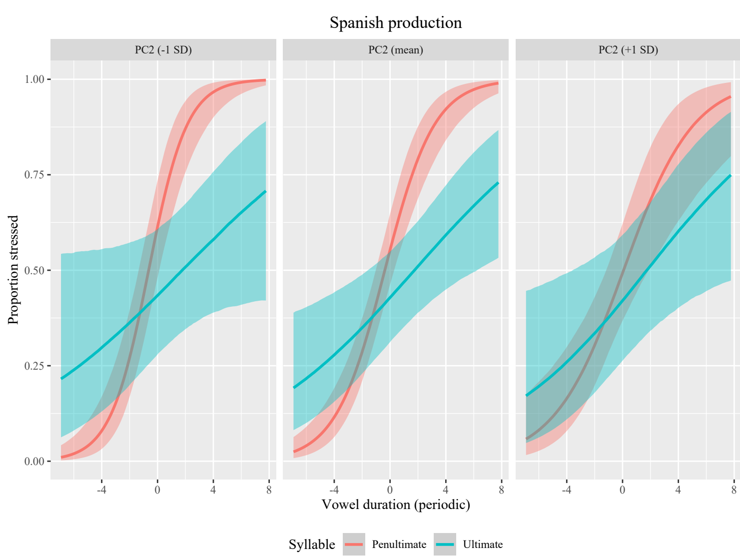



Figure 3.19: Predicted effects of vowel duration (periodic) across syllable and PC2 in Spanish production.

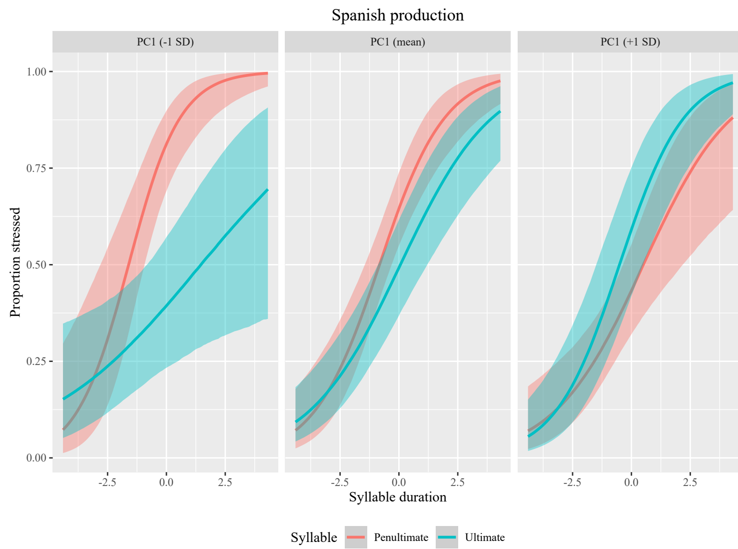



Figure 3.20: Predicted effects of syllable duration across syllable and PC1 in Spanish production.

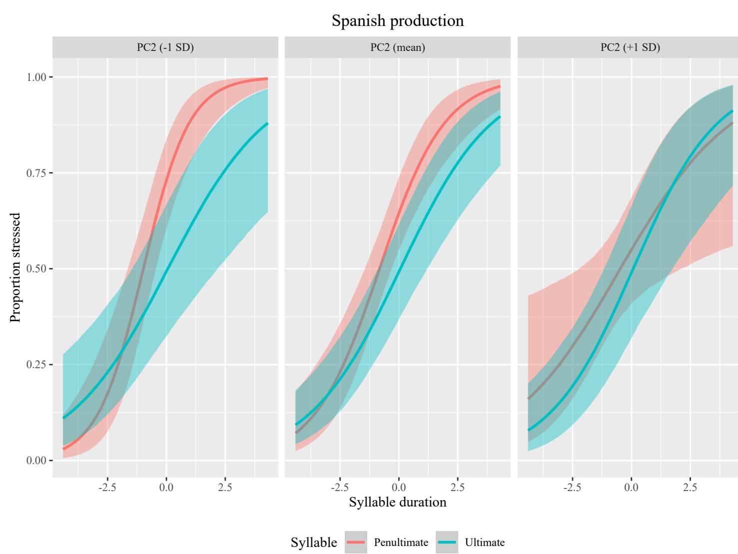



Figure 3.21: Predicted effects of syllable duration across syllable and PC2 in Spanish production.

Additionally, cue-weightings for syllable duration are heavier for speakers with less Catalan dominance relative to Spanish dominance (PC2; $\beta$ = -0.12 $\pm$ 0.05, 95% CI = [-0.22,

-0.02]), and these speakers also have higher cue-weightings in the penultimate syllable (β = 0.14 ± 0.06, 95% CI = [0.03, 0.25]), as shown in Figure 3.21.

Lastly, the significant interaction between PC1 (English dominance relative to Spanish and Catalan dominance) and syllable is again a by-product of unequal token counts across speakers. According to the model, an increase in PC1 correlates to a higher probability of stress in ultimate than penultimate syllables (β = -0.22, 95% CI = [-0.37, -0.08]). Since all speakers read from the same list of words, in theory the log likelihood of stress in each syllable type (penultimate and ultimate) should be the same across all speakers. However, words which were improperly stressed were removed from analysis, resulting in a token imbalance. Speakers with higher PC1 (greater than 0) had 605 stressed penultimate syllables and 954 stressed ultimate syllables, whereas speakers with lower PC1 (less than 0) had 886 stressed penultimate syllables and 1,013 stressed ultimate syllables.

## Catalan production by trilinguals

Production data from the Catalan elicitation task were submitted to a multilevel Bayesian model with separate interactions between acoustic measures, syllable, and each principal component from the BLP data, with random slopes of these interactions by participant (see Equation (3.4) for model formula). The logodds estimates from this model are shown in Table 3.8. As shown in the visual representation of logodds estimates of stressed syllables in Figure 3.22, normalized F1 (vowel height) is the most heavily weighted cue across all trilinguals (β = -3.89 ± 1.18, 95% CI = [-6.20, -1.60]), followed by syllable duration (β = 1.07 ± 0.15, 95% CI = [0.79, 1.38]), normalized F2 (β = 0.79 ± 0.23, 95% CI = [0.36, 1.27]), vowel duration (periodic) (β = 0.70 ± 0.08, 95% CI = [0.55, 0.85]), and F0 (β = 0.63 ± 0.18, 95% CI = [0.28, 0.99]), by decreasing cue-weightings.

A post-hoc analysis of F0 and syllable reveal that with an increase in F0, the probability of stress is higher in penultimate syllables than ultimate syllables (β = 1.37, 95% CI = [0.86, 1.88]). The relationship between F0 and syllable is visualized in Figure 3.23, where F0 is directly correlated to stress in penultimate syllables, but is inversely correlated to stress in ultimate syllables (slope = -0.76, 95% CI = [-1.05, -0.46]).

Normalized F1 interacts significantly with syllable, and a post-hoc analysis shows that with an increase in normalized F1 (vowel lowering), the probability of stress is higher in ultimate syllables (/a/) than in penultimate syllables (/i/; β = -16.40, 95% CI = [-21.60, -11.63]). This makes sense that more peripheral productions of /a/ (F1 lowering) are more likely to be stressed, given that Catalan has prescriptive reduction (i.e., F1 raising) of /a/ to /schwa/ in unstressed syllables. This interaction is visualized in Figure 3.24. Additionally, the slope of the effect of normalized F1 in penultimate syllables (/i/) is negative, where increasing F1 (vowel lowering, i.e., centralization) is correlated with a lower likelihood of stress (slope = -3.66, 95% CI = [-5.93, -1.37]).

Normalized F2 also interacts significantly with syllable, and a post-hoc analysis shows that with an increase in normalized F2 (vowel fronting), the probability of stress is higher in penultimate syllables (/i/) than in ultimate syllables (/a/; β = 3.6, 95% CI = [1.51,

| | Estimate | Error | L-95% CI | U-95% CI |
|---|---|---|---|---|
| Intercept | 0.06 | 0.28 | -0.48 | 0.61 |
| Vowel duration (periodic) | 0.70 | 0.08 | 0.55 | 0.85 |
| PC1 | -0.36 | 0.05 | -0.47 | -0.26 |
| Ultimate syllable | -0.78 | 0.43 | -1.64 | 0.05 |
| PC2 | 0.02 | 0.11 | -0.20 | 0.23 |
| F0 | 0.63 | 0.18 | 0.28 | 0.99 |
| Syllable duration | 1.07 | 0.15 | 0.79 | 1.38 |
| F1 | -3.89 | 1.18 | -6.20 | -1.60 |
| F2 | 0.79 | 0.23 | 0.36 | 1.27 |
| Vowel duration (periodic) : PC1 | -0.07 | 0.01 | -0.10 | -0.04 |
| Vowel duration (periodic) : Ultimate syllable | -0.58 | 0.09 | -0.76 | -0.40 |
| PC1 : Ultimate syllable | 0.43 | 0.08 | 0.26 | 0.59 |
| Vowel duration (periodic) : PC2 | 0.03 | 0.03 | -0.03 | 0.08 |
| Ultimate syllable : PC2 | -0.01 | 0.17 | -0.33 | 0.32 |
| PC1 : F0 | 0.04 | 0.04 | -0.03 | 0.12 |
| Ultimate syllable : F0 | -1.39 | 0.26 | -1.89 | -0.88 |
| PC2 : F0 | 0.03 | 0.07 | -0.11 | 0.17 |
| PC1 : Syllable duration | -0.09 | 0.03 | -0.15 | -0.03 |
| Ultimate syllable : Syllable duration | -0.93 | 0.16 | -1.25 | -0.62 |
| PC2 : Syllable duration | 0.04 | 0.06 | -0.08 | 0.15 |
| PC1 : F1 | -0.45 | 0.23 | -0.92 | -0.01 |
| Ultimate syllable : F1 | 16.90 | 2.55 | 12.10 | 22.21 |
| PC2 : F1 | -0.99 | 0.50 | -2.02 | -0.03 |
| PC1 : F2 | 0.06 | 0.04 | -0.03 | 0.15 |
| Ultimate syllable : F2 | -3.59 | 1.12 | -5.88 | -1.46 |
| PC2 : F2 | -0.10 | 0.08 | -0.26 | 0.06 |
| Vowel duration (periodic) : PC1 : Ultimate syllable | 0.06 | 0.02 | 0.02 | 0.09 |
| PC1 : Ultimate syllable : Syllable duration | 0.09 | 0.03 | 0.03 | 0.16 |

Table 3.8: Coefficients for the Bayesian multilevel model of data from the Catalan production task. The intercept represents the probability of stressed vowel productions in the penultimate syllable. To fit the table on one page, three-way interactions with a confidence interval spanning zero are not shown.

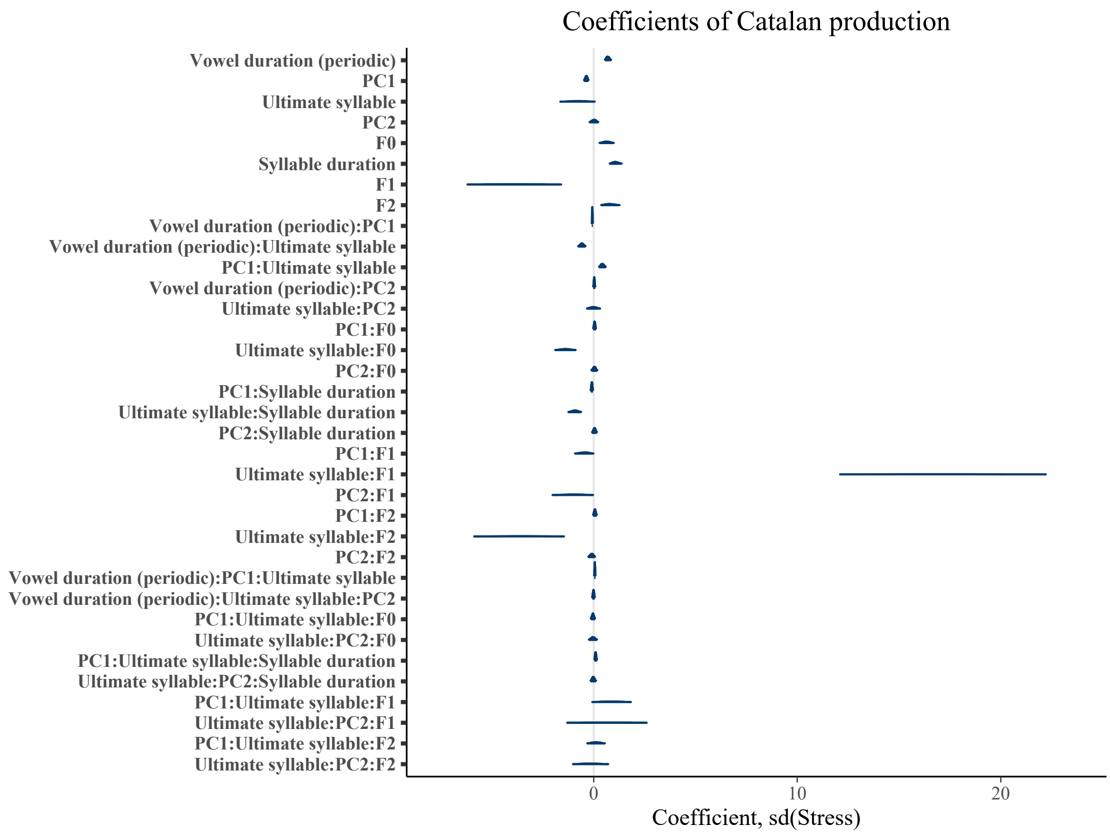



Figure 3.22: Coefficients from the Bayesian multilevel model of Catalan production. Coefficients with a distribution that does not include $x = 0$ are statistically distinct from 0 within a 95% confidence interval.

5.88]). This interaction is visualized in Figure 3.25. Additionally, the slope of the effect of normalized F2 in ultimate syllables (/a/) is negative, where increasing F2 (vowel fronting, i.e., centralization; see Figure 2.7 for reference) is correlated with a lower likelihood of stress (slope = -2.85, 95% CI = [-5.08, -0.90]).

The principal component interpreted as vowel duration and the degree of modal phonation interacts significantly with PC1 (English dominance relative to Spanish and Catalan dominance) and syllable. In the visualization of this interaction (Figure 3.26), cue-weightings for vowel duration and the degree of modal phonation are heavier in the penultimate syllable than in the ultimate syllables (β = 0.60, 95% CI = [0.43, 0.79]), however increasing English dominance relative to Spanish and Catalan dominance correlates to a decrease of cue-weighting for vowel duration and the degree of modal phonation in both syllables (β = -0.07 ± 0.01, 95% CI = [-0.10, -0.04]).

Similarly, syllable duration interacts with PC1 (English dominance relative to Spanish and Catalan dominance) and syllable. In the visualization of this interaction (Figure 3.27),

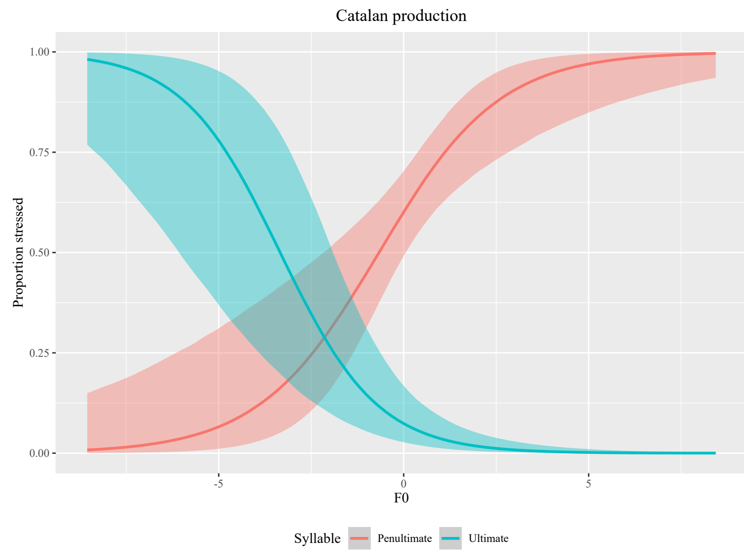



Figure 3.23: Predicted effects of F0 across syllable in Catalan production.

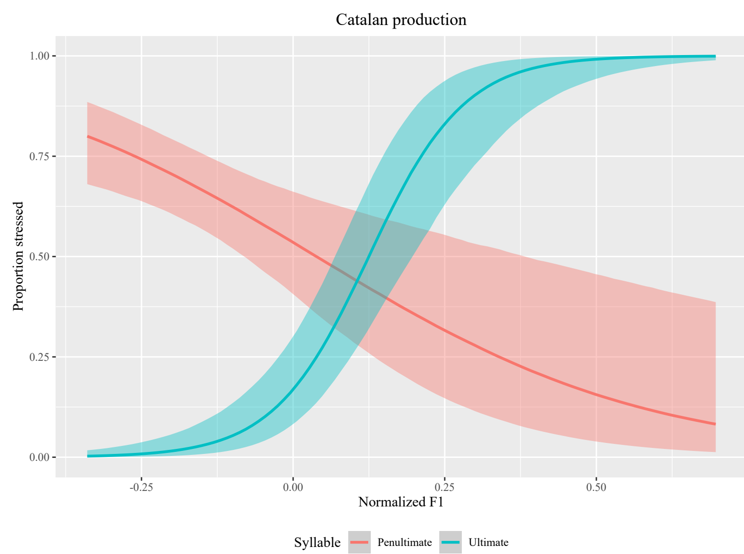



Figure 3.24: Predicted effects of normalized F1 across syllable in Catalan production.

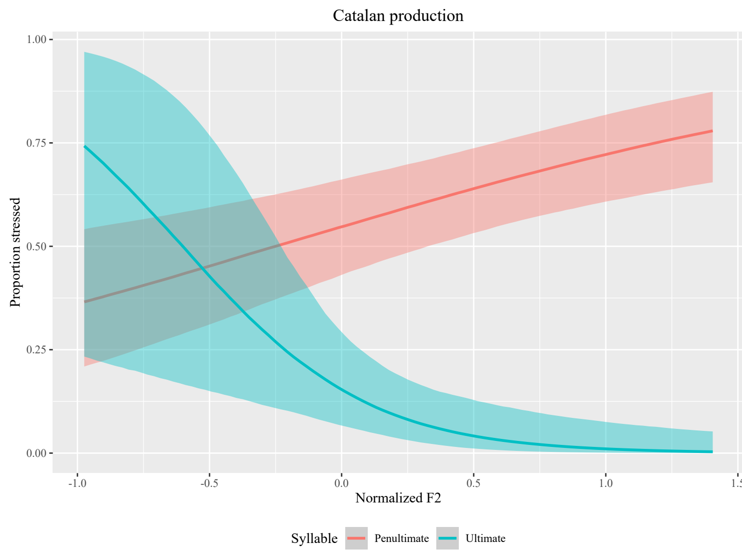



Figure 3.25: Predicted effects of normalized F2 across syllable in Catalan production.

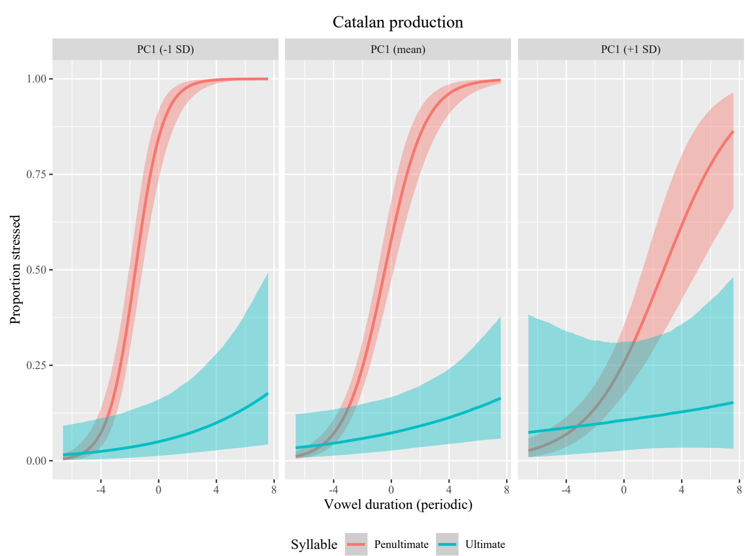



Figure 3.26: Predicted effects of duration (periodic) across syllable and PC1 in Catalan production.

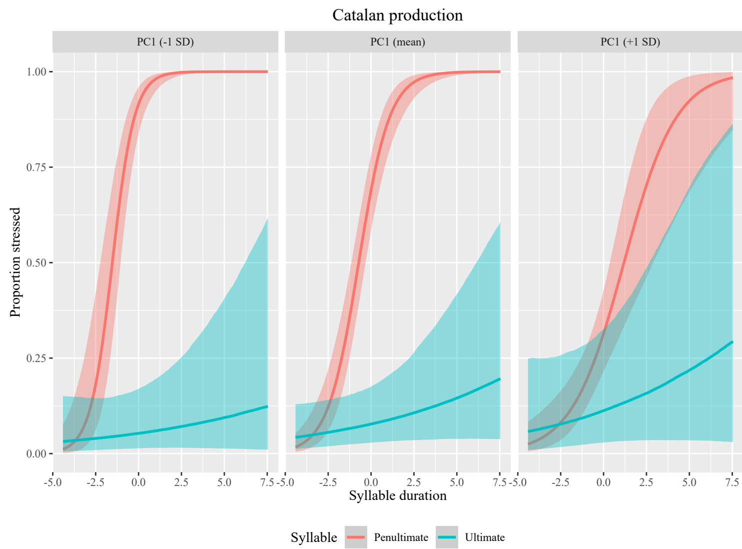



Figure 3.27: Predicted effects of syllable duration across syllable and PC1 in Catalan production.

cue-weightings for duration are heavier in the penultimate syllable than in the ultimate syllables (β = 0.96, 95% CI = [0.65, 1.29]), however increasing English dominance relative to Spanish and Catalan dominance correlates to a decrease of cue-weighting for duration in both syllables.

Lastly, the significant interaction between PC1 (English dominance relative to Spanish and Catalan dominance) and syllable is again a by-product of unequal token counts across speakers. According to the model, with an increase in PC1, the probability of stress is higher in ultimate than penultimate syllables (β = -0.367, 95% CI = [-0.591, -0.161]). Since all speakers read from the same list of words, in theory the log likelihood of stress in each syllable type (penultimate and ultimate) should be the same across all speakers. However, words which were improperly stressed were removed from analysis, resulting in a token imbalance. Speakers with higher PC1 (greater than 0) had 715 stressed penultimate syllables and 952 stressed ultimate syllables, whereas speakers with lower PC1 (less than 0) had 983 stressed penultimate syllables and 994 stressed ultimate syllables.

## 3.4 Discussion and conclusions

In this experiment, L1 English-L2 Spanish-L3 Catalan and early Spanish-Catalan bilinguals that learned English as an L3 (henceforth, L3 English trilinguals) produced sentences with stress minimal pairs in low pitch accents in English, Spanish, and Catalan. To observe potential regressive crosslinguistic influence, monolingual English speakers also completed the English sentence elicitation task. The main findings will now be summarized to address the research questions posed at the beginning of the chapter.

*RQ1*: How does speakers' relative experience with the three languages affect their cue-weighting in each language? Relative language dominance impacts cue-weightings for measures of duration in all three languages, but the directionality of impact differs across languages. In English, both vowel duration and syllable duration cue-weightings increase in the penultimate syllable as the amount of English dominance relative to Spanish and Catalan dominance increases (higher PC1). In Spanish and Catalan, however, higher PC1 corresponds to a *decrease* in cue-weightings for vowel and syllable duration in the penultimate syllables. Duration is the only cue that is consistently used across the three languages, but each language shows a different pattern of use based on syllable and whether vowel duration or syllable duration is specifically measured. Therefore, there are more opportunities for differences due to relative language dominance to emerge.

One possible issue with the duration cue-weightings observed is that the target words are phrase-final in English and phrase-medial in Spanish and Catalan. Perhaps there was more phrase-final syllable lengthening in English, thus obscuring the role of duration to cue stress shifts in the ultimate syllable. Had the target words been phrase-medial in all three languages, further conclusions could possibly be drawn regarding the role of relative language dominance and the use of duration cue-weightings across syllable in the three languages.

Apart from duration, the only other cue affected by relative language dominance is normalized F2 in English, where cue-weightings for F2 increase with increasing relative dominance in English, particularly in ultimate syllables, and with increasing dominance in Catalan relative to Spanish. Since F2 is a measure of the degree of vowel frontness/backness, it primarily captures the degree of centralization of mid vowels (i.e., those that move on the horizontal axis when centralizing) versus high or low vowels (i.e., those that move mainly on the vertical axis when centralizing). Whereas centralization can occur for both front and back mid vowels in English, centralization of mid vowels in unstressed syllables is not especially notable in Spanish (e.g., Quilis & Esgueva, 1983). In Catalan, only front mid vowels centralize in unstressed syllables, but back mid vowels do not (see Figure 2.7). Whereas the penultimate syllables almost exclusively have front mid vowels, the ultimate syllables have a lot more back mid vowels. Therefore, all speakers use F2 relatively similarly to cue stress shifts in penultimate vowels because centralization of front mid vowels occurs in English and Catalan (and to a lesser degree, in Spanish). Additionally, speakers with greater relative dominance in Catalan than Spanish have higher cue-weightings for F2 since vowel reduction occurs prescriptively in Catalan but not Spanish. However, because centralization of back mid vowels is unique to English, speakers with greater relative dominance in English utilize F2 to cue stress shifts in ultimate syllables more than speakers with less relative dominance in English.

*RQ2*: Do L1 English trilinguals use suprasegmental cues to a greater extent than English monolinguals when producing stress contrasts in English? I originally hypothesized that L1 English trilinguals would not pattern differently from monolingual speakers in terms of suprasegmental cues since the functional load of these cues in English is far less than that of vowel quality. Based on the results from the English production task, however, L1 English speakers with experience in the L2 and L3 do in fact use suprasegmental cues (duration) to a greater extent than English monolinguals. L1 English trilinguals have higher cue-weightings for syllable duration in penultimate syllables than monolingual speakers, and whereas L1 English trilinguals do not use this cue in ultimate syllables, monolingual speakers show a negative correlation between syllable duration and stress in ultimate syllables. This suggests that cue-weighting transfer can be regressive, where cue-weightings in subsequently acquired languages can impact cue-weightings in the L1. Additionally, cue-weighting transfer is not necessarily motivated by the functional load of a cue. Even though duration is not the most heavily weighted cue in English, it can still be susceptible to transfer and therefore increase in functional load. However, this does not occur unilaterally, but occurs feature-by-feature. For example, differences between cue-weightings for L1 English trilinguals and monolingual English speakers were only observed for syllable duration, not in other suprasegmental cues such as vowel duration and F0.

What do these results say about how cue-weighting transfer applies to L3 phonetics and phonology? A primary finding is that trilingual speakers with sufficient exposure to additional languages can use cues in a nativelike manner to cue lexical stress. For all trilingual speakers, normalized F1 and vowel duration were the most heavily weighted cues in English, whereas F0 and syllable duration were the heaviest-weighted cues in Spanish, and normalized F1 and syllable duration were the heaviest-weighted cues in Catalan. Therefore, trilinguals can make use of cues that are most informative in their additional languages. However, their relative language dominance can affect how these cues are used on a finer scale. For example, all speakers used duration to cue lexical stress shifts in the three languages, but the weight of duration across syllables differed based on language dominance. In English, speakers with greater Spanish and Catalan dominance relative to English dominance used duration in ultimate syllables to a greater degree than speakers with greater English dominance relative to Spanish and Catalan. In Spanish and Catalan, speakers with more English dominance relative to Spanish and Catalan dominance weighted duration more heavily in ultimate syllables than in penultimate syllables, where the reverse was true for speakers with greater Spanish and Catalan dominance relative to English dominance. Therefore, in studies of cue-weighting transfer, it is important to analyze cues on a more detailed level, such as how cues are weighted across syllables or different ways of signalling the same class of cue (e.g., syllable durations versus vowel durations). In the same vein, keeping the effects of F1 and F2 separate allowed for more information to be gathered, rather than combining the two measures into one measure of vowel peripherality. Secondly, the transfer of cue-weightings can be bidirectional, where cue-weightings in an L2 or L3 can affect cue-weightings in an L1 (regressive transfer). This finding underscores that all languages are active and available for multilinguals to use, and that it is important to observe production in all of a trilingual's languages, not just their L3.

Returning to the ranking of cue-weightings found in the literature, participants generally conformed to expectations. A measure of vowel quality (normalized F1) was the strongest cue in English (Campbell & Beckman, 1997; Chrabaszcz et al., 2014; Zhang & Francis, 2010), followed by a measure of duration (Beckman & Edwards, 1994; Campbell & Beckman, 1997). In Spanish, duration was the most prominent cue (Ortega-Llebaria & Prieto, 2011; Romanelli et al., 2018), but the relative weight of duration was not heavier for word-final vowels, as was found by Prieto and Ortega-Llebaria (2006). In Catalan, cue-weightings were vowel-specific, where duration was the most prominent cue for penultimate syllables (/i/), and a measure of vowel quality (normalized F1) was most prominent for ultimate syllables (/a/) (Ortega-Llebaria & Prieto, 2009). Spectral tilt as a correlate of articulatory effort did not surface in these data, but rather as a correlate of phonation type.

Another interesting finding from this task was the cue-weighting of F0 across syllable in each language. In English, higher F0 cued stress shifts in ultimate syllables, whereas the inverse relation was observed in penultimate syllables. However, in Spanish and Catalan, higher F0 cued stressed shifts in *penultimate* syllables and the reverse was true in ultimate syllables. One possible explanation is the difference in the sentences used to elicit lexical stress contrasts across the three languages. In English, target words appeared in post-focal position in a sentence of the form: "Not *X*, but *Y*," where the target words were often the last word of phrase *X*. In contrast, the Spanish and Catalan sentences had target words that always appeared phrase-medially in reporting clauses. Therefore, the difference in F0 cue-weightings across languages may be attributed to whether the target words appeared in phrase-final or phrase-medial position.

Beyond this crosslinguistic difference, it is interesting that F0 played a role at all, given that target words were elicited in low pitch accent conditions and F0 has generally been shown to cue pitch accent, rather than lexical stress (Beckman & Edwards, 1994, p. 13). Despite best efforts to control for pitch accent through the phrasal structure of the elicited sentences and the removal of incorrect productions, some participants may have deviated from this prominence structure.

Throughout the results section, when speaking of the principal component 'vowel duration (periodic)', vowel duration was mainly referenced, although the degree of modal phonation also covaries within this principal component. The strength of correlation between the raw measure of vowel duration and the vowel duration (periodic) principal component is very high (0.76), however, harmonic-amplitude measures are also present in this principal component with strong correlations. It is not clear whether creaky phonation is in and of itself a predictor of lexical stress shifts for these speakers, or whether creaky phonation covaries with short vowel durations, which alone cue lexical stress shifts. Impressionistically, creaky phonation was more common in the English production data than in the Spanish and Catalan production data, but with the current experimental design, the roles of modal phonation and vowel duration cannot fully be teased apart. A future perception study which isolates vowel duration from creaky phonation could provide more insight into the relative role of each acoustic measure in signalling stress shifts.

Given the strong patterns of vowel reduction in English and Catalan, it was not surprising that normalized F1 and F2 were heavily weighted in both languages. It is also not surprising that these cues signalled stress shifts in Spanish since phonetic reduction in unstressed syllables has been documented for both high and mid vowels (e.g., Quilis & Esgueva, 1983;

Romanelli et al., 2018; Ronquest, 2013). However, one might expect that relative dominance in English or in Catalan might mediate the cue-weightings for F1 or F2 in Spanish, should cue-weightings have been transferred from English or Catalan. The lack of an interaction with relative language dominance may suggest either that the amount of phonetic reduction in unstressed syllables is inherent to Spanish production, or that it is a result of crosslinguistic transfer since vowel reduction is present in both English and Catalan and all participants in the Spanish task speak both English and Catalan. A replication of this production task with monolingual Spanish speakers could provide more insight into this question and tease apart these possible explanations.

The next chapter introduces the task used to observe the perception of lexical stress shifts in each language by these trilingual speakers.

# Chapter 4

# Study 2: Perception of lexical stress

## 4.1 Introduction

In this experiment, trilinguals' lexical stress identifications in English, Spanish, and Catalan were obtained in three separate sessions. Though the stimuli for each language task have a similar phonological frame, language-specific phonetic cues were used for each language and the tasks were conducted in a language-specific environment to induce as much of a unilingual language mode as was possible. The stimuli were manipulated across three acoustic measures – duration, vowel quality, and spectral tilt – in order to address the following research questions:

*RQ1*: How does listeners' relative experience with the three languages affect their cue-weighting in each language?

In line with many representational models of bilingual phonetics (e.g., de Leeuw & Chang, in press; Flege & Bohn, 2021; Strange, 2011), I hypothesize that the effect of language experience can manifest in any of their languages, regardless of the order of acquisition. I predict that L1 English trilinguals will show great variability in their perception of L2 Spanish and L3 Catalan, due to their diversity of experience across the three languages. That is, their cue-weightings in their L2 and L3 will not be fixed, but will be vulnerable to the effects of language experience and relative language dominance.

*RQ2*: Do L1 English trilinguals use suprasegmental cues to a greater extent than English monolinguals when perceiving stress in English?

Because crosslinguistic influence can be bidirectional given the context (e.g., Birdsong, 2018; Cook, 2003; Grosjean, 1989), I hypothesize that increased experience with L2 Spanish and L3 Catalan can allow for transfer from the L2 and L3 of heavier cue-weightings for duration and spectral tilt to L1 English when these cues are more contextually informative, such as when listeners can no longer use vowel quality as a cue to stress.

The sections that follow describe the perception study designed to provide insight to these research questions, the speakers that participated in this study, and the results.

## 4.2 Methods

### Participants

The same participants that are described in Section 3.2 also participated in the perception experiment. During each scheduled language session, participants completed the production task, followed by the perception task.

### Materials

In each perception task, participants were instructed to listen carefully to nonce words produced in isolation, and for each one, to indicate the corresponding orthographic representation from the two choices provided (paroxytone or oxytone). Because listeners do not attend to only one acoustic cue, but rather they attend simultaneously to multiple cues, weighting the cues according to context (Chrabaszcz et al., 2014; Ortega-Llebaria et al., 2013), the voiced segments from each syllable in each token were manipulated so that each segment varied by two dimensions at a time. In this type of repetitive task, it is likely that listeners are engaged in signal detection, rather than lexical retrieval, so nonwords were used in each language session. This also serves to minimize lexical interference and focus on phonological and phonetic knowledge. Additionally, the use of real stress minimal pairs in each language would limit the comparability of results crosslinguistically and would introduce lexical frequency differentials between the two words (e.g., 1SG present *determino* 'I determine' may have a higher frequency than 3SG preterit *determinó* '(s)he determined'). As Gonzales and Lotto (2013) demonstrated, using language-specific phonetic cues in stimuli (e.g., use of English [ɹ] or Spanish [ɾ] in nonword continuum *bafri—pafri*), despite a differing language context in the experiment, can induce language-specific processing. Accordingly, the same nonword stress pairs were used in each language, with minor adaptations to account for phonotactics and orthographic norms in each language (English: *pondiss* /ˈpɑndɪs/ – *pondiss* /pənˈdɪs/; Spanish: *pondis* /ˈpondis/ – *pondís* /ponˈdis/; Catalan: *pondis* /ˈpondis/ – *pondís* /punˈdis/).

Ortín and Simonet (2022) caution against perception tasks in which participants must identify the stressed syllable in a word, arguing that these tasks measure phonological awareness rather than perception. They provide an example of participants who consistently can distinguish between *caso* [ˈkaso] 'case' and *casó* [kaˈso] '(s)he married' and yet cannot identify the contrastive phonological feature. However, given the increased metalinguistic knowledge of trilinguals relative to second language learners and the extensive amount of L2 classroom experience of the trilinguals that participated in this task (7.8 ± 5.4 years of L2 Spanish), the trilingual listeners studied here are assumed to have sufficient phonological knowledge to allow for this task to serve as a measure of perception, rather than phonological awareness.[1]

[1]In fact, following the Catalan perception task, many participants commented that they knew they were supposed to be listening for changes in vowel quality in the first syllable to help with their word identification judgments.

| Language | Stress | Sentence |
|---|---|---|
| English | paroxytone | No, I didn't BUY **pon**diss today, but I made it. |
| | oxytone | No, SHE won't pon**diss** today, but I will. |
| Spanish | paroxytone | "Lee estos artículos" − dice el pondis bajo. |
| | oxytone | "Lee estos artículos" − dice el pondís bajo. |
| Catalan | paroxytone | "Llegiu aquests articles" − diu el pondis baix. |
| | oxytone | "Llegiu aquests articles" − diu el pondís baix. |

Table 4.1: Sentences used to elicit endpoints for perception stimuli.

To create the endpoint stimuli, a male native English speaker, a male native Spanish speaker, and a female native Catalan speaker produced the target stimuli in unaccented positions in English, Spanish, and Catalan, respectively. For Spanish and Catalan, the target words appeared in reporting clauses, and for English, the target words appeared in post-focal positions. In the English sentences, bolded and underlined letters were used to signal which syllable to stress in each instance of the nonword. Capital letters were used to make clear where the prosodic nucleus of the sentence was. In the Spanish and Catalan sentences, speakers were not given additional indications of which syllables to stress since lexical stress is orthographically transparent in both languages. Table 4.1 shows the sentences used in each language to elicit the target stimuli.

One clean recording of each token was selected as the base for acoustic manipulation and the tokens were visually checked for formant and pitch tracking errors in Praat (Boersma & Weenink, 2019). The overall intensity of each token was scaled to 73 dB. Duration, average F1, average F2, average F3, average pitch, average intensity, the average intensity of the frequency band that included F0 and F1 (Catalan: 0−700 Hz; Spanish: 0−500 Hz; English: 0−900 Hz), and the average intensity of the frequency band that included formants beginning with F2 (Catalan: above 700 Hz; Spanish: above 500 Hz; English: above 900 Hz) were obtained from the voiced portion of each syllable (Table 4.2). Auditory stimuli were created to vary between initial stress (step 1) and final stress (step 5) across three dimensions: duration, vowel quality, and spectral tilt.

In each language, three $5 \times 5$ matrices were created where two dimensions were simultaneously manipulated from step 1 to step 5 in equidistant steps while the third was neutralized (following the procedure in Tremblay et al., 2021). The three matrices are: vowel quality and spectral tilt; vowel quality and duration; spectral tilt and duration. To produce the manipulations across vowel quality, a five-step continuum between the stress-initial and stress-final endpoints was generated using TANDEM-STRAIGHT (Kawahara et al., 2008) − a graphical user interface that generates continua nonparametrically, meaning that steps along the continuum do not differ across only one acoustic metric (e.g., F1), rather, the spectra of the endpoint tokens are manipulated in their entirety. TANDEM-STRAIGHT has been used to produce more natural-sounding stimuli (D. Kim et al., 2018), and resulting

| | | PONdis(s) | | ponDIS(S) | |
|---|---|---|---|---|---|
| Language | Acoustic Measure | First vowel | Second vowel | First vowel | Second vowel |
| English | Duration (ms) | 123.18 | 73.21 | 34.73 | 70.55 |
| | F0 (Hz) | 137.76 | 119.19 | 148.52 | 128.48 |
| | F1 (Hz) | 788.08 | 470.05 | 422.66 | 462.41 |
| | F2 (Hz) | 1268.37 | 1685.29 | 1489.45 | 1789.10 |
| | F3 (Hz) | 2516.64 | 2638.98 | 2675.71 | 2703.04 |
| | Intensity (dB) | 77.66 | 70.71 | 72.80 | 72.87 |
| | Intensity 0−900 Hz (dB) | 75.94 | 73.36 | 78.07 | 74.40 |
| | Intensity 900+ Hz (dB) | 75.46 | 59.17 | 65.06 | 72.34 |
| Spanish | Duration (ms) | 49.14 | 72.12 | 53.8 | 83.29 |
| | F0 (Hz) | 134.94 | 131.59 | 135.65 | 138.12 |
| | F1 (Hz) | 436.36 | 311.92 | 395.73 | 277.45 |
| | F2 (Hz) | 883.19 | 1961.01 | 1027.01 | 2136.02 |
| | F3 (Hz) | 2469.64 | 2465.94 | 2511.05 | 2506.38 |
| | Intensity (dB) | 77.93 | 74.14 | 77.91 | 75.22 |
| | Intensity 0−500 Hz (dB) | 80.50 | 77.71 | 80.87 | 79.10 |
| | Intensity 500+ Hz (dB) | 72.99 | 62.06 | 70.55 | 58.73 |
| Catalan | Duration (ms) | 90.41 | 95.63 | 54.45 | 118.73 |
| | F0 (Hz) | 171.37 | 162.70 | 173.71 | 164.53 |
| | F1 (Hz) | 615.77 | 293.86 | 379.79 | 311.21 |
| | F2 (Hz) | 1278.85 | 2511.13 | 1074.59 | 2479.54 |
| | F3 (Hz) | 3161.08 | 2843.71 | 2740.01 | 2755.08 |
| | Intensity (dB) | 76.12 | 73.55 | 74.64 | 74.27 |
| | Intensity 0−700 Hz (dB) | 78.80 | 77.11 | 83.92 | 84.34 |
| | Intensity 700+ Hz (dB) | 74.89 | 47.66 | 65.86 | 65.03 |

Table 4.2: Acoustic measures for vowels in endpoint words.

stimuli were in fact judged as sounding more natural by native speakers than an earlier set of stimuli produced with scripts in Praat. Figure B.1 shows the anchoring interface in TANDEM-STRAIGHT, in which phonologically similar segments in the two endpoints (e.g., the first vowel in each word) were anchored in preparation for nonparametric continua generation. Afterwards, pitch contours and vowel durations were neutralized at each step, using the 'Manipulate' function in Praat. The resulting formant tracks are shown in Figures B.2 to B.4.

To produce the vowel quality by spectral tilt matrix, intensities for the lower and higher frequency ranges were linearly interpolated from the endpoint measures in Table 4.2. From the formant continua that were neutralized for pitch and duration, spectral tilt was manipulated. First, the vowels were extracted and a pass Hann band filter was used to separately obtain the lower and higher frequency ranges. The intensity of the low and high bands was manually scaled using the scale intensity function in Praat. The bands were combined to stereo then converted back to mono and spliced into the stimulus frame.

To produce the spectral tilt by duration matrix, the step 3 stimuli on the vowel quality continua from the vowel quality by spectral tilt matrix were used as the base stimuli, since vowel quality was neutralized at this step. For each spectral tilt step, five different duration ratios were applied to the vowels with the manipulation feature using the PSOLA algorithm in Praat (Figure B.5; D. Kim et al., 2018). The durations of each vowel at each step on the duration continuum in each language can be seen in Figure B.6.

To produce the vowel quality by duration matrix, the step 3 stimuli on the spectral tilt continua from the vowel quality by spectral tilt matrix were used as the base stimuli, since spectral tilt was neutralized at this step. For each vowel quality step, five different duration ratios were applied to the vowels with the manipulation feature using the PSOLA algorithm in Praat (D. Kim et al., 2018).

Stimuli from the three matrices (25 x 3) were presented in three blocks, once per block, for a total of 225 trials in each of the three language tasks. The English perception task additionally included six practice trials to train participants to associate capital letters with stressed syllables, as detailed below.

## Procedure

As described in Section 3.2, separate sessions were scheduled in Catalan, English, and Spanish over Zoom (in that order), with at least 48 hours between each scheduled session. Each session was conducted completely in the target language by a native speaker. For each experimental session, participants completed the production task (described in Chapter 3), followed by a perception task (described in this chapter). The perception task was hosted on a remote server online and participants were provided a URL to complete the task during each Zoom call. Participants were told that the use of headphones was required in order to complete the perception experiment. In the English task, participants were trained in practice trials to associate capital letters with stressed syllables. They were instructed to categorize audio clips as 'CONvict' or 'conVICT' and were provided with feedback, as shown

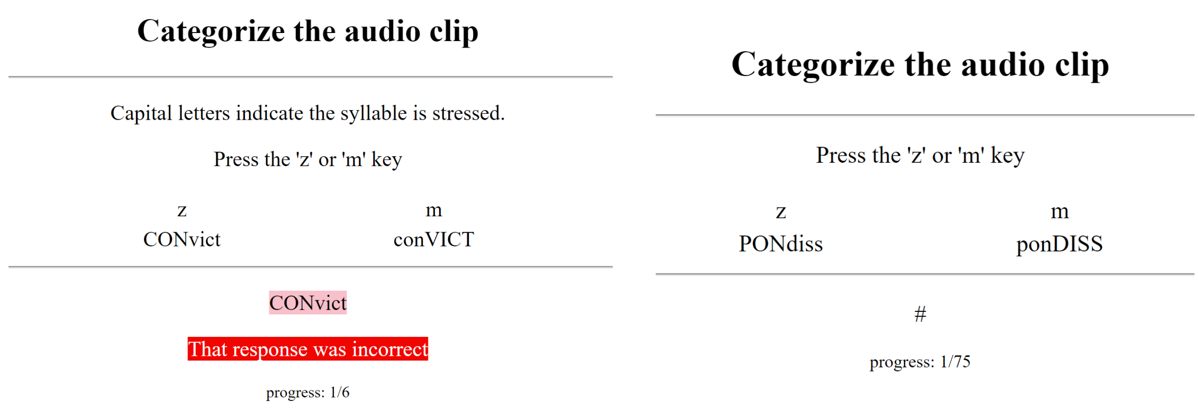



Figure 4.1: *Left:* Corrective feedback for English practice trials; *Right:* Main trial of English perception experiment.

in Figure 4.1 (*Left*). Next, participants completed the word identification task, with stimuli appearing in three blocks separated by short breaks. Each stimulus from the three matrices appeared in each block in a randomized order. For each trial, participants heard an audio clip and were told to categorize the word as one of the two options printed on the screen using the 'z' and 'm' keys on their keyboard (English: 'PONdiss' or 'ponDISS'; Spanish/Catalan: 'pondis' or 'pondís'). Responses that were faster than 200 ms or slower than 4000 ms triggered a warning message. The interface for the main trials in English is shown in Figure 4.1 (*Right*).

## Analysis

For all three languages, paroxytone (PONdis) responses were coded as '0' and oxytone (ponDIS) responses were coded as '1'. Excessively quick responses ($< 200$ ms) and excessively long responses ($> 2500$ ms) were removed from analysis (English task: $n = 443$, Spanish task: $n = 168$, Catalan task: $n = 409$) (Reinisch & Holt, 2014).[2] The steps on the continuum for each acoustic measure were centered on zero and jitter of 0.1% was added to each step to stabilize the data for Bayesian analysis.[3] Responses from the first block of each experiment were removed from the data set since participants reported hearing no differences between stimuli until the second block. After this adjustment, 10,217 English judgments, 7,278 Spanish judgments, and 7,260 Catalan judgments were submitted to statistical analysis. For the trilingual perception data in each language, a Bayesian multilevel

[2]Reaction times were only used as inclusion criteria, but further analysis of reaction times was outside of the scope of the present study.

[3]Without this small amount of jitter added to each step size, the ordinal values of step manipulations resulted in a problematic geometry of the model posterior.

model was fitted in R with Equation (4.1) using `brms` (Bürkner, 2017). The interactions between the three acoustic measures and each BLP component model how participants with different relative language dominance weight acoustic measures in isolation or in tandem when making stress-based word identity judgments. The by-participant random slopes allow for individual variation in perceptual bias towards one endpoint word or the other, as well as individual variation in cue-weighting. To examine how language profile affects English perception, Equation (4.2) was used in a separate Bayesian multilevel model with a binned factor of language profile (with levels of 'Monolingual', 'L1 English', and 'L3 English'). The interaction between each acoustic measure and language profile models how participants with different language profiles weight acoustic measures in isolation or in tandem when making stress-based word identity judgments. The by-participant random slopes allow for individual variation in perceptual bias towards one endpoint word or the other, as well as individual variation in cue-weighting. In all models, positive estimates are interpreted as increasing probability of ponDIS (oxytone) responses.

$$\begin{aligned} \text{ultimate stress} \sim\ & \text{duration} * \text{vowel quality} * \text{spectral tilt} * \text{PC1} + \\ & \text{duration} * \text{vowel quality} * \text{spectral tilt} * \text{PC2} + \\ & (\text{duration} * \text{vowel quality} * \text{spectral tilt} \mid \text{participant}) \end{aligned} \tag{4.1}$$

$$\begin{aligned} \text{ultimate stress} \sim\ & \text{duration} * \text{vowel quality} * \text{spectral tilt} * \text{language profile} + \\ & (\text{duration} * \text{vowel quality} * \text{spectral tilt} \mid \text{participant}) \end{aligned} \tag{4.2}$$

## 4.3 Results

### English perception compared to monolinguals

Perception data from the English task completed by monolingual and trilingual listeners were isolated and submitted to a multilevel Bayesian model with a four-way interaction between the three acoustic measures and language profile, and by-participant random slopes of the three-way interaction of the acoustic measures (see Equation (4.2) for model formula). The logodds estimates from this model are shown in Table 4.3.

As shown in the visual representation of logodds estimates of oxytone ponDIS responses in Figure 4.2, duration is the most heavily weighted cue across all listeners ($\beta = 1.02 \pm 0.19$, 95% CI = [0.66, 1.40]), followed by vowel quality ($\beta = 0.35 \pm 0.13$, 95% CI = [0.09, 0.62]), then spectral tilt ($\beta = 0.20 \pm 0.06$, 95% CI = [0.08, 0.31]). A post-hoc test of the interaction of vowel quality and language profile shows that monolingual listeners have heavier cue-weightings for vowel quality than L3 English trilinguals ($\beta = -0.60$, 95% CI = [-0.97, -0.20]) and L1 English trilinguals ($\beta = -0.39$, 95% CI = [-0.77, -0.01]). Differences in cue-weightings for vowel quality across language profile are shown in Figure 4.3.

| | Estimate | Error | L-95% CI | U-95% CI |
|---|---|---|---|---|
| (Intercept) | -0.83 | 0.29 | -1.40 | -0.26 |
| Vowel quality | 0.35 | 0.13 | 0.09 | 0.62 |
| Spectral tilt | 0.20 | 0.06 | 0.08 | 0.31 |
| Duration | 1.02 | 0.19 | 0.66 | 1.40 |
| L3 English | 0.53 | 0.42 | -0.31 | 1.32 |
| Monolingual | 0.81 | 0.43 | -0.04 | 1.67 |
| Vowel quality : Spectral tilt | 0.06 | 0.04 | -0.02 | 0.14 |
| Vowel quality : Duration | -0.03 | 0.06 | -0.14 | 0.09 |
| Spectral tilt : Duration | -0.08 | 0.05 | -0.18 | 0.02 |
| Vowel quality : L3 English | -0.21 | 0.19 | -0.57 | 0.16 |
| Vowel quality : Monolingual | 0.39 | 0.20 | 0.00 | 0.79 |
| Spectral tilt : L3 English | 0.01 | 0.08 | -0.14 | 0.17 |
| Spectral tilt : Monolingual | -0.01 | 0.08 | -0.17 | 0.15 |
| Duration : L3 English | 0.12 | 0.27 | -0.39 | 0.65 |
| Duration : Monolingual | 0.05 | 0.28 | -0.51 | 0.60 |
| Vowel quality : Spectral tilt : Duration | -107.32 | 44.97 | -195.88 | -19.06 |
| Vowel quality : Spectral tilt : L3 English | -0.02 | 0.05 | -0.12 | 0.09 |
| Vowel quality : Spectral tilt : Monolingual | -0.12 | 0.06 | -0.24 | -0.01 |
| Vowel quality : Duration : L3 English | 0.16 | 0.08 | 0.01 | 0.32 |
| Vowel quality : Duration : Monolingual | -0.07 | 0.08 | -0.23 | 0.08 |
| Spectral tilt : Duration : L3 English | 0.05 | 0.07 | -0.09 | 0.19 |
| Spectral tilt : Duration : Monolingual | -0.05 | 0.07 | -0.19 | 0.09 |
| Vowel quality : Spectral tilt : Duration : L3 English | 196.63 | 60.82 | 76.31 | 315.85 |
| Vowel quality : Spectral tilt : Duration : Monolingual | 108.07 | 63.70 | -15.99 | 232.13 |

Table 4.3: Coefficients for the Bayesian multilevel model of data from the English perception task, including responses from monolingual listeners. The intercept represents the logodds ratio of ponDIS (oxytone) responses with neutralized vowel quality, duration, and spectral tilt by L1 English trilinguals.

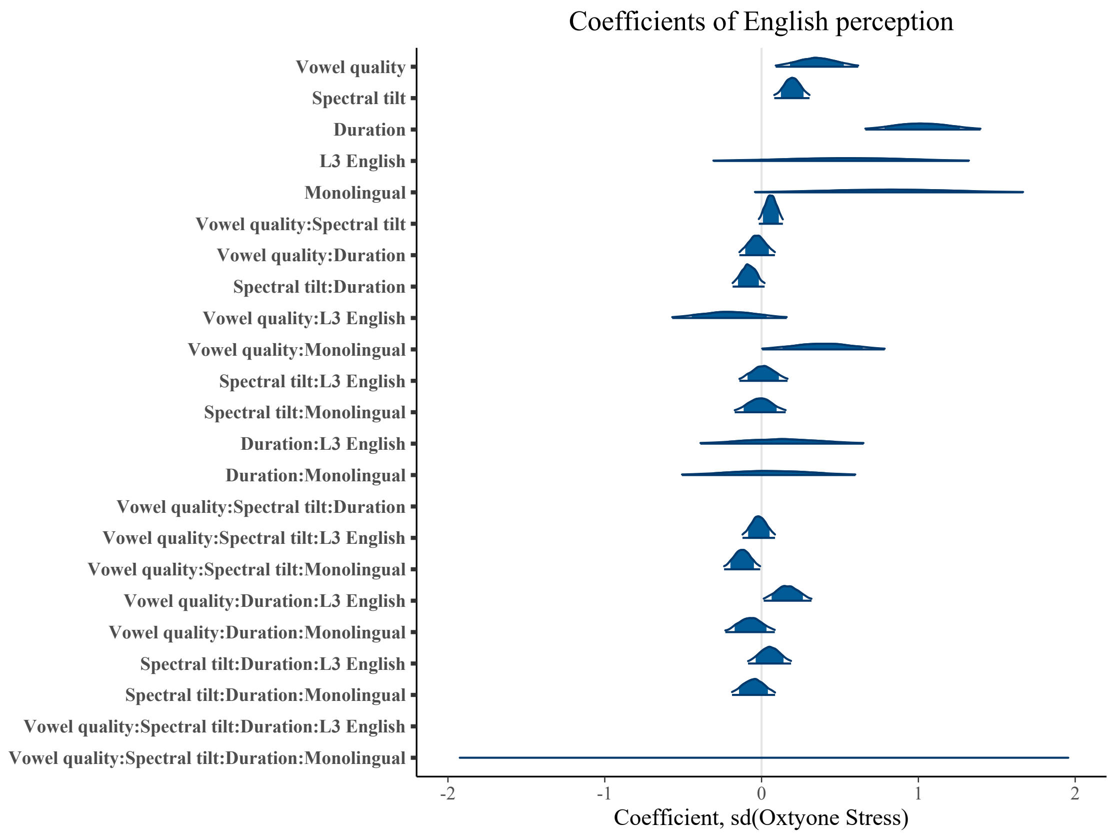



Figure 4.2: Coefficients from the Bayesian multilevel model of English perception with trilingual and monolingual data. Coefficients with a distribution that does not include $x = 0$ are statistically distinct from 0 within a 95% confidence interval.

Figure 4.4 shows that L3 English listeners attend to duration and vowel quality differently than L1 English trilinguals (β = 0.16 ± 0.08, 95% CI = [0.01, 0.32]). Whereas L3 English trilingual listeners weight vowel quality fairly equally across the steps of the duration continuum, L1 English trilingual listeners have heavier cue-weightings for vowel quality when the duration manipulations are on the ponDIS endpoint of the continuum (the endpoint where vowel reduction prescriptively occurs). Therefore, it seems that L1 English listeners' cue-weightings of vowel quality covary with duration, whereas L3 English listeners' cue-weightings of vowel quality are independent from duration. Figure 4.5 shows another visualization of the relationship between duration and vowel quality cues, where cue-weightings of vowel quality are represented across the vertical axis and cue-weightings of duration are represented across the horizontal axis. For L3 English listeners, there is no clear gradient across vowel quality steps, indicating that this cue is weighted relatively minimally. However, for L1 English listeners, there is a gradient for vowel quality which is more clear at higher steps on the duration continuum.

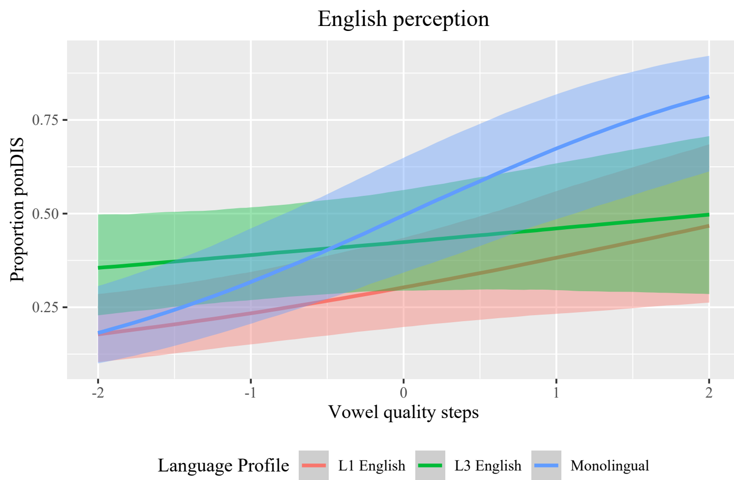



Figure 4.3: Predicted effects of vowel quality cue-weighting across language profile in English perception.

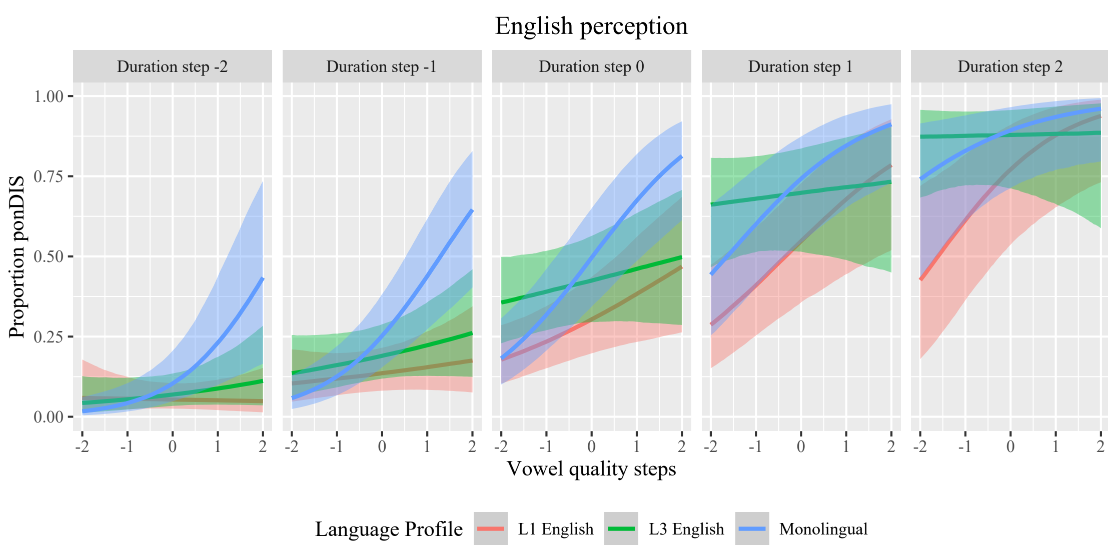



Figure 4.4: Predicted effects of vowel quality and duration cue-weighting across language profile in English perception.

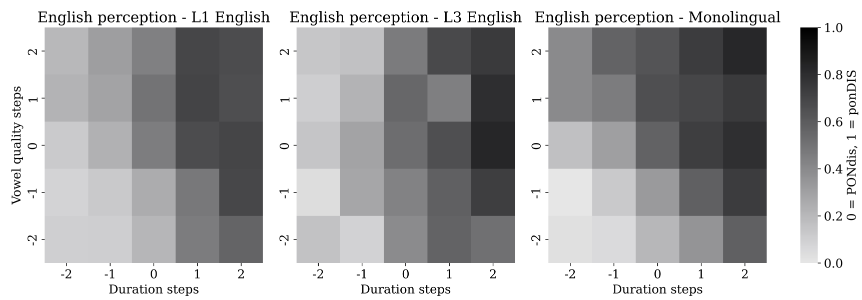



Figure 4.5: Proportion of 'ponDISS' responses in the English task across vowel quality and duration.

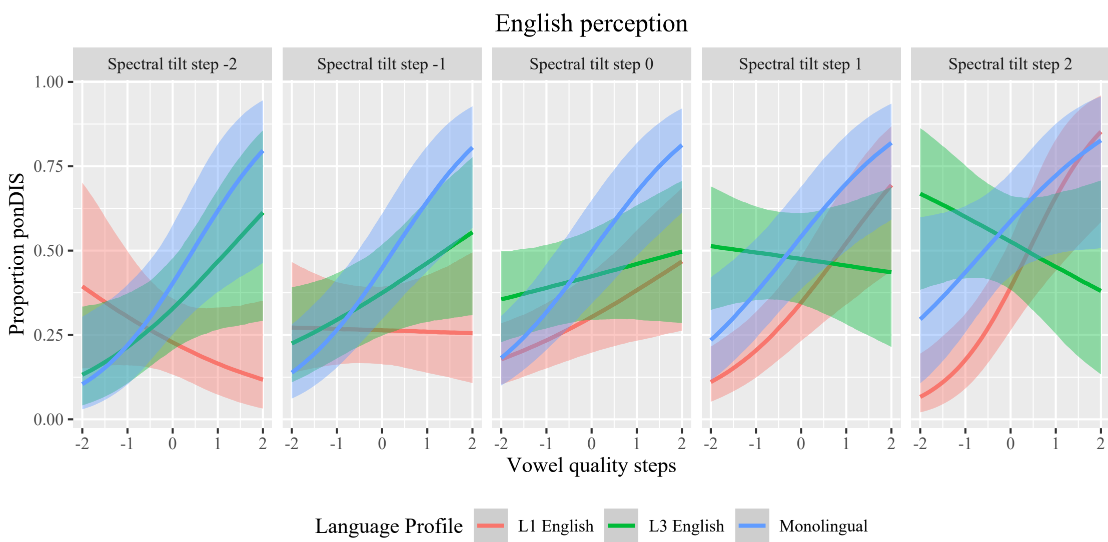



Figure 4.6: Predicted effects of vowel quality and spectral tilt cue-weighting across language profile in English perception.

For simultaneous manipulations of vowel quality and spectral tilt, monolingual listeners maintain similar cue-weightings for vowel quality regardless of spectral tilt information, whereas L1 English trilinguals have low cue-weightings for vowel quality when spectral tilt cues favor the PONdis endpoint and higher cue-weightings for vowel quality when spectral tilt cues favor the ponDIS endpoint (the endpoint where vowel reduction prescriptively occurs; $\beta$ = -0.12 ± 0.06, 95% CI = [-0.24, -0.01]). This covarying relationship between vowel quality and spectral tilt can be seen in Figure 4.6. Figure 4.7 also shows this trend where cue-weightings of spectral tilt are represented on the horizontal axis and cue-weightings for vowel quality are represented across the vertical axis. The heatmap for monolingual listeners has a fairly consistent gradient for vowel quality across all steps on the spectral tilt continuum, whereas the L1 English trilinguals have a stronger gradient for vowel quality at higher steps on the spectral tilt continuum. In comparison, there are no clear vowel quality or spectral tilt gradients seen among the L3 English trilinguals. Duration, however, is at-

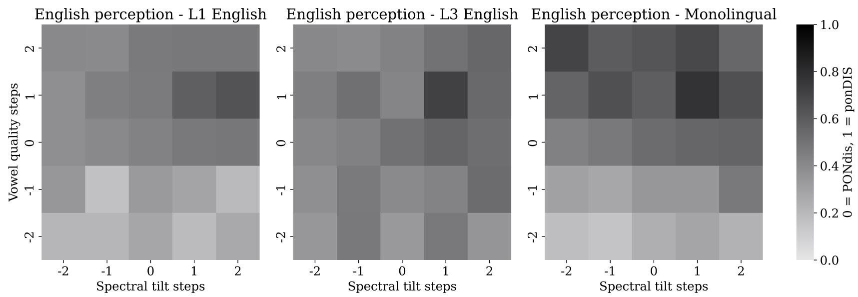



Figure 4.7: Proportion of 'ponDISS' responses in the English task across vowel quality and spectral tilt.

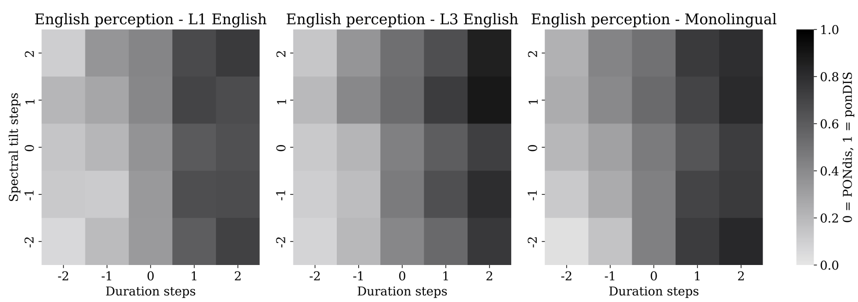



Figure 4.8: Proportion of 'ponDISS' responses in the English task across duration and spectral tilt.

tended to independently of spectral tilt for all listeners, as can be seen in Figure 4.8, where the gradient for duration on the x-axis is fairly consistent for all listener profiles across all spectral tilt steps.

## English perception by trilinguals

Perception data from the English task completed by trilingual listeners were isolated and submitted to a multilevel Bayesian model with a four-way interaction between the three acoustic measures and PC1, a four-way interaction between the three acoustic measures and PC2, and by-participant random slopes of the three-way interaction of the acoustic measures (see Equation (4.1) for model formula). The logodds estimates from this model are shown in Table 4.4.

| | Estimate | Error | L-95% CI | U-95% CI |
|---|---|---|---|---|
| (Intercept) | -0.59 | 0.23 | -1.05 | -0.13 |
| Vowel quality | 0.25 | 0.09 | 0.08 | 0.43 |
| Spectral tilt | 0.20 | 0.04 | 0.12 | 0.28 |
| Duration | 1.09 | 0.15 | 0.80 | 1.39 |
| PC1 | -0.06 | 0.05 | -0.16 | 0.03 |
| PC2 | 0.14 | 0.09 | -0.04 | 0.32 |
| Vowel quality : Spectral tilt | 0.05 | 0.03 | -0.00 | 0.11 |
| Vowel quality : Duration | 0.06 | 0.04 | -0.03 | 0.15 |
| Spectral tilt : Duration | -0.05 | 0.04 | -0.12 | 0.02 |
| Vowel quality : PC1 | 0.02 | 0.02 | -0.01 | 0.06 |
| Spectral tilt : PC1 | -0.00 | 0.01 | -0.02 | 0.01 |
| Duration : PC1 | -0.01 | 0.03 | -0.07 | 0.04 |
| Vowel quality : PC2 | 0.04 | 0.03 | -0.03 | 0.10 |
| Spectral tilt : PC2 | 0.01 | 0.02 | -0.03 | 0.03 |
| Duration : PC2 | -0.01 | 0.06 | -0.12 | 0.10 |
| Vowel quality : Spectral tilt : Duration | -32.18 | 30.70 | -92.44 | 28.50 |
| Vowel quality : Spectral tilt : PC1 | 0.00 | 0.01 | -0.01 | 0.01 |
| Vowel quality : Duration : PC1 | -0.01 | 0.01 | -0.03 | 0.00 |
| Spectral tilt : Duration : PC1 | -0.01 | 0.01 | -0.02 | 0.01 |
| Vowel quality : Spectral tilt : PC2 | 0.02 | 0.01 | -0.00 | 0.04 |
| Vowel quality : Duration : PC2 | -0.01 | 0.02 | -0.04 | 0.03 |
| Spectral tilt : Duration : PC2 | -0.02 | 0.01 | -0.04 | 0.01 |
| Vowel quality : Duration : PC1 | 3.11 | 5.98 | -8.90 | 14.77 |
| Vowel quality : Spectral tilt : Duration : PC2 | 10.41 | 10.73 | -10.33 | 31.65 |

Table 4.4: Coefficients for the Bayesian multilevel model of data from the English perception task. The intercept represents the logodds ratio of ponDIS (oxytone) responses with neutralized vowel quality, duration, and spectral tilt.

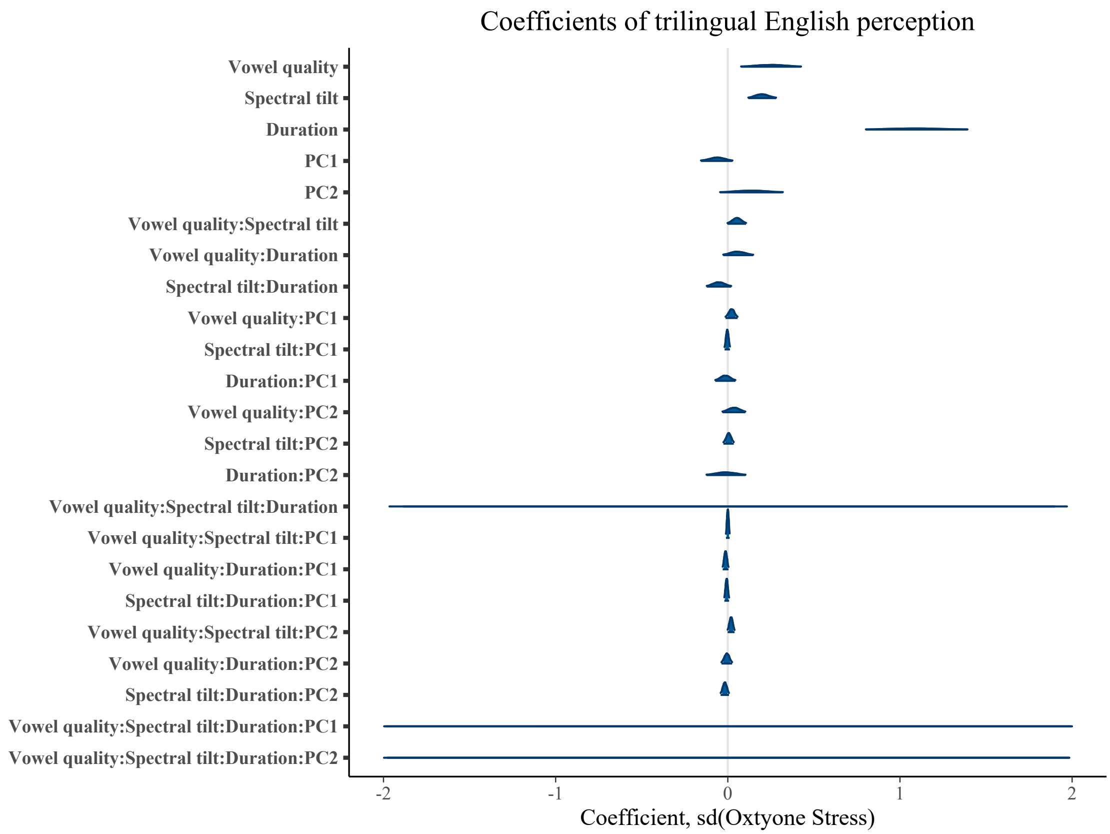



Figure 4.9: Coefficients from the Bayesian multilevel model of English perception. Coefficients with a distribution that does not include $x = 0$ are statistically distinct from 0 within a 95% confidence interval.

As shown in the visual representation of logodds estimates of oxytone ponDIS responses in Figure 4.9, duration is the most heavily weighted cue across all trilinguals ($\beta = 1.09 \pm 0.15$, 95% CI = [0.80, 1.39]), followed by vowel quality ($\beta = 0.25 \pm 0.09$, 95% CI = [0.08, 0.43]), then spectral tilt ($\beta = 0.20 \pm 0.04$, 95% CI = [0.12, 0.28]). These cue-weightings are not mediated by relative language dominance (PC1 or PC2). As the heatmaps from the previous section show, for the trilingual listeners, word identification boundaries are sharpest when duration is present as a cue (Figures 4.5 and 4.8), but the perceptual boundary is less distinct when duration is neutralized (Figure 4.7).

## Spanish perception

Perception data from the Spanish task completed by trilingual listeners were isolated and submitted to a multilevel Bayesian model with a four-way interaction between the three acoustic measures and PC1, a four-way interaction between the three acoustic measures and

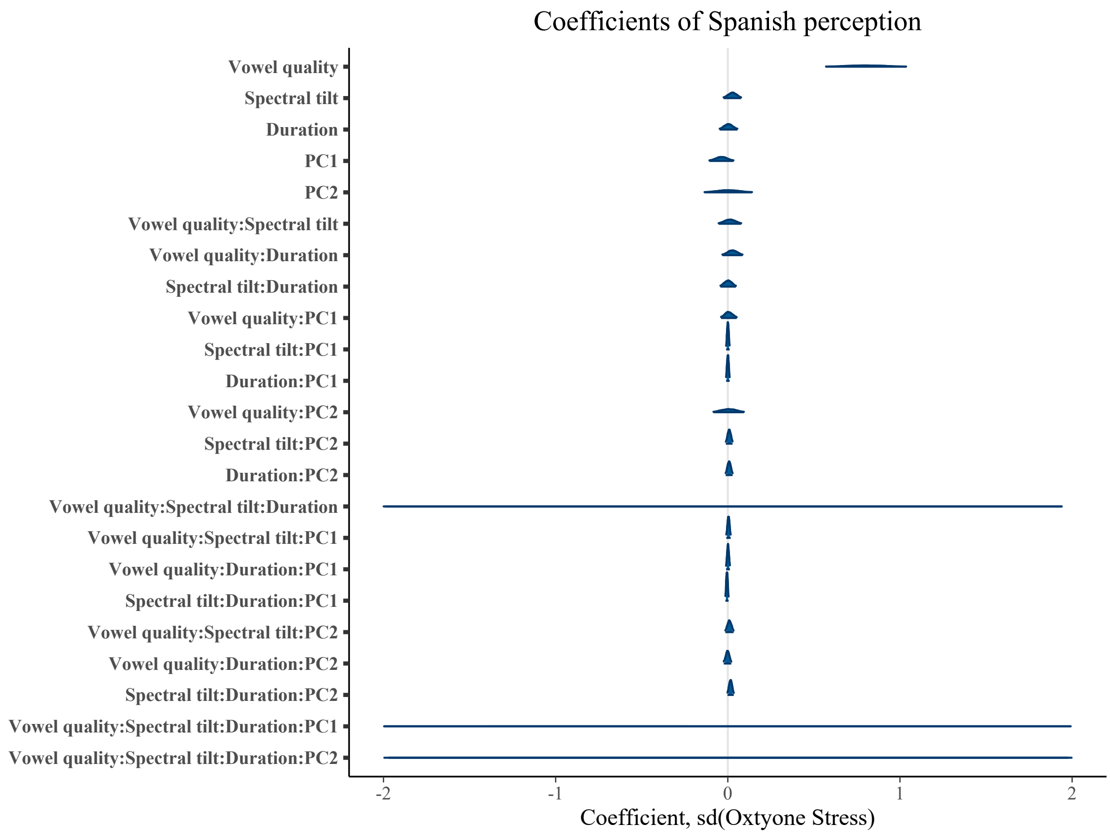



Figure 4.10: Coefficients from the Bayesian multilevel model of Spanish perception. Coefficients with a distribution that does not include $x = 0$ are statistically distinct from 0 within a 95% confidence interval.

PC2, and by-participant random slopes of the three-way interaction of the acoustic measures (see Equation (4.1) for model formula). The logodds estimates from this model are shown in Table 4.5. For the purposes of visualization, word identifications produced by L1 English and L3 English trilinguals across the three acoustic manipulation matrices were grouped and plotted into heatmaps (Figures B.7 to B.9).

As shown in the visual representation of logodds estimates of oxytone ponDIS responses in Figure 4.10, vowel quality is the most heavily weighted cue across all trilinguals ($\beta = 0.80 \pm 0.12$, 95% CI = [0.57, 1.04]). In fact, this cue is exclusively used, and durational cues and spectral tilt cues are not attended to. Neither PC1 nor PC2 mediate cue-weighting for vowel quality.

| | Estimate | Error | L-95% CI | U-95% CI |
|---|---|---|---|---|
| (Intercept) | 0.04 | 0.17 | -0.29 | 0.39 |
| Vowel quality | 0.80 | 0.12 | 0.57 | 1.04 |
| Spectral tilt | 0.03 | 0.03 | -0.02 | 0.08 |
| Duration | 0.00 | 0.03 | -0.05 | 0.06 |
| PC1 | -0.04 | 0.04 | -0.11 | 0.03 |
| PC2 | 0.00 | 0.07 | -0.14 | 0.14 |
| Vowel quality : Spectral tilt | 0.01 | 0.03 | -0.05 | 0.08 |
| Vowel quality : Duration | 0.03 | 0.03 | -0.03 | 0.09 |
| Spectral tilt : Duration | 0.00 | 0.02 | -0.04 | 0.05 |
| Vowel quality : PC1 | 0.01 | 0.02 | -0.04 | 0.05 |
| Spectral tilt : PC1 | 0.00 | 0.01 | -0.01 | 0.01 |
| Duration : PC1 | 0.00 | 0.01 | -0.01 | 0.01 |
| Vowel quality : PC2 | 0.01 | 0.05 | -0.08 | 0.09 |
| Spectral tilt : PC2 | 0.01 | 0.01 | -0.01 | 0.03 |
| Duration : PC2 | 0.01 | 0.01 | -0.01 | 0.03 |
| Vowel quality : Spectral tilt : Duration | 23.04 | 25.69 | -26.91 | 73.81 |
| Vowel quality : Spectral tilt : PC1 | 0.00 | 0.01 | -0.01 | 0.02 |
| Vowel quality : Duration : PC1 | 0.00 | 0.01 | -0.01 | 0.01 |
| Spectral tilt : Duration : PC1 | -0.00 | 0.00 | -0.01 | 0.00 |
| Vowel quality : Spectral tilt : PC2 | 0.01 | 0.01 | -0.01 | 0.03 |
| Vowel quality : Duration : PC2 | -0.00 | 0.01 | -0.02 | 0.02 |
| Spectral tilt : Duration : PC2 | 0.02 | 0.01 | -0.00 | 0.03 |
| Vowel quality : Spectral tilt : Duration : PC1 | -1.51 | 5.10 | -11.41 | 8.69 |
| Vowel quality : Spectral tilt : Duration : PC2 | -3.51 | 9.59 | -22.31 | 15.39 |

Table 4.5: Coefficients for the Bayesian multilevel model of data from the Spanish perception task. The intercept represents the logodds ratio of ponDIS (oxytone) responses with neutralized vowel quality, duration, and spectral tilt.

## Catalan perception

Perception data from the Catalan task completed by trilingual listeners were isolated and submitted to a multilevel Bayesian model with a four-way interaction between the three acoustic measures and PC1, a four-way interaction between the three acoustic measures and PC2, and by-participant random slopes of the three-way interaction of the acoustic measures (see Equation (4.1) for model formula). The logodds estimates from this model are shown in Table 4.6. For the purposes of visualization, word identifications produced by L1 English and L3 English trilinguals across the three acoustic manipulation matrices were grouped and plotted into heatmaps (Figures B.10 to B.12). As shown in the visual representation of logodds estimates of oxytone ponDIS responses in Figure 4.11, vowel quality is the most heavily weighted cue across all trilinguals (β = 0.73 ± 0.11, 95% CI = [0.52, 0.95]), followed by spectral tilt (β = 0.33 ± 0.08, 95% CI = [0.18, 0.49]), and duration (β = 0.31 ± 0.06, 95% CI = [0.20, 0.43]), in decreasing order of cue-weightings.

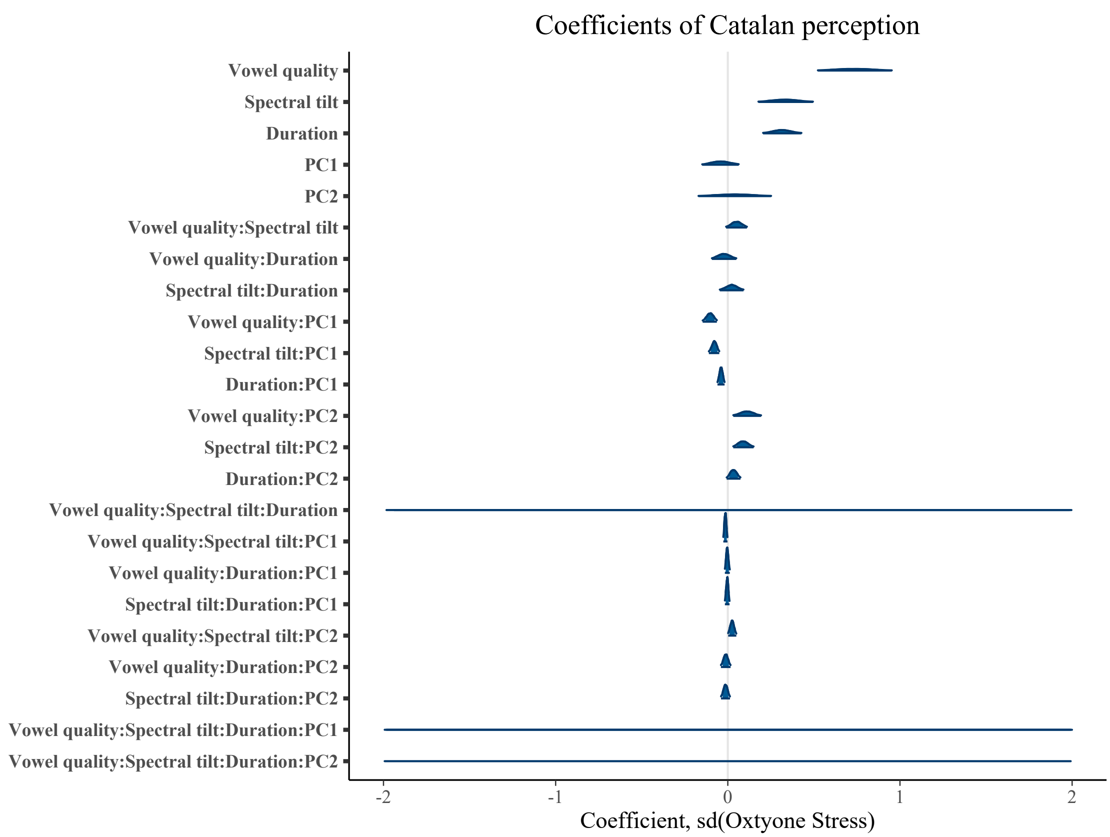



Figure 4.11: Coefficients from the Bayesian multilevel model of Catalan perception. Coefficients with a distribution that does not include $x = 0$ are statistically distinct from 0 within a 95% confidence interval.

| | Estimate | Error | L-95% CI | U-95% CI |
|---|---|---|---|---|
| (Intercept) | -0.94 | 0.27 | -1.47 | -0.41 |
| Vowel quality | 0.73 | 0.11 | 0.52 | 0.95 |
| Spectral tilt | 0.33 | 0.08 | 0.18 | 0.49 |
| Duration | 0.31 | 0.06 | 0.20 | 0.43 |
| PC1 | -0.04 | 0.05 | -0.15 | 0.06 |
| PC2 | 0.04 | 0.11 | -0.17 | 0.25 |
| Vowel quality : Spectral tilt | 0.05 | 0.03 | -0.01 | 0.11 |
| Vowel quality : Duration | -0.02 | 0.04 | -0.09 | 0.05 |
| Spectral tilt : Duration | 0.02 | 0.03 | -0.05 | 0.09 |
| Vowel quality : PC1 | -0.10 | 0.02 | -0.15 | -0.06 |
| Spectral tilt : PC1 | -0.08 | 0.02 | -0.11 | -0.05 |
| Duration : PC1 | -0.04 | 0.01 | -0.06 | -0.02 |
| Vowel quality : PC2 | 0.11 | 0.04 | 0.03 | 0.19 |
| Spectral tilt : PC2 | 0.09 | 0.03 | 0.03 | 0.15 |
| Duration : PC2 | 0.03 | 0.02 | -0.01 | 0.07 |
| Vowel quality : Spectral tilt : Duration | -6.84 | 29.30 | -63.88 | 51.12 |
| Vowel quality : Spectral tilt : PC1 | -0.01 | 0.01 | -0.03 | -0.00 |
| Vowel quality : Duration : PC1 | -0.00 | 0.01 | -0.02 | 0.01 |
| Spectral tilt : Duration : PC1 | -0.00 | 0.01 | -0.02 | 0.01 |
| Vowel quality : Spectral tilt : PC2 | 0.03 | 0.01 | 0.00 | 0.05 |
| Vowel quality : Duration : PC2 | -0.01 | 0.01 | -0.04 | 0.02 |
| Spectral tilt : Duration : PC2 | -0.01 | 0.01 | -0.04 | 0.01 |
| Vowel quality : Spectral tilt : Duration : PC1 | 0.26 | 5.85 | -11.13 | 11.70 |
| Vowel quality : Spectral tilt : Duration : PC2 | -18.30 | 11.83 | -41.13 | 4.68 |

Table 4.6: Coefficients for the Bayesian multilevel model of data from the Catalan perception task. The intercept represents the logodds ratio of ponDIS (oxytone) responses with neutralized vowel quality, duration, and spectral tilt.

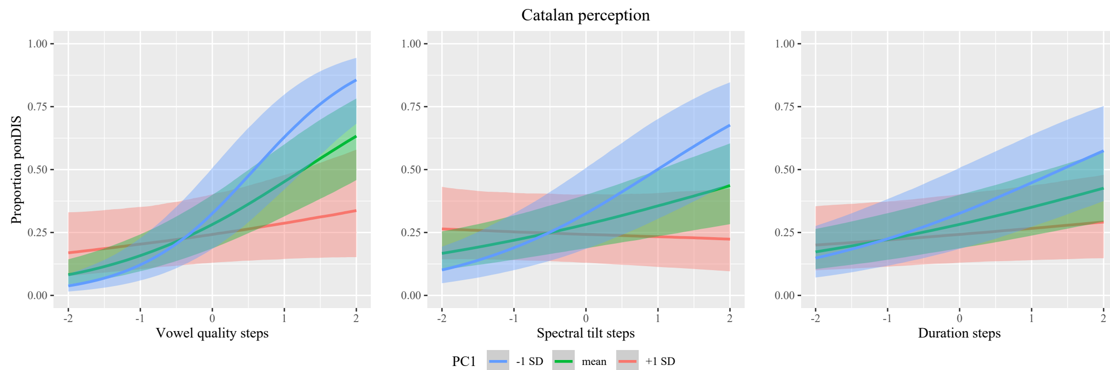



Figure 4.12: Predicted effects of vowel quality, duration, and spectral tilt cue-weighting across PC1 in Catalan perception.

As English dominance relative to Spanish and Catalan dominance (PC1) decreases, cue-weightings for vowel quality increase (β = -0.10 ± 0.02, 95% CI = [-0.15, -0.06]). Similarly, as PC1 decreases, cue-weightings for both spectral tilt (β = -0.08 ± 0.02, 95% CI = [-0.11, -0.05]) and duration (β = -0.04 ± 0.01, 95% CI = [-0.06, -0.02]) increase. These relationships are visualized in Figure 4.12. As Catalan dominance relative to Spanish dominance (PC2) increases, cue-weightings for vowel quality (β = 0.11 ± 0.04, 95% CI = [0.03, 0.19]) and spectral tilt (β = 0.09 ± 0.03, 95% CI = [0.03, 0.15]) increase. These relationships are visualized in Figure 4.13. Lastly, with increasing PC2, cue-weightings for spectral tilt and vowel quality covary (β = 0.03 ± 0.01, 95% CI = [0.01, 0.05]). For these listeners, vowel quality is more heavily weighted and yields increased discrimination when spectral tilt favors the oxytone (ponDIS) response (see Figure 4.14).

## 4.4 Discussion and conclusions

In this experiment, L1 English-L2 Spanish-L3 Catalan and early Spanish-Catalan bilinguals that learned English as an L3 (henceforth, L3 English trilinguals) completed word identification tasks with nonword stress minimal pairs in English, Spanish, and Catalan. To observe potential regressive crosslinguistic influence, monolingual English listeners also completed the English task. The main findings will be summarized to address the research questions posed at the beginning of the chapter.

*RQ1*: How does listeners' relative experience with the three languages affect their cue-weighting in each language? Based on the results of this experiment, relative experience with the three languages does affect cue-weighting in English and Catalan. In English, listeners that have English as an L1 responded differently to simultaneous manipulations of vowel quality and duration than listeners that have English as an L3. When duration manipulations favored the ponDIS endpoint, The L1 English trilinguals used vowel reduction in the first

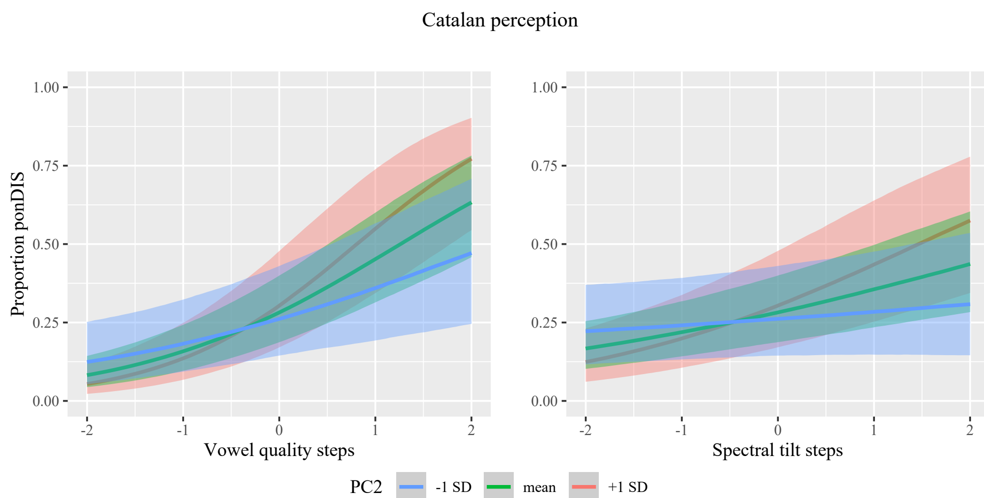



Figure 4.13: Predicted effects of vowel quality and spectral tilt cue-weighting across PC2 in Catalan perception.

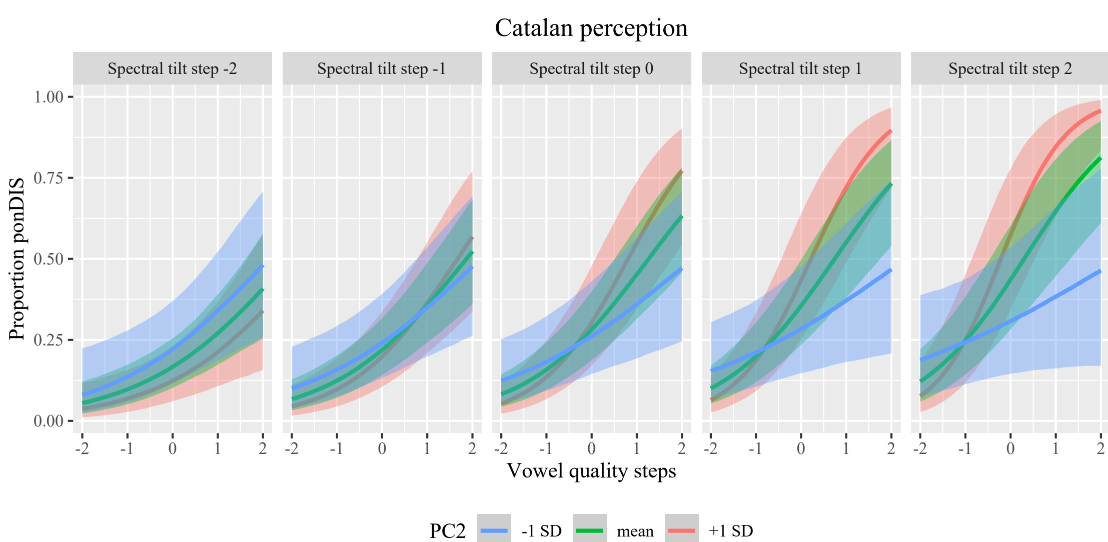



Figure 4.14: Predicted effects of vowel quality steps, spectral tilt steps, and PC2 on Catalan perception.

syllable to a greater extent as a cue to stress than listeners that have English as an L3. However, when the more fine-grained measure of relative language dominance is examined (rather than order of acquisition), there was no evidence that language experience affected cue-weightings in English among trilinguals. This could be due to the fact that the two groups of trilingual participants were fairly homogeneous, and a continuum of language dominance scores did not offer any finer detail than the grouped language profiles of 'L1 English' and 'L3 English', especially when the monolingual baseline was removed. In Catalan, as English dominance relative to Spanish and Catalan dominance increases, cue-weightings for duration, vowel quality, and spectral tilt all decrease. For all listeners, vowel quality is the heaviest weighted cue. As Catalan dominance relative to Spanish dominance increases, the cue-weightings for vowel quality and spectral tilt increase. Additionally, as Catalan dominance relative to Spanish dominance increases, cue-weightings for spectral tilt and vowel quality covary, where vowel quality is more heavily weighted when spectral tilt favors the oxytone (ponDIS) responses – the same pattern that is seen in English perception for trilingual but not monolingual listeners.

*RQ2*: Do L1 English trilinguals use suprasegmental cues to a greater extent than English monolinguals when perceiving stress in English? The results of the English perception task indicate that trilingual L1 English listeners appear to have vowel quality cue-weightings that are dependent on spectral tilt (a suprasegmental cue), whereas monolingual English listeners attend to vowel quality consistently across all steps on the spectral tilt continuum. Trilinguals also use vowel quality less than the monolingual English listeners do, indicating that exposure to additional languages has affected cue-weighting in their first language.

What do these results say about how cue-weighting transfer applies to L3 phonetics and phonology? Principally, the transfer of cue-weightings can be bidirectional, where cue-weightings in an L2 or L3 can affect cue-weightings in an L1 (i.e., regressive transfer). Additionally, there is a wide range of cue-weightings across the trilingual listeners, especially in Catalan, repudiating the notion that perception can be grouped into "nativelike" versus "L2" perception. This underscores that cue-weightings in the L2 and L3 are not rigid nor are they unchangeable copies of L1 cue-weightings. Rather, language experience and dominance can greatly shift cue-weightings. Most of the literature on cue-weighting transfer, however, examines naïve listeners or L2 learners who have had minimal experience with the L2. The trilinguals in the present study, however, have had significant exposure to and experience in the L2, and thus provide new insight into the role of language experience in cue-weighting transfer.

Another interesting finding is that in English, all listeners weight duration most heavily (for trilinguals this is above and beyond the highest cue). This seems to contradict the evidence in the literature that vowel quality is the heaviest cue to lexical stress in English. One possible explanation for this seemingly deviant result is the nature of the duration manipulations of the stimuli. As seen in Figure B.5, the English endpoint stimuli differed according to the duration of the first vowel whereas the duration of the second vowel remained relatively constant. The effect of this is that the overall duration of the word changes across the steps of the duration continuum, where the duration of the word at step 5 is almost half the duration of the word at step 1. In contrast, the total duration of the penultimate and ultimate vowels for the Spanish and Catalan stimuli is relatively constant across the duration continuum, but the ratio of the duration of the first and second vowel changes across the continuum. Because of this, in the English task, listeners may have been attending to the overall duration of the words when giving their word identification responses, rather than the relative duration of the first and second vowels. What is consistent with the literature is that monolingual English listeners attend to vowel quality even when duration cues are neutralized Figure 4.7. The L3 English trilinguals, however, have much lower cue-weightings for vowel quality when duration cues are neutralized.

In the Spanish perception task, vowel quality was the heaviest cue to lexical stress for all listeners. This result is not wholly unexpected given that the step size on the vowel quality continuum is much larger than the step sizes on the duration and spectral tilt continua (compare the endpoint values in Table 4.2 for vowel quality, duration, and spectral tilt). Although vowel quality (specifically for mid vowels) has been shown to vary across lexical stress in Spanish in production (e.g., Romanelli et al., 2018), attention to this cue in perception has been suggested to be rather weak (Ortega-Llebaria & Prieto, 2007). It is not clear from the results of this experiment whether vowel quality would have been used to disambiguate lexical stress pairs should duration have been a more informative cue, as it usually is in Spanish. It could be an idiosyncrasy of the speaker from whom the endpoints were elicited to vary vowel quality strongly for /i/, or perhaps vowel quality variations are more common – even for high vowels – than has been shown in the literature.

Trilingual listeners mainly differed in their perception of English and Catalan, however, little difference was found in the Spanish perception task. This makes sense based on the nature of the stimuli and the composition of the trilingual participant pool. The Spanish endpoint tokens (Table 4.2) are more similar across vowel quality, duration, and spectral tilt than the English and Catalan endpoints, yielding fewer possible cues to attend to and diminishing variance across the listeners. Additionally, the trilingual listeners mostly differ according to their relative English and Catalan dominance, rather than their relative Spanish dominance. This can be seen in the interpretation of the two principal components obtained from the BLP data with PCA, where the first component which explains 40.36% of the variance across participants relates to English dominance relative to Spanish and Catalan dominance. The second component which explains 10.32% of the variance relates to Catalan dominance relative to Spanish dominance. Given that the L1 English trilinguals are highly proficient in their L2 Spanish and that the L3 English trilinguals are early simultaneous or sequential Spanish bilinguals, the groups mainly differ according to their relative Catalan and English dominance.

Since vowel quality has been reported to be heavily weighted in both English and Catalan (e.g., Chrabaszcz et al., 2014; Ortega-Llebaria & Prieto, 2009), at first glance it's surprising that more English-dominant trilinguals (higher PC1) do not attend to vowel quality in Catalan as much as trilinguals with greater Catalan dominance relative to Spanish dominance (higher PC2). However, in English, vowel centralization (towards /ə/) cues lexical stress shifts, whereas in Catalan vowel quality changes are more complex (see Figure 2.7). Whereas some Catalan vowels, such as /a/, reduce to /ə/, others do not change phonemically in vowel quality, and still others undergo alternation in which they move further from /ə/ (such as /o/, which is realized as /u/ in unstressed positions). For the nonword stress pair used in the Catalan perception task, *pondis-pondís*, the penultimate vowel alternates from stressed /o/ to unstressed /u/ and the ultimate vowel is phonemically constant across stress.

Perhaps, then, L1 English listeners attend not so much to mere changes in vowel quality in their L1, but specifically to vowel centralization in unstressed syllables, since this is the predominate pattern they have been exposed to. Since the stressed /o/ to unstressed /u/ vowel quality change in Catalan is not centralization, but rather alternation, this pattern is perhaps unexpected and not as informative as the consistent reduction patterns in their L1 English. In contrast, because Catalan vowels are affected both by centralization and alternation, early Catalan bilingual listeners' cue-weighting is vowel-dependent (Ortega-Llebaria & Prieto, 2009). For some Catalan vowels, more centralized productions cue *stressed* syllables (e.g., /o/ in stressed syllable, /u/ in unstressed syllable), whereas for other vowels, more centralized productions cue *unstressed* syllables (e.g., /a/ in stressed syllable, /ə/ in unstressed syllable). This vowel-dependent cue-weighting may transfer to their L3 English, where vowel reduction/centralization may cue the shift from stressed to unstressed vowels only for certain vowels.

In the nonword stress pair used in English, /ɑ/ reduces to /ə/ in unstressed syllables (more central production). Although /ɑ/ is phonetically most similar to Catalan /a/ (which centralizes to /ə/ in unstressed syllables in Catalan), orthographically it is most similar to Catalan /o/ or /ɔ/ (both which alternate with /u/ in unstressed syllables in Catalan). Because these trilinguals have the phonological systems and writing systems of multiple languages, there can be orthographic effects on their phonological systems (Bassetti, in press). Therefore, vowel quality may be less informative for L3 English trilinguals in the English task since the English vowel reduction pattern opposes that which occurs in their L1 or L2 Catalan for /o/ and /ɔ/. To test this theory, one could create a follow-up word identification task in Catalan and English. In the Catalan task, the nonword stress pair would contain a vowel that centralizes in unstressed syllables and the English task would contain a nonword stress pair with the grapheme ⟨a⟩ which centralizes in unstressed syllables. If L1 English trilinguals attend more to vowel quality in the Catalan task than they do in the present study, this could indicate the facilitative transfer of vowel centralization cue-weighting. If L3 English trilinguals attend more to vowel quality in the English task, similarly this could indicate the facilitative transfer of vowel centralization cue-weighting that is vowel-dependent, reinforced by orthography.

The next chapter examines these findings in light of the findings from the production experiment in Chapter 3 to make some concluding remarks about cue-weighting transfer in the context of L3 phonetics and phonology.

# Chapter 5

# Conclusion

## 5.1 Summary of findings

This dissertation analyzes the production and perception of lexical stress in trilinguals' first, second, and third languages (L1, L2, and L3) to evaluate how the cue-weighting transfer hypothesis applies to L3 acquisition. Acoustic correlates of lexical stress were analyzed in English, Spanish, and Catalan, which all have lexical stress as indicated by the existence of stress minimal pairs and which are reported to belong to different rhythm classes (Prieto et al., 2012). L1 English-L2 Spanish-L3 Catalan trilinguals, early Spanish-Catalan bilinguals who learned English as an L3 (henceforth, L3 English trilinguals), and English monolinguals completed a sentence elicitation task and a word identification task in the respective languages.

The production task demonstrates that trilingual speakers with sufficient exposure to additional languages can use cues in a nativelike manner to cue lexical stress shifts. However, relative language dominance can affect how these cues are used in all three languages. For example, all speakers used duration to cue lexical stress shifts in the three languages, but the weight of duration across syllables in each language differed based on language dominance. Therefore, in bilingual and trilingual studies of cue-weighting transfer, it is important to measure acoustic cues to stress shifts on a more detailed level, taking into account how cues are weighted across syllables and different ways of signalling the same type of cue (e.g., syllable durations versus vowel durations). Additionally, the transfer of cue-weightings is demonstrated to be bidirectional, where cue-weightings in an L2 or L3 can affect cue-weightings in an L1 (i.e., regressive transfer). This finding provides further evidence that all languages in a multilingual's repertoire are active and available for use, and underscores the importance of analyzing trilingual production in all of a trilingual's languages, not just the L3.

The perception task demonstrates that relative experience with the three languages affects cue-weighting in English and Catalan, the trilinguals' respective L3s. In English, listeners that have English as an L1 respond differently to simultaneous manipulations of vowel quality and duration than listeners that have English as an L3. Whereas L3 English trilingual listeners weight vowel quality fairly lightly as duration is manipulated, L1 English trilingual listeners have heavier cue-weightings for vowel quality when the duration manipulations are on the ponDIS endpoint of the continuum. Therefore, it seems that L1 English listeners' cue-weightings of vowel quality covary with duration, whereas L3 English listeners' cue-weightings of vowel quality are independent from duration. In Catalan, as English dominance relative to Spanish and Catalan dominance increases, cue-weightings for duration, vowel quality, and spectral tilt all decrease. Secondly, the perception task demonstrates that exposure to additional languages has affected cue-weighting in L1 English trilinguals' first language, and therefore, cue-weighting transfer can also be regressive. For example, trilingual L1 English listeners appear to have vowel quality cue-weightings that are dependent on spectral tilt (a suprasegmental cue), whereas monolingual English listeners attend to vowel quality consistently across all steps on the spectral tilt continuum. Trilinguals also use vowel quality less than the monolingual English listeners do, indicating that exposure to additional languages has affected cue-weighting in their first language.

## 5.2 Returning to theory

Although these findings are, in part, specific to this group of participants and this grouping of languages, these results are not altogether surprising when they are considered through the frameworks of L3 models of phonetics and phonology. Based on the findings from these results of trilingual production and perception of lexical stress shifts, a theoretical framework of L3 phonetics and phonology needs to account for the following findings:

Finding 1. The L1 sound system is not fixed, but rather is influenced by subsequently learned languages (regressive transfer occurs);

Finding 2. Subsequent sound systems are not fixed copies of the L1. The individual's dynamic experience with each language affects how features are produced and perceived across their languages;

Finding 3. Regressive transfer and the effects of language experience can be manifested in both production and perception;

Finding 4. Both segmental knowledge and suprasegmental knowledge are represented in the multilingual repertoire and are susceptible to regressive transfer and the effects of language experience.

Of the models of bilingual phonetics and phonology which were discussed in Section 1.2, the Attrition & Drift in Access, Production, and Perception Theory (ADAPPT; de Leeuw & Chang, in press) makes predictions for multilingual production and perception which are most closely aligned with the above findings. As the SLM/SLM-r, PAM-L2, and L2LP are the predominant theoretical frameworks used in studies of bilingual phonetics and phonology, specific comparisons will be drawn between the predictions of these models and the predictions of ADAPPT. Regarding Finding 1, one of the most important differences between ADAPPT and other theoretical models is the treatment of regressive transfer. The SLM, SLM-r, and PAM-L2 acknowledge that crosslinguistic influence can be bidirectional (and therefore, that regressive transfer can occur), but these models make no claims as to the prevalence or implications of such transfer for the relationship between the L1 and L2 within bilinguals. Moreover, L2LP refutes the possibility of regressive transfer in perception (Escudero, 2005, p. 121). In contrast, ADAPPT theorizes that the L1 will always be impacted in some manner through the acquisition of and exposure to the L2, and that regressive transfer is a normal part of bilingual speech.

Finding 2 is supported by most theoretical frameworks of bilingual phonetics and phonology. The SLM-r, in fact, departs from the SLM in large part due to the central role the SLM-r places on the quantity and quality of language input in the varied outcomes of L2 production and perception. L2LP states that although L2 perception begins with a copy of L1 perception, language exposure can shape L2 perception so that it "reach[es] the optimal target L2 perception level" (Escudero, 2005, p. 121). PAM-L2 states that "[p]erceptual learning occurs for some L2 contrasts, but seems to depend on their phonological and phonetic relationship to the L1, specifically on perceived similarities vs. dissimilarities to L1 phonemes" (Best & Tyler, 2007, p. 20). ADAPPT, in turn, states that the L1 and L2 are dynamic and can change at any time in a person's life. Specifically, the nature of input (e.g., recency, duration, environment) can affect the nature of crosslinguistic influence (e.g., whether change is short-term or long-term).

Regarding Finding 3, ADAPPT addresses both production and perception, stipulating that all dimensions of bilingual speech can be affected in a bilingual's repertoire. Further, the model states that there is no straightforward link between production and perception and both systems may show different effects of attrition and drift. In contrast, specific relationships between production and perception are theorized by other models of bilingual phonetics and phonology (PAM-L2, L2LP, and the SLM: accurate perception precedes accurate production). Although the methodology of this dissertation was not designed to probe potential links between perception and production, the results show that both systems evidenced regressive transfer and were affected by language experience, but in different ways. Whereas gradient scores of relative language dominance were correlated with both English and Catalan cue-weightings in production, gradient scores of relative language dominance were only correlated with Catalan cue-weightings in perception. The less granular approach of grouping participants based on language profiles, however, was shown to be correlated with English cue-weightings in perception. Additionally, relative language dominance was only correlated with cue-weightings for duration and F2 in the production task, but was correlated with cue-weightings for duration, vowel quality, and spectral tilt in the perception task. Therefore, the results do not point to symmetrical effects of relative language experience on trilingual cue-weighting across production and perception.

Lastly, regarding Finding 4, the primary models of bilingual phonetics (SLM, SLM-r, PAM-L2, and L2LP) focus on segmental, rather than suprasegmental representation and use. In contrast, ADAPPT is one of the few theoretical models that touches on suprasegmental contrasts and processes as well as segmental contrasts. The authors state that there are many important contributors to the speech signal beyond segments, for example prosodic dimensions and phonotactic rules, and theories that explain speech must take all of these elements into account. The focus of the present work is to examine the production and perception of lexical stress, which is manifested by an array of segmental and suprasegmental features. Therefore, a theoretical framework that incorporates segmental and suprasegmental representation and use is not only preferable, but is necessary.

ADAPPT is relatively new to the field of bilingual phonetics and phonology, but its main principles are grounded in empirical studies of bilingualism and bidialectalism. That these principals align with the findings of this study not only gives further empirical weight to ADAPPT as is, but also supports an extension of this model to the L3. Future studies that examine production and perception in speakers with two or more languages, and additionally which incorporate segmental and suprasegmental features, can continue to shed light on the relative merits of this theoretical model compared to other, more established theories of bilingual phonetics and phonology.

## 5.3 Using the BLP with trilinguals

Principal component analysis (PCA) offers a promising method of analysis of the BLP, and other measures of relative language dominance, for trilinguals. Trilingualism is complex and we cannot expect one number to explain all the variation present in relative language dominance. However, PCA can significantly reduce the dimensionality of language dominance data while still accounting for the variation present. Although the participants used in this study were fairly homogeneous, where there were only L1 English-L2 Spanish-L3 Catalan speakers and early Spanish-Catalan bilinguals that learned English as an L3, the two components identified through PCA were interpretable and logical given the profiles of the participants. The interpretation of components may be a little less clear when trilinguals are more heterogeneous, but interpretations based on individual components of the BLP can still offer a means to analyze variation across participants. With more diversity in the order of acquisition and the incorporation of different combinations of sequential versus simultaneous bilingualism, a greater number of principal components may be identified, which will in turn increase the complexity of statistical models. However, we return again to the reality that trilingualism is complex and the relationships between language dominance and its various components (language history, use, attitudes, and proficiency) are complex. Rather than seeking simpler ways to conceptualize of and measure relative language dominance for multilinguals, we need to turn towards statistical models that can handle greater complexity, such as Bayesian modeling.

## 5.4 Concluding remarks

This dissertation has been situated within the field of L3 phonetics and phonology, and some issues within the field to date have been noted; specifically, the inconsistent treatment of language mode, an overt focus on production in the L3 (rather than equal treatment of perception and the L1 and L2), and a bias towards production studies which examine segmental features. This research design was intended to remedy some of these issues. A unilingual language mode was experimentally induced in each experimental session so that the source and direction of CLI could be more clearly observed. Data was obtained from all three languages in question so that potential regressive influence and the impact of language dominance on the fullness of the multilingual repertoire could be observed. Additionally, a continuous measure of relative language dominance was employed rather than preconceived categories such as 'L1 English' and 'L3 English' to tease apart the role of relative language experience and order of language acquisition. Lastly, both the production and perception of segmental and suprasegmental acoustic cues of a suprasegmental feature (lexical stress) were examined to contribute to the diversity of the field of L3 phonetics and phonology and provide a more well-rounded analysis of current theoretical frameworks.

As research in L3 phonetics and phonology becomes more prevalent, and as studies continue to minimize these issues and analyze a diverse range of segmental and suprasegmental features – as well as more typologically diverse language groupings – our understanding of multilingual use and representation of sound systems will continue to be informed by and to shape the growing body of theory.

# Bibliography


Amengual, M. (2012). Interlingual influence in bilingual speech: Cognate status effect in a continuum of bilingualism. *Bilingualism*, *15*(3), 517–530. https://doi.org/10.1017/S1366728911000460

Amengual, M. (2021). The acoustic realization of language-specific phonological categories despite dynamic cross-linguistic influence in bilingual and trilingual speech. *The Journal of the Acoustical Society of America*, *149*(2), 1271–1284. https://doi.org/10.1121/10.0003559

Amengual, M. (2023). Cross-language influences in the acquisition of L2 and L3 phonology. In E. Irina, M. Brysbaert, & A. Siyanova-Chanturia (Eds.), *Cross-language Influences in Bilingual Processing and Second Language Acquisition* (pp. 74–99). John Benjamins Publishing Company. https://doi.org/10.1075/bpa.16.04ame

Astruc, L., Payne, E., Post, B., Prieto, P., & Vanrell, M. d. M. (2010). Word prosody in early child Catalan, Spanish and English. *Speech Prosody 2010*, *100173*. https://www.isca-speech.org/archive/pdfs/speechprosody_2010/astruc10_speechprosody.pdf

Bardel, C., & Sánchez, L. (2017). The L2 status factor hypothesis revisited: The role of metalinguistic knowledge, working memory, attention and noticing in third language learning. In T. Angelovska & A. Hahn (Eds.), *L3 Syntactic Transfer: Models, New Developments and Implications* (pp. 85–101, Vol. 3). John Benjamins. https://doi.org/10.1075/bpa.5.05bar

Bassetti, B. (in press). Orthographic effects in the phonetics and phonology of second language learners and users. In M. Amengual (Ed.), *The Cambridge Handbook of Bilingual Phonetics and Phonology*. Cambridge University Press.

Beckman, M., & Edwards, J. (1994). Articulatory evidence for differentiating stress categories. In P. Keating (Ed.), *Phonological Structure and Phonetic Form: Papers in Laboratory Phonology III* (pp. 7–33). Cambridge University Press. https://doi.org/10.1017/CBO9780511659461.002

Best, C. T. (1991). The emergence of native-language phonological influences in infants: A perceptual assimilation model. *Haskins Laboratories Status Report on Speech Research*, *107*, 1–30. https://www.haskins.yale.edu/SR/SR107/SR107_01.pdf

Best, C. T., & Tyler, M. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O.-S. Bohn & M. J. Munro (Eds.), *Language*

*Experience in Second Language Speech Learning: In Honour of James Emil Flege* (pp. 13–34). John Benjamins. https://doi.org/10.1075/lllt.17.07bes

Birdsong, D. (2018). Plasticity, variability and age in second language acquisition and bilingualism. *Frontiers in Psychology*, *9*, 1–17. https://doi.org/10.3389/fpsyg.2018.00081

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

Boersma, P., & Weenink, D. (2019). Praat: doing phonetics by computer [Computer Program]. Version 6.1.47.

Brown, M., & Chang, C. B. (2023). Regressive cross-linguistic influence in multilingual speech rhythm: The role of language similarity. In M. M. Brown-Bousfield, S. Flynn, & É. Fernández-Berkes (Eds.), *L3 Development After the Initial State*. John Benjamins.

Bruggeman, L., Yu, J., & Cutler, A. (2022). Listener adjustment of stress cue use to fit language vocabulary structure. *Proc. Speech Prosody 2022*, 264–267. https://doi.org/10.21437/SpeechProsody.2022-54

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28. https://doi.org/10.18637/jss.v080.i01

Bybee, J. (2000). Lexicalization of sound change and alternating environments. In M. Broe & J. Pierrehumbert (Eds.), *Papers in Laboratory Phonology V: Acquisition and the Lexicon* (pp. 250–68). Cambridge: Cambridge University Press.

Bybee, J. (2002). Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change*, *14*(3), 261–290. https://doi.org/10.1017/S0954394502143018

Byers, E., & Yavas, M. (2017). Vowel reduction in word-final position by early and late Spanish-English bilinguals. *PloS one*, *12*(4), e0175226. https://doi.org/10.1371/journal.pone.0175226

Cabrelli Amaro, J. (2017). Testing the phonological permeability hypothesis: L3 phonological effects on L1 versus L2 systems. *International Journal of Bilingualism*, *21*(6), 698–717. https://doi.org/10.1177/1367006916637287

Cabrelli Amaro, J., & Rothman, J. (2010). On L3 acquisition and phonological permeability: A new test case for debates on the mental representation of non-native phonological systems. *International Review of Applied Linguistics in Language Teaching*, *48*(2–3). https://doi.org/10.1515/iral.2010.012

Cabrelli Amaro, J., & Wrembel, M. (2016). Investigating the acquisition of phonology in a third language–a state of the science and an outlook for the future. *International Journal of Multilingualism*, *13*(4), 395–409. https://doi.org/10.1080/14790718.2016.1217601

Campbell, N., & Beckman, M. (1997). Stress, prominence, and spectral tilt. In A. Botinis, G. K. Kouroupetroglou, & G. Carayiannis (Eds.), *Intonation: Theory, Models and Applications* (pp. 67–70). European Speech Communication Association.

Chan, I. L., & Chang, C. B. (2019). Perception of nonnative tonal contrasts by Mandarin-English and English-Mandarin sequential bilinguals. *The Journal of the Acoustical Society of America*, *146*(2), 956–972. https://doi.org/10.1121/1.5120522

Chrabaszcz, A., Winn, M., Lin, C. Y., & Idsardi, W. J. (2014). Acoustic cues to perception of word stress by English, Mandarin, and Russian speakers. *Journal of Speech, Language, and Hearing Research*, *57*(4), 1468–1479. https://doi.org/10.1044/2014_JSLHR-L-13-0279

Cook, V. (2003). *Effects of the Second Language on the First* (Vol. 3). Multilingual Matters.

Cruttenden, A. (1993). The de-accenting and re-accenting of repeated lexical items. *ESCA Workshop on Prosody*, 16–19.

Cutler, A. (1986). Forbear is a homophone: Lexical prosody does not constrain lexical access. *Language and Speech*, *29*(3), 201–220. https://doi.org/10.1177/002383098602900302

Cutler, A. (2005). Lexical stress. In D. B. Pisoni & R. E. Remez (Eds.), *The Handbook of Speech Perception* (pp. 264–289). Blackwell Publishing.

Cutler, A. (2015). Lexical stress in English pronunciation. In M. Reed & J. M. Levis (Eds.), *The Handbook of English Pronunciation* (pp. 106–124). Wiley Online Library. https://doi.org/10.1002/9781118346952.ch6

Cutler, A., Norris, D., & Sebastián-Gallés, N. (2004). Phonemic repertoire and similarity within the vocabulary. *8th International Conference on Spoken Language Processing (Interspeech 2004-ICSLP)*, 65–68.

Davidson, L. (2021). The versatility of creaky phonation: Segmental, prosodic, and sociolinguistic uses in the world's languages. *Wiley Interdisciplinary Reviews: Cognitive Science*, *12*(3), e1547. https://doi.org/10.1002/wcs.1547

De Bot, K. (2012). Rethinking multilingual processing: From a static to a dynamic approach. In J. Cabrelli Amaro, S. Flynn, & J. Rothman (Eds.), *Third Language Acquisition in Adulthood* (pp. 79–94, Vol. 46). John Benjamins Publishing Co Amsterdam. https://doi.org/10.1075/sibil.46.07bot

De Bot, K., Lowie, W., & Verspoor, M. (2007). A dynamic systems theory approach to second language acquisition. *Bilingualism: Language and Cognition*, *10*(1), 7–21. https://doi.org/10.1017/S1366728906002732

De Jong, K. J. (1995). The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation. *The Journal of the Acoustical Society of America*, *97*(1), 491–504. https://doi.org/10.1121/1.412275

de Leeuw, E., & Chang, C. B. (in press). Phonetic and Phonological L1 Attrition and Drift in Bilingual Speech. In M. Amengual (Ed.), *The Cambridge Handbook of Bilingual Phonetics and Phonology*. Cambridge University Press.

Domene Moreno, C. (2021). *Beyond transfer? The acquisition of an L3 phonology by Turkish-German bilinguals* [Doctoral dissertation, Julius-Maximilians-Universität Würzburg].

Escudero, P. (2005). *Linguistic perception and second language acquisition: Explaining the attainment of optimal phonological categorization* [Doctoral dissertation]. Utrecht University.

Escudero, P. (2009). Linguistic perception of "similar" L2 sounds. In P. Boersma & S. Hamann (Eds.), *Phonology in Perception* (pp. 151–190). Mouton de Gruyter Berlin. https://doi.org/10.1515/9783110219234.151

Escudero, P., Broersma, M., & Simon, E. (2013). Learning words in a third language: Effects of vowel inventory and language proficiency. *Language and Cognitive Processes*, *28*(6), 746–761. https://doi.org/10.1080/01690965.2012.662279

Escudero, P., Smit, E. A., & Mulak, K. E. (2022). Explaining L2 Lexical Learning in Multiple Scenarios: Cross-Situational Word Learning in L1 Mandarin L2 English Speakers. *Brain Sciences*, *12*(12), 1618. https://doi.org/10.3390/brainsci12121618

Evans, B. G., & Iverson, P. (2007). Plasticity in vowel perception and production: A study of accent change in young adults. *The Journal of the Acoustical Society of America*, *121*(6), 3814–3826. https://doi.org/10.1121/1.2722209

Ferrand, C. T. (2002). Harmonics-to-noise ratio: An index of vocal aging. *Journal of Voice*, *16*(4), 480–487. https://doi.org/10.1016/S0892-1997(02)00123-6

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

Flege, J. E., & Bohn, O.-S. (2021). The Revised Speech Learning Model (SLM-r). In R. Wayland (Ed.), *Second language speech learning: Theoretical and empirical progress* (pp. 3–83). Cambridge University Press. https://doi.org/10.1017/9781108886901.002

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. https://doi.org/10.1080/14790710408668175

Fraile, R., & Godino-Llorente, J. I. (2014). Cepstral peak prominence: A comprehensive analysis. *Biomedical Signal Processing and Control*, *14*, 42–54. https://doi.org/10.1016/j.bspc.2014.07.001

Francis, A. L., Baldwin, K., & Nusbaum, H. C. (2000). Effects of training on attention to acoustic cues. *Perception & Psychophysics*, *62*(8), 1668–1680.

Francis, A. L., & Nusbaum, H. C. (2002). Selective attention and the acquisition of new phonetic categories. *Journal of Experimental Psychology: Human perception and performance*, *28*(2), 349. https://doi.org/10.1037/0096-1523.28.2.349

Fry, D. B. (1958). Experiments in the perception of stress. *Language and Speech*, *1*(2), 126–152. https://doi.org/10.1177/002383095800100207

Gertken, L. M., Amengual, M., & Birdsong, D. (2014). Assessing language dominance with the bilingual language profile. *Measuring L2 proficiency: Perspectives from SLA*, 208–225. https://doi.org/10.21832/9781783092291-014

Goldman, J.-P. (2011). EasyAlign: An Automatic Phonetic Alignment Tool Under Praat. *INTERSPEECH*.

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

Grosjean, F. (1989). Neurolinguists, beware! The bilingual is not two monolinguals in one person. *Brain and Language*, *36*(1), 3–15.

Grosjean, F. (2011). An attempt to isolate, and then differentiate, transfer and interference. *International Journal of Bilingualism*, *16*(1), 11–21. https://doi.org/10.1177/1367006911403210

Grosjean, F. (2001). The bilingual's language modes. In J. Nicol (Ed.), *Explaining Linguistics. One Mind, Two Languages: Bilingual Language Processing* (pp. 1–22). Blackwell.

Gut, U. (2010). Cross-linguistic influence in L3 phonological acquisition. *International Journal of Multilingualism*, *7*(1), 19–38. https://doi.org/10.1080/14790710902972248

Holt, L. L., & Lotto, A. J. (2006). Cue weighting in auditory categorization: Implications for first and second language acquisition. *The Journal of the Acoustical Society of America*, *119*(5), 3059–3071. https://doi.org/10.1121/1.2188377

Hualde, J. I., & Prieto, P. (2015). Intonational variation in Spanish: European and American varieties. In S. Frota & P. Prieto (Eds.), *Intonation in Romance*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199685332.003.0010

Ingvalson, E. M., Holt, L. L., & McClelland, J. L. (2012). Can native Japanese listeners learn to differentiate /r–l/ on the basis of F3 onset frequency? *Bilingualism: Language and Cognition*, *15*(2), 255–274. https://doi.org/10.1017/S1366728911000447

Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Kettermann, A., Siebert, C., et al. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition*, *87*(1), B47–B57. https://doi.org/10.1016/S0010-0277(02)00198-1

Johnson, K. (1997). Speech perception without speaker normalization: An exemplar model. In K. Johnson & J. W. Mullennix (Eds.), *Talker Variability in Speech Processing* (pp. 145–166). Morgan Kaufmann Publishers.

Johnson, K. (2020). The ΔF method of vocal tract length normalization for vowels. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, *11*(1). https://doi.org/10.5334/labphon.196

Kawahara, H., Morise, M., Takahashi, T., Nisimura, R., Irino, T., & Banno, H. (2008). Tandem-STRAIGHT: A temporally stable power spectral representation for periodic signals and applications to interference-free spectrum, F0, and aperiodicity estimation. *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*, 3933–3936. https://doi.org/10.1109/ICASSP.2008.4518514

Keating, P. A., Garellek, M., & Kreiman, J. (2015). Acoustic properties of different kinds of creaky voice. In The Scottish Consortium for ICPhS 2015 (Ed.), *ICPhS* (pp. 1–5).

Kendall, T., & Fridland, V. (2012). Variation in perception and production of mid front vowels in the US Southern Vowel Shift. *Journal of Phonetics*, *40*(2), 289–306. https://doi.org/10.1016/j.wocn.2011.12.002

Kim, D., Clayards, M., & Goad, H. (2018). A longitudinal study of individual differences in the acquisition of new vowel contrasts. *Journal of Phonetics*, *67*, 1–20. https://doi.org/10.1016/j.wocn.2017.11.003

Kim, J. Y. (2015). Perception and production of Spanish lexical stress by Spanish heritage speakers and English L2 learners of Spanish. *Selected proceedings of the 6th Conference on Laboratory Approaches to Romance Phonology*, 106–128.

Kim, J. Y. (2016). *The Perception and Production of Prominence in Spanish by Heritage Speakers and L2 Learners* [Doctoral dissertation, University of Illinois at Urbana-Champaign].

Kopečková, R. (2016). The bilingual advantage in L3 learning: A developmental study of rhotic sounds. *International Journal of Multilingualism*, *13*(4), 410–425. https://doi.org/10.1080/14790718.2016.1217605

Kopečková, R., Marecka, M., Wrembel, M., & Gut, U. (2016). Interactions between three phonological subsystems of young multilinguals: The influence of language status. *International Journal of Multilingualism*, *13*(4), 426–443. https://doi.org/10.1080/14790718.2016.1217603

Law, W. L., Dmitrieva, O., & Francis, A. L. (2019). Language attitudes modulate phonetic interactions between languages in bilingual speakers in diglossic settings. *Linguistic Approaches to Bilingualism*, *11*(3), 289–322. https://doi.org/10.1075/lab.18085.law

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. https://doi.org/10.1080/14790710902972255

Llama, R., & López-Morelos, L. P. (2016). VOT production by Spanish heritage speakers in a trilingual context. *International Journal of Multilingualism*, *13*(4), 444–458. https://doi.org/10.1080/14790718.2016.1217602

Llama, R., & López-Morelos, L. P. (2020). On Heritage Accents: Insights from Voice Onset Time Production by Trilingual Heritage Speakers of Spanish. In E. Babatsouli & M. J. Ball (Eds.), *An Anthology of Bilingual Child Phonology* (p. 209, Vol. 142). Multilingual Matters Limited.

Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research*, *50*(4). https://doi.org/10.1044/1092-4388(2007/067)

McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., & Sonderegger, M. (2017). Montreal Forced Aligner: Trainable Text-Speech Alignment Using Kaldi. *Interspeech*, *2017*, 498–502. https://doi.org/https://doi.org/10.21437/interspeech.2017-1386

Nadeu, M. (2014). Stress-and speech rate-induced vowel quality variation in Catalan and Spanish. *Journal of Phonetics*, *46*, 1–22. https://doi.org/10.1016/j.wocn.2014.05.003

Nadeu, M. (2016). Phonetic and phonological vowel reduction in Central Catalan. *International Phonetic Association. Journal of the International Phonetic Association*, *46*(1), 33. https://doi.org/10.1017/S002510031500016X

Nalborczyk, L., Batailler, C., Lœvenbruck, H., Vilain, A., & Bürkner, P.-C. (2019). An introduction to Bayesian multilevel models using brms: A case study of gender effects on vowel variability in standard Indonesian. *Journal of Speech, Language, and Hearing Research*, *62*(5), 1225–1242. https://doi.org/10.1044/2018_JSLHR-S-18-0006

Olson, D. J. (2023). Measuring bilingual language dominance: An examination of the reliability of the Bilingual Language Profile. *Language Testing*, *40*(3), 521–547. https://doi.org/10.1177/02655322221139162

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

Ortega-Llebaria, M., del Mar Vanrell, M., & Prieto, P. (2010). Catalan speakers' perception of word stress in unaccented contexts. *The Journal of the Acoustical Society of America*, *127*(1), 462–471. https://doi.org/10.1121/1.3268506

Ortega-Llebaria, M., Gu, H., & Fan, J. (2013). English speakers' perception of Spanish lexical stress: Context-driven L2 stress perception. *Journal of Phonetics*, *41*(3-4), 186–197. https://doi.org/10.1016/j.wocn.2013.01.006

Ortega-Llebaria, M., & Prieto, P. (2007). Disentangling stress from accent in Spanish: Production patterns of the stress contrast in deaccented syllables. In P. Prieto, J. Mascaró, & M.-J. Solé (Eds.), *Segmental and Prosodic Issues in Romance Phonology* (pp. 155–176). John Benjamins.

Ortega-Llebaria, M., & Prieto, P. (2009). Perception of word stress in Castilian Spanish. In M. Vigário, S. Frota, & M. João Freitas (Eds.), *Phonetics and Phonology: Interactions and Interrelations* (pp. 35–50). John Benjamins Amsterdam.

Ortega-Llebaria, M., & Prieto, P. (2011). Acoustic correlates of stress in Central Catalan and Castilian Spanish. *Language and Speech*, *54*(1), 73–97. https://doi.org/10.1177/0023830910388014

Ortega-Llebaria, M., Prieto, P., & del Mar Vanrell, M. (2007). Perceptual evidence for direct acoustic correlates of stress in Spanish. *Proceedings of the 16th International Congress on Phonetic Sciences, Saarbrücken*, 1121–1124.

Ortín, R., & Simonet, M. (2022). Phonological processing of stress by native English speakers learning Spanish as a second language. *Studies in Second Language Acquisition*, *44*(2), 460–482. https://doi.org/10.1017/S0272263121000309

Patience, M. (2018). Acquisition of the tap-trill contrast by L1 Mandarin–L2 English–L3 Spanish Speakers. *Languages*, *3*(4), 42.

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

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

Pierrehumbert, J. (2001). Exemplar dynamics: Word frequency, lenition and contrast. In J. Bybee & P. Hopper (Eds.), *Frequency and the Emergence of Linguistic Structure* (pp. 137–157). Amsterdam: John Benjamins.

Prieto, P., del Mar Vanrell, M., Astruc, L., Payne, E., & Post, B. (2012). Phonotactic and phrasal properties of speech rhythm. Evidence from Catalan, English, and Spanish.

*Speech Communication*, *54*(6), 681–702. https://doi.org/10.1016/j.specom.2011.12.001

Prieto, P., & Ortega-Llebaria, M. (2006). Stress and Accent in Catalan and Spanish: Patterns of duration, vowel quality, overall intensity, and spectral balance. *Proceedings of Speech Prosody*, 337–340.

Quilis, A., & Esgueva, M. (1983). Realización de los fonemas vocálicos españoles en posición fonética normal. *Estudios de fonética*, *1*, 159–252.

Rallo Fabra, L. (2015). Can nonnative speakers reduce English vowels in a native-like fashion? Evidence from L1-Spanish L2-English bilinguals. *Phonetica*, *72*(2-3), 162–181. https://doi.org/10.1159/000430920

Reinisch, E., & Holt, L. L. (2014). Lexically guided phonetic retuning of foreign-accented speech and its generalization. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(2), 539–555. https://doi.org/10.1037/a0034409

Revelle, W. (2023). *Psych: Procedures for psychological, psychometric, and personality research* [R package version 2.3.3]. Northwestern University. Evanston, Illinois. https://CRAN.R-project.org/package=psych

Romanelli, S., Menegotto, A., & Smyth, R. (2018). Stress-induced acoustic variation in L2 and L1 Spanish vowels. *Phonetica*, *75*(3), 190–218. https://doi.org/10.1159/000484611

Ronquest, R. (2013). An acoustic examination of unstressed vowel reduction in Heritage Spanish. *Selected Proceedings of the 15th Hispanic Linguistics Symposium*, 157–171.

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. https://doi.org/10.1017/S136672891300059X

Salomé, F., Casalis, S., & Commissaire, E. (2022). Bilingual advantage in L3 vocabulary acquisition: Evidence of a generalized learning benefit among classroom-immersion children. *Bilingualism: Language and Cognition*, *25*(2), 242–255. https://doi.org/10.1017/S1366728921000687

Shue, Y.-L. (2010). *The voice source in speech production: Data, analysis and models* [Doctoral dissertation]. Univeristy of California, Los Angeles.

Shure, Inc. (2022). ShurePlus MOTIV. https://shure.com/en-US/products/software/shure_plus_motiv

Silva, E. J. C. (2016). Multilingual-g2p. https://github.com/jcsilva/multilingual-g2p

Sluijter, A. M., & van Heuven, V. J. (1996a). Acoustic correlates of linguistic stress and accent in Dutch and American English. *Proceeding of Fourth International Conference on Spoken Language Processing. ICSLP'96*, *2*, 630–633. https://doi.org/10.1109/icslp.1996.607440

Sluijter, A. M., & van Heuven, V. J. (1996b). Spectral balance as an acoustic correlate of linguistic stress. *The Journal of the Acoustical Society of America*, *100*(4), 2471–2485. https://doi.org/10.1121/1.417955

Strange, W. (2011). Automatic selective perception (ASP) of first and second language speech: A working model. *Journal of Phonetics*, *39*(4), 456–466. https://doi.org/10.1016/j.wocn.2010.09.001

Sun, X. (2002). Pitch determination and voice quality analysis using subharmonic-to-harmonic ratio. *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 333–336. https://doi.org/10.1109/ICASSP.2002.5743722

Tremblay, A., Broersma, M., Zeng, Y., Kim, H., Lee, J., & Shin, S. (2021). Dutch listeners' perception of English lexical stress: A cue-weighting approach. *The Journal of the Acoustical Society of America*, *149*(6), 3703–3714. https://doi.org/10.1121/10.0005086

Tremblay, A., & Spinelli, E. (2014). English listeners' use of distributional and acoustic-phonetic cues to liaison in French: Evidence from eye movements. *Language and Speech*, *57*(3), 310–337. https://doi.org/10.1177/0023830913504569

Van Geert, P. (2008). The dynamic systems approach in the study of L1 and L2 acquisition: An introduction. *The Modern Language Journal*, *92*(2), 179–199. https://doi.org/10.1111/j.1540-4781.2008.00713.x

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

Vitevitch, M. S., Luce, P. A., Charles-Luce, J., & Kemmerer, D. (1997). Phonotactics and syllable stress: Implications for the processing of spoken nonsense words. *Language and Speech*, *40*(1), 47–62. https://doi.org/10.1177/002383099704000103

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

Wrembel, M. (2015). Cross-linguistic influence in second vs. third language acquisition of phonology. In U. Gut, R. Fuchs, & E.-M. Wunder (Eds.), *Universal or Diverse Paths to English Phonology* (pp. 41–70). De Gruyter Mouton. https://doi.org/10.1515/9783110346084-004

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. https://doi.org/10.3390/languages5040052

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

Yazawa, K., Whang, J., Kondo, M., & Escudero, P. (2020). Language-dependent cue weighting: An investigation of perception modes in L2 learning. *Second Language Research*, *36*(4), 557–581. https://doi.org/10.1177/0267658319832645

Zhang, Y., & Francis, A. (2010). The weighting of vowel quality in native and non-native listeners' perception of English lexical stress. *Journal of Phonetics*, *38*(2), 260–271. https://doi.org/10.1016/j.wocn.2009.11.002

Zhang, Y., Nissen, S. L., & Francis, A. L. (2008). Acoustic characteristics of English lexical stress produced by native Mandarin speakers. *The Journal of the Acoustical Society of America*, *123*(6), 4498–4513. https://doi.org/10.1121/1.2902165

# Appendix A

# Chapter 3 materials

| | Paroxytones - primarily nouns | Oxytones - primarily verbs |
|---|---|---|
| | billow [ɪ, oʊ] | below [ɪ, oʊ] |
| | discount [ɪ, aʊ] | discount [ɪ, aʊ] |
| | forebear [ɔː, ɛ] | forbear [ɔː, ɛ] |
| | import [ɪ, ɔː] | import [ɪ, ɔː] |
| A | impress [ɪ, ɛ] | impress [ɪ, ɛ] |
| | insight [ɪ, aɪ] | incite [ɪ, aɪ] |
| | permit [ə, ɪ] | permit [ə, ɪ] |
| | relay [i, eɪ] | relay [i, eɪ] |
| | trusty [ə, i] | trustee [ə, i] |
| | combine [ɑ, aɪ] | combine [ə, aɪ] |
| | compact [ɑ, æ] | compact [ə, a] |
| | contract [ɑ, æ] | contract [ə, a] |
| | convert [ɑ, ə] | convert [ə, ə] |
| B | convict [ɑ, ɪ] | convict [ə, ɪ] |
| | desert [ɛ, ə] | desert [ə, ə] |
| | project [ɑ, ɛ] | project [ə, ɛ] |
| | object [ɑ, ə] | object [ə, ɛ] |
| | record [ɛ, ə] | record [ə, ɔ] |
| | subject [ə, ɪ] | subject [ə, ɛ] |

Table A.1: Experimental stimuli for English production task, with stressed syllables underlined. Set A contains words where vowel quality is not expected to differ and Set B contains words where vowel quality is expected to differ

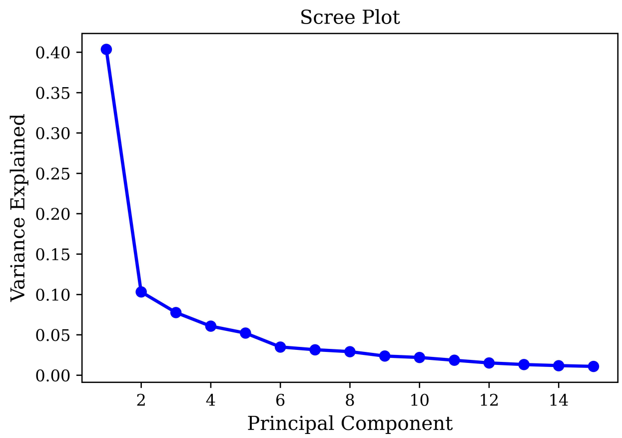



Figure A.1: Scree plot for PCA of BLP data. Based on the sharp decrease in explained variance, two components will be used.

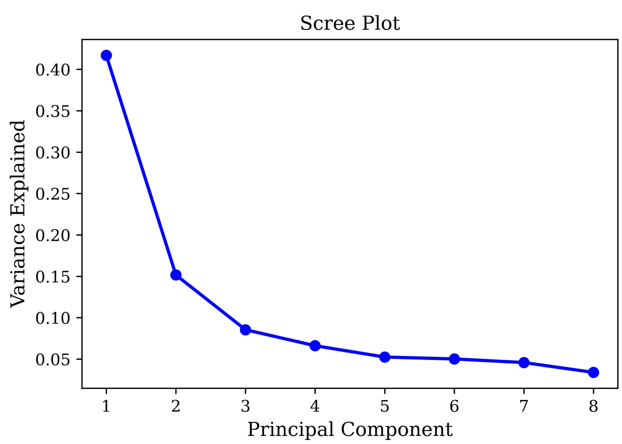



Figure A.2: Scree plot for PCA of a selection of acoustic measures from the production data. Based on the sharp decrease in explained variance, three components will be used.

| Word | Stress | Sentence |
|---|---|---|
| billow | paroxytone | No, the DRESS didn't billow in the wind, but the curtains did. |
| below | oxytone | No, her CALCIUM isn't below normal, but her iron is. |
| combine | paroxytone | No, there wasn't a SPIDER on the combine windshield, but there was a fly. |
| combine | oxytone | No, don't use a SPOON to combine the ingredients, use a beater. |
| compact | paroxytone | No, she doesn't OWN the compact car, she leases it. |
| compact | oxytone | No, she didn't use her FINGER to compact the soil, but she used her hand. |
| contract | paroxytone | No, HE didn't sign the contract today, but she did. |
| contract | oxytone | No, WE won't contract the disease, but you will. |
| convert | paroxytone | No, there weren't FIVE converts that year, but there were two. |
| convert | oxytone | No, they don't use MICROSOFT to convert files; they use Adobe. |
| convict | paroxytone | No, he didn't CATCH the convict yesterday, but he saw him. |
| convict | oxytone | No, the JUDGE will not convict him, but the jury will. |
| desert | paroxytone | No, she doesn't LOVE the desert landscape, she hates it. |
| desert | oxytone | No, the FATHER won't desert the army, but his son will. |
| discount | paroxytone | No, MONDAY isn't the discount day, but Friday is. |
| discount | oxytone | No, THEY don't discount his abilities, we do. |
| forebear | paroxytone | No, he didn't SHAME his forebear today, he honored him. |
| forbear | oxytone | No, the PARENTS won't forbear the mortgage, but the bank will. |
| import | paroxytone | No, the DRINKS weren't of great import to him, but the dinner was. |
| import | oxytone | No, THEY won't import more cars, but the government will. |
| impress | paroxytone | No, THEY don't want an impress of their hands, but the kids do. |
| impress | oxytone | No, his SINGING doesn't impress us, but his cooking does. |
| insight | paroxytone | No, I didn't want YOUR insight on the project, but his. |
| incite | oxytone | No, the LEADER doesn't incite violence, but his advisors do. |
| object | paroxytone | No, he didn't LOOK at the object on the shelf, he stared at it. |
| object | oxytone | No, THEY will not object to our methods, but he will. |
| permit | paroxytone | No, HER permit wasn't suspended, but mine was. |
| permit | oxytone | No, her FATHER won't permit her to go, but her mother will. |
| project | paroxytone | No, the SCIENCE project isn't due tomorrow, but the math project is. |
| project | oxytone | No, the PARENTS can't project their voices, but their son can. |
| record | paroxytone | No, they didn't BUY the record in the shop, but they listened to it. |
| record | oxytone | No, he doesn't HATE to record his voice, he loves it. |
| relay | paroxytone | No, I didn't RUN in the relay, but I walked. |
| relay | oxytone | No, my PARENTS won't relay my message, but my sister will. |
| subject | paroxytone | No, SHE wasn't the subject of the painting, but I was. |
| subject | oxytone | No, THEY won't subject you to questioning, but I will. |
| trusty | paroxytone | No, I didn't SELL my trusty gps, but I broke it. |
| trustee | oxytone | No, my MOTHER wasn't a trustee of the company, but my sister was. |

Table A.2: Sentences used to elicit English target words in a low pitch accent (post-focal) condition. Participants were told to emphasize the words written in all capital letters.

| | Paroxytones | Oxytones |
|---|---|---|
| abominar 'to hate' | abomina | abominar |
| denominar 'to name' | denomina | denominar |
| desanimar 'to discourage' | desanima | desanimar |
| descaminar 'to cheat' | descamina | descaminar |
| determinar 'to determine' | determina | determinar |
| discriminar 'to discriminate' | discrimina | discriminar |
| disseminar 'to spread' | dissemina | disseminar |
| eliminar 'to eliminate' | elimina | eliminar |
| encaminar 'to guide' | encamina | encaminar |
| examinar 'to examine' | examina | examinar |
| exterminar 'to kill' | extermina | exterminar |
| il·luminar 'to light' | il·lumina | il·luminar |
| incriminar 'to incriminate' | incrimina | incriminar |
| recriminar 'to reprimand' | recrimina | recriminar |

Table A.3: Experimental stimuli for Catalan production tasks, with stressed syllables underlined

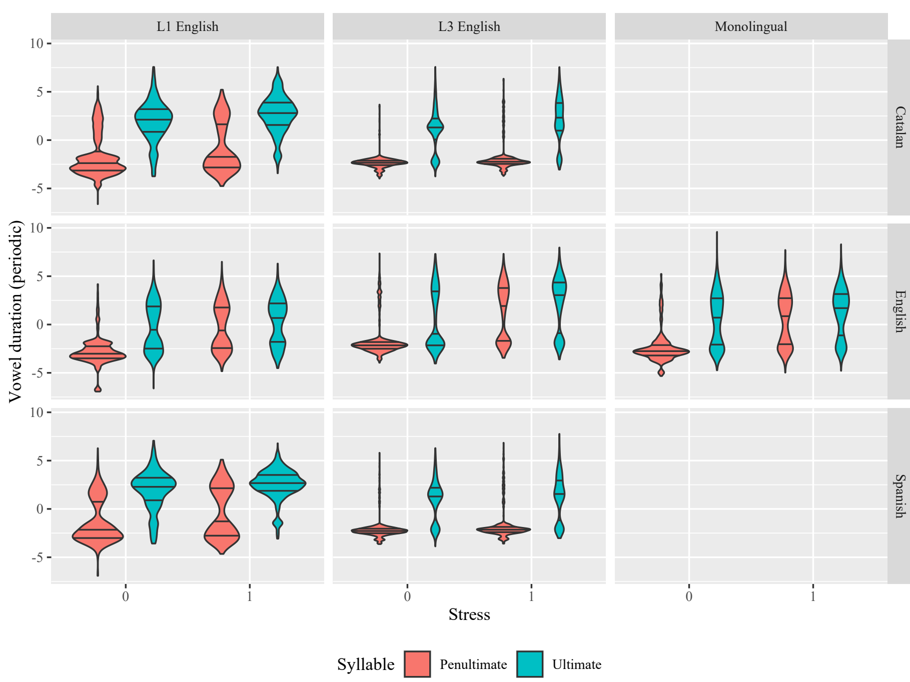



Figure A.3: Violin plots for vowel duration (periodic) across language, language profile, syllable, and stress.

| Verb | Stress | Sentence and gloss |
|---|---|---|
| abominar | paroxytone | "No m'agrada el te de camamilla" – abomina plena. *"I don't like chamomile tea," – she complains thoroughly.* |
| abominar | oxytone | "No m'agraden gens les curses de braus" – va abominar plena. *"I don't like bullfights at all," – she complained thoroughly.* |
| denominar | paroxytone | "Aquest gen es diu beta" – denomina concret. *"We'll call this gene beta," – he says concretely.* |
| denominar | oxytone | "Retrat d'un home trist" – va denominar concret. *"Portrait of a sad man," – he concretely named it.* |
| desanimar | paroxytone | "No passaré mai l'examen" – es desanima seriós. *"I'll never pass the exam," – he discourages himself seriously.* |
| desanimar | oxytone | "No puc cuinar sense cremar el menjar" – es va desanimar seriós. *"I can't cook without burning the food," – he discouraged himself seriously.* |
| descaminar | paroxytone | "Aquest no 'es el meu número de telèfon" – la descamina complagut. *"Here's my phone number," – he lies to her while smiling.* |
| descaminar | oxytone | "Aquest és el camí més curt" – la va descaminar complagut. *"This is the shortcut," – he lied to her while smiling.* |
| determinar | paroxytone | "La massa de l'àtom és mesurable" – determina complaguda. *"We can measure the mass of the atom," – she determines while smiling.* |
| determinar | oxytone | "Amb la pluja el vent és molt fort" – va determinar complaguda. *"With the rain the wind is very strong," – she determined while smiling.* |
| discriminar | paroxytone | "Aquest té menys sucre" – discrimina tímid. *"This one has less sugar," – he discriminates timidly.* |
| discriminar | oxytone | "Aquest és més fort" – va discriminar tímid. *"This one is stronger," – he discriminated timidly.* |
| disseminar | paroxytone | "La senyora Martí renuncia al seu càrrec" – dissemina prematura. *"Mrs. Martí has resigned," – she spreads the news prematurely.* |
| disseminar | oxytone | "El president de l'estat està malalt" – va disseminar prematura. *"The president of the state is unwell," – she spread the news prematurely.* |
| eliminar | paroxytone | "No són necessaris" – elimina segura. *"They aren't necessary," – she eliminates them with certainty.* |
| eliminar | oxytone | "Érem a casa a l'hora del crim" – va eliminar segura. *"We were at home at the time of the crime," – she narrows down with certainty.* |
| encaminar | paroxytone | "Confirmin la cita per demà" – les encamina tímida. *"Confirm the appointment for tomorrow," – she advises them timidly.* |
| encaminar | oxytone | "Llegiu aquests articles" – les va encaminar tímida. *"Read these articles," – she advised them timidly.* |
| examinar | paroxytone | "Escriguin la resposta" – els examina clarament. *"Write the answer," – he examines them clearly.* |
| examinar | oxytone | "Facin aquesta equació" – els va examinar clarament. *"Solve this equation," – he examined them clearly.* |
| exterminar | paroxytone | "No hi haurà més bestioles" –extermina complaguda. *"There won't be anymore bugs," – she exterminates happily.* |
| exterminar | oxytone | "Aquest verí va molt bé" – va exterminar complaguda. *"This insecticide works very well," – she exterminated happily.* |
| il·luminar | paroxytone | "Bon Nadal a tots" – il·lumina content. *"Merry Christmas to all," – he says while lighting it up.* |
| il·luminar | oxytone | "Acabo de comprar aquest llum" – va il·luminar content. *"I just bought this lamp," – he lit it contently.* |
| incriminar | paroxytone | "Vostès són les culpables" – les incrimina confiat. *"You are guilty," – he incriminates them confidently.* |
| incriminar | oxtyone | "Van entrar sense demanar permís" – les va incriminar confiat. *"They entered without asking permission," – he incriminated them confidently.* |
| recriminar | paroxytone | "No m'agrada que em truquis" – li recrimina seriosa. *"I don't like when you call me," – she reprimands him seriously.* |
| recriminar | oxytone | "És culpa teva" – li va recriminar seriosa. *"It's your fault," – she reprimanded him seriously.* |

Table A.4: Sentences used to elicit Catalan target words in low pitch accent (reporting clause).

| | Paroxytones | Oxytones |
|---|---|---|
| abominar ‘to hate’ | abomino | abominó |
| denominar ‘to name’ | denomino | denominó |
| desanimar ‘to discourage’ | desanimo | desanimó |
| descaminar ‘to cheat’ | descamino | descaminó |
| determinar ‘to determine’ | determino | determinó |
| discriminar ‘to discriminate’ | discrimino | discriminó |
| diseminar ‘to spread’ | disemino | diseminó |
| eliminar ‘to eliminate’ | elimino | eliminó |
| encaminar ‘to guide’ | encamino | encaminó |
| examinar ‘to examine’ | examino | examinó |
| exterminar ‘to kill’ | extermino | exterminó |
| iluminar ‘to light’ | ilumino | iluminó |
| incriminar ‘to incriminate’ | incrimino | incriminó |
| recriminar ‘to reprimand’ | recrimino | recriminó |

Table A.5: Experimental stimuli for Spanish production tasks, with stressed syllables underlined

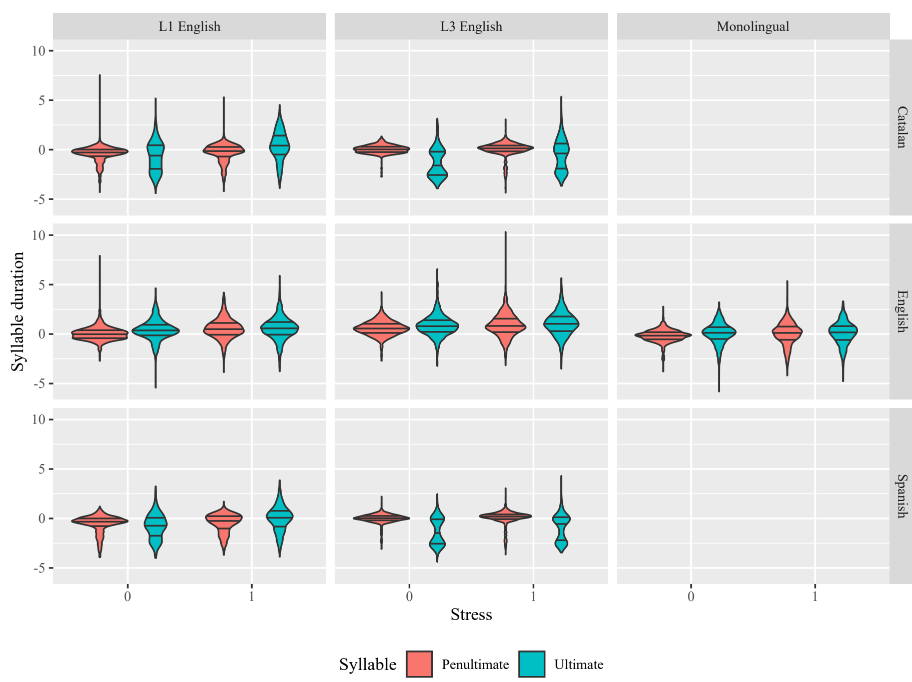



Figure A.4: Violin plots for syllable duration across language, language profile, syllable, and stress.

| Verb | Stress | Sentence and gloss |
|---|---|---|
| abominar | paroxytone | "No me gusta el té de manzanilla" – abomino con certeza. *"I don't like chamomile tea," – she complains with certainty.* |
| abominar | oxytone | "No me gustan las corridas" – abominó con certeza. *"I don't like bullfights," – she complained with certainty.* |
| denominar | paroxytone | "Se llamará beta" – lo denomino complacido. *"We'll call this beta," – he says pleased.* |
| denominar | oxytone | "Retrato de un hombre triste" – lo denominó complacido. *"Portrait of a sad man," – he concretely named it.* |
| desanimar | paroxytone | "No puedo estudiar matemáticas" – me desanimo compungida. *"I can't do math," – she discourages herself remorsefully.* |
| desanimar | oxytone | "No puedo cocinar sin quemar la comida" – me desanimó compungida. *"I can't cook without burning the food," – she discouraged herself remorsefully.* |
| descaminar | paroxytone | "Este no es mi número de teléfono" – la descamino complacido. *"This is not my phone number," – I lie to her happily.* |
| descaminar | oxytone | "Este es el camino más corto" – la descaminó complacido. *"This is the shortcut," – he lied to her happily.* |
| determinar | paroxytone | "La masa del átomo es medible" – determino complacida. *"We can measure the mass of the atom," – I determine happily.* |
| determinar | oxytone | "Con la lluvia el viento es muy fuerte" – determinó complacida. *"With the rain the wind is very strong," – she determined happily.* |
| discriminar | paroxytone | "Este tiene menos azúcar" – discrimino tímida. *"This one has less sugar," – I discriminate timidly.* |
| discriminar | oxytone | "Este es más fuerte" – discriminó tímida. *"This one is stronger," – she discriminated timidly.* |
| diseminar | paroxytone | "La directora Pérez renuncia a su cargo" – disemino compungido. *"Mrs. Pérez has resigned," – I spread the news remorsefully.* |
| diseminar | oxytone | "El presidente del estado se encuentra enfermo" – diseminó compungido. *"The president of the state is unwell," – he spread the news remorsefully.* |
| eliminar | paroxytone | "No son necesarios" – los elimino segura. *"They aren't necessary," – I eliminate them with certainty.* |
| eliminar | oxytone | "Tarjeta roja" – los eliminó segura. *"Red card," – she ejected them with certainty.* |
| encaminar | paroxytone | "Confirma la cita para mañana" – le encamino con ilusión. *"Confirm the appointment for tomorrow," – I advise him excitedly.* |
| encaminar | oxytone | "Lee estos artículos" – le encaminó con ilusión. *"Read these articles," – she advised excitedly.* |
| examinar | paroxytone | "Escriban la respuesta" – los examino seriamente. *"Write the answer," – I examine them clearly.* |
| examinar | oxytone | "Desarrollen esta ecuación" – los examinó seriamente. *"Solve this equation," – he examined them clearly.* |
| exterminar | paroxytone | 'No habrá más cucarachas" – extermino complacida. *"There won't be anymore cockroaches," – I exterminate happily.* |
| exterminar | oxytone | "Este insecticida es muy eficaz" – exterminó complacida. *"This insecticide works very well," – she exterminated happily.* |
| iluminar | paroxytone | "Estas son las luces nuevas" – ilumino contenta. *"These are the new lights," – I say happily while lighting them.* |
| iluminar | oxytone | "Compré esta lámpara ayer" – iluminó contenta. *"I bought this lamp yesterday," – she lit it contently.* |
| incriminar | paroxytone | "Los Osvald, culpables" – les incrimino seguro. *"The Osvalds are guilty," – I incriminate them confidently.* |
| incriminar | oxtyone | "Entraron sin pedir permiso" – les incriminó seguro. *"They entered without asking permission," – he incriminated them confidently.* |
| recriminar | paroxytone | "No me gusta su conducta" – les recrimino con intensidad. *"I don't like your behavior," – I reprimand them intensely.* |
| recriminar | oxytone | "Me robaron" – les recriminó con intensidad. *"They robbed me," – she reprimanded them intensely.* |

Table A.6: Sentences used to elicit Spanish target words in low pitch accent (reporting clause).

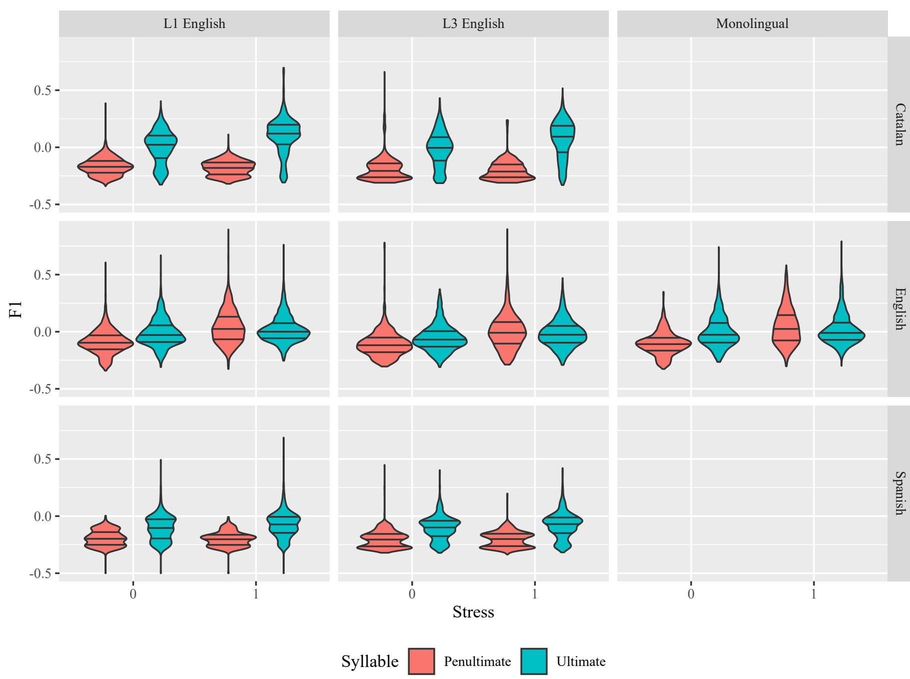



Figure A.5: Violin plots for centered and normalized F1 across language, language profile, syllable, and stress.

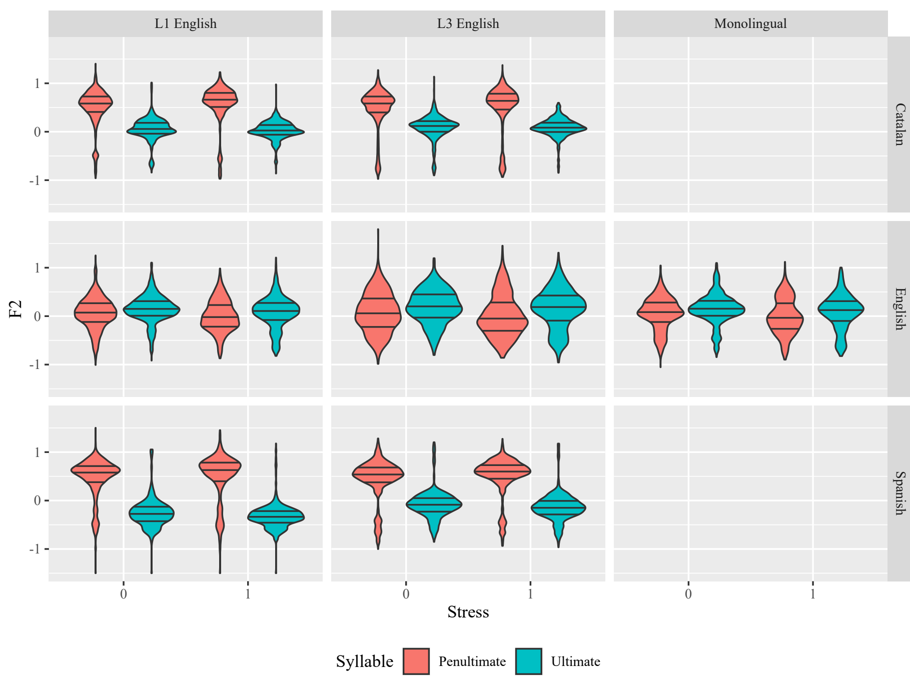



Figure A.6: Violin plots for centered and normalized F2 across language, language profile, syllable, and stress.

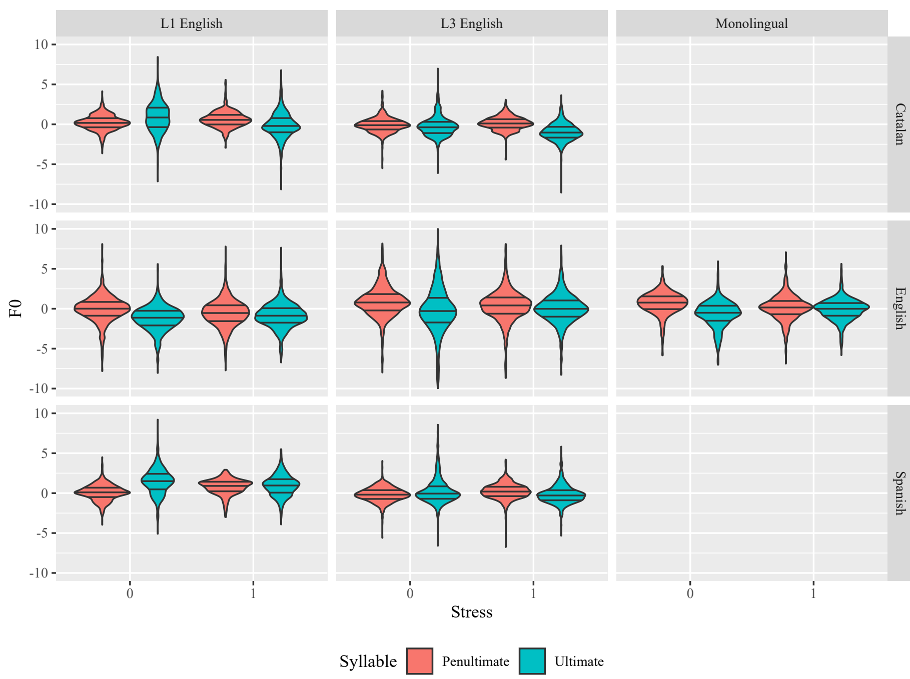



Figure A.7: Violin plots for F0 across language, language profile, syllable, and stress.

# Appendix B

# Chapter 4 materials

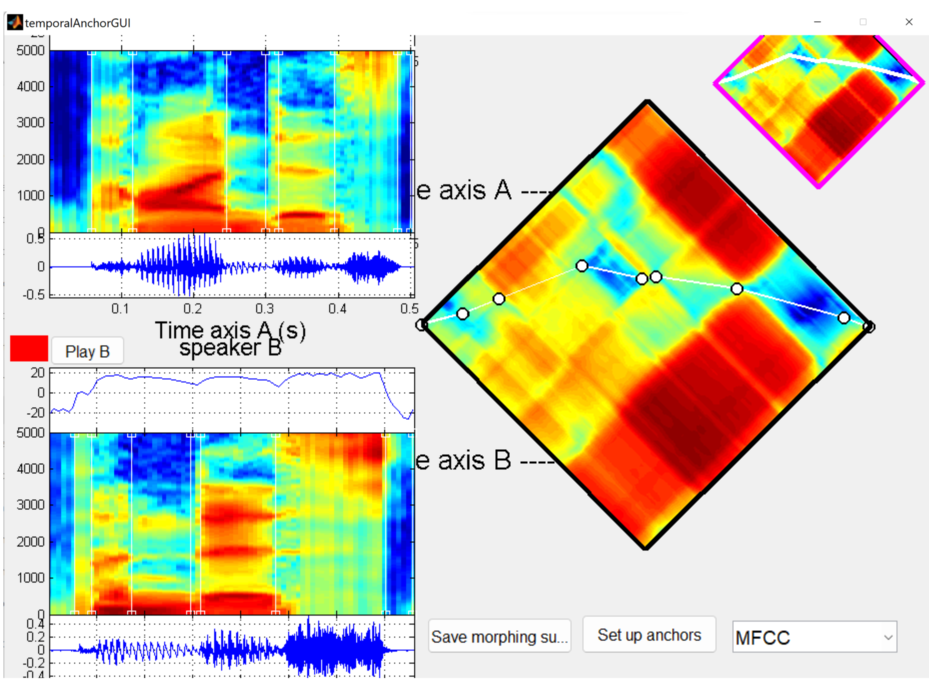
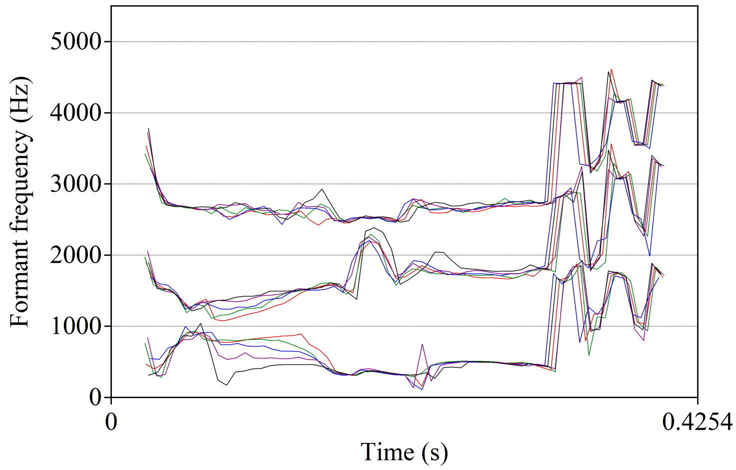
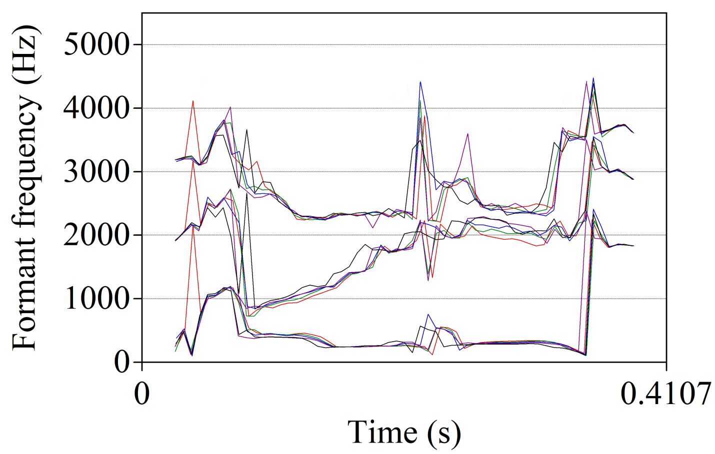



Figure B.1: TANDEM-STRAIGHT anchoring interface used to time-align similar segments for the endpoints in each language. The English stimuli are shown here.




Step 1 Step 2 Step 3 Step 4 Step 5

Figure B.2: Formant contours for English stimuli, where Step 1 is PONdiss and Step 5 is ponDISS. The biggest source of difference is in F1 and F2 during the first vowel.




Step 1 Step 2 Step 3 Step 4 Step 5

Figure B.3: Formant contours for Spanish stimuli, where Step 1 is PONdis and Step 5 is ponDIS. The biggest source of difference is in F2 during the second vowel.

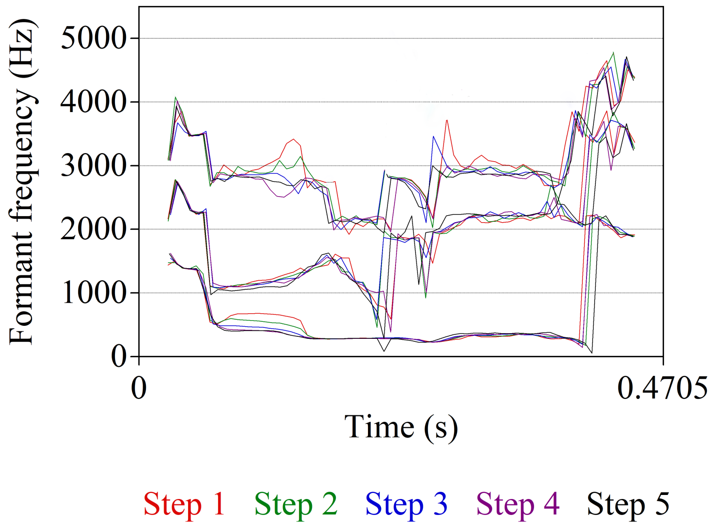



Figure B.4: Formant contours for Catalan stimuli, where Step 1 is PONdis and Step 5 is ponDIS. The biggest source of difference is in F1 during the first vowel.

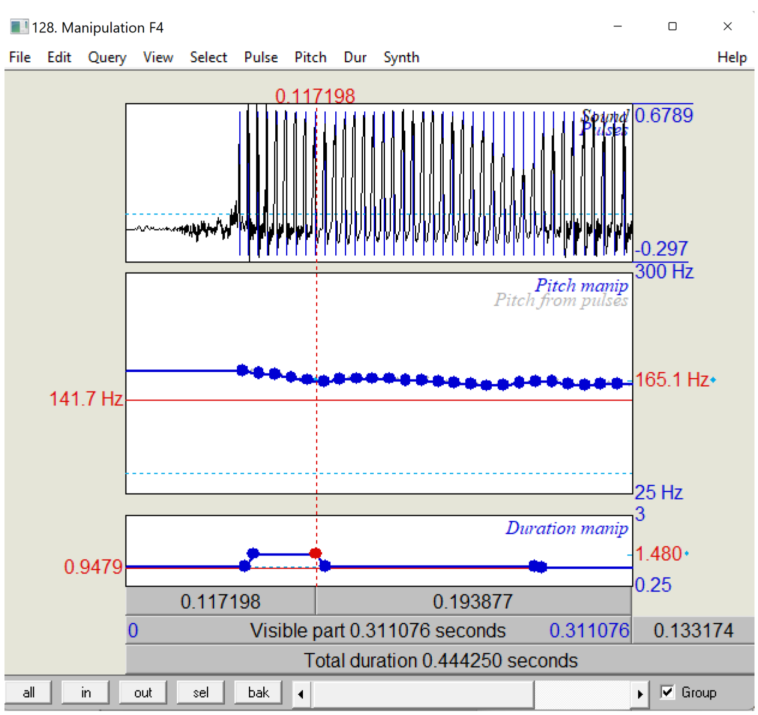



Figure B.5: Manipulation of duration using the PSOLA algorithm in Praat

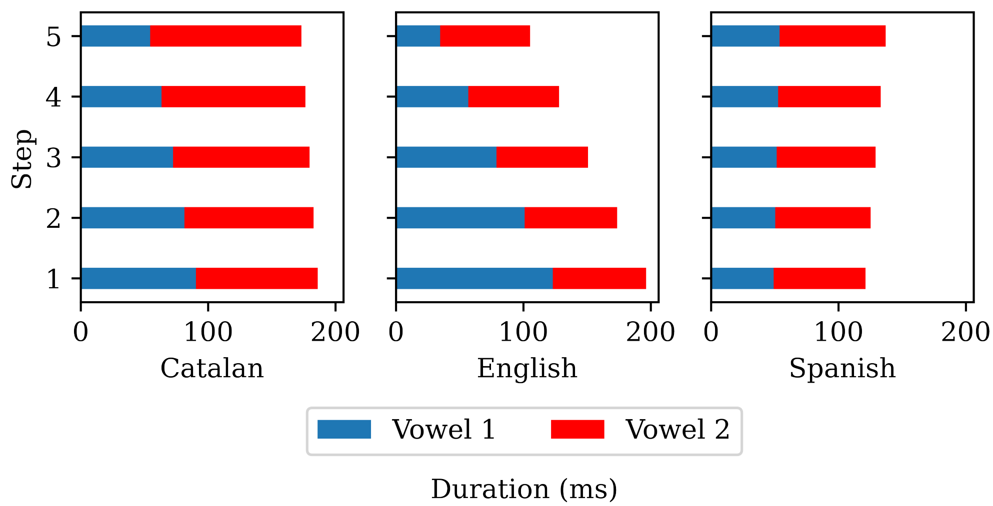



Figure B.6: Durations of syllables at each step, where Step 1 is stress-initial *PONdis* and Step 5 is stress-final *ponDIS*.

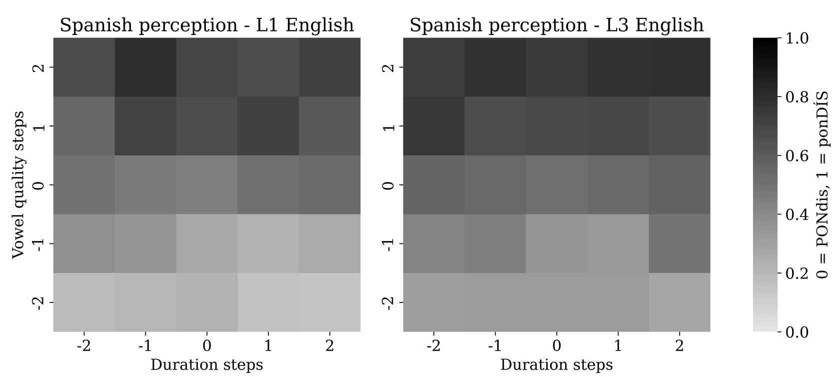



Figure B.7: Proportion of 'ponDIS' responses in the Spanish task across vowel quality and duration.

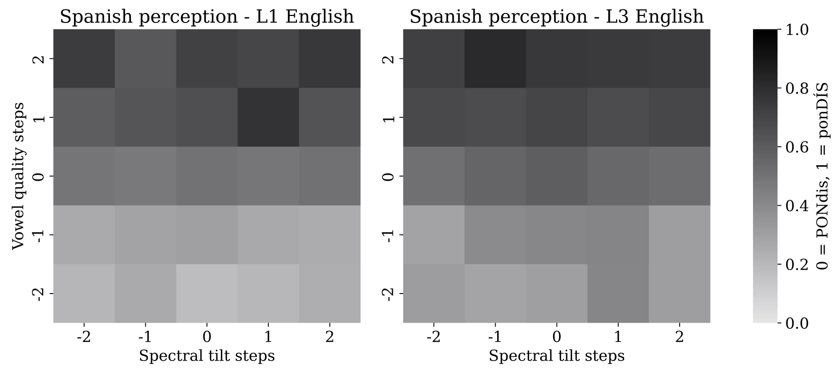



Figure B.8: Proportion of 'ponDIS' responses in the Spanish task across vowel quality and spectral tilt.

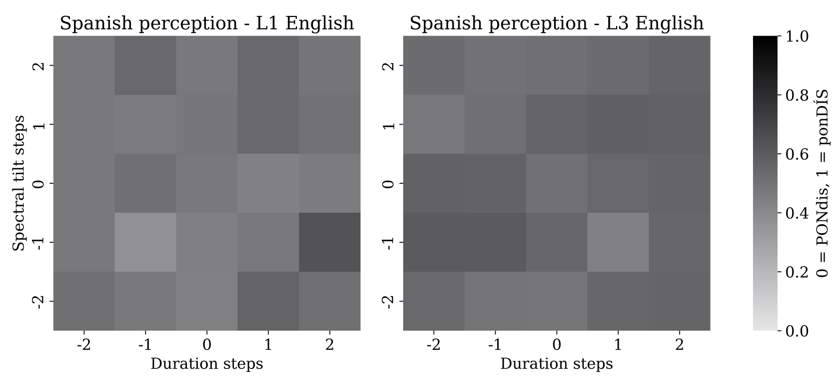



Figure B.9: Proportion of 'ponDIS' responses in the Spanish task across duration and spectral tilt.

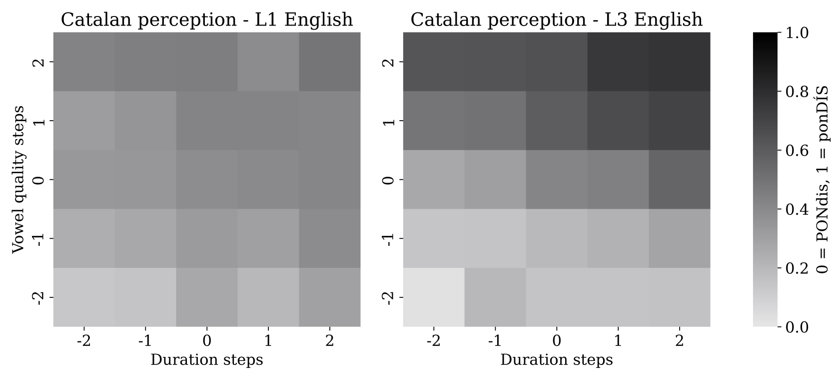



Figure B.10: Proportion of 'ponDIS' responses in the Catalan task across vowel quality and duration.

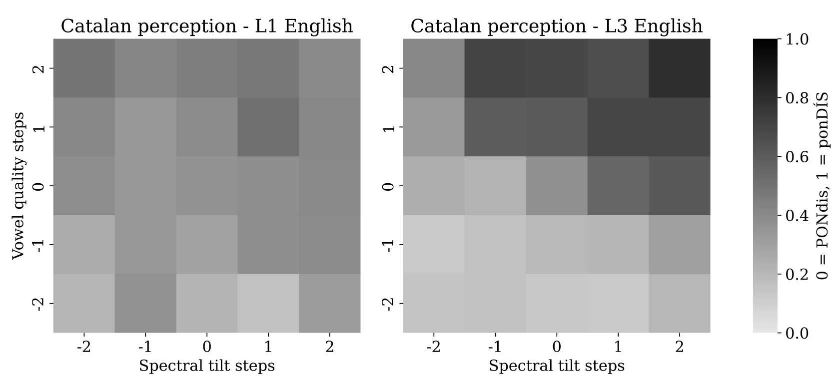



Figure B.11: Proportion of 'ponDIS' responses in the Catalan task across vowel quality and spectral tilt.

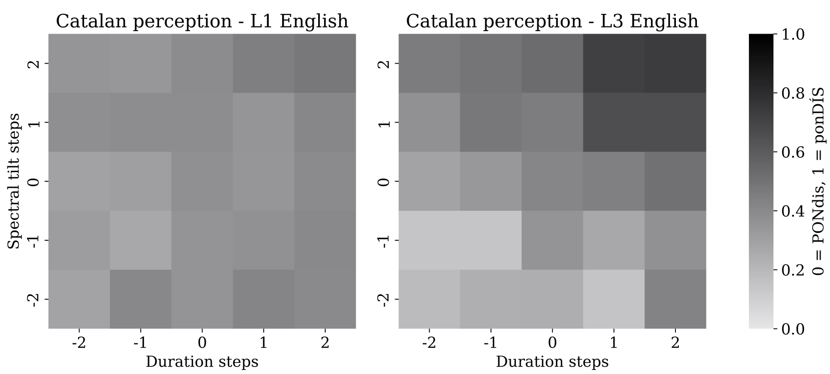



Figure B.12: Proportion of ‘ponDIS’ responses in the Catalan task across duration and spectral tilt.