CONTRAST PRESERVATION AND CONSTRAINTS ON INDIVIDUAL PHONETIC VARIATION

BY

CHRISTOPHER DAVID EAGER

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Spanish
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2017

Urbana, Illinois

Doctoral Committee:

Professor José Ignacio Hualde, Chair
Professor Jennifer S. Cole, Northwestern University
Associate Professor Anna María Escobar
Associate Professor Ryan K. Shosted
Dr. Joseph Roy

Abstract

Ferdinand de Saussure, one of the founders of modern Linguistics, described language as a system where everything holds together. Regarding the sounds of language, this has led to the current view that the phonology of a language consists of a complex system of relations between contrastive phonemes. In this dissertation, I test whether there are constraints on individual phonetic variation from a multivariate perspective due to this system of relations, and how these constraints interact with contrast preservation. Two main views of contrast preservation are considered. The first view is that contrast preservation is merely an outcome of other regular phonetic processes that affect multiple consonants simultaneously. The second view is that contrast preservation acts as a constraint on the phonetic realization of phonemes. To this end, two phonetic experiments are performed. In both experiments, multiple acoustic measures of intervocalic consonant strength are taken, and PCA is used for dimensionality reduction, resulting in measures of overall consonant strength. These measures are then analyzed with Bayesian linear mixed effects regression (using weakly informative priors and maximal random effects structures) in order to obtain distributional information about both populations and individual speakers.

In the first experiment, word-medial intervocalic /s/ and /f/ are compared for Valladolid Spanish and Barcelona Catalan. Both Catalan and Spanish have the fricatives /s/ and /f/, neither has /v/ contrasting with /f/, and only Catalan has /z/ contrasting with /s/. The results show that Catalan /s/ is stronger than Spanish /s/, but there is no evidence for a difference between the two language's /f/ strengths, with strong evidence that the magnitude of the difference between

Catalan and Spanish /s/ is larger than the magnitude of the difference between Catalan and Spanish /f/. I argue that these results are consistent with a role for contrast preservation as a constraint, with Catalan having stronger /s/ than Spanish because lenition of Catalan /s/ causes phonetic overlap with a contrasting phoneme, while lenition of Spanish /s/ does not.

In the second experiment, the simultaneous lenition of Spanish intervocalic /ptk/ and /bdg/ in three dialects (Cuzco, Peru; Lima, Peru; and Valladolid, Spain) is examined. Cuzco is found to have the strongest productions for both /ptk/ and /bdg/, Lima the weakest for both, and Valladolid in between for both. That is, the same hierarchy of strength applies in both cases, though the evidence for the difference between Valladolid and Lima /ptk/ is considerably weaker than the evidence for the other differences. I argue that the results are consistent with constraints on multivariate variation at the dialectal level, but that further research is required to see how constraints at the individual level relate to population differences.

Examining individual variation in both experiments, I find that the degree to which an individual speaker lenites /f/ is correlated with the degree to which they lenite /s/, and that the degree to which they lenite /ptk/ is correlated with both the degree to which they lenite /bdg/ and the degree to which they lenite /sf/. These correlations represent a significant constraint on individual phonetic variation from a multivariate perspective.

While a connection between individuals' /ptk/ and /bdg/ lenitions can be explained by both the constraint and outcome views of contrast preservation, the correlation between /sf/ and /ptk/ and the correlation between /s/ and /f/ lend support to the outcome view, and Catalan having stronger /s/ than Spanish but not stronger /f/ lends support to the constraint view. I argue for a framework in which acoustic lenition in a variety of intervocalic consonants may share a common articulatory source of lenition, giving rise to constraints on individual phonetic variation that may lead to contrast preservation as an outcome, but where there may additionally be a role for contrast preservation as a constraint. I conclude by discussing the importance of further acoustic studies that use the methodologies employed here, and studies that explore the articulatory and perceptual implications of the results.

Acknowledgments

Collection of the Peruvian speakers' data was supported by a fellowship from the Dorothea S. and Norman E. Whitten Endowment Fund of the Center for Latin American and Caribbean Studies at the University of Illinois at Urbana-Champaign. I'm very grateful to the members of my committee (José Ignacio Hualde, Jennifer Cole, Ryan Shosted, Anna María Escobar, and Joseph Roy) for their mentorship during my time here in Urbana; it is impossible to quantify how much I have learned by taking classes from and working with them. I'd also like to thank Mark Hasegawa-Johnson for advice on intensity measurements, and Juan María Garrido for giving me access to the Glissando corpus.

To my parents, Tim and Pam,

for always being supportive and believing in me.

And to my fiancé, Austin,

for putting up with me while I wrote this.

Table of Contents

Chapter 1. Introduction

1.1. Goal of this dissertation

Ferdinand de Saussure, one of the founders of modern Linguistics, described language as a system where everything holds together (Saussure, 1986[1916]). Regarding the sounds of language, this has led to the current view that the phonology of a language consists of a complex system of relations between contrastive phonemes. The goal of this dissertation to demonstrate that there are constraints on individual phonetic variation from a multivariate perspective due to this system of relations, and that these constraints interact with contrast preservation. In so doing, I also make methodological contributions to phonetics by increasing the degree to which measures are automated and employing dimensionality reduction on multivariate response variables, and to statistical modeling in linguistics through the use of Bayesian mixed effects regressions to obtain distributional information about individual speakers. These methodological contributions allow inference to be made on both populations and individuals using all of the information available in a dataset.

In this chapter, I begin in Section 1.2 with a discussion of differing views of contrast preservation (contrast preservation as a constraint and as an outcome) and why we might expect individual phonetic variation to be constrained from a multivariate perspective. In Section 1.3, I introduce the phonemic inventories for two related Western Romance languages, Spanish and Catalan, which serve as the languages of study in this dissertation. In Section 1.4, I describe the motivations for the first experiment, which compares Spanish and Catalan /s/ and /f/ lenition. In Section 1.5, I describe the motivations for the second experiment, which compares /ptk/ and

/bdg/ lenition in three dialects of Spanish. In Section 1.6, I give specific hypotheses concerning the results of these experiments.

In Chapter 2, I discuss the phonetic measurements of consonant strength taken as dependent variables, and the factors and covariates included as fixed effects in the analyses. I also discuss the following three phonetic methodological contributions I make in detail: (1) the automation of a duration measurement that can apply to any intervocalic consonant; (2) the treatment of zero-valued observations and non-zero-valued observations in a continuous fashion when a value of zero has meaning; and (3) the use of principal components analysis to reduce the dimensionality of multivariate dependent variables rather than running multiple regressions, which would inflate the rate of false discoveries.

In Chapter 3, I describe Bayesian mixed effects regressions, how they differ from the frequentist mixed effects regressions commonly used in linguistics, and how continuous variables and factors are treated in the regressions. I also make two major methodological contributions to the statistical analysis of linguistic data that I explain in detail: (1) the use of Bayesian mixed effects regressions to make probabilistic statements about individuals and the differences among them, rather than attempting to make currently widely-used methods fit with research questions they cannot adequately answer; and (2) the use of `NA` values to incorporate all available information in a dataset into a single regression. Mixed effects models have become widely used in Linguistics and other language sciences because they yield to better estimates for population effects (i.e. fixed effects) and lower Type I error rates when there are repeated measures on members of a grouping factor, with the differences between members considered as a noise component (e.g. Barr, Levy, Scheepers, & Tily, 2013). In this dissertation, there are hypotheses regarding population effects, but also hypotheses regarding individual variation from a multivariate perspective. To answer these questions, we need to employ a method that allows us to make probabilistic statements about individuals, not just the population. However, the standard frequentist approach to mixed effects models (e.g. `lme4`) does not allow us to make probabilistic statements about individuals in the same way that it allows us to make probabilistic statements about the population. While one may be tempted to simply run separate regressions on each individual speaker instead, this ignores that the speaker belongs to the population; better estimates are obtained by examining the random effects estimates in a mixed effects model in conjunction with the fixed effects. For this reason I use Bayesian mixed effects regression, which *does* allow such probabilistic statements to be made, representing a novel application of these models in Linguistics. That is, I model the individual variance as a noise component, but also examine these components as an object of inquiry. To obtain the best estimates possible for both populations and individuals, in addition to using maximal random effects structures (Barr et al., 2013), every attempt should be made to analyze all of the data together; binning data and running multiple regressions should be avoided to the greatest extent possible. To that end (in addition to treating 0-valued response variables as part of a continuum and using PCA as described in Section 2.5), I introduce the coding of factors as `NA` in subsets of the data where they are not contrastive. Provided that appropriate contrasts are used, these `NA` values can be set to 0 in the regression's model matrix, allowing all of the data to be analyzed in one regression, with the relevant factors only affecting the fitted values for observations in the subsets where they are contrastive.

In Chapter 4, I present the results for an experiment on Spanish and Catalan /s/ and /f/. I find that Catalan has stronger (i.e. longer, more voiceless) /s/ realizations than Spanish, but the languages do not differ in the strength of their /f/ productions. Spanish and Catalan both have the voiceless fricative phonemes /s/ and /f/, neither has the voiced fricative phoneme /v/, and Catalan (but not Spanish) has the voiced fricative phoneme /z/. In light of this, I argue that the results support a role for contrast preservation as a constraint.

In Chapter 5, I present the results for an experiment on /ptk/ and /bdg/ in the Spanish of Valladolid, Spain; Cuzco, Peru; and Lima, Peru. Both sets of plosive phonemes are subject to lenition in the intervocalic environment, with substantial variation across dialects. I find that the same dialectal hierarchy of strength obtains for both the voiceless and voiced plosives: Cuzco > Valladolid > Lima; however, the evidence for a difference between Valladolid and Lima /ptk/ is substantially weaker than for the other comparisons made. I argue that these results are consistent with the view that constraints on individual multivariate phonetic variation may give rise to constraints on dialectal multivariate phonetic variation, but only when overlap between speakers of the dialects is small enough.

In Chapter 6, I compare individual speakers' average consonant strengths from the two experiments. I find that individuals with relatively stronger /s/ also have relatively stronger /f/, that speakers with relatively stronger /ptk/ also have relatively stronger /bdg/, and that speakers with relatively stronger /sf/ also have relatively stronger /ptk/. Additionally, this relationship at the individual level is strongest for /ptk/ vs. /bdg/, weaker for /s/ vs. /f/, and weakest for /sf/ vs. /ptk/. I note that these relationship strengths reflect how phonetically similar the sets of consonants compared are, and argue that the results offer strong evidence for the existence of constraints on individual multivariate phonetic variation.

And in Chapter 7, I discus all of the results in relation to the hypotheses laid out in this chapter (Section 1.6). Because constraints exist on individual multivariate phonetic variation both when neutralization between the phonemes considered is plausible and when neutralization is implausible, I argue that contrast preservation in many circumstances can be explained as a result of other natural phonetic processes. However, given the population results for the Spanish and Catalan /s/ and /f/ experiment, I further argue that there is still a role for contrast preservation as a constraint within this framework.

1.2. Contrast preservation and constraints on individual phonetic variation

In this dissertation, I examine the phenomenon of contrast preservation, specifically in the context of intervocalic consonant lenition. I follow Lavoie (2001) in defining intervocalic consonant lenition as the phonetic process by which consonants become more similar to the surrounding vowels due to gestural overlap (e.g. Browman & Goldstein, 1986, 1992) and aerodynamic constraints on consonant voicing (e.g. Ohala & Riordan, 1979). In this framework, then, a consonant is stronger when it is more constricted and more voiceless.

The term *contrast preservation* is most often associated with the theory that the presence of phonological contrasts in a language conditions phonetic outcomes (e.g. if a language has /p/ but not /b/, the phonetic implementation of /p/ will be different than in a language with both /p/ and /b/). However, in this dissertation, I will refer to this theory as *contrast preservation as a constraint*, and use the term *contrast preservation* more generally as the opposite of neutralization; that is, regardless of why contrasting phonemes are not neutralized, I will describe a system of phonemic contrasts where neutralization is not occurring synchronically, or did not occur diachronically, as exhibiting contrast preservation. Historical examples of contrast preservation in this use of the term would thus include both chain shifts (Martinet, 1952), where contrasts are maintained, but their phonetic expressions change in tandem (e.g. /p b/ > /b β/), and also cases where a system of contrasts remains stable across time. Under this definition, we can then distinguish between phonological theories where contrast preservation is merely an outcome of purely phonetic processes and phonological theories where contrast preservation imposes a constraint on phonetic variation.

I will use the term *constraint* to describe a probabilistic pressure against a certain phonetic outcome. If the phonetic realization of a phoneme *A* is in part determined by the presence or absence of a phonetically similar contrastive phoneme *B*, then this would be contrast preservation acting as a constraint on *A* (e.g. if a language with both /p/ and /b/ voices /p/ less than a language with /p/ but not /b/, then contrast preservation constrains the phonetic realization of /p/). At the individual level, in a language with /p/, /b/, and /t/, if speakers with relatively stronger /p/ realizations also tend to have relatively stronger /b/ and /t/ realizations, this would represent a constraint on individual phonetic variation. That is, while speakers may vary widely in the strength of /p/, /b/, and /t/ considered individually, when the three are considered jointly, the speaker variation in the multivariate space may be probabilistically constrained. This use of the term *constraint* is thus different than the use of the term in other theoretical frameworks such as Optimality Theory, where constraints are violable and rankable, and theories that posit constraints that determine which outcomes are possible and which are not; as I use the term, constraints are always probabilistic.

In Sections 1.2.1 and 1.2.2, respectively, I discuss phonological theories where contrast preservation is merely an outcome of purely phonetic processes and phonological theories where contrast preservation imposes a constraint on phonetic variation. In Section 1.2.3, I discuss the predictions these two views make and how we can reasonably accept both of them at the same time. In Section 1.2.4, I argue that both of these perspectives imply constraints on individual phonetic variation, a possibility that is understudied in phonetics.

1.2.1. Contrast preservation as an outcome

We can explain both contrast preservation and neutralization as occurring due to the common articulatory gestures shared by the consonants in question.[1] In the case of neutralization, the explanation is fairly straightforward: two consonants that differ only in one or two respects become more similar to each other in those respects until they are indistinguishable; this can be explained by common phonetic reduction phenomena. In the case of contrast preservation, it is at first less obvious that a high degree of articulatory similarity could help preserve the contrast between two consonants. The reason that their similarity can lead to contrast preservation is that we find numerous examples of gestural differences and manipulations affecting multiple phonemes in a systematic way.

For example, Torreira and Ernestus (2011) compare intervocalic /ptk/ voicing and vowel devoicing between voiceless plosives in Spanish and French, and find that French has both less /ptk/ voicing and more vowel devoicing than Spanish. They argue that this may be indicative of a difference in the coarticulatory strategy for voiceless gestures in the two languages. In other words, perfect timing of consonantal closing gestures and voiceless gestures is unlikely in casual speech, but languages may differ in the direction in which they tend to overlap these gestures, and this affects multiple parts of the phonology.

Nielsen (2007, 2008, 2011) also finds evidence that manipulation of sub-phonemic features affects multiple parts of the phonology in an imitation study using the shadowing paradigm (Goldinger, 1998). Participants first read a list of words aloud, then heard a series of words, some of which started with /p/ and had artificially lengthened VOT, and repeated each word as quickly as possible upon hearing it (i.e. shadowing), and then re-read the series of

[1] The data in this dissertation are all acoustic, not articulatory. However, the discussion of articulation here offers a theoretical motivation for acoustic consequences we may expect to see in intervocalic consonant production, as discussed in Chapter 2.

words. The specific words that were shadowed were produced with longer VOT by the participants after the shadowing task than they were before; but other words that started with /p/ that were *not* shadowed were also produced with longer VOT, and so were words that started with /k/, even though *none* of the shadowed words had /k/ with lengthened VOT. Furthermore, the degree to which participants lengthened VOT for /k/ was the same as the degree to which they lengthened VOT for /p/, indicating that participants manipulated a phonetic gesture that impacted more than one phoneme simultaneously.

Still more evidence for systematic effects of changes in gestural coordination comes from articulatory studies on inter-speech postures (i.e. the neutral position of the articulators between utterances when preparing to speak). Research has shown that these postures are articulatory targets that differ between languages (Gick, Wilson, Koch, & Cook, 2005), that some bilingual speakers use different postures based on which language they are speaking, with speakers who only use one posture being perceived as less native-like in at least one of their languages (Wilson & Gick, 2014), and that these postures differ between read and spontaneous speech (Ramanarayanan et al., 2010). Thus some of the systematic acoustic differences that we see between registers, languages, and (perhaps) individuals could be due to a difference in these postures and how they interact with articulation once speech begins.

Given then that gestural differences can lead to systematic phonetic effects, it can be argued that contrast preservation occurs for much the same reason that neutralization occurs: because the sounds are similar. We might expect /p/ and /b/ to be ripe for neutralization (i.e. /p b/ > /b/ or /p b/ > /β/) because all that differentiates them is voicing, and intervocalic voicing of voiceless plosives is a common phenomenon motivated by gestural overlap. But we might for the same reason expect that they both lenite at the same time due to their sharing a bilabial closing gesture and maintain contrast (i.e. a chain shift /p b/ > /b β/) if what is really being lenited is the gesture they share, and the contrast between them is simply relative in this regard. Thus we can explain contrast preservation and neutralization in these cases from a purely phonetic perspective, motivated by well-known phenomena like gestural overlap and the aerodynamic voicing constraint (Browman & Goldstein, 1986, 1992; Gafos, 2002; Ohala & Riordan, 1979; Stevens, 2000, pp. 465–468). Whether neutralization or contrast preservation occurs in the presence of lenition, from this perspective, would be probabilistic.

1.2.2. Contrast preservation as a constraint

Contrast preservation as a constraint is central to phonological theories in which the phonetic realization of a phoneme is conditioned by the other phonemes in the language's inventory (Flemming, 2002; Lubowicz, 2003), though with a slightly different meaning for *constraint* than the one given in Section 1.2. In dispersion theory, Flemming (2002) proposes three constraints whose variable ranking play a role in the maintenance or loss of phonological contrasts: (1) a constraint that favors maximizing the number of phonological contrasts; (2) a constraint that favors maximizing the acoustic difference between contrasting phonemes; and (3) a constraint that favors minimizing articulatory effort. To see how these constraints may play out with intervocalic consonants, consider a language with a voicing contrast between plosives /p/ and /b/. Constraint (2) would disfavor voicing of intervocalic /p/, as this would lessen its acoustic distinction from /b/. Constraint (3) would favor voicing of intervocalic /p/, as this would lessen the articulatory effort for the production of /p/ (if we define effort as producing a /p/ that is maximally distinct from the surrounding vowels; i.e. a strongly articulated /p/ would have full lip closure and be entirely voiceless). Constraint (1) would favor maintaining the distinction between the two phonemes regardless of their phonetic realization. If the constraints were ranked (1, 2) > (3), we would expect /p/ and /b/ to maintain their canonical pronunciations; if they were ranked (3) > (2) > (1), we would expect /p/ and /b/ to neutralize as /b/ (or perhaps /β/); and if they were ranked (1) > (3) > (2), we would expect a chain shift, with /p b/ > /b β/. Another prediction of this view of contrast preservation is that a language with both /p/ and /b/, with proper constraint rankings, would voice intervocalic /p/ less than a language that had /p/ but not /b/.

Take, as another example, Kirchner's (1996) analysis of synchronic chain shifts in OT. The author argues that synchronic chain shifts can be explained with distance-based faithfulness constraints implemented through local conjunction. For example, to explain why /p/ ~ /b/ chain shifts to [b] ~ [β] rather than neutralizing to [β], we could posit constraints favoring lenition in voicing and constriction that are out-ranked by a conjoined constraint that disfavors lenition in both voicing and constriction at the same time.

In the constraint framework that I use in this dissertation, we could instead posit that contrast preservation acts as a probabilistic constraint on phonetic outcomes. That is, in general, the higher the functional load of a contrast, the more likely it is to be maintained in some fashion (whether this be through no lenition at all, or simultaneous lenition); but no contrast is entirely guaranteed to be preserved or lost. Some linguists have argued that the numerous examples of neutralization in the literature cast doubt on contrast preservation as a constraint that prevents lenition (Hock, 1991, pp. 150–151), arguing that it should only play a role as a repair strategy once lenition has begun and causes acoustic overlap (Hock, 1991, pp. 164–166). Others have argued that contrast preservation should not act as a constraint at all because it implies an improper teleological role for the speaker (Ohala, 1983). However, recent diachronic studies have shown that the vast majority of lenitions do not result in neutralization (Gurevich, 2004), and that the higher the functional load of a phonological contrast is, the less likely it is to be lost (Wedel, Kaplan, & Jackson, 2013). From a synchronic perspective, support for contrast preservation as a constraint comes from a second group of shadowing participants from the same experiment described in the previous section (Nielsen, 2007, 2008, 2011). The second group of participants shadowed stimuli that had *shortened* VOT for /p/ rather than lengthened VOT. While lengthening VOT for /p/ in English does not cause phonetic overlap with contrastive categories, shortening of VOT leads to overlap with /b/. For this group of shadowers, the author found that they did *not* imitate the shortened VOT of /p/, which the author argues is due to the presence of phonemic /b/. Thus, overall it seems that there may in fact be a role for contrast preservation as a constraint in phonology, and this role is probabilistic (i.e. we can still get neutralization, but the probability that we do is to some degree predictable).

1.2.3. Predictions of contrast preservation as a constraint and as an outcome

As outlined above, both views of contrast preservation (constraint and outcome) can account for known diachronic chain shifts, (unchanged) maintenance of contrasts, and neutralization. The theories are also not mutually exclusive in most respects (i.e. we can posit that the same mechanisms underlie multiple types of lenitions and still posit a role for contrast preservation as a constraint within this framework). One important prediction for which the two views of contrast preservation differ is with regards to the synchronic phonetic realizations of the same consonant in languages where lenition would result in neutralization and languages where it would not. Under the constraint view (or the intermediate view where contrast preservation can be merely an outcome, but also a constraint), a language with /s/ and /z/ would be less subject to intervocalic voicing of /s/ than a language that has /s/ but not /z/ in order to maintain the contrast. Under the outcome *only* view, the language that has both /s/ and /z/ would maintain the contrast specifically because its coarticulatory strategies do not lend themselves to the phonemes' neutralization. To tease these two possibilities apart, in the first experiment in this dissertation, Catalan and Spanish are compared for /s/ strength, and also /f/ strength. As described in Sections 1.3.2 and 1.4, Central Catalan has /s f z/ and Spanish has /s f/, with neither language having /v/, and there being no evidence that Central Catalan /s/ and /z/ are neutralizing. Under the constraint view, we should expect Catalan to have stronger /s/ than Spanish, but Catalan and Spanish /f/ to either not differ in strength, or for the magnitude of the difference for /s/ to be greater than that for /f/; under the outcome only view, we should expect Catalan to either have both stronger /s/ and /f/ than Spanish, or for the two languages to be the same for both fricatives.

1.2.4. Constraints on individual phonetic variation

Regardless of the perspective we take on neutralization and contrast preservation, we should expect to see constraints on individual variation from a multivariate perspective; that is, for example, we should expect speakers with relatively weaker /p/'s to have relatively weaker /b/'s in a gradient manner. This correlation, if present, would indicate that the full range of possible /p/ and /b/ variation is not found (i.e. when you plot speakers' /p/ strengths by their /b/ strengths, there are empty areas in the plot, showing a constraint on individual variation from a multivariate perspective).

Under the outcome *only* view of contrast preservation, we should expect these correlations for both chain shift situations and situations where neither a shift nor neutralization is occurring. In both cases, we simply extend the reasoning in Section 1.2.1 to individuals. That is, speakers who devoice vowels more ought to voice intervocalic /ptk/ less, as they involve the same gesture. In the case of a consonantal chain shift (or other simultaneous lenitions), lenition in the shared gestures of the consonants would affect both of them simultaneously; e.g. /p/ and /b/ are both bilabial consonants, and so a lenition of the bilabial closing gesture would affect both of them, and so individuals who lenite this gesture more would lenite both consonants more, not just one. This same reasoning applies to /s/ and /f/ lenition. If the same glottal voiceless gesture underlies both consonants (to some degree), then speaker variation in /s/ strength should be correlated with speaker variation in /f/ strength. We may even expect some degree of correlation between a speaker's /s/ strength and the same speaker's /p/ strength for similar reasons, but perhaps a weaker correlation due to differences in the pressures the aerodynamic voicing constraint applies to fricatives and plosives (Stevens, 2000, pp. 465–483).

Under the constraint view, in the context of chain shifts, the change must be simultaneous, such that speakers who participate to a greater degree in the first change A > B will also participate to a greater degree in the second change B > C, or else neutralization would occur (Carvalho, 2008; Gordon, 2013). Some evidence for such individual correlations have been found in studies on vowels: Gordon (2001) found some evidence of chain shifts at the individual level in a study on the Northern Cities vowel shift, but the results were not entirely consistent across individuals; and Langstrof (2006), in a study on the New Zealand front vowel shift, found that speakers with higher TRAP vowels also had significantly higher DRESS vowels and more centralized KIT vowels. This prediction of individual correlation can also be expected, for the same reasons, in cases of stable variation of simultaneous lenitions that result in contrast preservation.

Thus, constraints on individual variation from a multivariate perspective are predicted by both views of contrast preservation. If these constraints exist, understanding them is crucial to our understanding of phonology and phonetics, as language at its most fundamental level occurs in individuals; however, studies that test for the existence of these constraints are underrepresented in the literature. Correlation at the individual level (regardless of its theoretical motivation) may give rise to the same correlation at the dialectal or language level if the overlap between individuals from the different dialects or languages is small enough, but the correlation at the individual level may exist even if there is no evidence at the dialectal or language level. The opposite, however, does not have sound theoretical grounding: correlation at the dialectal or language level should not exist without correlation at the individual level. In the second experiment in this dissertation, the simultaneous lenition of intervocalic /ptk/ and /bdg/ in three dialects of Spanish is examined to test this possibility (described in Section 1.5). In the following sections, I describe some basic phonological properties of Spanish and Catalan, the two languages studied in this dissertation, and explain how dialectal, language, and individual comparisons in their intervocalic consonant lenitions can offer us insight into the role of contrast preservation and the existence of constraints on individual variation.

1.3. Spanish and Catalan

Spanish and Catalan are two related Western Romance languages; Spanish is spoken mainly in Spain and, since the 16th century, in the Americas; Catalan is spoken mainly in northeastern Spain, southern France, the Balearic Islands, Andorra, and Sardinia (Hualde, 2005, pp. 19–31; Penny, 2002, pp. 20–26; Wheeler, 2005, pp. 1–3). In this dissertation, data are taken from three dialects of Spanish: the standard Iberian dialect as spoken in Valladolid, Spain; Spanish as spoken in Lima, Peru; and the Andean dialect as spoken in Cuzco, Peru. The Catalan data are all from speakers of Central Catalan as spoken in Barcelona, Spain, where speakers of Catalan are universally bilingual in Spanish as well. I begin with a general description of the languages' phoneme inventories, and some important historical sound changes that gave rise to the main differences between them. In the case of Catalan, all phonemic descriptions should be understood to be for the standard Central Catalan dialect.

1.3.1. Vowels

Catalan has 8 vowel phonemes /i e ɛ ə a u o ɔ/; in stressed syllables, /ə/ does not occur, and in unstressed syllables, only /i ə u/ occur, with the stressed to unstressed correspondences given in Table 1.1. When an unstressed high vowel borders another vowel, it becomes a glide, allowing for diphthongs and triphthongs (Wheeler, 2005, pp. 24, 37–38, 54–55).

Table 1.1 Catalan vowel stress alternations.

<table>
<tr><th>Stressed</th><th>Unstressed</th></tr>
<tr><td>/i/</td><td>/i/</td></tr>
<tr><td>/e/</td><td rowspan="3">/ə/</td></tr>
<tr><td>/ɛ/</td></tr>
<tr><td>/a/</td></tr>
<tr><td>/u/</td><td rowspan="3">/u/</td></tr>
<tr><td>/o/</td></tr>
<tr><td>/ɔ/</td></tr>
</table>

Spanish has five vowel phonemes /i e a o u/ with contrastive syllabic stress (e.g. *hablo* /áblo/ 'I speak' versus *habló* /abló/ 's/he spoke'); vowels in unstressed syllables are shorter in duration, but do not centralize (Nadeu, 2014), and unstressed high vowels in contact with another vowel become glides to form diphthongs and triphthongs, with some exceptional hiatuses (Hualde, 2005, pp. 52–55). In the development of Spanish from Western Romance, which had the same vowel phonemes as Catalan has in stressed syllables, stressed /ɛ/ and /ɔ/ diphthongized to /ie/ and /ue/, respectively, and unstressed /ɛ/ and /ɔ/ neutralized with /e/ and /o/, respectively, giving the current vowel system (Penny, 2002, pp. 44–60).

1.3.2. Consonants

The consonant phonemes of Catalan and Spanish are given in Table 1.2 and Table 1.3, respectively (Hualde, 2005, p. 53; Wheeler, 2005, p. 11).

Table 1.2 Catalan consonant phonemes.

	Labial	Dental	Alveolar	Palatal	Velar
Plosive	p b	t d			k g
Affricate			ʦ ʣ	ʧ ʤ	
Fricative	f		s z	ʃ ʒ	
Nasal	m		n	ɲ	
Lateral			l	ʎ	
Tap			ɾ		
Trill			r		
Approximant				j	

Table 1.3 Spanish consonant phonemes.

	Labial	Dental	Alveolar	Palatal	Velar
Plosive	p b	t d			k g
Affricate				ʧ	
Fricative	f	(θ)	s		x
Nasal	m		n	ɲ	
Lateral			l	(ʎ)	
Tap			ɾ		
Trill			r		
Approximant				j	

The main difference between the two consonant inventories is that Catalan has a robust voicing distinction for sibilants, while Spanish does not. In Old Spanish, the sibilant system was much more similar to the system that Catalan has today, including /ʦ ʣ s z ʃ z/. This system underwent voicing neutralization in favor of the voiceless consonant, and the dental-alveolar affricate deaffricated, leaving a system with three phonemes /s̪ s ʃ/ (Penny, 2002, pp. 96–104). In the standard Castilian dialect (including Valladolid), dissimilation in place of articulation occurred, yielding the modern day system (/s̪ s ʃ/ > /θ s x/); in other varieties of Spanish, the distinction between /s̪/ and /s/ was lost, leaving only two phonemes, either /θ x/ (parts of Spain) or /s x/ (parts of Spain and all of Latin America, including Lima and Cuzco). Many Spanish dialects have also merged /ʎ/ and /ʝ/ as /ʝ/ (including Lima), with speakers in regions in contact with a language with phonemic /ʎ/ (including Cuzco) and some older monolingual speakers in Spain (including Valladolid) maintaining the distinction.

1.4. Intervocalic /s/ and /f/ voicing in Catalan and Spanish

Previous research has found that in Spanish, both intervocalic /s/ (Caravedo, 1990; A. Escobar, 1978; Hualde & Prieto, 2014; Torreblanca, 1976; Torreira & Ernestus, 2012) and /f/ (Blecua & Rost, 2013; Caravedo, 1990) can variably and gradiently voice in intervocalic position. In the case of /s/, Catalan has a voiced counterpart /z/ while Spanish does not, and in the case of /f/, neither language has a phonemically voiced counterpart.[2] To my knowledge, no study on the voicing of intervocalic Catalan /f/ has been done, and Hualde and Prieto (2014) conducted the only phonetic study of intervocalic voicing of Catalan /s/, and compared this with Spanish /s/, and also with Catalan /z/. The authors found that word-medial intervocalic /s/ is significantly more likely to be completely voiced in Spanish than in Catalan (5.9% and 2.7%, respectively), and that Spanish has shorter /s/ durations than Catalan. They also found that word-medial /z/ was at least partially devoiced in 40.6% of cases. They argue that these results are consistent with contrast preservation as a constraint; that is, because there is a phonemic opposition between /s/ and /z/ in Catalan, but not Spanish, Catalan has stronger /s/ productions than Spanish. However, as the authors note and as explained in Section 1.3.2, Modern Spanish /s/ is the result of neutralization between Old Spanish /s/ and /z/, while Catalan /s/ is historically more similar to Old Spanish /s/, and so it may simply be that when the Old Spanish fricatives neutralized, the result was somewhere in between the two original distinct phonemes in its

[2] Some dialects of Catalan, such as Majorcan, have /v/, but Central Catalan (the dialect under study) does not (Wheeler, 2005, p. 13).

phonetic realization. They argue that all of the matched voiceless consonants between the languages should be examined in order to test this (/ptksf/).[3]

In the first experiment in this dissertation, I take this framework and examine /s/ and /f/ in both languages, comparing Valladolid Spanish to Barcelona Catalan. While the comparison of Catalan and Spanish /ptk/ is certainly an interesting comparison, the most direct comparison we can make is limiting our scope to the fricatives /s/ and /f/ since, as Hualde and Prieto (2014) note, the aerodynamic voicing constraint (Ohala & Riordan, 1979; Stevens, 2000, pp. 465–468) behaves quite differently for fricatives and plosives. Gestural overlap (Browman & Goldstein, 1986, 1992) causes voiceless plosives to voice, and then constriction is difficult to maintain in the presence of voicing, leading to constriction weakening and spirantization. For fricatives, however, while this same gestural overlap exists motivating intervocalic fricative voicing, the maintenance of a constriction that creates turbulent airflow makes maintenance of voicing difficult, leading to devoicing or constriction weakening for voiced fricatives. So, while we might expect there to be some level of correlation between the voicing of phonemically voiceless intervocalic plosives and fricatives since they both involve the implementation of a voiceless gesture in a VCV sequence, we would expect the strongest correlation to occur for /s/ and /f/, which are the only two voiceless fricatives that the languages share. If contrast preservation is a constraint in its own right, we should expect Catalan /s/ to be more voiceless and longer in duration than Spanish /s/, but either no difference between the two languages for /f/, or a difference of smaller magnitude. If, on the other hand, contrast preservation is entirely an outcome of language-specific coarticulatory strategy, then we would expect Catalan to have both stronger /s/ and /f/ than Spanish. In either case, we should expect to see a correlation between

[3] The languages also share /ʧ/; however, /ʧ/ does not voice in either of the languages, except in Canary Island and Cuban Spanish, where it is pronounced as a palatal stop [c] which can voice intervocalically to [ɟ] (Hualde, 2005, p. 152; Trujillo, 1980)

individuals' mean /s/ strength and mean /f/ strength if the mechanisms underlying the individual variation in their strengths are the same.

In Catalan, the phonemic contrast between /s/ and /z/ only occurs only in onset position, and only word-initially or word-medially, with word-initial /z/ being exceedingly rare (Recasens, 1993). In word-final position, the two are neutralized, with voicing determined by the following context (Wheeler, 2005, pp. 162–164). In Spanish, coda /s/ and, in some dialects also word-final /s/ even when prevocalic, can variably aspirate or delete entirely (Hualde, 2005, pp. 161–165). Limiting the study to intervocalic fricatives, this makes the word-medial intervocalic position the best option for comparing Catalan /s/ to Spanish /s/. Similarly, in the historical development of Spanish, word-initial /f/ aspirated to /h/ before vowels (but not before liquids, the glide [w] in sequences like /fue/, and some words with glide [j]) and then eventually elided, making word-initial prevocalic /f/ uncommon in Spanish except in later borrowings from Latin (Penny, 2002, pp. 90–94, 103–104), again making the word-medial context the best option for comparison.[4] Additionally, while the aspiration and deletion of /s/ in coda and word-final prevocalic position acts as an overt sociolinguistic marker in many dialects of Spanish (see Lipski, 2011), and the degree of (de)voicing of Argentine Spanish /ʒ/ also acts as an overt sociolinguistic marker (Rohena-Madrazo, 2013), there are no reports that the voicing of word-medial intervocalic /s/ and /f/ in Spanish and Catalan does (i.e. phonetic variation in this context is below-the-radar).[5] For these reasons, I limit the fricative analysis to the word-medial intervocalic environment.

[4] For example, in the Spanish CREA corpus (Real Academia Española, 2016), word-initial prevocalic /s/ occurs 10,119,430 times, while word-initial prevocalic /f/ occurs 2,341,520 times.

[5] Davidson (2015) finds that voicing of word-final intervocalic /s/ in the Spanish of Catalan-Spanish bilinguals is covertly positively associated with solidarity with the Catalan-speaking community; but this has not been reported for word-medial position, where we should not expect it to be the case due to the presence of contrastive /z/ in this position, and it has not been reported to be recognized by monolingual speakers of Spanish.

1.5. Simultaneous lenition of /ptk/ and /bdg/ in Spanish

In Spanish, the phonemically voiceless plosives /ptk/ have been classically described as plain voiceless plosives (Navarro Tomás, 1977). However, recent acoustic studies on a variety of dialects have cast doubt on this characterization, finding that intervocalic /ptk/ may, with varying frequency and to varying degrees, voice and spirantize, making [ptk], [bdg], and [βðɣ] all possible pronunciations. This phenomenon has been documented in Spain (Hualde, Simonet, & Nadeu, 2011; Lewis, 2000, 2001; Machuca-Ayuso, 1997; Martínez-Celdrán, 2009; Munday, 2001; Torreblanca, 1976; Torreira & Ernestus, 2011), the Canary Islands (Oftedal, 1985; Trujillo, 1980), the Caribbean (Guitart, 1977), Chile (Figueroa, 2016; Poblete, 1992), Colombia (Lewis, 2000, 2001), and Lima, Peru (Caravedo, 1990; Lipski, 1994, pp. 321–322). This type of lenition is cross-linguistically common and has been argued to be phonetically motivated by gestural overlap and the aerodynamic voicing constraint (e.g. Browman & Goldstein, 1986, 1992; Ohala & Riordan, 1979).

Intervocalic /bdg/ are conventionally realized as approximants [βðɣ] in all dialects by an allophonic rule, and in most dialects also after glides and nonhomorganic consonants (Carrasco, 2008; Carrasco, Hualde, & Simonet, 2012; Hualde, 2005, p. 138). The degree of constriction of these approximants is highly variable, and can lead to complete elision (Carrasco, 2008; Carrasco et al., 2012; Cole, Hualde, & Iskarous, 1999; Eddington, 2011; Figueroa, 2016; Hualde, Simonet, et al., 2011; Ortega-Llebaria, 2004; Soler & Romero, 1999). Constriction weakening of intervocalic approximants is likely motivated by gestural overlap with the surrounding vowels (Browman & Goldstein, 1986, 1992).

While the allophonic description implies possible overlap between underlyingly voiced and voiceless plosives (e.g. *mido* 'I measure' /mído/ → [míðo] vs. *mito* 'myth' /míto/ → [míðo]), the simultaneous lenition of both sets of plosives may allow the contrast to be maintained in terms of constriction degree; that is, /d/ → [ð] appears to be less constricted than /t/→ [ð] (Hualde, Simonet, et al., 2011), and no dialect has been reported to have lost contrast between the sets. Additionally, there is substantial dialectal and individual variation in both lenitions (Carrasco, 2008; Cepeda, 1991; Lewis, 2000, 2001; Pérez, 2007 among others); but, just as for the voicing of intervocalic /s/ and /f/ described in Section 1.4, there are also no reports that intervocalic /ptk/ and /bdg/ lenition act as sociolinguistic markers, even though vowel devoicing bordering voiceless consonants and pause (which, as discussed below, should be correlated with /ptk/ voicing) does in some dialects (e.g. Delforge, 2009, 2012). This raises the question of whether the lenitions are connected; that is, does the degree to which a dialect lenites /ptk/ correlate with the degree to which the same dialect lenites /bdg/? This possibility also raises a second more fundamental question regarding the individual: does the degree to which a speaker lenites one set of plosives correlate with the degree to which they lenite the other? It is possible that the correlation exists at the level of the individual even if the overlap between dialects is great enough for dialectal differences to not show evidence of a correlation. Furthermore, given that previous research finds that, at the population level, the hierarchy of plosive strength by place of articulation differs by underlying voicing (i.e. /p/ > /t/ > /k/ but /b/ > /g/ > /d/; e.g. Hualde et al. (2011); Torreira and Ernestus (2011); among others), we may expect that the correlation at the individual level is not the same for the three places of articulation. As argued in Section 1.2.4, these predictions are consistent with both views of contrast preservation, but serve as an excellent test of whether there are constraints on individual multivariate variation.

To the author's knowledge, no study has tested for a connection between the two lenitions at the dialectal or individual level. The goal of the second experiment in this dissertation is to test whether this relationship exists in order to increase our understanding of lenition and variation at the dialectal and individual levels. To accomplish this, dialects that can be reasonably assumed to cover a wide range of the spectrum of lenition need to be examined. To that end, I analyze data from the Spanish spoken in Cuzco, Peru; Lima, Peru; and Valladolid, Spain.

Cuzco is located in the highlands of the Andean region of Peru and, having developed in contact with Quechua, has many Spanish-Quechua bilingual speakers (A. M. Escobar, 2011). Cuzco Quechua has three vowel phonemes /ɪ ʊ æ/, with /ɪ/ and /ʊ/ having lower allophones [e] and [o] in contact with uvular consonants (but not having phonemic status as do Spanish /e/ and /o/) and glide allophones [j] and [w] between vowels and in coda position, and the consonants given in Table 1.4 (Delforge, 2009; A. Escobar, 1978; Gallagher, 2010; Parker & Weber, 1996; Pérez-Silva, Palma, & Araujo, 2008):

Table 1.4 Cuzco Quechua consonant inventory. Parentheses indicate the coda allophone of the plosives at the same place of articulation.

	Bilabial	Alveolar	Palatal	Velar	Uvular	Glottal
Plosive / Affricate	p p’ pʰ	t t’ tʰ	ʧ ʧ’ ʧʰ	k k’ kʰ	q q’ qʰ	
Fricative	(ɸ)	s (θ)	(ʃ)	(x)	(χ)	h
Nasal	m	n	ɲ			
Lateral		l	ʎ			
Tap		ɾ				

Unlike Spanish, Cuzco Quechua has no voiced obstruents, but does have a three-way contrast in the voiceless plosive system between plain, aspirated, and ejective plosives in syllable onset with a restriction of no more than one non-plain stop per word, and neutralization to voiceless fricatives in coda (Gallagher, 2010; Parker & Weber, 1996). Delforge (2009, 2012) finds that unstressed vowels in both bilingual and monolingual Cuzco Spanish can devoice between voiceless consonants and before a pause. While, to the author's knowledge, there are no previous acoustic studies on Cuzco Spanish /ptk/ lenition, vowel-devoicing may imply a coarticulatory strategy inconsistent with high degrees of intervocalic /ptk/ voicing (e.g. Torreira and Ernestus' (2011) study comparing French and Spanish intervocalic /ptk/ voicing and vowel devoicing between two voiceless plosives, as discussed in Section 1.2.1), and so we cannot exclude the possibility that in the dialect's development, Quechua may have influenced Spanish /ptk/ (but see Hock (1991, pp. 481–485) for an argument against substrate influences). The voiced plosives /bdg/ have been described as resisting lenition in the Andean highlands of Peru (Lipski, 1994, pp. 319–320), but as Quechua has no voiced obstruents, there is not an analogous argument for substrate influence. Whereas Cuzco Spanish may have been influenced by Quechua in its historical development, the reason it is included in the study is because it appears to have particularly strong intervocalic plosives (especially when compared to Lima). The origin of this pattern is not important for the study as bilingualism is not directly relevant, and the speakers analyzed are either monolingual speakers of Spanish or Spanish-dominant Quechua bilingual speakers (as explained in further detail in Section 2.1.2). Overall, then, the literature suggests that Cuzco Spanish has a relatively lower degree lenition for both /ptk/ and /bdg/.

Lima is located in coastal Peru, with a different dialect than Cuzco (though the two have had increased contact in recent decades; see A. Escobar (1977)). Caravedo (1990) reports in an impressionistic study both voiced approximant realizations of /ptk/ and elided /bdg/ (with /d/ having the greatest rate of elision), and Lipski (1994, pp. 321–322) also notes frequent elision of both /b/ and /d/, making the Lima dialect likely to represent a relatively greater degree of lenition.

Valladolid, located in northwestern Spain, is considered to have the standard Castilian dialect, with realizations of intervocalic /bdg/ that vary from an approximant with considerable constriction to elision, with /d/ being especially susceptible to lenition (Williams, 1987). Munday (2001) found that /ptk/ can be partially or fully voiced in conversation, but are more resistant to voicing in more formal styles of speech. Valladolid may thus have a more moderate degree of lenition, somewhere between Cuzco and Lima.

Phonemically, the only differences between the dialects of Cuzco, Lima, and Valladolid is that Valladolid maintains the distinction between /θ/ and /s/ while Lima and Cuzco do not, and that Cuzco and (variably) Valladolid maintain the distinction between /ʎ/ and /ʝ/ while Lima does not (Hualde, 2005, pp. 19–31; Lipski, 1994, pp. 319–322).[6] The social pressures on lenition in these three dialects may well not be the same (as is the case whenever dialects are compared), but they all maintain the same phonemic contrast in plosive voicing between /ptk/ and /bdg/ in intervocalic position, and the question the second experiment seeks to answer is whether the degree to which one set is lenited is correlated with the other set's degree of lenition.

1.6. Hypotheses

If contrast preservation acts as a constraint, we should expect correlations at the individual level when neutralization is plausible but not occurring (i.e. Spanish /ptk/ and /bdg/ strength), but not necessarily in cases where neutralization is not plausible (i.e. /s/ and /f/ strength; /sf/ and /ptk/ strength); and we should expect the presence of phonetically similar phonologically contrastive phonemes to condition phonetic outcomes (i.e. Spanish /s/ weaker than Catalan /s/, but no difference for Spanish and Catalan /f/). If contrast preservation can occur as an outcome of entirely phonetic processes (i.e. individual correlations result in contrasts not being lost), then we should find these correlations not only when neutralization is plausible, but

[6] While the distinction between /ʎ/ and /ʝ/ may be maintained in Valladolid for some older speakers, Scarpace, Beery, and Hualde (2015) found no evidence for a contrast between the phonemes in the same corpus from which the Valladolid speakers in this dissertation are taken.

also when phonemes not at risk for neutralization share phonetic features and articulatory gestures in common. Based on the hypothesis that contrast preservation can occur as an outcome, but that there is also at least some role for contrast preservation as a constraint on top of this, I have the following concrete hypotheses concerning the data.

Hypothesis 1. Catalan /s/ will be stronger than (Valladolid) Spanish /s/, but Catalan /f/ will not be stronger than (Valladolid) Spanish /f/. I expect this outcome from the perspective that the existence or absence of phonological contrast can play an independent role in phonetic outcomes, and previous research that finds that Catalan has stronger /s/ productions than Spanish (Hualde & Prieto, 2014). Both Spanish and Catalan have /s/ and /f/, neither has /v/ contrasting with /f/, and Catalan has /z/ contrasting with /s/ but Spanish does not. If the presence of a contrastive phoneme can condition the phonetic realization of the phonemes with which it contrasts, then we should expect that Catalan /s/ is stronger than Spanish /s/ due to the presence of Catalan /z/, but that this pressure should not exist for /f/. If there is no role for contrast preservation as a constraint, we should expect Catalan /s/ and Catalan /f/ to either both be equally stronger than Spanish /s/ and /f/, or for the two to be the same in strength in both languages.

Hypothesis 2. In Spanish, for both the voiceless plosives /ptk/ and the voiced plosives /bdg/, the dialects will show the same strength hierarchy of Cuzco > Valladolid > Lima. I expect this outcome based on the hypothesis that /ptk/ lenition and /bdg/ lenition are connected. We should expect dialects that have stronger /ptk/ to also have stronger /bdg/ as well, and previous research on these three dialects imply the most likely hierarchy is Cuzco > Valladolid > Lima. This result is expected regardless of whether contrast preservation acts as a constraint or not.

Hypothesis 3. The degree to which individual speakers of Catalan and Spanish lenite /s/ will correlate with the degree to which the same speakers lenite /f/. I expect this outcome based on previous research that shows that changes in a gesture can affect multiple phonemes that both use the gesture. The contrast preservation as an outcome view implies that such relationships between phonemes exist regardless of whether or not neutralization of the phonemes is plausible. Thus, while /s/ and /f/ are not at risk of neutralization, they do share common glottal gestures, and so we should expect individuals who produce relatively stronger (i.e. longer, more voiceless) /s/ to also produce stronger /f/ if there are constraints on individual phonetic variation from a multivariate perspective.

Hypothesis 4. In Spanish, the degree to which individual speakers lenite /ptk/ will correlate with the degree to which the same speakers lenite /bdg/, and this relationship may differ by place of articulation. I expect this outcome for the same reasons as Hypothesis 2. The simultaneous lenition of /ptk/ and /bdg/ is predicted to occur under both views of contrast preservation at the individual level in a gradient, correlated manner, regardless of whether this is to prevent them from neutralizing, because they involve common gestures, or both.

Hypothesis 5. The degree to which Valladolid speakers lenite voiceless fricatives /sf/ will correlate with the degree to which they lenite voiceless plosives /ptk/. While the consonants /sf/ are less similar to /ptk/ than the comparisons in Hypotheses 3 and 4, we should still expect some degree of correlation between the two because they are both voiceless consonants. That is, I expect some degree of individual variation in intervocalic /ptk/ lenition to be due to speakers' having an overall tendency to voice (or not) intervocalic voiceless consonants in general.

Chapter 2. Phonetic methods

2.1. Datasets

2.1.1. Recording conditions

The Spanish data for Valladolid and Catalan data for Barcelona are taken from the Glissando corpus of Spanish and Catalan speech (Garrido et al., 2013). Recordings were made on Marantz PMD670W1B and PMD560 recorders, using a Mackie CR1604-VLZ mixer, at sampling frequency of 44 kHz, with the participants wearing a headset wireless Senheisser EW100-G2 microphone (Garrido et al., 2013).[7] The Spanish data from Cuzco and Lima were collected by the author in Peru in the summer of 2014. Recordings were made using a ZOOM H4n recorder and AKG C520 head-worn unidirectional condenser microphone placed approximately 2 cm from the right corner of the participant's mouth at a sampling rate of 44 kHz. Recordings were conducted at the participants' place of work, home, or study before the work-day started, which was always a quiet room with no echo. While no difference in quality between the recordings from the three dialects was noticed by the author during segmentation, to account for possible small differences in the level of background noise, intensity measurements were computed subtracting out the RMS amplitude from a silent interval in each recording to ensure their comparability (see Section 2.4).

[7] The Glissando recordings took place with both speakers wearing a head-worn microphone and additionally with a stationary microphone between them; the head-worn microphone recordings are used to ensure valid intensity measurements and comparability with the Cuzco and Lima recordings.

2.1.2. Speaker demographics and tasks

The Glissando corpus contains read speech from 8 trained speakers for both Spanish and Catalan (radio news broadcasters and advertising professionals), as well as spontaneous speech from these professional speakers and 20 non-professional speakers for both Spanish and Catalan (all university students, balanced for sex) participating in task-oriented dialogues and informal dialogues, for a total of 56 speakers. In this study, I use data from the non-professional speakers' task-oriented dialogues. The task-oriented dialogues consist of three subjects for each pair of speakers: travel, transportation and university information. For example, in the university information dialogue, one speaker plays the role of a university administrator and the other speaker plays the role of a student requesting information about taking courses abroad. The corpus has been force-aligned at the word, phoneme, and syllable level, but the alignment is too imperfect for analysis without manual correction. In the case of one set of task-oriented dialogues in the Spanish sub-corpus (representing the data for one male and one female speaker), the forced alignment was too imperfect to reliably locate segments of interest in the audio signal, and so these speakers' data were removed, resulting in 18 Spanish speakers in the Valladolid dialect, and 20 Catalan speakers from Barcelona, balanced for sex.

The data for Cuzco and Lima are taken from a read speech task and informal interviews conducted by the author, with participants classified demographically based on a questionnaire (see Section B.1) filled out prior to recording.[8] For the Lima dialect, there were 8 speakers (4 female), all of whom were monolingual speakers of Spanish born and raised in Lima with parents who were also monolingual speakers of Spanish, and all were studying at Pontífica Universidad del Perú. Lima is home to multiple dialects of Spanish (e.g. A. Escobar, 1977), and

[8] The Valladolid task-oriented dialogues and Cuzco and Lima interviews are not exactly the same in their level of task formality, but do both represent spontaneous speech, and so for the purposes of this dissertation are considered comparable tasks.

so I make no claim as to any specific dialect these speakers belong to. Rather, I will simply refer to these speakers being representative of the varieties spoken in Lima as the ‘Lima dialect’. There were 30 speakers for the Cuzco dialect, split between monolingual speakers of Spanish (16) and Spanish-dominant Quechua-bilingual speakers (14). The bilingual speakers were classified as Spanish-dominant based on responding that their parents spoke more Spanish than Quechua at home, that they attended school in Spanish, and that they currently used more Spanish than Quechua both at home and at work. The monolingual group was balanced for sex, education level (university or secondary), and two age groups (older than 40 years of age or 40 years of age or younger), with two speakers for each combination of social factors. The bilingual speakers were balanced for these same criteria except there was only one older male speaker who attended university and only one older female speaker who did not attend university (see Table B.1 for Cuzco participant demographics).

For the read speech task, words containing word-medial intervocalic /ptk/ and /bdg/ were placed into meaningful sentences (three words for each combination of phoneme and three stress conditions: in the onset of a stressed syllable, henceforth “tonic”; in the onset of a post-tonic syllable, henceforth “post-tonic”; or in between two unstressed vowels, henceforth “unstressed”) for a total of 54 planned observations per speaker, and 2,052 planned observations total (see Section B.2 for a list of the sentences and planned observations). For example, the sentence *El doctor administró la medicina al paciente* ‘The doctor gave the medicine to the patient’ contains a planned observation of unstressed intervocalic /d/ in the word *me**d**icina* ‘medicine.’ The task was to simply read each sentence aloud two times, and the second reading was analyzed. In the informal interviews, I simply asked open-ended questions about the participants’ work, studies, etc.

2.1.3. *Data distribution*

2.1.3.1. Spanish /ptk/ and /bdg/

For the read speech task in Cuzco and Lima, all planned observations of /ptk/ and /bdg/ were segmented. Because there were tens of thousands of possible /ptk/ and /bdg/ observations in the spontaneous speech data (Valladolid task-oriented dialogues and Cuzco and Lima interviews) and all observations required manual segmentation, stratified random sampling of the possible observations obtained by crawling the forced-aligned Valladolid TextGrids and my transcripts of the Cuzco and Lima interviews was performed.[9] As the planned observations in the read speech task were balanced for each of the three stress conditions described in Section 2.1.2 (in addition to plosive phoneme identity), the spontaneous speech sampling was stratified by plosive phoneme identity and stress condition. An R script written by the author was used to take a stratified random sampling of the content words in the list of possible spontaneous speech observations, taking 5 instances in each of the resulting 18 combinations of phone and stress (unless fewer than 5 plosives were available for a given combination), for an initial maximum of 90 plosives per speaker balanced across plosive phoneme and stress (6 plosive phonemes x 3 stress conditions x 5 tokens). The distribution of these randomly sampled plosives with respect to the other fixed effect factors included in the regression (described in Section 2.3.2.2) was examined, and when there were low cell counts for a speaker, additional observations were sampled. The suffix /-ado/ was not used, since it is conventionally deleted in some dialects, and the suffixes /-iko/ and /-igo/ were not used, because they may be neutralized for some speakers

[9] With simple random sampling, imbalances in the distribution of the consonants of interest would be reflected in the distribution of the randomly sampled consonants for analysis. With stratified random sampling, the possible observations are binned based on the category they belong to, and then within each category, the same number of observations is randomly sampled, leading to better balance across categories in the sample for analysis. Also note that because the observations are chosen randomly, they are equally likely to come from all parts of the recording (i.e. the analyzed observations are not biased towards consonants which occur early or late in the recordings).

(Hualde, Simonet, et al., 2011). Voiced plosives following a glide were also not used, since some dialects have fortition in this environment (Carrasco, 2008; Carrasco et al., 2012). As the only plosive that occurs word-finally in non-loan words with any frequency is /d/, and its allophony is different from word-initial and word-medial /d/ (Hualde & Eager, 2016), only word-initial and word-medial positions were considered.

Tokens that occurred bordering voiceless vowels or that preceded a vowel with creaky voice were excluded, as relative intensity measures and voicing measures cannot be reliably obtained in these cases. This resulted in a total of 5281 Spanish plosives for analysis, with distribution by phoneme, dialect, and task given in Table 2.1. The analysis of these 5281 Spanish /ptk/ and /bdg/ observations will be referred to simply as the "plosive analysis" throughout.

Table 2.1 Distribution of Spanish plosives for analysis. Numbers in parentheses indicate the number of speakers for the dialect.

Phoneme	Valladolid (18)	Cuzco (30)		Lima (8)	
		Read	Spontaneous	Read	Spontaneous
/p/	228	261	179	72	56
/t/	275	266	223	71	67
/k/	248	268	218	72	83
/b/	258	267	281	72	78
/d/	293	270	253	72	72
/g/	245	262	124	66	81

2.1.3.2. Spanish and Catalan /s/ and /f/

For Spanish and Catalan /s/ and /f/, the task-oriented dialogues for Valladolid and Barcelona were used. Because, as discussed in Section 1.4, word-initial /f/ is rare in Spanish, only content words that contained word-medial intervocalic /s/ and /f/ were considered. The forced-aligned TextGrids were crawled to create a list of observations meeting these criteria, and all observations were considered.[10] As for the plosives, tokens that occurred bordering voiceless vowels or that preceded a vowel with creaky voice were excluded. This resulted in a total of 2163 observations for analysis, with counts by language and phoneme given in Table 2.2. The analysis of these 2163 Spanish and Catalan /s/ and /f/ observations will be referred to simply as the "fricative analysis" throughout.

Table 2.2 Distribution of Spanish and Catalan fricatives for analysis. Numbers in parentheses indicate the number of speakers for the language.

Phoneme	Valladolid Spanish (18)	Barcelona Catalan (20)
/f/	165	269
/s/	765	964

2.2. Dependent variables: acoustic measures of plosive and fricative strength

2.2.1. Intensity

Several related relative intensity measures of consonant constriction have been used in previous studies on Spanish plosive lenition: the difference between the minimum intensity in the consonant and the maximum in the following vowel, the ratio of the consonant minimum and vowel maximum, and the maximum velocity of the rise in intensity from the consonant minimum to the vowel maximum. However, as one might expect, these measures are all highly positively correlated ($r > 0.9$; Hualde, Simonet, et al., 2011), and so their analysis needs to be carried out in a principled way that does not inflate the rate of false positive discoveries (discussed in detail in Section 2.5.2). In this study, the difference between the minimum intensity of the consonant and the maximum intensity in the following vowel measured in dB (henceforth "intensity difference") and maximum consonant to vowel intensity velocity measured in dB/ms (henceforth "intensity velocity") are used (Carrasco et al., 2012; Hualde,

[10] Foreign toponyms in the tourist dialogue were excluded.

Shosted, & Scarpace, 2011; Hualde, Simonet, et al., 2011; Hualde, Simonet, Shosted, & Nadeu, 2010; Parrell, 2011; Recasens, 2015; among others), with higher values indicating a more constricted consonant. It should also be noted that both of these measures involve taking the difference of two logs, meaning that they represent the log of the ratio of the corresponding RMS values. Intensity difference and velocity were taken for plosives /ptk/ and /bdg/, but not for fricatives /s/ and /f/, as in the case of fricatives, the minimum intensity of the consonant would get higher with stronger constriction, and be very different for /s/ and /f/, as the former is a sibilant and the latter is not; additionally, the hypotheses for the fricative analysis are specific to voicing.

2.2.2. Duration

Previous studies have measured plosive duration based on acoustic landmarks associated with closure and release, such as the absence of F2 or presence of a burst, with these landmarks not reliably available for approximants, and fricative duration based on the onset and offset of aperiodic noise in the waveform and spectrogram (Hualde & Prieto, 2014; Torreira & Ernestus, 2011; Turk, Nakai, & Sugahara, 2006). Duration for Spanish approximants has been measured based on the manual segmentation of intensity-based criteria (Hualde, Simonet, et al., 2011) by placing a marker where the intensity curve begins to decrease from the previous vowel and another where it stops increasing to the following vowel. In this paper, the approximant duration measurement is automated (see Section 2.4) and applied to all consonants (/ptk/, /bdg/, and /sf/), rather than using a separate F2-based closure-related criterion for /ptk/ that is not applicable to /bdg/, and aperiodic noise onset and offset criteria for /sf/, as high vowels can sometimes have superimposed frication, making the criteria inconsistently reliable. In this way, the exact same measurement is applied to all consonants, making them comparable. This has the additional benefit of eliminating the need to identify the exact time of closure for /ptk/, which in Spanish may not occur at all and would lead to an unnecessary dichotomization of the voiceless plosives into two groups based on whether they achieved full closure (ignoring the truly gradient nature of closure reduction in the process). To be clear, this measure, which I will simply refer to as *consonant duration*, includes portions of the surrounding vowels, which some other measures of consonant duration aim to avoid. This is intentional, and to control for the difference in intrinsic vowel durations, preceding and following vowel height are included in all analyses. Some manual segmentation is still required, but the precision required is substantially reduced. The researcher needs only to place boundaries that contain the intensity dip of a consonant without containing the surrounding vowel intensity maxima. From there, the exact location of the consonant minimum can be located using the intensity contour, and the preceding and following vowel intensity maxima can be located using the first-differentiated intensity contour (i.e. intensity velocity curve), searching for the first instance where the velocity changes sign (described in greater detail Section 2.4), rather than determining these locations manually.

2.2.3. Voicing

The degree of intervocalic /ptk/ voicing has been measured in relation to closure and release landmarks in the acoustic signal. For example, Torreira and Ernestus (2011) measure the duration of voicing in the closure of /ptk/ and VOT, and others measure the percentage of the consonant’s closure that is voiced (e.g. Hualde, Simonet, et al., 2011; Lewis, 2000). For Spanish and Catalan /sf/, voicing has been measured as the percentage of the fricative that is voiced (e.g. Hualde & Prieto, 2014). In this study, pitch tracking (K. Johnson, 2012, pp. 64–68) is used to detect a gap in voicing (Torreira & Ernestus, 2011) within the interval of the consonant's total duration as described above. The duration of this interval, which I will call voiceless duration, measures the total duration of the absence of periodicity in the VCV sequence, regardless of its relation to other less reliable acoustic landmarks (i.e. the voiceless period may contain possible pre-closure aspiration and VOT for plosives, and voicelessness that occurs before or after the onset of uninterrupted aperiodic noise for fricatives). This measure thus has the same advantages as the total consonant duration measure described above, in that no binning of the plosives or fricatives into separate groups is required. The voiceless duration was then divided by the consonant's total duration to obtain the percentage of the consonant that was voiceless, and both the durational and percentage measures were considered in analysis (note that the percent voiceless will never be exactly 1 since the consonant duration, as measured here, contains portions of the surrounding vowels). Some authors have measured the percentage of a consonant that is voiceless using the "fraction of locally unvoiced frames" in Praat's voice report (Eager, 2015; Hualde, Eager, & Nadeu, 2015; Hualde & Prieto, 2014), while others have used the pitch tracker directly as is done here (e.g. Torreira & Ernestus, 2011). The measures taken here can also be obtained by first taking the percentage of the consonant's total duration that is voiceless using the voice report and then multiplying this percentage by the consonants duration to obtain the voiceless duration; the two approaches are based on equivalent aspects of the acoustic signal and should not differ substantially in their outcomes.

2.3. Predictors

2.3.1. Predictors relevant to the hypotheses

In the fricative analysis, the main hypotheses to be tested are that Catalan has stronger /s/ productions than Spanish, but the languages do not differ in the strength of /f/, and that a speaker's /f/ strength will be correlated with their /s/ strength. For this reason, the full interaction of language (Catalan or Spanish) and fricative phoneme identity (/f/ or /s/) was included. In the plosive analysis, the main hypotheses are that the dialects will show a strength hierarchy of Cuzco > Valladolid > Lima for both /bdg/ and /ptk/, and that a speaker's voiceless plosive strength will be correlated with their voiced plosive strength, with the strength of this relationship varying by place of articulation. For this reason, dialect (Cuzco, Lima, or Valladolid), place of articulation (bilabial, dental, or velar), and underlying voicing (voiced or voiceless) were included, along with the interaction of dialect and underlying voicing, and the interaction of place of articulation and underlying voicing.

2.3.2. Control predictors

2.3.2.1. Social factors

Speaker sex (female or male) was included in both analyses. For the fricative analysis, the interaction of sex and language was included, and for the plosive analysis, the interaction between sex and dialect was included. Within the Cuzco plosive data, age group (older or younger), education level (secondary or university), and Quechua bilingualism (yes or no) were also included (coding of these factors is detailed in Section 3.4.2). While the analyses carried out in this dissertation are not sociolinguistic in nature, it is still important to include these predictors and interactions as controls.

2.3.2.2. Linguistic factors

Linguistic factors that have been shown to affect consonant strength in Spanish and Catalan included in both analyses were stress (tonic, post-tonic, or unstressed) and preceding and following vowel height (high or non-high); for the plosive analysis, position in the word (initial or medial) and task (read speech or spontaneous) were also included (Carrasco, 2008; Colantoni & Marinescu, 2010; Cole et al., 1999; Delforge, 2009; Hualde, Simonet, et al., 2011; Munday, 2001; Nadeu, 2014; Soler & Romero, 1999; Torreira & Ernestus, 2011; Williams, 1987; among others). With regards to the preceding and following vowels, previous research has coded for vowel phonemic identity (e.g. for Spanish, a factor with five levels /a/, /e/, /i/, /o/, and /u/). In this study, vowels are coded as high or non-high (/i/ and /u/ being high, and all other vowels in both languages being non-high), as height distinctions can be expected to affect all of the measures taken: non-high vowels have longer durations, which should lead to longer consonantal durations as measured here; voiceless periods tend to be longer bordering high vowels than non-high vowels; and non-high vowels have higher intensity than high vowels (Delforge, 2009, 2012; Torreira & Ernestus, 2011; among others).[11] For each of these control factors, the interaction with underlying voicing in the plosive analysis and the interaction with language in the fricative analysis were included when the descriptive statistics warranted inclusion (described in more detail in Chapter 4 and Chapter 5).

2.3.2.3. Linguistic covariates

To control for the possible effect of word frequency (Pierrehumbert, 2001), the natural logarithm of the number of times a consonant's word occurred in the CREA corpus (Real Academia Española, 2016) for Spanish and the IEC corpus for Catalan (IEC, 2016) was taken as a measure of word frequency, with words not found in the corpora given a value of $\ln(1) = 0$. As a final linguistic control, speech rate was measured, as it may affect all five phonetic measures of consonant strength taken in this study (intensity difference, intensity velocity, duration, voiceless duration, and percent voiceless). Speech rate was measured using the Praat script written by De Jong and Wempe (2009) to measure intensity maxima corresponding to syllable nuclei. The script first locates all intensity maxima over a recording-specific relative threshold, then discards those that either are not voiced or do not follow a dip in intensity of at

[11] It is possible that low and mid vowels may also differ in this regard, but I leave this to future research.

least 2 dB. The remaining maxima are added to a TextGrid and can be used to measure speech rate. In the case of the voiceless consonants /ptksf/, it seems reasonably clear that a faster speech rate will correspond to a weaker consonant. However, for the voiced plosives /bdg/, the exact role of speech rate, regardless of how it is measured, is somewhat unclear given that the plosives under study are leniting to the point of elision on a gradient scale. For a VCV sequence containing /b/, /d/, or /g/, if the consonant is weak enough (i.e. has a low enough intensity difference), then the two vowels in the sequence will be identified as a single nucleus rather than two separate nuclei. As the strength of the consonant increases to an intensity difference over 2 dB, then two syllable nuclei are more likely to be counted. In other words, a relatively greater number of syllable nuclei may be positively correlated with plosive strength in this case.[12] The effect of this measure would thus possibly differ for phonemically voiceless and voiced plosives, and so in the Spanish /ptk/ and /bdg/ analysis, its interaction with underlying voicing was included. The role of speech rate is also likely to vary considerably by speaker. To that end, the measure was speaker-normalized (for each speaker, the mean for that speaker was subtracted from the speaker's values, and divided by their standard deviation, yielding a measure that represents how quickly or slowly a speaker was talking with respect to their own speech rate tendencies on unit scale).

2.3.3. Random effects grouping factors

In both the plosive and fricative analyses, speaker was considered as a random grouping factor. For the Cuzco and Lima read speech plosives, experimental item was also included as a

[12] Even if a researcher used phone-rate and syllable-rate measures with audio files which were fully segmented at the phone and syllable level, they would run into the same issue: how do you count the number of phones and syllables in a VCV sequence when the consonant is elided? This will have an effect on the resulting measure, and counting the VCV sequence as two syllables begs the question with regards to whether or not speech rate can explain the presence of the elision.

grouping factor (coding detailed in Section 3.4.2). Other authors have included word (or similar groups) as a random effect grouping factor in spontaneous language production (e.g. Bresnan, Cueni, Nikitina, & Baayen, 2007), but in this study it is not included for two reasons: (1) the vast majority of words only have one observation, meaning their intercept would be collinear with their errors in the model (i.e. there aren't actually repeated measures on most of them); and (2) the sense in which there are repeated measures on words in spontaneous speech is not the same as is the case for experimental items (i.e. for the read speech items, the same plosive in the same word is occurring in the exact same sentence that was not created by the speaker, but rather by the experimenter, and so we can expect that repeated measures on this item will introduce an additional variance component). For all grouping factors, the maximal random effects structure, as described in Section 3.2.4, was included.

2.4. Segmentation and measure extraction

Segmentation and measure extraction were performed in Praat (Boersma & Weenink, 2016), using a script written by the author. De Jong and Wempe's (2009) script was also applied to each audio file to produce a TextGrid tier containing points at each syllable nucleus. For each audio file, a cross-correlated pitch object was created, following the Praat manual's recommendation that cross-correlation be used in the analysis of voicing rather than auto-correlation, with a time step of 1 ms. Sex-specific pitch ranges of 100-300 Hz for female speakers and 70-250 Hz for male speakers were used, as these ranges produce values that are not significantly different from using speaker-specific ranges (Eager, 2015; Vogel, Maruff, Snyder, & Mundt, 2009). All other parameters were left at their defaults. The pitch values from each pitch object were then written to a table, and a new column was created with the time gap between pitch points, obtained by first-differentiating the time stamps of the pitch points.

For intensity difference and intensity velocity, using the raw intensity measurements would cause the intensity from voicing to raise the minimum intensity in the consonant, causing a fully voiced segment and a fully voiceless segment of the same constriction degree to have different intensity differences and velocities. Intensity at very high frequencies may also be related to background noise rather than constriction, and so its exclusion is also desirable. For this reason, the raw audio files were filtered with a band-pass filter and von Hann window (Hualde, Nadeu, & Simonet, 2010; K. Johnson, 2012, pp. 61–62, 68–71; Recasens, 2015). Hualde et al. (2010) use a range of 500 - 10,000 Hz for the band-pass filter, while Recasens (2015) uses a range of 250 - 10,000 Hz, arguing that the higher floor of 500 Hz may remove intensity associated with F1, which contains important information about constriction. As the goal of the filter's floor is to remove intensity associated with F0 while maintaining intensity associated with F1, and these values are both higher for female speakers than for male speakers, in this study the floor was set to the sex-specific ceiling used in the creation of the pitch objects; that is, the band-pass filter was set to 300 - 10,000 Hz for female speakers and 250 - 10,000 Hz for male speakers. This filtered audio file was then used to create an intensity object with a time step of 1 ms. According to the Praat manual, the minimum pitch setting for the creation of intensity objects should be set as high as possible while still being lower than the minimum pitch that can be expected from the speaker, thus ensuring that the curve is neither pitch-synchronous nor smeared. For this reason, the minimum pitch was set to 100 Hz for female speakers and 70 Hz for male speakers (the same as for the cross-correlated pitch objects), making the effective analysis window length 8 ms for female speakers and 11.4 ms for male speakers. The values of the audio file in each window are “first squared, then convolved with a Gaussian analysis window (Kaiser-20; sidelobes below -190 dB” (Boersma & Weenink, 2016). The resulting intensity values (in dB) were then written to a table for each audio file. As an estimation of the amount of background noise in each filtered audio file, the intensity during a silent interval was taken, and recorded as $10\log_{10} RMS_{noise}$, with the raw values in the intensity tables representing $10\log_{10}(RMS_{speech} + RMS_{noise})$ (see K. Johnson (2012, pp. 59–60, 87–88) for details on the relationship between RMS and dB). The noise was then subtracted out of the raw values by first converting both to RMS, then subtracting the noise RMS out, and then converting the resulting values back to dB. This is similar to spectral subtraction (Boll, 1979), but as only the total intensity in the band-pass filtered signal is being used, the total noise is subtracted from the total raw values rather than at separate frequency components. These adjusted values were then first-differentiated to create a second column with intensity velocities in dB/ms.

The filtered audio file was used for segmentation. For each consonant, on the first tier of the TextGrid, boundaries were manually placed at the beginning and end of the word the consonant occurred in and the previous word. On the second tier, boundaries were placed at the beginning and end of the VCV sequence, with additional boundaries placed around the minimum intensity of the consonant.[13] The resulting three intervals were labeled with their phonemic identity (the first interval with the preceding vowel's, the second with the consonant’s, and the third with the following vowel's), and following each vowel's phonemic label, an *s* was entered to indicate the vowel belonged to a stressed syllable and a *u* was entered to indicate the vowel belonged to an unstressed syllable. If no evidence of a consonant was present (no dip in the intensity contour), the boundaries for the consonant were placed arbitrarily so that the script would be able to record phoneme identity, and a point was placed on the third tier within this interval and labeled *elided*. Silent intervals near each consonant (if any) were segmented on tier

[13] In the case that intensity reached a minimum then raised temporarily before falling again to a second minimum (as may occur in the case of high burst amplitude and long VOT for plosives, and in most fricatives), both consonant minima were ensured to be contained by the boundaries.

four and marked as *sil*. Finally, the syllable nucleus points from De Jong and Wempe's (2009) script were added as a fifth tier.[14]

A Praat script written by the author was then used to compute the acoustic measures of consonant strength described in Section 2.2 along with speech rate as described in Section 2.3.2.3. The observations' coding for the factors described in Sections 2.3.2.1 and 2.3.2.2 were obtained from interval labels and audio file names. For non-elided consonants, the script first found the minimum intensity value in the frequency-filtered noise-adjusted table within the consonant's segmented interval and placed a point on the third tier at this time (labeled *cmin*). Then, starting from the start point of the plosive interval in the intensity velocity table, the script worked backwards in time until the first non-negative velocity was found (representing the previous vowel's maximum intensity), recorded the time of this maximum, and added a point to the third tier at its time stamp (labeled *d1*). Similarly, starting from the end point of the consonant interval, the script worked forwards until the first non-positive velocity was found (representing the following vowel's maximum intensity), recorded the time at which the maximum occurred, and placed a point in tier three at this time (labeled *d2*). For plosives, the maximum value in the intensity velocity table that occurred between *cmin* and *d2* was also recorded, and a point was added to the third tier at the time at which the maximum velocity occurred (labeled *vel*). The script then found the pitch points in the pitch table that occurred between *d1* and *d2*, and searched for gaps greater than 2 time steps (that is, 2 ms).[15] The time stamps of the point at the beginning of the gap and of the following pitch point were recorded,

[14] The exact locations of the syllable nuclei and the consonant duration markers may be slightly different because different criteria are applied in determining their location.

[15] Due to small differences in the exact times of the pitch points, a fully voiced consonant was found to be able to have gaps in the table between 1 and 2 ms, while a threshold of 2 ms successfully identified voiceless periods.

and points were added to the third tier (labeled *vp1* and *vp2*).[16] If no voiceless period was found, these points were not placed. Lastly, points were placed in the second tier 500 ms before and after *cmin* and labeled *srw1* and *srw2* to create a 1-second speech rate window. For elided consonants, the script only placed points for the speech rate window, centering it at the mid-point of the VCV sequence. If any points overlapped in their exact times (as may be the case for *d2* from one consonant and *d1* from the following consonant when two observations occurred near one another), the labels were added to a single point and separated with an underscore (e.g. *d2_d1*). Examples of the resulting segmentations are provided in Figure 2.1 through Figure 2.6. In all of the example segmentations, the intervals in tiers one, two, and four are manually segmented, the points in tier five are the syllable nuclei detected by De Jong and Wempe's (2009) script, the points in tier three were placed by the author's script, and the solid black line is the intensity contour.

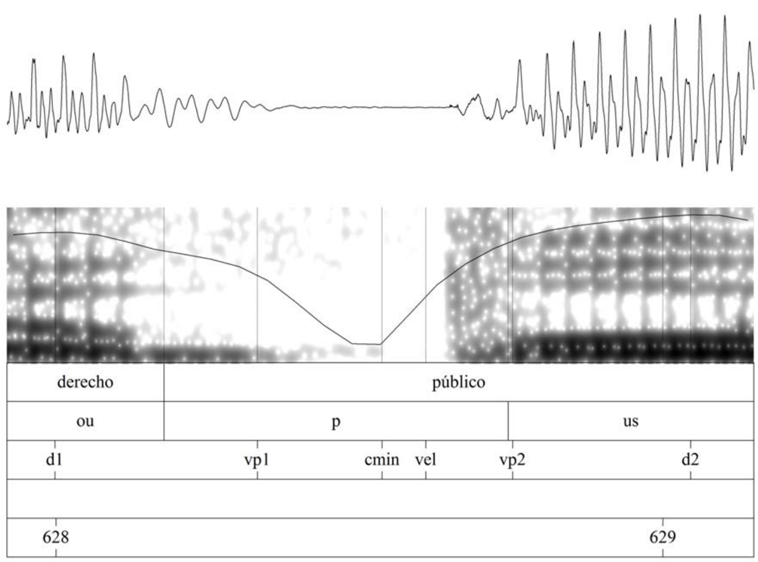

Figure 2.1 Example segmentation of a word-initial tonic /p/.

[16] Examination of the results of the script showed that in some cases, for plosives, a double burst (most often for /k/) was identified by the pitch tracker as being voiced due to bursts' temporal spacing, resulting in two voiceless periods. In these cases, the first *vp2* point and the second *vp1* point were deleted, yielding the desired interval.

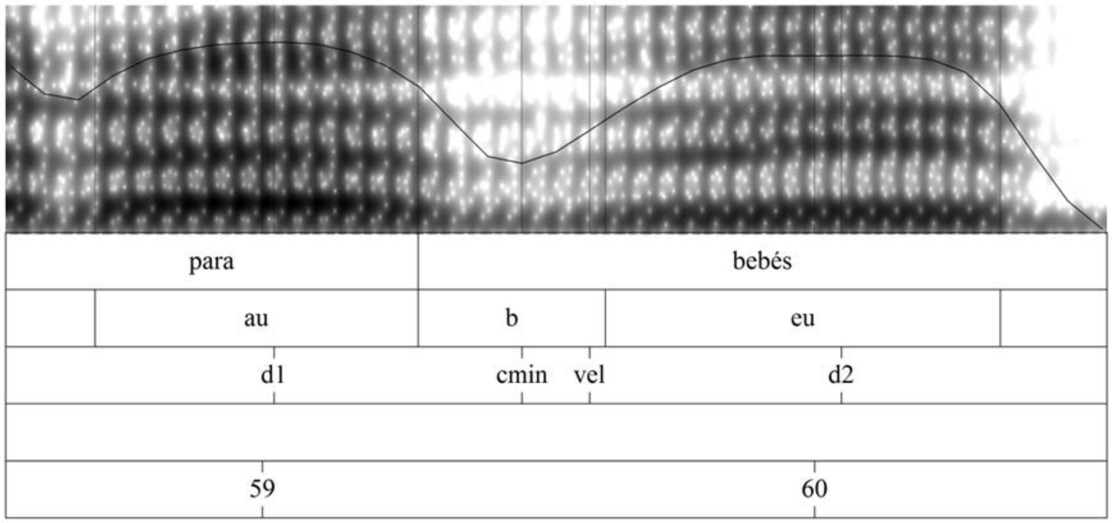

Figure 2.2 Example segmentation of a voiced approximant realization of a word-initial unstressed /b/.

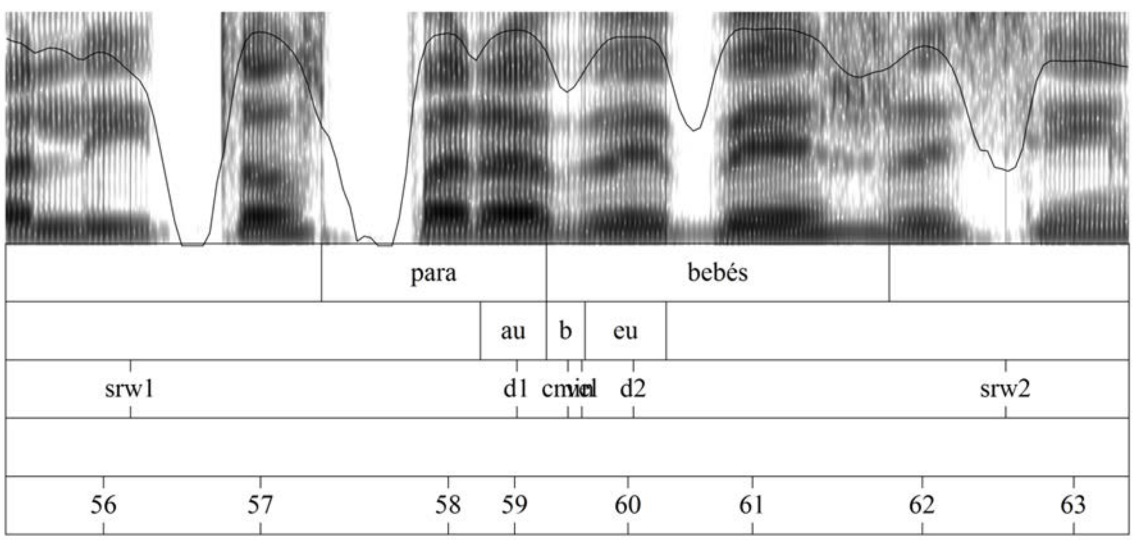

Figure 2.3 Example automated segmentation of the 1-second speech rate window centered at the consonant minimum intensity for the word-initial unstressed /b/ in Figure 2.2. In this case, the speech rate is 6 nuclei/s.

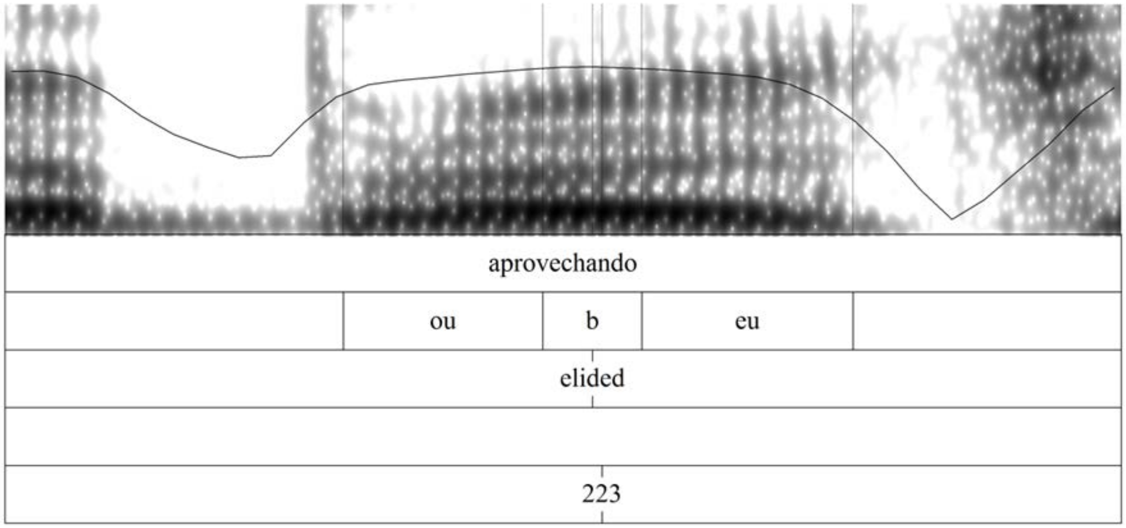

Figure 2.4 Example segmentation of an elided word-medial unstressed /b/.

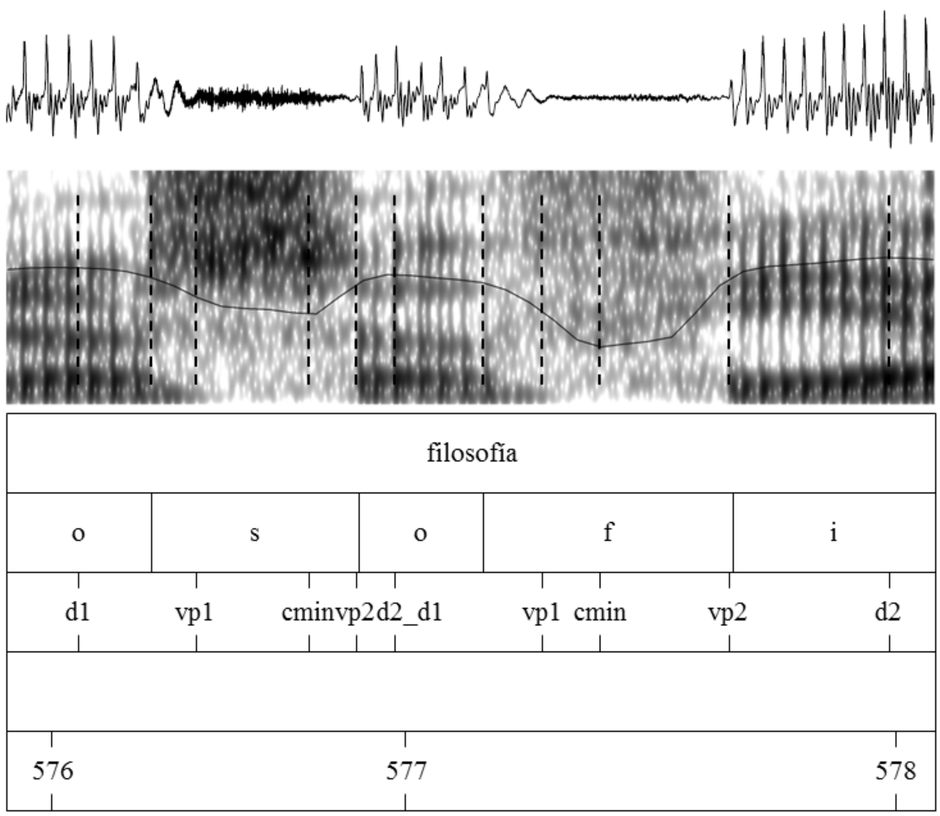

Figure 2.5 Example segmentation of /s/ and /f/, both with voiceless periods.

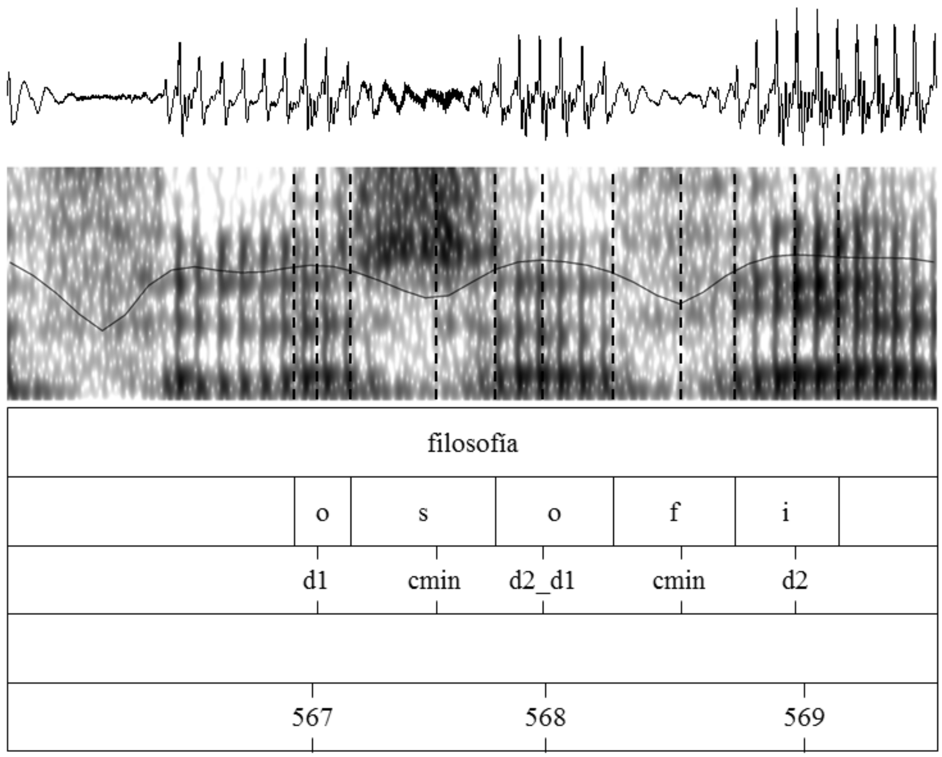

Figure 2.6 Example segmentation of fully voiced /s/ and /f/.

For non-elided consonants, consonant duration was measured as *d2 – d1*, and voiceless duration was measured as *vp2 – vp1* if these points were present or zero if they were not present (i.e. if the consonant was fully voiced). The percentage of the consonant that was voiceless was then computed by dividing the voiceless duration by the consonant duration (i.e. (*vp2 – vp1*) / (*d2 – d1*)). For non-elided plosives, intensity difference was measured as the intensity at *cmin* subtracted from the intensity at *d2*, and intensity velocity was measured as the velocity at *vel*. For elided consonants, all relevant dependent variables were set to zero. For all consonants (elided or not), the number of nuclei occurring between *srw1* and *srw2* was recorded, and the duration of the window between *srw1* and *srw2* that did not overlap with silent intervals in tier four was recorded. The number of nuclei was then divided by the non-silent duration to obtain the local speech rate measured in nuclei per second (this was then speaker-normalized as described in Section 2.3.2.3 in R). The phonemic identity of the consonant and the preceding and following vowels, the stress condition, the word the plosive occurred in, read speech item (from phonemic identity and word, if applicable), and, where relevant, whether the consonant was word-medial or word-initial were also extracted from the two manually segmented interval tiers, and speaker identifier and demographic information was obtained from the file names. These results were then stored in a CSV file. Log word frequency was then computed as described in Section 2.3.2.3 and added to the CSV file.

2.5. Dependent variable dimensionality reduction

2.5.1. Dependent variable descriptive statistics

Descriptive statistics for each measure of plosive strength (percent voiceless, voiceless duration, plosive duration, intensity difference, and intensity velocity) by underlying voicing are provided in Table 2.3. Recall that elided tokens were assigned a value of 0 for all measures, and non-elided tokens that were fully voiced were assigned a value of 0 for voiceless duration and percent voiceless.

Table 2.3 Descriptive statistics for acoustic measures of plosive strength by underlying voicing.

Voicing	Measure	Minimum	Median	Maximum	Mean	SD
Voiced (N = 2694)	Intensity Difference	0.00	6.27	44.26	8.75	8.59
	Intensity Velocity	0.00	0.21	1.70	0.31	0.33
	Duration	0.00	97.00	262.00	91.30	51.24
	Voiceless Duration	0.00	0.00	66.00	0.58	4.84
	Percent Voiceless	0.00	0.00	0.57	0.01	0.04
Voiceless (N = 2587)	Intensity Difference	0.64	36.43	73.95	36.21	9.39
	Intensity Velocity	0.05	1.28	5.56	1.38	0.64
	Duration	50.00	155.00	298.00	157.94	36.66
	Voiceless Duration	0.00	82.00	212.00	79.45	33.01
	Percent Voiceless	0.00	0.52	0.91	0.49	0.17

There were no elided tokens of /ptk/ in the dataset, and 15.5% of the /bdg/ tokens were elided (Cuzco: 2.8% in read speech and 8.4% in spontaneous speech; Lima: 34.3% in read speech and 43.7% in spontaneous speech; Valladolid: 21.1% in spontaneous speech). Additionally, 1.8% of the /bdg/ tokens were found to be partially devoiced (Cuzco: 3.9% in read speech and 2.3% in spontaneous speech; Lima: none; Valladolid: 0.3% in spontaneous speech), and 4.8% of the /ptk/ tokens were fully voiced (Cuzco: 0.6% in read speech and 2.7% in spontaneous speech; Lima: 7.0% in read speech and 20.4% in spontaneous speech; Valladolid: 6.1% in spontaneous speech). The five acoustic measures are (as should be expected) highly positively correlated, as can be seen in Figure 2.7.

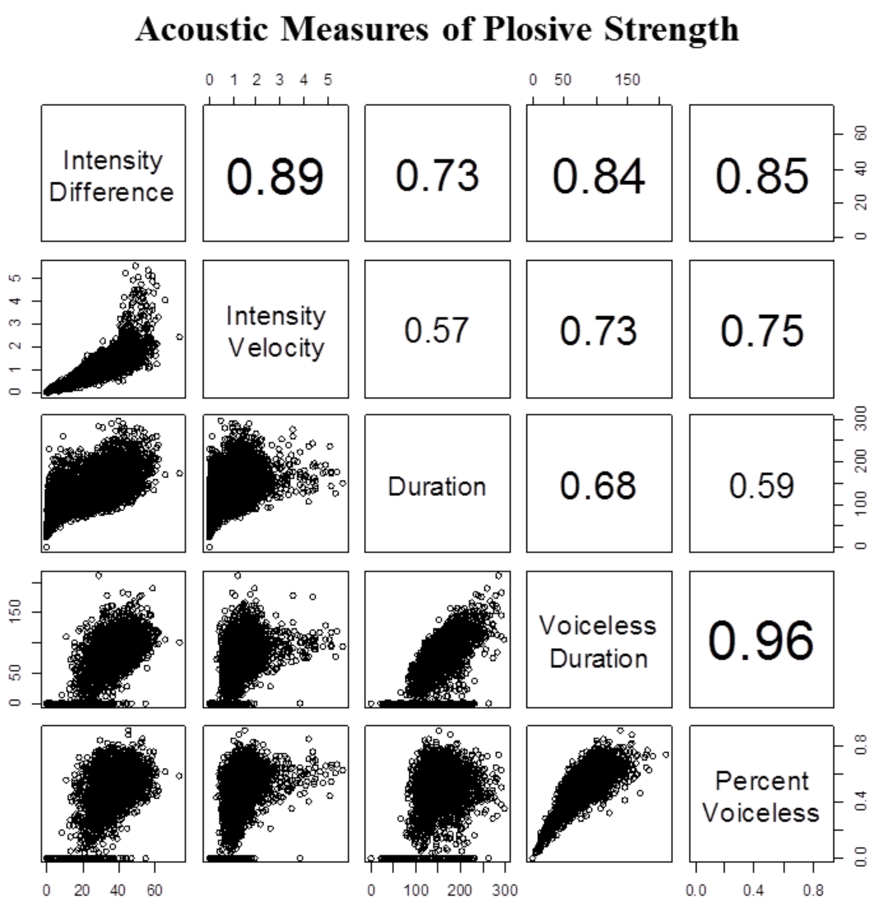

Figure 2.7 Scatterplots and sample correlations for acoustic measures of plosive strength.

Descriptive statistics for each measure of fricative strength (percent voiceless, voiceless duration, and fricative duration) by language and phoneme identity are provided in Table 2.4.

Table 2.4 Descriptive statistics for acoustic measures of fricative strength by language and phoneme.

Language	Fricative	Measure	Minimum	Median	Maximum	Mean	SD
Catalan	/f/ (N = 269)	Duration	64.000	143.000	266.000	147.297	33.092
		Voiceless Duration	0.000	72.000	158.000	67.082	33.823
		Percent Voiceless	0.000	0.482	0.790	0.441	0.201
	/s/ (N = 964)	Duration	61.000	154.500	310.000	158.749	37.961
		Voiceless Duration	0.000	78.000	169.000	73.381	36.243
		Percent Voiceless	0.000	0.495	0.797	0.446	0.191
Spanish	/f/ (N = 165)	Duration	73.000	142.000	271.000	145.242	35.544
		Voiceless Duration	0.000	73.000	160.000	66.921	35.027
		Percent Voiceless	0.000	0.519	0.794	0.445	0.209
	/s/ (N = 765)	Duration	45.000	139.000	288.000	143.141	32.526
		Voiceless Duration	0.000	71.000	159.000	64.825	35.956
		Percent Voiceless	0.000	0.504	0.780	0.434	0.221

There were no elided fricatives in the dataset. For Catalan, 12.6% of /f/ were fully voiced and 10.4% of /s/ were fully voiced. For Spanish, 14.5% of /f/ were fully voiced and 16.3% of /s/ were fully voiced. Just as for the plosives, the measures of fricative strength are also highly positively correlated, as can be seen in Figure 2.8.

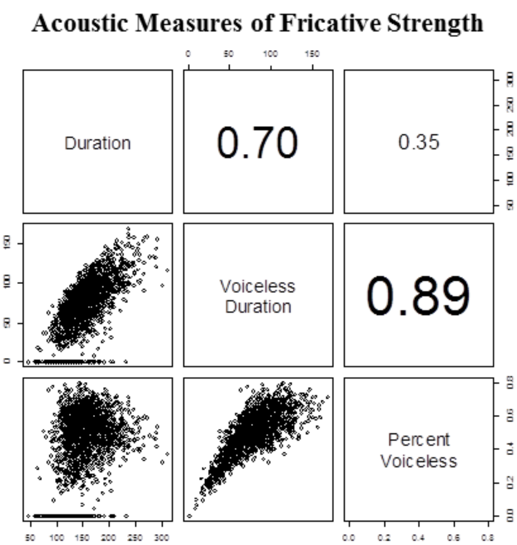

Figure 2.8 Scatterplots and sample correlations for acoustic measures of fricative strength.

2.5.2. Principal component analysis

Analyzing the measures separately introduces a researcher degrees of freedom issue that increases the likelihood of a false-positive result (Simmons, Nelson, & Simonsohn, 2011). In addition to this issue, analyzing them separately also misses the overall pattern in the datasets. A better alternative is to use a principled method to reduce the dimensionality of the data such as principal component analysis (PCA; see Baayen, 2008; R. A. Johnson & Wichern, 2002). PCA rotates the correlated multidimensional space to create a new set of uncorrelated variables. It starts by finding the linear combination of the variables that has the highest variance (and therefore explains the most variance in the data), and applies this transformation to create the first principal component (PC1), assigning a score to each observation based on the weight given to each variable in the rotation that produced the component (called loadings). It then repeats this process, finding the linear combination of the variables that explains the most of the variation that was not explained by PC1, under the constraint that the second component PC2 is orthogonal to PC1 (completely uncorrelated with a sample correlation coefficient of zero). This process continues until there are as many PCs as there are variables in the original dataset, with each subsequent PC explaining less of the total variance in the data than the previous PC. It is often the case that the majority of the variance in the original data can be explained with one or two PCs. A standard approach to PCA is that the first *n* PCs that cumulatively account for 80% of the variance in the data can be used to replace the original dataset without much loss of information (R. A. Johnson & Wichern, 2002, p. 422). PCs that account for large amounts of variance also tend to have loadings (the weights each variable is multiplied by before summing them to create the PC scores) that have an intuitive interpretation that explains overall patterns in the data. For these reasons, PCA was performed on the five acoustic measures of plosive strength and on the three acoustic measures of fricative strength, using the `prcomp` function in R, first mean-centering the variables and dividing them by their respective standard deviations to put them on unit scale, as PCA is sensitive to differences in variable scaling. The zero-valued tokens (arising from elisions and fully voiced consonants) were included in the analysis for two reasons: (1), the zero values are not missing values, but rather the natural end of a continuum that provides information to listeners (not hearing a drop in intensity in what is phonemically a VCV sequence is an event, and the listener has no way of knowing whether the elision was intentional or the result of articulatory undershoot; Ohala (1983)); and (2), the assumptions of the linear model are on the residuals, not the raw data. The results of the PCA on the plosive strength measures are presented in Table 2.5 (variable loadings) and Table 2.6 (PC importance), and the results of the PCA on the fricative strength measures are presented in Table 2.7 (variable loadings) and Table 2.8 (PC importance).

Table 2.5 Plosive principal component variable loadings.

Measure	PC1	PC2	PC3	PC4	PC5
Duration	0.39	0.89	-0.02	0.17	0.16
Voiceless Duration	0.47	-0.14	-0.49	0.22	-0.69
Percent Voiceless	0.46	-0.33	-0.44	-0.01	0.69
Intensity Difference	0.48	-0.01	0.29	-0.82	-0.15
Intensity Velocity	0.44	-0.28	0.69	0.50	0.02

Table 2.6 Plosive principal component importance measures.

	PC1	PC2	PC3	PC4	PC5
Standard Deviation	2.02	0.69	0.60	0.27	0.17
Proportion of Variance	0.81	0.10	0.07	0.01	0.01
Cumulative Proportion	0.81	0.91	0.98	0.99	1.00

Table 2.7 Fricative principal component variable loadings.

Measure	PC1	PC2	PC3
Duration	0.50	-0.80	-0.33
Voiceless Duration	0.65	0.09	0.75
Percent Voiceless	0.57	0.59	-0.57

Table 2.8 Fricative principal component importance measures.

	PC1	PC2	PC3
Standard Deviation	1.52	0.81	0.14
Proportion of Variance	0.77	0.22	0.01
Cumulative Proportion	0.77	0.99	1.00

In both PCAs, the first PC assigns nearly equal positive weights to all acoustic measures, giving PC1 the intuitive interpretation of measuring overall consonant strength in each case. For plosives, PC1 accounts for 81% of the variance in the raw data, and for fricatives, PC1 accounts for 77% of the variance in the raw data, with subsequent PCs explaining substantially less variance, meaning that in both cases the multivariate phonetic space can be reduced to a single variance component without a significant loss of information (R. A. Johnson & Wichern, 2002,

p. 422).[17] For this reason, for both the plosive and fricative datasets, PC1 will be used as the sole dependent variable in subsequent analyses.

[17] For the fricatives, PC1 accounts for slightly less than 80% of the variance, but its loadings indicate that it describes the aspects of the data we are interested in, and the cutoff is an approximant guideline.

Chapter 3. Statistical methods

3.1. Introduction

In the phonetic analyses carried out in this dissertation, all statistics are run in R (R Core Team, 2016). I use Bayesian mixed effects regression to obtain estimates for effects at the population level (i.e. the fixed effects), and the individual level (using both the fixed and random effects). In Section 3.2, I discuss the general form of mixed effects regressions and their estimation with frequentist methods in `lme4`, why mixed effects models are necessary, why maximal random effects structures should be the default implementation, and the why a Bayesian approach to these models allows us to answer research questions that a frequentist approach cannot. This discussion is crucial to understanding why a Bayesian approach better suits the research questions, and to understanding the differences in the interpretation of the results as compared to methods that linguists are more familiar with. In Section 3.3, I detail some of the basic concepts of Bayesian inference, as well as estimation of Bayesian mixed effects models with Markov Chain Monte Carlo. In Section 3.4, I discuss best statistical practices with respect to continuous variable scaling and factor contrasts, and priors that are weakly informative when these practices are followed. In Section 3.5, I explain how regression results will be presented, and the differences in interpretation between the results of Bayesian and frequentist regressions. In Section 3.6, I conclude with summary remarks on the statistical methods employed throughout this dissertation.

3.2. Frequentist mixed effects regression

The statistical notation I use for parameters is slightly different than that of the authors cited, so that the same notation can be used throughout. In quantitative studies, the researcher forms a hypothesis, collects data, measures this data in any number of ways, and then uses statistics to evaluate to what extent the data support the hypothesis. The goal of the quantitative analysis is to be able to make a probabilistic statement about the hypothesis. Mathematically, what we are interested in is $P(hypothesis \mid data)$, where $P()$ is a probability function and the '$|$' means *given* or *conditioned on*, making this statement read "the probability of the hypothesis given the data". In order to evaluate this probability, the hypothesis and the data need to be mapped numerically onto parameters. A common analysis method (and the one used in this dissertation) is linear mixed effects regression, which allows us to model both fixed effects (i.e. population level effects) and random effects (i.e. the noise introduced by individual variation when there are repeated measures on members of a grouping factor). When referring to a parameter in the model (e.g. the intercept term, the difference between the mean /p/ strengths of Cuzco and Lima, etc.), the true parameter (the actual population parameter whose value is by definition unknowable; e.g. the true mean duration of intervocalic /p/ in Lima Spanish) will be denoted with a Greek or Roman character, and a regression estimate for the parameter will be denoted with the same symbol with a hat (e.g. for the true parameter β, the regression estimate is $\hat{\beta}$). The '$\sim$' symbol denotes a probability distribution statement and is followed by the abbreviation for a distribution name and the parameters that define that distribution. For example, the normal distribution is defined by its mean and standard deviation, and $x \sim N(2, 3)$ reads "x is normally distributed with mean 2 and standard deviation 3".

3.2.1. Ordinary least squares regression and p-values

In ordinary least squares (OLS) regression (that is, a linear regression with only fixed effects estimated with a frequentist approach), the continuous dependent variable y is modeled as a linear function of a matrix of predictor features X, with the linear function expressed through a coefficient vector β, and with the errors in the predictions made by the model, ϵ, being independent of one another (i.e. not correlated with one another) and normally distributed with mean zero and unknown standard deviation σ_ϵ (Demidenko, 2013, pp. 2–3). That is,

$$y = X\beta + \epsilon, \qquad \epsilon \sim N(0, \sigma_\epsilon) \quad \rightarrow \quad y \sim N(X\beta, \sigma_\epsilon).$$

The regression estimates $\hat{\beta}$ are chosen such that $\hat{\sigma}_\epsilon$ is minimized. The variance of each element of $\hat{\beta}$, and the covariance of each pair of estimates in $\hat{\beta}$, are contained in the estimated covariance matrix $\hat{\Sigma}_\beta$ (whose true value Σ_β is also unknown). The estimates for these parameters have closed-form solutions (i.e. they are the result of simple matrix algebra). Letting $\hat{y}$ be the predicted values of the dependent variable, n be the number of observations (i.e. the number of rows in X and elements in y and ϵ), and p be the number of fixed effects (i.e. the number of columns in X and elements in β), we have:[18]

$$\hat{\beta} = (X^T X)^{-1} X^T y, \qquad \hat{y} = X\hat{\beta}, \qquad \hat{\epsilon} = y - \hat{y}, \qquad \hat{\sigma}_\epsilon^2 = \frac{\sum_{i=1}^{n} \hat{\epsilon}_i^{\,2}}{n-p}, \qquad \hat{\Sigma}_\beta = \hat{\sigma}_\epsilon^2 (X^T X)^{-1}.$$

These equations yield a unique solution so long as X is full rank (i.e. there is no complete collinearity in the predictors; Weisberg (2005, pp. 54–58)). These estimates are also known as maximum likelihood estimates, because they yield the highest possible value for the likelihood function, which is simply the joint probability of all of the data for a given set of values for β, written $P(y \mid \beta)$. Thus, in this context, $\hat{\beta}$ can be interpreted as "the data are most likely when we

---

[18] X^T is the transpose of X, which can also be written X'.

set β equal to $\hat{\beta}$". Note, however, that this is not the same as what we set out for, $P(hypothesis \mid data)$, but rather the reverse: $P(data \mid hypothesis)$. With our current estimates $\hat{\beta}$, we cannot actually quantify $P(\beta \mid y) \rightarrow P(hypothesis \mid data)$; that is, we cannot obtain a distribution from which we can make probabilistic statements about β.

Instead, we can set up a counterfactual statement where we assume a particular true value for β and assess the probability that we would obtain a $\hat{\beta}$ with at least as great a magnitude as our OLS $\hat{\beta}$, a procedure known as null hypothesis significance testing (NHST). Most often what is done is to assume that the true value for one element of β (say, the j'th coefficient) is zero: H_0: $\beta_j = 0$, with the corresponding alternative hypothesis being H_A: $\beta_j \neq 0$. In this case, if H_0 is true, then our estimate $\hat{\beta}_j$ has a Student's t-distribution (which is similar to a normal distribution but with more probability in the tails), and we can make a probabilistic statement using this distribution (Weisberg, 2005, pp. 31–34, 63, 74). While the normal distribution is defined by a mean parameter and a standard deviation parameter (i.e. $N(\mu, \sigma)$), the non-central t-distribution is defined by three parameters: the degrees of freedom ν, center m, and scale s. The ν parameter (which must be positive) determines how heavy the tails are, with lower values indicating heavier tails; at $\nu = \infty$, the t-distribution is the same as the normal distribution; that is, $t(\nu = \infty, m, s) = N(\mu = m, \sigma = s)$. When $\nu \leq 1$, the tails of the distribution are so heavy that there is no mean or variance for the distribution. When $1 < \nu \leq 2$, the distribution has mean m, but the variance is still ∞, and when $\nu > 2$, the distribution additionally has standard deviation $s\sqrt{\nu/(\nu - 2)}$. To visualize the difference between the two, density plots for the $N(0, 1)$ and $t(1, 0, 1)$ distributions are given in Figure 3.1.

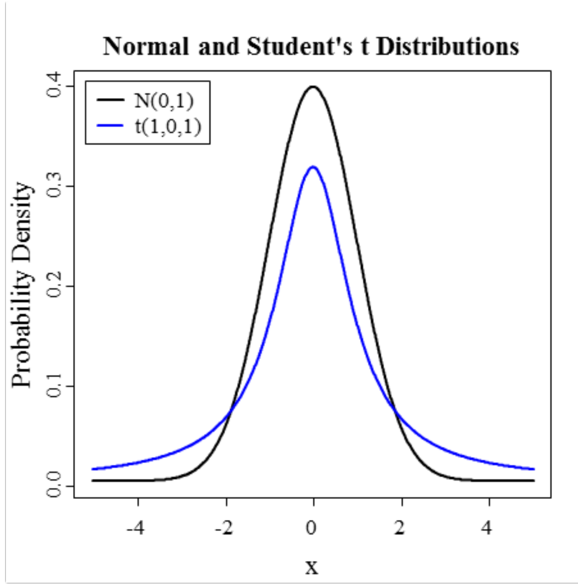

Figure 3.1 Comparison of a $N(0,1)$ distribution (black) and $t(1,0,1)$ distribution (blue).

For our probabilistic statement, we know that, if in reality $\beta_j = 0$, then we might still get a large magnitude for $\hat{\beta}_j$ by random chance, and that this random value for $\hat{\beta}_j$ has a t-distribution with degrees of freedom equal to the sample size minus the number of parameters, center zero, and scale equal to the standard error of $\hat{\beta}_j$ obtained from $\hat{\Sigma}_\beta$:

$$\hat{\beta}_j \mid \beta_j = 0 \;\sim\; t\left(n-p, 0, \sqrt{\widehat{\Sigma_{\beta[J,J]}}}\right).$$

We can standardize this distribution by dividing the estimate by its standard error:

$$T = \frac{\hat{\beta}_j}{\sqrt{\widehat{\Sigma_{\beta[J,J]}}}} \quad \rightarrow \quad T \mid \beta_j = 0 \;\sim\; t(n-p, 0, 1).$$

Using this distribution, we can find the probability that the absolute value of T would be at least a big as we found in our data if the true effect of β_j were zero. This probability is called a p-value. Consider the concrete values $\nu = 20$, $\hat{\beta}_j = 0.5$, and $\sqrt{\widehat{\Sigma_{\beta[J,J]}}} = 0.2$, which give us a value of $T^* = 2.5$ (where T^* is the particular value of T that we have observed in our data). In Figure 3.2, the plot of a $t(20, 0, 1)$ distribution is given, with the area under the curve where $|T| > |T^*|$ shaded in. The total area of the two shaded regions is the p-value, which in this case is approximately $p = .011$.

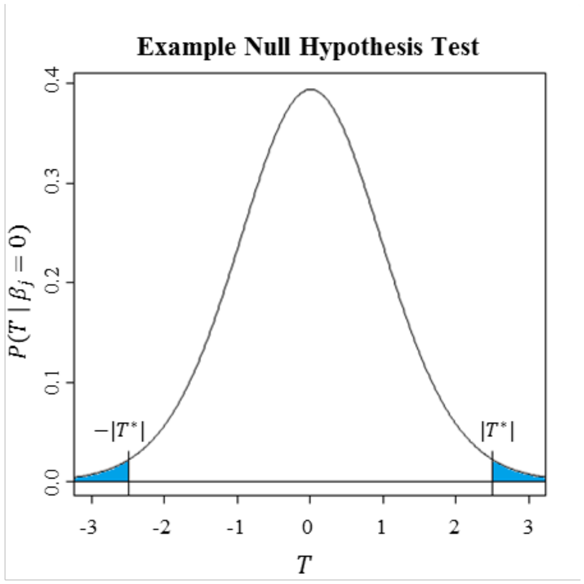

Figure 3.2 Visualization of what a frequentist p-value represents. The density curve is the distribution of $\hat{\beta}_j$ with its standard error divided out (that is, T) under $H_0\colon \beta_j = 0$. The area of the shaded regions where the absolute value of T is greater than the test statistic T^* is the p-value.

This p-value is then used to make a decision of whether to reject the null hypothesis or accept it based on a pre-defined α-level (in linguistics, almost always $\alpha = .05$). If $p < \alpha$ (as it is in this example), we reject the null hypothesis and argue that the data support the alternative hypothesis that $\beta_j \neq 0$; if we are incorrect in this decision, we have committed a “Type I error” (i.e. false positive). If $p \geq \alpha$, we accept (or “fail to reject”) the null hypothesis and argue that we don’t have strong evidence that $\beta_j \neq 0$; if we are incorrect in this decision, we have committed a “Type II error” (i.e. a false negative) (Weisberg, 2005, p. 31). Importantly, the interpretation of the p-value is *not* the probability that our hypothesis is incorrect. The p-value is only interpretable within the counterfactual statement that we set up. It tells us “if there were no effect, the probability I would get at least this extreme of an estimate is <p-value>”; it does *not* tell us “the probability that there is no effect is <p-value>”. In other words, the curve in Figure 3.2 is not the distribution for $P(\beta_j \mid y)$, but rather the distribution for $P(\hat{\beta}_j \mid \beta_j = 0)$ with the standard error divided out. This distinction is important, because in Bayesian regression, we *can* obtain the probability that an effect is positive or negative, as described in Section 3.5.2. The elements just described are what can be found in standard regression output from the `lm` function in R. The *estimate* column is $\hat{\beta}$, the *standard error* column is the square root of the diagonal elements of $\hat{\Sigma}_\beta$, the *t value* column is T^*, and the *p* column is the p-value under the null Hypothesis $H_0\colon \beta_j = 0$ for each coefficient.

3.2.2. Linear mixed effects regression

An assumption of OLS is that the errors for individual observations are independent of one another, and that σ_ϵ is the only source of variance in the dependent variable (Demidenko, 2013, pp. 2–3). In any study that involves repeated measures on members of a group (e.g. subjects and/or items), these assumptions are violated, as differences among individual members of the grouping factor introduce additional variance components, and when these variance components are not modeled, the errors for individual group members are correlated. The differences between members of the grouping factor must thus be accounted for in a mixed effects model (i.e. one that contains both fixed and random effects). That is, in addition to β, Σ_β, and σ_ϵ, we must also estimate the random effects coefficients γ (i.e. the random intercepts and random slopes for individual group members, which are the noise introduced by individual variation for the intercept term and any non-intercept term, respectively), which are assumed to be multivariate normally distributed with mean 0 and unknown covariance matrix Σ_γ. Letting Z be the matrix of the random effects features and MN be the multivariate normal distribution, this results in the following more complex model (Demidenko, 2013, pp. 5–7):

$$y = X\beta + Z\gamma + \epsilon, \qquad \gamma \sim MN(0, \Sigma_\gamma), \qquad \epsilon \sim N(0, \sigma_\epsilon) \quad \rightarrow \quad y \sim N(X\beta + Z\gamma, \sigma_\epsilon).$$

3.2.3. Iterative algorithmic estimation

In mixed effects models, no closed-form solution exists (i.e. estimation of the parameters is not a simple math problem as it is in OLS), and instead the solution is arrived at iteratively. In the R package `lme4` (Bates, Maechler, Bolker, & Walker, 2015), which is the package that has become standard in linguistics and other language sciences and takes a frequentist approach, the estimates of the fixed effects $\hat{\beta}$ and the overall error covariance matrix (which contains both the random effects covariance matrix $\hat{\Sigma}_\gamma$ and the residual variance $\hat{\sigma}_\epsilon^2$) are optimized iteratively, and the estimation stops when consecutive iterations change very little (i.e. the change in the gradient falls below a certain tolerance). The default in `lmer` is to use restricted maximum likelihood estimation (REML). Once a solution for $\hat{\beta}$, $\hat{\Sigma}_\gamma$, and $\hat{\sigma}_\epsilon^2$ is arrived at, $\hat{\Sigma}_\beta$ is computed (it deterministically related to the estimated parameters), and $\hat{\gamma}$ are estimated as ancillary parameters (Bates, Maechler, et al., 2015). There are then several functions in the `afex` (Singmann, Bolker, & Westfall, 2015), `pbkrtest` (Halekoh & Højsgaard, 2014) and `lsmeans` (Lenth & Hervao, 2015) packages that allow frequentist inference (i.e. p-values as described for OLS but with some additional algorithmic estimation required) to be made based on the model. There are many different optimization algorithms that can be used with `lme4`, and also entirely different approaches to linear mixed effects model estimation that do not use REML such as

Bayesian mixed effects regression, which is used in this dissertation and described in detail in Section 3.3.

3.2.4. Maximal random effects and model identifiability

Following Barr et al. (2013), I argue that the random effects structure in a mixed effects regression should be kept maximal. The maximal random effects structure is defined by random intercepts for the members of a grouping factor (e.g. subjects or items), as well as random slopes for any fixed effect that can vary within individual members of the grouping factor (e.g. in an analysis of spontaneous speech where F0 and speaker sex are included as fixed effects, speakers should have random intercepts and random slopes for F0, but not random slopes for sex, since each speaker has more than one value for F0 but only one value for sex). Some authors argue that as the random effects structure becomes more complex, the maximal model becomes unidentifiable, and simplification is required (e.g. Bates, Kliegl, Vasishth, & Baayen, 2015). However, from a statistical perspective, a linear mixed effects model is identifiable so long as the random effect feature matrix is full rank (i.e. there is no collinearity) for at least one random effect group member (Demidenko, 2013, pp. 117–120). Thus, we must be careful to separate *model identifiability* from *algorithmic estimability*; the former is a theoretical issue and the latter is a computational issue.

The maximal model is only unidentifiable when any of the following do not hold: (1) X is full rank (i.e. does not contain collinear terms); (2) there are more observations than parameters; and (3) Z is full rank for at least one group member for each grouping factor. In the case that the maximal model is truly unidentifiable, then I argue that a principled dimensionality reduction method should be applied to the fixed effects so that a maximal model can be fit. The reason for this is that fitting a less than maximal random effects structure implies prior knowledge on the part of the researcher that the variance in the random slopes for some of the fixed effects is exactly zero, which is not reasonable *a priori*. If the fixed effects structure the researcher wants to model can only be modeled based on this assumption, then the dataset is simply too small or too sparse to properly evaluate the researchers' hypotheses. If the maximal model is identifiable, it is still possible that the algorithm described in Section 3.2.3 will fail to converge to a solution, or converge to a statistically invalid solution. In this case, rather than allowing the algorithm to dictate the use of a less than maximal model, I argue that another approach to estimation of the maximal model should be taken, such as the Bayesian approach taken in this dissertation and described in detail in Section 3.3. For further discussion and examples of maximal models that fail to converge with frequentist estimation, but do converge with Bayesian estimation, see Kimball, Shantz, Eager, and Roy (2016).

3.2.5. Reasons for taking a Bayesian rather than frequentist approach

For the research questions raised in this dissertation, the random effects are not ancillary (as they are in the frequentist approach just described), but rather objects of inquiry. While maximal random effects models give better, more principled estimates of the fixed effects parameters, the variation among individual speakers is also crucial to the hypotheses laid out in Section 1.6. For this purpose, mixed effects models also offer better estimates of individual variation than running regressions on individuals, as fitting a mixed effects model treats all speakers as coming from a common population and estimates individual variation using all of the information available in a dataset, and running individual regressions would treat each speaker as their own population (which is not a reasonable assumption). Some previous research in sociolinguistics makes use of random intercepts from frequentist mixed effects regressions for inference about individual group member variation (e.g. Drager & Hay, 2012). However, with the standard frequentist approach, there is no p-value associated with the correlation parameters in the random effects, and no distributional information for the individual elements of $\hat{\gamma}$ (the individual group member behavior), but rather only point estimates.[19] For this reason, in this dissertation a Bayesian approach to mixed effects model estimation is used, as they include estimation of the individual random effects in the model rather than only the covariance matrix, and are better suited to answering the questions at hand. As described in Section 3.5.2, taking a Bayesian approach also allows us to move past a NHST approach; that is, rather than making probabilistic statements about $P(data \mid hypothesis)$, a Bayesian approach allows us to make probabilistic statements about $P(hypothesis \mid data)$, which is what we are actually interested in, because it gives us distributional information about $P(\beta \mid y)$ and $P(\gamma \mid y)$.

3.3. Bayesian mixed effects regression

3.3.1. Bayesian statistics

In frequentist statistics (e.g. mixed effects regression in `lme4`), the p-values obtained represent the probability of having data that show as extreme or more extreme of an effect as the data in the model when the true effect of a parameter is assumed to be exactly zero (null hypothesis significance testing). They are *not* the probability that an effect is non-zero, nor the probability that an alternative hypothesis is correct, as discussed in the previous section (Cowles, 2013, pp. 52–53). In Bayesian statistics, the law of total probability is used to obtain what researchers often assume frequentist p-values refer to (i.e. the probability of the hypothesis conditioned on the data). Letting $\theta = \{\beta,\ \Sigma_\beta,\ \gamma,\ \Sigma_\gamma,\ \sigma_\epsilon\}$ (i.e. all of the model parameters

---

[19] In the case of random intercepts, under certain conditions, p-values can be obtained, and a null hypothesis significance test can be obtained for the entire covariance matrix (Demidenko, 2013, pp. 133–137), but the hypotheses related to the random effects in this dissertation cannot.

considered jointly) in a linear mixed effects regression as described in Section 3.2.2, the law of total probability gives us (Cowles, 2013, pp. 7–8):

$$P(\theta \mid y) = \frac{P(y \mid \theta)P(\theta)}{P(y)},$$

where $P(y \mid \theta)$ is the likelihood function just as before (the probability of the data given a value for the model parameters), $P(\theta)$ is the prior probability distribution for the parameters, $P(y)$ is a normalizing constant obtained by integrating over all possible values of the parameters (this is in practice not necessary to estimate), and $P(\theta \mid y)$ is the posterior distribution of the parameters of interest (the probability distribution of the parameters given the data; i.e. what we are actually interested in). The basic idea is that if we can make reasonable *a priori* assumptions about the probability distributions for the parameters in the mixed effects model and express these mathematically through $P(\theta)$, then we can use the data to update our assumptions through the likelihood function $P(y \mid \theta)$, resulting in *posterior* distributions for the parameters (i.e. posterior in the sense of distributional information after observing the data), expressed mathematically through $P(\theta \mid y)$.

3.3.2. Weakly informative priors

In this dissertation, I use weakly informative prior distributions for the parameters (Gelman, 2006; Gelman, Jakulin, Pittau, & Su, 2008; Stan Development Team, 2016b, pp. 123–127). By “weakly informative”, I mean that, before running the regression, we assume the most likely scenario is that there are no effects at all (i.e. zero is the most likely value for all of the parameters), that larger effect magnitudes are less likely than smaller effect magnitudes (i.e. as we move away from zero, the probability of the effect becomes gradiently less likely), and positive and negative effects are equally likely (i.e. the distribution is symmetric). In this section, I discuss the general shape of reasonable weakly informative prior distributions, and return to the exact values I set for their parameters in Section 3.4.3, after discussing the continuous variable scaling and factor contrasts that make the parameter values reasonable in Sections 3.4.1 and 3.4.2. The alternatives to weakly informative priors are *informative* priors and *non-informative* priors. With informative priors, the mean for a parameter's prior is non-zero; in the absence of an extensive literature offering a reasonable value for an informative prior's mean, this risks biasing the results towards our hypotheses, and so is not done here. Non-informative priors treat all possible parameter values as equally likely (similar to frequentist approaches); as explained throughout this section, calling this approach "non-informative" is a misnomer; they are actually informative in an unreasonable manner.

For the fixed effects β, the possible values for the parameters range from $-\infty$ to $+\infty$, so we want a distribution that is symmetric around a peak at zero, slopes off in both directions, and has tails that extend infinitely. The t-distribution described in Section 3.2.1 is a very good candidate that matches this description. Setting aside the values for the degrees of freedom ν and scale s for now (we will return to the choice for these parameters in Section 3.4.3), we get a prior distribution with the shape shown in Figure 3.3 when we center the distribution at zero.

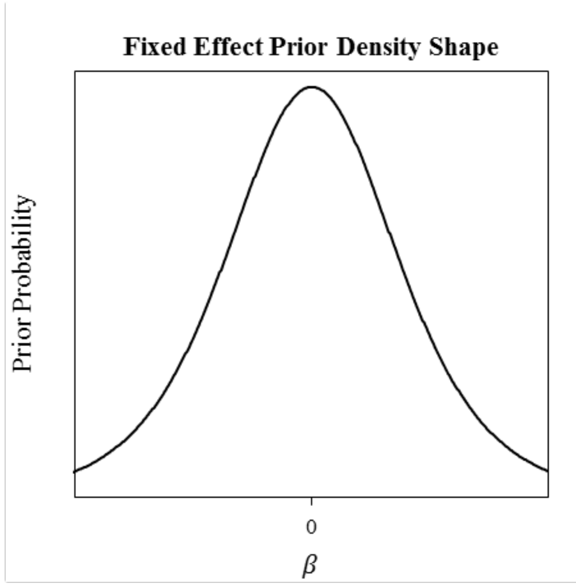

Figure 3.3 General shape of a reasonable prior for the fixed effects. The most likely effect is zero, negative and positive effects are equally likely, larger effect magnitudes are less likely than smaller effect magnitudes, and all effects are possible.

In this way, we are making a reasonable assumption about what the relative values of the fixed effects are likely to be *a priori*: we expect them to be random noise distributed about zero. In a frequentist regression, the distribution is implicitly assumed to be uniform across all possible values (i.e. all values for β are equally likely; an effect of +100 is just as likely as an effect of +1 or 0). This assumption is often referred to as "non-informative", but is in fact informative in the sense that it still makes a claim about the relative prior probability of different possible values for a parameter; it is merely the case that these claims are unreasonable. Another way to think of the weakly informative fixed effects prior distribution is that we are assuming that the null hypothesis is the most likely scenario absent data, while not actually assuming that it is true in our inference (i.e. we also assume that the non-zero values are possible, but less likely in a gradient manner as effect magnitude increases). We then allow the data to inform our *posterior* beliefs given this reasonable prior assumption (i.e. we require the data to show evidence that our prior belief that the parameters are just noise is wrong). The likelihood function $P(y \mid \theta)$ thus pulls the posterior $P(\theta \mid y)$ toward itself and away from the prior $P(\theta)$, and the more data we have, the stronger the effect of the likelihood's pull is (Cowles, 2013). We do not place a direct prior on Σ_β, and so our prior for β implies no correlation among the fixed effects; however, this does not imply that in the posterior they are uncorrelated; we are simply allowing the data to determine their correlation. We apply similar reasoning as we did for β to the random effects covariance matrix $\hat{\Sigma}_\gamma$, the residual standard error σ_ϵ, and the random effect coefficients γ.

Rather than modeling the random effects covariance matrix Σ_γ directly, we decompose it into its two components: the vector of standard deviations for each effect, σ_γ, and the correlation matrix for these effects, Ω_γ; any covariance matrix is uniquely defined by these two components (i.e. $\Sigma = \text{diag}(\sigma)\Omega\text{diag}(\sigma)^T$), and modeling them separately is more computationally efficient (Stan Development Team, 2016b, pp. 326–332); also note that in standard `lmer` output, these two components are presented separately. In the case of the correlation matrix, Ω_γ, we want a similar distribution as we found for β, except the range of values needs to be constrained to $[-1, 1]$, as values outside of this range are statistically invalid. That is, we want a distribution with the shape shown in Figure 3.4 (we will return to this distribution's parameterization in Section 3.4.3). In frequentist regression, the distribution for these correlation parameters is, again, uniform: the implicit assumption is that two random effects are just as likely to be perfectly correlated as they are to be weakly correlated or completely uncorrelated.

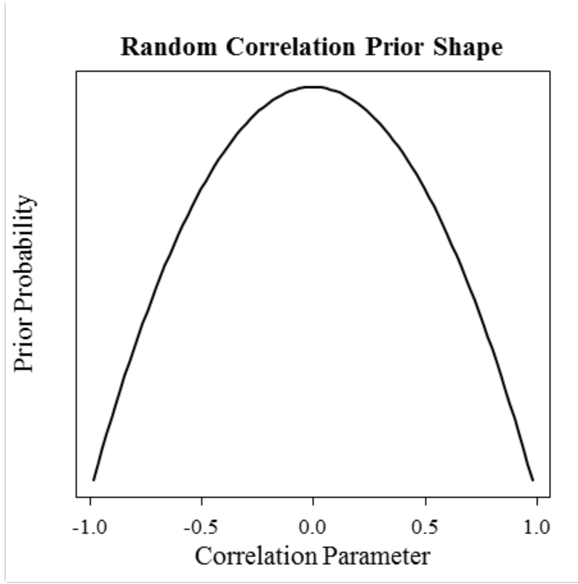

Figure 3.4 General shape of a reasonable prior for a correlation coefficient. The most likely correlation is no correlation, positive and negative correlations are equally likely, larger correlations are less likely than smaller correlations, and all valid parameter values (i.e. [-1, 1]) are possible.

For the standard deviations in the random effects, σ_γ, and the residual standard error, σ_ϵ, only positive values are possible by definition, so our distribution will not be symmetric about zero, but our other requirements (peak at zero and gradient drop off as magnitude increases) are still easily met with a half-normal distribution (i.e. the shape is the same as the positive half of the normal distribution), as shown in Figure 3.5. Again, in frequentist regression the implied assumption is that all possible values for the standard deviations are equally likely (i.e. uniformly distributed from 0 to $+\infty$).

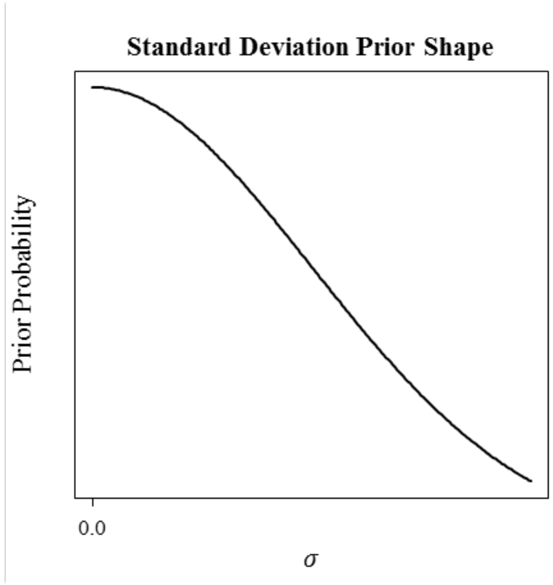

Figure 3.5 General shape of a reasonable prior for a standard deviation parameter. The most likely standard deviation is zero, larger standard deviations are less likely than smaller standard deviations, and all valid parameter values (i.e. positive values) are possible.

For the random intercepts and slopes themselves, γ, we use the same prior assumption as in frequentist regression. That is, the random effects are multivariate normally distributed with mean zero and covariance Σ_γ: $\gamma \sim MN(0, \Sigma_\gamma)$. The question that then remains is how to choose reasonable numeric values for the parameters of these prior distributions. This is indeed possible, provided that the dependent variable y and the predictors X and Z are all transformed to be on the same scale, as described in the following section.

3.4. Scaling, contrasts, and weakly informative prior parameter values

The scaling and contrast practices described in this section can be implemented with the `standardize` (Eager, 2017b) and `nauf` (Eager, 2017a) packages in R. The goal of all of these practices is to keep the regression parameters on the same scale, and to ensure that the intercept (which is the predicted value of an observation when all other coefficients are multiplied by zero) represents the corrected mean (i.e. the predicted value for an observation that is average in every way, holding covariates at their mean values and averaging over group differences in factors).

3.4.1. *Continuous variable scaling*

Continuous variables include covariates (i.e. fixed effects that take on continuous values) and the dependent variable in linear regression (Weisberg, 2005). All continuous variables are placed on unit scale by subtracting the mean of the variable and dividing by the standard deviation of the variable (also sometimes called z-scoring or simply scaling). The result is that the values in the transformed variable have the same relationship to one another as in the untransformed variable, but the transformed variable has mean 0 and standard deviation 1. This places all of the regression coefficients for covariates on the same scale, with the regression coefficient representing the predicted change in standard deviations of the dependent variable associated with a one standard deviation increase in the covariate (i.e. a 1-SD increase in the covariate leads to a predicted $\hat{\beta}$-SD change in the dependent variable).

3.4.1.1. Dependent variables

For both the plosive and fricative analyses, the acoustic measures were found to be highly correlated in Section 2.5, and PCA was performed. In both cases, higher values of PC1 indicate stronger consonants than lower values. Looking first at plosive PC1, Figure 3.6 and Table 3.1 provide descriptive statistics for PC1 by underlying voicing, and show that the PCA successfully combined the acoustic measures into a continuum of plosive strength that easily distinguishes underlyingly voiced plosives from underlyingly voiceless plosives and has good coverage across its range of values.

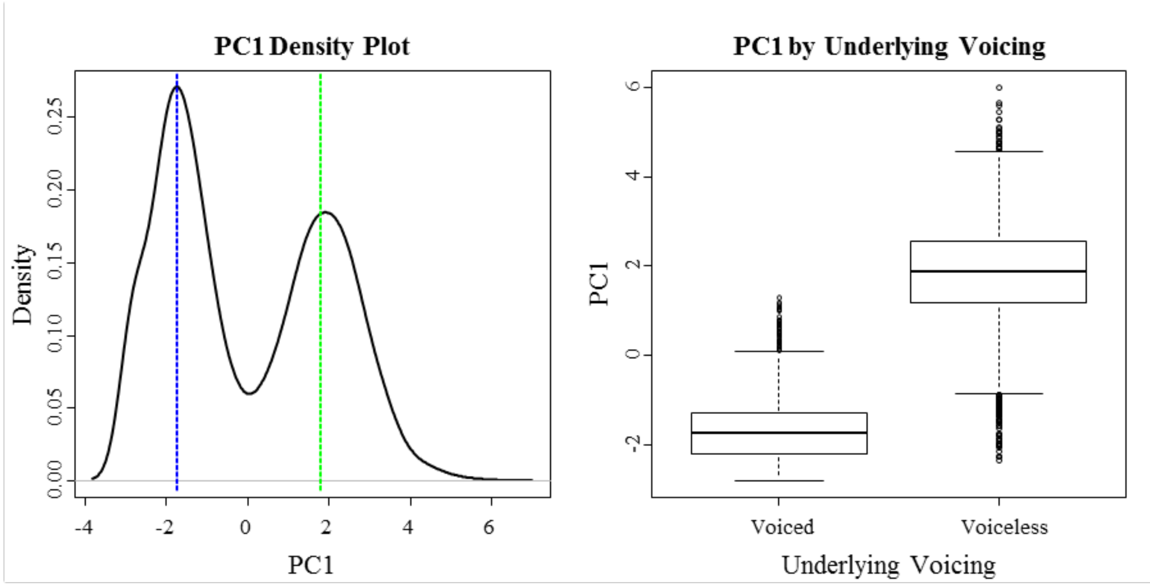

Figure 3.6 Density plot of plosive PC1 with verticle lines at the mean for undelrying voiced plosives (blue line on left side) and underlyingly voiceless plosives (green line on right side) (left panel) and boxplot of plosive PC1 by underlying voicing (right panel).

Table 3.1 Descriptive statistics for plosive PC1 by underlying voicing.

Voicing	N	Minimum	Median	Maximum	Mean	SD
Voiced	2694	-2.816	-1.755	1.292	-1.729	0.734
Voiceless	2587	-2.334	1.870	6.010	1.801	1.172

As can be seen in Figure 3.6, PC1 has two modes, one centered around -1.75 corresponding to mostly underlyingly voiced plosives and one around +1.75 corresponding to mostly underlyingly voiceless plosives. However, the standard deviations for the two groups are very different, with /ptk/ having a standard deviation nearly one and a half times that of /bdg/ (Table 3.1). Even if PC1 were put on unit scale, a one standard deviation increase in scaled PC1 would mean very different things for voiced and voiceless plosives (i.e. an effect size of +1 in a regression on scaled PC1 would mean an increase of 1.72 within-underlying-voicing standard deviations for /ptk/ and 2.75 within-underlying-voicing standard deviations for /bdg/). As the goal of the plosive study is to analyze the factors that affect the *relative* lenition of the two sets of plosives within-speaker and across dialects, the PC1 values for each observation were mean-centered and scaled according to the observation's underlying voicing, resulting in a measure that I will call "Voicing-Normalized PC1" and abbreviate VNPC1 (i.e. using the values in Table 3.1, for /bdg, $VNPC1 = (PC1 + 1.729)\ /0.734$, and for /ptk/, $VNPC1 = (PC1 - 1.801)/1.172$). VNPC1 measures how strong a plosive is given its underlying voicing, with zero indicating an average strength, positive values indicating a stronger than average plosive, and negative values indicating a weaker than average plosive, with magnitude being interpreted in terms of within-underlying-voicing standard deviations. In a regression, this has the effect of removing the main effect of underlying voicing (which is already clearly substantial in Figure 3.6 and Table 3.1) and making the interaction of underlying voicing with other predictors straightforwardly interpreted. Thus, VNPC1 will serve as the dependent variable for the remainder of the plosive analyses.

Turning now to fricative PC1, Table 3.2 provides descriptive statistics for PC1 by language and fricative phoneme, with corresponding boxplots in Figure 3.7. The standard deviation of fricative PC1 is nearly identical across the four groups, and so PC1 was scaled across the entire fricative dataset. I will refer to this dependent variable as "Normalized PC1" and abbreviate it as NPC1.

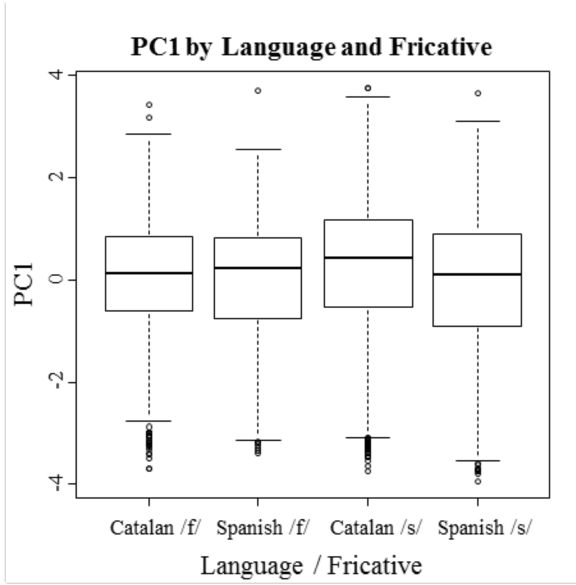

Figure 3.7 Boxplot of fricative PC1 by language and fricative phoneme.

Table 3.2 Descriptive statistics for fricative PC1 by language and fricative phoneme.

Language	Fricative	N	Minimum	Median	Maximum	Mean	SD
Catalan	/f/	269	-3.684	0.138	3.433	-0.085	1.437
	/s/	964	-3.725	0.431	3.777	0.202	1.519
Spanish	/f/	165	-3.366	0.218	3.725	-0.104	1.501
	/s/	765	-3.946	0.107	3.666	-0.203	1.536

3.4.1.2. Covariates

In both the plosive and fricative regressions, there are two covariates: log word frequency and speech rate. Because speech rate was speaker-normalized, it is already on unit scale. For log word frequency, however, this is not the case. For example, in the plosive data, the mean log word frequency is 8.2, the standard deviation is 2.4, and a value of 0 indicates that a word occurred only once in the CREA corpus. If left unscaled, the intercept would represent the predicted strength for a plosive whose word is as infrequent as possible, but was produced with the speaker's average speech rate. While the coefficient for speech rate would represent the change in plosive strength for a 1-SD increase in the speaker's speech rate, the coefficient for log word frequency would represent the change in plosive strength for an increase of 1 log-count, which is an increase of 0.42 standard deviations. By scaling log word frequency, the intercept again represents the corrected mean: the predicted strength for a plosive whose word is of average frequency, and that was produced with the speaker's average speech rate; and the regression coefficient for log word frequency represents the change in standard deviations of plosive strength predicted for a 1-SD increase in log word frequency, making it directly comparable to the coefficient for speech rate (i.e. the coefficients are on the same scale). For the plosive regression, log word frequency was scaled for the entire data set. For the fricative dataset, however, the log-counts for Spanish and Catalan come from different corpora (with different total numbers of words), and so for the fricative data, log word frequency was scaled within language, so that the covariate is on unit scale, and represents how frequent a word is in relation to the other words from that corpus in the data.

3.4.2. Factor contrasts

Factors are variables that take on a defined set of categorical values called levels rather than continuous values. In regression, a factor with K levels is modeled through the use of $K-1$ dummy variables (Agresti, 2002; Davis, 2010). Each level of the factor is assigned a value for each dummy variable based on a contrast matrix. So, for example, a factor with four levels has a contrast matrix with four rows (one for each level) and three columns (one for each dummy variable), with the values in the cells of the matrix determining the numerical expression for the factor levels in the dummy variables. There are two general types of factors, ordered and unordered, whose contrasts are treated differently. In this dissertation, all factors are unordered.

Unordered factors take on two or more categorical values that are not intrinsically ordered (or have a somewhat ordered interpretation but there are only two categories, as is sometimes the case with factors coded as false vs. true, 0 vs. 1, or no vs. yes). The factors included in the fricative analysis are language, sex, fricative phoneme identity, stress, preceding vowel height, and following vowel height. The factors included in the plosive analysis are dialect, sex, underling voicing, place of articulation, stress, preceding vowel height, following vowel height, word position, task, age group, education level, and Quechua bilingualism. For unordered factors, the default in R is to use treatment contrasts, where the first level is coded as 0 for all of the dummy variables, and the remaining levels each have a dummy variable for which they are coded +1, and are then coded as 0 for the other dummy variables. Using stress as an example, this would lead to the contrasts in Table 3.3 (recall from Chapter 2 that here stress refers to consonants, and so the levels of the factor describe the syllabic stress context).

Table 3.3 Treatment contrasts for stress.

Stress	Stress, Tonic	Stress, Unstressed
Post-Tonic	0	0
Tonic	1	0
Unstressed	0	1

With treatment contrasts, the intercept loses the interpretation of the corrected mean, since when all of the dummy variables in Table 3.3 are multiplied by zero, the resulting value corresponds to post-tonic consonants. To avoid this (and to ensure that the coefficients for the dummy variables stay, on average, closer to zero, but without altering the ultimate interpretation of the results), sum contrasts are used. With sum contrasts in R, the first $K - 1$ levels each get a dummy variable for which they are coded +1, and then are valued 0 for the other dummy variables. The last level is assigned a value of -1 for all of the dummy variables. Sum contrasts also have additional computational benefits in comparison to treatment contrasts for similar reasons as covariate scaling. For our stress example, this results in the contrast matrix given in Table 3.4.

Table 3.4 Sum contrasts for stress.

Stress	Stress, Post-Tonic	Stress, Tonic
Post-Tonic	1	0
Tonic	0	1
Unstressed	-1	-1

In the regressions in this dissertation, sum contrasts are used for all factors. With sum contrasts, the intercept maintains the interpretation of the corrected mean, since when all of the dummy variable coefficients are multiplied by zero, it averages over their effects (note that no row in the contrast matrix in Table 3.4 has all zeros, and thus multiplying all of the coefficients by zero cannot describe any one level; rather, the mean of the values in each column is zero, and so multiplying all of the dummy variable coefficients by zero averages over their effects). Another important advantage of using sum contrasts in this dissertation is that it allows for observations to be coded as `NA` (in sociolinguistics this is often referred to as slashing).

In the context of factors, I use `NA` to mean "not applicable", which is conceptually different than "not available" or "missing at random". The concept applies only to unordered factors, and indicates that the factor is simply not meaningful for an observation (either for theoretical reasons, or for practical reasons related to the sampling scheme), or that while the observation may technically be definable by one of the factor levels, the interpretation of its belonging to that group isn't the same. While no factors in the fricative analysis require `NA` coding, several factors in the plosive analysis do. First and foremost, the social factors of age group, education level, and Quechua bilingualism are only contrastive in the subset of the plosive data pertaining to Cuzco. The Lima and Valladolid speakers are all younger, university educated, and monolingual speakers of Spanish; however, coding them as such creates collinearity in the posterior estimates of these social factors and dialect. Additionally, in the case of Quechua bilingualism, the factor does not have the same interpretation across dialects. In

Cuzco, a monolingual speaker is *monolingual in Spanish as opposed to bilingual in Spanish and Quechua*, while in Valladolid, for example, the interpretation is not the same. For this reason, age group, education level, and Quechua bilingualism were coded as NA for all observations pertaining to the Lima and Valladolid dialects. In addition to these social factors, task and word position also require some NA coding. Due to the imbalanced nature of the data and the sampling scheme, task is only contrastive for the Cuzco and Lima subsets, and word position is only contrastive for spontaneous speech (in all three dialects), since all of the planned observations in the read speech task were word-medial. For this reason, task was coded as NA for all Valladolid observations, and as either "read speech" or "spontaneous speech" for all Cuzco and Lima observations, and word position was coded as either "initial" or "medial" for all spontaneous speech observations (Valladolid task-oriented dialogues and Cuzco and Lima interviews), and as NA in the Cuzco and Lima read speech data.

For factors with NA values, sum contrasts can be used ignoring the NA values, and then afterwards the NA values can be set to zero in the model matrix (i.e. X and Z) for all dummy variables. For dialect, age group, education level, and Quechua bilingualism, this results in the contrasts in Table 3.5, Table 3.6, and Table 3.7. For dialect, task, and word position, this results in the contrasts in Table 3.8.

Table 3.5 Contrasts for age group.

Dialect	Age Group	Dialect, Cuzco	Dialect, Lima	Age Group, Older
Cuzco	Older	1	0	1
	Younger	1	0	-1
Lima	NA	0	1	0
Valladolid	NA	-1	-1	0

Table 3.6 Contrasts for education level.

Dialect	Education Level	Dialect, Cuzco	Dialect, Lima	Education Level, Secondary
Cuzco	Secondary	1	0	1
	University	1	0	-1
Lima	NA	0	1	0
Valladolid	NA	-1	-1	0

Table 3.7 Contrasts for Quechua bilingualism.

Dialect	Quechua Bilingual	Dialect, Cuzco	Dialect, Lima	Quechua Bilingual, Yes
Cuzco	Yes	1	0	1
	No	1	0	-1
Lima	NA	0	1	0
Valladolid	NA	-1	-1	0

Table 3.8 Contrasts for task and word position.

Dialect	Task	Word Position	Dialect, Cuzco	Dialect, Lima	Task, Read Speech	Word Position, Initial
Cuzco	Read	NA	1	0	1	0
	Spontaneous	Initial	1	0	-1	1
		Medial	1	0	-1	-1
Lima	Read	NA	0	1	1	0
	Spontaneous	Initial	0	1	-1	1
		Medial	0	1	-1	-1
Valladolid	NA	Initial	-1	-1	0	1
		Medial	-1	-1	0	-1

This setup allows the regression coefficients to only affect the predicted value for observations where the factor is contrastive. For example, for all Lima and Valladolid observations, the coefficient for “Age Group, Older” is multiplied by zero, and it never adds or subtracts from their predicted values. For Cuzco observations, the value obtained by adding “Dialect, Cuzco” to the intercept represents the corrected mean for the entire Cuzco dialect, averaging over the effect of age group, and the coefficient for “Age Group, Older” represents the difference between the predicted value for Cuzco observations pertaining to speakers in the older group and the corrected mean for Cuzco. Similar logic applies to all other contrasts involving `NA` values. When examining descriptive statistics in Chapter 5 for a factor that has NA values, only the subset of the data where the factor is contrastive is considered, and when generating posterior estimates of group differences, care is taken to use contrasts that properly represent the relevant subset of the data (all specific contrasts applied are given in Section B.5).

This same methodology can be extended to the random effects structure, where experimental item is only applicable in the read speech task. In the model matrix for a mixed effects regression, the random effects features Z are represented with a column for each intercept or slope for each group member. These columns are set to 0 for any observation that does not pertain to the particular group member, and are set to the same as the fixed effects values for observations that do pertain to the group member. By setting the value for all item effects to zero for all spontaneous speech observations, we can allow the item effects to apply only to the relevant subset of the data (which is not possible in `lme4`), as demonstrated in Table 3.9, where the random intercept coding and the coding for the random slope for sex are given for different combinations of task, dialect, and sex.

Table 3.9 Example random effects coding for item.

Item	Task (Dialect)	Sex	Item i01		Item i02	
			Intercept	Sex, Female	Intercept	Sex, Female
i01	Read (Cuzco/Lima)	Female	1	1	0	0
i01	Read (Cuzco/Lima)	Male	1	-1	0	0
NA	Spontaneous (Cuzco/Lima)	Female	0	0	0	0
NA	Spontaneous (Cuzco/Lima)	Male	0	0	0	0
NA	NA (Valladolid)	Female	0	0	0	0
NA	NA (Valladolid)	Male	0	0	0	0
i02	Read (Cuzco/Lima)	Female	0	0	1	1
i02	Read (Cuzco/Lima)	Male	0	0	1	-1
NA	Spontaneous (Cuzco/Lima)	Female	0	0	0	0
NA	Spontaneous (Cuzco/Lima)	Male	0	0	0	0
NA	NA (Valladolid)	Female	0	0	0	0
NA	NA (Valladolid)	Male	0	0	0	0

One final modification is required to contrast coding in order to include a dialect slope for item. Because only a subset of the levels for the dialect factor are applicable in the read speech task (when item is not NA, dialect is always either Cuzco or Lima and never Valladolid), the dialect contrasts used in the fixed effects cannot be applied to the dialect slope for items. Taking only the relevant subset of the fixed effects dialect contrast matrix, we have the item dialect slope contrasts in Table 3.10. These, however, would be collinear with the random item intercepts (the item intercept would always be exactly equal to sum of the item slopes "Dialect, Cuzco + Dialect, Lima"). For this reason, sum contrasts were reapplied to dialect within this context, as in Table 3.11. To be clear, the fixed effects contrasts are not altered, but rather only the item dialect slope contrasts are altered. This yields the contrasts in Table 3.12.

Table 3.10 Item dialect slope contrasts that are collinear with item intercepts.

Dialect	Dialect, Cuzco	Dialect, Lima
Cuzco	1	0
Lima	0	1

Table 3.11 Item dialect slope coding that avoids collinearity with item intercepts.

Dialect	Dialect, Cuzco (Item Coding)
Cuzco	1
Lima	-1

Table 3.12 Example of fixed and random item effect coding of dialect.

Item	Dialect	Task	Intercept	Dialect, Cuzco	Dialect, Lima	Task, Read Speech	i01 Intercept	i01 Dialect, Cuzco
i01	Cuzco	Read	1	1	0	1	1	1
i01	Lima	Read	1	0	1	1	1	-1
NA	Cuzco	Spontaneous	1	1	0	-1	0	0
NA	Lima	Spontaneous	1	0	1	-1	0	0
NA	Valladolid	NA	1	-1	-1	0	0	0

The use of `NA` values for the fixed and random effects makes the code necessary to run the model more complicated (though I have created an R package, `nauf` (Eager, 2017a), which can apply the treatment of `NA` values discussed here automatically), and the interpretation of the output also becomes more complex. However, these contrasts importantly allow all of the data to be considered in a single regression. Otherwise the data would need to be binned and analyzed separately, which would ignore, for instance, that both the read speech and interviews provide information about the same speaker effects, and that both Cuzco speakers who match the Lima and Valladolid speakers demographically and Cuzco speakers who do not match them provide information about the Cuzco dialect as a whole. By running a single regression with `NA` coding, all of the information in the data can be modeled simultaneously, yielding better posterior estimates.

3.4.3. Weakly informative prior parameter values

Provided that the guidelines for scaling and contrasts given in Sections 3.4.1 and 3.4.2 are followed, then we can express the prior distributions from Section 3.3.2 mathematically in a principled way, based on those in Gelman (2006), Gelman et al. (2008), and Stan Development Team (2016b, pp. 143–148). The prior $\beta \sim t(5, 0, 2)$ assigns independent Student's t-distribution priors with center 0, scale 2, and 5 degrees of freedom to each fixed effect coefficient. A probability density plot of this distribution is given in Figure 3.8.

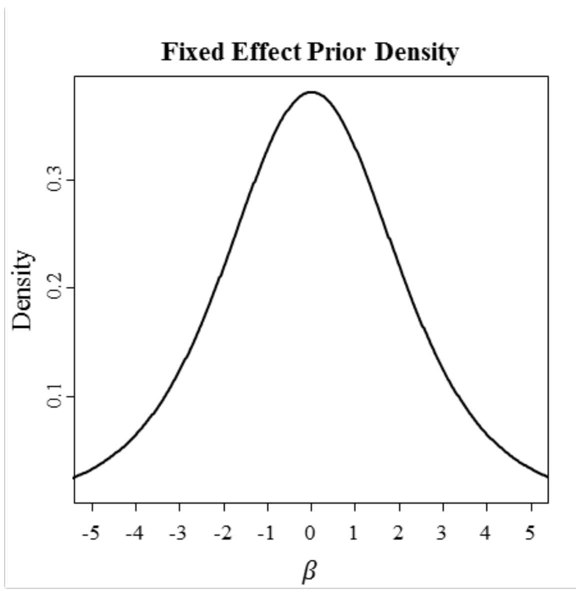

Figure 3.8 Prior density for the fixed effects regression coefficients: $\beta \sim t(5, 0, 2)$.

This fixed effects t-distribution prior has mean 0 and standard deviation of 2.58, making it similar to a normal distribution with mean 0 standard deviation 2.5, but with heavier tails (i.e. allowing for larger effects to be more likely than a normal prior would). When the scaling and contrast guidelines laid out in the previous sections are followed, an effect magnitude greater than 5 is highly unlikely, but we still want to allow for this possibility. The reason such a large effect is unlikely is that, for linear models, an effect of +5 for a covariate would indicate that an increase of 1 standard deviation in the covariate corresponds to an increase of 5 standard deviations in the dependent variable (which would be a huge effect; it would imply that an increase in 2 standard deviations results in an increase of 10 standard deviations, etc.). This reasoning also applies to factors, where this effect would correspond to a departure of 5 standard deviations from the corrected mean (e.g. Stan Development Team, 2016b, p. 126).

Moving on to the prior for σ_ϵ (the residual standard error), with a scaled dependent variable, an intercept-only model (i.e. just guessing 0 for every observation) would result in a residual standard error of 1, with lower values of the standard error indicating a better model fit. For this reason, I use a half-normal prior with scale of 0.5, resulting in the distribution shown in Figure 3.9.

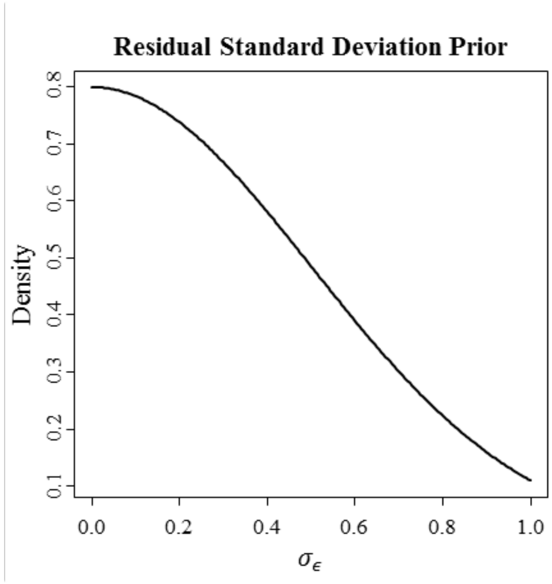

Figure 3.9 Prior density for residual standard error: $\sigma_\epsilon \sim HN(0, 0.5)$.

As mentioned in Section 3.3.2, for the random effects, rather than placing a prior on the covariance matrix Σ_γ directly, priors are placed on the correlation matrix Ω_γ and standard deviations σ_γ, with the covariance matrix Σ_γ being a by-product of the two (Stan Development

Team, 2016b, pp. 143–148, 326–332). The standard deviation in the random intercepts is assigned a half-normal prior with a scale of 1.5 and the random slopes are assigned a half-normal prior with a scale of 1, as it is often the case that the intercepts have more variance than the slopes. Figure 3.10 shows a density plot of these prior distributions.

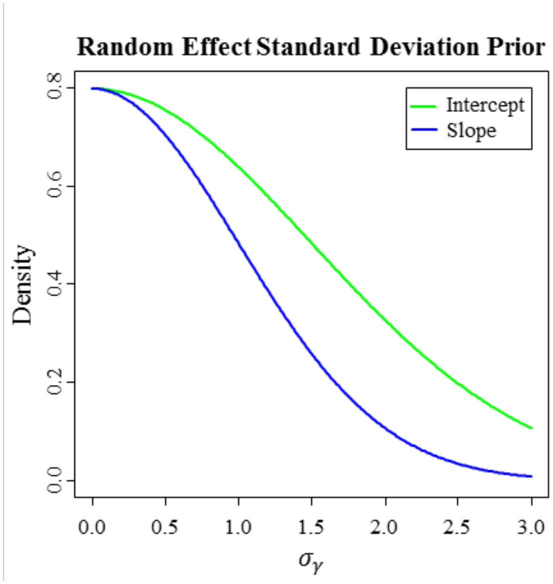

Figure 3.10 Prior density for random effects standard deviations. For intercepts (green), $\sigma_\gamma \sim HN(0, 1.5)$ and for slopes (blue), $\sigma_\gamma \sim HN(0, 1)$.

In the regressions carried out in this dissertation, the correlation parameters in Ω_γ are not directly relevant (or even directly interpretable in the case of factors), but they are important because they underlie the correlations of interest (i.e. they will not tell us directly if speakers' /p/ strengths correlate with their /b/ strengths, but they underlie the posterior estimate of the correlations of interest). The $LKJ(2)$ prior (Lewandowski, Kurowicka, & Joe, 2009) on the random effect correlation matrix favors zero correlation over positive and negative correlations, with positive and negative correlations being equally likely, yielding a distribution that has the properties discussed in Section 3.3.2. The prior is best visualized as the marginal distribution of a single correlation coefficient in the correlation matrix having a symmetric beta prior (adjusted to range -1 to +1), as shown in Figure 3.11.

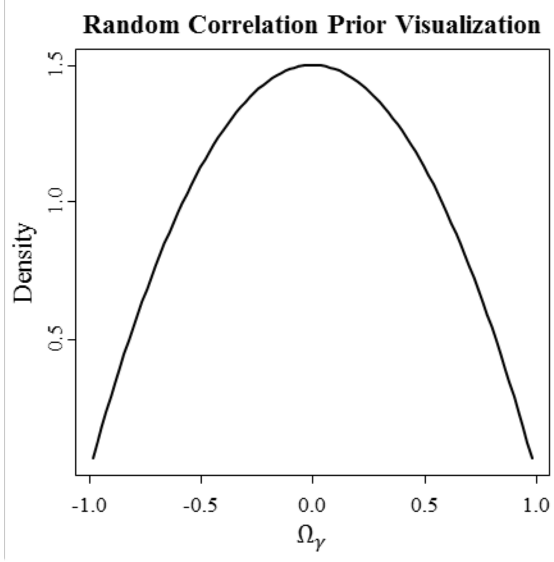

Figure 3.11 Visual approximation of the prior density on a random effects correlation: $\Omega_\gamma \sim LKJ(2)$.

The prior on the random intercepts and slopes themselves is then a multivariate normal distribution with mean 0 and covariance matrix Σ_γ (the same as in `lme4`), with Σ_γ defined by the standard deviations σ_γ and correlation matrix Ω_γ. Using these weakly informative priors, we do not bias the results towards our hypothesis, but rather place reasonable constraints on the distributions of the model parameters (Gelman et al., 2008). The more data there is, the less the priors matter; that is, the data push the posterior away from the prior and toward the likelihood function when there is enough evidence for an effect (Cowles, 2013). To summarize, then, I fit the following model for both the plosive and fricative data:

$$\beta \sim t(5, 0, 2), \qquad \gamma \sim MN(0, \Sigma_\gamma), \qquad \Sigma_\gamma = \text{diag}(\sigma_\gamma)\Omega_\gamma\text{diag}(\sigma_\gamma)^T,$$

$$\sigma_{\gamma_0} \sim HN(0, 1.5), \qquad \sigma_{\gamma_i} \sim HN(0, 1)\ i > 0, \qquad \Omega_\gamma \sim LKJ(2), \qquad \hat{y} = X\beta + Z\gamma,$$

$$\sigma_\epsilon \sim HN(0, 0.5), \qquad y \sim N(\hat{y}, \sigma_\epsilon).$$

3.4.4. *Algorithmic estimation of the posterior distribution*

In simple problems, the exact posterior distribution of the parameters can be determined (i.e. the distribution $P(\theta \mid y)$ has a name and parameters and we can simply use what we know about the distribution to make probabilistic statements about the parameters and our hypotheses). However, in complex problems like mixed effects regression, the posterior distribution does not have a name, and cannot be directly expressed mathematically, and so the posterior distribution is estimated algorithmically with Markov Chain Monte Carlo (Cowles, 2013, pp. 111–124).

3.4.4.1. Monte Carlo estimation

Given independent and identically distributed (iid) random samples from a distribution, the sample characteristics can be used as estimates of the distribution characteristics, which is a technique known as Monte Carlo estimation (Caflisch, 1998; Cowles, 2013, pp. 118–120; Geyer, 2011, pp. 6–7). For example, we know that a $N(3, 2)$ distribution has mean 3 and variance 2^2 = 4; but say that we didn't know this. If we could obtain iid random samples from the distribution, then the sample mean and variance would approximate the true mean and variance, with the standard error in the estimate of the mean being a function of the variance estimate and the sample size. In R, we can do this by executing the following code (where `rnorm` gives random samples from the normal distribution):[20]

[20] The numbers are, of course, pseudo-random, but will be referred to as random throughout the dissertation with the understanding that they are pseudo-random (Kroese, Taimre, & Zdravko, 2011, pp. 1–20, 44–66).

```
n <- 10000
x <- rnorm(n, 3, 2)
mc_mean <- mean(x)
mc_var <- var(x)
mc_se_mean <- sqrt(mc_var / n)
mc_mean
#> 2.986926
mc_var
#> 4.099462
mc_se_mean
#> 0.02024713
```

In this case, our estimate of the mean is 2.987 with a standard error of 0.020, and our estimate of the variance is 4.099. Both of these estimates are close to the true value, and are based on a sample size of 10,000. We could further decrease the error in the estimates by increasing the sample size. Because the numbers are random, the estimates will be slightly different each time, but in each case, provided a large enough sample, the estimates are reasonable approximations of the true values. The samples can similarly be used to estimate quantiles (e.g. a 95% interval), medians, modes, and any other aspect of the distribution. While in this case the true answers are known, Monte Carlo estimation is particularly useful in high-dimensional cases (i.e. when there is a large number of parameters to be estimated simultaneously) where the true distribution is analytically intractable, as is the case for posterior distributions in Bayesian mixed effects regression. But, in order for this estimation method to be applied, a random sample from the distribution must first be obtained. In Bayesian mixed effects regression, these samples are obtained by simulating Markov Chains.

3.4.4.2. Markov Chains

A Markov Chain is a vector (or chain) of temporally related random variables where the probability distribution of the variable at time $n + 1$ depends only on the value (or state) at time n, and not on any previous states, with the initial state at time $n = 1$ having its own distribution

(Geyer, 2011, p. 4; Norris, 1997).[21] For example, consider the simple Markov Chain graphed in Figure 3.12. The Markov Chain has three states (possible values that it can take) that are represented by numbered boxes: 1, 2, and 3. There are arrows going from every box to every box, with the fractions for each arrow representing transition probabilities. So, for example, if at time n, the Markov Chain is at state 3, then at time $n + 1$, the probability the chain stays at state 3 is ¼, the probability that it moves to state 2 is ¼, and the probability that it moves to state 1 is ½. At time $n - 1$, the chain could have been at any of the three states, and which state it was at has no effect on the transition probabilities for moving to time $n + 1$; only the state at time n determines the transition probabilities for the state at time $n + 1$.

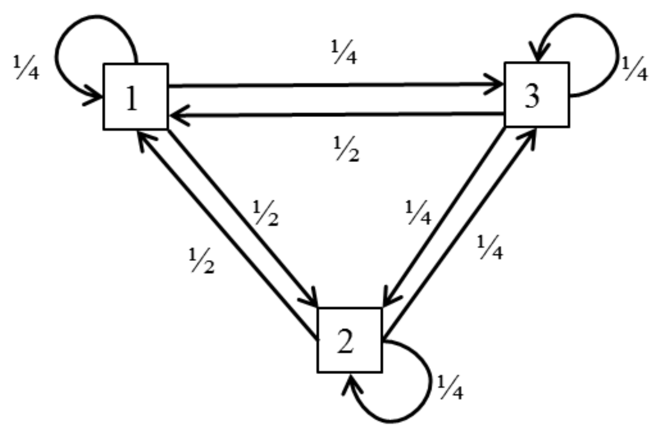

Figure 3.12 Example of a simple Markov Chain with three states.

If a Markov Chain meets certain requirements (it is irreducible (Norris, 1997, pp. 10–12), positive-recurrent (Norris, 1997, pp. 33–39, 118–121), and aperiodic (Norris, 1997, pp. 40–47)), then, regardless of the initial state at time $n = 1$, the chain will converge to a stationary or equilibrium distribution π; that is, given enough time, the overall proportion of time that is spent

[21] For the discussion of Markov Chains in this dissertation, it is more convenient to define the chain as starting at time $n = 1$, but often when talking about the mathematical definition of a Markov Chain, the initial state time is referenced as $n = 0$.

at each state converges to a distribution called the equilibrium distribution (Geyer, 2011, pp. 5–6; Norris, 1997, pp. 33–47, 117–123). If the transition probabilities satisfy the detailed balance equations, then Markov Chain is further called reversible (Norris, 1997, pp. 47–51, 123–125). In Markov Chain Monte Carlo, a much more complex Markov Chain than the one given in Figure 3.12 is simulated, where the state of the Markov Chain is a vector of values for all of the parameters in the model (i.e. the fixed and random effects), and the transition probabilities are drawn at each step (or iteration) from a proposal distribution that relies on the values of the parameters at the current state of the chain (Geyer, 2011).

3.4.4.3. Markov Chain Monte Carlo

The goal of Markov Chain Monte Carlo (MCMC) is to set up a reversible Markov Chain whose equilibrium distribution π is the posterior distribution of the parameters of interest (i.e. $\pi = P(\theta \mid y)$). The Markov Chain is given a random initial state (a set of random values for all of the parameters being estimated) and then a proposal for a possible next state is randomly generated based on the current values of the parameters, and the chain then either stays at the current state or moves to the proposed state. The probability that the chain moves to the proposed state is based on the relative probability of the two states as determined by the priors and the likelihood function jointly.[22] This process is then repeated, treating the new state as if it were the initial state, and ignoring the previous state. The evolution of the chain is simulated for a large number of iterations, discarding a number of initial iterations in which the chain may not yet have reached the equilibrium distribution (known as burn-in or warm-up iterations), resulting

[22] For some MCMC algorithms, such as the Gibbs sampler, all proposals are accepted by design, and for others, such as the independence sampler, the proposal distribution does not depend on the current state, but neither of these is discussed here; see (Geyer, 2011).

in random samples from the posterior distribution $P(\theta \mid y)$. The characteristics of the posterior are then estimated by applying Monte Carlo methods to the samples (Geyer, 2011).

To give the researcher confidence that the equilibrium distribution (i.e. posterior distribution) has been reached, multiple chains are run from diffuse starting points (i.e. starting points that are not too similar too each other, or else running multiple chains doesn't really increase our confidence; in practice these starting points are usually randomly generated), and convergence statistics (described in Section 3.5.1) are applied to determine whether it can be reasonably believed that all of the chains reached the same distribution (Gelman & Shirley, 2011). However, the samples generated by the Markov Chain are not independent of one another (that is, not iid), but rather are temporally auto-correlated within-chain, and so the error in the Monte Carlo estimate of the mean for each parameter (as described in Section 3.4.4.1) needs to be adjusted by first calculating the *effective* sample size for that parameter, which accounts for this auto-correlation (Geyer, 2011, pp. 8–9). Holding the sample size constant, the greater the auto-correlation in the samples is, the lower the effective sample size is, and the higher the Monte Carlo error is. Thus, when implementing MCMC algorithmically, the goal is to minimize the computational time required to get from one sample to an effectively independent sample. To that end, in the studies carried out in this dissertation, Hamiltonian Monte Carlo (HMC) with the No-U-Turns Sampler (NUTS) as implemented in Stan is used (Carpenter et al., 2017; Hoffman & Gelman, 2014; Neal, 2011).

HMC differs from other approaches to MCMC in that it makes use of Hamiltonian dynamics to simulate the movement of the model about the posterior distribution (Neal, 2011). Hamilton's equations are usually used to describe a particle's movement in space through time given its position and momentum (e.g. an air particle in three-dimensional space has a three-dimensional position vector and a three-dimensional momentum vector that change across time together). In Hamiltonian Monte Carlo, the current state of the chain is treated as a position vector, and a momentum vector of the same size is randomly generated at each iteration (e.g. if we have 100 total parameters combining the fixed and random effects, then we are treating the model as a fictional 100-dimensional particle). Then, given the iteration's position and randomly generated momentum, the evolution of the fictional particle is simulated in discrete time steps to reach a new state. NUTS is a type of HMC that is especially efficient and does not require manual tuning of algorithmic parameters; for a detailed description of the algorithm, see Hoffman and Gelman (2014). The advantage of HMC with NUTS is that it tends to produce less correlated posterior samples; that is, it is more efficient in achieving effectively independent samples of the posterior distribution of the model parameters. The NUTS algorithm is implemented in Stan (Carpenter et al., 2017), a Bayesian modeling language that interfaces with R through the `rstan` package (Stan Development Team, 2016a). Stan code is provided in Sections A.3 and B.6, and is written following the guidelines the Stan language reference manual to allow for optimal performance (Stan Development Team, 2016b, pp. 316–341).[23]

3.5. Regression output and interpretation

Throughout this section, I will use the regression output for stress in the plosive regression as an example (discussed and interpreted in Chapter 5). Recall that, as we are discussing consonants, stress refers to the syllabic environment (tonic, post-tonic, or unstressed).

[23] There are several R packages which allow Bayesian mixed effects regressions to be fit using an `lme4`-like formula and implement weakly informative priors automatically, including `rstanarm` (Gabry & Goodrich, 2016) and `brms` (Bürkner, in press). However, these do not allow for `NA` values to be implemented directly, and so I have written my own Stan code.

3.5.1. Assessing model convergence

As with any statistical model, the results of a Bayesian mixed effects regression in Stan need to be checked for reasonableness; that is, the convergence of the model needs to be assessed. For each regression in this dissertation, I ran four chains, each with 1000 warm-up iterations and 2500 post-warm-up iterations (resulting in 10,000 posterior samples), and with the target acceptance rate at the default value of 0.8 (the only algorithmic parameter that may require some tuning, but usually does not, and did not in the regressions carried out here).[24] I implemented the following three convergence checks. First, I ensured that that there were no divergent transitions post-warm-up, which would indicate that the target acceptance rate algorithmic parameter needs to be increased (Stan Development Team, 2016b). Second, I ensured that the Gelman-Rubin $\hat{R}$ statistic, which measures the potential scale reduction that could be achieved if the chains were allowed to run for more iterations, was under 1.1 for all parameters (Gelman & Rubin, 1992). Third, I examined trace plots to ensure good mixing for the chains and posterior density plots to ensure there was no multimodality (Cowles, 2013, pp. 111–144). The trace plot and posterior density plot for the regression coefficient "Stress, Post-Tonic" (i.e. the difference between post-tonic plosives and the corrected mean in terms of VNPC1) are provided in Figure 3.13.

[24] The target acceptance rate, δ, ranging 0 to 1, exclusive, controls the size of the steps taken in the evolution of the Hamiltonian system. The higher δ is, the smaller the steps are, and the longer sampling takes. If δ is too small, divergent transitions can occur, which invalidate the results.

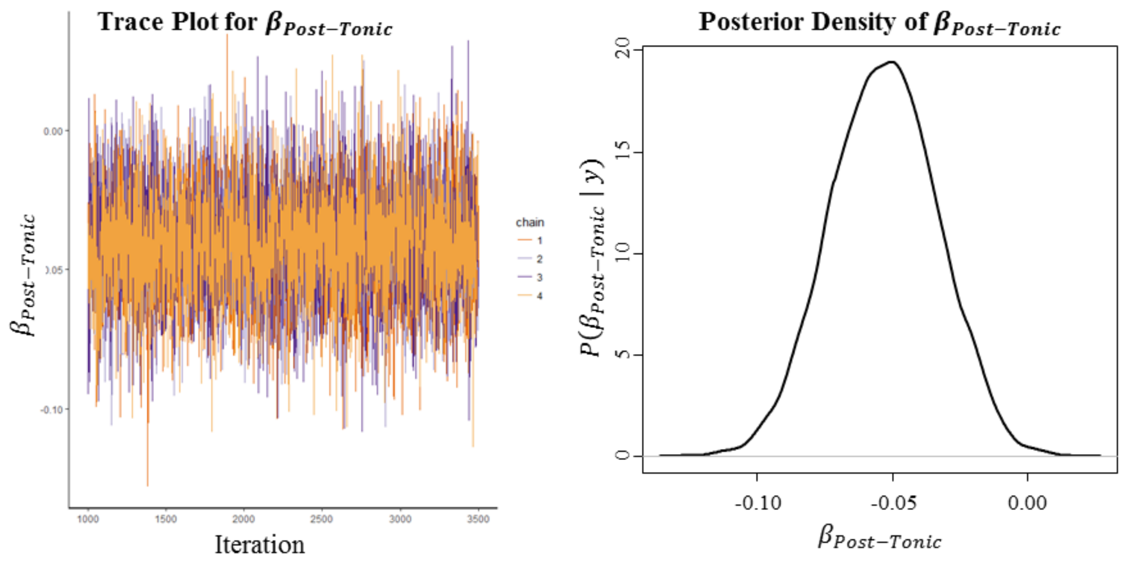

Figure 3.13 Trace plot (left panel) and density plot (right panel) of the posterior samples for $\beta_{Post-Tonic}$.

The trace plot in Figure 3.13 (left panel) shows the path that each of the four Markov Chains took for the "Stress, Post-Tonic" fixed effects parameter, superimposed on one another, excluding the warm-up iterations. What we are looking for in this plot is that the chains are randomly moving around the same area of the possible parameter values; and they are (i.e. we don't see any evidence that different colors occupy different areas of the plot, and the plot overall looks like a fuzzy flat horizontal bar). The posterior density plot in Figure 3.13 (right panel) shows the density of all of the values visited by all four chains in the trace plot, and represents our Monte Carlo estimate of the distribution $P(\beta_{Post-Tonic} \mid y)$. This distribution does not have a name (i.e. it isn't normally- or t-distributed, it is $P(\beta_{Post-Tonic} \mid y)$-distributed). This density plot represents a very rich source of information that we have about the fixed effects parameter, and we can summarize this information using the distributional characteristics of the samples themselves to obtain Monte Carlo estimates of the true distributional characteristics of $P(\beta_{Post-Tonic} \mid y)$, as described in the following section.

3.5.2. Fixed effects posterior distribution summaries

3.5.2.1. Fixed effect parameters

The regression tables presented in this dissertation are similar in format to regression tables used with frequentist methods, but with some important differences in interpretation (explanation of the output draws from Cowles (2013)). Continuing with the plosive stress example, I use a subset of Table 5.1 from the results in Chapter 5 to illustrate how the tables are interpreted. Table 3.13 shows the regression output for stress only, along with the intercept.

Table 3.13 Example fixed effect regression output.

Fixed Effect	Mean	SD	2.5%	97.5%	Neff	P(sign)	
Intercept (Corrected Mean)	-0.196	0.051	-0.297	-0.095	3575	> .999	***
Stress, Post-Tonic	-0.041	0.019	-0.079	-0.002	10000	.982	***
Stress, Tonic	0.219	0.020	0.180	0.257	10000	> .999	

Comparing the columns in Table 3.13 to those from a frequentist regression (in R), the posterior *Mean* is most similar to the *estimate*, the posterior standard deviation *SD* is most similar to the *standard error*, the posterior 95% equal-tailed credible interval (*2.5%* and *97.5%* columns) is most similar to a 95% confidence interval, and the posterior probability of the sign of the mean, *P(sign)*, is most similar to a *p-value*. The *Neff* column is the effective sample size of the parameter's posterior distribution (with a max of 10,000 in this case), as described in Section 3.4.4.3. The $\hat{R}$ statistic is omitted, as when rounded to two decimals, it was exactly 1.00 for all parameters in all regressions, and the standard error of the mean (as described in Section 3.4.4.1) is omitted, as it is recoverable from $SD/\sqrt{Neff}$. As for differences in interpretation, all of the output in the Bayesian regression table is based on the sample characteristics of the 10,000 posterior samples from the Markov Chains. In Figure 3.14, I show the same density plot from

Figure 3.13 (right panel), but now with the summary statistics visualized. I then discuss each summary statistic in detail.

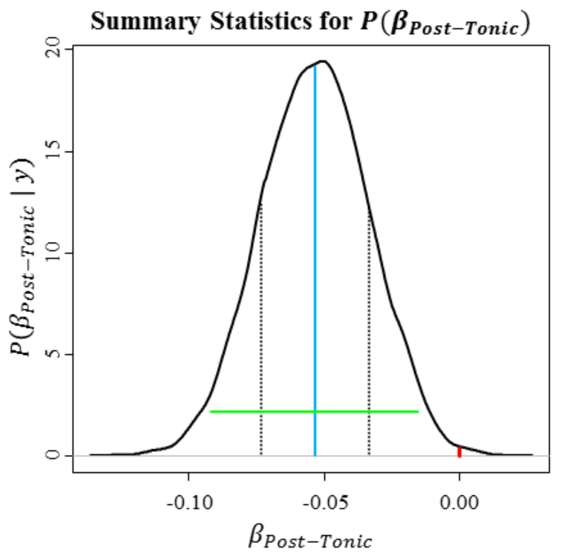

Figure 3.14 Summary statistics of the posterior distribution of $\beta_{Post-Tonic}$: the blue vertical line is the posterior mean; the vertical dotted black lines are one posterior SD on either side of the mean; the green bar is the 95% credible interval; and the area under the curve to the left of the red bar at zero is P(sign).

The posterior mean is not a (restricted) maximum likelihood estimate of the effect size, but rather the expected value of the parameter given the data, obtained by taking the sample mean of the posterior samples for each parameter. There are other options besides the posterior mean that can be used as the point estimate of the parameters, such as the posterior median and posterior mode. However, the posterior mean is the most widely used and most directly comparable to the frequentist estimate in interpretation because it minimizes a squared error loss function with respect to the posterior estimates for the parameter rather than an absolute difference loss function (posterior median) or curvature-based loss function (posterior mode) (Berger, 1993; Lehmann & Casella, 1998).

The posterior standard deviation is the standard deviation of the 10,000 posterior samples, not an ancillary estimate generated based on the (restricted) maximum likelihood estimate of the random effects covariance matrix. In frequentist statistics, a 95% confidence interval is often misinterpreted along the lines of meaning "there is a 95% probability that the true value of the parameter is contained in the interval." This, however, is not a correct interpretation of a confidence interval. The correct interpretation is "if the experiment were repeated a large number of times and a 95% confidence interval were generated each time, 95% of the time the confidence interval would contain the true parameter." In Bayesian statistics, the 95% credible interval has the interpretation often mis-assigned to frequentist confidence intervals: given the data, there is a 95% probability that the parameter falls within the interval.

Similarly, a p-value does *not* represent the probability that the null hypothesis is correct, but rather the probability that an effect as extreme or more extreme than that found in the data by (restricted) maximum likelihood estimation would be found assuming that the true effect is zero, as discussed in Section 3.2.1. For a more in-depth discussion of what a frequentist p-value means and some common misinterpretations, see Cohen (1994). The probability of the sign of the mean, P(sign), is the proportion of the posterior samples for an estimate that have the same sign as the posterior mean. That is, if the posterior mean is positive, then P(sign) is the probability that the effect is positive given the data, and if the posterior mean is negative, then P(sign) is the probability that the effect is negative given the data. The higher the value of P(sign), the greater certainty there is for a non-zero effect with the same sign as the posterior mean. While technically the range of P(sign) is 0 to 1, in practice the lower bound is 0.5, with values lower than 0.5 only arising in the case of highly skewed or multimodal posterior distributions (which would be detected by the convergence checks described in Section 3.5.1).

As this probability value describes what researchers often want to know but are unable to obtain with frequentist statistics (i.e. the probability of a hypothesis given the data), it is not necessary to take a NHST approach to the data, and in this dissertation I do not take such an approach.

With a NHST approach, an alpha level is chosen (usually $\alpha = .05$), and then two-tailed tests are performed to obtain the probability of the data given the null hypothesis, and an accept or reject decision is made with regards to the null hypothesis. If the null hypothesis is true and is not rejected, this is referred to as a Type I error. If the null hypothesis is false but accepted, this is referred to as a Type II error. As I do not take a NHST approach in this dissertation, instead of Type I and Type II errors, we refer to Type S (sign) errors and Type M (magnitude) errors (e.g. Gelman & Tuerlinckx, 2000). A Type S error occurs when we say that we are confident that a parameter is positive, but is in fact negative, or vice versa, and a Type M error occurs when we say that we are confident that an effect size is large when it is in fact small, or vice versa. The P(sign) estimate is thus the posterior probability that we are not committing a Type S error. In Table 3.13, a Bayesian NHST with $\alpha = .05$ would be declaring effects with P(sign) ≥ .975 "significant" and effects with P(sign) < .975 "not significant". In this dissertation, I will refer to this degree of certainty with regards to the effect direction as "strong evidence" for the effect direction (denoted '***' in the P(sign) column). But, rather than rejecting as insignificant all other effects, I will discuss them as there being "some evidence" for the effect direction when .950 ≤ P(sign) < .975 (denoted '**' in the P(sign) column), "weak evidence" for the effect direction when .900 ≤ P(sign) < .950 (denoted '*' in the P(sign) column), and "little evidence" for the effect direction when P(sign) < .900 (denoted 'x' in the P(sign) column). In other words, while binning the degree of confidence is to some extent necessary for purposes of discussion, the actual posterior probability should always be considered in a gradient manner.

3.5.2.2. Group means and pairwise comparisons

In frequentist statistics, group means for factor levels and pairwise comparisons of the groups are performed using least-squares means (e.g. via the `lsmeans` package (Lenth & Hervao, 2015)), with the p-values for the pairwise comparisons adjusted to account for multiple comparisons (most often with the Tukey method). These comparisons are only made if an omnibus test first shows that the factor accounted for a significant amount of variance overall, which is done by comparing nested models with ANOVA. With Bayesian inference, there is no issue of multiple comparisons in this classical sense (Gelman, Hill, & Yajima, 2012; Gelman & Tuerlinckx, 2000), as P(sign) consistently represents the probability of a Type S error across all comparisons made. That is, if, for example, a total of 100 claims of strong evidence for the signs of effects are made throughout the dissertation, then we expect on average 2.5 false claims of effect sign (Type S errors), regardless of whether or not the claims were for covariates, binary factors, or pairwise comparisons for factors with three or more levels. Continuing with the Spanish stress example, the contrasts applied to obtain the group means for the different stress categories are given Table 3.14.

Table 3.14 Example contrasts for posterior group means.

Stress	Contrasts applied to coefficients
Tonic	Intercept (Corrected Mean) + Stress, Tonic
Post-Tonic	Intercept (Corrected Mean) + Stress, Post-Tonic
Unstressed	Intercept (Corrected Mean) - Stress, Tonic - Stress, Post-Tonic

These contrasts are applied to each posterior sample of the model parameters individually, yielding an estimate of the group means for each iteration (in this case, 10,000 posterior estimates). These estimates can then be subtracted from one another to obtain an estimate for each group difference for each iteration, and inference can be made on the group means and pairwise comparisons just as described for the fixed effect parameters in Section

3.5.2.1. For group means, I report the posterior mean, standard deviation, and 95% credible interval, and for pairwise comparisons, I additionally provide the posterior probability of the sign of the contrast mean. For our stress example, these are given in Table 3.15 (also Table 5.4) and Table 3.16 (also Table 5.5), respectively.

Table 3.15 Example output for fixed effect factor group means.

Stress	Mean	SD	2.5%	97.5%
Tonic	0.023	0.052	-0.082	0.126
Post-Tonic	-0.237	0.055	-0.345	-0.127
Unstressed	-0.374	0.056	-0.485	-0.264

Table 3.16 Example output for fixed effect factor pairwise comparisons.

Contrast	Mean	SD	2.5%	97.5%	P(sign)	
Tonic - Post-Tonic	0.260	0.033	0.195	0.324	> .999	***
Tonic - Unstressed	0.397	0.036	0.328	0.467	> .999	***
Post-Tonic - Unstressed	0.137	0.036	0.067	0.207	> .999	***

3.5.3. Random effects

3.5.3.1. Standard deviations

To summarize the posterior distribution of the random effects standard deviations, I use the same format as described for the fixed effects in Section 3.5.2.1, except for two alterations. First, because standard deviations cannot be negative, P(sign) is meaningless and therefore omitted. Second, because the tail of the posterior distributions of variance components tends to be very heavy (which can raise the value of the posterior mean, making it perhaps not the best posterior point estimate), I provide the posterior median in addition to the posterior mean. For our stress example, the standard deviation for the random speaker slopes for stress are given in Table 3.17 (taken from Table B.2).

Table 3.17 Example regression output for random effect standard deviations.

Random Speaker Effect	Mean	SD	Median	2.5%	97.5%	Neff
Intercept (Corrected Mean)	0.304	0.035	0.301	0.244	0.381	4484
Stress, Post-Tonic	0.031	0.021	0.029	0.001	0.076	4298
Stress, Tonic	0.075	0.018	0.075	0.038	0.111	4302

3.5.3.2. Correlations

The correlation in the random effects can be summarized in the same way as described for the fixed effects in Section 3.5.2.1. However, as discussed in Section 3.4.3, the correlations in Ω_γ are not directly useful. I am interested in the correlation between certain quantities generated for each subject, such as the correlation between the predicted means of the strength of their /k/ and /g/ productions. This information is not directly expressed in the correlation parameters in Ω_γ. I therefore generate estimates for such quantities using both the fixed and random effects at each iteration of the model, and then generate an estimate for the correlation between the quantities at each iteration, in a similar fashion as described for group means and pairwise comparisons in Section 3.5.2.2.

3.6. Conclusion

Bayesian mixed effects regressions with maximal random effects structures and weakly informative priors allow hypotheses concerning individual differences to be tested while still making use of all of the information available in the data in a single model. The estimation of these models is done using Markov Chain Monte Carlo (and in this dissertation specifically, Hamiltonian Monte Carlo with the No-U-Turns sampler in Stan). When continuous variables are scaled, factors are coded with sum contrasts, and factor `NA` values are coded as zero, reasonable weakly informative priors can be applied. These weakly informative priors do not bias the results in favor of the hypotheses, but rather place reasonable constraints on the model parameters, which can be summarized as follows: (1) the most likely effect is no effect; (2) for fixed effects and random effect correlations, positive and negative values are equally likely (for standard deviations this is not applicable); (3) effects that are larger in magnitude are gradiently less likely than effects that are smaller in magnitude; and (4) effects with a magnitude greater than 5 are very unlikely. Importantly, the data can override the priors when there is strong enough evidence in the data, with the posterior approaching the likelihood as the sample size increases. Taking a Bayesian approach to mixed effects regression also allows us to move past a null hypothesis significance testing approach, since the posterior distribution of the model gives us what we are actually looking for: the probability of our hypothesis given the data.

Chapter 4. Spanish and Catalan fricative results

In this chapter, I present the results of the Bayesian mixed effects regression on Valladolid Spanish and Catalan /sf/ NPC1 (normalized PC1, which measures the strength of a fricative in terms of duration, voiceless duration, and percent voiceless on unit scale). The fixed effects included in the regression were language, fricative phoneme identity, stress, sex, preceding vowel height, following vowel height, log word frequency, speech rate, and the interaction of language with each of fricative phoneme identity, stress, sex, preceding vowel height, and following vowel height (that is, the interaction of language with all of the factors was included based on examining descriptive statistics). Speaker was included as a random effect grouping factor, with the maximal random effects structure, which included random intercepts and random slopes for fricative phoneme identity, stress, preceding vowel height, following vowel height, log word frequency, and speech rate. The posterior distribution of the fixed effects parameters are given in Table 4.1 (the random speaker effect variances and residual standard error are provided in Section A.1). In Section 4.1, the posterior distribution of the control predictors are presented, and in Section 4.2, the posterior distribution of the interaction of language and fricative phoneme identity are presented. For each predictor, descriptive statistics and plots are provided along with effect estimates and, when relevant, pairwise comparisons.

Table 4.1 Posterior distribution of the fixed effects.

Fixed Effect	Mean	SD	2.5%	97.5%	Neff	P(sign)	
Intercept (Corrected Mean)	-0.033	0.067	-0.165	0.099	2770	.692	x
Language, Catalan	0.075	0.067	-0.057	0.204	2834	.872	x
Fricative, /f/	0.015	0.043	-0.070	0.098	5359	.637	x
Stress, Post-Tonic	-0.035	0.037	-0.109	0.038	7315	.830	***
Stress, Tonic	0.151	0.036	0.078	0.222	10000	> .999	
Sex, Female	0.098	0.065	-0.029	0.223	2894	.936	*
Preceding Vowel, High	-0.084	0.026	-0.135	-0.032	10000	.999	***
Following Vowel, High	0.100	0.028	0.045	0.155	8250	> .999	***
Log Word Frequency	-0.054	0.027	-0.110	-0.003	8361	.980	***
Speech Rate	-0.246	0.021	-0.287	-0.205	10000	> .999	***
Language, Catalan : Fricative, /f/	-0.093	0.043	-0.179	-0.008	4727	.986	***
Language, Catalan : Stress, Post-Tonic	-0.138	0.038	-0.213	-0.063	6940	> .999	***
Language, Catalan : Stress, Tonic	0.039	0.036	-0.032	0.108	10000	.867	
Language, Catalan : Sex, Female	0.012	0.065	-0.117	0.141	2816	.579	x
Language, Catalan : Preceding Vowel, High	-0.034	0.026	-0.084	0.018	10000	.905	*
Language, Catalan : Following Vowel, High	0.028	0.029	-0.030	0.084	6713	.826	x

4.1. Control predictors

4.1.1. Linguistic covariates

As log word frequency and speaker-normalized speech rate are covariates and are not involved in any interactions, the posterior estimate of their effects can be taken directly from Table 4.1.

4.1.1.1. Word frequency

A scatterplot of NPC1 by normalized log word frequency is provided in Figure 4.1, where the sample correlation is -0.064.

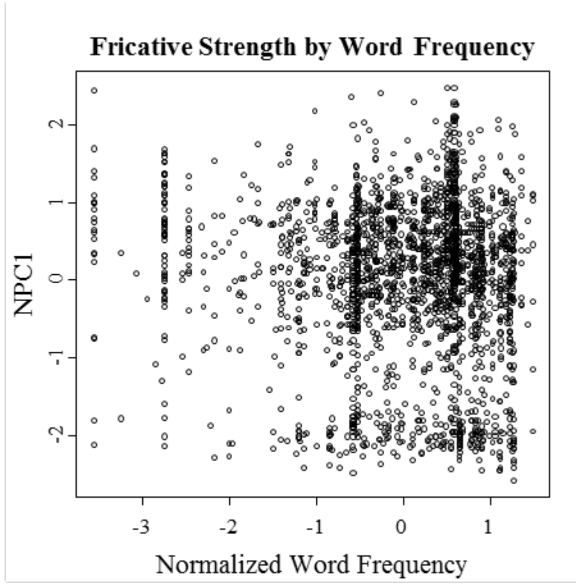

Figure 4.1 Scatterplot of NPC1 by normalized log word frequency ($r = -0.064$).

The posterior estimate is that a 1-SD increase in log word frequency leads to an estimated decrease in NPC1 of 0.054, and we have strong evidence for this effect being negative (P(sign) = .980). That is, the more frequent a word is, the weaker fricatives in the word are predicted to be, consistent with previous research on word frequency and lenition (e.g. Pierrehumbert, 2001).

4.1.1.2. Speech rate

A scatterplot of NPC1 by speaker-normalized speech rate is provided in Figure 4.2, where the sample correlation is -0.258.

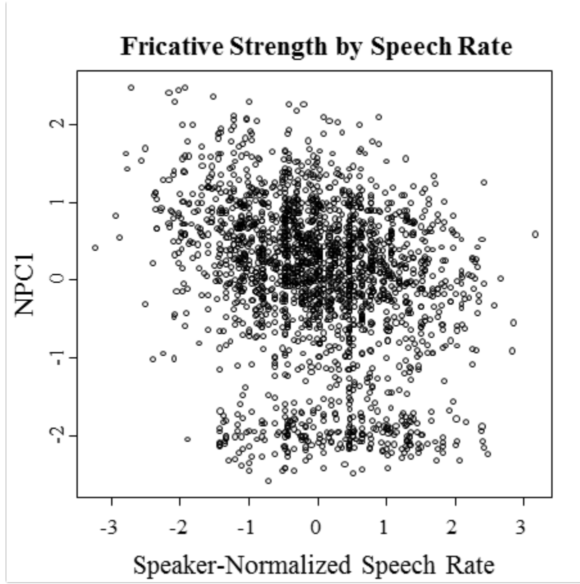

Figure 4.2 Scatterplot of NPC1 by speaker-normalized speech rate ($r = -0.258$).

The posterior estimate is that a 1-SD increase in speaker-normalized speech rate leads to an estimated decrease in NPC1 of 0.246, and we have strong evidence for this effect being negative (P(sign) > .999). This is consistent with previous research on Spanish /s/ lenition where speech rate was included as a covariate (Torreira & Ernestus, 2012), and also exactly what we would expect the general effect of increased speech rate to be (i.e. increased articulatory undershoot leading to acoustically weaker consonants).

4.1.2. Linguistic factors

4.1.2.1. Stress

Descriptive statistics for NPC1 by language and stress are provided in Table 4.2, with corresponding boxplots in Figure 4.3.

Table 4.2 Descriptive statistics for NPC1 by language and stress.

Language	Stress	N	Minimum	Median	Maximum	Mean	SD
Catalan	Tonic	283	-2.263	0.223	2.253	0.112	0.929
	Post-Tonic	351	-2.444	0.087	2.351	-0.039	0.980
	Unstressed	599	-2.417	0.268	2.478	0.159	1.013
Spanish	Tonic	214	-2.453	0.150	2.444	-0.008	0.972
	Post-Tonic	399	-2.589	0.210	2.405	-0.061	1.070
	Unstressed	317	-2.480	-0.050	1.994	-0.274	0.919

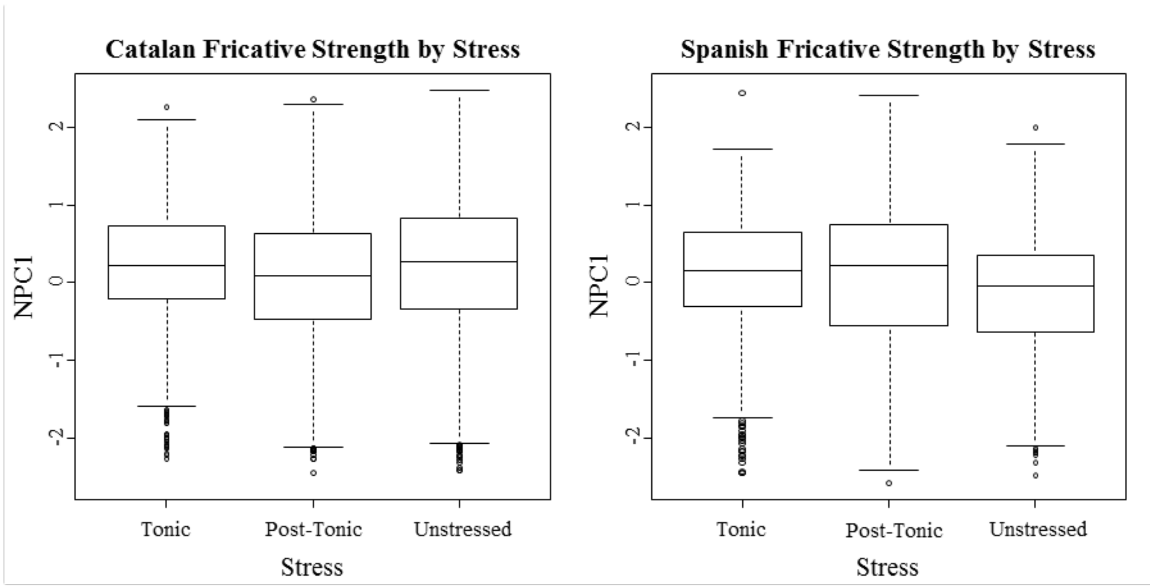

Figure 4.3 Boxplots of NPC1 by stress for Catalan (left panel) and Spanish (right panel) fricatives.

There is strong evidence for an interaction between stress and language (maximum P(sign) > .999 in Table 4.1). Posterior group means for NPC1 by language and stress are given in Table 4.3, with corresponding pairwise comparisons within language in Table 4.4.

Table 4.3 Posterior group means for NPC1 by language and stress.

Language	Stress	Mean	SD	2.5%	97.5%
Catalan	Tonic	0.232	0.106	0.023	0.438
	Post-Tonic	-0.131	0.101	-0.327	0.066
	Unstressed	0.025	0.108	-0.189	0.235
Spanish	Tonic	0.004	0.115	-0.224	0.227
	Post-Tonic	-0.004	0.107	-0.217	0.202
	Unstressed	-0.323	0.113	-0.549	-0.103

Table 4.4 Posterior group differences for NPC1 by stress within language.

Language	Contrast	Mean	SD	2.5%	97.5%	P(sign)	
Catalan	Tonic - Post-Tonic	0.363	0.090	0.185	0.539	> .999	***
	Tonic - Unstressed	0.207	0.086	0.037	0.375	.992	***
	Post-Tonic - Unstressed	-0.156	0.086	-0.325	0.017	.962	**
Spanish	Tonic - Post-Tonic	0.008	0.088	-0.162	0.181	.538	x
	Tonic - Unstressed	0.327	0.094	0.140	0.511	> .999	***
	Post-Tonic - Unstressed	0.318	0.100	0.120	0.513	.998	***

For Catalan fricatives, the posterior estimates show a hierarchy of Tonic > Unstressed > Post-Tonic, with strong evidence for fricatives in the onset of a tonic syllable being stronger than fricatives in the onset of the post-tonic syllable or between two unstressed vowels, and some evidence for the difference between the unstressed and post-tonic conditions. In Spanish, the posterior estimates show a different hierarchy of Tonic, Post-Tonic > Unstressed, with strong evidence for fricatives in the unstressed condition being weaker than the other two conditions, and little evidence for a difference between the tonic and post-tonic conditions. This is different from the findings of Torreira and Ernestus (2012), who found no significant effect for stress in Spanish for intervocalic /s/ voicing; this may be perhaps due to post-tonic fricatives patterning with tonic fricatives here, while post-tonic fricatives were classified with unstressed fricatives by Torreira and Ernestus (2012).

4.1.2.2. Preceding vowel height

Descriptive statistics for NPC1 by language and preceding vowel height are provided in Table 4.5, with corresponding boxplots in Figure 4.4.

Table 4.5 Descriptive statistics for NPC1 by language and preceding vowel height.

Language	Preceding Vowel	N	Minimum	Median	Maximum	Mean	SD
Catalan	Non-High	939	-2.444	0.267	2.478	0.143	0.993
	High	294	-2.417	0.110	1.974	-0.072	0.954
Spanish	Non-High	722	-2.589	0.096	2.041	-0.110	0.980
	High	208	-2.363	0.032	2.444	-0.161	1.083

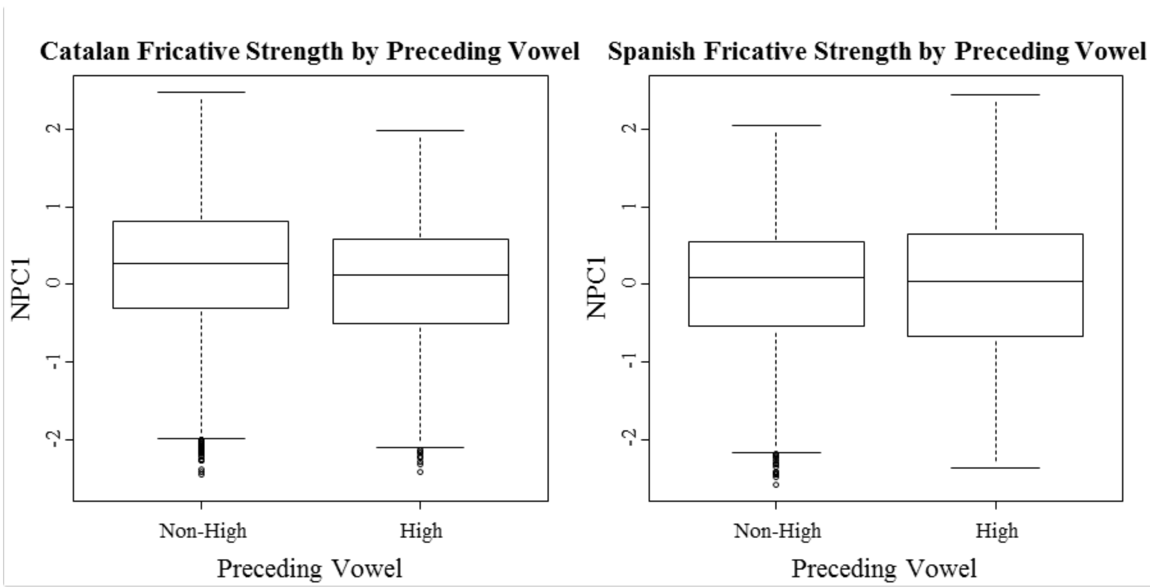

Figure 4.4 Boxplots of NPC1 by preceding vowel height for Catalan (left panel) and Spanish (right panel) fricatives.

While the descriptive statistics show a possible interaction between preceding vowel height and language, there is not strong evidence for this interaction (P(sign) = .905 in Table 4.1). For this reason, in Table 4.6, the group means for NPC1 are given by preceding vowel height overall (averaging over the interaction term), with corresponding group difference estimate in Table 4.7.

Table 4.6 Posterior group means for NPC1 by preceding vowel height.

Preceding Vowel	Mean	SD	2.5%	97.5%
High	-0.117	0.075	-0.264	0.032
Non-High	0.051	0.068	-0.084	0.183

Table 4.7 Posterior group differences for NPC1 by preceding vowel height.

Contrast	Mean	SD	2.5%	97.5%	P(sign)	
High - Non-High	-0.167	0.052	-0.270	-0.064	.999	***

The posterior estimates show a moderate negative effect, with fricatives after a high vowel being 0.167 standard deviations weaker than those following a non-high values, and the evidence for the direction of this effect is very strong (P(sign) = .999). This is different from Torreira and Ernestus' (2012) study on Spanish /s/ lenition, where the effect of preceding vowel identity was not found to have a significant effect for any measure taken. The effect of preceding vowel height on overall consonant strength is somewhat difficult to predict. From a duration perspective, a preceding high vowel should lead to a shorter consonant duration as measured in this study (preceding vowel intensity maximum to following vowel maximum); however, when an underlyingly voiceless obstruent voices bordering sonorants, the voicing tends to bleed in from the preceding vowel rather than the following vowel (L. Davidson, 2016), and this effect should be lesser for high vowels. For NPC1, which incorporates both of these measures, the model is telling us that there is strong evidence that the overall effect of a preceding high vowel (vs. non-high vowel) is negative.

4.1.2.3. Following vowel height

Descriptive statistics for NPC1 by language and following vowel height are provided in Table 4.8, with corresponding boxplots in Figure 4.5.

Table 4.8 Descriptive statistics for NPC1 by language and following vowel height.

Language	Following Vowel	N	Minimum	Median	Maximum	Mean	SD
Catalan	Non-High	475	-2.272	0.023	2.351	-0.107	0.962
	High	758	-2.444	0.345	2.478	0.216	0.984
Spanish	Non-High	605	-2.589	0.107	2.444	-0.113	1.035
	High	325	-2.480	0.026	1.994	-0.137	0.943

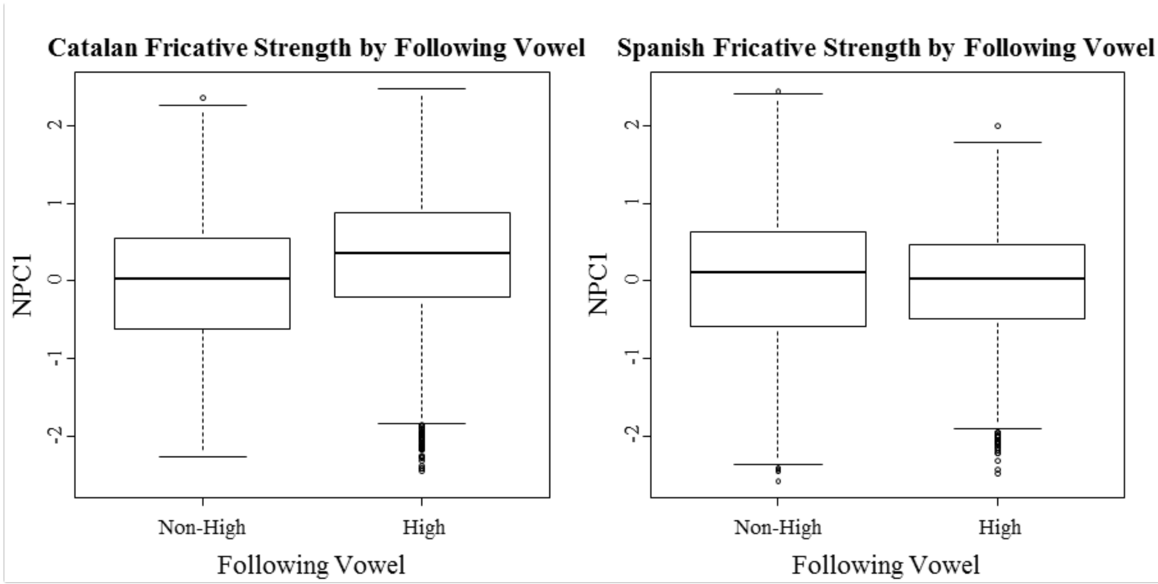

Figure 4.5 Boxplots of NPC1 by following vowel height for Catalan (left panel) and Spanish (right panel) fricatives.

While the descriptive statistics show a possible interaction between following vowel height and language, there is little evidence for this interaction (P(sign) = .826 in Table 4.1). For this reason, in Table 4.9, the group means for NPC1 are given by following vowel height overall (averaging over the interaction term), with corresponding group difference estimate in Table 4.10

Table 4.9 Posterior group means for NPC1 by following vowel height.

Following Vowel	Mean	SD	2.5%	97.5%
High	0.067	0.074	-0.080	0.215
Non-High	-0.133	0.070	-0.271	0.006

Table 4.10 Posterior group differences for NPC1 by following vowel height.

Contrast	Mean	SD	2.5%	97.5%	P(sign)	
High - Non-High	0.200	0.056	0.090	0.310	> .999	***

The posterior estimates show a moderate positive effect, with fricatives after a high vowel being 0.200 standard deviations stronger than those following a non-high values, and the evidence for the direction of this effect is very strong (P(sign) > .999). This is different from Torreira and Ernestus' (2012) study on Spanish /s/ lenition, where the effect of following vowel identity was not found to have a significant effect for any measure taken. However, the result found here is consistent with a number of studies that find that high vowels are more likely to devoice after voiceless consonants than non-high vowels (e.g. Delforge, 2009, 2012; Torreira & Ernestus, 2011), which would correspond to a longer voiceless duration and higher percentage of the overall consonant duration being voiceless, leading to a higher NPC1.

4.1.3. Social factor (speaker sex)

Descriptive statistics for NPC1 by language and sex are provided in Table 4.11, with corresponding boxplots in Figure 4.6.

Table 4.11 Descriptive statistics for NPC1 by language and sex.

Language	Sex	N	Minimum	Median	Maximum	Mean	SD
Catalan	Female	602	-2.444	0.288	2.469	0.175	0.990
	Male	631	-2.417	0.176	2.478	0.012	0.980
Spanish	Female	468	-2.453	0.129	1.994	-0.080	0.970
	Male	462	-2.589	-0.009	2.444	-0.164	1.036

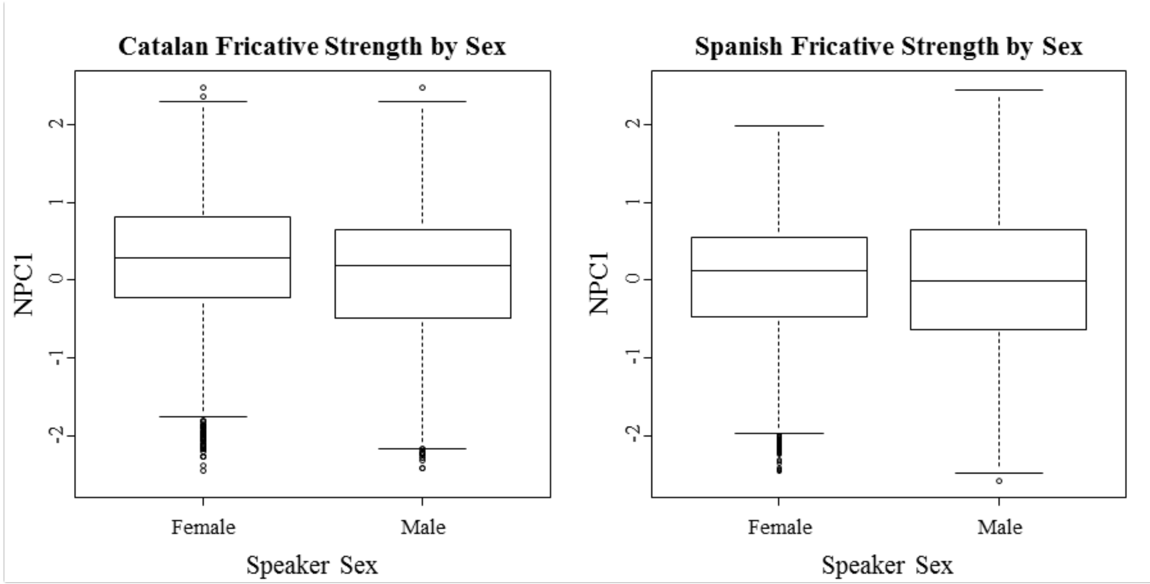

Figure 4.6 Boxplots of NPC1 by sex for Catalan (left panel) and Spanish (right panel) fricatives.

While the descriptive statistics show a possible interaction between speaker sex and language, there is little evidence for this interaction (P(sign) = .579 in Table 4.1). For this reason, in Table 4.12, the group means for NPC1 are given by sex overall (averaging over the interaction term), with corresponding group difference estimate in Table 4.13.

Table 4.12 Posterior group means for NPC1 by sex.

Sex	Mean	SD	2.5%	97.5%
Female	0.066	0.093	-0.120	0.248
Male	-0.131	0.093	-0.314	0.053

Table 4.13 Posterior group differences for NPC1 by sex.

Contrast	Mean	SD	2.5%	97.5%	P(sign)	
Female - Male	0.197	0.131	-0.058	0.447	.936	*

The posterior estimates show a moderate positive effect, with female speakers producing fricatives that are 0.197 standard deviations stronger than fricatives produced by male speakers. However, we have only weak evidence for this effect (P(sign) = .936).

4.2. Language and fricative phoneme identity

Descriptive statistics for NPC1 by language and fricative phoneme identity are provided in Table 4.14, with corresponding boxplots in Figure 4.7.

Table 4.14 Descriptive statistics for NPC1 by language and phoneme identity.

Fricative	Language	N	Minimum	Median	Maximum	Mean	SD
/f/	Catalan	269	-2.417	0.091	2.253	-0.056	0.943
	Spanish	165	-2.209	0.143	2.444	-0.068	0.985
/s/	Catalan	964	-2.444	0.283	2.478	0.133	0.997
	Spanish	765	-2.589	0.070	2.405	-0.133	1.008

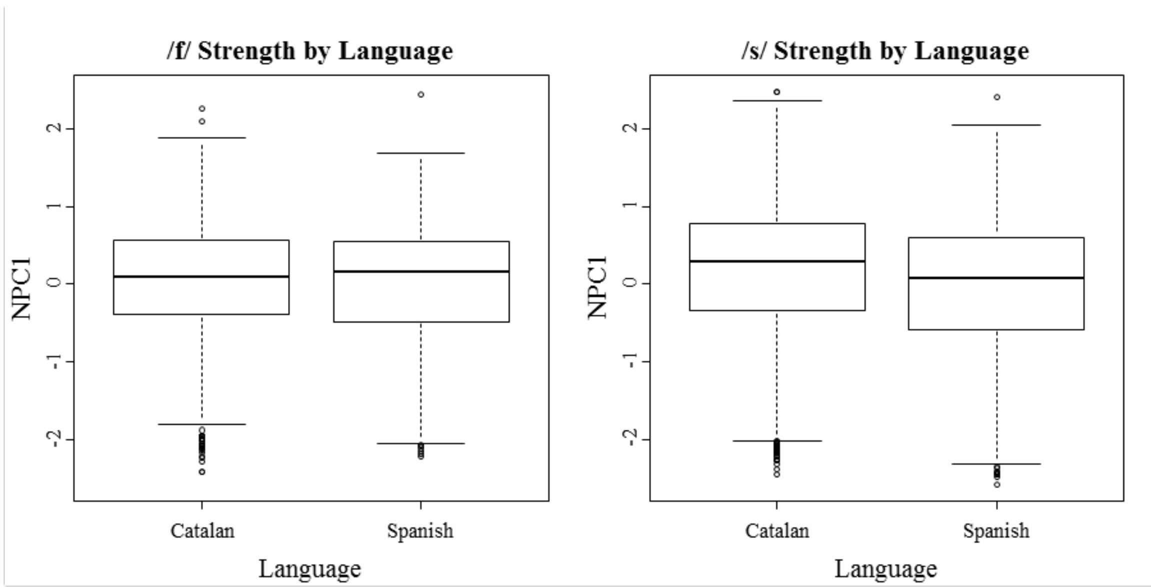

Figure 4.7 Boxplots of NPC1 by language for /f/ (left panel) and /s/ (right panel).

There is strong evidence for an interaction between phoneme identity and language (P(sign) = .986 in Table 4.1). Posterior group means for NPC1 by language and phoneme identity are given in Table 4.15, with corresponding pairwise comparisons within phoneme identity in Table 4.16.

Table 4.15 Posterior group means for NPC1 by language and phoneme identity.

Fricative	Language	Mean	SD	2.5%	97.5%
/f/	Catalan	-0.036	0.104	-0.243	0.169
	Spanish	0.000	0.118	-0.233	0.235
/s/	Catalan	0.120	0.112	-0.103	0.342
	Spanish	-0.215	0.114	-0.439	0.010

Table 4.16 Posterior group differences for NPC1 by language within phoneme identity.

Contrast	Fricative	Mean	SD	2.5%	97.5%	P(sign)	
Catalan - Spanish	/f/	-0.036	0.159	-0.351	0.270	.592	x
	/s/	0.335	0.160	0.022	0.653	.981	***

The posterior pairwise comparisons show little evidence for a difference between Catalan and Spanish /f/ (P(sign) = .592), but strong evidence that Catalan /s/ is stronger than Spanish /s/ (P(sign) = .981). The magnitude of this effect is also stronger than nearly any of the other factors included in the regression (the comparison for Catalan Tonic vs. Post-Tonic in Table 4.4 is slightly larger at 0.363). In addition to having strong evidence for Catalan /s/ being stronger than Spanish /s/, we also have strong evidence that the magnitude of the difference for /s/ is greater than the magnitude of the difference for /f/, as shown in Table 4.17.

Table 4.17 Posterior estimate for the difference in magnitude for Catalan vs. Spanish /s/ and /f/.

Contrast	Mean	SD	2.5%	97.5%	P(sign)	
(Catalan /s/ - Spanish /s/) - (Catalan /f/ - Spanish /f/)	0.371	0.174	0.032	0.716	0.986	***

These results strongly support Hypothesis 1 (from Section 1.6: *Catalan /s/ will be stronger than (Valladolid) Spanish /s/, but Catalan /f/ will not be stronger than (Valladolid) Spanish /f/.*) Hualde and Prieto (2014) found that Catalan /s/ was significantly longer than Spanish /s/, and argued that this may be indicative of an effect of contrast preservation, but noted that comparison of other voiceless segments in the languages was necessary to further support the claim. The results found here are consistent with their findings for /s/, and further show that there is little evidence for a difference in /f/ strengths between the languages, and strong evidence that the difference between the languages' /s/ strengths is larger than the difference between their /f/ strengths, consistent with their predictions for /f/. Given that Spanish and Catalan both have /s/ and /f/, neither has /v/, and Catalan has /z/ while Spanish does not, these results are consistent with a role for contrast preservation as a constraint in its own right. That is, if Catalan simply had stronger voiceless fricatives in Spanish, we would expect to also find evidence that Catalan has stronger /f/ realizations than Spanish, not just stronger /s/ realizations. One possible explanation for this asymmetry is that the presence of the contrastive consonant /z/ in Catalan constrains the degree to which Catalan /s/ lenites, while this constraint does not exist in Spanish.

Chapter 5. Spanish plosive results

In this chapter, I present the results of the Bayesian mixed effects regression on Spanish /ptk/ and /bdg/ VNPC1 (voicing-normalized PC1, which measures the strength of a plosive given its underlying voicing in terms of duration, voiceless duration, percent voiceless, intensity difference, and intensity velocity on unit scale). The fixed effects included in the regression were underlying voicing, place of articulation, stress, word position, preceding vowel height, following vowel height, log word frequency, speech rate, task, dialect, sex, age group , education level, Quechua bilingualism, the interaction of dialect and sex, and the interaction of underlying voicing with each of place of articulation, word position, preceding vowel height, following vowel height, speech rate, dialect, sex, and Quechua bilingualism (see Section B.4 for descriptive statistics for interactions with underlying voicing that were not included in the model). Speaker and item were included as random effect grouping factors, with maximal random effects structures. For speaker, this included random intercepts and random slopes for underlying voicing, place of articulation, stress, word position, preceding vowel height, following vowel height, log word frequency, speech rate, task, and the interaction of underlying voicing with each place of articulation, word position, preceding vowel height, following vowel height, and speech rate. For item, this included random intercepts and random slopes for speech rate, dialect, sex, age group, education level, Quechua bilingualism, and the interaction of dialect and sex. The posterior distribution of the fixed effects parameters are given in Table 5.1 (the random effect variances and residual standard error are provided in Section B.3). In Section 5.1, the posterior distributions of the control predictors are presented. In Section 5.2, the posterior distributions of the interaction of underlying voicing and place of articulation are presented. And in Section 5.3, the posterior distributions of the interaction of underlying voicing and dialect are presented. For each predictor, descriptive statistics and plots are provided along with effect estimates and, when relevant, pairwise comparisons.

Table 5.1 Posterior distribution of the fixed effects.

Fixed Effect	Mean	SD	2.5%	97.5%	Neff	P(sign)	
Intercept (Corrected Mean)	-0.196	0.051	-0.297	-0.095	3575	> .999	***
Voicing, Voiced	0.027	0.034	-0.039	0.093	5329	.784	x
Place, Biliabial	0.063	0.017	0.030	0.097	10000	> .999	***
Place, Dental	-0.064	0.022	-0.106	-0.020	10000	.997	
Stress, Post-Tonic	-0.041	0.019	-0.079	-0.002	10000	.982	***
Stress, Tonic	0.219	0.020	0.180	0.257	10000	> .999	
Word Position, Initial	0.085	0.020	0.048	0.125	10000	> .999	***
Preceding Vowel, High	0.050	0.016	0.019	0.082	10000	.999	***
Following Vowel, High	-0.032	0.015	-0.062	-0.002	10000	.981	***
Log Word Frequency	-0.060	0.013	-0.087	-0.033	10000	> .999	***
Speech Rate	-0.071	0.012	-0.095	-0.049	10000	> .999	***
Task, Read Speech	0.147	0.028	0.092	0.204	6899	> .999	***
Dialect, Cuzco	0.610	0.070	0.475	0.749	2610	> .999	***
Dialect, Lima	-0.450	0.077	-0.600	-0.296	4843	> .999	
Sex, Female	0.213	0.051	0.113	0.312	3109	> .999	***
Age Group, Older	0.011	0.055	-0.096	0.123	4323	.582	x
Education Level, Secondary	0.013	0.054	-0.096	0.118	4314	.601	x
Quechua Bilingual, Yes	0.023	0.057	-0.089	0.132	4524	.656	x
Voicing, Voiced : Place, Biliabial	-0.071	0.017	-0.105	-0.037	10000	> .999	***
Voicing, Voiced : Place, Dental	-0.038	0.022	-0.082	0.007	10000	.953	
Voicing, Voiced : Word Position, Initial	0.054	0.017	0.022	0.088	10000	.999	***
Voicing, Voiced : Preceding Vowel, High	0.013	0.014	-0.014	0.039	10000	.825	x
Voicing, Voiced : Following Vowel, High	0.063	0.015	0.035	0.092	10000	> .999	***
Voicing, Voiced : Speech Rate	0.128	0.012	0.104	0.151	10000	> .999	***
Voicing, Voiced : Dialect, Cuzco	0.055	0.045	-0.033	0.144	3466	.896	***
Voicing, Voiced : Dialect, Lima	-0.106	0.051	-0.207	-0.005	5256	.980	
Voicing, Voiced : Sex, Female	-0.052	0.028	-0.106	0.004	4572	.965	**
Voicing, Voiced : Quechua Bilingual, Yes	-0.025	0.037	-0.098	0.046	5087	.756	x
Dialect, Cuzco : Sex, Female	0.122	0.059	0.003	0.241	3710	.978	***
Dialect, Lima : Sex, Female	-0.127	0.078	-0.283	0.027	4218	.951	

5.1. Control predictors

5.1.1. Linguistic covariates

5.1.1.1. Word frequency

A scatterplot of VNPC1 by normalized log word frequency is provided in Figure 5.1, where the sample correlation is -0.075.

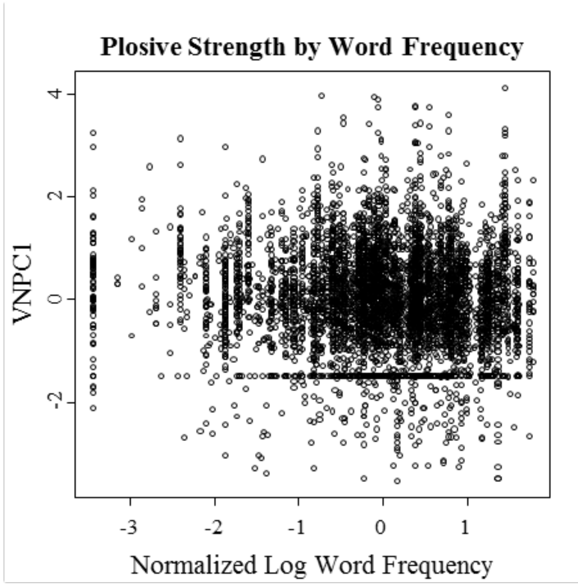

Figure 5.1 Scatterplot of VNPC1 by normalized log word frequency ($r = -0.075$).

As log word frequency is a covariate and is not involved in any interactions, the posterior estimate of its effect can be taken directly from Table 5.1. A 1-SD increase in log word frequency leads to an estimated decrease in VNPC1 of 0.060, and we have strong evidence for this effect being negative (P(sign) > .999). That is, the more frequent a word is, the weaker plosives in the word are predicted to be, consistent with previous research on word frequency and lenition (e.g. Pierrehumbert, 2001). This is also consistent with the results found for Spanish and Catalan fricatives in Section 4.1.1.1, with the effect sizes even being nearly the same (-0.054 vs. -0.060).

5.1.1.2. Speech rate

Scatterplots of VNPC1 by speaker-normalized speech rate are provided separately by underlying voicing in Figure 5.2, with a sample correlation of 0.112 for voiced plosives and a sample correlation of -0.268 for voiceless plosives.

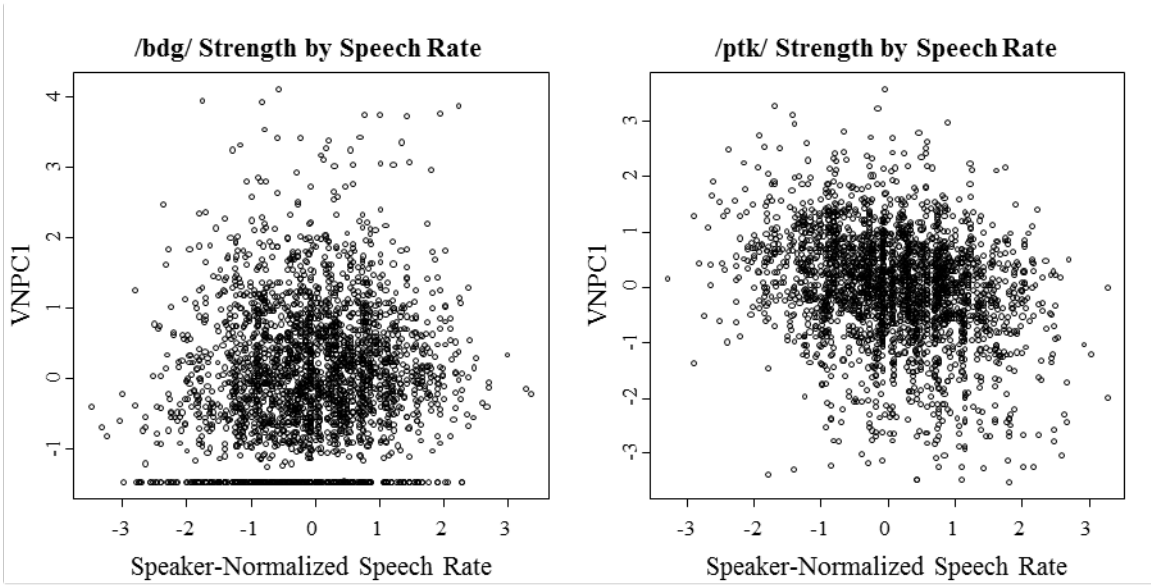

Figure 5.2 Scatterplots of VNPC1 by speaker-normalized speech rate for underlyingly voiced (left panel; $r = 0.112$) and voiceless (right panel; $r = -0.268$) plosives.

The posterior estimate for the interaction of underlying voicing and speech rate in Table 5.1 shows strong evidence for the interaction (P(sign) > .999). Table 5.2 provides posterior estimates of the effect of a 1-SD increase in speaker-normalized speech rate on VNPC1 by underlying voicing.

Table 5.2 Posterior estimate of the effect of a 1-SD increase in speaker-normalized speech rate by underlying voicing.

Voicing	Mean	SD	2.5%	97.5%	P(sign)	
Voiced	0.057	0.016	0.025	0.087	> .999	***
Voiceless	-0.199	0.017	-0.233	-0.165	> .999	***

The descriptive statistics show a strong negative relationship between speech rate and plosive strength for underlyingly voiceless plosives (as one would expect; also consistent with the results for fricatives in Section 4.1.1.2, but with a smaller effect size here), and a weaker positive relationship for underlyingly voiced plosives. The posterior estimates show that these relationships hold (with smaller magnitudes), with strong evidence for the effect direction in each case. While the positive relationship between speech rate and voiced plosive strength may seem somewhat counterintuitive, this is likely the result of intervocalic voiced plosives in Spanish leniting to the point of elision in a gradient manner (as discussed in Section 2.3.2.3). As the intensity difference of the plosive increases, it is more likely that two syllable nuclei will be identified for the VCV sequence.

5.1.2. Linguistic factors

5.1.2.1. Stress

Descriptive statistics for plosive strength by stress are provided in Table 5.3, with corresponding boxplots in Figure 5.3.

Table 5.3 Descriptive statistics for VNPC1 by stress.

Stress	N	Minimum	Median	Maximum	Mean	SD
Tonic	2007	-3.232	0.226	4.115	0.222	1.029
Post-Tonic	1629	-3.462	-0.036	3.922	-0.076	0.966
Unstressed	1645	-3.528	-0.142	3.718	-0.196	0.943

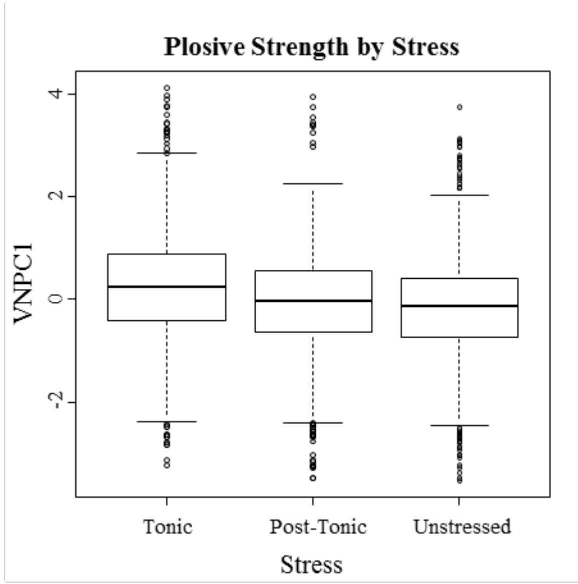

Figure 5.3 Boxplots of VNPC1 by stress.

Posterior group means for VNPC1 by stress are given in Table 5.4, with pairwise comparisons given in Table 5.5.

Table 5.4 Posterior group means for VNPC1 by stress.

Stress	Mean	SD	2.5%	97.5%
Tonic	0.023	0.052	-0.082	0.126
Post-Tonic	-0.237	0.055	-0.345	-0.127
Unstressed	-0.374	0.056	-0.485	-0.264

Table 5.5 Posterior group differences for VNPC1 by stress.

Contrast	Mean	SD	2.5%	97.5%	P(sign)	
Tonic - Post-Tonic	0.260	0.033	0.195	0.324	> .999	***
Tonic - Unstressed	0.397	0.036	0.328	0.467	> .999	***
Post-Tonic - Unstressed	0.137	0.036	0.067	0.207	> .999	***

The descriptive statistics show a strength hierarchy by stress of Tonic > Post-Tonic > Unstressed. The posterior pairwise provide strong evidence for the overall hierarchy found in the descriptive statistics (P(sign) > .999 for all contrasts). This is consistent with previous research on the effect of Spanish stress on plosive production (e.g. Carrasco et al., 2012; Hualde, Simonet, et al., 2011; Torreira & Ernestus, 2011). However, the hierarchy found for plosives here is different than that found for Spanish fricatives in Section 4.1.2.1, where the trend was still the same, but there was very little evidence for a difference between tonic and post-tonic fricatives. These findings support differentiating between the Post-Tonic and Unstressed conditions in research on Spanish phonetics, as the treatment of post-tonic consonants may differ by consonant type.

5.1.2.2. Word position

For the analysis of word position, only the spontaneous speech data are considered (Cuzco and Lima interviews and Valladolid task-oriented dialogues), as this is the only case in which the factor is contrastive (the read speech task only contained planned observations of word-medial plosives and the word position factor was coded as `NA` for that subset). Descriptive statistics for plosive strength by word position and underlying voicing are provided in Table 5.6, with corresponding boxplots in Figure 5.4.

Table 5.6 Descriptive statistics for VNPC1 by underlying voicing and word position.

Voicing	Word Position	N	Minimum	Median	Maximum	Mean	SD
Voiced	Initial	444	-1.481	-0.119	4.115	-0.059	1.026
	Medial	1241	-1.481	-0.271	3.955	-0.241	0.899
Voiceless	Initial	722	-3.475	-0.153	2.749	-0.214	0.997
	Medial	855	-3.528	-0.159	2.375	-0.244	1.003

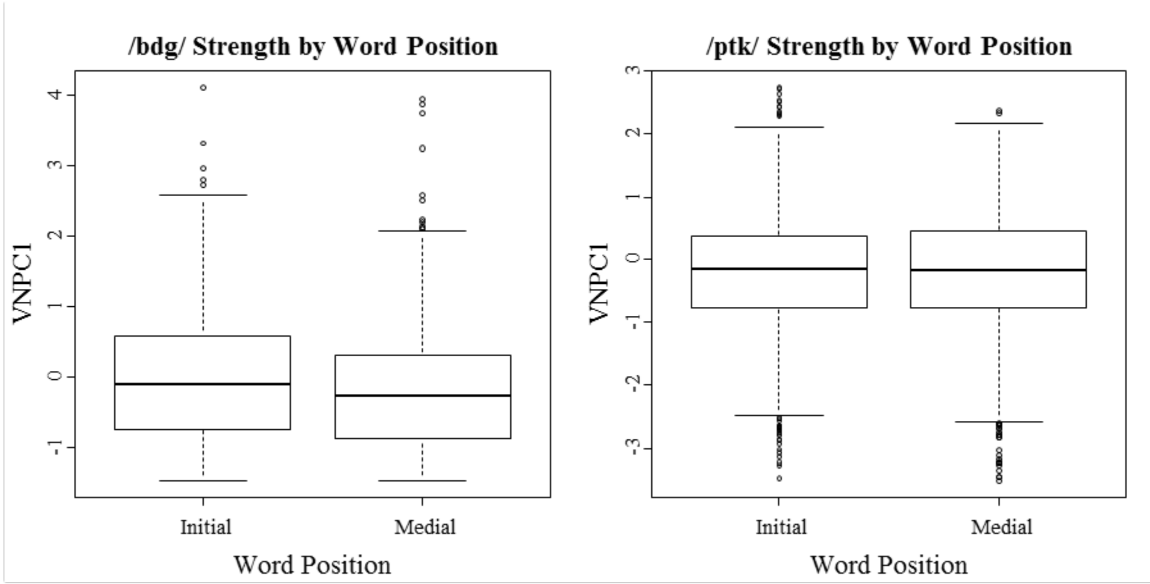

Figure 5.4 Boxplots of VNPC1 by word position for underlyingly voiced (left panel) and voiceless (right panel) plosives.

There is strong evidence for an interaction between word position and underlying voicing (P(sign) = .999 in Table 5.1). Posterior group means for VNPC1 by word position and underlying voicing are given in Table 5.7, with corresponding effect estimates for word position within underlying voicing in Table 5.8.

Table 5.7 Posterior group means for VNPC1 by underlying voicing and word position.

Voicing	Word Position	Mean	SD	2.5%	97.5%
Voiced	Initial	-0.129	0.065	-0.253	0.001
	Medial	-0.407	0.057	-0.517	-0.296
Voiceless	Initial	-0.289	0.074	-0.433	-0.145
	Medial	-0.352	0.071	-0.489	-0.213

Table 5.8 Posterior group differences for VNPC1 by word position within underlying voicing.

Voicing	Contrast	Mean	SD	2.5%	97.5%	P(sign)	
Voiced	Initial - Medial	0.278	0.054	0.174	0.386	> .999	***
Voiceless	Initial - Medial	0.063	0.049	-0.032	0.161	.899	x

For underlyingly voiced plosives, there is strong evidence (P(sign) > .999) that word-initial plosives are stronger than word-medial plosives, with a moderate effect size of 0.278. For underlyingly voiceless plosives, the effect direction is the same, but much smaller in magnitude (0.063), and there is little evidence for the direction of the effect (P(sign) = .899). The literature shows conflicting evidence for an effect of word position on plosive voicing, with some studies finding no effect for either set of plosives (e.g. Hualde, Simonet, et al., 2011), and others finding an effect for voiced plosives (e.g. Carrasco, 2008). To my knowledge, no study has found that word position has an effect on the strength of /ptk/. The data presented here show an effect for voiced plosives, and possibly a much smaller difference for voiceless plosives.

5.1.2.3. Preceding vowel height

Descriptive statistics for plosive strength by preceding vowel height are provided in Table 5.9, with corresponding boxplots in Figure 5.5.

Table 5.9 Descriptive statistics for VNPC1 by underlying voicing and preceding vowel height.

Voicing	Preceding Vowel	N	Minimum	Median	Maximum	Mean	SD
Voiced	High	892	-1.481	0.019	3.955	0.038	1.039
	Non-High	1802	-1.481	-0.052	4.115	-0.019	0.980
Voiceless	High	769	-3.528	0.182	2.798	0.103	0.998
	Non-High	1818	-3.475	0.024	3.592	-0.043	0.998

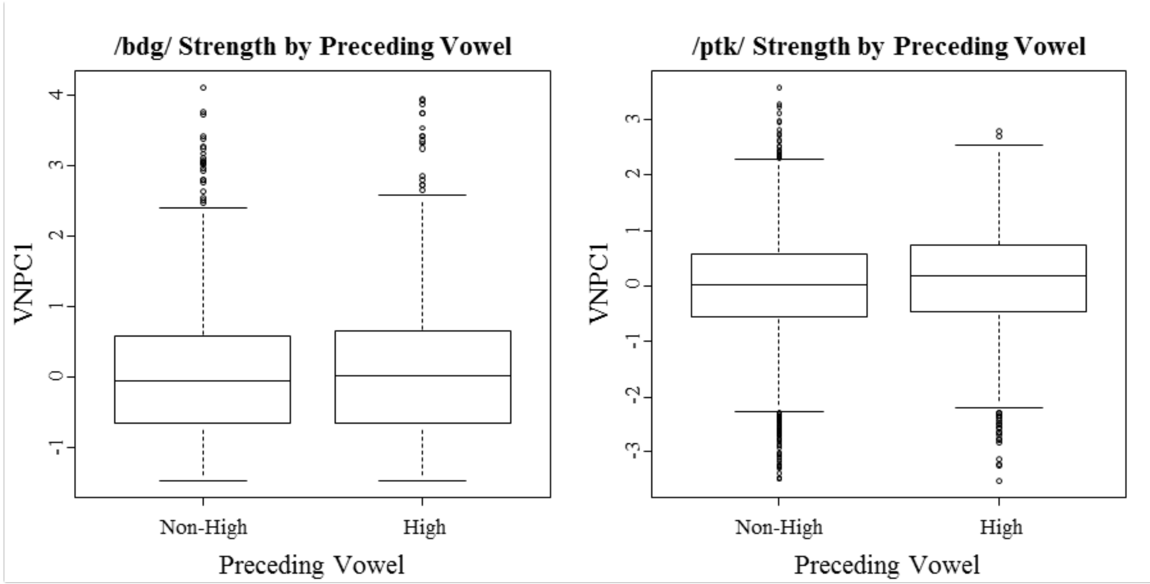

Figure 5.5 Boxplots of VNPC1 by preceding vowel height for underlyingly voiced (left panel) and voiceless (right panel) plosives.

There is little evidence in the data for an interaction between underlying voicing and preceding vowel height (P(sign) = .825 in Table 5.1), but strong evidence for the main effect (P(sign) = .999 in Table 5.1). For this reason, in Table 5.10, the interaction term is averaged over, and posterior group means are given for preceding high and non-high vowels' VNPC1 overall. Table 5.11 provides the corresponding effect estimate.

Table 5.10 Posterior group means for VNPC1 by preceding vowel height.

Preceding Vowel	Mean	SD	2.5%	97.5%
High	-0.146	0.054	-0.253	-0.040
Non-High	-0.246	0.053	-0.350	-0.142

Table 5.11 Posterior group differences for VNPC1 by preceding vowel height.

Contrast	Mean	SD	2.5%	97.5%	P(sign)	
High - Non-High	0.100	0.032	0.038	0.164	> .999	***

There is strong evidence for a small effect of preceding vowel height that does not differ in effect magnitude for underlying voiced and voiceless plosives, with plosives following high vowels being stronger than plosives following non-high vowels. This is different from the effect found for fricative NPC1 in Section 4.1.2.2, where a preceding high vowel led to a shorter more voiced fricative than a preceding non-high vowel. This may be because plosive VNPC1 incorporates intensity-based acoustic measures of consonant constriction, while fricative NPC1 does not. Of the five measures that make up VNPC1, we would only expect a preceding high vowel to lower overall consonant duration; for voiceless duration and percent voiceless, we would expect less voicing bleed-in from a preceding high vowel than non-high vowel, as discussed in Section 4.1.2.2; for intensity difference and intensity velocity, we would expect a preceding high vowel to lead to less articulatory undershoot since the articulators are closer together during the production of a high vowel than a non-high vowel. Overall then, for VNPC1, we have strong evidence for plosives following high vowels being (slightly) stronger than plosives following non-high vowels.

5.1.2.4. Following vowel height

Descriptive statistics for plosive strength by following vowel height are provided in Table 5.12, with corresponding boxplots in Figure 5.6.

Table 5.12 Descriptive statistics for VNPC1 by underlying voicing and following vowel height.

Voicing	Following Vowel	N	Minimum	Median	Maximum	Mean	SD
Voiced	High	738	-1.481	0.072	4.115	0.149	1.066
	Non-High	1956	-1.481	-0.068	3.922	-0.056	0.968
Voiceless	High	484	-3.232	-0.036	2.375	-0.109	0.902
	Non-High	2103	-3.528	0.082	3.592	0.025	1.020

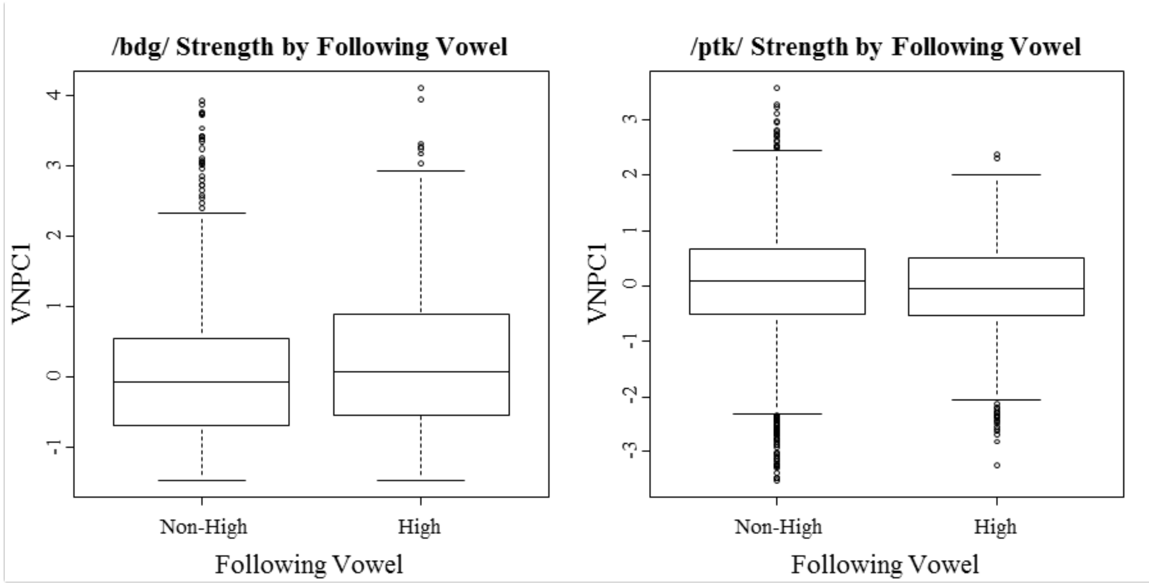

Figure 5.6 Boxplots of VNPC1 by following vowel height for underlyingly voiced (left panel) and voiceless (right panel) plosives.

The estimates in Table 5.1 show strong evidence for an interaction between following vowel height and underlying voicing (P(sign) > .999), as is also evident in Figure 5.6. Posterior group means for VNPC1 by underlying voicing and following vowel height are given in Table 5.13, with corresponding pairwise comparisons given in Table 5.14.

Table 5.13 Posterior group means for VNPC1 by underlying voicing and following vowel height.

Voicing	Following Vowel	Mean	SD	2.5%	97.5%
Voiced	High	-0.139	0.059	-0.256	-0.022
	Non-High	-0.201	0.054	-0.305	-0.093
Voiceless	High	-0.318	0.076	-0.468	-0.169
	Non-High	-0.128	0.068	-0.261	0.008

Table 5.14 Posterior group differences for VNPC1 by following vowel height within underlying voicing.

Voicing	Contrast	Mean	SD	2.5%	97.5%	P(sign)	
Voiced	High - Non-High	0.062	0.040	-0.016	0.141	.937	*
Voiceless	High - Non-High	-0.190	0.044	-0.275	-0.102	> .999	***

There is strong evidence that voiceless plosives that precede high vowels are moderately weaker than voiceless plosives that precede non-high vowels, and weak evidence for an effect in the opposite direction with a smaller magnitude for voiced plosives. As was the case for the effect of preceding vowel height, the effect of following vowel height for voiceless plosive VNPC1 is the opposite of the effect found for voiceless fricative NPC1 in Section 4.1.2.3. Here there is no clear phonetic explanation, as we would expect the effect of a following high vowel on the voicing measures to be similar for voiceless fricatives and voiceless plosives, and the effect of a following high vowel on constriction and duration to be similar for voiceless plosives and voiced plosives, meaning that if a following high vowel has a positive effect for fricative NPC1 and for voiced plosive VNPC1, then we should see a positive effect for voiceless plosives as well, but we do not. Previous research on Spanish consonant lenition that includes the surrounding vowels qualities as predictors has not given full descriptive statistics for the effect of following vowel height or the estimates from the regressions since they were included as controls, so there are no numbers to compare these findings to, and I leave the reasons for this unexpected affect aside for future research.

5.1.2.5. Task

For the analysis of task, only the word-medial plosives from spontaneous speech from Lima and Cuzco and the data from the read speech task were considered, as this is the only case where the factor is contrastive (Valladolid speakers only participated in one task and the read speech task had no word-initial planned observations). Descriptive statistics for plosive strength by task and underlying voicing are provided in Table 5.15, with corresponding boxplots in Figure 5.7.

Table 5.15 Descriptive statistics for VNPC1 by task.

Task	N	Minimum	Median	Maximum	Mean	SD
Read Speech	2019	-3.034	0.348	3.922	0.341	0.955
Spontaneous Speech	1071	-3.528	-0.076	3.955	-0.126	1.006

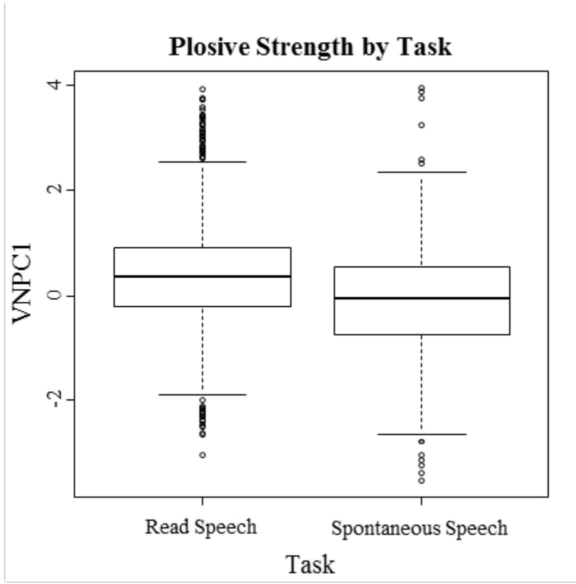

Figure 5.7 Boxplots of VNPC1 by task.

Posterior group means for VNPC1 for the read speech task and Cuzco and Lima interviews are given in Table 5.16, with the corresponding comparison in Table 5.17.

Table 5.16 Posterior group means for VNPC1 by task.

Task	Mean	SD	2.5%	97.5%
Read Speech	0.031	0.076	-0.119	0.179
Spontaneous Speech	-0.348	0.067	-0.479	-0.212

Table 5.17 Posterior group differences for VNPC1 by task.

Contrast	Mean	SD	2.5%	97.5%	P(sign)	
Read - Spontaneous	0.380	0.058	0.264	0.496	> .999	***

The descriptive statistics show, and the posterior estimates confirm, strong evidence for an effect of task: read speech plosives are 0.380 standard deviations stronger than spontaneous speech plosives (P(sign) > .999). This is consistent with a number of studies on Spanish plosive lenition that have examined speech in different tasks (e.g. Hualde, Simonet, et al., 2011; Lewis, 2000, 2001; Munday, 2001), and we now also have evidence that the effect of task is not merely due to a difference in speech rate, since it was also included in the regression.

5.1.3. Social factors

5.1.3.1. Cuzco age group, education level, and Quechua bilingualism

For the analysis of the social factors of age group, education level, and Quechua bilingualism, only the data from the Cuzco speakers are considered, as this is the only subset of the data where the factor is contrastive. Descriptive statistics for plosive strength by age group are provided in Table 5.18 (with corresponding boxplots in Figure 5.8), descriptive statistics for plosive strength by education level are provided in Table 5.19 (with corresponding boxplots in Figure 5.9), and descriptive statistics for Quechua bilingualism and underlying voicing are provided in Table 5.20 (with corresponding boxplots in Figure 5.10).

Table 5.18 Descriptive statistics for VNPC1 by age group.

Age Group	N	Minimum	Median	Maximum	Mean	SD
Older	1362	-2.690	0.464	4.115	0.466	0.934
Younger	1510	-2.548	0.330	3.922	0.373	0.844

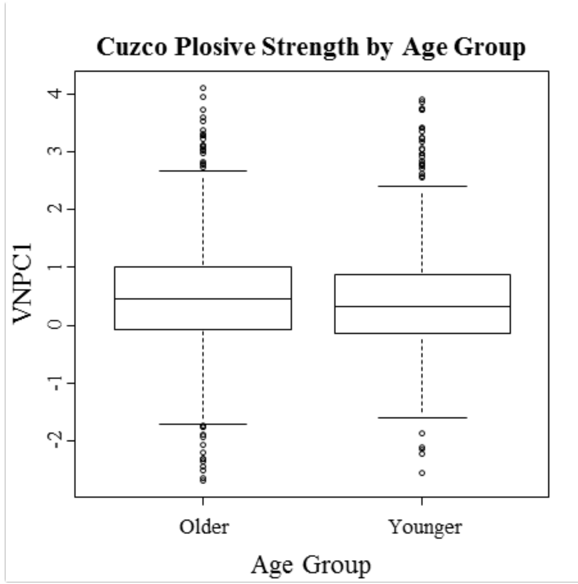

Figure 5.8 Boxplots of VNPC1 by age group.

Table 5.19 Descriptive statistics for VNPC1 by education level.

Education Level	N	Minimum	Median	Maximum	Mean	SD
Secondary	1445	-2.636	0.417	3.869	0.433	0.931
University	1427	-2.690	0.368	4.115	0.401	0.844

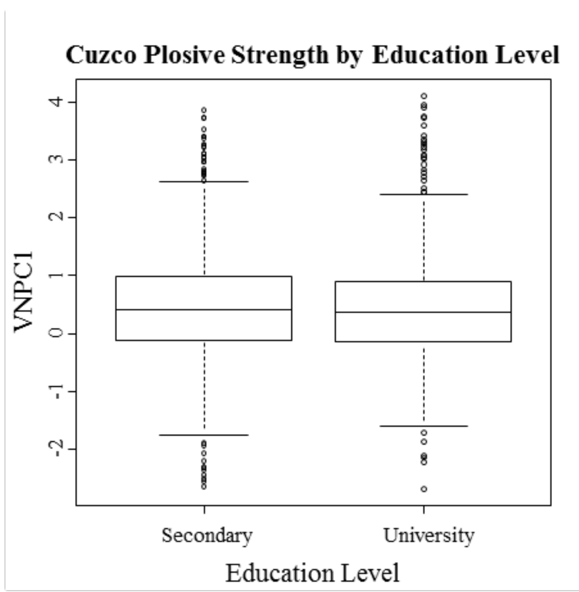

Figure 5.9 Boxplots of VNPC1 by education level.

Table 5.20 Descriptive statistics for VNPC1 by underlying voicing and Quechua bilingualism.

Voicing	Quechua Bilingual	N	Minimum	Median	Maximum	Mean	SD
Voiced	Yes	693	-1.481	0.417	3.746	0.463	0.869
	No	764	-1.481	0.397	4.115	0.446	1.013
Voiceless	Yes	686	-2.548	0.430	3.285	0.456	0.722
	No	729	-2.690	0.330	3.592	0.307	0.904

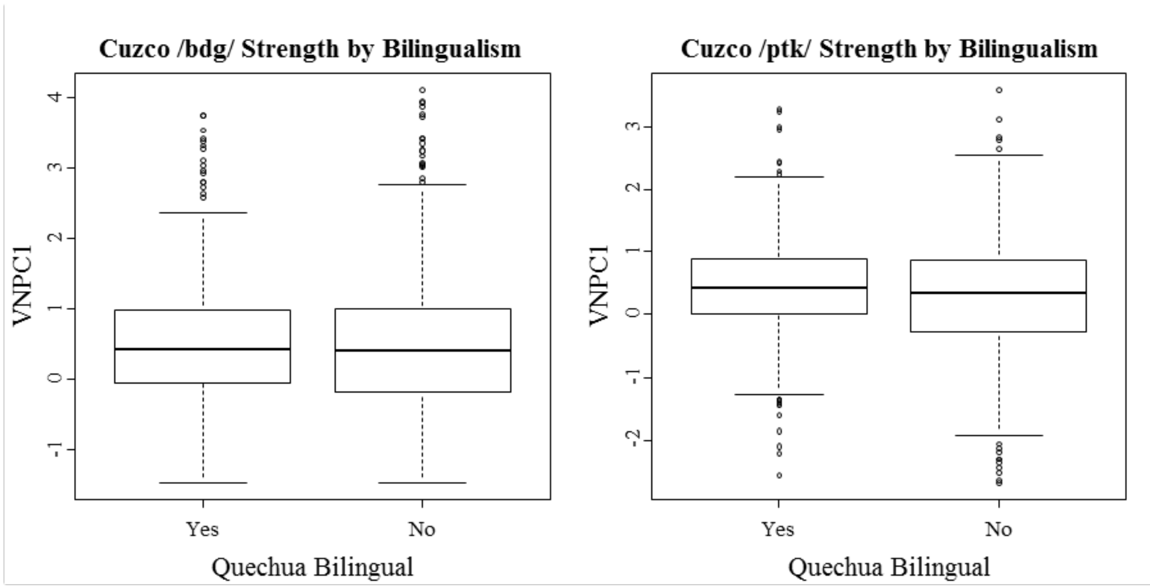

Figure 5.10 Boxplots of VNPC1 by Quechua bilingualism for underlyingly voiced (left panel) and voiceless (right panel) plosives.

The descriptive statistics show essentially no trends for any of the Cuzco-only social factors except a small trend for Quechua-bilingual speakers having stronger /ptk/ than monolingual Cuzco speakers. There was little evidence for any of these factors having main effects, and also little evidence for an interaction between underlying voicing and Quechua bilingualism (of the four relevant estimates in Table 5.1, maximum P(sign) = .756). That is, at least in the data collected for this study in Cuzco, age group, education level, and Quechua bilingualism did not prove to be important from a control perspective (as this is not a sociolinguistic study, and the number of speakers in each cell of the full contingency table of these factors is small, I will refrain from making a sociolinguistic analysis of these results).

5.1.3.2. Speaker sex

As there is not strong evidence for an interaction between sex and underlying voicing (P(sign) = .965 in Table 5.1), but there is strong evidence for an interaction between sex and dialect (maximum P(sign) = .978 in Table 5.1), I provide descriptive statistics for VNPC1 by sex and dialect in Table 5.21, with corresponding boxplots in Figure 5.11 (now considering all of the data again).

Table 5.21 Descriptive statistics for VNPC1 by dialect and sex.

Dialect	Sex	N	Minimum	Median	Maximum	Mean	SD
Cuzco	Female	1429	-2.209	0.638	4.115	0.693	0.842
	Male	1443	-2.690	0.133	3.534	0.144	0.850
Lima	Female	380	-3.034	-0.473	1.949	-0.535	0.844
	Male	482	-3.528	-0.868	2.706	-0.826	0.936
Valladolid	Female	790	-3.265	-0.174	2.749	-0.160	0.816
	Male	757	-3.475	-0.521	2.718	-0.621	0.848

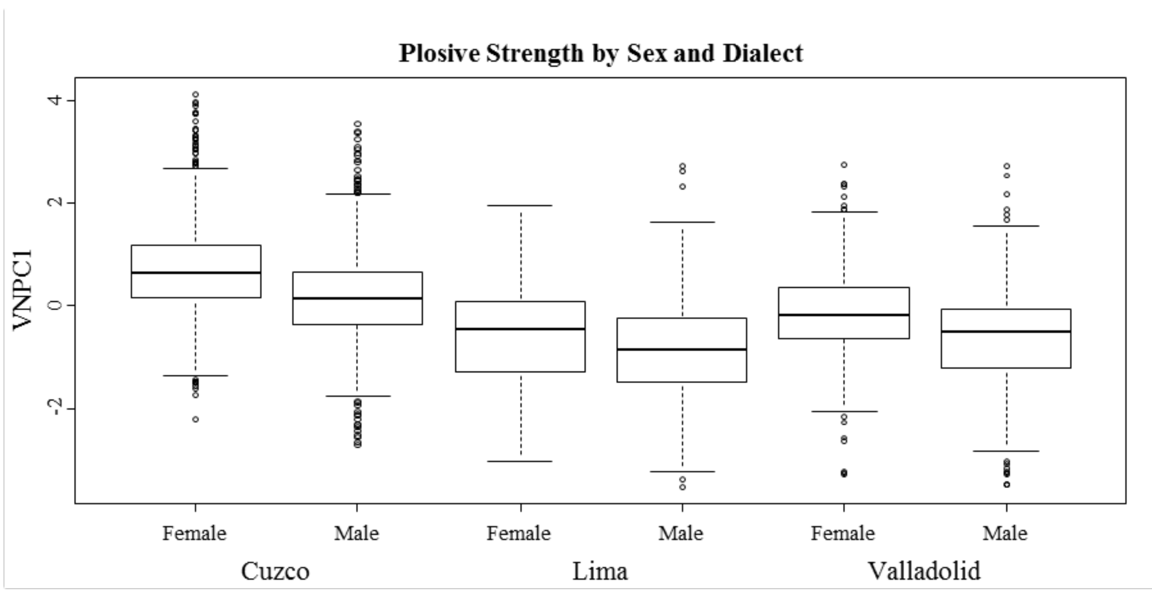

Figure 5.11 Boxplots of VNPC1 by dialect and sex.

Posterior group means for VNPC1 by sex and dialect are given in Table 5.22, with pairwise comparisons for sex within dialect given in Table 5.23.

Table 5.22 Posterior group means for VNPC1 by dialect and sex.

Dialect	Sex	Mean	SD	2.5%	97.5%
Cuzco	Female	0.749	0.093	0.564	0.931
	Male	0.079	0.087	-0.095	0.245
Lima	Female	-0.559	0.155	-0.865	-0.253
	Male	-0.732	0.156	-1.038	-0.425
Valladolid	Female	-0.139	0.116	-0.368	0.085
	Male	-0.574	0.116	-0.803	-0.349

Table 5.23 Posterior group differences for VNPC1 by sex within dialect.

Dialect	Contrast	Mean	SD	2.5%	97.5%	P(sign)	
Cuzco	Female - Male	0.670	0.117	0.436	0.899	> .999	***
Lima	Female - Male	0.173	0.224	-0.270	0.614	.786	x
Valladolid	Female - Male	0.435	0.147	0.147	0.726	.999	***

The effect direction for sex is the same in all three dialects: female speakers have stronger plosives on average than male speakers. However, while there is strong evidence for this effect in Cuzco and Valladolid, there is little evidence for the effect direction in the Lima data. Additionally, the effect magnitude is substantially larger for Cuzco than for Valladolid, and substantially stronger for Valladolid than for Lima.

5.2. Place of articulation

Descriptive statistics for plosive strength by place of articulation and underlying voicing are provided in Table 5.24, with corresponding boxplots in Figure 5.12.

Table 5.24 Descriptive statistics for VNPC1 by underlying voicing and place of articulation.

Voicing	Place	N	Minimum	Median	Maximum	Mean	SD
Voiced	Bilabial	956	-1.481	-0.021	3.955	0.001	0.951
	Dental	960	-1.481	-0.175	4.115	-0.057	1.128
	Velar	778	-1.481	0.068	3.922	0.069	0.881
Voiceless	Bilabial	796	-3.364	0.217	3.592	0.138	1.022
	Dental	902	-3.239	-0.021	3.124	-0.021	0.963
	Velar	889	-3.528	0.008	2.749	-0.103	1.004

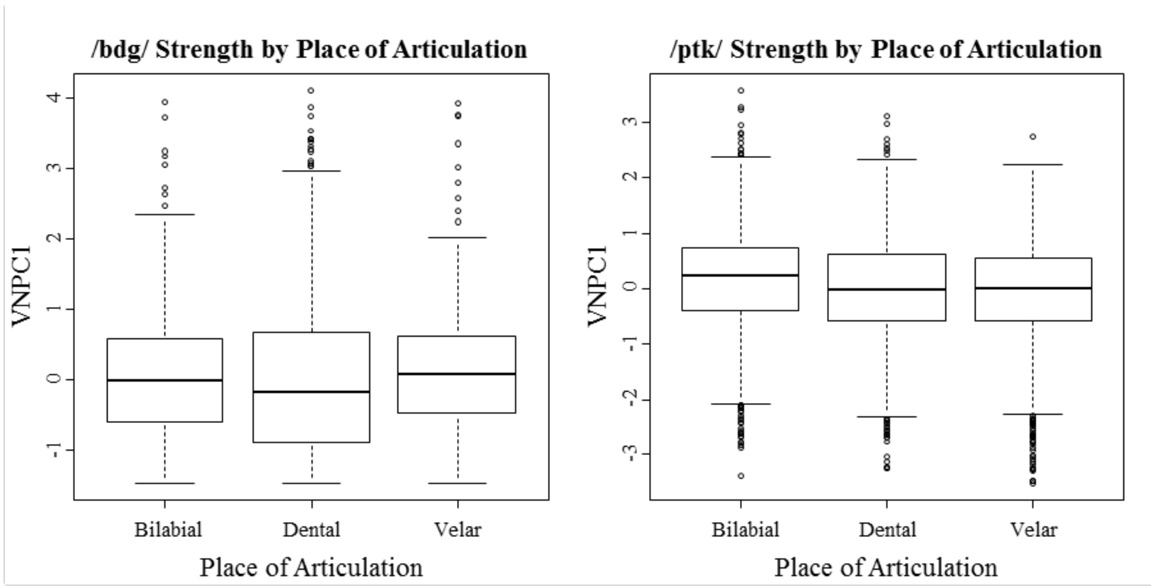

Figure 5.12 Boxplots of VNPC1 by place of articulation for underlyingly voiced (left panel) and voiceless (right panel) plosives.

There is strong evidence for an interaction between underlying voicing and place of articulation (maximum P(sign) > .999 in Table 5.1). Posterior group means for VNCP1 by underlying voicing and place of articulation are provided in Table 5.25, with corresponding pairwise comparisons within underlying voicing in Table 5.26.

Table 5.25 Posterior group means for VNPC1 by underlying voicing and place of articulation.

Voicing	Place	Mean	SD	2.5%	97.5%
Voiced	Bilabial	-0.178	0.058	-0.290	-0.063
	Dental	-0.271	0.065	-0.399	-0.143
	Velar	-0.060	0.061	-0.179	0.060
Voiceless	Bilabial	-0.088	0.074	-0.234	0.058
	Dental	-0.248	0.075	-0.396	-0.103
	Velar	-0.332	0.073	-0.476	-0.187

Table 5.26 Posterior group differences for VNPC1 by place of articulation within underlying voicing.

Voicing	Contrast	Mean	SD	2.5%	97.5%	P(sign)	
Voiced	Bilabial - Dental	0.094	0.045	0.004	0.183	.980	***
	Bilabial - Velar	-0.117	0.050	-0.215	-0.019	.991	***
	Dental - Velar	-0.211	0.064	-0.335	-0.083	.999	***
Voiceless	Bilabial - Dental	0.161	0.046	0.069	0.250	> .999	***
	Bilabial - Velar	0.244	0.046	0.153	0.334	> .999	***
	Dental - Velar	0.083	0.053	-0.022	0.186	.940	*

The descriptive statistics show, and the posterior estimates confirm, that the effect of place of articulation on plosive strength is not the same for voiceless and voiced plosives. For the voiceless plosives, there is a hierarchy /p/ > /t/ > /k/, with strong evidence for the bilabial plosive being stronger than the dental and velar plosives (P(sign) > .999 in both cases), and weak evidence for the dental being stronger than the velar (P(sign) = .940). That is, the further backward in the vocal tract the voiceless plosive is articulated, the weaker it is. This is in line with what is predicted by the aerodynamic voicing constraint (Ohala & Riordan, 1979), and also with previous research on Spanish that has shown /k/ to be the most likely to lenite (e.g. Hualde, Simonet, et al., 2011; Torreira & Ernestus, 2011). For the voiced plosives, there is a hierarchy /b/ > /g/ > /d/, with strong evidence for all group differences (P(sign) ≥ .980 in all cases). This is consistent with previous research on Spanish that has shown /d/ to be the most likely to lenite, and delete in the sequence /ado/ in words like /lado/ ‘side’ and the past participle /-ado/

(Caravedo, 1990; Hualde, Simonet, et al., 2011; Lipski, 1994; Williams, 1987; among others), and with the historical resilience of Spanish /b/ (intervocalic Western Romance */d/ and */g/ were elided in the historical development of Spanish, while */b/ was not, instead merging with */β/ in Old Spanish; Penny (2002)). The difference between the ranking of the dental and velar plosives based on underlying voicing may indicate a difference in the relationship between voiceless and voiced plosive strength at the individual level, a possibility that is revisited in the analysis of individual differences in Section 6.2.

5.3. Dialect differences

For the analysis of dialect, only data from the spontaneous speech data are considered (that is, the read speech data from Lima and Cuzco are not considered), and only data from those Cuzco speakers who match the Lima and Valladolid speakers demographically are considered (younger age group, university educated, and monolingual).[25] Descriptive statistics for plosive strength by dialect and underlying voicing are provided in Table 5.27, with corresponding boxplots in Figure 5.13.

Table 5.27 Descriptive statistics for VNPC1 by underlying voicing and dialect.

Voicing	Dialect	N	Minimum	Median	Maximum	Mean	SD
Voiced	Cuzco	73	-1.481	-0.164	2.723	0.050	1.019
	Lima	231	-1.481	-0.944	1.524	-0.875	0.671
	Valladolid	796	-1.481	-0.371	2.337	-0.383	0.778
Voiceless	Cuzco	64	-2.134	-0.031	1.608	-0.074	0.686
	Lima	206	-3.528	-0.898	2.335	-0.933	1.099
	Valladolid	751	-3.475	-0.329	2.749	-0.388	0.945

[25] While these Cuzco-only social factors had no effect on VNPC1, comparing only the demographically matched speakers is more principled; using all of the Cuzco to make the comparison does not change the outcome or interpretation (as should be expected).

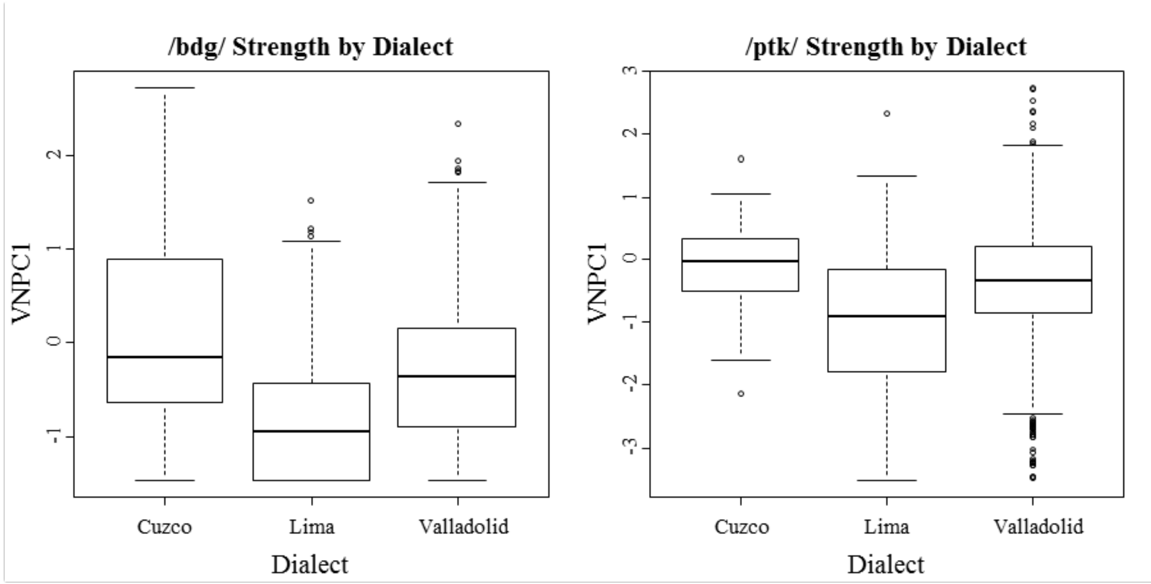

Figure 5.13 Boxplots of VNPC1 by dialect for underlyingly voiced (left panel) and voiceless (right panel) plosives.

Posterior group means for VNPC1 by dialect and underlying voicing are given in Table 5.28, with corresponding pairwise comparisons for dialect within underlying voicing in Table 5.29.

Table 5.28 Posterior group means for VNPC1 by underlying voicing and dialect.

Voicing	Dialect	Mean	SD	2.5%	97.5%
Voiced	Cuzco	0.326	0.121	0.081	0.563
	Lima	-0.873	0.114	-1.094	-0.648
	Valladolid	-0.279	0.094	-0.462	-0.094
Voiceless	Cuzco	0.114	0.136	-0.157	0.382
	Lima	-0.714	0.144	-0.996	-0.419
	Valladolid	-0.435	0.119	-0.670	-0.204

Table 5.29 Posterior group differences for VNPC1 by dialect within underlying voicing.

Voicing	Contrast	Mean	SD	2.5%	97.5%	P(sign)	
Voiced	Cuzco - Lima	1.199	0.161	0.880	1.514	> .999	***
	Cuzco - Valladolid	0.605	0.161	0.290	0.925	> .999	***
	Lima - Valladolid	-0.594	0.143	-0.876	-0.311	> .999	***
Voiceless	Cuzco - Lima	0.827	0.194	0.436	1.217	> .999	***
	Cuzco - Valladolid	0.548	0.191	0.177	0.925	.999	***
	Lima - Valladolid	-0.279	0.183	-0.642	0.085	.939	*

The descriptive statistics show the same hierarchy of strength for both /bdg/ and /ptk/: Cuzco > Valladolid > Lima. The posterior estimates show the same hierarchy, with strong evidence for all group differences for /bdg/, strong evidence for Cuzco having stronger /ptk/ than both Valladolid and Lima, and weak evidence for Valladolid having stronger /ptk/ than Lima. The results for Valladolid are consistent with Munday (2001) and Williams (1987), and the results for Cuzco and Lima confirm the qualitative observations of Caravedo (1990) and Lipski (1994), and add a thorough quantitative analysis of plosive strength to our knowledge of Spanish in Peru. The resistance to intervocalic /ptk/ voicing in Cuzco is also consistent with the findings of Torreira and Ernestus (2011), who argue that unstressed vowel devoicing (which Delforge (2009, 2012) finds in Cuzco) implies a coarticulatory strategy incompatible with extensive intervocalic /ptk/ voicing.

The results also offer some evidence for Hypothesis 2 (from Section 1.6: *In Spanish, for both the voiceless plosives /ptk/ and the voiced plosives /bdg/, the dialects will show the same strength hierarchy of Cuzco > Valladolid > Lima.*). For /bdg/, the hypothesized hierarchy is very clear, with the pairwise comparisons all showing very large effects for which we have strong evidence. For /ptk/, however, while the hierarchy in the posterior group means is the same, the pairwise comparisons show smaller magnitudes of difference between the dialects in all three comparisons. And while the evidence for Cuzco having stronger /ptk/ than both

Valladolid and Lima is strong, the evidence for Valladolid having stronger /ptk/ than Lima is considerably weaker (i.e. we cannot state with confidence that the two are different). Thus, overall the evidence for a correlation between intervocalic /ptk/ lenition and intervocalic /bdg/ lenition in Spanish at the dialectal level is not entirely convincing. However, as discussed in Section 1.2.4, we may still expect to see strong evidence for a correlation between the lenitions at the individual level, since the reason we would expect to see the correlation at the dialectal level is based on correlation at the individual level in combination with a small enough overlap between speakers of different dialects. This possibility is explored in Section 6.2.

Chapter 6. Individual variation results

In this chapter, I examine constraints on individual variation. In Section 6.1, I compare the Valladolid Spanish and Catalan speakers' posterior estimates for /s/ and /f/ strength by language. In Section 6.2, I compare the Spanish speakers' posterior estimates for voiced and voiceless plosive strength by place of articulation. And in Section 6.3, I compare the Valladolid Spanish speakers' posterior estimates for /ptk/ and /sf/ strength. In each case, posterior estimates are generated for each speaker and phoneme by applying appropriate contrasts to both the fixed and random effects at each iteration of the relevant Bayesian mixed effects regression, averaging over the effects of all control predictors. I then run linear regressions on the speaker estimates at each iteration to obtain posterior samples of the relationship between the estimates. Strong evidence for a relationship thus implies a constraint on individual phonetic variation from a multivariate perspective.

6.1. Spanish and Catalan /s/ and /f/

Considering now all data from the fricative experiment, boxplots of each speaker's /s/ and /f/ strengths are given in Figure 6.1.

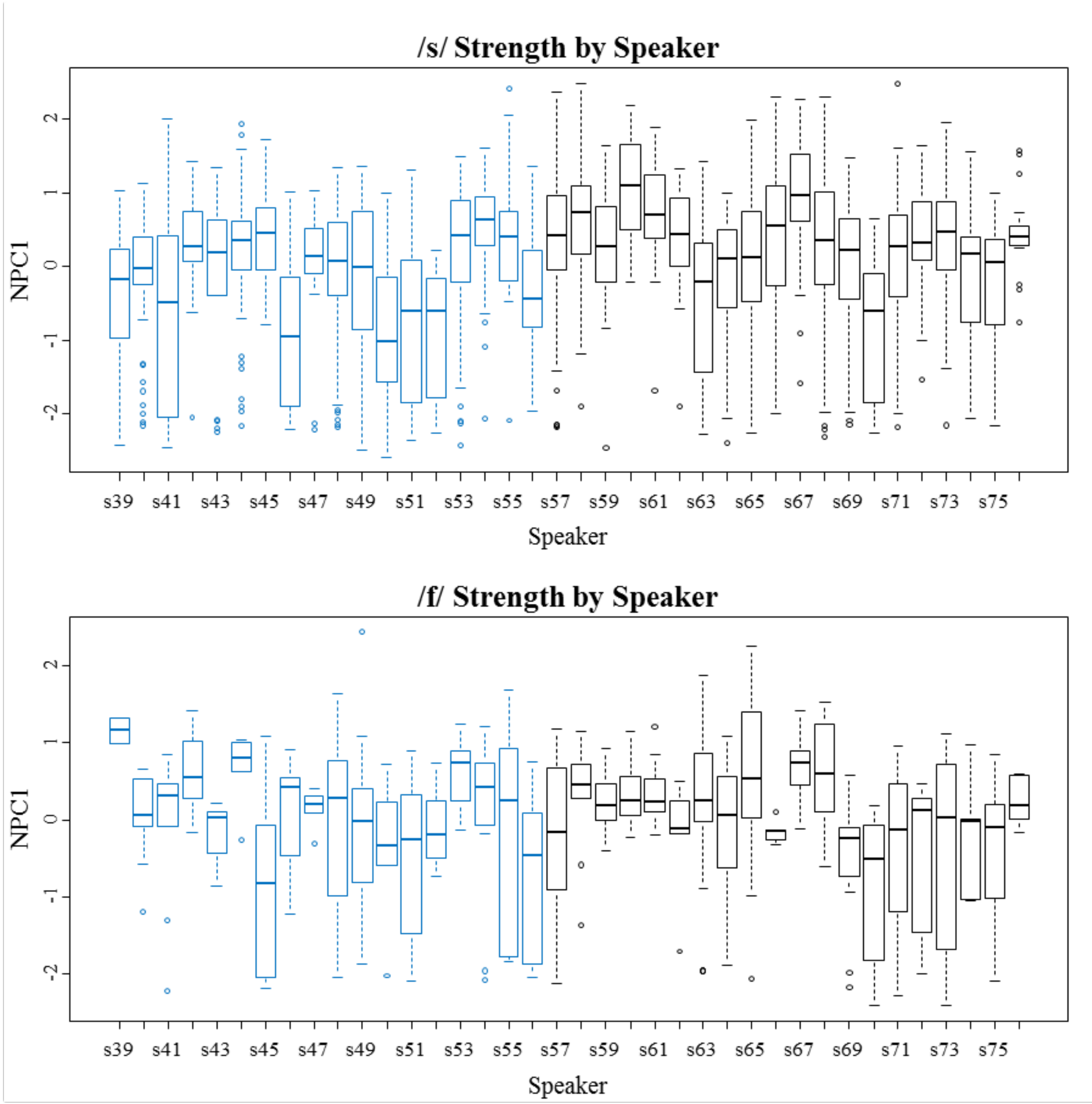

Figure 6.1 Boxplots of NPC1 by speaker for /s/ (top panel) and /f/ (bottom panel). Spanish speakers from Valladolid are in blue and Catalan speakers are in black.

Using the Bayesian mixed effects regression from Chapter 4, for each iteration, I generated a posterior estimate of each speaker's mean fricative strength for each of /s/ and /f/, resulting in a data frame with a column for the /s/ strength estimates and a column for the /f/ strength estimates, and a row for each speaker (for a total of 38 rows). For each of these data frames, an ordinary least squares regression was run on /s/ strength with the full interaction of /f/

strength and language as predictors. The regression estimates and R^2 at each iteration were logged. The posterior mean for R^2 was .400 (posterior standard deviation .114; 95% credible interval [.196, .642]). The posterior distribution of the regression estimates is given in Table 6.1, and Figure 6.2 shows a scatterplot of the posterior mean of the individual estimates with the posterior mean regression line.

Table 6.1 Posterior distribution of the regression on individual speakers' /s/ estimates.

Fixed Effect	Mean	SD	2.5%	97.5%	P(sign)	
Intercept (Corrected Mean)	-0.034	0.043	-0.114	0.055	.796	x
/f/ Strength	0.648	0.177	0.324	1.022	> .999	***
Language, Catalan	0.166	0.042	0.083	0.249	.999	***
/f/ Strength : Language, Catalan	0.027	0.118	-0.211	0.257	.594	x

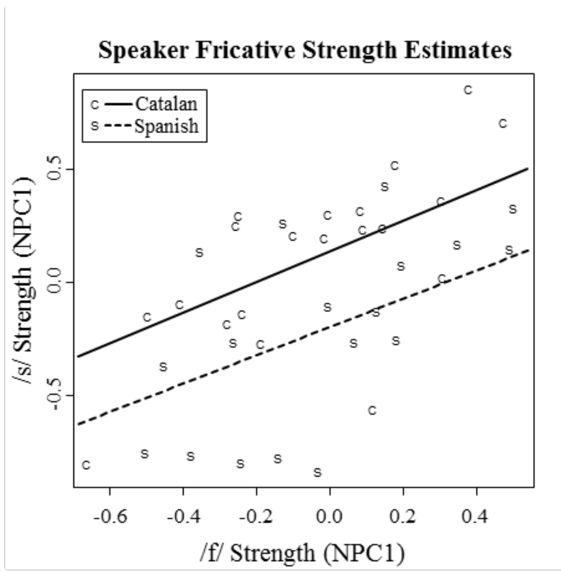

Figure 6.2 Scatterplot of /s/ speaker estimates by /f/ speaker estimates with regression lines by language.

There is strong evidence for a positive relationship between speakers’ /s/ strength and /f/ strength (mean = 0.648, P(sign) > .999). However, there is little evidence for a difference between Spanish and Catalan in the magnitude of this effect (P(sign) = .594). Knowing speakers’ mean /f/ strengths, 40.0% of the variance same speakers’ mean /s/ strength were able to be predicted. These results offer strong support for Hypothesis 3 (from Section 1.6: *The degree to which individual speakers of Catalan and Spanish lenite /s/ will correlate with the degree to which the same speakers lenite /f/.*). From the perspective of articulatory phonology, this should be expected, as both /s/ and /f/ would be defined by a glottal gesture that is voiceless and produces enough airflow across the constriction to create turbulent noise. While /s/ and /f/ differ in other important ways that affect their production (e.g. place of articulation, involvement of the tongue, and stridency), we are seeing strong evidence that speakers do not manipulate independently the degree to which they voice the two consonants intervocalically. The lack of evidence for a difference in the strength of this relationship for Spanish and Catalan is also interesting, given that Catalan has, on average, stronger /s/ realizations than Spanish (Section 4.2). It would thus seem that Catalan and Spanish differ in the target strength for /s/, but not for /f/, and conditioning on these targets speakers deviate in systematic ways.

6.2. Spanish /ptk/ and /bdg/

Considering the /ptk/ and /bdg/ data from all Spanish speakers' spontaneous speech (that is, excluding the read speech data from Cuzco and Lima), boxplots for each speaker's plosive strengths are given in Figure 6.3.

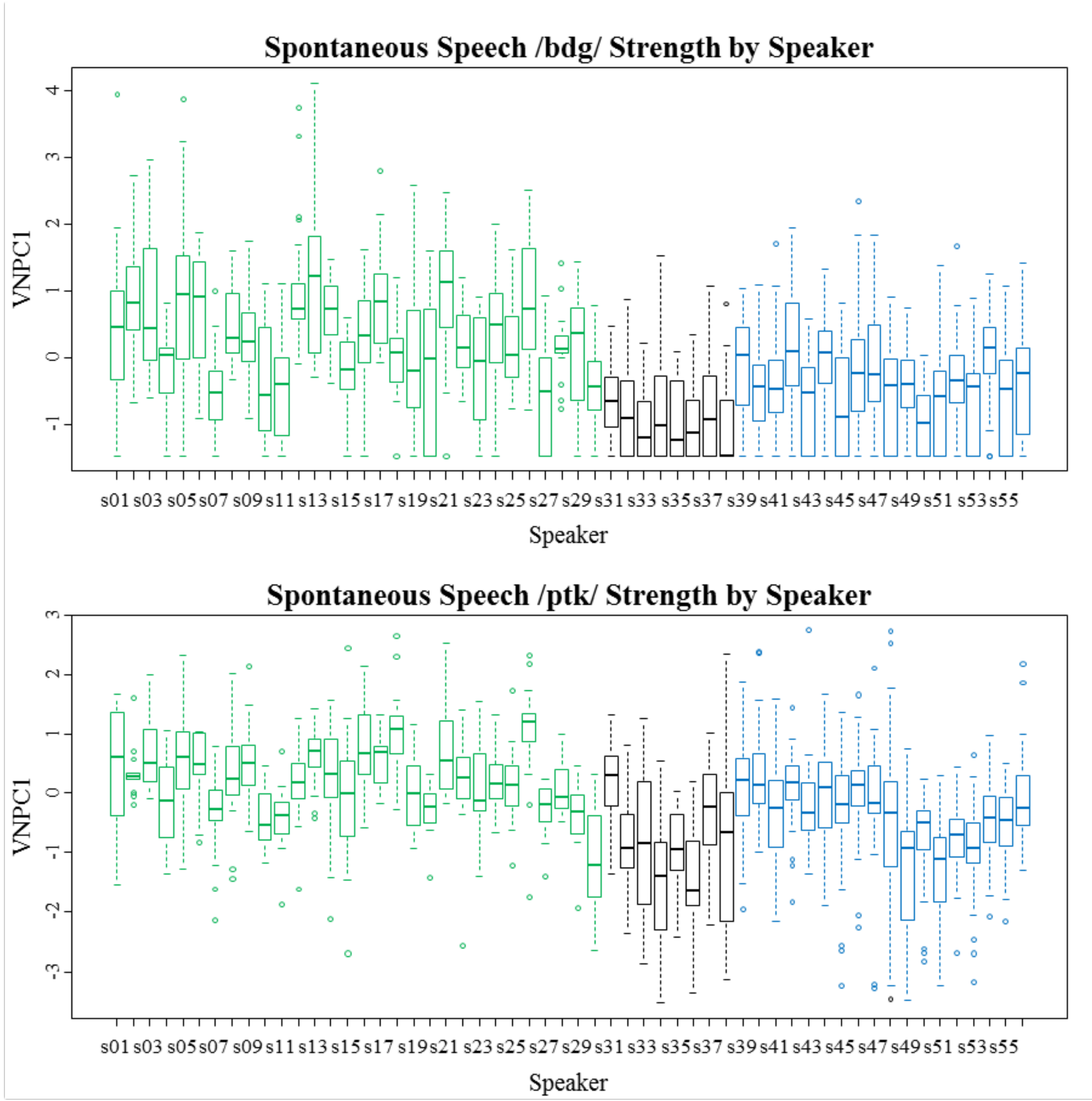

Figure 6.3 Boxplots of VNPC1 by speaker for underlyingly voiced (top panel) and voiceless (bottom panel) plosives. Cuzco speakers are in green; Lima speakers are in black; and Valladolid speakers are in blue.

A trend is apparent in the boxplots, where speakers who produce relatively stronger voiced plosives also produce relatively stronger voiceless plosives. Using the Bayesian mixed effects regression from Chapter 5, for each iteration, I generated a posterior estimate of each speaker's mean plosive strength for each of the six plosive phonemes, resulting in a data frame with a column for the voiced plosive strength estimates and a column for the voiceless plosive strength estimates, and a row for each place of articulation for each speaker (for a total of 168 rows). For each of these data frames, an ordinary least squares regression was run on voiced plosive strength with the full interaction of voiceless strength and place of articulation as predictors. The regression estimates and R^2 at each iteration were logged. The posterior mean for R^2 was .602 (posterior standard deviation .037; 95% credible interval [.528, .674]). The posterior distribution of the regression estimates is given in Table 6.2, the posterior estimate of the slope for each place of articulation is given in Table 6.3, and pairwise comparisons of these slopes are given in Table 6.4. Figure 6.4 shows a scatterplot of the posterior mean of the individual estimates with the posterior mean regression line.

Table 6.2 Posterior distribution of the regression on individual speakers' /bdg/ estimates.

Fixed Effect	Mean	SD	2.5%	97.5%	P(sign)	
Intercept (Corrected Mean)	0.079	0.029	0.022	0.136	.997	***
Voiceless Strength	0.777	0.038	0.704	0.852	> .999	***
Place, Bilabial	-0.109	0.030	-0.168	-0.048	> .999	***
Place, Dental	-0.056	0.032	-0.120	0.008	.959	
Voiceless Strength : Place, Bilabial	0.003	0.023	-0.052	0.042	.630	***
Voiceless Strength : Place, Dental	0.130	0.039	0.055	0.210	> .999	

Table 6.3 Posterior slope for individuals' voiceless plosives strength by place of articulation.

Voiceless Strengh Slope	Mean	SD	2.5%	97.5%	P(sign)	
Bilabial	0.781	0.045	0.691	0.867	> .999	***
Dental	0.908	0.057	0.799	1.022	> .999	***
Velar	0.644	0.052	0.545	0.748	> .999	***

Table 6.4 Posterior contrasts for voiceless plosive strength slope by place of articulation.

Voiceless Strength Slope Contrast	Mean	SD	2.5%	97.5%	P(sign)	
Bilabial - Dental	-0.127	0.052	-0.238	-0.035	.996	***
Bilabial - Velar	0.137	0.051	0.031	0.232	.993	***
Dental - Velar	0.264	0.075	0.121	0.414	> .999	***

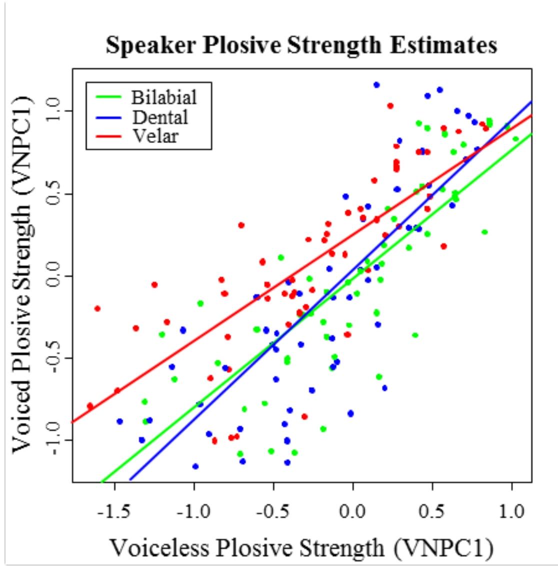

Figure 6.4 Scatterplot of voiced plosive speaker estimates by voiceless plosive speaker estimates, with regression lines by place of articulation.

As can be seen in Figure 6.4, there is a very strong relationship between a speaker's mean voiced plosive strength and the same speaker's mean voiceless plosive strength conditioning on place of articulation. As can be seen in Table 6.3, there is strong evidence that the correlation is positive for all three places of articulation, and, as can be seen in Table 6.4, there is also strong evidence that the strength of the correlation is different for all three places of articulation. Knowing the mean strength of the speakers' /p/, /t/, and /k/ productions, 60.2% of the variance in the mean strength of the same speakers' /b/, /d/, and /g/ productions was able to be predicted. The relationship between the two sets of plosives at the level of the individual is strongest for dental plosives (0.908), weaker for bilabial plosives (0.781), and weakest for velar plosives (0.644).

These results thus offer very strong support for Hypothesis 4 (from Section 1.6: *In Spanish, the degree to which individual speakers lenite /ptk/ will correlate with the degree to which the same speakers lenite /bdg/, and this relationship may differ by place of articulation.*).

While the correlation at the individual level is very clear, at the dialectal level the evidence was less clear (Section 5.3), reinforcing the importance of examining individuals when dealing with questions of systematicity in phonological systems. The differences in the correlation strength by place of articulation at the individual level also offer an explanation for why, a the population level, we find a strength hierarchy of /p/ > /t/ > /k/ for voiceless plosives, but a hierarchy of /b/ > /g/ > /d/ for voiced plosives. The correlation between /k/ and /g/ strength is the weakest and the correlation between /t/ and /d/ strength is the strongest, which could explain why the ordering of the dental and velar plosives is different for the two hierarchies.

These results are, like the results for fricatives in Section 6.1, consistent with what we would expect from articulatory phonology. Conditioning on place of articulation, the voiced and voiceless plosive sets share similar oral gestures, and the results here support the hypothesis that the relative magnitude of these gestures (as determined by acoustic evidence at least), is not manipulated in an entirely independent manner for different phonemes by the same individual. This in turn, then, serves as a good illustration of how contrast preservation can be an outcome from purely phonetic processes. It may well be the case that one speaker weakens /ptk/ to the point where they resemble the /bdg/ productions of another speaker, but the reduction in gestural magnitude that leads to the speaker’s weak /ptk/ productions is also present in their /bdg/ productions, such that for each speaker the two sets remain distinct.

6.3. Valladolid /ptk/ and /sf/

Now considering all data from Valladolid from both experiments, boxplots of /ptk/ and /sf/ strength by speaker are given in Figure 6.5.

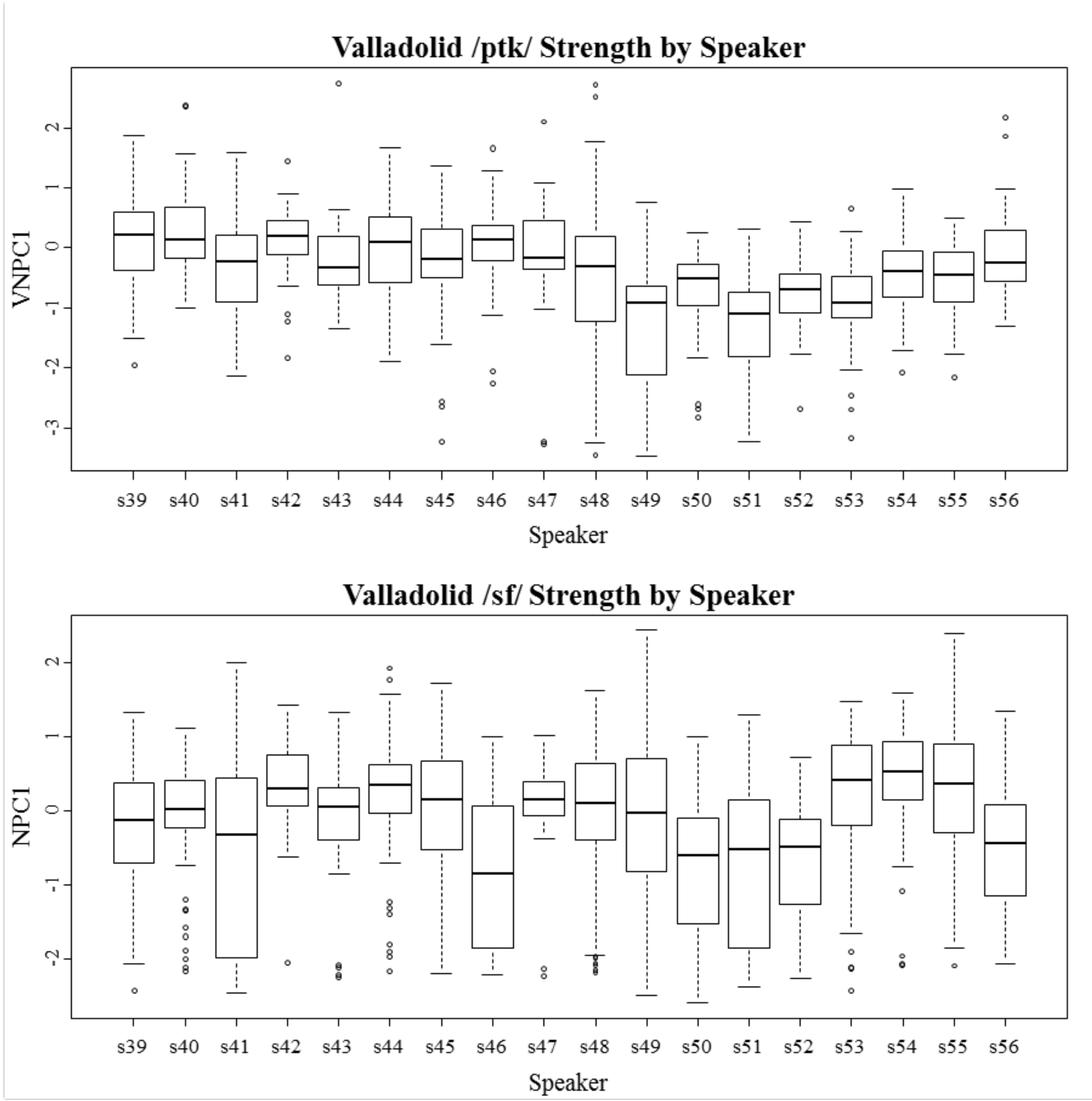

Figure 6.5 Boxplots of voiceless plosive strength (top panel) and voiceless fricative strength (bottom panel) by speaker.

I tested for a relationship between the Valladolid speakers' /ptk/ strength estimates and /sf/ strength estimates by using the estimates described in Sections 6.2 and 6.1. As the posterior samples are randomly shuffled when returned by Stan, and 10,000 posterior samples were obtained from each regression, they can be used jointly for Monte Carlo inference. For each posterior sample, I created a data frame with a column for average voiceless plosive strength (by averaging over the effect of place of articulation for each speakers’ estimates in the data frames described in Section 6.2) and a column for average voiceless fricative strength (by averaging over the effect of fricative phoneme identity for each speakers’ estimates in the data frames described in Section 6.1), and one row per Valladolid speaker (18 rows total). For each of these data frames, an ordinary least squares regression was run on /ptk/ strength with /sf/ strength as the predictor. The regression estimates and R^2 at each iteration were logged. The posterior mean for R^2 was .114 (posterior standard deviation .074; 95% credible interval [.011, .276]). The posterior distribution of the regression estimates is given in Table 6.5, and Figure 6.6 shows a scatterplot of the posterior mean of the individual estimates with the posterior mean regression line.

Table 6.5 Posterior distribution of the regression comparing Valladolid speakers' /ptk/ and /sf/ estimates.

Fixed Effect	Mean	SD	2.5%	97.5%	P(sign)	
Intercept (Corrected Mean)	-0.363	0.041	-0.442	-0.280	> .999	***
Fricative Strength	0.418	0.150	0.137	0.728	.998	***

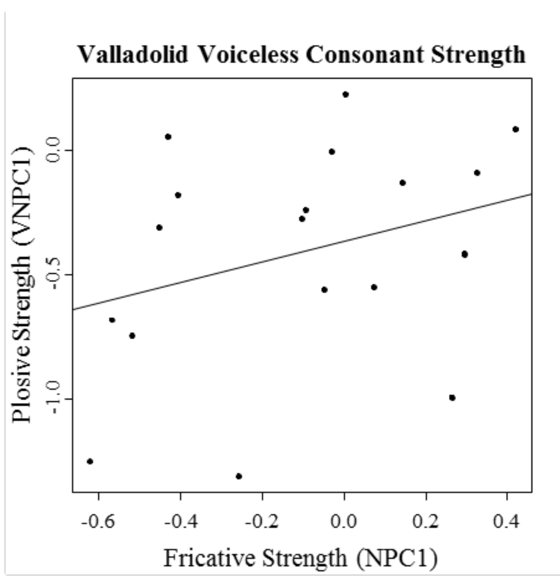

Figure 6.6 Scatterplot of /ptk/ strength speaker estimates by /sf/ strength speaker estimates with regression line.

There is strong evidence for a positive relationship between speakers’ /ptk/ strength and /sf/ strength (mean = 0.418, P(sign) = .998). Knowing the speakers’ mean /sf/ strengths, 11.4% of the variance same speakers’ mean /ptk/ strengths was able to be predicted. These results thus offer strong support for Hypothesis 5 (from Section 1.6: *The degree to which Valladolid speakers lenite voiceless fricatives /sf/ will correlate with the degree to which they lenite voiceless plosives /ptk/.*).

6.4. Strength of the correlations at the individual level

While there were no specific hypotheses related to the strength of the relationship at the individual level for each of the comparisons made in this chapter, it is interesting to note that we were able to explain the most variance when using speakers’ /ptk/ strengths to predict their /bdg/ strengths (60.2%), followed by when using speakers’ /f/ strengths to predict their /s/ strengths (40.0%), followed by when using speakers’ /sf/ strengths to predict their /ptk/ strengths (11.4%). This hierarchy of variance explained matches up with how similar to sets of consonants being compared is. For /ptk/ and /bdg/, conditioning on place of articulation, the consonants differ in constriction and voicing, but share the same general oral gestures. For /s/ and /f/, the consonants differ in place of articulation, but both require a glottal gesture that is voiceless but produces enough airflow to create turbulent noise. For /sf/ and /ptk/, the consonants only share in common the fact that they are voiceless obstruents; while they both involve a voiceless glottal gesture, these gestures are not the same (e.g. Stevens, 2000, pp. 379–380). Thus, the more similar the gestures were between the sets of consonants being compared, the stronger the relationship was.

Chapter 7. Discussion

7.1. Hypothesis evaluation

Hypothesis 1. Catalan /s/ will be stronger than (Valladolid) Spanish /s/, but Catalan /f/ will not be stronger than (Valladolid) Spanish /f/. This hypothesis is strongly supported by the data (Section 4.2). Catalan /s/ was found to be 0.335 standard deviations stronger than Spanish /s/, with strong evidence for the effect direction (P(sign) = .981). For /f/, the posterior mean for the difference was that Catalan /f/ was 0.036 standard deviations weaker than Spanish /f/, with little evidence for this effect direction (P(sign) = .592). This does not indicate that we have great certainty that there is no difference between Catalan and Spanish /f/ strength (the 95% credible interval for the difference is [-0.351, 0.270]), but rather that there is a large amount of uncertainty for the effect; however, we do have strong evidence that the difference between Catalan /s/ and Spanish /s/ is larger than the difference between Catalan /f/ and Spanish /f/ (mean 0.371; P(sign) = .986).

Hypothesis 2. In Spanish, for both the voiceless plosives /ptk/ and the voiced plosives /bdg/, the dialects will show the same strength hierarchy of Cuzco > Valladolid > Lima. This hypothesis is partially supported by the data (Section 5.3). There is strong evidence for this hierarchy for /bdg/, with all pairwise dialect comparisons having magnitude greater than or equal to 0.594 and P(sign) > .999. However, for /ptk/, while the posterior means for the pairwise comparisons support this strength hierarchy, we only have strong evidence for Cuzco having stronger /ptk/ than both Valladolid and Lima, with weak evidence for Valladolid having stronger /ptk/ than Lima. This may be due to the smaller sample size for Lima (only 8 speakers), but may also be due to too much overlap between the speakers of the two dialects. Additionally, all three contrast magnitudes are smaller for /ptk/ than for /bdg/.

Hypothesis 3. The degree to which individual speakers of Catalan and Spanish lenite /s/ will correlate with the degree to which the same speakers lenite /f/. This hypothesis is strongly supported by the data (Section 6.1). A 1-SD increase in a speaker's /f/ strength was found to correspond to a 0.648-SD increase in their /s/ strength (P(sign) > .999). There was little evidence that this effect was different for Spanish and Catalan speakers (P(sign) = .594), and overall, knowing the speakers' /f/ strengths, 40.0% of the variation in their /s/ strengths was able to be predicted.

Hypothesis 4. In Spanish, the degree to which individual speakers lenite /ptk/ will correlate with the degree to which the same speakers lenite /bdg/, and this relationship may differ by place of articulation. This hypothesis is strongly supported by the data (Section 6.2). Knowing speakers' voiceless plosive strengths, 60.2% of the variance in their voiced plosive strengths was able to be predicted. A 1-SD increase in /t/ strength corresponded to a 0.908-SD increase in /d/ strength (P(sign) > .999); a 1-SD increase in /p/ strength corresponded to a 0.781-SD increase in /b/ strength (P(sign) > .999); a 1-SD increase in /k/ strength corresponded to a 0.644-SD increase in /g/ strength; and there is strong evidence that these slopes differ in magnitude (P(sign) ≥ .993 for all pairwise comparisons).

Hypothesis 5. The degree to which Valladolid speakers lenite voiceless fricatives /sf/ will correlate with the degree to which they lenite voiceless plosives /ptk/. The hypothesis is strongly supported by the data (Section 6.3). A 1-SD increase in the Valladolid speakers' voiceless fricative strengths corresponded to a 0.418-SD increase in their voiceless plosive strengths, with 11.4% of the variance in their voiceless plosive strengths being able to be predicted.

7.2. Contrast preservation and constraints on individual phonetic variation

The results of the two experiments carried out in this dissertation strongly support the hypothesis that individual phonetic variation is constrained from a multivariate perspective. This occurred both when contrast preservation could arguably be acting as a constraint (Spanish /ptk/ and /bdg/ comparison), and when contrast preservation cannot be acting as a constraint (Spanish and Catalan /s/ and /f/ comparison, and Valladolid /ptk/ and /sf/ comparison). This lends substantial support to the view that contrast preservation can be an outcome. That is, since these correlations at the individual level lead to contrast preservation (i.e. they exist between Spanish /ptk/ and /bdg/ and the result is that speakers are not neutralizing), but also exist between consonants where there is no risk of neutralization, then in cases where neutralization is a possibility, a natural conclusion is that processes that are not motivated by contrast preservation as a constraint can produce it anyways.

However, the results of the fricative experiment also support a role for contrast preservation as a constraint. Catalan and Spanish both have /s/ and /f/, neither have /v/, and only Catalan has /z/. The strength of the relationship between individuals' /s/ strengths and /f/ strengths was not found to differ for the two languages, and there was little evidence for a population difference in /f/ strength for the languages, but there was strong evidence for a population difference in /s/ strength for the languages, and strong evidence that this difference was greater than the difference for /f/. That is, /s/ and /f/ are just as related in relative strength in the two languages, but on average Catalan /s/ is stronger. This supports a role for contrast preservation as a constraint, with the interpretation of the results presented here being that Catalan /s/ is stronger than Spanish /s/ because lenition of Catalan /s/ causes phonetic overlap with Catalan /z/, while in Spanish, /s/ lenition does not lead to overlap with any contrasting categories. In the case of the simultaneous lenition of Spanish /ptk/ and /bdg/, we cannot say to what extent contrast preservation is acting as a constraint, or whether contrast preservation is simply an outcome achieved through lenition of gestures common to the consonants; it could also be a bit of both. Both experiments presented in this dissertation can also serve to give us a window into how similar diachronic changes (or lack thereof) occurred.

The plosive experiment gives us an idea of what diachronic chain shifts may have looked like (e.g. the chain shift from Latin to Western Romance, whereby intervocalic geminate consonants simplified, non-geminate voiceless consonants voiced, and non-geminate voiced consonants spirantized; Penny (2002, pp. 74–84)). It may be that such changes occurred through individuals exhibiting constrained variation in lenition, with the entire system moving together due to the same underlying gestural lenitions. The plosive experiment results also suggest focusing on individual variation with respect to phonological change; while the correlation between /ptk/ and /bdg/ strength was robust at the individual level, the evidence was less clear when examining dialect population differences (i.e. language occurs first and foremost in individual speakers).

The fricative experiment offers a synchronic look at the phenomenon Wedel et al. (2013) studied from a diachronic perspective. The authors found that the higher the functional load of a contrast is, the less likely it was to be lost. If the role for contrast preservation as a constraint found in the fricative study is taken to be gradient, then we might expect a correspondence between functional load and the phonetic strength of consonants that contrast with a weaker counterpart (i.e. in the experiment I looked at *no contrast* vs. *contrast*, but we could conduct similar experiments where there is a contrast in both languages but the functional load is lower in one language than in the other, a possibility that should be explored in future research).

Overall, then, the results show the importance of approaching phonology as a complex system that occurs first and foremost within individuals. That is, when we examine a small subset of sounds and fail to explore how these sounds are related to each other and to other sets of sounds within the individual speaker, we are missing out on information that is crucial to our understanding of phonology. For the three within-individual correlations explored, the more similar the sets of sounds were, the more individual variance we were able to explain from a multivariate perspective: 60.2% variance explained for /ptk/ and /bdg/, 40.0% variance explained for /s/ and /f/, and 11.4% variance explained for /ptk/ and /sf/. The general picture then is that the variance in the production of individual phonemes that we see is indicative of some smaller subset of variance components that affect multiple phonemes simultaneously, with the relationships between the variance in individual phonemes' productions being determined by how related the phonemes are to one another articulatorily.

Within this framework, we should also leave room for a possible teleological role for contrast preservation (i.e. contrast preservation as a constraint). Hock (1991, pp. 164–166) defines a role for teleology in phonology that consists of repair strategies employed gradually over time in reaction to one phoneme encroaching on another phoneme acoustically (as may occur in a chain shift). However, he argues that contrast preservation merely acts as a repair strategy, and does not play a role in preventing a change from occurring in the first place (Hock, 1991, pp. 150–151). The synchronic results presented in this dissertation for Catalan and Spanish /s/ and /f/, and diachronic research by Wedel et al. (2013) who found functional load to play a gradient role in the probability that a contrast is lost, suggest that teleology may in fact have a role not only in terms of repair strategies when sounds do change, but also in terms of preventing sounds from changing in the first place.

While the data in the experiments carried out here show clear evidence of constraints on individual variation and support an intermediate approach to the issue of contrast preservation, they are of course only two datasets. Further research needs to be done, with more studies looking at acoustic correlations with large samples of speakers as was done here. Even more importantly, studies also need to look at these constraints on individual variation from an articulatory perspective. The acoustic data presented here show evidence that the variation we see in individuals' phoneme productions are the result of a smaller interacting subset of variance components, and I argue that these are very likely to be tied to articulatory gestures. Another interesting implication of the results presented in this dissertation is that a phonetic phenomenon that could in theory affect two phonemes, but where the phenomenon is a sociolinguistic marker for only one of the phonemes, we should expect the phenomenon should still be correlated at the individual level for the two phonemes, but the variation in the phoneme that is not a sociolinguistic marker should not be assigned any social meaning. As a concrete example, consider /ʒ/ and /s/ in Argentine Spanish. The /ʒ/ phoneme can be variably (and gradiently) realized as either [ʒ] or [ʃ], and this is a sociolinguistic marker (Rohena-Madrazo, 2013). We should expect speakers who produce /ʒ/ as relatively more voiceless to also produce /s/ as relatively more voiceless, but for the difference in /s/ voicing to not be noticed by speakers in the same way as differences in /ʒ/ voicing. This dissertation has thoroughly laid out widely applicable phonetic and statistical methodologies that can be used in such future research.

7.3. Measure automation and dimensionality reduction

A phonetic measure for the duration of intervocalic consonants based on the preceding and following vowels' intensity maxima was automated and employed to measure a wide range of consonants (plosives, fricatives, and approximants). In addition to this measure, two voicing measures (one duration-based and one percentage-based) and two relative intensity measures (on a difference and one a velocity) were taken. All of these measures had the advantage of relying on minimal amounts of manual segmentation, with only four boundaries being necessary (beginning and end of the VCV sequence, and boundaries containing the consonant's minima if not elided); all other durational, voicing, and intensity measures were then able to be automatically extracted from the acoustic signal.

Previous researchers have examined some subset of these measures individually when examining /ptk/, /bdg/, or /sf/. For example, Hualde et al. (2011) measured the percentage of /ptk/ closures that were voiced and classified them into two voicing categories (voiceless vs. partially or fully voiced), and compared three different intensity-based measurements and a duration measurement for /bdg/, voiceless /ptk/, and partially or fully voiced /ptk/, finding significant differences in each case, and specifically noting that even when /ptk/ are voiced, they are different in their constriction from /bdg/. As shown in this study, all five measures are highly correlated for plosives (Figure 2.7), and a principal component analysis shows that 81% of the variance in the five measures can be explained by a single variance component (Table 2.5 and Table 2.6) that very successfully separates underlyingly voiceless plosives from underlyingly voiced plosives (Figure 3.6 and Table 3.1). Similarly, for fricatives, the voicing and duration measures were found to be correlated (Figure 2.8), and a principal component analysis shows that 77% of the variance in the three measures can be explained by a single variance component (Table 2.7 and Table 2.8) that gives about equal weights to all three measures of strength.

Given that these measures are a simplification of the acoustic signal, that they share common articulatory origins, that listeners receive all of these inputs at once rather than separately, and that the majority of their variance can be explained by a single component, it seems more reasonable that the distinction between /ptk/ and /bdg/ in Spanish and the strength of /sf/ in Spanish and Catalan lies not in a single measure of consonant strength, or in several measures separately, but rather in the totality of the multivariate space they create. Using the PCA approach also provides a principled way to avoid inflating false discovery rates by running multiple regressions on highly correlated measures, which has become an increasing problem in many fields of research (Simmons et al., 2011). This dissertation thus makes a methodological contribution that can be employed in the measurement of consonant lenition, and in other areas of phonetic research where such high correlations exist as well.

7.4. Bayesian mixed effects regression and `NA` coding

The use of Bayesian mixed effects regression as opposed to a frequentist approach (as is more commonly used in linguistics) allowed for full distributional information about individual speakers to be obtained, contributing a novel use of these models in the analysis of linguistic data. The Bayesian approach also allowed the strength of evidence in the data for hypotheses to be evaluated directly, rather than taking a null hypothesis significance testing approach. The use of `NA` values to code unordered factors for observations in subsets of the data where the factor is not contrastive allowed for each dataset to be modeled in its entirety with a single regression, rather than binning the data and running multiple regressions. In this way, all of the information was able to be incorporated at once, providing a further methodological contribution that can be used in future linguistic research, especially in observational data where there tends to be a greater degree of imbalance.

7.5. Conclusion

Two phonetic experiments were carried out in this dissertation in order to test whether individual phonetic variation is constrained from a multivariate perspective, and to test how this relates to contrast preservation. In both experiments, multiple phonetic measures of consonant strength were automatically extracted from the acoustic signal, and dimensionality reduction was performed with PCA. A single component was found to be able to explain the vast majority of the variance in each case, and Bayesian mixed effects regressions were run on these components in order to obtain posterior distributions for population estimates and also for individual speakers.

In the first experiment, word-medial intervocalic /s/ and /f/ were compared for Valladolid Spanish and Barcelona Catalan. The relative strength of /s/ and /f/ were found to be correlated at the individual level in both languages, with no evidence that the magnitude of the correlation differed by language. While strong evidence was found for Catalan /s/ being stronger than Spanish /s/, and for this difference being much larger than the difference between Catalan /f/ and Spanish /f/, little evidence was found for a difference between Spanish and Catalan /f/.

In the second experiment, the simultaneous lenition of Spanish /ptk/ and /bdg/ in three dialects (Cuzco, Lima, and Valladolid) was examined. While there was very strong evidence for voiceless and voiced plosive lenition being correlated at the individual level (and for this relationship varying by place of articulation), the evidence for this correlation at the dialectal level was considerably weaker. Valladolid /ptk/ strength and /sf/ strength were further found to be correlated at the individual level.

Overall, the results support the hypothesis that individual phonetic variation is, to some extent, constrained. It was argued that this is likely due to the acoustic lenition of multiple segments sharing a common articulatory source of lenition, which is supported by the correlations existing both when neutralization is a possibility (i.e. /ptk/ vs. /bdg/) and when it is not a possibility (i.e. /s/ vs. /f/ and /ptk/ vs. /sf/). Due to presence of correlations in both cases, it was argued that the view that contrast preservation can be merely an outcome of other phonetic forces is strongly supported. However, based on the results of the fricative experiment, it was argued that within this larger framework (constraints on individual variation and contrast preservation as an outcome), there may also be an additional role for contrast preservation as a constraint in its own right.

Future research should look at other situations like the ones studied here from an acoustic perspective, as no single study can serve as the sole evidence for a hypothesis. The phonetic and statistical methods employed here should be used to study other chain shifts (including stable phonetic variation along a continuum where contrast is maintained) and differences in synchronic lenition based on a gradient measure of contrast (e.g. functional load, as opposed to here where *no contrast* was compared to *contrast*). The perceptual implications of constraints on individual phonetic variation should also be explored. Specifically, are speakers aware, on some level, of the correlation between these different segments? Can hearing the production of a consonant prime the listener for a specific degree of lenition in a seemingly unrelated consonant? Finally, studies should examine how the acoustic correlations found in this dissertation relate back to articulation.

References

Agresti, A. (2002). *Categorical Data Analysis* (2nd ed.). Wiley.

Baayen, R. H. (2008). *Analyzing linguistic data: a practical introduction to statistics using R*. Cambridge, UK ; New York: Cambridge University Press.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278.

Bates, D., Kliegl, R., Vasishth, S., & Baayen, R. H. (2015). Parsimonious Mixed Models. *ArXiv*. Retrieved from https://arxiv.org/abs/1506.04967

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

Berger, J. O. (1993). *Statistical decision theory and Bayesian analysis* (2nd ed). New York: Springer-Verlag.

Blecua, B., & Rost, A. (2013). Perception and Production of Spanish /f/: the role of the listener. *Poster Presented at Phonetics and Phonology in Iberia*.

Boersma, P., & Weenink, D. (2016). Praat: doing phonetics by computer (Version 6.0.21). Amsterdam. Retrieved from http://www.fon.hum.uva.nl/praat/

Boll, S. F. (1979). Suppression of acoustic noise in speech using spectral subtraction. *IEEE Transactions on ASSP*, *27*, 113–120.

Bresnan, J., Cueni, A., Nikitina, T., & Baayen, R. H. (2007). Predicting the dative alternation. *Cognitive Foundations of Interpretation*, 69–94.

Browman, C. P., & Goldstein, L. (1986). Towards an articulatory phonology. *Phonology Yearbook*, *3*(21), 9–252.

Browman, C. P., & Goldstein, L. (1992). Articulatory phonology: An overview. *Phonetica*, *49*(3–4), 155–180.

Bürkner, P.-C. (in press). brms: An R Package for Bayesian Multilevel Models using Stan. *Journal of Statistical Software*. Retrieved from https://CRAN.R-project.org/package=brms

Caflisch, R. E. (1998). Monte carlo and quasi-monte carlo methods. *Acta Numerica*, *7*, 1–49.

Caravedo, R. (1990). *Sociolingüística del español de Lima*. Pontificia Universidad Católica del Perú.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., … Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, *76*(1).

Carrasco, P. (2008). *An acoustic study of voiced stop allophony in Costa Rican Spanish*. University of Illinois at Urbana-Champaign.

Carrasco, P., Hualde, J. I., & Simonet, M. (2012). Dialectal Differences in Spanish Voiced Obstruent Allophony: Costa Rican versus Iberian Spanish. *Phonetica*, *69*(3), 149–179.

Carvalho, J. (2008). Western Romance. *Lenition and Fortition*, 207–233.

Cepeda, G. (1991). *Las consonantes de Valdivia*. Valdivia: Fondecyt, Universidad Austral de Chile.

Cohen, J. (1994). The Earth Is Round (p < .05). *American Psychologist*, *49*(12), 997–1003.

Colantoni, L., & Marinescu, I. (2010). The scope of stop weakening in Argentine Spanish. In *Selected proceedings of the 4th conference on laboratory approaches to Spanish phonology* (pp. 100–114).

Cole, J., Hualde, J. I., & Iskarous, K. (1999). Effects of prosodic and segmental context on /g/-lenition in Spanish. In *Proceedings of the fourth international linguistics and phonetics conference* (Vol. 2, pp. 575–589).

Cowles, M. K. (2013). *Applied Bayesian Statistics with R and OpenBUGS Examples*. Springer.

Davidson, J. (2015). *Social dynamics of Catalan-Spanish contact in the evolution of Catalonian Spanish*. University of Illinois at Urbana-Champaign.

Davidson, L. (2016). Variability in the implementation of voicing in American English obstruents. *Journal of Phonetics*, *54*, 35–50.

Davis, M. J. (2010). Contrast Coding in Multiple Regression Analysis: Strengths, Weaknesses, and Utility of Popular Coding Structures. *Journal of Data Science, 8,* 61–73.

De Jong, N. H., & Wempe, T. (2009). Praat script to detect syllable nuclei and measure speech rate automatically. *Behavior Research Methods*, *41*(2), 385–390.

Delforge, A. M. (2009). *The Rise and Fall of Unstressed Vowel Reduction in the Spanish of Cusco, Peru: A Sociophonetic Study.* ERIC.

Delforge, A. M. (2012). “Nobody wants to sound like a provinciano”: The recession of unstressed vowel devoicing in the Spanish of Cusco, Perú1. *Journal of Sociolinguistics, 16*(3), 311–335.

Demidenko, E. (2013). *Mixed models: theory and applications with R* (2nd ed.). Hoboken: Wiley.

Drager, K., & Hay, J. (2012). Exploiting random intercepts: Two case studies in sociophonetics. *Language Variation and Change*, *24*(01), 59–78.

Eager, C. D. (2015). Automated voicing analysis in Praat: Statistically equivalent to manual segmentation. In The Scottish Consortium for ICPhS 2015 (Ed.), *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow, UK: University of Glasgow.

Eager, C. D. (2017a). nauf: Regression with Not Applicable Unordered Factors in R (Version 0.1.0). Retrieved from https://github.com/CDEager/nauf

Eager, C. D. (2017b). standardize: Tools for Standardizing Variables for Regression in R (Version 0.1.0). Retrieved from https://github.com/CDEager/standardize

Eddington, D. (2011). What are the contextual phonetic variants of /b d g/ in colloquial Spanish? *Probus*, *23*(1), 1–19.

Escobar, A. (1977). Existe el castellano limeño? *Lexis*, *1*(1), 39–49.

Escobar, A. (1978). Variaciones sociolingüísticas del castellano en el Perú.

Escobar, A. M. (2011). Spanish in contact with Quechua. *The Handbook of Hispanic Sociolinguistics*, 321–352.

Figueroa, M. (2016). *Lenition in the production and perception of Chilean spanish approximant consonants: Implications for lexical access models*. University College London.

Flemming, E. S. (2002). *Auditory representations in phonology*. New York: Routledge.

Gabry, J., & Goodrich, B. (2016). rstanarm: Bayesian Applied Regression Modeling via Stan (Version 2.14.1). Retrieved from http://mc-stan.org

Gafos, A. I. (2002). A grammar of gestural coordination. *Natural Language & Linguistic Theory*, *20*(2), 269–337.

Gallagher, G. E. S. (2010). *The perceptual basis of long-distance laryngeal restrictions*. Massachusetts Institute of Technology.

Garrido, J. M., Escudero, D., Aguilar, L., Cardeñoso, V., Rodero, E., de-la-Mota, C., … Bonafonte, A. (2013). Glissando: a corpus for multidisciplinary prosodic studies in Spanish and Catalan. *Language Resources and Evaluation*, *47*(4), 945–971.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, *1*(3), 515–534.

Gelman, A., Hill, J., & Yajima, M. (2012). Why We (Usually) Don't Have to Worry About Multiple Comparisons. *Journal of Research on Educational Effectiveness*, *5*, 189–211.

Gelman, A., Jakulin, A., Pittau, M. G., & Su, Y.-S. (2008). A weakly informative default prior distribution for logistic and other regression models. *The Annals of Applied Statistics*, *2*(4), 1360–1383.

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, *7*, 457–472.

Gelman, A., & Shirley, K. (2011). Inference from Simulations and Monitoring Convergence. In S. Brooks (Ed.), *Handbook for Markov chain Monte Carlo* (pp. 163–174). Boca Raton: Taylor & Francis.

Gelman, A., & Tuerlinckx, F. (2000). Type S error rates for classical and Bayesian single and multiple comparison procedures. *Computational Statistics*, *15*, 373–390.

Geyer, C. J. (2011). Introduction to Markov Chain Monte Carlo. In S. Brooks (Ed.), *Handbook for Markov chain Monte Carlo* (pp. 3–48). Boca Raton: Taylor & Francis.

Gick, B., Wilson, I., Koch, K., & Cook, C. (2005). Language-specific articulatory settings: Evidence from inter-utterance rest position. *Phonetica*, *61*(4), 220–233.

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

Gordon, M. J. (2001). *Small-town values and big-city vowels: A study of the Northern Cities Shift in Michigan*. Duke University Press.

Gordon, M. J. (2013). Investigating Chain Shifts and Mergers. In J. K. Chambers & N. Schilling (Eds.), *The handbook of language variation and change* (pp. 203–219). John Wiley & Sons.

Guitart, J. M. (1977). Aspectos del consonantismo habanero: reexamen descriptivo. In *Boletín de la Academia Puertorriqueña de la Lengua Española 7*. ERIC.

Gurevich, N. (2004). *Lenition and contrast: the functional consequences of certain phonetically conditioned sound changes*. Psychology Press.

Halekoh, U., & Højsgaard, S. (2014). A Kenward-Roger Approximation and Parametric Bootstrap Methods for Tests in Linear Mixed Models: The R Package pbkrtest. *Journal of Statistical Software*, *59*(9), 1–30.

Hock, H. H. (1991). *Principles of historical linguistics*. New York: Walter de Gruyter.

Hoffman, M. D., & Gelman, A. (2014). The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, *15*(Apr), 1593–1623.

Hualde, J. I. (2005). *The sounds of Spanish*. Cambridge, UK ; New York: Cambridge University Press.

Hualde, J. I., & Eager, C. D. (2016). Final devoicing and deletion of /-d/ in Castilian Spanish. *Studies in Hispanic and Lusophone Linguistics*, *9*(2), 329–353.

Hualde, J. I., Eager, C. D., & Nadeu, M. (2015). Catalan voiced prepalatals: Effects of nonphonetic factors on phonetic variation? *Journal of the International Phonetic Association*, *45*(3), 243–267.

Hualde, J. I., Nadeu, M., & Simonet, M. (2010). Lenition and phonemic contrast in Majorcan Catalan. In *Romance Linguistics 2009: Selected papers from the 39th Linguistic Symposium on Romance Languages (LSRL)* (p. 63). Tucson, Arizona: John Benjamins Publishing.

Hualde, J. I., & Prieto, P. (2014). Lenition of intervocalic alveolar fricatives in Catalan and Spanish. *Phonetica*, *71*(2), 109–127.

Hualde, J. I., Shosted, R., & Scarpace, D. (2011). Acoustics and articulation of Spanish /d/ spirantization. In *Proceedings of the International Congress of Phonetic Sciences XVII* (pp. 906–909).

Hualde, J. I., Simonet, M., & Nadeu, M. (2011). Consonant lenition and phonological recategorization. *Laboratory Phonology*, *2*(2), 301–329.

Hualde, J. I., Simonet, M., Shosted, R., & Nadeu, M. (2010). Quantifying Iberian spirantization: Acoustics and articulation. In *40th Linguistic Symposium on Romance Languages, Seattle, WA* (pp. 26–28).

IEC. (2016). Diccionari de freqüències del diccionari descriptiu de la llengua catalana. Retrieved from http://dcc.iec.cat/ddlci/scripts/indexA.asp

Johnson, K. (2012). *Acoustic and auditory phonetics* (3rd ed). Malden, MA: Wiley-Blackwell.

Johnson, R. A., & Wichern, D. W. (2002). *Applied multivariate statistical analysis* (Vol. 5). Prentice hall Upper Saddle River, NJ.

Kimball, A., Shantz, K., Eager, C. D., & Roy, J. (2016). Beyond Maximal Random Effects for Logistic Regression: Moving Past Convergence Problems. *ArXiv*. Retrieved from https://arxiv.org/abs/1611.00083

Kirchner, R. (1996). Synchronic chain shifts in Optimality Theory. *Linguistic Inquiry*, *27*(2), 341–350.

Kroese, D. P., Taimre, T., & Zdravko, I. B. (2011). *Handbook of monte carlo methods*. Wiley.

Langstrof, C. (2006). Acoustic evidence for a push-chain shift in the intermediate period of New Zealand English. *Language Variation and Change*, *18*(02), 141–164.

Lavoie, L. M. (2001). *Consonant strength: phonological patterns and phonetic manifestations*. New York: Garland Pub.

Lehmann, E. L., & Casella, G. (1998). *Theory of point estimation* (2nd ed). New York: Springer.

Lenth, R. V., & Hervao, M. (2015). *lsmeans: Least-Squares Means*. Retrieved from http://CRAN.R-project.org/package=lsmeans

Lewandowski, D., Kurowicka, D., & Joe, H. (2009). Generating random correlation matrices based on vines and extended onion method. *Journal of Multivariate Analysis*, *100*, 1989–2001.

Lewis, A. (2000). Acoustic variability of intervocalic voiceless stop consonants in three Spanish dialects. In *Hispanic Linguistics at the Turn of the Millennium: Papers from the 3rd Hispanic Linguistics Symposium* (pp. 101–114).

Lewis, A. (2001). *Weakening of intervocalic /ptk/ in two Spanish dialects: Toward the quantification of lenition processes*. University of Illinois at Urbana-Champaign.

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

Lipski, J. M. (2011). Socio-phonological Variation in Latin American Spanish. In M. Diaz-Campos (Ed.), *The Handbook of Hispanic Sociolinguistics* (pp. 72–97). John Wiley & Sons.

Lubowicz, A. (2003). *Contrast preservation in phonological mappings*. Rutgers University.

Machuca-Ayuso, M. J. (1997). *Las obstruyentes no continuas del español: relación entre las categorías fonéticas y fonológicas en habla espontánea*. Publicacions de la Universitat Autònoma de Barcelona.

Martinet, A. (1952). Function, structure, and sound change. *Word*, *8*(1), 1–32.

Martínez-Celdrán, E. (2009). Sonorización de las oclusivas sordas en una hablante murciana: problemas que plantea. *Estudios de Fonética Experimental*, *18*, 254–271.

Munday, P. (2001). *An acoustic study of Spanish voiceless stops in three speaking styles*. New York University.

Nadeu, M. (2014). Stress- and speech rate-induced vowel quality variation in Catalan and Spanish. *Journal of Phonetics*, *46*, 1–22.

Navarro Tomás, T. (1977). *Manual de pronunciación española* (17th ed.). Madrid.

Neal, R. M. (2011). MCMC Using Hamiltonian Dynamics. In S. Brooks (Ed.), *Handbook for Markov chain Monte Carlo* (pp. 113–162). Boca Raton: Taylor & Francis.

Nielsen, K. Y. (2007). Implicit phonetic imitation is constrained by phonemic contrast. In *Proceedings of the 16th International Congress of the Phonetic Sciences*. Saarbrcken, Germany: Citeseer.

Nielsen, K. Y. (2008). *Word-level and feature-level effects in phonetic imitation*. University of California, Los Angeles.

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

Norris, J. R. (1997). *Markov Chains*. Cambridge University Press.

Oftedal, M. (1985). *Lenition in Celtic and in Insular Spanish: the secondary voicing of stops in Gran Canaria*. Oxford University Press.

Ohala, J. J. (1983). The origin of sound patterns in vocal tract constraints. In *The production of speech* (pp. 189–216). Springer.

Ohala, J. J., & Riordan, C. J. (1979). Passive vocal tract enlargement during voiced stops. *The Journal of the Acoustical Society of America*, *65*(1).

Ortega-Llebaria, M. (2004). Interplay between phonetic and inventory constraints in the degree of spirantization of voiced stops: Comparing intervocalic /b/ and intervocalic /g/ in Spanish and English. In T. L. Face (Ed.), *Laboratory approaches to Spanish phonology* (pp. 237–253). Berlin ; New York: Mouton de Gruyter.

Parker, S., & Weber, D. (1996). Glottalized and aspirated stops in Cuzco Quechua. *International Journal of American Linguistics*, 70–85.

Parrell, B. (2011). Dynamical account of how /b, d, g/ differ from /p, t, k/ in Spanish: Evidence from labials. *Laboratory Phonology*, *2*(2), 423–449.

Penny, R. (2002). *A History of the Spanish Language* (2 edition). Cambridge, UK; New York: Cambridge University Press.

Pérez, H. E. (2007). Estudio de la variación estilística de la serie/bdg/en posición intervocálica en el habla de los noticieros de la televisión chilena. *Estudios de Fonética Experimental*, *16*, 228–259.

Pérez-Silva, J. I., Palma, J. A., & Araujo, R. B. (2008). *Contra el prejuicio lingüístico de la motosidad: un estudio de las vocales del castellano andino desde la fonética acústica*. Pontificia Universidad Católica del Perú, Instituto Riva-Agüero.

Pierrehumbert, J. B. (2001). Exemplar dynamics: Word frequency, lenition and contrast. In B. Joan & H. Paul (Eds.), *Frequency and the emergence of linguistic structure* (pp. 137–157). Amsterdam: John Benjamins.

Poblete, M. T. (1992). La sonorización de las obstruyentes sordas/p, t, k/en el habla urbana de Valdivia. *Estudios Filológicos*, *27*, 73–94.

R Core Team. (2016). R: A language and environment for statistical computing (Version 3.3.2). Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/

Ramanarayanan, V., Byrd, D., Goldstein, L., Narayanan, S. S., Kobayashi, T., Hirose, K., & Nakamura, S. (2010). Investigating articulatory setting-pauses, ready position, and rest-using real-time MRI. In *INTERSPEECH* (pp. 1994–1997).

Real Academia Española. (2016). Corpus de referencia del español actual (CREA). *Banco de Datos*. Retrieved from http://www.rae.es

Recasens, D. (1993). *Fonètica i fonologia*. Enciclopèdia Catalana.

Recasens, D. (2015). The Effect of Contextual Consonants on Voiced Stop Lenition: Evidence from Catalan. *Language and Speech*, *59*(1), 139–161.

Rohena-Madrazo, M. (2013). Variación y cambio de sonoridad de la fricativa postalveolar del español de Buenos Aires. In L. Colantoni & C. Rodríguez-Louro (Eds.), *Perspectivas teóricas y experimentales sobre el español de la Argentina* (pp. 37–57). Iberoamericana Vervuert.

Saussure, F. de. (1986). *Course in General Linguistics*. (C. Bally, A. Sechehaye, & A. Riedlinger, Eds.). LaSalle, Ill: Open Court.

Scarpace, D., Beery, D., & Hualde, J. I. (2015). Allophony of /ʝ/ in Peninsular Spanish. *Phonetica*, *72*(2–3), 76–97.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*.

Singmann, H., Bolker, B., & Westfall, J. (2015). *afex: Analysis of Factorial Experiments*. Retrieved from http://CRAN.R-project.org/package=afex

Soler, A., & Romero, J. (1999). The role of duration in stop lenition in Spanish. In *Proceedings of the 14th International Congress of Phonetic Sciences* (pp. 483–486). San Francisco.

Stan Development Team. (2016a). RStan: the R interface to Stan (Version 2.14.1). Retrieved from http://mc-stan.org

Stan Development Team. (2016b). Stan Modeling Language Users Guide and Reference Manual (Version 2.14.0). Retrieved from http://mc-stan.org

Stevens, K. N. (2000). *Acoustic Phonetics*. Cambridge, Mass.: The MIT Press.

Torreblanca, M. (1976). La sonorización de las oclusivas sordas en el habla toledana. *Boletín de La Real Academia Española*, *56*(207), 117–146.

Torreira, F., & Ernestus, M. (2011). Realization of voiceless stops and vowels in conversational French and Spanish. *Laboratory Phonology*, *2*(2), 331–353.

Torreira, F., & Ernestus, M. (2012). Weakening of intervocalic /s/ in the Nijmegen Corpus of Casual Spanish. *Phonetica*, *69*(3), 124–148.

Trujillo, J. R. (1980). Sonorización de sordas en Canarias. *Anuario de Letras: Lingüística Y Filología*, (18), 247–254.

Turk, A., Nakai, S., & Sugahara, M. (2006). Acoustic segment durations in prosodic research: A practical guide. *Methods in Empirical Prosody Research*, *3*, 1–28.

Vogel, A. P., Maruff, P., Snyder, P. J., & Mundt, J. C. (2009). Standardization of pitch-range settings in voice acoustic analysis. *Behavior Research Methods*, *41*(2), 318–324.

Wedel, A., Kaplan, A., & Jackson, S. (2013). High functional load inhibits phonological contrast loss: A corpus study. *Cognition*, *128*(2), 179–186.

Weisberg, S. (2005). *Applied linear regression* (3rd ed.). Hoboken, NJ: Wiley.

Wheeler, M. (2005). *The phonology of Catalan*. New York: Oxford University Press.

Williams, L. (1987). *Aspectos sociolingüísticos del habla de la ciudad de Valladolid*. Valladolid: Universidad de Valladolid.

Wilson, I., & Gick, B. (2014). Bilinguals use language-specific articulatory settings. *Journal of Speech, Language, and Hearing Research*, *57*(2), 361–373.

Appendix A. Fricative experiment supplement

A.1. Regression variance components

The posterior mean, standard deviation, median, 95% credible interval, and effective sample size of each variance component are provided in Table A.1 (speaker effects) and Table A.2 (residuals).

Table A.1 Posterior distribution of the random effect standard deviations for speaker.

Random Speaker Effect	Mean	SD	Median	2.5%	97.5%	Neff
Intercept (Corrected Mean)	0.363	0.056	0.359	0.268	0.487	3760
Fricative, /f/	0.199	0.037	0.197	0.133	0.279	4675
Stress, Post-Tonic	0.132	0.044	0.132	0.042	0.218	1367
Stress, Tonic	0.087	0.047	0.086	0.007	0.183	2661
Preceding Vowel, High	0.044	0.032	0.039	0.002	0.118	3430
Following Vowel, High	0.090	0.033	0.091	0.018	0.155	1717
Log Word Frequency	0.104	0.035	0.104	0.032	0.175	3286
Speech Rate	0.044	0.028	0.041	0.002	0.106	3523

Table A.2 Posterior distribution of the residual standard error.

Mean	SD	Median	2.5%	97.5%	Neff
0.834	0.013	0.834	0.808	0.860	10000

A.2. Contrasts for group means

The following list gives the contrasts applied to the fixed effects coefficient vector at each iteration of the Bayesian mixed effects regression to obtain the group mean estimates summarized in Sections 4.1 and 4.2. Contrast estimates were then obtained by simply subtracting the estimate vectors from one another.

Language, Catalan : Stress, Tonic = Intercept (Corrected Mean) + Language, Catalan + Stress, Tonic + Language, Catalan : Stress, Tonic

Language, Catalan : Stress, Post-Tonic = Intercept (Corrected Mean) + Language, Catalan + Stress, Post-Tonic + Language, Catalan : Stress, Post-Tonic

Language, Catalan : Stress, Unstressed = Intercept (Corrected Mean) + Language, Catalan - Stress, Tonic - Stress, Post-Tonic - Language, Catalan : Stress, Tonic - Language, Catalan : Stress, Post-Tonic

Language, Spanish : Stress, Tonic = Intercept (Corrected Mean) - Language, Catalan + Stress, Tonic - Language, Catalan : Stress, Tonic

Language, Spanish : Stress, Post-Tonic = Intercept (Corrected Mean) - Language, Catalan + Stress, Post-Tonic - Language, Catalan : Stress, Post-Tonic

Language, Spanish : Stress, Unstressed = Intercept (Corrected Mean) - Language, Catalan - Stress, Tonic - Stress, Post-Tonic + Language, Catalan : Stress, Tonic + Language, Catalan : Stress, Post-Tonic

Preceding Vowel, High = Intercept (Corrected Mean) + Preceding Vowel, High

Preceding Vowel, Non-High = Intercept (Corrected Mean) - Preceding Vowel, High

Following Vowel, High = Intercept (Corrected Mean) + Following Vowel, High

Following Vowel, Non-High = Intercept (Corrected Mean) - Following Vowel, High

Sex, Female = Intercept (Corrected Mean) + Sex, Female

Sex, Male = Intercept (Corrected Mean) - Sex, Female

Language, Catalan : Fricative, /f/ = Intercept (Corrected Mean) + Language, Catalan + Fricative, /f/ + Language, Catalan : Fricative, /f/

Language, Catalan : Fricative, /s/ = Intercept (Corrected Mean) + Language, Catalan - Fricative, /f/ - Language, Catalan : Fricative, /f/

Language, Spanish : Fricative, /f/ = Intercept (Corrected Mean) - Language, Catalan + Fricative, /f/ - Language, Catalan : Fricative, /f/

Language, Spanish : Fricative, /s/ = Intercept (Corrected Mean) - Language, Catalan - Fricative, /f/ + Language, Catalan : Fricative, /f/

A.3. Stan code

```
functions {
  matrix vec_to_mat_by_row(int R, int C, vector v){
    matrix[R,C] m;
    for(r in 1:R) m[r] = v[(C*(r-1)+1):(C*r)]';
    return m;
  }
}

data {
  int<lower=0> N;  // number of observations
  int<lower=0> K;  // number of coefficients

  int<lower=0> nz;  // num non-zero elements in model matrix
  vector[nz] w;  // non-zero elements in model matrix
  int<lower=0> v[nz];  // column indices for w
  int<lower=0> u[N+1];  // row-start indices for non-zero elements

  vector[N] y;  // scaled response

  int<lower=0> P;  // number of fixed effects
  int<lower=0> G;  // number of random effect groups
  int<lower=0> cindx[G,2];  // coefficient index for random effects
  int<lower=0> M_1;  // number of speaker members
  int<lower=0> Q_1;  // number of speaker effects per member

  // (hyper) priors
  real<lower=0> scale_beta;  // prior scale for betas
  real<lower=0> nu_beta;  // degrees of freedom for beta t-dist prior
  real<lower=0> sc_q0;  // prior scale for random intercept sds
  real<lower=0> sc_qs;  // prior scale for random slope sds
  real<lower=0> eta_q;  // shape for LKJ prior on random effects correlations
  real<lower=0> sc_res;  // prior scale for sd of the residuals
}

parameters {
  // all parameters sampled on unit scale or with cholesky factors
  // (as applicable) and reparameterized

  vector[P] beta_raw;

  matrix[Q_1,M_1] gamma_1_raw;
  vector<lower=0>[Q_1] sigma_1_raw;
  cholesky_factor_corr[Q_1] omega_1_raw;

  real<lower=0> sigma_res_raw;
}

transformed parameters {
  vector<lower=0>[Q_1] sigma_1;  // sd in the speaker effects
  real<lower=0> sigma_res;  // sd of the residuals

  vector[K] coef;  // all coefficients
  vector[N] y_hat;  // fitted values
```

```
  coef[1:P] = scale_beta * beta_raw;

  sigma_1[1] = sc_q0 * sigma_1_raw[1];
  sigma_1[2:Q_1] = sc_qs * sigma_1_raw[2:Q_1];
  coef[cindx[1,1]:cindx[1,2]]
    = to_vector(rep_matrix(sigma_1,M_1)
      .* (omega_1_raw * gamma_1_raw));

  sigma_res = sc_res * sigma_res_raw;

  y_hat = csr_matrix_times_vector(N,K,w,v,u,coef);
}

model {
  beta_raw ~ student_t(nu_beta,0,1);

  to_vector(gamma_1_raw) ~ normal(0,1);
  sigma_1_raw ~ normal(0,1);
  omega_1_raw ~ lkj_corr_cholesky(eta_q);

  sigma_res_raw ~ normal(0,1);
  y ~ normal(y_hat,sigma_res);
}

generated quantities {
  vector[N] log_lik;  // log-likelihod
  vector[P] beta;  // fixed effects
  matrix[M_1,Q_1] gamma_1;  // speaker effects
  matrix[Q_1,Q_1] omega_1;  // correlation in the speaker effects

  for(n in 1:N) log_lik[n] = normal_lpdf(y[n] | y_hat[n],sigma_res);
  beta = coef[1:P];
  gamma_1 = vec_to_mat_by_row(M_1,Q_1,coef[cindx[1,1]:cindx[1,2]]);
  omega_1 = tcrossprod(omega_1_raw);
}
```

Appendix B. Plosive experiment supplement

B.1. Demographic questionnaire and Cuzco participant demographics

The participants from Cuzco and Lima filled out a questionnaire with the following questions, with classification of the Cuzco participants along with occupation given in Table B.1:

1. ¿Cuántos años tiene usted?
2. ¿Cuál es su sexo?
3. ¿En qué trabaja usted?
4. ¿Cuál es el nivel de educación más alto que usted ha cumplido?
5. ¿En qué ciudad nació usted?
6. ¿Ha vivido usted en una ciudad diferente de la ciudad en que vive actualmente? ¿Cuántos años tenía y por cuánto tiempo vivió allí?
7. ¿Cuántos años tenía usted cuando se mudó a la ciudad en que vive actualmente?
8. ¿Qué lenguas habla usted? ¿Cuántos años tenía usted cuando empezó a aprender cada una y dónde las aprendió (en la escuela, en casa, etc.)?
9. ¿Qué lenguas habla usted en casa y con qué frecuencia?
10. ¿Qué lenguas habla usted en el trabajo y con qué frecuencia?
11. ¿Qué lenguas se hablaban en su casa antes de que empezara usted a asistir a la escuela?
12. ¿Cuántos años tenía usted cuando empezó a asistir a la escuela? ¿Dónde asistió a la escuela? ¿Qué lengua(s) se usaban en la escuela?
13. ¿Qué lenguas hablaban sus padres?
14. ¿Qué lenguas hablaban sus abuelos?

Table B.1 Cuzco participant demographics.

Speaker	Age	Quechua Bilingual	Age Group	Education Level	Sex	Occupation
s01	48	No	Older	University	F	Teacher
s02	30	No	Younger	University	F	Secretary
s03	37	Yes	Younger	Secondary	F	Gardener
s04	57	Yes	Older	University	M	Teacher
s05	37	No	Younger	Secondary	F	Tourism
s06	25	No	Younger	University	F	Tourism
s07	29	No	Younger	University	M	Teacher
s08	21	Yes	Younger	University	F	Teacher
s09	50	Yes	Older	University	F	Teacher
s10	34	No	Younger	University	M	House Cleaning
s11	32	Yes	Younger	University	M	Tourism
s12	48	Yes	Older	University	F	Small Business
s13	42	No	Older	University	F	Street Vendor
s14	20	Yes	Younger	University	M	Cook
s15	64	No	Older	University	M	Doctor
s16	65	Yes	Older	Secondary	M	Electrician
s17	44	Yes	Older	Secondary	F	Small Business
s18	46	No	Older	University	M	Small Business
s19	20	No	Younger	Secondary	F	Cook
s20	27	No	Younger	Secondary	M	Bank
s21	53	No	Older	Secondary	F	Street Vendor
s22	30	Yes	Younger	Secondary	M	Architecture
s23	28	No	Younger	Secondary	M	Cook
s24	19	Yes	Younger	University	F	Student
s25	30	Yes	Younger	Secondary	M	Tourism
s26	54	No	Older	Secondary	F	Street Vendor
s27	64	Yes	Older	Secondary	M	Small Business
s28	25	Yes	Younger	Secondary	F	Cook
s29	65	No	Older	Secondary	M	Small Business
s30	47	No	Older	Secondary	M	Street Vendor

B.2. Read speech task sentences

The following is the full list of sentences that the Cuzco and Lima participants read aloud in the read speech task, with planned observations analyzed in the Spanish experiment bolded and underlined (there are additional planned observations that occur post-nasal, post-lateral, and in coda, but these are not reported on here and are not bolded and underlined below).

1. El hombre calvo que vende mangos y mandarinas está en el rincón de la plaza.
2. La pe**t**ición no fue aceptada por el gobierno.
3. Las alpa**c**as no te a**t**a**c**arán a menos que las provoques.
4. No se pue**d**e comparar dos universidades tan diferentes.
5. Si quieres hacer un pastel, primero hay que mezclar el azú**c**ar con la mante**qu**illa en un bol y combinar el resto de los ingredientes se**c**os en un bol se**p**arado.
6. El despacho del a**b**o**g**a**d**o es difícil de encontrar porque está en una calle demasiado angosta y angular para carros.
7. El doctor administró la me**d**icina al paciente.
8. El discurso polí**t**ico arengará al partido para la elección en agosto.
9. No es necesario que le pongas sal a una so**p**a buena.
10. Mi hija es muy a**v**enturera, pero también pa**t**osa. Cada día le di**g**o que tenga cuidado.
11. Es absurdo mentir y prolongar lo ine**v**i**t**able. Álvaro, admi**t**e que tienes la culpa.
12. No me gusta el langostino.
13. El a**d**olescente tonto se rompió el to**b**illo montando en bicicleta.
14. “El mundo anti**g**uo” es un término usado para referirse a Euro**p**a.
15. Se**g**ún el alcalde, una erupción volcánica destruirá la aldea.
16. Los ingredientes princi**p**ales del gua**c**amole son paltas, tomates y cebollas.
17. Marisol colgará una alda**b**a en la puerta.
18. El go**b**ierno a veces su**b**e los impuestos para dar a**p**oyo a la soldadesca durante una guerra.
19. Es importante que cambien el ambiente caldeado de la oficina aunque sea difícil.
20. El chico simpático silbará mientras va al banco donde sa**c**ará cinco soles.
21. No le gusta el ju**g**o pulposo.
22. Le encanta pasar tiempo cabalgando en el campo.
23. No pude ir a la facultad porque había un tronco en el sendero.
24. Por favor, no olvi**d**e hacer las albóndi**g**as.
25. Juan contará el mi**t**o del genio malvado atra**p**ado en la lámpara encantada.
26. El lenguaje tiene tendencia a cambiar.
27. En la milpa los campesinos culti**v**an maíz.
28. Cui**d**ado que no vuelques el florero de ámbar.
29. El hombre alto alquila el a**p**artamento con el balcón.
30. No absolverán al hombre culpable de vender armas ilegales.
31. La semana que viene faltaré tres días de clase.
32. El ejército inva**d**ió el país.
33. Hubo un golpe de estado.
34. El estudiante a**g**arró el recibo que le enviaron e hizo un calco para sus archivos.
35. El atleta que hace alpinismo saltará sobre la piedra.
36. Mi hermana es banquera, pero no le gusta su trabajo.
37. El conductor se disculpará por la demora.
38. El álbum era po**p**ular entre la gente culta.
39. El albañil se lastimó la espalda mientras hacía un peldaño. Tuvimos que llamar una ambulancia.

40. Ana no comulgará con las ideas de Pedro.
41. No olvi**d**aré pa**g**ar el alquiler. Lo pa**g**aré cuando pueda.
42. El agua del río salpi**c**ó a la falda de algo**d**ón y ahora se ha reducido.
43. Ese perro está tratando de qui**t**arse las pulgas.
44. Se amoldará al sueldo de su empleo nuevo.
45. El sindi**c**ato y el gobierno firmaron varios acuerdos, pero algunos no se han implementado
46. todavía.
47. El cálculo es difícil porque el ámbito ma**t**emático es abstracto.
48. Huelga que me lo digas. Necesito alterar el documento.
49. La cerveza tiene menos alcohol que el vino.
50. El niño delgado brincará por la casa.
51. Los dos e**v**entos fueron simultáneos.
52. Pon las alcachofas en la sartén y tá**p**ala para que no salga el vapor.

B.3. Regression variance components

The posterior mean, standard deviation, median, 95% credible interval, and effective sample size of each variance component are provided in Table B.2 (speaker effects), Table B.3 (item effects), and Table B.4 (residuals).

Table B.2 Posterior distribution of the random effect standard deviations for speaker.

Random Speaker Effect	Mean	SD	Median	2.5%	97.5%	Neff
Intercept (Corrected Mean)	0.304	0.035	0.301	0.244	0.381	4484
Voicing, Voiced	0.187	0.023	0.185	0.147	0.235	5245
Place, Biliabial	0.026	0.018	0.023	0.001	0.068	4407
Place, Dental	0.105	0.017	0.105	0.074	0.142	6739
Stress, Post-Tonic	0.031	0.021	0.029	0.001	0.076	4298
Stress, Tonic	0.075	0.018	0.075	0.038	0.111	4302
Word Position, Initial	0.092	0.021	0.091	0.053	0.133	5390
Preceding Vowel, High	0.065	0.017	0.065	0.030	0.098	4026
Following Vowel, High	0.037	0.021	0.037	0.002	0.078	3257
Log Word Frequency	0.027	0.016	0.026	0.001	0.061	4138
Speech Rate	0.023	0.015	0.021	0.001	0.056	3433
Task, Read Speech	0.107	0.019	0.106	0.074	0.148	7412
Voicing, Voiced : Place, Biliabial	0.031	0.021	0.028	0.001	0.075	4226
Voicing, Voiced : Place, Dental	0.111	0.019	0.110	0.077	0.150	7041
Voicing, Voiced : Word Position, Initial	0.061	0.022	0.062	0.014	0.102	4109
Voicing, Voiced : Preceding Vowel, High	0.015	0.012	0.013	0.001	0.043	5943
Voicing, Voiced : Following Vowel, High	0.028	0.017	0.026	0.001	0.064	3922
Voicing, Voiced : Speech Rate	0.027	0.016	0.026	0.001	0.061	3856

Table B.3 Posterior distribution of the random effect standard deviations for item.

Random Item Effect	Mean	SD	Median	2.5%	97.5%	Neff
Intercept (Corrected Mean)	0.212	0.032	0.210	0.154	0.280	5189
Speech Rate	0.115	0.029	0.115	0.060	0.173	3549
Dialect, Cuzco (Item Coding)	0.104	0.030	0.104	0.044	0.164	3457
Sex, Female	0.021	0.016	0.018	0.001	0.058	7572
Age Group, Older	0.031	0.022	0.027	0.001	0.080	6044
Education Level, Secondary	0.029	0.022	0.025	0.001	0.080	5063
Quechua Bilingual, Yes	0.034	0.024	0.031	0.001	0.089	4662
Dialect, Cuzco (Item Coding) : Sex, Female	0.035	0.023	0.033	0.002	0.084	4836

Table B.4 Posterior distribution of the residual standard error.

Mean	SD	Median	2.5%	97.5%	Neff
0.673	0.007	0.673	0.660	0.687	10000

B.4. Additional descriptive statistics

The following tables give descriptive statistics by underlying voicing for control factors that showed no evidence of an interaction with underlying voicing, and were therefore included only as fixed effects in the regression on VNPC1.

Table B.5 Descriptive statistics for VNPC1 by underlying voicing and stress.

Voicing	Stress	N	Minimum	Median	Maximum	Mean	SD
Voiced	Tonic	1053	-1.481	0.161	4.115	0.202	1.057
	Post-Tonic	861	-1.481	-0.103	3.922	-0.091	0.947
	Unstressed	780	-1.481	-0.181	3.718	-0.173	0.929
Voiceless	Tonic	954	-3.232	0.278	3.592	0.244	0.998
	Post-Tonic	768	-3.462	0.075	3.238	-0.059	0.987
	Unstressed	865	-3.528	-0.112	2.979	-0.218	0.955

Table B.6 Descriptive statistics for VNPC1 by underlying voicing and task.

Voicing	Task	N	Minimum	Median	Maximum	Mean	SD
Voiced	Read Speech	1009	-1.481	0.301	3.922	0.323	1.019
	Spontaneous Speech	645	-1.481	-0.143	3.955	-0.132	0.977
Voiceless	Read Speech	1010	-3.034	0.396	3.592	0.359	0.888
	Spontaneous Speech	426	-3.528	-0.021	2.330	-0.117	1.050

Table B.7 Descriptive statistics for VNPC1 by underlying voicing and age group.

Voicing	Age Group	N	Minimum	Median	Maximum	Mean	SD
Voiced	Older	694	-1.481	0.451	4.115	0.508	0.935
	Younger	763	-1.481	0.350	3.922	0.405	0.955
Voiceless	Older	668	-2.690	0.487	3.592	0.423	0.932
	Younger	747	-2.548	0.317	3.238	0.340	0.712

Table B.8 Descriptive statistics for VNPC1 by underlying voicing and education level.

Voicing	Education Level	N	Minimum	Median	Maximum	Mean	SD
Voiced	Secondary	741	-1.481	0.416	3.869	0.475	0.985
	University	716	-1.481	0.390	4.115	0.433	0.905
Voiceless	Secondary	704	-2.636	0.421	3.124	0.389	0.870
	University	711	-2.690	0.357	3.592	0.369	0.776

B.5. Contrasts for group means

The following list gives the contrasts applied to the fixed effects coefficient vector at each iteration of the Bayesian mixed effects regression to obtain the group mean estimates summarized in Sections 5.1, 5.2, and 5.3. Contrast estimates were then obtained by simply subtracting the estimate vectors from one another.

Voicing, Voiced : Speech Rate = Speech Rate + Voicing, Voiced : Speech Rate

Voicing, Voiceless : Speech Rate = Speech Rate - Voicing, Voiced : Speech Rate

Stress, Tonic = Intercept (Corrected Mean) + Stress, Tonic

Stress, Post-Tonic = Intercept (Corrected Mean) + Stress, Post-Tonic

Stress, Unstressed = Intercept (Corrected Mean) - Stress, Tonic - Stress, Post-Tonic

Voicing, Voiced : Word Position, Initial = Intercept (Corrected Mean) - 2/3 * Task, Read Speech + Voicing, Voiced + Word Position, Initial + Voicing, Voiced : Word Position, Initial

Voicing, Voiced : Word Position, Medial = Intercept (Corrected Mean) - 2/3 * Task, Read Speech + Voicing, Voiced - Word Position, Initial - Voicing, Voiced : Word Position, Initial

Voicing, Voiceless : Word Position, Initial = Intercept (Corrected Mean) - 2/3 * Task, Read Speech - Voicing, Voiced + Word Position, Initial - Voicing, Voiced : Word Position, Initial

Voicing, Voiceless : Word Position, Medial = Intercept (Corrected Mean) - 2/3 * Task, Read Speech - Voicing, Voiced - Word Position, Initial + Voicing, Voiced : Word Position, Initial

Preceding Vowel, High = Intercept (Corrected Mean) + Preceding Vowel, High

Preceding Vowel, Non-High = Intercept (Corrected Mean) - Preceding Vowel, High

Voicing, Voiced : Following Vowel, High = Intercept (Corrected Mean) + Voicing, Voiced + Following Vowel, High + Voicing, Voiced : Following Vowel, High

Voicing, Voiced : Following Vowel, Non-High = Intercept (Corrected Mean) + Voicing, Voiced - Following Vowel, High - Voicing, Voiced : Following Vowel, High

Voicing, Voiceless : Following Vowel, High = Intercept (Corrected Mean) - Voicing, Voiced + Following Vowel, High - Voicing, Voiced : Following Vowel, High

Voicing, Voiceless : Following Vowel, Non-High = Intercept (Corrected Mean) - Voicing, Voiced - Following Vowel, High + Voicing, Voiced : Following Vowel, High

Task, Read Speech = Intercept (Corrected Mean) + 1/2 * Dialect, Cuzco + 1/2 * Dialect, Lima + Task, Read Speech

Task, Spontaneous Speech = Intercept (Corrected Mean) + 1/2 * Dialect, Cuzco + 1/2 * Dialect, Lima - Task, Read Speech

Sex, Female : Dialect, Cuzco = Intercept (Corrected Mean) + Sex, Female + Dialect, Cuzco + Sex, Female : Dialect, Cuzco

Sex, Female : Dialect, Lima = Intercept (Corrected Mean) + Sex, Female + Dialect, Lima + Sex, Female : Dialect, Lima

Sex, Female : Dialect, Valladolid = Intercept (Corrected Mean) + Sex, Female - Dialect, Cuzco - Dialect, Lima - Sex, Female : Dialect, Cuzco - Sex, Female : Dialect, Lima

Sex, Male : Dialect, Cuzco = Intercept (Corrected Mean) - Sex, Female + Dialect, Cuzco - Sex, Female : Dialect, Cuzco

Sex, Male : Dialect, Lima = Intercept (Corrected Mean) - Sex, Female + Dialect, Lima - Sex, Female : Dialect, Lima

Sex, Male : Dialect, Valladolid = Intercept (Corrected Mean) - Sex, Female - Dialect, Cuzco - Dialect, Lima + Sex, Female : Dialect, Cuzco + Sex, Female : Dialect, Lima

/b/ (Voicing, Voiced : Place, Bilabial) = Intercept (Corrected Mean) + Voicing, Voiced + Place, Bilabial + Voicing, Voiced : Place, Bilabial

/d/ (Voicing, Voiced : Place, Dental) = Intercept (Corrected Mean) + Voicing, Voiced + Place, Dental + Voicing, Voiced : Place, Dental

/g/ (Voicing, Voiced : Place, Velar) = Intercept (Corrected Mean) + Voicing, Voiced - Place, Bilabial - Place, Dental - Voicing, Voiced : Place, Bilabial - Voicing, Voiced : Place, Dental

/p/ (Voicing, Voiceless : Place, Bilabial) = Intercept (Corrected Mean) - Voicing, Voiced + Place, Bilabial - Voicing, Voiced : Place, Bilabial

/t/ (Voicing, Voiceless : Place, Dental) = Intercept (Corrected Mean) - Voicing, Voiced + Place, Dental - Voicing, Voiced : Place, Dental

/k/ (Voicing, Voiceless : Place, Velar) = Intercept (Corrected Mean) - Voicing, Voiced - Place, Bilabial - Place, Dental + Voicing, Voiced : Place, Bilabial + Voicing, Voiced : Place, Dental

Voicing, Voiced : Dialect, Cuzco = Intercept (Corrected Mean) + Voicing, Voiced + Dialect, Cuzco - Age Group, Older - Education Level, Secondary - Quechua Bilingual, Yes + Voicing, Voiced : Dialect, Cuzco - Voicing, Voiced : Quechua Bilingual, Yes

Voicing, Voiced : Dialect, Lima = Intercept (Corrected Mean) + Voicing, Voiced + Dialect, Lima + Voicing, Voiced : Dialect, Lima

Voicing, Voiced : Dialect, Valladolid = Intercept (Corrected Mean) + Voicing, Voiced - Dialect, Cuzco - Dialect, Lima - Voicing, Voiced : Dialect, Cuzco - Voicing, Voiced : Dialect, Lima

Voicing, Voiceless : Dialect, Cuzco = Intercept (Corrected Mean) - Voicing, Voiced + Dialect, Cuzco - Age Group, Older - Education Level, Secondary - Quechua Bilingual, Yes - Voicing, Voiced : Dialect, Cuzco + Voicing, Voiced : Quechua Bilingual, Yes

Voicing, Voiceless : Dialect, Lima = Intercept (Corrected Mean) - Voicing, Voiced + Dialect, Lima - Voicing, Voiced : Dialect, Lima

Voicing, Voiceles : Dialect, Valladolid = Intercept (Corrected Mean) - Voicing, Voiced - Dialect, Cuzco - Dialect, Lima + Voicing, Voiced : Dialect, Cuzco + Voicing, Voiced : Dialect, Lima

B.6. Stan code

```
functions {
  matrix vec_to_mat_by_row(int R, int C, vector v){
    matrix[R,C] m;
    for(r in 1:R) m[r] = v[(C*(r-1)+1):(C*r)]';
    return m;
  }
}

data {
  int<lower=0> N;  // number of observations
  int<lower=0> K;  // number of coefficients

  int<lower=0> nz;  // num non-zero elements in model matrix
  vector[nz] w;  // non-zero elements in model matrix
  int<lower=0> v[nz];  // column indices for w
  int<lower=0> u[N+1];  // row-start indices for non-zero elements

  vector[N] y;  // scaled response

  int<lower=0> P;  // number of fixed effects
  int<lower=0> G;  // number of random effect groups
  int<lower=0> cindx[G,2];  // coefficient index for random effects
  int<lower=0> M_1;  // number of speaker members
  int<lower=0> Q_1;  // number of speaker effects per member
  int<lower=0> M_2;  // number of item members
  int<lower=0> Q_2;  // number of item effects per member

  // (hyper) priors
  real<lower=0> scale_beta;  // prior scale for betas
  real<lower=0> nu_beta;  // degrees of freedom for beta t-dist prior
  real<lower=0> sc_q0;  // prior scale for random intercept sds
  real<lower=0> sc_qs;  // prior scale for random slope sds
  real<lower=0> eta_q;  // shape for LKJ prior on random effects correlations
  real<lower=0> sc_res;  // prior scale for sd of the residuals
}

parameters {
  // all parameters sampled on unit scale or with cholesky factors
  // (as applicable) and reparameterized

  vector[P] beta_raw;

  matrix[Q_1,M_1] gamma_1_raw;
  vector<lower=0>[Q_1] sigma_1_raw;
  cholesky_factor_corr[Q_1] omega_1_raw;

  matrix[Q_2,M_2] gamma_2_raw;
  vector<lower=0>[Q_2] sigma_2_raw;
  cholesky_factor_corr[Q_2] omega_2_raw;

  real<lower=0> sigma_res_raw;
}

transformed parameters {
```

```
  vector<lower=0>[Q_1] sigma_1;  // sd in the speaker effects
  vector<lower=0>[Q_2] sigma_2;  // sd in the item effects
  real<lower=0> sigma_res;  // sd of the residuals

  vector[K] coef;  // all coefficients
  vector[N] y_hat;  // fitted values

  coef[1:P] = scale_beta * beta_raw;

  sigma_1[1] = sc_q0 * sigma_1_raw[1];
  sigma_1[2:Q_1] = sc_qs * sigma_1_raw[2:Q_1];
  coef[cindx[1,1]:cindx[1,2]]
    = to_vector(rep_matrix(sigma_1,M_1)
      .* (omega_1_raw * gamma_1_raw));

  sigma_2[1] = sc_q0 * sigma_2_raw[1];
  sigma_2[2:Q_2] = sc_qs * sigma_2_raw[2:Q_2];
  coef[cindx[2,1]:cindx[2,2]]
    = to_vector(rep_matrix(sigma_2,M_2)
      .* (omega_2_raw * gamma_2_raw));

  sigma_res = sc_res * sigma_res_raw;

  y_hat = csr_matrix_times_vector(N,K,w,v,u,coef);
}

model {
  beta_raw ~ student_t(nu_beta,0,1);

  to_vector(gamma_1_raw) ~ normal(0,1);
  sigma_1_raw ~ normal(0,1);
  omega_1_raw ~ lkj_corr_cholesky(eta_q);

  to_vector(gamma_2_raw) ~ normal(0,1);
  sigma_2_raw ~ normal(0,1);
  omega_2_raw ~ lkj_corr_cholesky(eta_q);

  sigma_res_raw ~ normal(0,1);
  y ~ normal(y_hat,sigma_res);
}

generated quantities {
  vector[N] log_lik;  // log-likelihod
  vector[P] beta;  // fixed effects
  matrix[M_1,Q_1] gamma_1;  // speaker effects
  matrix[Q_1,Q_1] omega_1;  // correlation in the speaker effects
  matrix[M_2,Q_2] gamma_2;  // item effects
  matrix[Q_2,Q_2] omega_2;  // correlation in the item effects

  for(n in 1:N) log_lik[n] = normal_lpdf(y[n] | y_hat[n],sigma_res);
  beta = coef[1:P];
  gamma_1 = vec_to_mat_by_row(M_1,Q_1,coef[cindx[1,1]:cindx[1,2]]);
  omega_1 = tcrossprod(omega_1_raw);
  gamma_2 = vec_to_mat_by_row(M_2,Q_2,coef[cindx[2,1]:cindx[2,2]]);
  omega_2 = tcrossprod(omega_2_raw);
}
```